

Funktionaalisesti tuntemattomien *BRCA1*- ja *BRCA2*-varianttien tarkastelu *in silico* -menetelmillä ja *ANO7*-varianttien alleelisen epätasapainon tutkiminen droplet digital PCR -menetelmällä

Niko Hautanen

Pro gradu –tutkielma

Turun yliopisto
Biologian laitos

TIIVISTELMÄ

TURUN YLIOPISTO

Luonnontieteiden ja tekniikan tiedekunta

NIKO HAUTANEN: Funktionaalisesti tuntemattomien *BRCA1*- ja *BRCA2*-varianttien tarkastelu *in silico* -menetelmillä ja *ANO7*-varianttien alleelisen epätasapainon tutkiminen droplet digital PCR-menetelmällä

Pro gradu -tutkielma, 63 s, 24 liites.

Biologia

12/2020

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turunitin Originality Check -järjestelmällä.

Tutkimuksen tarkoituksena oli tarkastella Turun yliopistollisen keskussairaalan Perinnöllisyyslääketieteen poliklinikan potilailta sekvensoinnin perusteella löydettyjä tuntemattoman merkityksen varianteiksi, eli VUS-variantteiksi, kutsuttuja *BRCA1*- ja *BRCA2*-geenien emäsvariantteja. Tutkittavia *BRCA1*-variantteja oli neljä ja *BRCA2*-variantteja 18. Alun perin tutkimuksen tarkoitus oli tutkia näiden varianttien alleelista epätasapainoa droplet digital PCR- eli ddPCR -menetelmällä, mutta huonon näytesaatavuuden takia tutkimus rajattiin vain varianttien tarkasteluun *in silico* -menetelmillä. Varianttien tarkastelu suoritettiin erilaisilla tietokoneohjelmilla, joihin lukeutuivat Alamut, Human Splicing Finder, RBPmap ja RegRNA. Tietokoneanalyysien tuottamaa tietoa tarkasteltiin jokaisen variantin osalta, ja tämän perusteella varianttien joukosta löytyi kolme erityisen mielenkiintoista varianttia, jotka tietokoneanalyysien perusteella voisivat sopia hyvin jatkotutkimuksiin, ohjelmien havaitsemien geenin toiminnan kannalta mahdollisesti haitallisten ominaisuuksien perusteella. Nämä variantit olivat *BRCA2* c.6821G>T c.7712A>G, ja c.9872C>G.

Tämän lisäksi tutkimuksen tarkoituksena oli tutkia *ANO7*-geenin eturauhassyöpäriskiä lisäävän rs77559646-varianttia kantavien potilaiden cDNA-näytteiden alleelista epätasapainoa droplet digital PCR -menetelmällä. Tutkimukseen valittiin kymmenen näytettä ja kolme yhden nukleotidin polymorfian (SNP) emästä RNA-tason alleelisen epätasapainon määrittämistä varten. Tarkoituksena oli vertailla tällä menetelmällä saatuja tuloksia jo olemassa oleviin sekvensointituloksiin, ja tarkastella menetelmän soveltuvuutta alleelisen epätasapainon tutkimiseen RNA-tasolla. Tulosten perusteella ddPCR pystyi helposti havaitsemaan toisen alleelin puuttumisen homotsygoottisista näytteistä, mutta heterotsygoottisten näytteiden osalta menetelmän tarkkuus alleelisen epätasapainon määrittämisessä vaihteli tutkittavasta SNP:stä ja näytteestä toiseen. Joissain näytteissä sekvensointi ja ddPCR-tulokset tukivat toisiaan, mutta toisissa näytteissä menetelmillä saadut tulokset erosivat toisistaan merkittävästi. Tilanteessa, jossa sekvensoinnilla saatu tieto alleelien välisestä suhteesta ei ollut kattavaa, ddPCR pystyi todennäköisesti määrittämään alleelien välisen suhteen tarkemmin kuin sekvensointidataan perustuva tulkinta.

Avainsanat:

VUS-variantti, *in silico* -menetelmät, *BRCA1*, *BRCA2*, *ANO7*, alleelinen epätasapaino, droplet digital PCR, ddPCR

Sisällysluettelo

1	Johdanto	1
1.1	Rinta- ja eturauhassyöpä	1
1.2	Emäsvarianttien tarkastelu ja tulkinta	3
1.3	Mutaatioiden vaikutusten ennakoarviointi <i>in silico</i> -menetelmillä	6
1.4	Alleelinen epätasapaino.....	8
1.5	Droplet Digital PCR.....	9
1.6	Tutkimuksen tarkoitus	10
2	Aineisto ja menetelmät	11
2.1	Koehenkilöt	11
2.2	<i>BRCA</i> -näytteiden RNA-eristys ja cDNA-synteesi	11
2.3	<i>in silico</i> -analyysi Alamut® Visual-ohjelmalla	11
2.4	<i>in silico</i> -analyysi Human Splicing Finder -ohjelmalla	11
2.5	<i>in silico</i> -analyysi RBPmap-ohjelmalla	13
2.6	<i>in silico</i> -analyysi RegRNA-ohjelmalla.....	15
2.7	Alukkeiden ja koettimien suunnittelu.....	16
2.8	<i>ANO7</i> alukkeiden ja koettimien toimivuuden testaus	20
2.9	<i>ANO7</i> -näytteiden analysointi.....	28
3	Tulokset	30
3.1	Alamut® Visual -analyysin tulokset	30
3.2	Human Splicing Finder -analyysin tulokset	32
3.3	RBPmap -analyysin tulokset.....	33
3.4	RegRNA 2.0 -analyysin tulokset	35
3.5	<i>ANO7</i> ddPCR-mittausten tulokset.....	37
4	Pohdinta ja tulosten tarkastelu	39
4.1	<i>BRCA1</i> ja <i>BRCA2</i> -tutkimuksen ongelmakohdat.....	39
4.2	<i>BRCA1</i> ja <i>BRCA2</i> Alamut tuloksien tulkinta ja johtopäätökset	40
4.3	Mahdollisten <i>BRCA1</i> - ja <i>BRCA2</i> -varianttien silmukointimuutosten tarkastelu <i>in silico</i> -analyysien tuloksien perusteella.....	42
4.3.1	Silmukointiin tarvittavat sekvenssit RNA juosteessa	42
4.3.2	<i>BRCA1</i> - ja <i>BRCA2</i> -varianttien silmukointimuutosten tarkastelu.....	44
4.4	Muiden <i>BRCA1</i> ja <i>BRCA2</i> <i>in silico</i> -analyysien tulosten tarkastelu.....	46
4.4.1	Havaitut ribosomin sitoutumiskohdat. transkriptiotekijät, ESE-sekvenssit ja muut muutokset	46
4.4.2	Havaitut mikro-RNA ja RNA-editointisekvenssit	47
4.5	Yhteenveto <i>BRCA</i> -varianttien analyysistä, sekä esimerkkejä mielenkiintoisista varianteista	49

4.6	<i>ANO7</i> ddPCR-tulosten ja sekvenssidatan vertailu	51
4.6.1	Erot ja yhtäläisyydet ddPCR-mittauksien ja sekvensointidatan välillä	51
4.6.2	Yhteenveto ddPCR-menetelmän toimivuudesta alleelisen epätasapainon määrittämisessä.....	55
5	Kiitokset.....	57
6	Lähteet	57
7	Liitteet.....	64

1 Johdanto

1.1 Rinta- ja eturauhassyöpä

Rintasyöpä on suomalaisnaisten yleisin syöpätyyppi, johon sairastui Suomen syöpärekisterin (2020) mukaan vuonna 2018 4934 naista. Vuosittain rintasyöpään sairastuu Suomessa noin 5000 naista. Suomen rintasyöpäryhmän diagnostiikka- ja hoitosuosituksen mukaan (Joensuu 2019) syöpäriskiä vaikuttavat monet ympäristötekijät, kuten lapsettomuus, vaihdevuosisoireisiin käytetyt hormonihoidot, runsas alkoholin käyttö, tupakointi ja ylipaino, mutta useimmiten suoranaista syytä potilaan syövän ilmaantumiselle ei pystytä määrittämään. On arvioitu, että noin 5–10 % rintasyövistä voidaan selittää perinnöllisellä alttiudella.

Tunnetuimmat perinnölliseen alttiuteen liittyvät geenit ovat *BRCA1* ja *BRCA2*, joiden yhteys rintasyöpään todettiin jo yli kaksikymmentä vuotta sitten. Miki ym. (1994) toteivat *BRCA1*-mutaatioiden yhteyden rintasyöpäriskiä jo vuonna 1994, ja hieman myöhemmin Wooster ym. (1995) työryhmineen löysi yhteyden *BRCA2*-geenin ja rintasyöpäriskin välillä. *BRCA1* ja *BRCA2* ovat geenejä, jotka säätelevät DNA:n korjausmekanismeja ja transkriptiota DNA-vaurioiden yhteydessä. Yoshidan ja Mikin (2004) yhteenvedossa mainitaan, että geenien toiminta on yhteydessä myös kromosomaalisen stabiiliteetin ylläpitämiseen, sekä solusykliin ja apoptoosiin. Geenejä kutsutaan tuumorisuppressori- eli kasvurajoitegeneiksi, ja jos näissä geneissä esiintyy jokin haitallinen mutaatio, se voi aiheuttaa DNA:n korjausmekanismien häiriöitä, ja johtaa lisääntyneeseen spontaanien mutaatioiden ilmenemiseen soluissa ja mahdollisesti syövän kehittymiseen.

BRCA-mutaatioiden vaikutus rintasyöpäriskiä vaihtelee eri tutkimuksissa. Sining Chenin ja Giovanni Parmigianin (2007) meta-analyysin mukaan *BRCA1*-mutaatioiden kantajilla on 57 % ja *BRCA2*-mutaatioiden kantajilla 49 % riski sairastua rintasyöpään 70:een ikävuoteen mennessä, kun taas Kuchenbaeckerin ym. (2017) tutkimuksessa johon osallistui 3886 naista, syöpäriski 80:een ikävuoteen mennessä oli *BRCA1*-mutaatioiden kantajilla 72 % ja *BRCA2*-mutaatioiden kantajilla 69 %. Tiedetään myös, että eri alueilla on hieman erilainen mutaatioprofiili ja rintasyövän sairastumisriski. Rin-

tasyöpäriskin lisäksi *BRCAl*- ja *BRCA2*-mutaatioiden kantajilla on myös lisääntynyt riski sairastua munasarjasyöpään, jonka suuruus on noin 20—40 %. Pelkästään *BRCAl*-geenistä on vuosien aikana Petrucellin ym. (1998) mukaan löydetty yli 1800 patogeenistä varianttia.

Suomen syöpäjärjestöjen mukaan (2020 b) eturauhasyöpä on puolestaan miesten yleisin syöpä, johon sairastuu vuosittain Suomessa noin 5000 henkilöä. Suomen syöpärekisterin (2020) mukaan vuonna 2018 syöpään sairastui 5016 miestä. Sairauden tarkkaa aiheuttajaa ei tiedetä, mutta on havaittu, että syöpäriskiä lisäävät runsas rasvan ja lihan syöminen, liikunnan puute sekä tupakointi. Epidemiologisten analyysien perusteella on perinnöllisen alttiuden esitetty olevan syynä noin puolessa syöpätapauksia. Esimerkiksi Hjelmberg ym. (2014) havaitsivat pohjoismaalaisessa kaksostutkimuksessa, että syöpäriskistä noin 57 % voidaan mahdollisesti selittää perinnöllisillä tekijöillä. Kankuri-Tammilehdon, Vihisen ja Schleutkerin (2019) katsauksen mukaan näitä perinnöllisiä tekijöitä tunnetaan toistaiseksi vain vähän, mutta esimerkiksi *HOXB13*-geenin virheet ovat osoittautuneet syöpäriskiä kasvattavaksi tekijäksi erityisesti pohjoismaalaisessa väestössä. Myös *ATM*- ja *CHEK2*-geenien mutaatiot näyttävät lisäävän syöpäriskiä.

Yksi eturauhasyöpäriskiinkin vaikuttava tekijä näyttäisi olevan mutaatiot *ANO7*-geenissä. *ANO7* on erityisesti eturauhasessa esiintyvä anoktamiini-perheeseen kuuluva kalsiumin aktivoima transmembraaninen proteiini, jolla on fosfolipidi-skramblaasi -aktiivisuutta (phospholipid scramblase). Proteiinin tarkkaa merkitystä solun toiminnalle ei vielä tunneta. Kaikkonen ym. (2018) löysivät geenistä kaksi mutaatiota, jotka näyttävät merkittävästi lisäävän syöpäriskiä. Variantti rs77559646 on silmukointia häiritsevä variantti, joka tutkimuksen perusteella on yhteydessä lisääntyneeseen syöpäriskiin, sekä syövän aggressiivisuuteen. Tämä variantti aiheuttaa neljännen eksonin jälkeisen intronisen fragmentin säilymisen sekvenssissä. Variantti rs148609049 puolestaan on 5'-UTR alueella sijaitseva variantti, joka ei tutkimuksen mukaan lisää syöpäriskiä, mutta sillä näyttää olevan yhteys aikaisempaan kuolleisuuteen syöpään sairastuneilla. Tämän lisäksi tutkimus osoitti *ANO7*-geenin ekspression olevan korkeampaa syöpäkudoksessa, kuin normaalissa eturauhaskudoksessa, ja tämä korkeampi ekspressio oli yhteydessä korkeampaan potilaiden kuolleisuuteen.

1.2 Emäsvarianttien tarkastelu ja tulkinta

Jos henkilö epäilee kuuluvansa sukuun, jossa esiintyy perinnöllistä syöpää, hänellä on mahdollisuus hakeutua perinnöllisyysneuvontaan. Potilaan genomia tarkastelemalla on mahdollista selvittää, kantaako hän joitain vaarallista patogeenistä mutaatiota tai harmittonta varianttia geneeissä, joiden on osoitettu olevan yhteydessä syöpäriskiin. Näiden tutkimusten ongelmana on kuitenkin se, että esimerkiksi *BRCA*-geneistä on löydetty jopa tuhansia variantteja, ja edelleen löydetään uusia ennestään tuntemattomia variantteja. Näitä vaikutuksiltaan tuntemattomia variantteja kutsutaan tuntemattoman merkityksen varianteiksi, eli VUSeiksi (variant of uncertain/unknown significance). Jos potilas saa tutkimuksesta vastaukseksi, että hän kantaa varianttia, jonka merkitystä ei tarkkaan tiedetä, tämä vaikeuttaa perinnöllisyysneuvontaa, ja voi lisätä potilaan huolta, vaikka perinnöllisyysneuvonnassa selitetäänkin perusteellisesti se, mitä muutoksesta nykyään tiedetään. Asiaan joudutaan palaamaan muutaman vuoden kuluttua, jolloin on mahdollista tarkistaa, onko variantin merkityksestä tullut uutta tietoa. Ennaltaehkäiseviin toimenpiteisiin, kuten esimerkiksi rintojen poistoon, ei voi ryhtyä epävarman variantin havaitsemisen perusteella, koska VUS-muutos saattaa ollakin hyvänlaatuinen variantti, joka ei ollenkaan selitä potilaan sairastumisriskiä. Näistä syistä on tärkeää tutkia ja selvittää VUSien vaikutuksia geeneihin ja syöpäriskiin. Alati lisääntyvä tutkimustieto johtaa ajan kuluessa VUS-varianttien uudelleen luokitteluun haitallisiksi tai harmittomiksi, mikä helpottaa potilaiden hoidon suunnittelua, kuten Son ym. (2019) tutkimuksessa, jossa 75:stä *BRCA1* tai *BRCA2* VUS-varianttia kantavasta potilaasta kaksi määritettiin uudestaan päivitettyjen tietojen ja suositusten perusteella haitallisen variantin kantajiksi ja kolmekymmentä todennäköisesti harmittoman variantin kantajiksi. Jos variantin luokitus muuttuu patogeeniseksi eli mutaatioksi, niin tietoa voidaan käyttää sukulaisten syöpäriskin arvioinnissa: ne sukulaiset, joilla ei ole mutaatiota vapautuvat tiheemmästä seurannasta.

The American College Of Medical Genetics And Genomics (ACMG), Association for Molecular Pathology (AMP) ja College of American Pathologists ovat yhdessä antaneet ohjeistuksia varianttien patogeenisyyden asteen määrittämiseen. Jotta variantti voidaan luokitella luotettavasti patogeeniseksi tai harmittomaksi, on variantin vaikutuksesta oltava luotettavaa tietoa useasta lähteestä. Eri kriteerien perusteella varianttien patogeenisyys voidaan luokitella Richardsin ym. (2015) suositusten mukaan viiteen eri kategori-

aan: patogeeninen (pathogenic), todennäköisesti patogeeninen (likely pathogenic), epävarma (uncertain significance), todennäköisesti harmiton (likely benign) ja harmiton (benign). Patogeenisyydestä epäilyn herättävät esimerkiksi variantit, jotka aiheuttavat geenin transkriptiin ennenaikaisen lopetuskodonin, sekä emäsmuutokset, jotka muuttavat tietyn aminohapon samaksi, kuin jokin aikaisemmin todettu patogeeninen variantti. Myös proteiinin toiminnalle kriittisille domeeneille ilmaantuneet variantit voivat viitata patogeenisyyteen, varsinkin jos näillä alueilla ei ole aikaisemmin todettu harmittomia emäsmuutoksia. Tämän lisäksi siittiössä, munasolussa tai alkiokehityksessä voi tapahtua *de novo* -mutaation syntyminen, jolloin perinnöllistä sairautta sairastavan henkilön suvussa ei ole ennestään samaa sairautta, mutta tällaisen muutoksen osoittaminen perinnölliseksi on haasteellista. Kaikenlaisten emäsvarianttien kohdalla on tosin tärkeä myös tarkastella variantin mahdollisia vaikutuksia silmukointiin.

Geenituotteen tai transkriptin toimintaa voidaan myös tutkia *in vitro*- ja *in vivo* -menetelmillä, ja jos validoitu menetelmä toteaa variantin olevan proteiinille toiminnallisesti haitallinen, tai että variantti häiritsee transkriptin silmukointia, on tämä myös vahvaa todistusaineistoa variantin patogeenisyydelle, olettaen, että tutkimukseen käytetyt menetelmät on validoitu tähän käyttötarkoitukseen. Populaatiotasolla patogeenisyyteen voi viitata se, että variantti esiintyy sairailta ihmisillä huomattavasti useammin kuin kontrollipopulaatiossa. Tämän menetelmän ongelmana on kuitenkin se, että erittäin harvinaisten varianttien tutkimuksissa tilastollinen merkittävyys usein laskee niin paljon, että tuloksien tulkitseminen muuttuu kyseenalaiseksi. Myös erilaisten tietokoneohjelmien tulkintoja voidaan pitää merkinä patogeenisyydestä, jos usea eri ohjelma toteaa variantin haitalliseksi. Tosin tietokoneanalyysit yksin eivät koskaan riitä todistukseksi patogeenisyydestä, vaan tarvitsevat rinnalleen myös muuta todistusaineistoa.

Richardisin ym. (2015) suosituksista löytyy monia muita kriteerejä patogeenisyyden arvioimiseksi. Kirjoittajat myös painottavat, että tekstissä mainitut kriteerit eivät välttämättä sovi jokaisen geenin tutkimiseen. Tämä on erityisen tärkeää huomioida varsinkin silloin, kun todistusaineistona on ainoastaan tiettyyn sairauteen tai geenin erikoistuneen asiantuntijaryhmän omakohtainen tulkinta tietyn variantin patogeenisyydestä.

Richardisin ym. (2015) ohjeet ja niiden noudattaminen, sekä päivittäminen ovat tärkeitä, sillä standardoidut patogeenisyyteen viittaavat kriteerit ja sanasto helpottavat kliinisten

laboratorioiden ja muiden terveydenhuollon ammattilaisten välistä yhteistyötä ja -ymmärrystä, jos potilaan tutkimuksen osaksi kuuluu emäsvarianttien tarkastelu. Ymmärrettävä asiantuntijoiden laatima raportti potilaan kantamasta variantista helpottaa kaikkien osapuolien yhteistyötä, sekä potilaan hoidon ja seurannan suunnittelua.

Terveydenhuollon ammattilaisilla ja tutkijoilla on käytettävissä myös monia eri tietokantoja, joista löytyy ajankohtaista tietoa yksittäisten varianttien vaikutuksista. Yksi tunnetuimmista tällaisista tietokannoista on Landrumin ym. (2017) kuvaama ja NCBI:n ylläpitämä ClinVar. Tämän tietokannan tarkoitus on koota yhteen tietoa varianteista eri lähteistä, ja saatavilla olevan tiedon määrän ja uskottavuuden perusteella määritellä variantin mahdollinen vaikutus patogeeniseksi, harmittomaksi tai tuntemattomaksi. Kaikki tietyn variantin vaikutukseen viittaavien tutkimusten ja datasettien lähteet löytyvät variantin sivulta riippumatta siitä, ovatko tutkimusten johtopäätökset variantin vaikutuksista yhdenmukaisia vai ristiriitaisia. Tämän lisäksi jokaiselle variantille on tietokannassa määritetty oma variantin vaikutuksen johtopäätöksen luotettavuutta edustava ”review status” -pistemäärä, jossa suurempi luku viittaa luotettavampaan tulkintaan variantin vaikutuksesta, johtuen tiukemmista ja kattavammista arviointiperusteista. Nolla tähteä viittaa siihen, että variantti on vain havaittu ja sen vaikutuksen tulkinnalle ei ole todisteita, yksi tähti tarkoittaa, että variantti ja sen vaikutus on kuvailtu vain kerran, tai että useampi lähde on kuvaillut variantin vaikutuksen ristiriitaisesti. Jos variantin ”review status” -pistemäärä ylittää kolme tähteä, on variantin vaikutuksen tulkinta todennäköisesti luotettava, sillä tämän pistemäärän voi saada vain asiantuntijapaneelin arvioima variantti. ClinVarin arviointijärjestelmä ja pisteytys helpottaa eri varianttien vaikutusten tarkastelemista, ja samalla kokoaa yhteen varianttiin viittaavaa kirjallisuutta, sekä päivittää tietoja variantista uusien tutkimuslöydösten ja julkaisujen myötä.

Yleisesti variantteja arvioivien tietokantojen lisäksi on olemassa monia spesifisempiä tietokantoja, kuten esimerkiksi *BRCA*-geeneihin ja rintasyöpään keskittyvä ENIGMA consortium (2020), jonka asiantuntijaryhmä arvioi ja luokittelee aktiivisesti VUS-variantteja. Tämän tietokannan luokittelut ja tulkinnat ovat erityisen tärkeitä rintasyövän ja syöpäriskin tutkimuksessa.

1.3 Mutaatioiden vaikutusten ennakoarviointi *in silico* -menetelmillä

Eri variantteja kuvailevien tietokantojen tarkastelun lisäksi, on ennen varsinaisen tutkimustyön aloittamista usein myös hyvä tarkastella erilaisten emäsmuutosten vaikutuksia teoreettiselta pohjalta erilaisilla tietokoneohjelmilla. Tällä tavalla voidaan suuresta määrästä erilaisia emäsmuutoksia rajata tutkimukseen ne mutaatiot, jotka vaikuttavat mielenkiintoisimmilta, tai eroavat merkittävästi muista vaihtoehdoista. Mielenkiintoisia emäsmuutoksia ovat esimerkiksi sellaiset, jotka muuttavat jonkin proteiinin toiminnalle kriittisen aminohapon kemiallisesti erilaiseksi aminohapoksi, tai muutokset, jotka muuttavat jotenkin mRNA:n silmukointia, johtaen erilaiseen proteiinituotteeseen.

Tietynlaisten emäsmuutosten, kuten ennen aikaisten lopetuskodonien, lukuraamia muuttavien indel-mutaatioiden ja suurten deleetioiden vaikutukset ovat tyypillisesti haitallisia geenin toiminnalle. Yksittäisten emäksien muutokset, intronisten mutaatioiden ja geenin säätelyalueilla sijaitsevien mutaatioiden vaikutuksia on puolestaan vaikeampi ennakoida. Tällaiset mutaatiot saattavat muuttaa jonkin proteiinin toiminnalle kriittisen aminohapon toiseksi, häiritä RNA:n silmukointiprosessia, tai muuttaa jonkin DNA:han tai RNA:han sitoutuvan proteiinin tunnistussekvenssin toimimattomaksi. Toisaalta monet tällaisista mutaatioista voivat olla täysin harmittomia.

Mutaatioiden vaikutusten ennakoointia varten on kehitetty monia erilaisia tietokoneohjelmia, jotka tarkastelevat mutaatioiden mahdollisia vaikutuksia eri menetelmillä ja eri lähtökohdista. Yksi kliinisen tutkimuksen käytetyimpiä ohjelmia on Alamut® Visual, joka yhdistää monen erilaisen ennusteohjelman ja julkisten tietokantojen havainnot yhdeksi helppolukuiseksi tulokseksi. Alamut-ohjelmaan kuuluu neljä erilaista alaohjelmaa: Align-GVGD, PolyPhen-2, SIFT ja MutationTaster2. Alamutiin kuuluva Align-GVGD-ohjelma (Tavtigian ym. 2006) ennakoii missense-mutaatioiden haitallisuutta proteiinin toiminnalle aminohappotasolla, tarkastelemalla mutaation seurauksena muuttuneen aminohapon kemiallisten ominaisuuksien eroavaisuutta proteiinissa normaalisti ilmenevään aminohappoon verrattuna. PolyPhen-2 (Adzhubei ym. 2010) tarkastelee mutaatioiden vaikutuksia proteiinin kolmiulotteiseen rakenteeseen, sekä tutkii, miten yleistä aminohapon muuttuminen toiseksi tietyssä kohdassa aminohapposekvenssiä on homologisissa proteiineissa. Samanlaisilla menetelmillä toimii myös ohjelma nimeltä SIFT (Sim ym. 2012). MutationTaster2 (Schwarz ym. 2014) tarkastelee proteiinin rakenteen lisäksi mutatoituneen emäksen fylogeneettistä konservoitumista ja mutaation

aiheuttamia mahdollisia vaikutuksia silmukointiin, sekä geenin säätelyalueisiin. Näiden lisäksi Alamut-ohjelma tarkastelee myös muuttuneen nukleotidin ja aminohapon fylogeneettistä konservoitumista.

Internetistä löytyy monia muita erilaisia ilmaisia sekä maksullisia ennusteohjelmia, joita tutkijat voivat käyttää projekteissaan. Esimerkiksi Human Splicing Finder (Desemet ym. 2009) käyttää monia erilaisia algoritmeja löytääkseen silmukointiin tarvittavien sekvenssien häiriöitä. RBPmap (Paz ym. 2014) puolestaan etsii nukleiinihapposekvenssistä erilaisia RNA:ta sitovien proteiinien sitoutumiskohtia, ja vertailemalla mutaatiosekvenssiä normaaliin sekvenssiin, voidaan tarkastella mutaatioiden aiheuttamia mahdollisia muutoksia. RegRNA -ohjelma (Huang ym. 2006) tunnistaa mRNA-sekvenssistä erilaisia säätelyalueita, kuten silmukointiin tarvittavia sekvenssejä, miRNA kohdesekvenssejä ja transkriptiosäätelyyn tarvittavia sekvenssejä. Vertaamalla mutanttisekvenssiä normaalisekvenssiin voi ohjelmalla tarkastella mutaatioiden mahdollisia vaikutuksia.

Vaikka eri ennusteohjelmien tarkkuus paranee alati ohjelmien kehittyessä, emme pysty kuitenkaan korvaamaan oikeaa mutaatioiden tutkimusta näillä ohjelmilla. Ernst ym. (2018) selvittivät Alamut-ohjelman tarkkuutta *BRCA1* ja *BRCA2* missense-varianttien patogeenisyyden toteamiseen, ja kävi ilmi, että ohjelma ennusti runsaasti väärää positiivisia tuloksia, joiden mukaan harmiton variantti olisi ohjelman mukaan patogeeninen. Toisaalta tutkimuksessa kävi myös ilmi, että SIFT ja MutationTaster2 pystyivät sulkemaan pois patogeeniset variantit hyvin, eli nämä ohjelmat tulkitsivat patogeeniset variantit erittäin harvoin harmittomiksi.

Ennusteohjelmat tarjoavat hyvän tavan keskittää tutkimuksia todennäköisimmin merkittäviin emäsmuutoksiin, mutta lopullinen tulos mutaatioiden vaikutuksista on silti todettava muillakin tutkimusmenetelmillä. Geenit, geenituotteet ja niiden säätelyjärjestelmät ovat monimutkaisia verkostoja, joista emme vielä tiedä kaikkea, ja tämän takia ennusteohjelmat eivät mitä todennäköisimmin tule koskaan täysin korvaamaan perinteisiä tutkimusmenetelmiä.

1.4 Alleelinen epätasapaino

Alleelinen epätasapaino (allelic imbalance), josta käytetään myös nimitystä alleelispesifinen ekspressio (allele specific expression) on ilmiö, jossa ihmisen kahdesta vanhemmilta peritystä saman geenin alleeleista toinen ekspressoituu enemmän tai vähemmän kuin toinen. Bucklandin (2004) katsauksen mukaan, tämä ilmiö näyttää olevan yleinen ihmisgenomissa, ja johtuu usein muutoksista geenien säätely- ja promoottorialueilla, sekä epigeneettisistä muutoksista. Myös erilaiset mutaatiot itse geenissä näyttävät vaikuttavaan alleeliseen epätasapainoon. Jos toinen alleeli on täysin poissa genomista deleetion takia, geenin ekspressio odotetusti laskee. Mutaatiot, jotka aiheuttavat ennenaikaisia lopetuskodoneja, tai lukukehyksen muutoksia, näyttävät paikasta riippuen Rheen ym. (2017) mukaan laskevan geeniekspressiota paljon useammin, kuin silmukointiin vaikuttavat mutaatiot, missense mutaatiot, tai hiljaiset mutaatiot. Tämän lisäksi Mizuno ja Okada (2019) ovat havainneet, että jotkin mutaatiot vaikuttavat alleelien väliin ekspressioon eri tavalla eri kudoksissa.

Alleelisen epätasapainon vaikutus syöpäriskiä on kiistanalainen, ja näyttää riippuvan syöpätyypistä. Eri tekniikoilla tehdyt tutkimukset vaikeuttavat tuloksien vertailua ja saattaa osaltaan johtaa ristiriitaisiin tuloksiin. Pyrosekvensointia käyttäen ovat Tomsic ym. (2010) ja Segui ym. (2011) esimerkiksi todenneet, ettei *TGFBR1*-geenin alleelista epätasapainoa esiinny paksusuolensyöpään sairastuneilla ihmisillä merkittävästi enemmän kuin terveillä ihmisillä. Akuutin myeloidisen leukemian (AML) tapauksessa taas sekvenssidataan perustuvan Batchan ym. (2019) tutkimuksen mukaan alleelista epätasapainoa esiintyy syöpää sairastavilla ihmisillä yhdeksässä geenissä, jotka ovat usein mutatoituneita sairastuneilla ihmisillä. Merkittävin ero normaali- ja mutanttialleelin välisessä ekspressiossa havaittiin *GATA2*-geenissä.

