



**TURUN
YLIOPISTO**

Matemaattis-luonnontieteellinen
tiedekunta

DigitalMLPA:n käytön validointi kliinisessä työssä perinnöllisten syöpien tunnistuksessa

Ville Raappana

Biologia (Fysiologia ja genetiikka)

Pro gradu -tutkielma

Laajuus: 30 op

Ohjaaja(t):