Chen ym. (2008) havaitsivat alleelispesifistä realiaikaista PCR-tekniikkaa käyttäen, että alleelista epätasapainoa *BRCA1*- ja *BRCA2*-geenien ekspressiossa esiintyy rintasyöpää sairastavilla ihmisillä, mutta ei terveillä henkilöillä. Samaa tekniikkaa hyödyntäen Shen ym. (2011) havaitsivat, että *BRCA1*-geenin alleelista epätasapainoa esiintyy munasarjasyöpään sairastuneilla henkilöillä enemmän kuin terveillä. *BRCA2*-geenin alleelista epätasapainoa ei tutkimuksessa havaittu syöpään sairastuneilla henkilöillä. Näiden havaintojen lisäksi Huijmsansin ym. (2015) mukaan, myös *HER2*- ja *TOPA2*-geenien alleelista

epätasapainoa näyttää esiintyvän varsinkin *HER2*-geeniä yliekspressoivissa sekä tripla-negatiivisissa rintasyöpäkasvaimissa, mutta ei normaalissa rintasyöpäkuksessa.

”SNaPshot quantitative primer extension assay” -tekniikkaa käyttämällä ovat Caux-Moncoutier ym. (2009) myös tutkineet alleelista epätasapainoa rintasyöpää sairastavilla henkilöillä, jotka eivät kantaneet mitään tunnettuja haitallisia mutaatioita, mutta kantoi-vat jotain VUS-varianttia. Kuudestatoista *BRCA1* VUS-varianttia kantavasta potilaasta kahdella havaittiin normaalin vaihtelun ulkopuolella oleva ero alleelien välisessä eks-pressiossa. 37:stä *BRCA2* VUS-varianttia kantavasta potilaasta kolmella havaittiin nor-maalista vaihtelusta poikkeava ero alleelisessa ekspressiossa. Uudessa pyrosekvensoin-tia hyödyntävässä tutkimuksessa taas Jamarid ym. (2017) eivät löytäneet yhteyttä *BRCA1*- tai *BRCA2*-geenin alleelisen epätasapainon ja rinta- ja munasarjasyöpäriskin välillä. Eri tutkimusmenetelmien ja tutkimusten ristiriitaisten tuloksien johdosta alleeli-sen epätasapainon vaikutus rinta- ja munasarjasyöpäriskiinkin on vielä toistaiseksi epävar-ma.

1.5 Droplet Digital PCR

Droplet digital PCR eli ddPCR, on kvantitatiivinen menetelmä, jolla pystytään tarkasti toteamaan ja mittaamaan näytteessä olevia nukleinihappoja EvaGreen värillä tai fluore-soivia-koettimia (esim. Taqman) hyödyntäen. Menetelmän tarkkuuden ja spesifisyyden saa aikaan PCR-reaktion jakaminen 20 000:een erilliseen reaktioon, muodostamalla reaktioseoksesta kahden nanolitrin kokoisia pisaroita, joista jokaisessa tapahtuu erilli-nen PCR-reaktio. Menetelmällä voidaan muun muassa havaita tiettyjä harvinaisia sek-venssejä suuresta määrästä DNA:ta, ja verrata kahden eri alleelin välisiä ekspressiota-soja. PCR-lämpösyklin jälkeen käytetyn väriaineen aiheuttama signaali todetaan näyt-teestä mittaamalla pisaroiden aiheuttama fluoresenssi. DdPCR-laitteisto määrittää fluo-resenssin perusteella jokaisen pisaran negatiiviseksi tai positiiviseksi, ja laskee tilastolli-sesti pisaroiden perusteella näytteessä olevan DNA:n määrän. Jos samassa reaktiossa on kaksi eri fluoroforeilla merkattua koetinta, voidaan mitata näiden kahden koettimen välinen suhde (Hindson ym. 2011).

DdPCR-tekniikkaa on käytetty syöpägeenien alleelisen epätasapainon tutkimiseen. De Smith ym. (2015) ovat tutkineet menetelmällä kuuden eri geenin syöpäriskiinkin liittyvien

SNP-emäksien (single nucleotide polymorphism) alleelisen epätasapainon yhteyttä lapsuusajan akuuttisen lymfaattiseen leukemiaan (ALL). Tutkimuksessa selvitettiin RNA-tason ekspression sijaan sitä, miten syöpäkudoksessa normaali- ja riskialleelit säilyivät syöpäsoluissa sairaudelle yleisten deleetioiden johdosta. Tutkimuksessa ei löydetty näyttöä sille, että syöpäkudoksen kehitys suosisi normaalin tai riskialleelin säilymistä kudoksessa. Maeda ym. (2020) ovat puolestaan tutkineet *H3F3A*-geenin K27M-mutaation aiheuttamaa alleelista epätasapainoa diffuusien keskiviivan gliomien aivo-kasvaimissa, ja he totesivat, että joissakin mutaatiota kantavista kasvaimista mutanttialleelia ilmenee enemmän kuin normaalia alleelia, johtaen mutanttialleelin korkeampaan ekspression ja syöpäpotilaiden suurempaan kuolleisuuteen.

1.6 Tutkimuksen tarkoitus

Yksi tutkimuksen alkuperäisistä tarkoituksista oli tutkia tiettyjen Turun yliopistollisen keskussairaalan (TYKS) Kliinisen genetiikan vastualueen, Perinnöllisyyslääketieteen poliklinikan potilailla esiintyvien *BRCA1* ja *BRCA2* VUS-varianttien mahdollista yhteyttä geenien alleeliseen epätasapainoon ddPCR-menetelmällä, ja samalla tarkastella kyseisen menetelmän soveltuvuutta tutkimukseen, jossa on pieni koehenkilömäärä. Ennen varsinaista ddPCR-analyysiä, variantit oli tarkoitus rajata ja niiden soveltuvuus tutkimukseen määrittää erilaisilla *in silico* -analyysillä. Tutkimukseen sopivien tarkasteltavia variantteja kantavien potilasnäytteiden vähäisen määrän vuoksi, tutkimus rajattiin varianttien *in silico* -analyysien tarkasteluun.

Tutkimuksen toinen tavoite on tutkia tiettyjen *ANO7*-varianttien alleelista epätasapainoa samalla ddPCR-menetelmällä. RNA-sekvensoinnilla on tarkasteltu Kaikkosen ym. (2018) löytämän syöpäriskiä lisäävän rs77559646-variantin kantajien *ANO7* ekspressiota ja myös sitä, ekspressoituvatko alleelit yhtä paljon, laskemalla muiden eksonisten SNP-emäksien alleelifrekvenssit. Tarkoituksena on tutkia näitä samoja eksonisia SNP-emäksiä ddPCR-menetelmällä, ja verrata saatuja alleelifrekvenssien tuloksia RNA-sekvensoinnin antamiin tuloksiin, ja näin tarkastella ddPCR-menetelmän soveltuvuutta alleelisen epätasapainon tutkimiseen.

2 Aineisto ja menetelmät

2.1 Koehenkilöt

BRCA1 ja *BRCA2* -tutkimuksen koehenkilöt valittiin Turun yliopistollisen keskussairaalan (TYKS) Perinnöllisyyslääketieteen poliklinikan potilaista, joilla oli sekvensoinnin perusteella todettu jokin *BRCA1* tai *BRCA2* VUS-variantti. Nämä variantit on listattu taulukkoon 1. *ANO7* -tutkimusta varten puolestaan analysoitiin 10 rs77559646-varianttia kantavan potilaan eturauhaskudoksen RNA:sta tuotettua cDNA-näytettä. Yksi näyte oli peräisin normaalista kudoksesta, loput syöpäkudoksesta. Käytetty cDNA oli tuotettu oligo(dT)- ja satunnaisalukemenetelmän (random primer mix) yhdistelmällä.

2.2 *BRCA*-näytteiden RNA-eristys ja cDNA-synteesi

BRCA-tutkimusta varten koehenkilöiltä kerättiin suostumuksella kokoverinäytteet ”PAXgene® Blood RNA” -putkiin ja RNA eristettiin ”PAXgene® Blood RNA” -kitillä valmistajan ohjeiden mukaan. RNA:n eristyksen jälkeen suoritettiin cDNA-synteesi ”iScript™ Advanced cDNA Synthesis Kit for RT-qPCR” -kitillä valmistajan ohjeiden mukaan. Nämä työvaiheet tehtiin, vaikka kyseisiä näytteitä ei lopulta analysoitukaan.

2.3 *in silico* -analyysi Alamut® Visual-ohjelmalla

Tutkimuksessa tarkasteltiin Alamut-ohjelmalla saatuja raportteja Varsinais-Suomen alueella todetuista *BRCA1*- ja *BRCA2*-VUS-varianteista. DNA-tutkimusnäytteitä saatiin 14:ltä rintasyöpägenitutkimukseen osallistuneelta potilaalta, jotka olivat saaneet aikaisemmin perinnöllisyysneuvontaa Turun yliopistollisen keskussairaalan Perinnöllisyyslääketieteen poliklinikalla.

2.4 *in silico* -analyysi Human Splicing Finder -ohjelmalla

Tähän analyysiin käytetty ohjelma oli Human Splicing Finder versio 3.1 (HSF 3.1). Kyseinen ohjelmistoversio ei ole enää vapaasti saatavilla, vaan uusi päivitetty versio muuttui tutkimuksen aikana maksulliseksi. Analyysiä varten haettiin *BRCA* VUS-variantin rs-numeron perusteella VUSia ympäröivä geenisekvenssi ”Ensembl Genome

Browser” -tietokannasta. Rs-numeron perusteella haetun variantin sivuilta valittiin ”flanking sequence” -näkyvä, jossa näkyi VUS-variantti punaisella ja eksoninen sekvenssi keltaisella. VUS-varianttia ympäröivästä sekvenssistä kopioitiin sekvenssi, johon sisältyi eksoni ja 150–200 emästä intronista sekvenssiä. Jos VUSin sisältämä eksoni oli niin suuri, että eksonin ja intronin sekvenssin pituudeksi tuli yli 2500 emästä, VUSia ympäröivä sekvenssi rajattiin siten, että intronista sekvenssiä oli sekvenssissä mukana vain 150–200 emästä VUS-emästä lähempänä olevasta intronista. Rajaus 2500 emäseen perustui siihen, että Human Splicing Finder ei pysty käsittelemään tätä pidempiä sekvenssejä.

HSF-analyysityypiksi valittiin ”analyze a sequence” ja sekvenssi valittiin kopioimalla ohjelmaan oma sekvenssi. Ohjelman näkymään kopioitiin rajattu sekvenssi, jossa VUS-emäs määritettiin villityypin emäkseksi, ja sekvenssin eksoni rajattiin. ”Proceed to analysis” -toiminnon jälkeen valittiin ”perform a quick mutation”, jossa VUS-positiossa oleva villityypin emäs vaihdettiin mutanttiemäkseksi. Analyysin perusteella tulokset jaettiin kolmeen kategoriaan: uudet luovuttajasekvenssit (new donor sites), uudet eksoniset hiljentäjä- eli ESS-sekvenssit (new ESS sites) sekä rikkoutuneet eksoniset vahventaja- eli ESE-sekvenssit (broken ESE sites). Kuvassa 1 on esimerkki yhden variantin Human Splicing Finder analyysin tuloksien yhteenvedosta.

Reference sequence You analyze your own sequence.

```

1 aatgacaaaa agttactaaa tcactgccat cacacggttt atacagatgt caatgatgta ttgattatag aggttttcta ctgttgctgc atcttatttt
101 tatttgttta catgtctttt cttatttttag tgtccttaaa aggttgataa tcacttgctg agtgtgtttc tcaaaccaatt taatttcagG AGCCTACAAG
201 AAAGTACGAG ATTTAGTCAA CTTGTTGAAG AGCTATTGAA AATCATTGTG GCTTTTCAGC TTGACACAGG TTTGGAGTgt aagtgtttaa tatccaaga
301 atgacactca agtgcgtgcc atgaaaactc aggaagtgtt cacaattact ttctatgacg tggtgataag accttttagt ctagggttaa tttagttctg
401 tatctgtaat ctatttttaa aaaattactc ccaactgtct cacaccttat tttatcaatc gtaagtgca catttttcac

```

Total sequence length: 480 nucleotides

Mutant sequence Substitution of position 207 (C>T)

```

1 aatgacaaaa agttactaaa tcactgccat cacacggttt atacagatgt caatgatgta ttgattatag aggttttcta ctgttgctgc atcttatttt
101 tatttgttta catgtctttt cttatttttag tgtccttaaa aggttgataa tcacttgctg agtgtgtttc tcaaaccaatt taatttcagG AGCCTACAAG
201 AAAGTATGAG ATTTAGTCAA CTTGTTGAAG AGCTATTGAA AATCATTGTG GCTTTTCAGC TTGACACAGG TTTGGAGTgt aagtgtttaa tatccaaga
301 atgacactca agtgcgtgcc atgaaaactc aggaagtgtt cacaattact ttctatgacg tggtgataag accttttagt ctagggttaa tttagttctg
401 tatctgtaat ctatttttaa aaaattactc ccaactgtct cacaccttat tttatcaatc gtaagtgca catttttcac

```

Total sequence length: 480 nucleotides

The sequences analyzed in HSF are underlined.

Interpreted Data

This table shows only relevant results related to the mutation position and context.
The mutation occurs in the late exonic positions, the following table show results of donor splice sites, ESE and ESS that could be affected by the mutation

Predicted signal	Prediction algorithm	cDNA Position	Interpretation
New ESS Site	1 - ESR Sequences from Goren et al.		Creation of an exonic ESS site. Potential alteration of splicing.
	2 - IIEs from Zhang et al.		
ESE Site Broken	1 - PESE Octamers from Zhang & Chasin		Alteration of an exonic ESE site. Potential alteration of splicing.
	2 - HSF Matrices - 9G8		
	3 - EIEs from Zhang et al.		

Kuva 1. Esimerkki *BRCA1* c.230C>T-variantin Human Splicing Finder tuloksien yhteenvedosta. Kuvassa näkyy ylhäällä sekvenssi, jossa on vihreällä kirjaimella merkitty normaali alleelin C-emäs ja alemmassa sekvenssissä punaisella kirjaimella merkitty varianttialleelin T-emäs. Eksoninen sekvenssi on merkitty isoilla kirjaimilla ja introninen sekvenssi pienillä kirjaimilla. Kuvan alaosan ”interpreted data” -osiossa näkyy minkälaisia ja millä algoritmeilla ohjelma on löytänyt mahdollisia silmukointiin vaikuttavia muutoksia varianttisekvenssissä.

2.5 *in silico* -analyysi RBPmap-ohjelmalla

Tähän analyysiin käytetty ohjelma oli RBPmap version 1.1. Eksonisten VUSien analyysiä varten haettiin *BRCA*-geenin transkriptin referenssisekvenssi NCBI ”nucleotide” tietokannasta. *BRCA1*-transkripteista käytettiin *BRCA1*-transkripti varianttia 1 (NCBI Reference Sequence: NM_007294.4) ja *BRCA2*-geenin kohdalla ainutta tietokannassa olevaa referenssi transkriptia (NCBI Reference Sequence: NM_000059.3). Intronisten varianttien kohdalla haettiin samasta tietokannasta geenien referenssisekvenssit *BRCA1*:lle (NCBI Reference Sequence: NG_005905.2) ja *BRCA2*:lle (NCBI Reference Sequence: NG_012772.3).

Sekvensseistä etsittiin VUS-emäksen positio ja analyysiä varten kopioitiin VUS-emäksen molemmilta puolilta 10 emästä. Tämä sekvenssi syötettiin RBPmap ikkunaan

ja genomiksi valittiin human. ”Motif selection” -kohdasta valittiin RBPmap listalta ”all human/mouse motifs” -vaihtoehto, ja analyysi suoritettiin muuttamatta muita parametrejä. Jos 10 emästä ei tuottanut tuloksia, lisättiin VUS-emästä ympäröivään sekvenssiin emäksiä 25 emäkseen asti viisi kerrallaan, kunnes ohjelma antoi tuloksen. Jos tulosta ei saatu 25 emäksen kanssa, todettiin ettei sitoutumiskohtia löytynyt.

Tämän jälkeen normaali referenssisekvenssissä oleva emäs muutettiin VUS-variantti emäkseksi ja sama analyysi toistettiin. Molempien analyysien tulokset tallennettiin, ja sekvenssien eroja tutkittiin. Merkittävät muutokset villityypin sekvenssistä VUS-mutanttisekvenssiin rajattiin sitoutumiskohtiin, jotka sisälsivät VUS-emäksen. Muutoksia oli kahdenlaisia: uudet varianttisekvenssiin ilmestyneet sitoutumiskohdat, sekä normaalisekvenssistä hävinneet sitoutumiskohdat. Jos sitoutumiskohdan muutos oli sellaisessa positiossa, että sitoutumiskohta ei sisältänyt VUS-emästä, tällaista muutosta ei huomioitu tarkastelussa. Jos tapahtunut muutos oli sellainen, että sekvenssissä oli vieläkin usea saman proteiinin sitoutumiskohtaa, ja varianttisekvenssistä hävisi osa näistä sekvensseistä, ei tätä muutosta myöskään huomioitu tarkastelussa, koska sekvenssiin jäi edelleen sitoutumiskohtia tälle proteiinille.

Kun VUS-mutanttisekvenssiä verrattiin villityypin sekvenssiin, havainnot jaettiin neljään kategoriaan riippuen sitoutumiskohtaa vastaavan proteiinin toiminnasta: uudet ja hävinneet mahdollisesti silmukointiin liittyvät sitoutumiskohdat, sekä uudet ja hävinneet muut sitoutumiskohdat. Proteiinien toiminnan määrittämiseen haettiin tieto GeneCards tietokannasta. Kuvassa 2 esimerkki yhden variantin RBPmap analyysin tuloksista.

BRCA1 C.230C>T
Normaali alleeli

Results for sequence: sequence1
Genomic position: chr17:43104924-43104954 Strand: -
[View binding sites predictions summary](#)

Protein: MS1(Hs/Mm)	Position	Genomic coordinate	Motif	Occurrence	Z-score	P-value
	23	chr17:43104932	uaguwrg	gcccacaagaagaagucagaguu <u>u</u> gUCa <u>u</u> acu	1.955	2.53e-02
Protein: PABPC4(Hs/Mm)	7	chr17:43104948	aaaaaar	gccc <u>u</u> ca <u>g</u> g <u>g</u> aa <u>g</u> uacagaguuaguc <u>u</u> acu	1.882	2.99e-02
Protein: SNRNPT0(Hs/Mm)	24	chr17:43104931	rwucaag	gcccacaagaagaagucagaguu <u>u</u> UCA <u>u</u> CU	1.973	2.42e-02
Protein: SRSF3(Hs/Mm)	2	chr17:43104953	wcww	gCC <u>U</u> uCaagaagaagucagaguuaguc <u>u</u> acu	2.373	8.82e-03
	5	chr17:43104950	wcww	gCC <u>U</u> uCa <u>g</u> gaaagaagucagaguuaguc <u>u</u> acu	2.402	8.15e-03
	12	chr17:43104943	wcww	gcccacaaga <u>g</u> aa <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	2.314	1.03e-02
	26	chr17:43104929	wcww	gcccacaagaagaagucagaguuag <u>U</u> CA <u>u</u> CU	2.794	2.60e-03
Protein: SRSF5(Hs/Mm)	3	chr17:43104952	yywvsg	gCC <u>U</u> uCa <u>g</u> gaaagaagucagaguuaguc <u>u</u> acu	2.591	4.79e-03
Protein: SRSF7(Hs/Mm)	12	chr17:43104943	wgacta	gcccacaaga <u>g</u> aa <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.957	2.52e-02
	14	chr17:43104941	acgag	gcccacaagaaga <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.897	2.69e-02
	15	chr17:43104940	acgagagay	gcccacaagaaga <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	2.550	5.98e-03
Protein: TRA2B(Hs/Mm)	15	chr17:43104940	aaguguu	gcccacaagaaga <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.684	4.61e-02
	17	chr17:43104938	aaguguu	gcccacaagaaga <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.684	4.61e-02

[Download predictions summary as text file](#)

Variantti alleeli

Results for sequence: sequence1
Genomic position: chr17:43104924-43104954 Strand: -
[View binding sites predictions summary](#)

Protein: MS1(Hs/Mm)	Position	Genomic coordinate	Motif	Occurrence	Z-score	P-value
	23	chr17:43104932	uaguwrg	gcccacaagaagaagucagaguu <u>u</u> gUCa <u>u</u> acu	1.955	2.53e-02
Protein: PABPC4(Hs/Mm)	7	chr17:43104948	aaaaaar	gccc <u>u</u> ca <u>g</u> g <u>g</u> aa <u>g</u> uacagaguuaguc <u>u</u> acu	1.882	2.99e-02
Protein: SNRNPT0(Hs/Mm)	24	chr17:43104931	rwucaag	gcccacaagaagaagucagaguu <u>u</u> UCA <u>u</u> CU	1.973	2.42e-02
Protein: SRSF3(Hs/Mm)	2	chr17:43104953	wcww	gCC <u>U</u> uCaagaagaagucagaguuaguc <u>u</u> acu	2.314	1.03e-02
	5	chr17:43104950	wcww	gCC <u>U</u> uCa <u>g</u> gaaagaagucagaguuaguc <u>u</u> acu	2.373	8.82e-03
	26	chr17:43104929	wcww	gcccacaagaagaagucagaguuag <u>U</u> CA <u>u</u> CU	2.676	3.73e-03
Protein: SRSF5(Hs/Mm)	3	chr17:43104952	yywvsg	gCC <u>U</u> uCa <u>g</u> gaaagaagucagaguuaguc <u>u</u> acu	2.591	4.78e-03
Protein: SRSF7(Hs/Mm)	15	chr17:43104940	acgagagay	gcccacaagaaga <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.850	3.22e-02
Protein: TRA2B(Hs/Mm)	15	chr17:43104944	aaguguu	gcccacaaga <u>g</u> aa <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	2.105	1.76e-02
	17	chr17:43104938	aaguguu	gcccacaagaaga <u>g</u> aa <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.768	3.85e-02
	17	chr17:43104938	aaguguu	gcccacaagaaga <u>g</u> aa <u>g</u> U <u>u</u> CA <u>g</u> Uagaguuaguc <u>u</u> acu	1.768	3.85e-02

[Download predictions summary as text file](#)

Kuva 2. Esimerkki *BRCA1* C.230C>T-variantin RBPmap tuloksista. Kuvassa vasemmalla normaali alleelin tulokset ja oikealla varianttialleelin tulokset. Kuvassa punaisilla neliöillä merkitty ohjelman havaitsemat erot sekvenssien välillä. Merkittävä ero sekvenssien välillä näyttäisi olevan SRSF3-proteiinin sitoutumiskohdan häviäminen varianttisekvenssistä positiossa 12. Kyseisen proteiinin sitoutumiskohta on neljä emästä. Tämä on merkittävin ero, sillä varianttiemäs on tässä tutkimuksessa sekvenssissä positiossa 16, eli sitoutumiskohta sisältää VUS-emäksen. Varianttisekvenssistä näyttää hävinneen myös SRSF7-proteiinin sitoutumiskohdat positioista 12 ja 14, mutta näitä ei huomioitu, koska positiossa 15 oleva sitoutumiskohta säilyi varianttisekvenssissä. Sama pätee myös varianttisekvenssiin positioon 11 ilmestyneeseen uuteen TRA2B-proteiinin sitoutumiskohtaan. Muutosta ei huomioitu, sillä sekvenssistä löytyi jo TRA2B sitoutumiskohta positioista 15.

2.6 *in silico* -analyysi RegRNA-ohjelmalla

RegRNA-analyysi suoritettiin RegRNA 2.0 ohjelmistoversiolla. Ohjelman ”scan” -ikkunaan syötettiin samanlainen sekvenssi kuin Human Splicing Finder -ohjelmaan, mutta Human Splicing Finderistä poiketen, RegRNA-ikkunaan kopioitiin sekvenssi, jossa oli koko eksoni ja 150–200 emästä intronista sekvenssiä eksonin molemmilta puolilta, olettaen että sekvenssin pituus oli alle 10 kb. Ikkunan ”select RNA motifs” -valikosta valittiin ”select all” -vaihtoehto ja ikkunan alaosasta painettiin ”submit” -painiketta. Tämä analyysi tehtiin normaalille sekvenssille, jossa varianttiemäksen kohdalla oli normaali emäs, sekä varianttisekvenssille, jossa normaali emäs oli muutettu variantti emäkseksi. Analyysien tuloksia verrattiin toisiinsa, ja muutokset normaali- ja varianttisekvenssin välillä kirjattiin ylös. Kuvassa 3 on esimerkki yhden variantin RegRNA-analyysin tuloksista.

BRCA2	Oct-1	278–290	13	ttggtatgataag	BRCA2	Oct-1	278–290	13	ttggtatgataag
C.583T>G normaali alleeli	Bmi-2	274–289	16	ctctatggtatgataa	C.583T>G variantti alleeli	Bmi-2	274–289	16	ctctatggtatgataa
	TATA	133–147	15	ctctatggtatgataa		TATA	133–147	15	ctctatggtatgataa
	Pax-3	168–180	13	ttgttagatgataa		Pax-3	168–180	13	ttgttagatgataa
	Pax-4	249–260	12	atctctctctctct		Pax-4	249–260	12	atctctctctctct
	SMAD	173–181	9	agpctctaa		C/EBP	225–236	12	gctcttggatgag
	MYB	407–415	9	taactgtgca		SMAD	173–181	9	agpctctaa
	SMAD	168–178	11	ttgttagatgac		MYB	407–415	9	taactgtgca
	SOX	387–399	13	gctcttggatgag		SMAD	168–178	11	ttgttagatgac
	Klf3	250–254	5	ctaac		SOX	387–399	13	gctcttggatgag
	Klf3	254–258	5	ctaac		Klf3	250–254	5	ctaac
	Nkx3-2	79–84	6	taagtg		Klf3	254–258	5	ctaac
	Nkx3-2	395–400	6	taagtg		Nkx3-2	79–84	6	taagtg
Transcriptional regulatory motif	Nkx3-2	431–436	6	taagtg	Transcriptional regulatory motif	Nkx3-2	395–400	6	taagtg
	ZNF333	106–110	5	ataat		Nkx3-2	431–436	6	taagtg
	ZNF333	284–288	5	ataat		ZNF333	106–110	5	ataat
	ZNF333	327–331	5	ataat		ZNF333	284–288	5	ataat
	ZNF333	382–386	5	ataat		ZNF333	327–331	5	ataat
	ZNF333	416–420	5	ataat		ZNF333	382–386	5	ataat
	HOXA13	135–140	6	ataaa		ZNF333	416–420	5	ataat
	CDP	113–127	15	ctctatggtatgagpp		HOXA13	135–140	6	ataaa
	TCF-4	388–396	9	ttcttgata		CDP	113–127	15	ctctatggtatgagpp
	IRF-4	371–377	7	gaaata		TCF-4	388–396	9	ttcttgata
	IRF-7	80–86	7	taagtg		IRF-4	371–377	7	gaaata
	Smad3	169–181	13	ttgttagatgataa		IRF-7	80–86	7	taagtg
	PARP	151–156	6	ttctct		Smad3	169–181	13	ttgttagatgataa
	Cdx-1	421–426	6	ttctct		PARP	151–156	6	ttctct
						Cdx-1	421–426	6	ttctct

Kuva 3. Esimerkki *BRCA2* C.583T>G-variantin RegRNA tuloksista. Kuvassa näkyvillä vain tuloskatsauksen ”transcriptional regulatory motif” -osio, josta ainut ero sekvenssien välillä löytyi. Tuloksista näkyy, että variantti sekvenssiin on ilmestynyt uusi C/EBP sitoutumiskohta positioon 225–236. Tulos oletetaan merkittäväksi, sillä tutkitussa sekvenssissä variantti emäksen positio oli 232, eli juuri tämän uuden sitoutumiskohdan paikalla.

2.7 Alukkeiden ja koettimien suunnittelu

Alukkeet ja koettimet suunniteltiin ”Bio-Rad Droplet Digital™ PCR

Applications Guide” aluke- ja koetinsuunnitteluohjeiden, sekä Kamitakin ym. (2018) alleelispesifisen RNA-ekspression tutkimiseen tarkoitettun ddPCR-protokollan mukaisesti. Alukkeet ja koettimet suunniteltiin Primer3plus -ohjelman avulla siten, että alukkeiden suunnittelussa ja valitsemisessa PCR-tuotteen koko oli 60–200 emästä, sekvensseissä vältettiin kolmea emästä pidempiä G- ja C-toistoja ja alukkeiden 3’ -päähen suunniteltiin G- tai C-emäs, jos mahdollista. Koettimet suunniteltiin siten, että niiden pituus oli enintään 30 emästä, ja että tutkittava variantti emäs sijaitsi mahdollisimman keskellä koetinta. Jos tämä ei ollut mahdollista, variantti emäksen täytyi sijaita vähintään neljän emäksen päässä koettimen 3’ tai 5’ -päästä. Koettimessa ei myöskään saanut olla G-emästä 5’ -päässä. Koetinsekvenssi ei saanut olla päällekkäinen alukesekvenssin kanssa, mutta sai sijaita alukesekvenssin vieressä.

Jos alukesuunnittelu ei onnistunut Kamitakin ym. (2018) protokollan 60 °C annealing lämpötila-asetuksilla, kokeiltiin protokollan vaihtoehtoisia 55 °C asetuksia. Jos alukesuunnittelu ei onnistunut näinkään, kokeiltiin omia koetinsekvenssejä, jotka olivat enintään 30:n emäksen pituisia, ja joissa varianttiemäs oli vähintään neljän emäksen päässä koettimen 5’ ja 3’ -päästä. Jos omien koetinsekvenssien syöttäminen ohjelmaan ei myöskään toiminut, todettiin kyseinen variantti sopimattomaksi tutkimukseen alukesuunnittelun epäonnistumisen takia.

BRCA1- ja *BRCA2*-varianteista valittiin alukesuunnitteluun ensin ne variantit, joiden vaikutuksista ei ClinVar tietokannan mukaan ole tarkkaa tietoa, ja jotka vaikuttivat Alamut-ohjelman analyysien perusteella haitallisilta. Tämän lisäksi alukesuunnitteluun valittiin myös variantit, joiden osalta oli odotettavissa saatavilla olevia potilasnäytteitä. Tiettyjen varianttien kantajia pyydettiin tutkimuksen valmistelun aikana näytteenottoon, mutta näytesaatavuudesta tuli ongelma tutkimuksen kannalta: odotettua määrää potilasnäytteitä ei saatu, ja näytteitä tiettyjen varianttien kantajilta oli niin vähän, ettei niiden

tarkasteleminen olisi ollut tutkimuksen kannalta mielekäästä. Liian pieni koehenkilömäärä olisi tehnyt tuloksien tulkitsemisesta liian epävarmaa. Tiedot varianttien CinVar tulkinnoista, näytesaatavuudesta ja alukesuunnittelun onnistumisesta löytyvät taulukosta 1.

Taulukko 1. Kaikki tutkimuksessa käsitellyt *BRCA1*- ja *BRCA2*-variantit, näytesaatavuus, alukesuunnittelun onnistuminen ja ClinVar-tietokannan tulkinta variantin patogeenisyydestä.

Variantti	ClinVar tulkinta (Review status)	Potilasnäytteiden saatavuus	Alukesuunnittelun onnistuminen
<i>BRCA1</i>			
c.230C>T	2/4 epävarma	0	Ei
c.3257T>C	0/4 epävarma	0	-
c.3649T>C	2/4 epävarma	0	Ei
c.4676-8C>G	¼ mahdollisesti harmiton	1	Kyllä
<i>BRCA2</i>			
c.68-7T>A	¾ harmiton	2	Ei
c.583T>G	-	0	Ei
c.2884C>T	2/4 epävarma	1	Ei
c.3314C>T	-	1	Ei
c.4756A>C	-	1	Kyllä
c.5225A>T	2/4 epävarma	0	Ei
c.6821G>T	¼ mahdollisesti harmiton	1	Kyllä
c.7121A>G	¼ epävarma	0	-
c.7712A>G	¼ epävarma	0	Ei
c.7759C>T	¼ epävarma	0	-
c.7805+13A>G	¼ epävarma	0	-
c.8855T>C	-	0	-
c.8953+24T>C	-	1	Ei
c.9232G>T	-	2	Ei
c.9302T>C	¼ epävarma	0	Ei
c.9411T>G	¾ harmiton	4	Ei
c.9872C>G	2/4 epävarma	0	Ei

c.10234A>G	$\frac{3}{4}$ harmiton	0	-
------------	------------------------	---	---

Koska alukesuunnittelu epäonnistui suurimmalta osalta varianteista, *BRCA1*- ja *BRCA2*-varianttien lisäksi suunniteltiin myös SISu-tietokannasta löytyneille Suomessa yleisimmin esiintyville *BRCA2* SNP-emäksille alukkeet. Yleisimmistä SNP:istä rajattiin pois sellaiset, joiden alleelifrekvenssi oli alle 0.15, ja jotka sijaitsivat UTR- tai toistojakson alueella. Nämä SNP:t ja niiden alukesuunnittelun onnistuminen on listattu taulukkoon 2.

Taulukko 2. SISu-tietokannan yleisimmät *BRCA2* SNP:t ja näiden alukesuunnittelun onnistuminen.

<i>BRCA2</i> SNP	Alleelifrekvenssi	Alukesuunnittelun onnistuminen
rs9534262	0.578091	Kyllä
rs1801406	0.319613	Ei
rs1799955	0.237801	Ei
rs144848	0.196731	Kyllä
rs543304	0.186071	Ei

ANO7-tutkimukseen valittiin sekvenssidatan perusteella yhdeksän *ANO7*-varianttia, joille alukesuunnittelu tehtiin. Näistä varianteista vain neljälle alukkeiden ja koettimien suunnittelu onnistui tyydyttävällä tavalla, ja näistä neljästä variantista kolme valittiin tutkimukseen. Kyseiset variantit löytyvät taulukosta 3.

Muista varianteista poiketen, intronisen rs78154103 GC-variantin alukesuunnittelussa otettiin huomioon variantin aiheuttama häiriö intronin seitsemän silmukointiin. Sekvenssintidatan perusteella on havaittu, että jos henkilö kantaa alleelia G, tämä variantti aiheuttaa silmukoinnissa intronin seitsemän fragmentin säilymisen transkriptissa. Tämän variantin alukesuunnittelussa käytettiin seuraavaa sekvenssiä, jossa eksoninen sekvenssi on alleviivattu ja variantti emäs on punainen:

GCCGGTTCAGAGTGAACAAGCTGCCACGCTTCCTCGGGAGTGACAACCAGG
ACACCTTCTTCACAAGCACCAAGAGGCACCAAATTGTGAGTGGGGGTTCC
 TGCCGCGGGCCCCAAACCCTGACCTGGCTCTGAGGATCCTGCCAGCCCCCA
 GCCTCCATGGATGCAGAAAGGCCTGCTTGAGACCAGAGGGCCGAGGCC

Onnistuneesti suunnitellut alukkeet tilattiin Bio-Radin ”Customize PCR primers or probe assays” kautta valitsemalla ”ddPCR probe assay”. Jokaisen variantin HEX- ja FAM-leimatut alleelispesifiset koettimet alukkeineen tilattiin erikseen, asettamalla ”create custom assay” -ikkunaan oikea ja vasen aluke, haluttu koetin ja sen leima, ja asettamalla aluke-koetin suhde (primer-to-probe ratio) 1,8:aan. Tarkat tiedot tutkimukseen käytetyistä alukkeista ja koettimista löytyvät taulukosta 4.

Taulukko 3. *ANO7*-variantit, joille protokollan mukainen alukesuunnittelu tehtiin. Lihavoituna tutkimukseen valitut variantit.

<i>ANO7</i> -variantti	Alukesuunnittelun onnistuminen
rs77559646 GA	Kyllä
rs11695874 GT	Ei
rs147325507 CA	Ei
rs2074840 CT	Kyllä
rs57677160 CT	Ei
rs7593101 CT	Kyllä
rs78605785 GT	Ei
rs7590653 GA	Ei
rs78154103 GC	Kyllä

Taulukko 4. *ANO7*-varianteille suunnitellut alukkeet ja koettimet. 1: Vasen aluke, 2: Oikea aluke, 3: PCR-tuotteen koko, 4: Koetin, normaali alleeli, varianttiemäs punaisena alleviivattuna, suluissa leima, 5: Koetin, variantti alleeli, variantti emäs punaisena alleviivattuna, suluissa leima. Nukleotidisekvenssien yhteyteen on merkitty Primer3plus-ohjelman ennakoima sulamislämpötila (Tm). Eksoni 8 rs2074840 CT- ja eksoni 16 rs7593101 CT-varianttien alukkeet suunniteltiin käänteiskomplementaariselle sekvenssille.

	<i>ANO7</i> rs77559646 GA (eksoni/introni 4)	<i>ANO7</i> rs2074840 CT (eksoni 8)	<i>ANO7</i> rs7593101 CT (eksoni 16)
1	5'- GACAGAACAGACATGCACAG- 3' Tm 59.0	5'- TGCTGGAAAAGGACTTGGC -3' Tm 60.5	5'- AAACAAGGTGTGGTAGTTG-3' Tm 55.2
2	5'- GTCAGTACTGGATGCAGG-3' Tm 60.1	5'- AAAAACCTGCTTGGGATCC A-3' Tm 59.5	5'-CTCAAGGTGTTTCATCTTCC- 3' Tm 55.2
3	131 bp	140 bp	108 bp
4	5'- CCAGGTAC G TGGAGGCTGTCA TGGGC-3' Tm 64.5 (FAM)	5'- AGGGGCC G TCATGCAGGG GGA-3' Tm 64.0 (HEX)	5'- CAATGTAGACGGGTGAGGAG T AGAAGT-3' Tm 57.5 (HEX)
5	5'- CCAGGTAC A TGGAGGCTGTCA TGGGCAG-3' Tm 64.0 (HEX)	5'- TGGGGGCGTCTTGAAGGGG CC A TCAT-3' Tm 65.5 (FAM)	5'- AGGCAATGTAGACGGGTGAG GA A TAGAA-3' Tm 59.1 (FAM)

2.8 *ANO7* alukkeiden ja koettimien toimivuuden testaus

Kaikkien *ANO7* alukkeiden ja koettimien toimivuus testattiin tekemällä eturauhassyöpäsoluminjojen cDNA:lla ddPCR-reaktio. Optimaalinen PCR-syklin annealing-lämpötila testattiin lämpögradientilla, ja tämän jälkeen alukkeiden ja koettimien spesifisyys testattiin variantti DNA konsentraatiogradientilla.

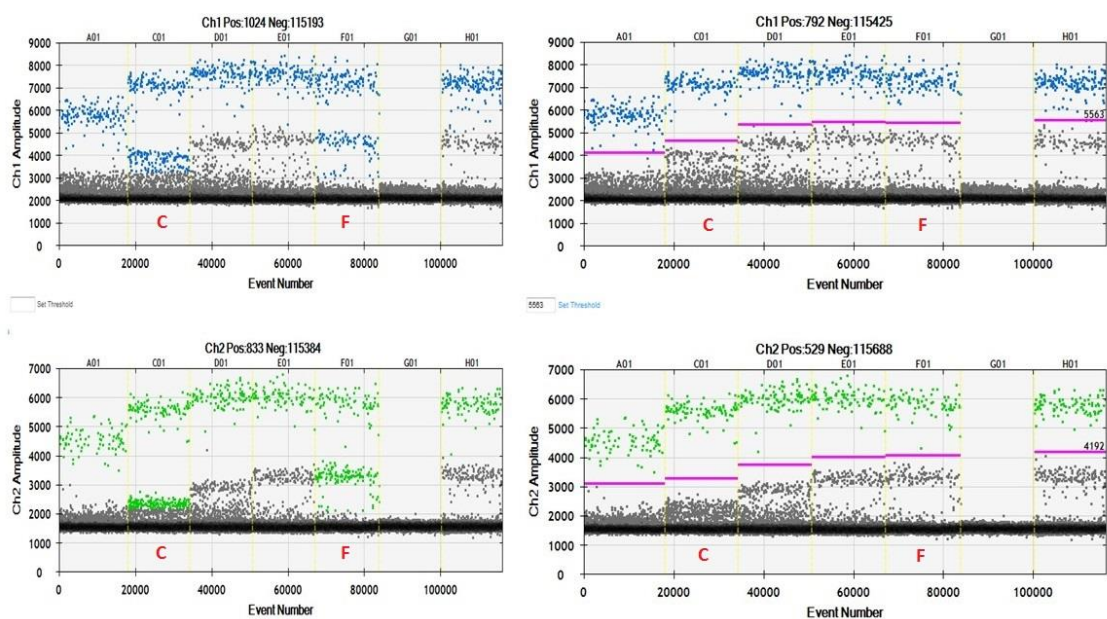
Testausta varten kolmen eturauhassyöpäsoluminjan eristetystä RNA:sta tehtiin cDNA samalla menetelmällä kuin *BRCA*-näytteille "iScript™ Advanced cDNA Synthesis Kit for RT-qPCR" -kitillä valmistajan ohjeiden mukaan. Jokainen *ANO7* aluke- ja koetinpari testattiin eri solulinjoilla, joissa tutkittava variantti esiintyi. *ANO7* eksoni 16 rs7593101 -variantin alukkeiden testaamiseen käytettiin MDA PCa 2b-soluminjan cDNA:ta, *ANO7* eksoni/introni 4 rs77559646 -variantin alukkeiden testaamiseen käytettiin 22Rv1-soluminjan cDNA:ta ja *ANO7* eksoni 8 rs2074840 -variantin alukkeiden testaamiseen käytettiin LNCaP-soluminjan cDNA:ta.

DdPCR-reaktiota varten näytteet valmisteltiin syväkaivoiselle PCR-levylle "PrimePCR™ ddPCR™ Gene Expression Probe Assays" -ohjeen mukaan. Reaktioon käytettiin ohjeen mukainen maksimaalinen 50 ng initial RNA määrä solulinjojen alhaisen *ANO7*-geenin ekspression takia. Protokollan ohjeiden lisäksi valmis näytelevy myös sekoitettiin vorteksoimalla ja sentrifugoitettiin "Eppendorf Centrifuge 5804" -sentrifugilla minuutin ajan 1000 rpm nopeudella ennen pisaroiden valmistamista. PCR-kuoppalevy myös sinetöitiin "Bio-Rad Piercable Foil Heat Seal" sinetöintiliuskalla "Bio-Rad PX1 PCR Plate Sealer" -laitteella 180 °C lämpötilassa viiden sekunnin ajan ennen vorteksointia ja sentrifugointia. DdPCR-reaktioon tarvittavat pisarat valmistettiin sentrifugoinnin jälkeen automaattisella "Bio-Rad AutoDG™ Instrument" -laitteella valmistajan ohjeen mukaan. Pisaroiden muodostamisen jälkeen valmiit pisarat sisältävä PCR-kuoppalevy sinetöitiin samalla tavalla kuin ennen sentrifugointia ja näytteet laitettiin PCR-koneeseen.

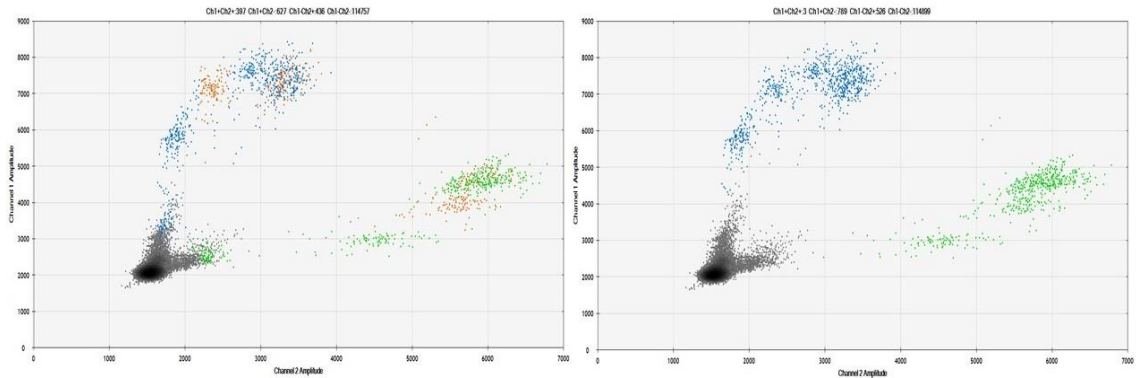
Lämpögradienttia varten "Bio-Rad C1000 Touch™ Thermal Cycler" PCR-koneeseen ohjelmoitiin "PrimePCR™ ddPCR™ Gene Expression Probe Assays" -ohjeen mukai-

nen PCR-sykli, jossa annealing/extension-lämpötilaksi asetettiin lämpögradientti, jolla jokaiseen pystyrivillä olevaan PCR-levyn kaivoon saatiin eri lämpötila. Yhteen kaivoon, jonka lämpötila ei eronnut merkittävästi muista, laitettiin negatiivinen kontrolli, jossa solulinjan cDNA korvattiin vedellä. Samasta syystä toiseen kaivoon laitettiin pelkästään laimennettu ”ddPCR™ Buffer Control for Probes” puskuriliuos, jota ei analysoitu (liite 1, taulukot 15 ja 16).

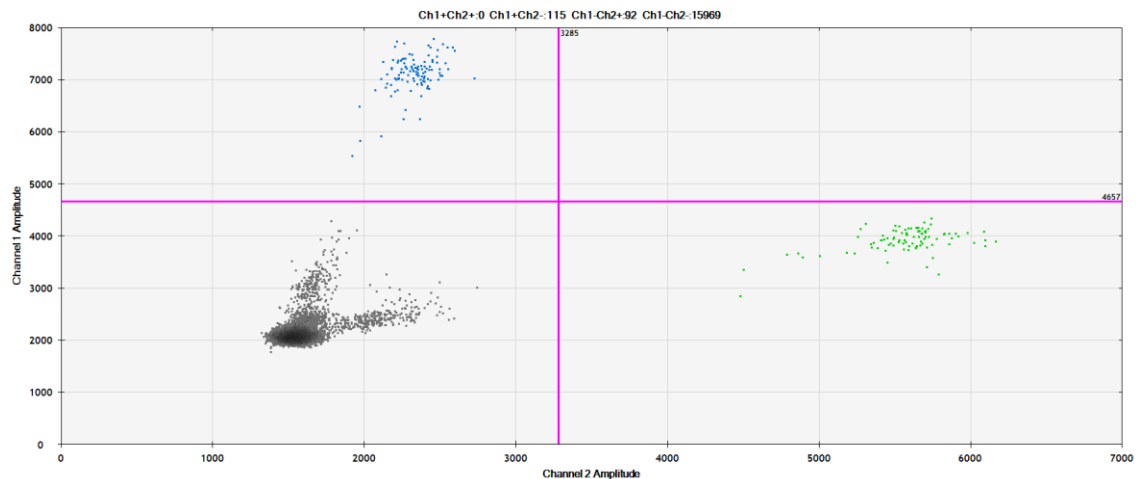
PCR-syklin jälkeen pisarat analysoitiin ”Bio-Rad QX200™ Droplet Reader” -laitteella, ja Quantasoft-ohjelmistolla. Kaikkien lämpötilagradienttien tapauksessa määritettiin itse erikseen jokaiselle kaivolle FAM ja HEX fluoresenssikanaville omat raja-arvot, joiden perusteella ohjelmisto määrittä pisarat negatiivisiksi tai positiivisiksi. Raja-arvot määritettiin käsin, koska ohjelmiston automaattinen analyysitoiminto ei pystynyt asettamaan raja-arvoja oikealla tavalla, mikä johti kyseenalaisiin tuloksiin (kuvat 4 ja 5). Tämä johtui todennäköisesti siitä, että laitteisto analysoi jokaisen kaivon erikseen omana mittauksena, eikä vertaa kaivoja toisiinsa. Raja-arvojen käsin asettelun jälkeen tulokset olivat kunnollisia (kuvat 4 ja 5). Raja-arvojen määrittämisen jälkeen valittiin optimaalinen annealing-lämpötila sen perusteella, missä kaivossa/lämpötilassa FAM-positiiviset, HEX-positiiviset, negatiiviset ja kaksoispositiiviset pisarapopulaatiot erosivat toisistaan parhaiten (kuvat 6 ja 10). Eksoni 16 rs7593101 alukkeiden ja koettimien optimaaliseksi annealing-lämpötilaksi valittiin 59 °C ja eksoni 8 rs2074840 alukkeiden ja koettimien optimaaliseksi annealing-lämpötilaksi valittiin 63 °C.



Kuva 4. ”1-D amplitude plot” esimerkki eksoni 16 rs7593101 näytelevyn lämpötilagradientin tuloksista. Siniset pisteet edustavat FAM-positiivisia pisaroita ja vihreät HEX-positiivisia pisaroita. Vasemman puolen kuvaajassa C ja F kaivojen keskimmäiset pisarapopulaatiot on määritetty automaattisesti positiivisiksi, toisin kuin muissa kaivoissa. Oikean puolen samoissa kuvaajissa raja-arvot on asetettu käsin (violetti viiva) siten, ettei keskimmäistä pisarapopulaatiota oleteta positiiviseksi.

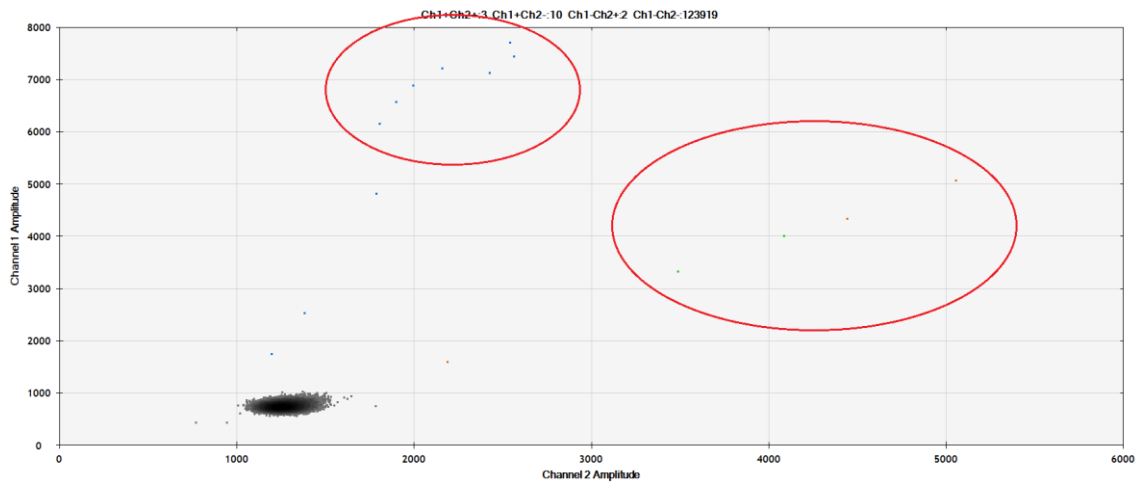


Kuva 5. ”2-D amplitude plot” esimerkki eksoni 16 rs7593101 näytelevyn lämpötilagradientin tuloksista. Kuvassa kaikki kaivot yhdistettynä, vasemmalla puolella automaattisilla raja-arvoilla ja oikealla puolella kuvan 4 mukaan asetetuilla raja-arvoilla. Automaattisilla raja-arvoilla vasemmassa kuvassa esiintyy satunnaisia oransseja kaksoispositiivisia pisaroita kohdissa, joissa niitä ei pitäisi esiintyä. Oikeassa kuvassa oikein asetetuilla raja-arvoilla siniset FAM ja vihreät HEX populaatiot näkyvät selvästi erillään.



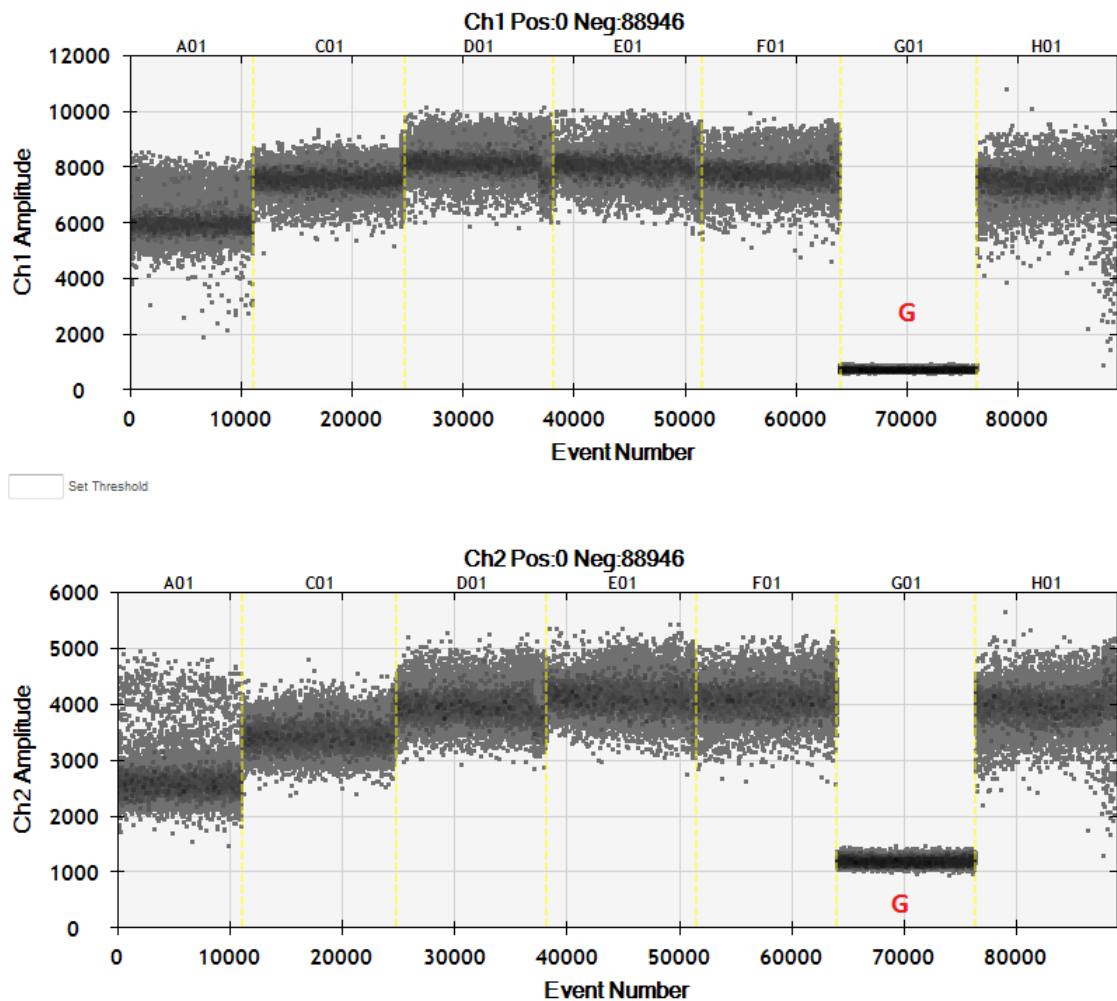
Kuva 6. ”2-D amplitude plot” Esimerkki eksoni 16 rs7593101 näytelevyn kaivosta C, jossa annealing-lämpötila oli 59.1 °C. Kuvassa näkyy miten FAM ja HEX pisarapopulaatiot ovat selvästi erillään toisistaan. MDA PCa 2b-solulinjan alhaisen *ANO7* ekspression takia kuvassa ei näy kaksoispositiivisten pisaroiden populaatiota, joka teoriassa ilmestyisi kuvaajan oikeaan yläkulmaan FAM ja HEX populaatioiden väliin. Kuvassa näkyy myös violeteilla viivoilla asetetut raja-arvot.

22Rv1-solulinjan ekspressio oli niin pientä, ettei eksoni/introni 4 rs77559646 alukkeiden ja koettimien toimivuudesta saatu luotettavaa tulosta (kuva 7). Koe toistettiin uudella PCR-ohjelmalla, jonka lämpötilagradientti oli 62—52 °C (liite 1, taulukko 17), mutta tämä ei parantanut tulosten luotettavuutta.



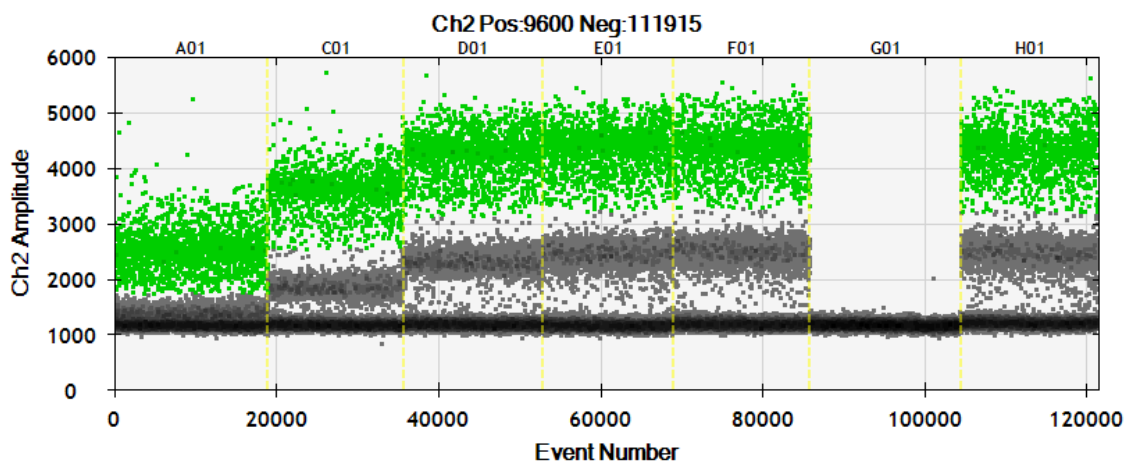
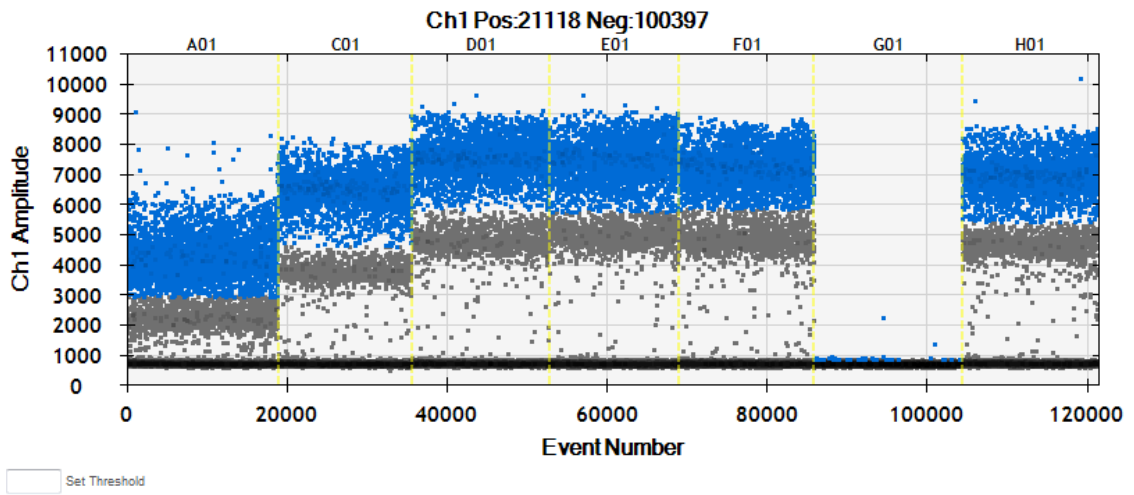
Kuva 7. ”2-D amplitude plot” eksoni/introni 4 rs77559646 näytelevystä, joka tehtiin liitteen 1 taulukon 17 PCR-ohjelman mukaan. Kuvassa kaikki kaivot. Mustalla negatiiviset pisarat, kuvassa näkyy yksittäisiä positiivisia pisaroita, mutta ei tarpeeksi optimaalisen annealing-lämpötilan määrittämistä varten.

Koska 22Rv1-solulinjan cDNA:lla ei pystytty optimaalista annealing-lämpötilaa määrittämään, *ANO7* eksoni/iontroni 4 koettimien ja alukkeiden testausta varten tehtiin *ANO7*:ää yliekspressoivien PC3-solulinjoista eristetystä RNA:sta cDNA:ta. Solulinjassa pLVX-IRES-mCherry-*ANO7S* esiintyy normaali alleeli G ja solulinjassa pLVX-IRES-mCherry-*ANO7S*-R158H esiintyy mutanttialleeli A. Solulinjojen cDNA:sta tehtiin 50/50 sekoitus sopivan annealing-lämpötilan määrittämistä varten. PCR-ohjelma oli sama kuin liitteen 1 taulukossa 17.

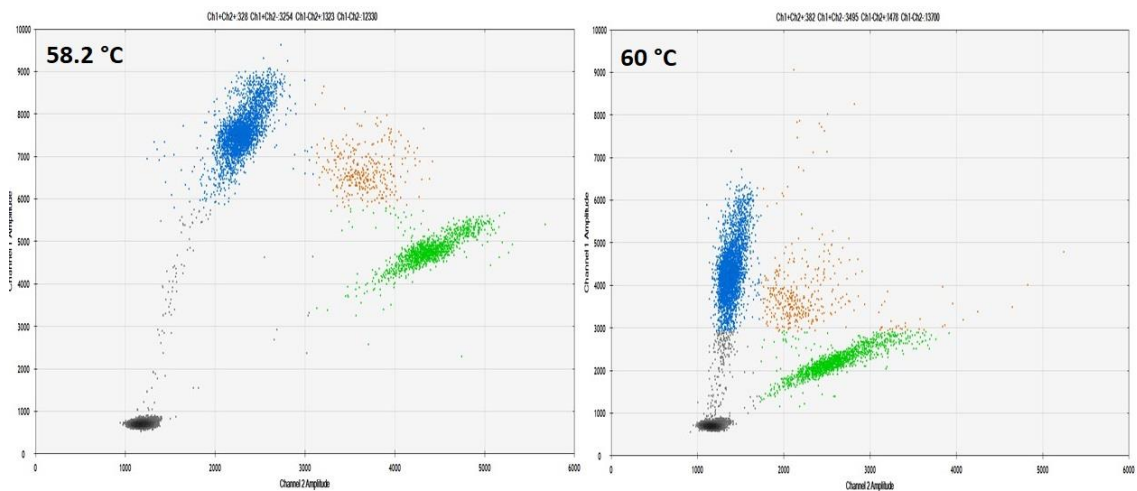


Kuva 8: ”1-D amplitude plot” PC3-solulinjojen 50/50 cDNA sekoituksesta. Kuvassa liian suuri cDNA:n määrä johtaa siihen, ettei negatiivisia pisaroita synny, kuten negatiivisessa kontrollikaivossa G.

PC3-solulinjojen yliekspressio johti siihen, ettei lukulaite pystynyt liian suuren ekspresion takia antamaan tuloksia (kuva 8). Tämä johtui siitä, että laite ei pysty toteamaan signaalia, jollei näytteessä ole positiivisten pisaroiden lisäksi myös negatiivisia pisaroita. Sekoituksesta tehtiin 1:50 laimennos ja koe toistettiin. Tällä laimennoksella saatiin tulos, mutta edelleen suuren ekspresion takia tehtiin toinen 1:250 laimennos, ja koe toistettiin. Tällä laimennoksella saatiin selvemmin näkyviin eri pisarapopulaatiot (kuva 9), ja eksoni/introni 4 alukkeiden ja koettimien annealing-lämpötilaksi valittiin 58 °C (kuva 10).



Kuva 9. ”1-D amplitude plot” PC3-solulinjojen 50/50 cDNA sekoituksen 1:250 laimennoksesta. Laimennoksen avulla ohjelma pystyy erottamaan negatiiviset ja positiiviset pisarat toisistaan, toisin kuin laimentamattomassa näytteessä (kuva 8). Lisäksi kuvassa näkyy selvästi myös kaksi positiivista pisarapopulaatiota, joista vain ylempi määritettiin positiiviseksi asettamalla omat raja-arvot pisarapopulaatioiden väliin.



Kuva 10. ”2-D amplitude plot” esimerkkikuvat PC3-solulinjojen 50/50 sekoituksen 1:250 laimennoksesta. Molemmassa annealing-lämpötiloissa HEX, FAM ja kaksoispositiiviset pisarapo-

pulaatiot ovat selvästi erillään toisistaan, mutta 58.2 °C annealing-lämpötilalla pisarapopulaatioiden välinen etäisyys johtaa selvempään eroon populaatioiden välillä. Tämän perusteella 58 °C valittiin eksoni/introni 4 alukkeiden ja koettimien optimimaaliseksi annealing-lämpötilaksi.

Tämän jälkeen koettimien spesifisyys määritettiin. Kamitakin ym. (2018) protokollassa ei koettimien spesifisyyttä määritetty lainkaan, vaan alukkeiden ja koettimien toimivuuden toteamiseen riitti vain parhaan annealing-lämpötilan määrittäminen. Tämä johtuu siitä, että koettimien samankaltaisuuden aiheuttaman ristireagoinnin takia spesifisyyden testaamista ei pidetä olennaisena. Menetelmällä tutkitaan alleelien suhteita, ja tämä ristireagointi ei protokollan mukaan vaikuta tuloksiin merkittävästi, kunhan ”2-D amplitude plot” -ikkunassa pisarapopulaatiot ovat selvästi erillään. Protokollasta poiketen päätimme kuitenkin tarkastella koettimien spesifisyyttä konsentraatiogradientteilla.

Eksoni/introni 4 koettimien spesifisyyden testaamista varten ANO7S- ja ANO7S-R158H-solulinjojen cDNA:sta tehtiin 1:250 laimennettu konsentraatiogradientti, sekoittamalla solulinjojen cDNA:ta eri suhteissa, jolloin normaalia G alleelia kantavan ANO7S-solulinjan cDNA:n määrää näytteessä kasvatettiin (liite 1, taulukko 18).

Taulukko 5. Yhteenveto eksoni/introni 4 rs77559646 konsentraatiogradientin tuloksista.

Normaali / variantti cDNA:n suhde näytteessä	Normaali / variantti cDNA:n konsentraatio (kopiota / μ l)	Normaali variantin cDNA:n suhde variantti cDNA:han	Normaalien FAM-positiivisten / variantti HEX-positiivisten pisaroiden määrä
50/50	260/107	2.5	2508/1074
50/50	279/110	2.53	2503/1060
75/25	401/55.7	7.2	3662/586
75/25	412/53.9	7.6	3545/538
90/10	500/26.8	18.6	4904/319
90/10	469/28	16.8	3581/256

Taulukossa 5 varianttialleelia vastaavan HEX-leimatun koettimen aiheuttaman signaalin selvä lasku tätä varianttia sisältävän cDNA:n määrän vähentyessä osoittaa, että koettimet ovat tarpeeksi spesifisiä tutkimustarkoitukseen. Taulukkoon kerätyt tiedot ja esimerkkikuvat löytyvät liitteen 1 kuvista 11–14.

Samanlaiset konsentraatigradientit tehtiin myös eksoni 8 rs2074840 ja eksoni 16 rs7593101 koettimien tarkastelua varten. Eksoni 8 konsentraatigradientissa vähennettiin heterotsygoottisen variantti alleelia sisältävän LnCap-solulinjan cDNA:ta korvaamalla se normaalia alleelia sisältävän homotsygoottisen MDA-solulinjan cDNA:lla. Eksoni 16 konsentraatigradientti tehtiin päinvastoin vähentämällä varianttialleelia sisältävän MDA-solulinjan cDNA:ta korvaamalla se normaalia alleelia sisältävän LnCap-solulinjan cDNA:lla. Vaikka konsentraatigradientit erosivat hieman toisistaan, johtuen solulinjojen eri DNA konsentraatioista, kaikissa näkyy, miten varianttialleelin konsentraatio laskee, kun variantti alleelia sisältävän cDNA:n määrä näytteessä vähenee. Konsentraatiogradienttien toistamista ei pidetty tutkimuksen kannalta kriittisenä. Taulukoiden 5, 6 ja 7 konsentraatiogradienttien tuloksien perusteella kaikkien varianttien koettimet todettiin toimiviksi, eli koettimet pystyvät erottamaan normaalin- ja varianttialleelin toisistaan, kunhan sopivat raja-arvot asetetaan.

Taulukko 6. Yhteenveto eksoni 8 rs2074840 konsentraatiogradientin tuloksista.

Normaali (MDA) / variantti (LnCap) cDNA:n suhde näytteessä	Normaali / variantti cDNA:n konsentraatio (kopiota / μ l)	Variantti cDNA:n suhde normaaliin cDNA:han	Normaalien HEX-positiivisten / variantti FAM-positiivisten pisaroiden määrä
0/100	5.4/2.6	0.49	72/35
0/100	6/2.9	0.48	77/37
50/50	12.2/1.6	0.13	132/17
50/50	11.5/2.1	0.18	125/23
90/10	16.4/0.65	0.04	174/7
90/10	17.6/0.52	0.03	200/6
100/0	16.9/0	-	164/0

Taulukko 7. Yhteenveto eksoni 16 rs7593101 konsentraatigradientin tuloksista.

Normaali (LnCap) / variantti (MDA) cDNA:n suhde näytteessä	Normaali / variantti cDNA:n konsent- raatio (kopiota / μ l)	Variantti cDNA:n suhde normaaliin cDNA:han	Normaalien HEX- positiivisten / vari- antti FAM- positiivisten pisa- roiden määrä
0/100	5/9	1.8	56/100
0/100	6.5/10.5	1.6	63/102
50/50	29/4.6	0.16	309/50
50/50	24.3/5	0.21	260/54
90/10	38.7/0.38	0.01	405/4
90/10	38.2/1.3	0.035	393/14

2.9 ANO7-näytteiden analysointi

Potilasnäytteiden *ANO7* koesarjat tehtiin samalla tavalla kuin alukkeiden optimointi ”PrimePCR™ ddPCR™ Gene Expression Probe Assays” -ohjeen mukaan. *ANO7*-analyysiin valittiin jo valmiiksi varastossa olevia cDNA-potilasnäytteitä. Tutkimuksessa käytetty cDNA oli tuotettu oligo(dT)- ja satunnaisalukemenetelmän (random primer mix) yhdistelmällä. Tutkittavia näytteitä oli kymmenen: BLCA10, RALP32, -35, -161, -202, -205, -207, -214, -378 ja -387. Kaikki näytteet olivat peräisin syöpäkudoksesta, paitsi näyte BLCA10, joka oli normaalia kudosta. Kaikista näytteistä tehtiin 1:10 cDNA laimennos, ja ensimmäiseksi tutkittiin eksoni 16 rs7593101 -variantti näytteistä 202, 205 ja 207. Onnistuneen kokeen jälkeen sama koe suoritettiin lopuille näytteille.

Kun loput näytteet tutkittiin eksoni 16 koettimilla, havaittiin että näytteiden 10, 161 ja 214 cDNA-konsentraatio ja positiivisten pisaroiden määrä oli erittäin alhainen. Luotettavamman tuloksen saamiseksi tehtiin näiden näytteiden cDNA:sta 1:3 laimennos, jota käytettiin seuraavissa koesarjoissa. Tämän lisäksi havaittiin, että näytteet 378 ja 387 eivät tuottaneet lainkaan pisaroita ddPCR-analyysissä, johtuen todennäköisesti siitä, etteivät näytteet sisältäneet cDNA:ta, tai siitä että, *ANO7*-geenin ekspressiotaso on kyseisissä näytteissä liian alhainen. Tämän takia näytteet 378 ja 387 todettiin sopimattomiksi tutkimukseen. Koesarjoja tehtiin yhteensä kolme, joista jokaisessa oli jokaisesta tutkittavasta näytteestä kaksi duplikaattia, paitsi näytteistä 378 ja 387, jotka tutkittiin

vain kerran. Joidenkin näytteiden käsittelyssä ilmeni tulosten tarkastelun yhteydessä tyhjiä mittauksia, johtuen todennäköisesti siitä, että PCR-levyn valmistelussa näyte unohdettiin pipetoida. Tämän takia osasta näytteitä ei saatu yhteensä kuutta mittausta.

Eksoni 16 koesarjan jälkeen näytteet 10, 32, 35, 161 202, 205, 207 ja 214 analysoitiin eksoni 8 rs2074840 ja eksoni/introni 4 rs77559646 alukkeilla ja koettimilla. Jos tietyn näytteen tulos oli, että näytteessä esiintyy vain toinen alleeli, ei katsottu tarpeelliseksi tutkia näytettä useampaa kertaa. Heterotsygoottiset näytteet tutkittiin kolme kertaa, jossa jokaisessa koesarjassa oli jokaisesta heterotsygoottisesta näytteestä kaksi duplikaattia. Kuten edellä, joidenkin näytteiden käsittelyssä ilmeni tulosten tarkastelun yhteydessä tyhjiä mittauksia, johtuen todennäköisesti siitä, että näyte unohtui pipetoida PCR-levylle. Kyseisistä näytteistä ei saatu tämän takia kuutta mittausta.

DdPCR-mittauksien jälkeen tarkasteltiin saatavilla olevaa tietoa tutkittavien SNP-emästen eri alleelien ilmenemisestä tutkittavissa näytteissä alleelien suhteiden määrittämistä varten. Tiedot perustuivat aikaisemmin tehtyihin sekvensointeihin. Kun alleelien väliset suhteet oli sekvensointidatan perusteella määritetty, verrattiin näitä tuloksia ddPCR-menetelmän antamiin tuloksiin.

Näytteen 35 tiedot oli kerätty mRNA-kirjaston datasetistä, jonka sekvensointikattavuus (coverage) oli noin 26–32 miljoonaa luentaa, ja kaikkien muiden näytteiden tiedot oli kerätty totaali RNA datasetistä, jonka sekvensointikattavuus oli näytteestä riippuen noin 32–54 miljoonaa luentaa (liite 5, taulukko 21). Muista näytteistä poiketen, näytteestä 35 oli sekvensoitu sekä benigniä, että syöpäkudosta.

3 Tulokset

3.1 Alamut® Visual -analyysin tulokset

Taulukko 8. Alamut® Visual -ohjelman antamat tulokset. Align GVGD -ohjelma antaa tuloksen numerona, jonka perusteella C0 häiritsee vähiten proteiinin toimintaa, ja C65 häiritsee proteiinin toimintaa merkittävästi. B = benigni, TH = todennäköisesti haitallinen, Pol. = harmiton polymorfismi, Kesk. = keskinkertainen, Ei kon. = ei konservoitumista.

Variantti	Poly-Phen-2 tulkin-ta	SIFT tulkin-ta	Mutati-on Tas-ter tul-kinta	Align GVGD tulkin-ta	Nukleotidin fylogeneettinen konservoitumi-nen	Aminohapon fylogeneettinen konservoitumi-nen
<i>BRCA1</i>						
c.230C>T	TH	B	TH	C0	Kesk.	Kesk.
c.3257T>C	TH	B	TH	C0	Heikko	Kesk.
c.3649T>C	B	B	Pol.	C0	Heikko	Kesk.
c.4676-8C>G	-	-	-	-	-	-
<i>BRCA2</i>						
c.68-7T>A	-	-	-	-	-	-
c.583T>G	TH	TH	Pol.	C0	Heikko	Kesk.
c.2884C>T	B	B	Pol.	C0	Ei kon.	Heikko
c.3314C>T	TH	B	TH	C0	Kesk.	Heikko
c.4756A>C	TH	B	Pol.	C0	Heikko	Heikko
c.5225A>T	TH	B	Pol.	C0	Heikko	Heikko
c.6821G>T	TH	B	TH	C0	Kesk.	Heikko
c.7121A>G	B	B	Pol.	C0	Ei kon.	Kesk.
c.7712A>G	TH	TH	TH	C0	Vahva	Vahva
c.7759C>T	TH	TH	TH	C15	Heikko	Vahva
c.7805+13A>G	-	-	-	-	-	-
c.8855T>C	B	TH	Pol.	C25	Vahva	Heikko
c.8953+24T>C	-	-	-	-	-	-
c.9232G>T	TH	TH	TH	C0	Kesk.	Vahva
c.9302T>C	TH	TH	TH	-	Kesk.	Kesk.

c.9411T>G	B	-	-	-	-	-
c.9872C>G	TH	TH	TH	C65	Vahva	Vahva
c.10234A>G	B	B	Pol.	C0	Ei kon.	Heikko

Taulukosta 8 näkee miten Alamut-ohjelman tulkinnat keskittyvät eksonisiin variantteihin, ohjelma ei antanut tuloksia neljästä intronisesta variantista. Monen variantin kohdalla esiintyi alaohjelmien tulkintojen välillä ristiriitoja siitä, onko emäsmuutoksen vaikutus todennäköisesti haitallinen vai harmiton. Ohjelman tuloksien perusteella 18:ta eksonisesta variantista neljän variantin tapauksessa kaikki ohjelmat tulkitsivat emäsmuutoksen todennäköisesti harmittomaksi. Näihin variantteihin kuuluivat *BRCA1* c.3649T>C, ja *BRCA2*-varianttien osalta variantit c.2884C>T, c.7121A>G, sekä c.10234A>G. Näiden varianttien nukleotidin ja aminohapon fylogeneettinen konservoituminen oli vain heikkoa tai keskivertoa. Variantti c.10234A>G on myös ClinVar-tietokannan asiantuntijapaneelin tarkastelun perusteella todettu hyvänlaatuiseksi, muiden varianttien ClinVar tulkinta on vielä epävarma.

Varianttien joukosta löytyi myös sellaisia, jotka näyttivät olevan kaikkien Alamut-alaohjelmien tulkintojen mukaan todennäköisesti haitallisia. Nämä variantit olivat *BRCA2* c.7759C>T sekä *BRCA2* c.9872C>G. Molempien varianttien ClinVar tulkinta on epävarma, ja molempien varianttien aminohapon fylogeneettinen konservoituminen on korkea. Lisäksi c.9872C>G variantin nukleotidin fylogeneettinen konservoituminen oli myös korkea.

Muista Alamut-alaohjelmista poiketen AlignGVGD tulkitsi vain kolme varianttia mahdollisesti haitalliseksi. AlignGVGD-ohjelma antaa emäsmuutokselle pistemäärän C0:sta C65:een, jossa pistemäärän kasvaessa mutaation aiheuttama häiritsevä vaikutus proteiinin toimintaan myös kasvaa. Vain kolme varianttia sai AlignGVGD-ohjelman mukaan C0:n ylittävän pistemäärän. Näihin variantteihin kuului edellä mainitut *BRCA2* c.7759C>T sekä *BRCA2* c.9872C>G variantit, sekä variantti *BRCA2* c.8855T>C, jossa muiden ohjelmien tulkinnat emäsmuutoksen vaikutuksesta olivat ristiriitaisia. Kaikki Alamut-analyysin tulokset on koottu yhteen muiden ohjelmien tuloksien kanssa liitteen 3 taulukkoon 20.

3.2 Human Splicing Finder -analyysin tulokset

Taulukko 9. Human Splicing Finder tulokset, jossa esitetään varianttisekvenssissä havaitut muutokset referenssisekvenssiin verrattuna. Sulkeissa muutoksen todenneiden eri algoritmien lukumäärä.

Variantti	Uusi luovuttajasekvenssi (new donor site)	Uusi eksoninen silmukoinnin hiltäjäsekvenssi (new ESS site)	Rikkoutunut eksoninen silmukoinnin vahvistajasekvenssi (new ESE site)
<i>BRCA1</i>			
c.230C>T	-	1 (2)	1 (3)
c.3257T>C	-	-	-
c.3649T>C	-	-	-
c.4676-8C>G	-	-	-
<i>BRCA2</i>			
c.68-7T>A	-	-	-
c.583T>G	1 (1)	1 (3)	-
c.2884C>T	-	1 (3)	-
c.3314C>T	-	-	1 (2)
c.4756A>C	-	-	1 (3)
c.5225A>T	-	-	1 (2)
c.6821G>T	-	-	1 (5)
c.7121A>G	1 (1)	1 (3)	-
c.7712A>G	-	1 (3)	1 (4)
c.7759C>T	1 (1)	-	-
c.7805+13A>G	-	1 (5)	1 (2)
c.8855T>C	-	-	1 (3)
c.8953+24T>C	-	-	-
c.9232G>T	-	1 (2)	-
c.9302T>C	-	-	1 (2)
c.9411T>G	1 (1)	-	-
c.9872C>G	2 (1)	1 (2)	1 (3)
c.10234A>G	-	-	1 (2)

Taulukosta 9 näkyy, että viiden variantin osalta Human Splicing Finder ei löytänyt mitään merkittäviä muutoksia normaalin ja varianttisekvenssin välillä. Kaikista yleisin muutos, jonka ohjelma havaitsi puolessa tutkituista varianttisekvensseistä, oli rikkoutunut ESE-sekvenssi. Uuden luovuttajasekvenssin ilmestyminen varianttisekvenssiin näyttää olevan harvinaisin muutos. Uuden luovuttajasekvenssin oli aina todennut vain yksi algoritmi, mutta ESS- ja ESE-sekvenssien muutosten tapauksessa usein muutoksen oli havainnut usea eri algoritmi. Ohjelma ei havainnut tutkituissa sekvensseissä lainkaan mahdollisesti rikkoutuneita ESS-sekvenssejä tai uusia ESE-sekvenssejä. Kaikki Human Splicing Finder analyysin tulokset on koottu yhteen muiden ohjelmien tuloksien kanssa liitteen 3 taulukkoon 20.

3.3 RBPmap -analyysin tulokset

Taulukko 10. RBPmap -analyysin tulokset, jossa esitetään varianttisekvenssissä ilmenneet muutokset proteiinien sitoutumiskohdissa referenssisekvenssiin verrattuna. Motif = sitoutumiskohta. Proteiinit on jaettu eri kategorioihin GeneCards tietokannan kuvauksien perusteella.

Variantti	Uusi silmukointi motif	Hävinnyt silmukointi motif	Uusi muu motif	Hävinnyt muu motif
<i>BRCA1</i>				
c.230C>T	-	SRSF3	-	-
c.3257T>C	-	HNRNPA1L2	-	HNRNPA1L2
c.3649T>C	-	SRSF3	-	MATR3
c.4676-8C>G	-	-	ZC3H14, RBM41	PCBP1, PCBP3
<i>BRCA2</i>				
c.68-7T>A	KHDRBS1, KHDRBS2	-	PABPC4	DAZAP1, RBMS1
c.583T>G	-	-	-	-
c.2884C>T	TRA2B	-	-	-
c.3314C>T	-	-	RBM42	HNRNPK
c.4756A>C	-	-	-	-
c.5225A>T	TRA2B	QKI, RBM42	G3BP2	QKI
c.6821G>T	-	HNRNPA2B1	-	HNRNPA2B1,

				LIN28A
c.7121A>G	-	-	-	PABPC5
c.7712A>G	RBM5	-	HNRNPA2B1	G3BP2
c.7759C>T	-	-	ZNF638	-
c.7805+13A>G	-	PABPC1, SART3	-	PABPC1, PABPC4, PUM2, SART3
c.8855T>C	-	-	-	-
c.8953+24T>C	-	HNRNPC, HNRNPCL1, RALY, U2AF2	-	CPEB2, CPEB4, TIA1
c.9232G>T	RALY, U2AF2	SFPQ, SRSF2	CPEB2, CPEB4, HNRNPC, HNRNPCL1, ZC3H14, ZNF638	-
c.9302T>C	-	HNRNPU	-	HNRNPU, PUM2
c.9411T>G	SRSF5	MBNL1, RBM41, U2AF2, SRSF3	-	A1CF, CPEB4
c.9872C>G	-	MBNL1, SRSF5, TARDBP	TIA1, ZC3H14	-
c.10234A>G	-	YBX1, ZCRB1, SRSF3	CNOT4	-

Taulukossa 10 esiintyvien proteiinien erottelu silmukointiin mahdollisesti vaikuttaviin ja muihin proteiineihin perustui GeneCards tietokannan tulkintoihin. Vaikka tietyissä varianteissa esiintyi paljon eroja referenssisekvenssin ja varianttisekvenssin välillä, millään listatulla proteiinilla ei havaittu *BRCA*-geeneihin, tai erityisesti rintasyöpään liittyviä ominaisuuksia, paitsi c.8953+24T>C ja c.9232G>T varianteissa havaitulla HNRPC-

proteiinilla, joka Anantha ym. (2013) mukaan saattaa säädellä *BRCA1*- ja *BRCA2*-geenien mRNA silmukointia, ja BRCA-proteiinien vuorovaikutuksia muiden proteiinien kanssa DNA-vaurioiden korjauksessa toimivissa nukleoproteiini komplekseissa. Kaikki RBPmap -analyysin tulokset on koottu yhteen muiden ohjelmien tuloksien kanssa liitteen 3 taulukkoon 20. Tarkempaa tietoa listalla olevien proteiinien ominaisuuksista löytyy liitteestä 4.

3.4 RegRNA 2.0 -analyysin tulokset

Taulukko 11. RegRNA 2.0 -analyysin tulokset, jossa esitetään varianttisekvenssissä ilmenneet muutokset referenssisekvenssiin verrattuna. TR. = transkriptiosäätelijäsekvenssi (transcriptional regulatory motif), ESE. = eksoninen silmukoinnin vahvistajasekvenssi (exon splicing enhancer motif), RBS = ribosomin sitoutumiskohta (ribosome binding site), ORF = avoin lukuraami (open reading frame).

Variantti	Uudet sitoutumiskohdat	Hävinneet sitoutumiskohdat	Muut muutokset
<i>BRCA1</i>			
c.230C>T	-	-	Hävinnyt RNA C-to-U editointikohta
c.3257T>C	-	GTF2IRD1-isoform 2 (TR.)	-
c.3649T>C	-	-	-
c.4676-8C>G	-	-	-
<i>BRCA2</i>			
c.68-7T>A	-	-	Uusi RBS ja ORF
c.583T>G	C/EBP (TR.)	-	-
c.2884C>T	aMEF-2 (TR.)	-	-
c.3314C>T	-	cd45 tyrosine phosphatase (ESE.)	-
c.4756A>C	-	-	Uusi hsa-miR-769-3p mikro-RNA-kohdesekvenssi
c.5225A>T	-	-	-
c.6821G>T	-	dystrophin gene	Hävinnyt hsa-miR-1273e

		exon 27 ja g-alpha-s exon 3 (ESE.)	mikro-RNA-kohdesekvenssi
c.7121A>G	-	-	-
c.7712A>G	MAF (TR.)	-	Hävinnyt RNA C-to-U editointikohta
c.7759C>T	-	-	-
c.7805+13A>G	-	Oct-1 (TR.)	-
c.8855T>C	Kid3 ja TEL1 (TR.)	-	-
c.8953+24T>C	-	TCF-3 (TR.)	-
c.9232G>T	-	-	Hävinnyt RNA C-to-U editointikohta
c.9302T>C	-	PLZF (TR.)	-
c.9411T>G	ATF2 (TR.)	Cdx-1 (TR.)	-
c.9872C>G	-	-	-
c.10234A>G	-	-	-

Kuten taulukossa 11 näkyy, RegRNA-ohjelma havaitsi monia muutoksia varianttisekvenssissä referenssisekvenssiin verrattuna, mutta millään havaitulla transkriptiotekijällä tai eksonisella silmukoinnin vahvistajalla ei havaittu *BRCA*-geeneihin tai rintasyöpään liittyviä ominaisuuksia Genecards tietokannan perusteella. Kaikki RegRNA -analyysin tulokset on koottu yhteen muiden ohjelmien tuloksien kanssa liitteen 3 taulukkoon 20. Tarkempaa tietoa listalla olevien proteiinien ja RNA-molekyylien ominaisuuksista löytyy liitteestä 4.

3.5 ANO7 ddPCR-mittausten tulokset

Taulukko 12. Eksoni 16 rs7593101 -variantin RNA sekvensointidatan ja ddPCR-tulosten vertailu. Sekvensointituloksissa tähdellä (*) merkittyjen näytteiden yksittäisten eri alleelien esiintyminen oletetaan virhelähteeksi, ja suhteen määrittämisessä tätä alleelia ei ole huomioitu.

Näyte ja kudostyyppi	Sekvensointitulokset, luentojen määrä (normaalialleeli C variantialleeli T)	Variantin suhde referenssialleeliin RNA sekvensoinnin perusteella	Variantin suhde referenssialleeliin suhde ddPCR-tuloksien perusteella
10 normaali	C19 T13	0.68	1,1 σ 0,28
32 syöpä	C78 T85	1.09	0,98 σ 0,06
35 benigni	C181 T184	1,02	-
35 syöpä*	C240 T316 A1*	1,32	Vain referenssi
161 syöpä	C91 T31	0.34	0,34 σ 0,05
202 syöpä	C292 T352	1.2	1,15 σ 0,06
205 syöpä	C127 T173	1,36	0,98 σ 0,08
207 syöpä	C77 T91	1.18	1,48 σ 0,17
214 syöpä	C48 T31	0.65	0,88 σ 0,21

Tulosten perusteella näytteen 35 sekvensoinnin antama tulos ei vastaa ddPCR-tulosta lainkaan. Taulukosta näkee, miten näytteiden 161 ja 202 sekvensointi- ja ddPCR-tulokset vastaavat hyvin toisiaan, kun taas näytteissä 10 ja 205 tulokset eroavat toisistaan merkittävästi. Näytteiden 10, 207 ja 214 ddPCR-mittauksien keskihajonnat ovat myös merkittävästi muita näytteitä suurempia. Tarkemmat tiedot ddPCR-mittauksien yksityiskohdista löytyvät liitteen 5 taulukosta 21.

Taulukko 13. Eksoni 8 rs2074840 -variantin RNA sekvensointidatan ja ddPCR-tulosten vertailu. Sekvensointituloksissa tähdellä (*) merkittyjen näytteiden yksittäisten eri alleelien esiintyminen oletetaan virhelähteeksi, ja suhteen määrittämisessä tätä alleelia ei ole huomioitu.

Näyte ja kudostyyppi	Sekvensointitulokset, luentojen määrä (normaalialleeli C variantialleeli T, muu alleeli G)	Variantin suhde referenssialleeliin RNA sekvensoinnin perusteella	Variantin suhde referenssialleeliin suhde ddPCR-tuloksien perusteella
10 normaali	C21 T20	0,95	1,56 σ 0,65
32 syöpä	T177	Vain variantti	Vain variantti
35 benigni	T194	Vain variantti	-
35 syöpä *	C2 T215*	Vain variantti	1,2 σ 0,06
161 syöpä*	C1 T126*	Vain variantti	Vain variantti
202 syöpä	C301 T430	1,43	1,45 σ 0,09
205 syöpä*	G1 T431*	Vain variantti	Vain variantti
207 syöpä	T193	Vain variantti	Vain variantti
214 syöpä	T69	Vain variantti	Vain variantti

Tulosten perusteella näytteen 35 sekvensoinnin antama tulos ei vastaa ddPCR-tuloksia. Taulukosta näkee hyvin, miten homotsygoottisten näytteiden tapauksessa, sekvensointi

ja ddPCR pystyvät molemmat helposti toteamaan toisen alleelin puuttumisen, olettaen että näytteiden 161 ja 205 sekvenssidatassa on virheellisesti määritetty yksittäinen havainto. Vain näytteen 10 sekvensointi ja ddPCR-tuloksissa on havaittavissa merkittävä ero, lisäksi näytteen ddPCR-mittauksien keskihajonta on erittäin suuri. Tarkemmat tiedot ddPCR-mittauksien yksityiskohdista löytyvät liitteen 5 taulukosta 22.

Taulukko 14. Eksoni/introni 4 rs77559646 -variantin RNA sekvensointidatan ja ddPCR-tulosten vertailu.

Näyte ja kudostyyppi	Sekvensointitulokset, luentojen määrä (normaalialleeli G varianttialleeli A)	Referenssin suhde varianttialleeliin RNA sekvensoinnin perusteella	Referenssin suhde varianttialleeliin suhde ddPCR-tuloksien perusteella
10 normaali	G1 A3	0,33	0,55 σ 0,06
32 syöpä	G8	Vain referenssi	Vain referenssi
35 benigni	G47	Vain referenssi	-
35 syöpä	G50	Vain referenssi	Vain referenssi
161 syöpä	G10	Vain referenssi	Vain referenssi
202 syöpä	G26 A19	1,37	0,41 σ 0,04
205 syöpä	G21	Vain referenssi	Vain referenssi
207 syöpä	G5 A9	0,56	0,29 σ 0,04
214 syöpä	G10	Vain referenssi	Vain referenssi

Kahdesta muun variantin tuloksista poiketen, tällä variantilla näytteen 35 sekvensointi- ja ddPCR-tulokset vastaavat toisiaan. Taulukosta <14 näkyy miten myös tämän variantin osalta molemmat menetelmät pystyvät toteamaan homotsygoottiset näytteet helposti. Toisaalta sekvensointi ja ddPCR-tulokset eroavat huomattavasti toisistaan heterotsygoottisten näytteiden osalta, ja monella näistä näytteistä, kuten 10 ja 207, ei sekvensointi ole tuottanut merkittävää määrää havaintoja alleelien välisen suhteen tarkkaa määrittämistä varten. Tarkemmat tiedot ddPCR-mittauksien yksityiskohdista löytyvät liitteen 5 taulukosta 23.

4 Pohdinta ja tulosten tarkastelu

4.1 *BRCA1* ja *BRCA2* -tutkimuksen ongelmakohdat

BRCA1- ja *BRCA2*-variantteja ei tutkittu ddPCR-menetelmällä lainkaan, sillä tutkimukseen ei saatu odotettua määrää potilasnäytteitä. *BRCA1*- ja *BRCA2*-variantit ovat tyypillisesti sukukohtaisia variantteja, joten siksi erilaisia tietyn tyypin variantteja esiintyy määrällisesti vähän, ja toisaalta suvutkin ovat nykyään tyypillisesti pieniä, joten siksikin saman variantin kantajia on vähän. Taloudellisista syistä, ei tutkimukseen tilattu alukkeita ja koettimia vain yhden tai kolmen potilasnäytteen tutkimista varten. Tämän takia tutkimus rajoittui *BRCA1*- ja *BRCA2*-varianttien osalta *in silico* -analyysiin.

Variantteja olisi voitu tutkia valmiiksi suunnitelluilla toimiviksi todetuilla kaupallisilla ”genotyping assay” -alukkeilla ja koettimilla, mutta näiden ongelmaksi nousi se, ettemme saaneet tietoa siitä, mihin alukkeet sitoutuvat kohdesekvenssissä. Tarkoituksena oli tutkia RNA-ekspressiota, joten alukkeet, jotka olisivat sitoutuneet introniseen sekvenssiin, eivät olisi tutkimuksessa toimineet. Tämän takia päätimme suunnitella kaikki alukkeet ja koettimet itse, ohjeita noudattaen.

Suurin ongelma alukkeiden ja koettimien suunnittelussa oli koetinsekvenssin liian alhainen sulamislämpötila, joka johtui AT-rikkaista sekvensseistä. Tämä johti monen koettimen tapauksessa siihen, että Primer3plus-ohjelman laskema ennakoitu sulamislämpötila oli liian alhainen. Ohjelma myös ilmoitti monen koettimen kohdalla ongelmaksi, että koettimella oli riski sitoutua epäspesifiseen sekvenssiin. Tämä ohjelman antama varoitus oletettiin harmittomaksi, johtuen tutkimuksen keskittymisestä RNA-tason tutkimiseen, ja siksi, että PCR-syklin jälkeen näytteessä tulisi olemaan merkityksettömän pieni määrä mahdollista epäspesifistä cDNA:ta, johon koetin voisi sitoutua. Lisäksi tutkimuksessa tutkittiin yksittäisiä emäksiä koettimilla, jotka suurilta osin erosivat toisistaan vain tutkittavan emäksen kohdalla. Tämä itsessään jo johtaa suureen ristireagoinnin mahdollisuuteen koettimien välillä, jonka käyttämämme Kamitakin ym. (2018) protokolla otti huomioon ohjeistuksessaan.

Alun perin *BRCA1*- ja *BRCA2*-varianttien osalta oli tarkoitus rajata tutkimukseen variantit täysin *in silico* -analyysien perusteella, mutta analyysien tuloksien valossa ei vari-

anttien joukosta löytynyt mitään merkittävästi muista varianteista eroavia vaihtoehtoja. Tästä syystä alukesuunnitteluun valittiin variantit näytesaatavuuden, ClinVar tietokannan tulkinnan, sekä Alamut-ohjelman tuloksien perusteella.

BRCA1- ja *BRCA2*-variantit olisi mahdollisesti voitu tutkia käyttämällä hyväksi yleisiä heterotsygoottisia SNP:jä, joiden avulla geenin eri alleelit olisi voitu erottaa toisistaan. Tätä menetelmää ei kuitenkaan käytetty, sillä pienen näytämäärän takia todennäköisyys tietyn yleisen SNP:n esiintymisestä tutkimuspopulaatiossa todettiin liian epävarmaksi. Tutkimuksen kannalta ei katsottu kannattavaksi ostaa mahdollisesti tarpeettomia ja toimimattomia alukkeita ja koettimia. Lisäksi tällä menetelmällä emme olisi saaneet tarkkaa tietoa alleelien ekspression suhteista, sillä meillä ei olisi ollut tietoa siitä, kumpi alleeli esiintyy normaalissa ja kumpi VUSia kantavassa kromosomissa. Tätä varten olisimme tarvinneet potilaiden sekvenssitietoja, mutta emme saaneet näitä tietoja käyttöömmä tutkimusasetelman puitteissa.

4.2 *BRCA1* ja *BRCA2* Alamut tuloksien tulkinta ja johtopäätökset

BRCA1- ja *BRCA2*-varianttien analyysi Alamut-ohjelmalla johti vaihteleviin tuloksiin ja tulkintoihin, kuten taulukossa 8 näkyy. Monen variantin kohdalla eri alaohjelmat olivat tulkinneet emäsmuutoksen mahdollisen vaikutuksen ristiriitaisesti, mutta joukosta löytyi myös variantteja, joilla kaikki ohjelmat olivat tulkinneet variantin haitalliseksi tai harmittomaksi. Variantti *BRCA1* c.3649 T>C, ja *BRCA2*-varianttien osalta variantit c.2884C>T, c.7121A>G sekä c.10234A>G olivat sellaisia variantteja, jotka kaikki alaohjelmat olivat tulkinneet mahdollisesti harmittomiksi ja c.10234A>G variantti on myös ClinVar tietokannan asiantuntijapaneelin tarkastelun perusteella todettu hyvänlaatuiseksi. Näiden tulosten perusteella ei voi kuitenkaan todeta kolmea muuta Alamut tulkintojen mukaan harmitonta, mutta ClinVarin mukaan epävarmaa varianttia varmasti harmittomiksi, vaan varianteista on saatava tarkempaa tutkimustietoa.

Ernstin ym. (2018) *BRCA1*- ja *BRCA2*-varianttien tietokoneanalyysissä todettiin, että Alamut-ohjelman SIFT- ja MutationTaster-alaohjelmat tulkitsivat mahdollisesti patogeeniset variantit erittäin harvoin harmittomiksi. Kun tutkimustuloksia eksonisten *BRCA1*- ja *BRCA2*-varianttien Alamut tuloksia tarkastellaan, käy kuitenkin ilmi, että 18:ta variantista kuuden tapauksessa ohjelmien antama tulkinta variantin mahdollisesta

vaikutuksesta on ristiriitainen. Jos Ernstin ym. tutkimustulosta voisi pitää varmana, kuusi varianttia voisi jättää SIFT ja MutationTaster tulkintojen perusteella pois mahdollisesti patogeenisten varianttien joukosta. Näihin mahdollisesti harmittomiin varianteihin kuuluivat *BRCA1* c.3649T>C ja *BRCA2*-varianteista c.2884C>T, c.4756A>C, c.5225A>T, c.7121A>G ja c.10234A>G. Aikaisemmista tutkimuksista huolimatta näiden varianttien harmittomuutta ei voi todeta pelkästään SIFT ja MutationTaster tulkinnoilla, vaan varianttien todellinen vaikutus on selvitettävä jatkotutkimuksilla.

Varianttien joukosta löytyi myös sellaisia, jotka olivat kaikkien Alamut-alaohjelmien mukaan mahdollisesti haitallisia. Näitä olivat *BRCA2*-variantit c.7759C>T sekä c.9872C>G. Muista alaohjelmista poiketen AlignGVGD oli tulkinnut vain kolme varianttia mahdollisesti haitalliseksi, ja näihin varianteihin lukeutui edellä mainittujen lisäksi myös *BRCA2* c.8855T>C. AlignGVGD-ohjelman harvat haitalliset tulkinnat saattavat selittyä sillä, että ohjelma tutkii vain variantissa olevan muuttuneen aminohapon ominaisuuksia normaalisekvenssin aminohappoon verrattuna, kun taas muut ohjelmat ottavat huomioon myös esimerkiksi proteiinin kolmiulotteisen rakenteen, variantin mahdolliset vaikutukset silmukointiin ja säätelyalueihin, sekä emäksen fylogeneettisen konservoitumisen.

Jos AlignGVGD tulokset jätetään huomiotta sen perusteella, ettei ohjelma näyttä muihin ohjelmiin verrattuna olevan yhtä herkkä haitallisten muutoksien toteamisessa, kolmen muun ohjelman tulkintojen perusteella myös *BRCA2*-variantit c.7712A>G, c.9232G>T, sekä c.9302 T>C ovat kaikkien ohjelmien tulkintojen perusteella mahdollisesti haitallisia. Varianttien c.7712A>G ja c.9302 T>C ClinVar tulkinta on epävarma ja variantilla c.9232G>T ei ole vielä edes omaa ClinVar sivua. Kaikkien kolmen variantin nukleotidin ja aminohapon fylogeneettinen konservoituminen oli vähintään keskivertoa ja variantilla c.7712A>G molemmat olivat korkeita.

Analyysien perusteella 18:ta eksonisesta variantista viisi saattavat olla haitallisia proteiinin toiminnalle Alamut-ohjelmien tulkintojen mukaan, jos Align-GVGD tulkinnat jätetään huomiotta. Nämä variantit ovat *BRCA2* c.7712A>G, c.7759C>T, c.9232G>T, c.9302 T>C ja c.9872C>G. Nämä variantit saattavat olla hyviä ehdokkaita jatkotutkimuksiin, varsinkin niiden osalta, joista ei ole vielä kattavaa tietoa ClinVar tietokannassa.

4.3 Mahdollisten *BRCA1*- ja *BRCA2*-varianttien silmukointimuutosten tarkastelu *in silico* -analyysien tuloksien perusteella

4.3.1 Silmukointiin tarvittavat sekvenssit RNA juosteessa

Silmukointia varten transkriptin introneista löytyy kolme splice site -sekvensseiksi kutsuttua silmukointiprosessille kriittistä rakennetta: intronin 5' -päässä oleva GT luovuttaja-/donori-sekvenssi (donor site), intronin 3' -päässä oleva AG vastaanottaja-/akseptori-sekvenssi (acceptor site), sekä 18–40 nukleotidia ylävirtaan vastaanottaja sekvenssistä sijaitseva polypyrimidiini alue ja branch point -emäs, joka on aina adeniini. GT-AG luovuttaja-vastaanottaja -sekvenssien lisäksi on olemassa myös harvinaisempia silmukointisekvenssejä kuten GC-AG rakenne, mutta Bursetin ym. (2001) nisäkkäiden silmukointisekvenssien tarkastelun perusteella lähes kaikki (98.71%) analysoiduista silmukointisekvensseistä ovat GT-AG tyyppiä. Silmukointiin tarvittavat proteiinit sitoutuvat näihin luovuttaja-, vastaanottaja- ja branch point -sekvensseihin, ja jos sekvenssit muuttuvat liikaa, se saattaa johtaa silmukoinnin häiriöihin ja mahdollisesti johonkin sairauteen.

Abramowiczin ja Gosin (2018) mukaan mutaatiot voivat vaikuttaa monella eri tavalla silmukointiin riippuen mutaation paikasta transkriptissa. Splice site -sekvenssit ovat lyhyitä ja tämän takia transkriptista saattaa löytyä myös kryptisiä splice site -sekvenssejä, jotka toimivat voimakkaampina kohteina silmukointiproteiineille oikeaan splice site -sekvenssiin verrattuna. Yleensä nämä kryptiset sekvenssit eivät kuitenkaan aiheuta silmukointimuutoksia, koska sekvenssistä löytyy usein myös silmukointia edistäviä voimistaja (enhancer) -sekvenssejä ja silmukointia estäviä vaimentaja-/hiljentäjä (silencer) -sekvenssejä, joihin tietyt silmukointia säätelevät proteiinit voivat sitoutua, ja näin säädellä kryptisten splice site -sekvenssien vaikutuksia. Mutaatiot saattavat heikentää normaalin splice site -sekvenssin voimakkuutta, tai luoda uuden voimakkaan splice site -sekvenssin, johtaen silmukoinnin mahdollisiin häiriöihin.

Abramowiczin ja Gosin (2018) mukaan luovuttaja- ja vastaanottajasekvenssien mutaatioista luovuttajasekvenssien mutaatiot ovat yleisempiä kuin vastaanottajasekvenssien mutaatiot. Jos splice site -sekvenssi on heikko ja mutaatio aktivoi voimakkaamman

kryptisen splice site -sekvenssin viereisessä eksonissa tai intronissa, saattaa mutaatio tällöin johtaa intronin säilymiseen transkriptissa, tai eksonisen fragmentin häviämiseen, riippuen mutaation sijainnista. Mutaation vaikutus riippuu myös muusta ympäröivästä sekvenssistä, ja mahdollisista voimistaja- ja vaimentajasekvensseistä ja niiden tiheydestä, sekä mahdollisesti muuttuneesta transkriptin sekundaarirakenteesta. Tästä johtuen samankaltainen mutaatio eri eksonien luovuttajasekvensseissä saattaa tapauksesta riippuen johtaa intronin säilymiseen, tai eksonin häviämiseen. Splice site -mutaatioiden vaikutukset saattavat olla myös kudskohtaisia, riippuen eri säätelyproteiinien esiintymisestä eri solutyypeissä. Luovuttaja- ja vastaanottajasekvenssien mutaatioiden lisäksi, voivat mutaatiot luoda transkriptiin uuden, tai häiritä olemassa olevaa eksonista voimistaja- tai vaimentajasekvenssiä. Varsinkin eksonisen silmukoinnin voimistajan (ESE, exonic splicing enhancer) häiriöt saattavat johtaa eksonin häviämiseen transkriptissa.

Introniset mutaatiot saattavat puolestaan johtaa Abramowiczin ja Gosin (2018) mukaan niin sanottujen pseudoeksonien säilymiseen transkriptissa, jolloin osa intronia säilyy transkriptissa. Tällöin mutaatio luo introniin kryptisen luovuttaja- tai vastaanottajasekvenssin, joka toimii silmukointiprosessissa normaalien luovuttaja- ja vastaanottajasekvenssien kanssa. Esimerkkejä tällaisista mutaatiosta ovat Santzin ym. (2017) tarkastelemat mutaatiot *CFTR*-geenistä, jossa c.3718-2477C>T ja c.3718-2477C>T introniset mutaatiot johtavat intronisen sekvenssin säilymiseen transkriptissa. Nämä introniset sekvenssit sisältävät ennenaikaisen lopetuskodonin, joka johtaa normaalia lyhyemmän ja tämän takia funktionaalisesti toimimattoman proteiinin synteisiin.

Eksonien oikeaan tunnistamiseen ei riitä vain splice site -sekvenssit, vaan tähän tarvitaan myös eksonisia voimistaja-, eli ESE-sekvenssejä (ESE, exonic splicing enhancer) ja eksonisia vaimentaja, eli ESS-sekvenssejä (ESS, exonic splicing silencer). ESE- ja ESS-sekvenssit ovat myös tärkeitä vaihtoehtoisessa silmukoinnissa. Zhun ym. (2001) mukaan ESE-sekvenssit edesauttavat silmukointiproteiinien kerääntymistä tietylle alueelle. Näille sekvensseille sitoutuvat SR-proteiinit, jotka edesauttavat muiden silmukointiin tarvittavien proteiinien kuten U1- ja U2AF-proteiinien kerääntymistä silmukointikompleksin muodostamista varten. ESS-sekvenssit puolestaan estävät silmukointiproteiinien kerääntymistä tietylle alueelle, ja näin estävät tiettyjen silmukointisekvenssien käytön silmukointiprosessin yhteydessä, johtaen intronin säilymiseen, tai tietyn

eksonin silmukointiin pois transkriptista. Näiden sekvenssien tarkkaa toimintatapaa ei vielä tiedetä, mutta ainakin hnRNP-proteiinien on todettu sitoutuvan näille sekvensseille, estäen silmukointia edistävien proteiinien sitoutumisen lähellä oleville eksonisille voimistajasekvensseille.

4.3.2 *BRCA1*- ja *BRCA2*-varianttien silmukointimuutosten tarkastelu

Kun Human Splicing Finder tuloksia tarkastellaan, niin nähdään, että 11:ta variantilla on ohjelma todennut rikkoutuneen ESE-sekvenssin. Monilla varianteilla esiintyi paljon muutoksia silmukointiin liittyvien proteiinien sitoutumiskohdissa, mutta juuri ESE-sekvensseihin sitoutuvien SR-proteiinien sitoutumiskohdat ovat tarkastelun kannalta mielenkiintoisia. Näistä varianteista vain kahdella RBPmap-ohjelma havaitsi varianttisekvenssistä hävinneen SR-proteiini sitoutumiskohdan. *BRCA1* c.230C>T -variantista hävisi SRSF3 sitoutumiskohta ja *BRCA2*-variantista c.9872C>G hävisi SRSF5 sitoutumiskohta. Nämä tulokset ovat yhteensopivia toistensa kanssa, mutta varianttien joukosta löytyi myös variantti *BRCA2* c.9232G>T, joilla Human Splicing Finder ei todennut rikkoutunutta ESE-sekvenssiä, mutta RBPmap totesi silti hävinneen SRSF2-proteiinin sitoutumiskohdan. Lisäksi varianttien joukosta löytyi mielenkiintoinen *BRCA2* c.9411T>G -variantti, jossa ei Human Splicing Finderin mukaan löydy rikkinäistä ESE-sekvenssiä, mutta RBPmap tuloksien mukaan variantista on hävinnyt SRSF3 sitoutumiskohta, mutta myös ilmestynyt uusi SRSF5 sitoutumiskohta. Se miten tämä SR-proteiinin sitoutumiskohdan mahdollinen muutos SRSF3:sta SRSF5:een vaikuttaisi silmukointiin, on epäselvä, sillä molemmat ovat SR-proteiineja. Voi olla mahdollista, että väärän SR-proteiinin sitoutuminen väärään kohtaan transkriptia, saattaa vaikuttaa silmukointiin häiritsevällä tavalla. Variantin RBPmap-analyysi tosin havaitsi tosin myös monen muun silmukointiin liittyvän proteiinin sitoutumiskohtien muutoksia, joten tämä SR-proteiinin vaihtuminen ei ole ainoa mahdollinen silmukointia häiritsevä tekijä.

Human Splicing Finder totesi myös kahdeksan variantin tapauksessa varianttisekvenssissä uuden ESS-sekvenssin. Nämä sekvenssit ovat ESE-sekvenssien lisäksi erityisen mielenkiintoisia, sillä ne ovat todennäköisimmin silmukointia estävien hnRNP-proteiinien sitoutumiskohtia. Kahdeksasta variantista, joilla Human Splicing Finder totesi uuden ESS sitoutumiskohdan, vain *BRCA2* c.7712A>G -varianttiin ilmestyi uusi HNRNPA2B1 sitoutumiskohta ja *BRCA2* c.9232G>T -varianttiin ilmestyvät uudet HNRNPC- sekä HNRNPCL1-proteiinien sitoutumiskohdat. Näistä HNRNPC sitoutu-

miskohdan ilmestyminen on erityisen mielenkiintoinen, sillä Ananthan ym. (2013) mukaan proteiini saattaa säädellä *BRCA1*- ja *BRCA2*-geenien mRNA silmukointia, ja BRCA-proteiinien vuorovaikutuksia muiden proteiinien kanssa DNA vaurioiden korjauksessa toimivissa nukleoproteiini komplekseissa.

Näiden lisäksi varianttien joukosta löytyi myös sellaisia, joilla hävisi hnRNP-sitoutumiskohta. Näihin variantteihin lukeutui *BRCA2*-variantit c.3314C>T josta hävisi HNRNPK sitoutumiskohta, c.6821G>T josta hävisi HNRNPA2B1 sitoutumiskohta, c.8953+24T>C, josta hävisi HNRNPC sitoutumiskohta, sekä c.9302T>C josta hävisi HNRNPU sitoutumiskohta. C.8953+24T>C variantin kohdalla Human Splicing Finder ei ollut todennut mitään silmukointimuutoksia, mutta muiden varianttien osalta ohjelma oli todennut vain rikkoutuneen ESE-sekvenssin.

Mahdollisten uusien luovuttajasekvenssien (donor site) vaikutuksia on vaikea arvioida, koska uuden luovuttajasekvenssin voimakkuus ei todennäköisesti ole suurempi kuin oikean 5' -luovuttajasekvenssin. Human Splicing Finder vain toteaa uuden mahdollisen luovuttajasekvenssin, mutta ei vertaa tämän sekvenssin voimakkuutta jo olemassa olevaan luovuttajasekvenssiin intronin 5' -päässä. Jos uusi luovuttajasekvenssi olisi voimakkaampi kuin normaali luovuttajasekvenssi, voisi olla mahdollista, että osa eksonin 3' -pään sekvenssistä jäisi transkriptista pois, kun silmukointikoneisto valitsee väärän sitoutumiskohdan. Tämä saattaisi mahdollisesti johtaa esimerkiksi lukuraamin muutokseen, tai proteiinin toiminnalle kriittisen domeenin rakenteen häiriöön.

Tuloksien perusteella on vaikea sanoa, mitkä emäsvariantit saattavat vaikuttaa silmukointiin merkittävästi, sillä eri ohjelmat analysoivat sekvenssejä eri menetelmillä, ja eri proteiinien tarkkoja vuorovaikutuksia *BRCA1*- ja *BRCA2*-transkriptien kanssa ei tiedetä.

4.4 Muiden *BRCA1* ja *BRCA2* *in silico* -analyysien tulosten tarkastelu

4.4.1 Havaitut ribosomin sitoutumiskohdat, transkriptiotekijät, ESE-sekvenssit ja muut muutokset

Silmukointimuutosten lisäksi *in silico* -analyysit havaitsivat myös muita erilaisia muutoksia varianttisekvensseissä, mutta nämä muutokset saattavat olla harmittomia. Esimerkiksi *BRCA2* c.68-7T>A -variantin RegRNA-analyysin toteama uusi ribosomin sitoutumiskohta (RBS) ja avoin lukuraami (ORF) saattavat olla harmittomia ohjelman havaitsemia muutoksia. Vaikka transkriptiin ilmestyisi uusi mahdollinen ribosomin sitoutumiskohta ja avoin lukuraami, se ei johtaisi ribosomin sitoutumiseen tähän vaihtoehtoiseen sitoutumiskohtaan, sillä sekvenssin ympäristöstä puuttuvat muut translaatioon tarvittavat tunnustussekvenssit, kuten 5' UTR-alue. Tämän lisäksi muutos on intronissa, eli ribosomi ei edes pystyisi tunnistamaan sekvenssiä, sillä sitä ei intronin pois silmukoinnin jälkeen enää olisi transkriptissa. Voi tosin olla, että havaittu muutos olisi uuden sisäisen ribosomin sitoutumiskohdan (internal ribosome entry site) syntyminen. Hellenin ja Sarnowin (2001) katsauksen mukaan tällaisia sekvenssejä on havaittu stressinsietoon ja apoptoosiin liittyvien geenien mRNA-molekyyleillä. On myös mahdollista, että ohjelma tulkitsee tämän muutoksen ribosomin sitoutumiskohdaksi ja uudeksi avoimeksi lukuraamiksi vain sen takia, että aminohappo on muuttunut metioniiniksi, ja ohjelma toteaa pelkästään tämän perusteella, että uusi ribosomin sitoutumiskohta on syntynyt.

RegRNA-ohjelma löysi sekvensseistä myös paljon erilaisia ESE-sekvenssejä, ja osalla varianteista löytyi eroja näiden sekvenssien ilmenemisessä normaalin- ja varianttisekvenssin välillä. Näitä havaintoja on vaikea tulkita, sillä ohjelma esittää sekvenssin sillä nimellä, mistä geenistä tämä kyseinen sekvenssi on aikaisemmin löytynyt. Monet geenit todennäköisesti käyttävät näitä samoja ESE-sekvenssejä silmukoinnissa, mutta tämän tutkielman puitteissa näiden mahdollisten muutosten vaikutuksia *BRCA*-transkripteihin on vaikea arvioida.

Sekä RBPmap että RegRNA löysivät silmukointiin liittyvien proteiinien lisäksi eroja transkriptiotekijöiden ja muiden proteiinien sitoutumiskohdissa, mutta näiden muutoksien merkityksiä ei pysty aina varmasti toteamaan. Näiden kyseisten proteiinien tarkas-

telussa ei löytynyt yhtäkään *BRCA1*- tai *BRCA2*-spesifistä proteiinia ja suurimmalla osalla proteiineista ei ollut yhteyttä syöpään. Tämä ei tarkoita, etteikö joillakin näistä proteiineista voisi olla vuorovaikutuksia *BRCA*-geenien kanssa.

Mistä tahansa emäsekvenssistä löytyy sattumalta todennäköisesti joitain eri proteiinien mahdollisia sitoutumissekvenssejä. Kaikista tutkituista sekvensseistä löytyi aina useita erilaisia mahdollisia sitoutumiskohtia, joten todennäköisyys sille, että mutaation aiheuttama satunnaisen proteiinin sitoutumiskohdan ilmestyminen tai häviäminen normaalsekvenssiin verrattuna aiheuttaisi merkittäviä haitallisia vaikutuksia on epävarma.

4.4.2 Havaitut mikro-RNA ja RNA-editointisekvenssit

RNA-editointi on ilmiö, jossa valmiin transkriptin sekvenssi muuttuu eri entsyymien toimesta transkription jälkeen, muuttamatta ympäröivää RNA-sekvenssiä. Yksi tällainen RNA-editoinnin muoto on C-to-U RNA editointi. Vun ja Tsukuharan (2017) katsauksessa mainitaan, että C-to-U RNA -editointisekvenssejä on löytynyt monien transkriptien 3' UTR-alueilta, varsinkin luuytimen makrofageista. Tämän lisäksi artikkelissa mainitaan myös apolipoproteiini B:n (apoB) vaihtoehtoinen editointi maksassa ja ohutsuolessa, jossa nukleotidi 6666:den muutos C:stä U:ksi luo transkriptiin enneaikaisen lopetuskodonin, ja johtaa erilaisten transkriptien esiintymiseen maksakudoksessa ja ohutsuolessa. ApoB:n RNA-editointia on pidetty pitkään ainoana konkreettisena esimerkkinä C-to-U editoinnista ihmisellä, mutta esimerkiksi Sharman ym. (2015) tutkimuksessa on löytynyt näyttöä satojen eri geenien C-to-U editoinnista APOBEC3A-proteiinin toimesta makrofagien M1-polarisaatiossa ja monosyyttien altistuessa interferoneille ja hypoksialle.

RegRNA löysi lähes kaikista sekvensseistä mahdollisia C-to-U RNA editoinnin kohtia, mutta *BRCA1*-variantilla c.230C>T sekä *BRCA2*-varianteilla c.7712A>G ja c.9232G>T varianttisekvenssistä hävisi tällainen sekvenssi varianttiemäksen kohdalla, kun varianttisekvenssiä verrattiin normaaliin sekvenssiin. C-to-U RNA-editoinnin täyttä merkitystä ihmiselle ei vielä tunneta, vaikutukset saattavat olla kudosis- sekä tilannekohtaisia kuten Sharman ym. (2014) tutkimuksessa, ja kaikkia editointiin osallistuvia eri proteiineja ja niiden käyttämiä tunnistussekvenssejä ei välttämättä vielä tiedetä. *BRCA*-geenien mahdollisesta RNA editoinnista ei ole tutkimustietoa, mutta Asaokan ym. (2019) tutkimuk-

sessä on havaittu, että APOBEC3-välitteistä RNA editointia esiintyy rintasyöpäkasvaimissa, ja lisääntyneellä editoinnin määrällä on havaittu yhteys korkeampaan immuunisolujen aktiivisuuteen ja potilaan selviytymiseen.

RegRNA havaitsi kahdella *BRC A2*-variantilla myös mikro-RNA kohdesekvenssien muutoksia, mutta myöskään näillä ei mahdollisesti ole funktionaalisia vaikutuksia. Varianttiin c.4756A>C ilmestyi uusi hsa-miR-769-3p mikro-RNA:n kohdesekvenssi. Xian ym. (2019) ovat havainneet tämän mikro-RNA:n yliekspression olevan yhteydessä maksasyöpään, ja Qiu ym. (2016) havaitsivat yliekspression olevan yhteydessä myös melanoomaan, mutta yhteyksiä rintasyöpään tai *BRC A2*-geeniin ei ole tutkittu. On mielenkiintoista, että kyseinen mikro-RNA on yhteydessä syöpäsairauksiin, mutta uuden tämän mikro-RNA:n kohdesekvenssin ilmestyminen ei tämänhetkisten tutkimusten perusteella vaikuta *BRC A2*-transkripteihin.

Variantista c.6821G>T puolestaan hävisi hsa-miR-1273e mikro-RNA:n kohdesekvenssi. Tämä muutos ei todennäköisesti vaikuta transkriptiin, sillä kyseisen mikro-RNA-sekvenssin funktiosta ei ole miRBase, NCBI tai Genecards tietokantojen mukaan tietoa, koska kyseinen sekvenssi on poistettu tietokannasta. Mikro-RNA-sekvenssejä analysoivat ohjelmat ovat saattaneet tulkita kyseisen sekvenssin funktionaaliseksi mikro-RNA-geeniksi, vaikka sekvenssillä ei tällaista toimintoa oikeasti ole.

Tällaisten mikro-RNA-sekvenssien muutoksia on vaikea arvioida ilman tarkempaa tutkimusta. Mikro-RNA:t toimivat usein sitoutumalla UTR-alueille, ja kummankaan mikro-RNA:n kohdesekvenssin muutos ei sijainnut UTR-alueella. Tosin Hsu ym. (2011) ovat havainneet myös koodaavalle sekvenssille sitoutuvia mikro-RNA-molekyylejä, joten ei ole mahdotonta, että tällaiset koodaavalle sekvenssille ilmaantuvat mikro-RNA kohdesekvenssien muutokset voisivat olla funktionaalisia.

4.5 Yhteenveto *BRCA*-varianttien analyysistä, sekä esimerkkejä mielenkiintoisista varianteista

BRCA-varianttien sekvenssianalyysit eri ohjelmilla tuottivat paljon mahdollisesti mielenkiintoista tietoa, mutta yksin tämä data ei hyödytä biologisesti funktionaalisten varianttien tunnistamista. Datasta on pystyttävä karsimaan pois merkityksettömät havainnot, ja koska ohjelmat keskittyvät vain mahdollisten muutosten havaitsemiseen, on tutkijan osattava päätellä ja selvittää, mitkä muutokset ovat oikeasti merkittäviä.

Eri ohjelmien käyttö saman variantin tarkasteluun mahdollistaa eri tuloksien vertailun, ja jos eri ohjelmien tulokset tukevat toisiaan, vahvistaa tämä mahdollisuutta siitä, että variantilla on oikeasti jokin vaikutus. Tämän takia varsinkin Human Splicing Finder tuloksien vertailu muiden ohjelmien havaintoihin on erityisen mielenkiintoista, vaikka toisiaan tukevia tuloksia mahdollisista silmukointiproteiinien sitoutumiskohtien muutoksista esiintyi vain parilla variantilla ESE-sekvenssien tapauksessa. Samaan tapaan Alamut-alaohjelmien samankaltaiset tulokset vahvistavat todennäköisyyttä siitä, että variantilla on tosiasiaa jokin vaikutus, varsinkin jos kaikki ohjelmat tulkitsevat variantin haitalliseksi, ja kyseinen emäs sekä aminohappo ovat fylogeneettisesti konservoituja.

Tietokoneanalyysien perusteella osalla varianteista ilmeni monia mahdollisia muutoksia, kun taas toisilla muutoksia ilmeni selvästi vähemmän. Esimerkki variantista, jossa ohjelmat havaitsivat hyvin vähän muutoksia, on *BRCA2* c.4756A>C, jolla havaittiin vain rikkoutunut ESE-sekvenssi, sekä uusi hsa-miR-769-3p mikro-RNA kohdesekvenssi. Kuten edellä mainittu, tämä mikro-RNA on kuitenkin mielenkiintoinen siksi, että sillä on todettu olevan yhteys syöpäsairauksiin.

Hyvä esimerkki variantista, jolla saattaa hyvinkin olla jokin funktionaalinen vaikutus, on *BRCA2* c.9872C>G. Kyseisellä variantilla havaittiin neljä erilaista silmukointimuutosta, joista yksi oli ESE-sekvenssin häviäminen, ja tätä muutosta tukee myös havainto ESE-sekvensseille sitoutuvan SRSF5-proteiinin sitoutumiskohdan häviämisestä RBPmap -analyysin perusteella. Variantilla havaittiin myös monen muun silmukointiin liittyvän proteiinin sitoutumiskohdan muutoksia. Näiden havaintojen lisäksi variantin ClinVar tulkinta on epävarma, kaikki Alamut-alaohjelmat tulkitsivat muutoksen haitalliseksi ja kyseisen emäksen sekä aminohapon fylogeneettinen konservoituminen oli

vahva. Näiden havaintojen perusteella tämä variantti voisi olla hyvä valinta syvempään tarkasteluun.

Toinen esimerkki mielenkiintoisesta variantista on *BRCA2* c.7712A>G. Tällä variantilla oli havaittu kaksi silmukointimuutosta, joista molemmat oli todennut usea Human Splicing Finderin käyttämä algoritmi. Myös proteiinien sitoutumiskohtien muutoksissa oli havaittu uudet silmukointiin ja mRNA aineenvaihduntaan liittyvien RBM5- ja HNRNPA2B1-proteiinien sitoutumiskohdat. Näiden havaintojen lisäksi myös kolme neljästä Alamutin alaohjelmasta oli tulkinnut variantin haitalliseksi, ja emäksen sekä aminohapon fylogeneettinen konservoituminen oli vahva. Variantti näyttää myös olevan erityisen harvinainen eurooppalaisessa populaatiossa, sillä gnomAD genomes ja ExAC tietokannoista ei löytynyt lainkaan havaintoja (liite 2, taulukko 19). Näistä syistä myös tätä varianttia voisi harkita jatkotutkimukseen.

Alleelifrekvenssin perusteella *BRCA2*-variantti c.6821G>T voisi myös olla hyvä valinta jatkotutkimukseen, sillä kyseisen variantin SISu tietokannan alleelifrekvenssi oli korkein (0,013) ja myös maailmanlaajuisesti tämän variantin frekvenssi oli eri tietokannoissa korkeampien frekvenssien joukossa (0,0014– 0,0024). Tällaisen hieman yleisemmän variantin vaikutuksen selvittäminen voisi helpottaa jatkotutkimuksia, sekä perinnöllisyysneuvontaa tulevaisuudessa. Variantissa havaittiin rikkoutunut ESE-sekvenssi ja hävinnyt pre-mRNA:n prosessointiin ja kuljetukseen, sekä silmukointiin liittyvän HNRNPA2B1-proteiinin sitoutumiskohta. Alamut tulkinnat variantista ovat ristiriitaisia ja nukleotidin fylogeneettinen konservoituminen on vain keskivertoa, mutta muista varianteista poiketen, tässä variantissa hävinneen ESE-sekvenssin oli todennut viisi eri Human Splicing Finder -ohjelman algoritmiä. Usean algoritmin toisiaan tukeva tulkinta tekee havainnosta luotettavamman kuin muiden hävinneiden ESE-sekvenssien tapauksessa. Vaikka tämänhetkinen ClinVar tulkinta olettaa variantin harmittomaksi, näyttöä tästä ei ole vielä tarpeeksi, ja tietokoneanalyysien perusteella variantilla saattaa hyvinkin olla jotain funktionaalisia vaikutuksia.

Varianttien tietokoneanalyysissä ja tulosten tarkastelussa on aina muistettava, että tulkintoja ja havaintoja tekevät ohjelmat ovat ihmisten suunnittelema ja olemassa olevan tiedon ympärille rakennettuja työkaluja. Ohjelmat helpottavat tutkijan työtä ja auttavat mielenkiintoisten varianttien rajauksessa, mutta ne eivät koskaan pysty suoraan totea-

maan varianttien biologisia vaikutuksia. Kaikki tietokoneohjelmien ennustamat vaikutukset tulee todistaa myös käytännössä. Kaikista varianteista löytyi jotain muutoksia, ja erilaisia proteiinien sitoutumiskohtien muutoksia olivat ohjelmat todenneet erityisen paljon. Kooste kaikista havaituista muutoksista, sekä lisätietoa ja tulkintoja jokaisesta variantista löytyy liitteistä 3 ja 4.

4.6 *ANO7* ddPCR-tulosten ja sekvenssidatan vertailu

4.6.1 Erot ja yhtäläisyydet ddPCR-mittauksien ja sekvensointidatan välillä

DdPCR-menetelmää ja sekvensointia on vertailtu useassa tieteellisessä julkaisussa. Aikaisemmin mainitussa de Smithin ym. (2015) lapsuusajan akuuttista lymfaattista leukemiaa (ALL) käsittelevässä tutkimuksessa tarkasteltiin tutkimusryhmän käyttämää ddPCR-tekniikkaa, vertailemalla ddPCR-tuloksia alleelisestä epätasapainosta perinteiseen sanger-sekvensoinnilla saatuihin tuloksiin, ja havaittiin, että menetelmien tulokset ovat hyvin yhteensopivia ainakin kahden tutkitun SNP:n tarkastelussa. Tutkimuksessa tosin selvitettiin RNA-tason ekspression sijaan sitä, miten syöpäkudoksessa normaali- ja riskialleelit säilyivät syöpäsoluissa sairaudelle yleisten deleetioiden johdosta. Tämän lisäksi Wang ym. (2019) ovat suoraan verranneet sanger-sekvensointia ja ddPCR-tekniikkaa papillaarisen kilpirauhassyövän riskiä kasvattavan *BRAF*^{V600E} -mutaation kliinisessä havaitsemisessa, ja he totesivat, että ddPCR oli sensitiivisempi tekniikka mutaation havaitsemiseen, varsinkin jos DNA-näytteen konsentraatio on matala, kuten kilpirauhassyövän tutkimiseen käytetyissä pienikokoisissa syöpänäytteissä. Tässäkään tutkimuksessa ei kuitenkaan tarkasteltu eri alleelien suhteita RNA-tasolla, vaan keskityttiin tietyn mutaation havaitsemiseen DNA-näytteestä. Eri tutkimuksia tarkastellessa ei löytynyt aikaisempaa tutkimustietoa, joka liittyisi suoranaisesti RNA-tason alleelisen epätasapainon tutkimiseen, sekä RNA-sekvensoinnin ja ddPCR-tekniikan tuloksien vertailuun. Tämän johdosta tuloksia tarkastellessa ei ole mahdollisuutta viitata aikaisempiin havaintoihin.

Yksi tärkeä tämän tutkimuksen tulosten tarkasteluun vaikuttava tekijä on se, ettei tutkimuksessa ollut varmuutta siitä, onko sekvensointiin ja ddPCR-mittauksiin käytetty täysin samaa lähtömateriaalia. Vaikka molemmilla menetelmillä olisi käytetty potilaan syöpäkudoksesta peräisin olevaa RNA:ta, on mahdollista, että eri menetelmillä on voitu

käyttää syöpäkudoksen RNA:ta, joka ei ole ollut peräisin täysin samasta näytteestä. Huomioitava on myös se, että tuumorit, joista näytteet on kerätty, ovat olleet heterogeenisiä, joten saman syöpäkasvaimen näytteiden välillä voi todennäköisesti ilmetä eroja. On mahdollista, että eri menetelmiin käytetty RNA on ollut peräisin eri syöpäkudospalasta, tai se on eristetty eri menetelmällä eri henkilön toimesta. kaikki nämä tekijät saattavat vaikuttaa eroihin menetelmien tuloksien välillä.

Taulukon 12 tuloksien perusteella näytteen 35 sekvensointidata ei näytä vastaavan tutkittua näytettä, koska ddPCR-mittauksen antama tulos eksoni 16 rs7593101 -variantin tarkastelussa on täysin erilainen. Sekvensoinnin perusteella näyte 35 on heterotsygootti, mutta ddPCR-mittauksen perusteella näytteessä olisi ainoastaan referenssialleeli. Myös taulukon 13 eksoni 8 rs2074840 -variantin tulos on menetelmien välillä täysin erilainen, sekvensoinnin perusteella näyte on homotsygootti, mutta ddPCR-mittauksen perusteella heterotsygootti. Mittauksen jälkeen kävi ilmi, että kyseinen näyte on sekvensoitu kahdesti eri menetelmillä, ja nämä sekvensointitulokset ovat vastanneet toisiaan. Tämän havainnon perusteella on erittäin todennäköistä, että cDNA-synteesiin käytetty RNA-näyte on ollut väärä.

Muiden näytteiden osalta sekvenssidatan ja ddPCR-mittauksien tulokset alleelien välisistä suhteista vaihtelevat rajusti tiettyjen varianttien ja näytteiden kohdalla. Esimerkiksi taulukossa 12 eksoni 16 rs7593101 -variantin näytteiden 205, 207 ja 214 alleelien väliset suhteet eroavat eri menetelmien tuloksissa merkittävästi toisistaan. Toisaalta samoja alukkeita ja koettimia käyttäen loppujen näytteiden tulokset ovat huomattavasti lähempänä toisiaan, erityisesti näytteiden 161 ja 202, joiden alleelien väliset suhteet ovat melkein samat, sekä sekvensoinnin, että ddPCR-mittauksen perusteella (sekvensointi/ddPCR: 0,34/0,34 ja 1,2/1,15) ja näytteiden ddPCR-mittauksien keskihajonnat ovat myös matalat (0,05 ja 0,06). Tämä ero tuloksissa eri näytteiden välillä ei selity näytteiden 205, 207 ja 214 alhaisella cDNA konsentraatiolla ddPCR-mittauksessa, sillä vain näytteessä 214 keskimääräinen konsentraatio on huomattavasti muista näytteistä poikkeava (4,57 kopiota/ μ l verrattuna 15,3–16,12 kopioon/ μ l). Näytteiden 207 ja 214 ddPCR-mittauksessa alleelien suhteen keskihajonta on merkittävä (0,17 ja 0,21), mutta toisaalta näytteen 205 keskihajonta on huomattavasti pienempi (0,08). Yleisesti pienempi alleelien välisen suhteen keskihajonta ddPCR-mittauksissa näyttää liittyvän yhtenevään tulokseen sekvenssidatan kanssa, paitsi näytteen 205 tapauksessa.

Suuret keskihajonnat saattavat tehdä ddPCR-mittauksista kyseenalaisia. Erot keskihajonnoissa voisi selittyä sillä, että pieni näytteen cDNA konsentraatio lisää mittausten epätarkkuutta, mutta jos tarkastellaan eksoni 16 rs7593101 -variantin osalta näytteiden 32 ja 207 ddPCR-mittausten keskihajontoja (0,06 ja 0,17), ero cDNA konsentraatioissa ei selitä eroa keskihajonnassa, sillä molemmissa näytteiden cDNA konsentraatiot ovat lähellä toisiaan (13,55 ja 16,12 kopiota/ μ l). Tämän lisäksi, vaikka näytteen 10 alleelien välisen suhteen keskihajonta ddPCR-mittauksen perusteella on eksoni 16 ja erityisesti eksoni 8 koettimilla suuri (0,28 ja 0,65), eksoni/introni 4 koettimilla taas ddPCR-mittausten keskihajonta on merkittävästi pienempi (0,06), vaikka näytteen cDNA konsentraatio ei ole muuttunut. Voi tosin olla, että suurempi cDNA konsentraatio yhdistettynä useampaan toistoon saattaisi pienentää mittausten keskihajontaa, ja lisätä kaikkien tuloksien luotettavuutta. Toisaalta toistojen lisääminen heterotsygoottisten näytteiden osalta kasvattaisi analyysin tekemiseen tarvittavaa työtaakkaa ja reagenssien kulutusta.

Taulukoiden 13 ja 14 tulosten valossa on selvää, että ddPCR-menetelmä pystyy tehokkaasti havaitsemaan toisen alleelin puuttumisen näytteestä. Kaikilla koettimilla ddPCR-mittaus pystyi helposti havaitsemaan, onko näytteessä molemmat, vai vain toinen alleeli. Nämä tulokset olivat myös yhteensopivia sekvensointituloksien kanssa. Molemmilla menetelmillä esiintyi tosin yksittäisiä toisen alleelin esiintymiseen viittaavaa datapistettä, mutta nämä ovat mitä todennäköisimmin vain menetelmien virrehavaintoja, jotka voidaan jättää tuloksien analysoinnissa huomiotta.

Toisaalta taulukon 14 näytteen 207 eksoni/introni 4 sekvensointidata tekee tällaisesta tulkinnasta haastavan. Sekvensoinnilla on saatu tulokseksi yhteensä vain 14 havaintoa. Havaintojen pienen määrän takia on vaikea sanoa, kuinka luotettava tulos alleelien välisestä suhteesta on. Erityisesti tämän näytteen eksoni/introni 4 ddPCR-mittauksen tulos saattaisi olla tarkempi kuin sekvensointitulos, sillä ddPCR-menetelmällä havaintoja on saatu merkittävästi enemmän (yhteensä keskimäärin 322 positiiviseksi määritettyä pisaraa), ja tämän lisäksi näiden havaintojen keskihajonta on pieni (0,04). Myös näytteen 10 eksoni/introni 4 on ongelmallinen samasta syystä, sillä kyseisestä näytteestä on saatu sekvensoinnilla vain neljä havaintoa, kun taas ddPCR on määrittänyt keskimäärin 317 positiivista havaintoa, ja ddPCR-mittauksen keskihajonta on pieni (0,06). Myös näytteistä 32, 161 ja 214 on saatu sekvensoinnilla erittäin pieni määrä havaintoja (8, 10 ja

10), mutta näiden näytteiden kannalta tämä ei ole ongelmallista, sillä näytteissä on havaittu vain toinen alleeli.

Eksoni/introni 4 variantin sekvensointidatan suppeus johtuu luultavasti siitä, että kyseessä on introninen variantti, joka on havaittu totaali-RNA-sekvensoinnilla. Tämä johtaa huomattavasti pienempään havaintojen määrään intronisissa sekvensseissä, koska RNA-sekvensointi keskittyy nimenomaan ekspressoituihin sekvensseihin. Eksoni/introni 4 rs77559646 -variantti on tosin erityistapaus, sillä tämä variantti aiheuttaa häiriöitä silmukoinnissa, ja johtaa intronisen sekvenssin säilymiseen transkriptissa. Tästä huolimatta sekvenssidata on tällä variantilla paljon suppeampaa kuin muilla varianteilla.

Mielenkiintoista tämän eksoni/introni 4 variantin kannalta on se, että heterotsygoottisten näytteiden ddPCR-mittauksien tuloksien keskihajonnat olivat kaikki pienempiä muiden varianttien tuloksiin verrattuna, riippumatta cDNA konsentraatiosta. Yksi selitys tälle saattaa olla, että suunnitellut koettimet ristireagoivat kohdesekvenssiensä kanssa vähemmän kuin muille varianteille suunnitellut koettimet, johtuen tarkempaan ja luotettavampaan eroon referenssi- ja varianttipositiivisten pisaroiden esiintymisessä ddPCR-mittauksessa.

Kaikissa ddPCR-mittauksissa on otettava huomioon, että positiiviset ja negatiiviset pisarat toisistaan erottava raja-arvo asetettiin itse tuloksien tarkastelun yhteydessä samalla tavalla kuin alukkeiden optimoinnissa. Kaikissa saman variantin analyysissä käytettiin samoja raja-arvoja, joilla vain ylin pisarapopulaatio todettiin positiiviseksi. Ylimmän ja keskimmäisen pisarapopulaation välillä sijaitsevat aina lukuisia yksittäisiä mahdollisesti positiivisia pisaroita, joista raja-arvon perusteella jotkin määrittyivät positiiviseksi ja osa negatiiviseksi. Voisi olettaa, että tämä vaikuttaisi tuloksiin merkittävästi, mutta tämän tutkimuksen kannalta tämä ei ole merkittävä virhelähde, sillä näitä pisaroita oli niin vähän, että raja-arvojen muuttaminen ei vaikuttanut merkittävästi alleelien väliseen suhteeseen. Jos näitä kyseenalaisia pisaroita olisi huomattavan paljon, luotettavan raja-arvon asettaminen olisi ongelmallisempaa. Tämän ongelman voisi tosin mahdollisesti ratkaista alentamalla näytteen cDNA konsentraatiota, valitsemalla PCR-lämpösykliin eri annealing lämpötilan, tai suunnittelemalla erilaiset alukkeet. Tässä tutkimuksessa ei kuitenkaan tällaista ongelmaa ilmennyt.

4.6.2 Yhteenveto ddPCR-menetelmän toimivuudesta alleelisen epätasapainon määrittämisessä

DdPCR-menetelmän soveltuvuus alleelien välisen suhteen määrittämiseen näyttää riippuvan näytteestä ja tutkittavasta variantista. Menetelmä pystyy helposti havaitsemaan toisen alleelin puuttumisen näytteestä, mutta sekvensointidataa ja ddPCR-mittauksia tarkastellessa ilmenee suuria eroavaisuuksia menetelmillä saaduissa tuloksissa. Taulukon 12 eksoni 16 variantin tuloksissa näkyy miten osalla näytteistä menetelmien tulokset ovat melkein identtiset, kun taas osalla näytteistä ero on merkittävä, johtuen suuresta ddPCR-mittauksien keskihajonnasta. Suuret keskihajonnat ddPCR-mittauksissa saattavat olla seuraus virheellisistä mittauksista, tai reaktioiden valmistelusta. Yksi mahdollinen syy saattaa myös olla sekvensointiin ja ddPCR:ään käytetyn RNA:n lähde. Jos menetelmissä on käytetty saman potilaan eri kudospalasta peräisin olevaa RNA:ta, voisi tämä selittää menetelmien antamat erilaiset tulokset.

Taulukon 13 eksoni 8 variantin tuloksissa referenssialleelin puuttuminen ilmenee selvästi molemmilla menetelmillä, ja heterotsygoottisen näytteen 202 tulokset alleelien välisestä suhteesta ovat yhteensopivia (1,43 ja 1,45). Toisaalta on mielenkiintoista, että samoilla koettimilla ja alukkeilla heterotsygoottisen näytteen 10 ddPCR-tulos eroaa sekvensointituloksesta merkittävästi (sekvensointi/ddPCR 0,95/1,56), mutta tämä johtuu mitä todennäköisimmin ddPCR-mittauksen erittäin suuresta keskihajonnasta (0,65).

Taulukon 14 eksoni/introni 4 variantin tuloksissa varianttialleelin puuttuminen ilmenee selvästi molemmilla menetelmillä, ja ddPCR-mittaukset ovat heterotsygoottisten näytteiden kohdalla kattavia ja keskihajontojen (0,04) perusteella tarkkoja. Toisaalta tämän variantin ongelmakohdaksi nousee sekvensointidatan kattavuus. Heterotsygoottisten näytteiden eri menetelmien tulokset alleelien välisistä suhteista eroavat toisistaan merkittävästi (sekvensointi/ddPCR 1,37/0,41 ja 0,56/0,29), mutta ddPCR-mittauksien pienten keskihajontojen ja sekvensointidatan suppeuden perusteella voisi tämän variantin tapauksessa todeta ddPCR-tuloksen luotettavammaksi alleelisen epätasapainon mittariksi.

Tulosten valossa on vaikea sanoa, kumpi menetelmä soveltuu paremmin alleelisen epätasapainon tarkasteluun. Eri varianteilla ja myös eri näytteillä ilmeni paljon eroavaisuuksia menetelmien tuloksissa, mutta myös erittäin hyvin toisiaan tukevia havaintoja, kuten eksoni 16 näytteet 161 ja 202. Yksi tutkimuksen ongelmakohta on se, että eksoni 8 ja eksoni/introni 4 varianteilla esiintyi monta homotsygoottista näytettä, jotka molemmat menetelmät pystyivät määrittämään helposti. Tutkimukseen olisi saatu parempaa tietoa, jos isompi osa tutkittavista näytteistä olisi ollut heterotsygoottisia näiden varianttien suhteen. Toinen tutkimuksen iso ongelma on eksoni/introni 4 näytteiden suppea sekvensointidata, joka vaikeuttaa menetelmien tuloksien vertailua. Toisaalta ddPCR-mittausten alhaiset keskihajonnat viittaavat menetelmän tuottavan tämän variantin osalta luotettavampaa tietoa sekvensointiin verrattuna. DdPCR-mittauksien ongelmakohtaksi nousi myös tiettyjen mittausten suuri keskihajonta, joka tekee tuloksista kyseenalaisia.

Tutkimusasetelmaa voisi parantaa monella tavalla mahdollisissa jatkotutkimuksissa. Tutkittavaksi voisi ottaa enemmän variantteja sekä näytteitä, ja ennalta varmistaa, että näytteissä tutkittavat variantit olisivat mieluummin heterotsygootteja kuin homotsygootteja. Lisäksi ennen tutkimusta pitäisi varmistaa näytteiden sekvenssidatan kattavuus, jotta menetelmiä voisi verrata paremmin toisiinsa. Menetelmien tarkkaa vertailua varten olisi myös erityisen tärkeää käyttää samaa lähdemateriaalia sekvensointiin ja ddPCR-mittaukseen. Myös syy tiettyjen näytteiden ddPCR-mittauksien korkeille keskihajonnoille olisi hyvä selvittää ja korjata. Mielenkiintoista voisi myös olla tutkia, minkälaisia tuloksia alleelisesta suhteesta antaisivat samalle variantille suunnitellut eri aluke- ja koetinparit. Jos tämänkaltaisessa tarkastelussa kävisi ilmi merkittäviä aluke- ja koetinriippuvaisia eroja, ddPCR-menetelmän luotettavuus kärsisi merkittävästi.

5 Kiitokset

Haluan kiittää lääketieteen professori Johanna Schleutkeria, VS ylilääkäri Minna Kan-kuri-Tammilehtoa, sekä erikoistutkija FT Gudrun Wahlströmiä pro gradu -projektin tarjoamisesta, sekä ammattitaitoisesta ohjauksesta ja palautteesta projektin aikana. Kii-tokset kuuluvat myös Turun yliopiston Luonnontieteen ja tekniikan tiedekunnan dosent-ti Christina Nokkalalle erinomaisesta pro gradu -prosessin ohjauksesta ja avustuksesta, sekä Turun yliopiston Biolääketieteen laitoksen laboratoriomestari Jukka Karhulle erin-omaisesta käytännön töiden ohjauksesta ja avustuksesta.

6 Lähteet

Abramowicz A & Gos M (2018) Splicing mutations in human genetic disorders: exam-ples, detection, and confirmation, s. 253–268, *Journal of Applied Genetics*, Vol 59

Adzhubei I A, Schmidt S, Peshkin L, Ramensky V E, Gerasimova A, Bork P, Kon-drashov A S, & Sunyaev S R (2010) A method and server for predicting damaging mis-sense mutations, s. 248–249, *Nature Methods*, Vol 7

Anantha R W, Alcivar A L, Ma J, Cai H, Simhadri S, Ule J, König J, & Xia B (2013) Requirement of Heterogeneous Nuclear Ribonucleoprotein C for BRCA Gene Expres-sion and Homologous Recombination, *PLoS ONE*, 8(4): e61368

Asaoka M, Ishikawa T, Takabe K, & Patnaik S K (2019) APOBEC3-Mediated RNA Editing in Breast Cancer is Associated with Heightened Immune Activity and Improved Survival, *International Journal of Molecular Sciences*, Vol 20

Batcha A M N, Bamopoulos S A, Kerbs P, Kumar A, Jurinovic V, Rothenberg-Thurley M, Ksienzyk B, Philippou-Massier J, Krebs S, Blum H et al. (2019) Allelic Imbalance of Recurrently Mutated Genes in Acute Myeloid Leukaemia, *Scientific Reports*, Vol 9, Article number: 11796

Buckland P R (2004) Allele-specific gene expression differences in humans, s. 255-260, *Human Molecular Genetics*, Vol 13, Issue supplement 2

Burset M, Seledtsov I A & Solovyev V V (2001) SpliceDB: database of canonical and non-canonical mammalian splice sites, s. 255-259, Nucleic Acids Research, Vol 29, No. 1

Caux-Moncoutier V, Page's-Berhouet S, Michaux D, Asselain B, Caste'ra L, Pauw A D, Buecher B, Gauthier-Villars M, Stoppa-Lyonnet D & Houdayer C (2009) Impact of *BRCA1* and *BRCA2* variants on splicing: clues from an allelic imbalance study, s. 1471-1480, European Journal of Human Genetics, Vol 17

Chen Sining & Parmigiani Giovanni (2007) Meta-Analysis of *BRCA1* and *BRCA2* Penetrance, s. 1329-1333, J. Clin. Oncol, Vol. 25, No. 11

Chen X, Weaver J, Bove B A , Vanderveer L A, Weil S C, Miron A, Daly M B & Godwin A K (2008) Allelic imbalance in *BRCA1* and *BRCA2* gene expression is associated with an increased breast cancer risk, s. 1336–1348, Human Molecular Genetics, Vol 17, No. 9

Desmet F O, Hamroun D, Lalande M, Collod-Be'roud G, Claustres M & Be'roud C (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals, Nucleic Acids Research, Vol 37, No. 9

De Smith A J, Walsh K M, Hansen H M, Endicott A A, Wiencke J K, Metayer C & Wiemels J L (2015) Somatic Mutation Allelic Ratio Test Using ddPCR (SMART-ddPCR): An Accurate Method for Assessment of Preferential Allelic Imbalance in Tumor DNA, PLoS One, Vol 10

ENIGMA consortium (2020) < <https://enigmaconsortium.org/> > [luettu 3.9.2020]

Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler R K & Hauke J (2018) Performance of in silico prediction tools for the classification of rare *BRCA1/2* missense variants in clinical diagnostics, BMC Medical Genomics, Vol 11, Article number: 35

- Hellen C U T & Sarnow P (2001) Internal ribosome entry sites in eukaryotic mRNA molecules, s. 1593–1612, *Genes And Development*, Vol 15
- Hindson B J, Ness K D, Masquelier D A, Belgrader P, Heredia N J, Makarewicz A J, Bright I J, Lucero M Y, Hiddessen A L, Legler T C et al. (2011) High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number, s. 8604–8610, *Analytical Chemistry*, Vol 83
- Hjelmberg J B, Scheike T, Holst K, Skytthe A, Penney K L, Graff R E, Pukkala E, Christensen K, Adami H O, Holm N V et al. (2014) The Heritability of Prostate Cancer in the Nordic Twin Study of Cancer, s.2303-2310, *Cancer Epidemiology Biomarkers & Prevention*, Vol 23, issue 11
- Hsu J B, Chiu C, Hsu S, Huang W, Chien C, Lee T & Huang H (2011) miRTar: an integrated system for identifying miRNA-target interactions in human, *BMC Bioinformatics*, Vol 12, Article number: 300
- Huang H Y, Chien C, Jen K & Huang H D (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements, *Nucleic Acids Research*, Vol 34, Web Server issue
- Huijismans C J J, Brule A J C, Rigter H, Poodt J, Linden J C, Savelkoul P H M, Hilbink M & Hermans M H A (2015) Allelic imbalance at the *HER2/TOP2A* locus in breast cancer, *Diagnostic Pathology*, Vol 10, Article number: 56
- Jamard E, Volard B, Dugue A E, Legros A, Leconte A, Clarisse B, Davy G, Polycarpe F, Dugast C, Abadie C et al. (2017) *BRCAl* allele-specific expression in genetic predisposed breast/ovarian cancer, s. 167–171, *Familial Cancer*, Vol 16
- Joensuu H (2019) Rintasyövän yleisyys ennuste, riskitekijät ja ehkäisy. Teoksessa: *Rintasyövän valtakunnallinen diagnostiikka- ja hoitosuositus* (Aittomäki K, Auvinen P, Heikkilä P, Hukkinen K, Huovinen R, Jahkola T, Joensuu H, Joukainen S, Jukkola A, Karihatala E ym.), s 6–7. Suomen Rintasyöpäryhmä ry

Kaikkonen E, Rantapero T, Zhang Q, Taimen P, Laitinen V, Kallajoki M, Jambulingam D, Ettala O, Knaapila J, Boström P J et al. (2018) ANO7 is associated with aggressive prostate cancer, s. 2479–2487, *International Journal of Cancer*, Vol 143, issue 10

Kamitaki N, Usher C L & McCaroll S A (2018) Using Droplet Digital PCR to Analyze Allele-Specific RNA Expression, In: *Digital PCR. Methods in Molecular Biology*, (Karlin-Neumann G., Bizouarn F. eds) s. 401-422, Vol 1768, Humana Press, New York, NY

Kankuri-Tammilehto M, Vihinen P & Schleutker J (2019) Syövän perinnöllisyys, *Lääkärilehti*, s. 880–886, Vol 74

Kuchenbaecker K B, Hopper J L, Barnes D R, Phillips K, Mooij T M, Roos-Blom M, Jervis S, Leeuwen F E, Milne R L, Andrieu N et al. (2017) Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers, *JAMA*, s. 2402-2416, Vol 317

Landrum M J, Lee J M, Benson M, Brown G R, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D & Jang W et al. (2017) ClinVar: Improving Access to Variant Interpretations and Supporting Evidence, *Nucleic Acids Research*, Vol 46, issue D1,

Maeda S, Ohka F, Okuno Y, Aoki K, Motomura K, Takeuchi K, Kusakari H, Yanagisawa N, Sato S, Yamaguchi J et al. (2020) *H3F3A* mutant allele specific imbalance in an aggressive subtype of diffuse midline glioma, H3 K27M-mutant, *Acta Neuropathologica Communications*, Vol 8, issue 1, Article number: 8

Miki Y, Swensen J, Shattuck-Eidens D, Futreal P A, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett L B, Ding W et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, s.66–71, *Science*, Vol 266

Mizuno A & Okada Y (2019) Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects, s. 1745–1756, *European Journal of Human Genetics*, Vol 27

Paz I, Kostı I, Ares Jr M, Cline M & Mandel-Gutfreund Y (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins, *Nucleic Acids Research*, Vol 42, Web Server issue

Petrucci N, Daly M B, Pal T, Adam M P, Ardinger H H, Pagon R A, Wallace S E, Bean L JH, Stephens K, Amemiya A (1998) BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer, *GeneReviews* (updated 12/2016)

Qiu H, Lu X, Yang S, Weng C, Zhang E & Chen F (2016) MiR-769 promoted cell proliferation in human melanoma by suppressing GSK3B expression, s. 117–123, *Biomedicine & Pharmacotherapy*, Vol 82

Rhee J, Lee S, Park W, Kim Y & Kim T (2017) Allelic imbalance of somatic mutations in cancer genomes and transcriptomes, *Scientific Reports*, Vol 7, Article number: 1653

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody W W, Hegde M, Lyon E, Spector E et al. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology, s. 405–424, *Genet Med*, Vol 17

Sanz D J, Hollywood J A, Scallan M F & Harrison P T (2017) Cas9/gRNA targeted excision of cystic fibrosiscausing deep-intronic splicing mutations restores normal splicing of CFTR mRNA, *PLoS One*, Vol 12

Schwarz J M, Cooper D N, Schuelke M & Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age, s.361–362, *Nature Methods*, Vol 11

Segui N, Stevens K N, Guino' E, Rozek L S, Moreno V R, Capella' G, Gruber S B & Valle L (2011) No association between germline allele-specific expression of *TGFBR1* and colorectal cancer risk in Caucasian and Ashkenazi populations, s. 735–740, *British Journal of Cancer*, Vol 104

Sharma S, Patnaik S K, Taggart R T, Kannisto E D, Enriquez S M, Gollnick P & Baysal B E (2015) APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages, *Nature Communications*, Vol 6, Article number: 6881

Shen J, Medico L, & Zhao H (2011) Allelic Imbalance in *BRCA1* and *BRCA2* Gene Expression and Familial Ovarian Cancer, *Cancer Epidemiology, Biomarkers & Prevention*, Vol 20, Issue 1

Sim N, Kumar P, Hu J, Henikoff S, Schneider G & Ng P C (2012) SIFT web server: predicting effects of amino acid substitutions on proteins, *Nucleic Acids Research*, Vol 40, Web Server issue

So M, Jeong T, Lim W, Moon B, Paik N S, Kim S C & Huh J (2019) Reinterpretation of BRCA1 and BRCA2 variants of uncertain significance in patients with hereditary breast/ovarian cancer using the ACMG/AMP 2015 guidelines, s. 510–519, *Breast Cancer*, Vol 26

Suomen syöpärekisteri 2020, syöpätilastosovellus <<https://syoparekisteri.fi/tilastot/tautitilastot/>> [luettu 2.9.2020]

Syöpäjärjestöt (2020a) Rintasyöpä, <<https://www.kaikkisyovasta.fi/tietoa-syovasta/syopataudit/rintasyopa/>> [luettu 10.3.2020]

Syöpäjärjestöt (2020b) Eturauhassyöpä, <<https://www.kaikkisyovasta.fi/tietoa-syovasta/syopataudit/eturauhassyopa/>> [luettu 10.3.2020]

Tavtigian S V, Deffenbaugh A M, Yin L, Judkins T, Scholl T, Samollow P B, de Silva D, Zharkikh A, & Thomas A (2005) Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral, s. 295–305, *Journal of Medical Genetics*, Vol 43

Tomsic J, Guda K, Liyanarachchi S, Hampel H, Natale L, Markowitz S D, Tanner S M & Chapelle A (2010) Allele-specific expression of *TGFBR1* in colon cancer patients, s. 1800–1804, *Carcinogenesis*, Vol 31, Issue 10

Vu L T & Tsukahara T (2017) C-to-U editing and site-directed RNA editing for the correction of genetic mutations, s. 243–253, *BioScience Trends*, Vol 11, Issue 3

Wang Z, Sun K, Jing C, Cao H, Ma R & Wu J (2019) Comparison of Droplet Digital PCR and Direct Sanger Sequencing for the Detection of the BRAF V600E Mutation in Papillary Thyroid Carcinoma, *Journal Of Clinical Laboratory Analysis*, Vol 33, issue 6

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G et al. (1995) Identification of the breast cancer susceptibility gene BRCA2, s.789–792, *Nature*, Vol 378

Xian Y, Wang L, Yao B, Yang W, Mo H, Zhang L & Tu K (2019) MicroRNA-769-5p contributes to the proliferation, migration and invasion of hepatocellular carcinoma cells by attenuating RYBP, *Biomedicine & Pharmacotherapy*, Vol 118

Yoshida K & Miki Y (2004) Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage, s.866–871, *Cancer Science*, Vol 95, no 11

Zhu J, Mayeda A & Krainer A R (2001) Exon Identity Established through Differential Antagonism between Exonic Splicing Silencer-Bound hnRNP A1 and Enhancer-Bound SR Proteins, s.1351–1361, *Molecular Cell*, Vol 8

7 Liitteet

Liite 1: Yksityiskohtaiset tiedot *ANO7* alukkeiden ja koettimien testaamisesta ja optimoinnista

Taulukko 15. PCR-koneeseen asetettu *ANO7* eksoni 16 rs7593101 lämpötilagradientin näytelevy. Tällä levyllä näytteitä oli vain pystyriivillä 1.

PCR levyn kaivo	MDA PCa 2b cDNA/rs7593101	Annealing lämpötila °C
A	Näyte	61
B	Puskuri, ei näytettä	60.4
C	Näyte	59.1
D	Näyte	57.2
E	Näyte	55
F	Näyte	53.1
G	Negatiivinen kontrolli	51.7
H	Näyte	51

Taulukko 16. PCR-koneeseen asetettu *ANO7* eksoni/introni 4 rs77559646 ja *ANO7* eksoni 8 rs2074840 lämpötilagradientin näytelevy. Tällä levyllä näytteitä oli pystyriiveillä 1 ja 2. Näytelevyn valmistelussa kaivoon D2 laitettiin vahingossa kaksinkertainen määrä ohjeen mukaan valmistettua master mix seosta.

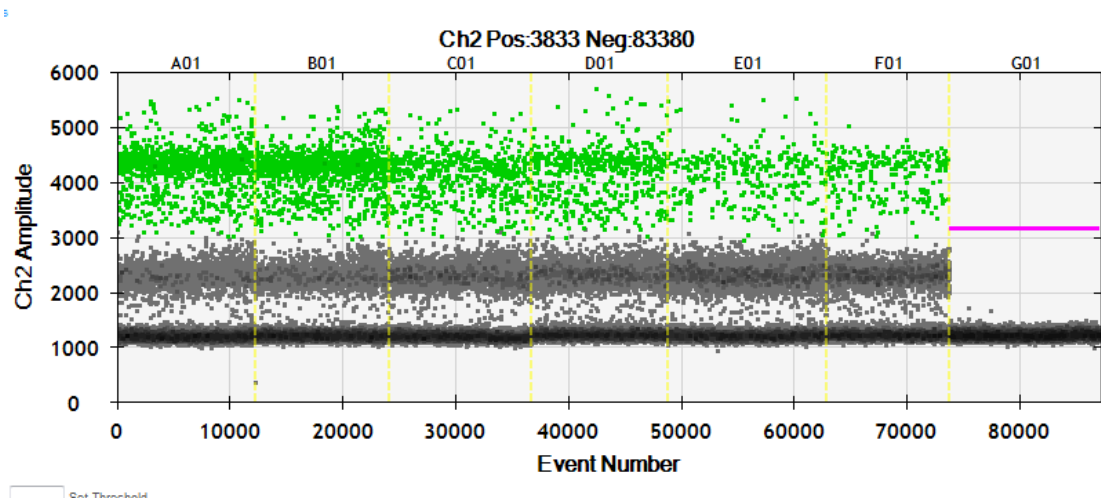
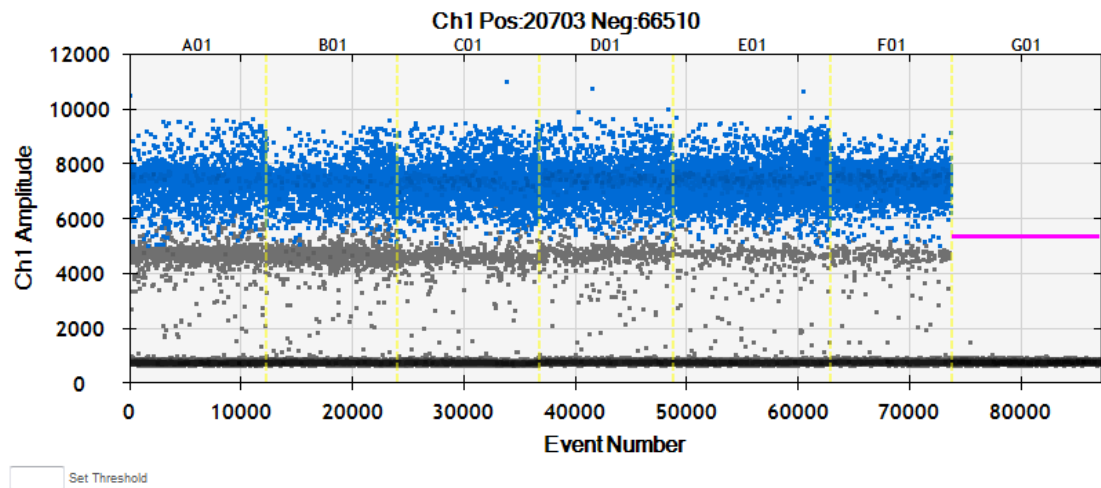
PCR levyn kaivo	1) 22Rv1 cDNA/ rs77559646	2) LNCaP cDNA/ rs2074840	Annealing lämpötila °C
A	Näyte	Näyte	67
B	Puskuri, ei näytettä	Puskuri, ei näytettä	66.3
C	Näyte	Näyte	64.9
D	Näyte	Näyte	62.8
E	Näyte	Näyte	60.4
F	Näyte	Näyte	58.3
G	Negatiivinen kontrolli	Negatiivinen kontrolli	56.8
H	Näyte	Näyte	56

Taulukko 17. Uusi PCR-koneeseen asetettu alhaisemman annealing lämpötilan *ANO7* eksoni/introni 4 rs77559646 lämpötilagradientin näytelevy. Tällä levyllä näytteitä oli vain pystyrivillä 1.

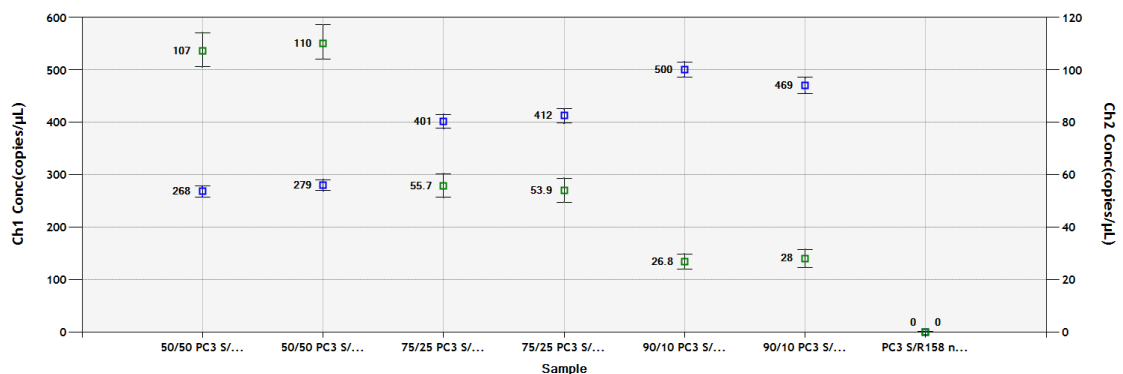
PCR levyn kaivo	22Rv1 cDNA/ rs77559646	Annealing lämpötila °C
A	Näyte	62
B	Puskuri, ei näytettä	61.4
C	Näyte	60.1
D	Näyte	58.2
E	Näyte	56
F	Näyte	54.1
G	Negatiivinen kontrolli	52.7
H	Näyte	52

Taulukko 18. Eksoni/introni 4 rs77559646 PCR-levyn ANO7S ja ANO7S-R158H 1:250 laimennoksen konsentraatiogradientti. Annealing lämpötila 58 °C.

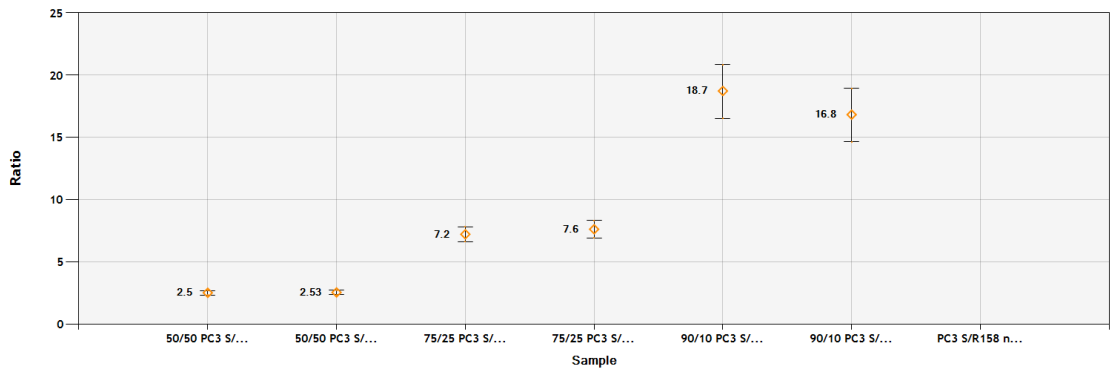
PCR levyn kaivo	ANO7S / ANO7S-R158 cDNA suhde
A	50/50
B	50/50
C	75/50
D	75/50
E	90/10
F	90/10
G	Negatiivinen kontrolli
H	Puskuri



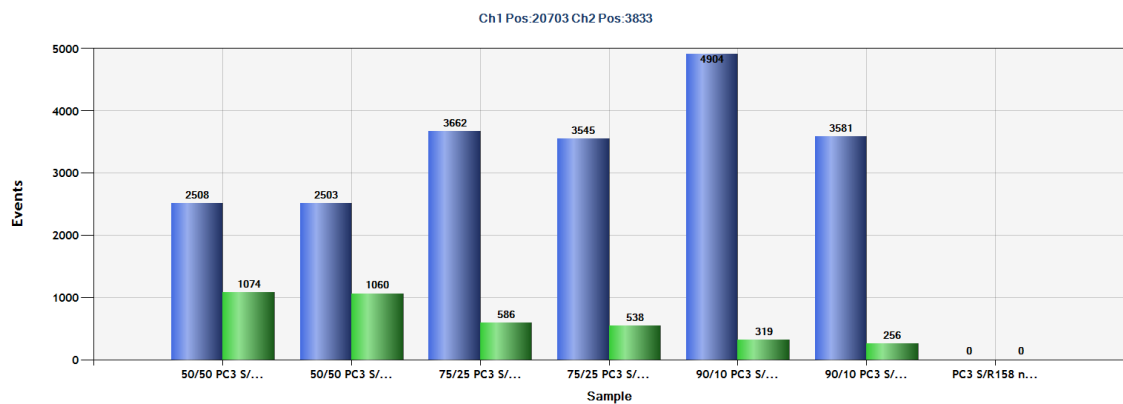
Kuva 11. "1-D amplitude plot" eksoni/introni 4 rs77559646 konsentraatigradientista. Siniset FAM-pisarot vastaavat normaalia G-alleelia ja vihreät HEX-pisarot vastaavat variantti A-alleelia. Kuvassa näkyy miten vihreiden HEX-pisaroiden määrä laskee kaivosta A kaivoon F kun varianttialleelin sisältävän cDNA:n määrä laskee liitteen 1 taulukon 18 mukaan.



Kuva 12. "Concentration" -ikkuna eksoni/introni 4 rs77559646 konsentraatigradientista. Kuvassa näkyy selvästi, miten vihreän mutanttialleelia sisältävän cDNA:n konsentraatio laskee vasemmalta oikealle kaivosta A kaivoon F liitteen 1 taulukon 18 mukaan.



Kuva 13. ”Ratio” -ikkuna eksoni/introni 4 rs77559646 konsentraatigradientista. Kuvassa näkyy, miten normaalialleelin suhde varianttialleeliin muuttuu selvästi, kun varianttialleelin sisältämän cDNA:n määrä laskee näytteessä vasemmalta oikealle kaivosta A kaivoon F liitteen 1 taulukon 18 mukaan.



Kuva 14. ”Events” -ikkuna eksoni/introni 4 rs77559646 konsentraatigradientista. Kuvassa näkyy miten itse asetettujen raja-arvojen perusteella droplet reader -laite on määrittänyt pisaroi- ta FAM- ja HEX-positiivisiksi. Varianttialleelia vastaavien vihreiden HEX-positiivisten pisaroi- den määrä laskee vasemmalta oikealle kaivosta A kaivoon F, kun tätä varianttia sisältämän cDNA:n määrä näytteessä laskee liitteen 1 taulukon 18 mukaan.

Liite 2: *BRCA*-varianttien alleelifrekvenssit

Taulukko 19. Tutkittavien *BRCA*-varianttien SISusta (The Sequencing Initiative Suomi) ja ClinVarista haetut eri tietokantojen mukaiset alleelifrekvenssit. GnomAD ja ExAC frekvensseissä on huomioitava eurooppalaisten frekvenssien monesti huomattavasti suurempi otanta muihin maailman väestöihin verrattuna.

Variantti	gnomAD genomes european/global	gnomAD exomes european/global	ExAC european/global	SISu, laatu-kontrollin jälkeinen frekvenssi
<i>BRCA1</i>				
c.230C>T	-	0.000022/0.000016	0.00004/0.000025	-
c.3257T>C	-	0.000022/0.000012	0.00005/0.000033	0.000048
c.3649T>C	-	-	-	-
c.4676-8C>G	0.00005/0.00006	0.000097/0.000064	0.00014/0.000084	0.00029
<i>BRCA2</i>				
c.68-7T>A	0.0035/0.0026	0.0031/0.0029	0.0032/0.0024	0.006
c.583T>G	-	-	-	-
c.2884C>T	0.00011/0.00010	0.00016/0.000097	0.00015/0.000091	0.00082
c.3314C>T	-	-	-	-
c.4756A>C	-	-	-	-
c.5225A>T	0.00005/0.00003	0.000037/0.000020	0.00004/0.000025	0.000048
c.6821G>T	0.0037/0.0024	0.0027/0.0015	0.0022/0.0014	0.013
c.7121A>G	0.00005/0.00003	0.0/0.000004	-	-
c.7712A>G	0.0/0.00022	0.000007/0.000032	0.0/0.000041	-
c.7759C>T	-	0.000066/0.000044	0.00007/0.000041	-
c.7805+13A>G	0.00005/0.00025	0.000022/0.000040	0.00003/0.000058	-
c.8855T>C	-	-	-	-
c.8953+24T>C	-	-	-	-
c.9232G>T	-	-	-	-
c.9302T>C	-	-	-	-
c.9411T>G	0.0012/0.00076	0.001/0.00055	0.00083/0.00052	0.0042
c.9872C>G	0.00021/0.00013	0.00016/0.000088	0.00005/0.000033	0.00081
c.10234A>G	0.0022/0.037	0.0026/0.023	0.0023/0.023	0.0084

Liite 3: Yhteenveto *BRCA1*- ja *BRCA2*-varianttien *in silico* -analyyseistä.

Taulukko 20. Yhteenveto *BRCA1*- ja *BRCA2*-varianttien *in silico* -analyyseistä. Ev. = epävarma, Hai = haitallinen, Har. = harmiton, U. = uusi, H. = hävinnyt, R. = rikkoutunut, Ei kon. = ei konservoitumista, Heik. = heikko, Kesk. = keskiverto, Vah. = vahva, ESS = exonic splicing silencer, ESE = exonic splicing enhancer

Variantti	ClinVar tulkinta	Alamut-ohjelmien tulkinnat	Havaitut silmukointimuutokset	Uudet sitoutumiskohdat	Hävinneet sitoutumiskohdat	Muut muutokset	Nukleotidin fylogeneettinen konservoituminen	Aminohapon fylogeneettinen konservoituminen
BRCA1								
c.230C>T	2/4 Ev.	2/4 Hai.	U. ESS site R. ESE site	-	SRSF3	H. C-to-U RNA editing site	Kesk.	Kesk.
c.3257T>C	0/4 Ev.	2/4 Hai.	-	-	GTF2IRD1-isoform 2, HNRNPA1L2	-	Heik.	Kesk.
c.3649T>C	2/4 Ev.	0/4 Hai.	-	-	SRSF3, MATR3	-	Heik.	Kesk.
c.4676-8C>G	¼ mahdollisesti Har.	-	-	RBM41, ZC3H14	PCBP1, PCBP3	-	-	-
BRCA2								
c.68-7T>A	¾ Har.	-	-	KHDRBS1, KHDRBS2, PABPC4	DAZAP1, RBMS1	Uusi ribosomin sitoutumiskohta ja avoin lukuraami	-	-
c.583T>G	-	2/4 Hai.	U. Donor site U. ESS site	C/EBP	SRSF3	-	Heik.	Kesk.
c.2884C>T	2/4 Ev.	0/4 Hai.	U. ESS site	TRA2B, aMEF-2	-	-	Ei kon.	Heik.
c.3314C>T	-	2/4 Hai.	R. ESE site	RBM42	HNRNPK, cd45 tyrosine phosphatase ESE	-	Kesk.	Heik.
c.4756A>C	-	¼ Hai.	R. ESE site	-	-	U. hsa-miR-769-3p micro RNA target site	Heik.	Heik.
c.5225A>T	2/4 Ev.	1/3 Hai.	R. ESE site	TRA2B, G3BP2	QKI, RBM42	-	Heik.	Heik.
c.6821G>T	¼ mahdollisesti Har.	2/4 Hai.	R. ESE site	-	HNRNPA2B1, LIN28A, dystrophin gene exon 27 ESE, g-alpha-s exon 3 ESE	H. hsa-miR-1273e micro RNA target site	Kesk.	Heik.
c.7121A>G	¼ Ev.	0/4 Hai.	U. Donor site	-	PABPC5	-	Ei kon.	Kesk.

			U. ESS site					
c.7712A>G	¼ Ev.	¾ Hai.	U. ESS site R. ESE site	MAF, RBM5, HNRN- PA2B1	G3BP2	H. RNA C-to-U editing site	Vah.	Vah.
c.7759C>T	¼ Ev.	4/4 Hai.	U. Donor site	ZNF638	-	-	Hei.	Vah.
c.7805+13 A>G	¼ Ev.	-	U. ESS site R. ESE site	-	Oct-1, PABPC1, PABPC4, PUM2, SART3	-	-	-
c.8855T> C	-	2/4 Hai.	R. ESE site	Kid3, TEL1	-	-	Vah.	Hei.
c.8953+24 T>C	-	-	-	-	HNRNPC, HNRN- PCL1, RALY, U2AF2, CPEB2, CPEB4, TIA1	-	-	-
c.9232G> T	-	¾ Hai.	U. ESS site	RALY, U2AF2, CPEB2, CPEB4, HNRNPC, HNRN- PCL1, ZC3H14, ZNF638	SFPQ, SRSF2	H. RNA C-to-U editing site	Kesk.	Vah.
c.9302T> C	¼ Ev.	3/3 Hai.	R. ESE site	-	HNRNPU, PUM2, PLZF	-	Kesk.	Kesk.
c.9411T> G	¾ Har.	0/1 Hai.	U. Donor site	ATF2, SRSF5	CDX1, MBNL1, RBM41, U2AF2, SRSF3, A1CF, CPEB4	-	-	-
c.9872C> G	2/4 Ev.	4/4 Hai.	2 U. Donor site U. ESS site R. ESE site	TIA1, ZC3H14	MBNL1, SRSF5, TARDBP	-	Vah.	Vah.
c.10234A> G	¾ Har.	0/4 Hai.	R. ESE site	CNOT4	YBX1, ZCRB1, SRSF3	-	E ikon.	Heik.

Liite 4: *BRCA1*- ja *BRCA2*-varianttien yksityiskohtainen tarkastelu

Tässä liitteessä tarkastelemme erikseen jokaisen tutkitun variantin. Tähän tarkasteluun sisältyy kuvaus jokaisesta proteiinista, jolla havaittiin jokin muutos varianttisekvenssissä. Proteiinien tiedot haettiin Genecards tietokannasta. Havaittujen silmukointimuutosten yhteydessä mainitaan sulkeissa Human Splicing Finderin käyttämien algoritmien määrä, jolla tulos saatiin.

***BRCA1*-variantit**

c.230C>T

Havaitut silmukointimuutokset: Uusi ESS-sekvenssi (2) ja rikkoutunut ESE-sekvenssi (3)

Hävinneet sitoutumiskohdat: SRSF3

Muut havainnot: Hävinnyt C-to-U RNA editointikohta

ClinVar tulkinta: 2/4 epävarma

Alamut-ohjelmien tulkinnat: 2/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Keskiverta

Aminohapon fylogeneettinen konservoituminen: Keskiverta

Tulkinta: Rikkoutunut ESE-sekvenssi tulos vastaa sekvenssistä hävinnyttä SRSF3 silmukointiproteiinin sitoutumiskohdan häviämistä hyvin. Variantin ClinVar tulkinta on epävarma ja neljästä Alamut-ohjelmasta puolet totesivat variantit haitalliseksi. Vaikka emäksen ja aminohapon fylogeneettinen konservoituminen on vain keskiverta, voisi silmukointitulosten ja hävinneen SRSF3 sitoutumiskohdan häviämisen perusteella todeta variantin mahdollisesti silmukointia häiritseväksi.

c.3257T>C

Ei havaittuja silmukointimuutoksia.

Hävinneet sitoutumiskohdat: GTF2IRD1-isoform 2, HNRNPA1L2

ClinVar tulkinta: 0/4 epävarma

Alamut-ohjelmien tulkinnat: 2/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Heikko

Aminohapon fylogeneettinen konservoituminen: Keskiverto

Tulkinta: GTF2IRD1 säätelee kognitiivista kehitystä, mutta sillä saattaa olla vaikutusta myös solusyklin säätelyssä, kun taas HNRNPA1L2 on silmukointiproteiini. Emäksen fylogeneettinen konservoituminen on heikko ja silmukointimuutoksia ei havaittu, vaikka hävinnyt silmukointiproteiinin sitoutumiskohta havaittiin. Variantin vaikutus on epävarma.

c.3649T>C

Ei havaittuja silmukointimuutoksia.

Hävinneet sitoutumiskohdat: SRSF3, MATR3

ClinVar tulkinta: 2/4 epävarma

Alamut-ohjelmien tulkinnat: 0/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Heikko

Aminohapon fylogeneettinen konservoituminen: Keskiverto

Tulkinta: Silmukointitulkinnaasta huolimatta sekvenssissä havaittiin hävinnyt SRSF3 silmukointiproteiinin sitoutumiskohta. MATR3 saattaa stabiloida tiettyjä mRNA-molekyylejä, ja olla vastuussa viallisten mRNA-molekyylien säilyttämisestä tumassa, mutta tämä ei ole varma. Alamut tulkintojen, nukleotidin heikon fylogeneettisen konservoitumisen ja silmukointimuutosten perusteella variantin voisi mahdollisesti todeta harmittomaksi, tosin ClinVar tietokannan mukaan tilanne on vielä epävarma.

c.4676-8C>G

Ei havaittuja silmukointimuutoksia

Uudet sitoutumiskohdat: RBM41, ZC3H14

Hävinneet sitoutumiskohdat: PCBP1, PCBP3

ClinVar tulkinta: ¼ mahdollisesti harmiton

Tulkinta: RBM41 on RNA:ta sitova proteiini, jolla on todettu yhteys Ebsteinin anomaliaan. ZC3H14 on taas poly-A-sekvenssiä sitova proteiini, joka saattaa vaikuttaa mRNA:n stabiliteettiin, kuljetukseen ja translaatioon varsinkin hermosoluissa. PCBP1 proteiini taas on polioviruksen translaatiota säätelevä proteiini, joilla on todettu myös

yhteys ihmisen 15-lipoxygenaasin translaation säätelyssä. PCBP3 on C-rikkaisiin sekvensseihin sitoutuva RNA:ta sitova proteiini, jolla on transkription jälkeisiä vaikutuksia.

Variantissa on havaittu paljon eri proteiinien sitoutumiskohtien muutoksia, mutta näiden vaikutuksia on vaikea ennakoida, varsinkin koska kyseessä on introninen variantti. Silmukointimuutoksia ei havaittu ja ClinVar on tulkinnut variantin mahdollisesti harmittomaksi, joten tämä variantti voisi mahdollisesti olla harmiton.

***BRC42*-variantit**

c.68-7T>A

Ei havaittuja silmukointimuutoksia.

Uudet sitoutumiskohdat: KHDRBS1, KHDRBS2, PABPC4

Hävinneet sitoutumiskohdat: DAZAP1, RBMS1

Muut havainnot: Uusi ribosomin sitoutumiskohta ja avoin lukuraami

Tulkinta: KHDRBS1 ja KHDRBS2 ovat silmukointiproteiineja, joilla on todettu spesifisiä silmukointivaikutuksia eri geeneissä. PABPC4 on poly-A-sekvenssiin sitoutuva proteiini, jolla saattaa olla vaikutuksia transkription säätelyyn T-soluissa, verihiiutaleissa ja megakaryosyyteissä. RBMS1-proteinilla on todettu yhteys *MYC*-geenin toimintaan, kun taas DAZAP1 on RNA:ta sitova proteiini, jota tarvitaan spermatogeneesissä.

Vaikka Human Splicing Finder ei havainnut silmukointimuutoksia, sekvenssissä on silti havaittu silmukointiin liittyvien proteiinien sitoutumiskohtien muutoksia. Variantin mahdollinen vaikutus on epäselvä.

c.583T>G

Havaitut silmukointimuutokset: Uusi luovuttajasekvenssi (1) ja ESS-sekvenssi (3)

Uudet sitoutumiskohdat: C/EBP

Hävinneet sitoutumiskohdat: SRSF3

ClinVar tulkinta: Ei tietoa

Alamut-ohjelmien tulkinnat: 2/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Heikko

Aminohapon fylogeneettinen konservoituminen: Keskiverto

Tulkinta: C/EBP on transkriptiotekijä, joka säätelee monia immuunipuolustukseen ja tulehdukseen liittyviä geenejä, kun taas SRSF3 on silmukointiproteiini. Silmukointitilojen havaitsema uusi ESS-sekvenssi ja hävinnyt SRSF3 sitoutumiskohta saattavat vaikuttaa häiritsevästi silmukointiin. Lisäksi variantista ei ole ClinVar tietokannassa mitään tietoa ja Alamut-ohjelmien tulkinnat ovat ristiriitaisia. Näiden havaintojen perusteella variantin vaikutus on hyvin epävarma.

c.2884C>T

Havaitut silmukointimuutokset: Uusi ESS-sekvenssi (3)

Uudet sitoutumiskohdat: TRA2B, aMEF-2

ClinVar tulkinta: 2/4 epävarma

Alamut-ohjelmien tulkinnat: 0/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Ei konservoitumista

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: TRA2B on silmukointiproteiini, kun taas aMEF-2 on lihassoluspesifinen transkriptiotekijä, joka säätelee kasvutekijöiden ja stressin aktivoimia geenejä.

Variantissa oli yksi silmukointiin liittyvän proteiinin muutos, sekä uusi havaittu ESS-sekvenssi. Toisaalta Alamut-ohjelmat ovat tulkinneet variantin harmittomaksi ja nukleotidi sekä aminohappo eivät ole vahvasti fylogeneettisesti konservoituneita. Variantin vaikutus on epävarma, kuten ClinVar tulkinta.

c.3314C>T

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (2)

Uudet sitoutumiskohdat: RBM42

Hävinneet sitoutumiskohdat: HNRNPK, cd45 tyrosine phosphatase ESE

ClinVar tulkinta: Ei tietoa

Alamut-ohjelmien tulkinnat: 2/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Keskierto

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: RBM42 sitoutuu *CDKN1A* mRNA:han ja HNRNPK on ribonukleoproteiini, joka sitoutuu poly-C-sekvensseihin ja on vastuussa mRNA:n prosessoinnista ja kuljuksesta, mutta jolla ei ole havaittu silmukointiin liittyviä ominaisuuksia. Koska Alamut tulokset ovat ristiriitaisia ja variantilla ei ole omaa ClinVar sivua, variantin mahdollista vaikutusta on näiden tuloksien perusteella vaikea ennakoida.

c.4756A>C

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (3)

Muut havainnot: Uusi hsa-miR-769-3p mikro-RNA-kohdesekvenssi

ClinVar tulkinta: Ei tietoa

Alamut-ohjelmien tulkinnat: ¼ haitallinen

Nukleotidin fylogeneettinen konservoituminen: Heikko

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: Vaikka Human Splicing Finder on havainnut silmukointimuutoksen, muut ohjelmat eivät ole havainneet mitään silmukointiproteiinien sitoutumiskohtien muutoksia. Valtaosa Alamut-ohjelmista tulkitse variantin harmittomaksi, ja tämän lisäksi nukleotidin sekä aminohapon fylogeneettinen konservoituminen on heikkoa. Vaikka ClinVar tietokannassa ei ole tietoa tästä variantista, voisi näiden tulosten perusteella pitää varianttia mahdollisesti harmittomana.

c.5225A>T

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (2)

Uudet sitoutumiskohdat: TRA2B, G3BP2

Hävinneet sitoutumiskohdat: QKI, RBM42

ClinVar tulkinta: 2/4 epävarma

Alamut-ohjelmien tulkinnat: 1/3 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Heikko

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: TRA2B on silmukointiproteiini, ja QKI on RNA:ta sitova proteiini, jolla on silmukointiominaisuuksia ja tärkeä tehtävä myelinisaation säätelyssä. RBM42 sitoutuu

spesifisesti *CDKN1A* mRNA:han. G3BP2 on tukiproteiini, joka liittyy antiviraaliseen signalointiin.

Alamut tuloksien ja nukleotidin sekä aminohapon fylogeneettisen konservoitumisen, sekä silmukointituloksien perusteella tämä variantti saattaa olla mahdollisesti harmiton.

c.6821G>T

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (5)

Hävinneet sitoutumiskohdat: HNRNPA2B1, LIN28A, dystrophin gene exon 27 ESE, g-alpha-s exon 3 ESE

Muut havainnot: Hävinnyt hsa-miR-1273e mikro-RNA-kohdesekvenssi

ClinVar tulkinta: ¼ mahdollisesti harmiton

Alamut-ohjelmien tulkinnat: 2/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Keskierto

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: HNRNPA2B1 on proteiini, joka toimii per-mRNA:n prosessoinnissa ja kuljetuksessa, ja jolla on myös silmukointia edistäviä vaikutuksia. LIN28A puolestaan säätelee tiettyjä kantasolujen toiminnalle tärkeitä geenejä. Vaikka moni algoritmi on tulkinnut variantissa olevan rikkoutunut ESE-sekvenssi, ei minkään SR-proteiinin sitoutumispaikan häviämistä todettu. ClinVar tietokannan mukaan variantti saattaa olla harmiton ja aminohapon fylogeneettinen konservoituminen on heikko, mutta Alamut-ohjelmien tulkinnat ovat ristiriitaisia. Variantin vaikutus silmukoinnin tai aminohappomuutoksen osalta on epävarma.

c.7121A>G

Havaitut silmukointimuutokset: Uusi luovuttajasekvenssi (1) sekä uusi ESS-sekvenssi (3)

Hävinneet sitoutumiskohdat: PABPC5

ClinVar tulkinta: ¼ epävarma

Alamut-ohjelmien tulkinnat: 0/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Ei konservoitumista

Aminohapon fylogeneettinen konservoituminen: Keskierto

Tulkinta: PABPC5 on mRNA aineenvaihduntaa säätelevä proteiini, joka sitoutuu mRNA:n poly-A-sekvenssiin. Vaikka Human Splicing Finder tulkitsi variantissa olevan silmukointimuutoksia, nämä muutokset eivät vastaa proteiinien sitoutumiskohtien muutoksia. Tämän lisäksi kaikki Alamut-ohjelmat ovat tulkinneet variantin harmittomaksi. Tämän lisäksi kyseisen nukleotidin fylogeneettinen konservoituminen on heikko. Näiden tietojen perusteella voi variantti olla mahdollisesti harmiton.

c.7712A>G

Havaitut silmukointimuutokset: Uusi ESS-sekvenssi (3) ja rikkoutunut ESE-sekvenssi (4)

Uudet sitoutumiskohdat: MAF, RBM5, HNRNPA2B1

Hävinneet sitoutumiskohdat: G3BP2

Muut havainnot: Hävinnyt RNA C-to-U editointikohta

ClinVar tulkinta: ¼ epävarma

Alamut-ohjelmien tulkinnat: ¾ todennäköisesti haitallinen

Nukleotidin fylogeneettinen konservoituminen: Vahva

Aminohapon fylogeneettinen konservoituminen: Vahva

Tulkinta: MAF on transkriptiofaktori, joka säätelee T-solujen apoptoosia, sekä linssin kuitujen kehitystä alkiossa. RBM5 on silmukointiproteiini, jonka on todettu olevan tärkeä tuumorisuppressorigeeni, joka säätelee solusyklin pysäyttämistä ja apoptoosia varsinkin *p53*, *FAS* ja *CASP2/caspase-2*-geenien vaihtoehdoisen silmukoinnin avulla. HNRNPA2B1 on proteiini, joka säätelee monia mRNA aineenvaihdunnan prosesseja, kuten mRNA stabiliteettia ja kuljetusta. G3BP2 puolestaan on antiviraaliselle signaaloinnille tärkeä sytoplasmisten stressijyvästen muodostamiseen tarvittava tukiproteiini.

Havaittujen silmukointimuutosten ja RBM5 sitoutumiskohdan ilmestyminen varianttisekvenssin on mielenkiintoinen havainto, tämän lisäksi valtaosa Alamut-ohjelmista on tulkinnut variantin haitalliseksi ja sekä nukleotidin, että aminohapon fylogeneettinen konservoituminen ovat vahvoja. Näiden tietojen perusteella variantti voisi olla mahdollisesti haitallinen.

c.7759C>T

Havaitut silmukointimuutokset: Uusi luovuttajasekvenssi (1)

Uudet sitoutumiskohdat: ZNF638

ClinVar tulkinta: ¼ epävarma

Alamut-ohjelmien tulkinnat: 4/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Heikko

Aminohapon fylogeneettinen konservoituminen: Vahva

Tulkinta: ZNF638 on sytidiinirikkaisiin sekvensseihin sitoutuva proteiini, joka edistää retroviraalisen genomiin integroimattoman DNA:n repressiota, ja säätelee myös rasvasolujen kehitystä. Vaikka proteiinien sitoutumiskohtien muutokset eivät tue silmukointituloksia, saattaa variantti olla proteiinitasolla haitallinen Alamut-ohjelmien tulkintojen sekä aminohapon korkean fylogeneettisen konservoitumisen perusteella.

c.7805+13A>G

Havaitut silmukointimuutokset: Uusi ESS-sekvenssi (5) ja rikkoutunut ESE-sekvenssi (2)

Hävinneet sitoutumiskohdat: Oct-1, PABPC1, PABPC4, PUM2, SART3

ClinVar tulkinta: ¼ epävarma

Tulkinta: Oct-1, joka tunnetaan myös nimellä POU2F1, on transkriptiofaktori, joka säätelee tiettyjen snRNA-geenien ja immunoglobuliinien, sekä histoni H2B:n ekspressiota. PABPC1 on poly-A-sekvensseihin sitoutuva proteiini, joka säätelee translaatiota ja silmukointia. PABPC4 sitoutuu myös poly-A-sekvensseihin, ja proteiinin on todettu säätelevän mRNA-molekyylien stabiliteettia aktivoituneissa T-soluissa. PUM2 on transkription jälkeinen repressori, joka säätelee alkion kehitystä, solujen erilaistumista, sekä virusinfektioiden tunnistamista. SART3 on silmukointikompleksin kokoonpanoa ja hajoamista edistävä proteiini.

Silmukointitulosten ja proteiinien sitoutumiskohtien häviämisten perusteella on vaikea sanoa, onko variantilla haitallista vaikutusta silmukointiin, sillä kaksi kolmesta havaitusta silmukointiproteiinin sitoutumiskohdan muutoksesta kuuluivat poly-A-sekvensseihin sitoutuville proteiineille.

c.8855T>C

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (3)

Uudet sitoutumiskohdat: Kid3 and TEL1 (TR.)

ClinVar tulkinta: Ei tietoa

Alamut-ohjelmien tulkinat: 2/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Vahva

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: Kid3, toiselta nimeltä ZNF354C, on transkriptiotekijä, joka mahdollisesti säätelee osteoblastien erilaistumista, sekä syntymän jälkeistä myogeneesiä. TEL1, toiselta nimeltään ATM, on DNA:n kaksoisjuosteen vaurioihin reagoiva kinaasi, joka säätelee muun muassa p53 ja BRCA1-proteiinien välityksellä solusykliä ja apoptoosia.

Proteiinien sitoutumiskohtien muutokset eivät vastaa silmukointituloksia, mutta ristiriitaiset Alamut tulokset sekä nukleotidin vahva fylogeneettinen konservoituminen saattavat viitata silmukointimuutokseen. Kokonaisuudessaan tämän variantin vaikutusta on vaikea arvioida.

c.8953+24T>C

Ei havaittuja silmukointimuutoksia

Hävinneet sitoutumiskohdat: HNRNPC, HNRNPCL1, RALY, U2AF2, CPEB2, CPEB4, TIA1

ClinVar tulkinta: Ei tietoa

HNRNPC on proteiini, joka sitoutuu mRNA:n UTR-alueille ja säätelee mRNA aineenvaihduntaa ja kuljetusta, sekä saattaa säädellä spesifisesti *BRCA1*- ja *BRCA2*-geenien mRNA silmukointia, ja BRCA-proteiinien vuorovaikutuksia muiden proteiinien kanssa DNA-vaurioiden korjauksessa toimivissa nukleoproteiinikomplekseissa. HNRNPCL1 on kuvailtu silmukointiproteiiniksi, kun taas RALY on silmukointiproteiini, joka toimii myös transkriptionaalisenä kofaktorina maksassa kolesterolin synteesissä. U2AF2 on intronin branch site -sekvenssille sitoutuva silmukointiproteiini, ja CPEB2 on proteiini, jonka oletetaan olevan tärkeä transkriptionaalisesti epäaktiivisten haploidisten spermati-

dien toiminnalle. CPEB4 on mRNA:n 3' UTR-alueelle sitoutuva proteiini, joka säätelee UPR (unfolded protein response) -vastetta maksassa, ja proteiinia tarvitaan myös solusyklin etenemiseen. TIA1 proteiinilla on silmukointiin ja mahdollisesti apoptoosiin liittyviä toimintoja. TIA1 sitoutuu 3' UTR-alueelle.

Tulkinta: Variantissa tapahtui erityisen paljon silmukointiin liittyvien proteiinien sitoutumiskohtien muutoksia, toisaalta Human Splicing Finder ei todennut mitään silmukointimuutoksia, joten variantin vaikutus silmukointiin jää epäselväksi.

c.9232G>T

Havaitut silmukointimuutokset: Uusi ESS-sekvenssi (2)

Uudet sitoutumiskohdat: RALY, U2AF2, CPEB2, CPEB4, HNRNPC, HNRNPCL1, ZC3H14, ZNF638

Hävinneet sitoutumiskohdat: SFPQ, SRSF2

Muut havainnot: Hävinnyt RNA C-to-U editointikohta

ClinVar tulkinta: Ei tietoa

Alamut-ohjelmien tulkinnat: $\frac{3}{4}$ haitallinen

Nukleotidin fylogeneettinen konservoituminen: Keskiaverto

Aminohapon fylogeneettinen konservoituminen: Vahva

Tulkinta: RALY on silmukointiproteiini, joka toimii myös transkriptionaalisenä kofaktorina maksassa kolesterolin synteesissä. U2AF2 on intronin branch site -sekvenssille sitoutuva silmukointiproteiini, kun taas CPEB2 on proteiini, jonka oletetaan olevan tärkeä transkriptionaalisesti epäaktiivisten haploidisten spermatidien toiminnalle. CPEB4 puolestaan on mRNA:n 3' UTR-alueelle sitoutuva proteiini, joka säätelee UPR (unfolded protein response) -vastetta maksassa, ja proteiinia tarvitaan myös solusyklin etenemiseen.

HNRNPC on proteiini, joka sitoutuu mRNA:n UTR-alueille ja säätelee mRNA aineenvaihduntaa ja kuljetusta, sekä saattaa säädellä spesifisesti *BRCA1*- ja *BRCA2*-geenien mRNA silmukointia, ja BRCA-proteiinien vuorovaikutuksia muiden proteiinien kanssa DNA vaurioiden korjauksessa toimivissa nukleoproteiinikomplekseissa. HNRNPCL1 on kuvailtu silmukointiproteiiniksi. ZC3H14 on poly-A-sekvensseihin sitoutuva proteiini, joka säätelee mRNA stabiiliteettia, kuljetusta ja translaatiota. ZNF638 on transkrip-

tiofaktori, joka säätelee rasvasolujen kehittymistä, sekä genomiin integroimattoman retroviraalisen DNA:n hiljentämistä. SFPQ proteiini säätelee monia tuman prosesseja, spesifisiä geenejä ja sillä on myös vaikutuksia silmukointiin. SRSF2 on silmukointiin tarvittava proteiini.

Variantissa havaittiin useita silmukointiin liittyvien proteiinien sitoutumiskohtien muutoksia, sekä uusia, että hävinneitä. Mielenkiintoinen havainto oli uuden HNRNPC- proteiinin sitoutumiskodan ilmestyminen varianttisekvenssiin, tällä proteiinilla saattaa olla *BRCA1*- ja *BRCA2*-spesifisiä silmukointivaikutuksia. Tosin proteiini sitoutuu UTR-alueelle, joten sillä ei välttämättä ole vaikutusta tämän variantin tapauksessa. Human Splicing Finder myös havaitsi mahdollisen uuden ESS sitoutumiskohdan ilmestymisen sekvenssiin, ja tämä tulos saattaa liittyä usean uuden silmukointiproteiinin sitoutumiskohdan ilmestymiseen varianttisekvenssiin. Variantti saattaa tulosten perusteella vaikuttaa silmukointiin, ja tulosten varjossa tätä olisi mielenkiintoista tutkia tarkemmin. Silmukoinnin lisäksi variantti saattaa olla myös proteiinitasolla haitallinen, sillä valtaosa Alamut-ohjelmista tulkitsi variantin haitalliseksi ja aminohapon fylogeneettinen konservoituminen on korkea.

c.9302T>C

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (2)

Hävinneet sitoutumiskohdat: HNRNPU, PUM2, PLZF

ClinVar tulkinta: ¼ epävarma

Alamut-ohjelmien tulkinnat: 3/3 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Keskiverto

Aminohapon fylogeneettinen konservoituminen: Keskiverto

Tulkinta: HNRNPU säätelee muun muassa telomeerien pituutta, sekä monien geenien transkriptiota ja silmukointia. PUM2 on UTR-alueille sitoutuva repressori, joka säätelee alkionkehitystä ja solujen erilaistumista. PLZF, joka on tunnettu myös nimellä ZBTB16, on transkriptiofaktori, joka säätelee solusyklin etenemistä, ja jolla on vuorovaikutuksia histonideasetyylaasin kanssa.

Tuloksien perusteella variantin vaikutuksia on vaikea arvioida. Proteiinitasolla variantti saattaa olla haitallinen, mutta silmukointitulokset ovat epämääräisiä.

c.9411T>G

Havaitut silmukointimuutokset: Uusi luovuttajasekvenssi (1)

Uudet sitoutumiskohdat: ATF2, SRSF5

Hävinneet sitoutumiskohdat: CDX1, MBNL1, RBM41, U2AF2, SRSF3, A1CF, CPEB4

ClinVar tulkinta: $\frac{3}{4}$ harmiton

Alamut-ohjelmien tulkinnat: 0/1 haitallinen (vain PolyPhen-2 tulkinta)

Tulkinta: ATF2 on transkriptiofaktori, joka säätelee monia solujakautumiseen ja DNA-vahingon korjaamiseen liittyviä geenejä. SRSF5 ja SRSF3 ovat tärkeitä silmukointiproteiineja. CDX1 on transkriptiotekijä, joka säätelee suolispesifistä geeniekspressiota, sekä enterosyyttien erilaistumista. MBNL1 on vaihtoehtoisen silmukoinnin säätelijä, kun taas RBM41 on RNA:han sitoutuva proteiini, jolla on yhteys Ebsteinin anomaliaan. U2AF2 on silmukointiproteiini, ja A1CF on keskeinen apolipoproteiini B:n RNA editoinnissa tarvittava proteiini. CPEB4 on mRNA:n 3' UTR-alueelle sitoutuva proteiini, joka säätelee UPR (unfolded protein response) -vastetta maksassa, ja proteiinia tarvitaan myös solusyklin etenemiseen.

Variantista löytyi paljon silmukointiin liittyvien proteiinien sitoutumispaikkojen muutoksia, sekä uusi mahdollinen luovuttajasekvenssi. Nämä muutokset voivat mahdollisesti olla silmukointia häiritseviä muutoksia, mutta toisaalta ClinVar tietokannan tulkinta on, että variantti olisi mahdollisesti harmiton.

c.9872C>G

Havaitut silmukointimuutokset: Kaksi uutta luovuttajasekvenssiä (1), uusi ESS-sekvenssi (2) sekä rikkoutunut ESE-sekvenssi (3)

Uudet sitoutumiskohdat: TIA1, ZC3H14

Hävinneet sitoutumiskohdat: MBNL1, SRSF5, TARDBP

ClinVar tulkinta: 2/4 epävarma

Alamut-ohjelmien tulkinnat: 4/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Vahva

Aminohapon fylogeneettinen konservoituminen: Vahva

Tulkinta: TIA1 proteiinilla on silmukointiin ja mahdollisesti apoptoosiin liittyviä toimintoja. TIA1 sitoutuu 3' UTR-alueelle. ZC3H14 on poly-A-sekvensseihin sitoutuva proteiini, joka säätelee mRNA stabiliteettia, kuljetusta ja translaatiota. MBNL1 on vaihtoehdoisen silmukoinnin säätelijä, ja SRSF5 on tärkeä silmukointiproteiini. TARDBP on 3' UTR-alueille sitoutuva silmukointia säätelevä proteiini. Tämän transkriptien lisäksi proteiini säätelee mitokondrionaalisia transkripteja.

Havaittujen silmukointimuutosten ja näitä havaintoja tukevien silmukointiproteiinien sitoutumispaikkojen muutosten, Alamut-ohjelmien tulkintojen, sekä nukleotidin ja aminohapon vahvan fylogeneettisen konservoitumisen perusteella, tämä variantti saattaa olla monella tavalla geenin toiminnalle haitallinen.

c.10234A>G

Havaitut silmukointimuutokset: Rikkoutunut ESE-sekvenssi (2)

Uudet sitoutumiskohdat: CNOT4

Hävinneet sitoutumiskohdat: YBX1, ZCRB1, SRSF3

ClinVar tulkinta: $\frac{3}{4}$ harmiton

Alamut-ohjelmien tulkinnat: 0/4 haitallinen

Nukleotidin fylogeneettinen konservoituminen: Ei konservoitumista

Aminohapon fylogeneettinen konservoituminen: Heikko

Tulkinta: CNOT4 on transkription säätelijä, joka on osa apoptoosia, immunitettä sekä kasvaimen syntyä säätelevää JAK-STAT viestintäjärjestelmää. YBX1 on DNA:ta sekä RNA:ta sitova proteiini, jolla on RNA stabiliteettiin, transkriptioon ja silmukointiin liittyviä vaikutuksia. ZCRB1 on silmukointiproteiini, joka on osa U12-tyypin splicingosomia.

Variantti vaikuttaa proteiinitasolla harmittomalta, mutta rikkoutunut ESE-sekvenssi saattaa liittyä havaittuihin silmukointiproteiinien sitoutumiskohtien häviämiseen jollain tavalla. Toisaalta ClinVar tulkinnan mukaan variantti on harmiton ja nukleotidin sekä aminohapon fylogeneettinen konservoituminen on heikkoa, joten tämä variantti ei kokonaisuudessaan todennäköisesti ole haitallinen.

Liite 5: Potilasnäytteiden sekvensointidata

Taulukko 21. Näytteiden sekvensointitiedot ovat peräisin Turun yliopiston bioinformatiikkayksikön Arfa Mehmoodin, Asta Laihon ja Laura Elon suorittamasta geeniekspression profiloinnin tutkimuksesta Elina kaikkoselle ja Johanna Schleutkerille (Project 150067, January 28 2016). Taulukkoon merkitty kaikkien luentojen määrä. Kyseessä on totaali RNA kirjasto.

Näytteen 35 tulokset ovat peräisin ”Project 150048_II Jenni Mäki-Jouppila, Stranded mRNA Sequencing” -sekvensoinnista. Sekvensointi tehtiin eturauhasen kudoksenäytteistä, joista toinen oli syöpäkudosta ja toinen normaalia kudosta. Kyseessä on mRNA kirjasto.

Näyte ja kudostyyppi	Sekvensointityyppi	Kaikki luennat (total reads, miljoonaa)	Kartoitetut luennat (total mapped reads, miljoonaa)
10 normaali	Totaali RNA kirjasto	40	37,7
32 syöpä	Totaali RNA kirjasto	36,4	33,1
35 normaali	mRNA kirjasto	31,7	29,1
35 syöpä	mRNA kirjasto	26,2	23,7
161 syöpä	Totaali RNA kirjasto	37,4	34,9
202 syöpä	Totaali RNA kirjasto	54,2	51
205 syöpä	Totaali RNA kirjasto	47,6	45,8
207 syöpä	Totaali RNA kirjasto	32,1	30,8
214 syöpä	Totaali RNA kirjasto	32,9	30,8

Liite 6: ANO7 potilasnäytteiden ddPCR-mittauksien yksityiskohtaiset tulokset

Taulukko 22. Eksoni 16 rs7593101 -variantin potilasnäytteiden ddPCR-mittauksien yhteenveto. Taulukkoon merkitty varianttialleelin suhde referenssialleeliin on Quantasoft-ohjelman ”ratio” -ikkunan antama tulos. Tuloksiin listattu mittauksien keskiarvo \bar{x} , keskihajonta σ ja vaihteluväli W. Tähdellä (*) merkityillä näytteillä kaksi 1:10 laimennettua mittausta 1:3 laimennuksien mitausten joukossa.

Näyte, sulkeissa mittauksien määrä	Varianttialleelin konsentraatio (kopiota/ μ l)	Referenssialleelin konsentraatio (kopiota/ μ l)	Varianttisaroiden määrä	Referenssaroiden määrä	Varianttialleelin suhde referenssialleeliin
10* (6)	\bar{x} 6,66 σ 2,74 W 3,15-9,47	\bar{x} 6,47 σ 3,04 W 2,01-10	\bar{x} 71,5 σ 30,66 W 33-104	\bar{x} 69,5 σ 33,69 W 21-103	\bar{x} 1,1 σ 0,28 W 0,78-1,57
32 (6)	\bar{x} 13,55 σ 0,58 W 12,9-14,3	\bar{x} 13,92 σ 0,93 W 12,4-14,9	\bar{x} 140,67 σ 13,78 W 133-157	\bar{x} 144 σ 11,1 W 125-159	\bar{x} 0,98 σ 0,06 W 0,89-1,05
35 (4)	\bar{x} 0 σ 0 W 0	\bar{x} 48,68 σ 4,06 W 42,60-51	\bar{x} 0 σ 0 W 0	\bar{x} 465,50 σ 60,18 W 396-529	Vain referenssi
161* (6)	\bar{x} 9,83 σ 4,36 W 3,96-14,3	\bar{x} 29,50 σ 13,57 W 11,60-40,3	\bar{x} 106,67 σ 50,68 W 38-157	\bar{x} 318,83 σ 160,32 W 118-481	\bar{x} 0,34 σ 0,05 W 0,28-0,42
202 (5)	\bar{x} 15,3 σ 1,59 W 13,4-17,1	\bar{x} 13,34 σ 1,21 W 11,5-14,5	\bar{x} 189,6 σ 38,08 W 156-	\bar{x} 147,6 σ 17,27 W 135-	\bar{x} 1,15 σ 0,06 W 1,07-1,21

			245	178	
205 (6)	\bar{x} 22,95 σ 2 W 19,9-25,1	\bar{x} 23,47 σ 2,07 W 20,9-26,4	\bar{x} 264,67 σ 25,91 W 241-315	\bar{x} 270 σ 9,25 W 254-279	\bar{x} 0,98 σ 0,08 W 0,92-1,13
207 (6)	\bar{x} 16,12 σ 1,35 W 14,6-17,6	\bar{x} 11,03 σ 1,77 W 8,65-13,7	\bar{x} 187,67 σ 23,39 W 157-221	\bar{x} 127,67 σ 16 W 107-155	\bar{x} 1,48 σ 0,17 W 1,29-1,69
214* (6)	\bar{x} 4,57 σ 2,69 W 1,2-6,94	\bar{x} 4,86 σ 2,23 W 1,4-6,71	\bar{x} 57 σ 35,74 W 12-90	\bar{x} 60,17 σ 30,89 W 14,87	\bar{x} 0,88 σ 0,21 W 0,52-1,06

Taulukko 23. Eksoni 8 rs2074840 -variantin potilasnäytteiden ddPCR-mittauksien yhteenveto. Taulukkoon merkitty varianttialleelin suhde referenssialleeliin on Quantasoft-ohjelman ”ratio” -ikkunan antama tulos. Tuloksiin listattu mittauksien keskiarvo \bar{x} , keskihajonta σ ja vaihteluväli W. Tähdellä (*) merkityillä näytteillä ilmeni mittauksissa vain yksi poikkeava pisara, joka oletetaan virhelähteeksi, tämän takia alleelin suhteen arvoja ei ole taulukkoon laskettu tarkasti.

Näyte, sulkeissa mittauksien määrä	Varianttialleelin konsentraatio (kopiota/ μ l)	Referenssialleelin konsentraatio (kopiota/ μ l)	Variantti pisaroiden määrä	Referenssi pisaroiden määrä	Varianttialleelin suhde referenssialleeliin
10 (6)	\bar{x} 4,31 σ 1,06 W 3,07-6,01	\bar{x} 2,97 σ 0,84 W 2,2-4,24	\bar{x} 49,67 σ 11 W 34-60	\bar{x} 35 σ 12,52 W 22-51	\bar{x} 1,56 σ 0,65 W 1,06-2,73
32* (4)	\bar{x} 23 σ 0,88 W 22-24,1	\bar{x} 0,02 σ 0,04 W 0-0,07	\bar{x} 277,75 σ 29,79 W 245-309	\bar{x} 0,25 σ 0,5 W 0-1	Vain variantti*
35 (6)	\bar{x} 23,45 σ 0,67 W 22,6-24,3	\bar{x} 19,63 σ 1,07 W 18-20,9	\bar{x} 305,83 σ 25,32 W 270-332	\bar{x} 256,83 σ 26,22 W 211-279	\bar{x} 1,2 σ 0,06 W 1,13-1,28
161 (4)	\bar{x} 32,98 σ 1,14 W 32,1-34,5	\bar{x} 0	\bar{x} 407,25 σ 43,71 W 371-462	\bar{x} 0	Vain variantti
202 (6)	\bar{x} 13,37 σ 1,47 W 10,7-14,9	\bar{x} 9,23 σ 1,09 W 7,08-10	\bar{x} 178,5 σ 22,15 W 145-204	\bar{x} 123,33 σ 14,46 W 96-134	\bar{x} 1,45 σ 0,09 W 1,33-1,54
205 (4)	\bar{x} 45,93 σ 5,79 W 38,8-51,7	\bar{x} 0	\bar{x} 560,5 σ 15,97 W 540-579	\bar{x} 0	Vain variantti
207 (4)	\bar{x} 27,78 σ 2,17 W 26,2-30,8	\bar{x} 0	\bar{x} 398 σ 19,03 W 380-421	\bar{x} 0	Vain variantti
214* (4)	\bar{x} 9,22 σ 1,24 W 8,09-10,9	\bar{x} 0,02 σ 0,03 W 0-0,07	\bar{x} 136 σ 16,39 W 119-156	\bar{x} 0,25 σ 0,5 W 0-1	Vain variantti*

Taulukko 24. Eksoni/introni 4 rs77559646 -variantin potilasnäytteiden ddPCR-mittauksien yhteenvedo. Taulukkoon merkitty referenssialleelin suhde varianttialleeliin on Quantasoft-ohjelman ”ratio” -ikkunan antama tulos. Tuloksiin listattu mittauksien keskiarvo \bar{x} , keskihajonta σ ja vaihteluväli W. Tähdellä (*) merkityillä näytteillä ilmeni mittauksissa vain yksi poikkeava pisara, joka oletetaan virhelähteeksi, tämän takia alleelien suhteen arvoja ei ole taulukkoon laskettu tarkasti.

Näyte, sulkeissa mittauksien määrä	Referenssialleelin konsentraatio (kopiota/ μ l)	Varianttialleelin konsentraatio (kopiota/ μ l)	Referenssi pisaroiden määrä	Varianttipisaroiden määrä	Referenssin suhde varianttialleeliin
10 (6)	\bar{x} 8,99 σ 2,54 W 5,99-13,4	\bar{x} 16,55 σ 4,98 W 12,8-26,2	\bar{x} 111,5 σ 40,54 W 78-191	\bar{x} 205,67 σ 81,5 W 159-371	\bar{x} 0,55 σ 0,06 W 0,46-0,65
32* (2)	\bar{x} 17,35 σ 0,78 W 16,8-17,9	\bar{x} 0,04 σ 0,05 W 0-0,07	\bar{x} 229,5 σ 21,92 W 214-245	\bar{x} 0,5 σ 0,7 W 0-1	Vain referenssi*
35 (2)	\bar{x} 31,35 σ 1,63 W 30,2-32,5	\bar{x} 0	\bar{x} 405 σ 2,83 W 403-407	\bar{x} 0	Vain referenssi
161 (2)	\bar{x} 23,85 σ 1,34 W 22,9-24,8	\bar{x} 0	\bar{x} 316,5 σ 34,65 W 292-341	\bar{x} 0	Vain referenssi
202 (6)	\bar{x} 5,65 σ 0,64 W 4,85-6,48	\bar{x} 13,83 σ 1,42 W 12-15,4	\bar{x} 79,83 σ 7,63 W 71-92	\bar{x} 194,83 σ 14,63 W 176-215	\bar{x} 0,41 σ 0,04 W 0,35-0,48
205* (2)	\bar{x} 22,3 σ 1,56 W 21,2-23,4	\bar{x} 0,03 σ 0,04 W 0-0,06	\bar{x} 333 σ 2,83 W 331-335	\bar{x} 0,5 σ 0,7 W 0-1	Vain referenssi*
207 (6)	\bar{x} 5,28 σ 0,42 W 5,05-6,13	\bar{x} 18,35 σ 1,83 W 16,3-20,6	\bar{x} 72,83 σ 11,92 W 63-96	\bar{x} 249,5 σ 19,9 W 218-268	\bar{x} 0,29 σ 0,04 W 0,25-0,35
214 (2)	\bar{x} 6,39 σ 0,66 W 5,92-6,85	\bar{x} 0	\bar{x} 90 σ 0 W 90-90	\bar{x} 0	Vain referenssi

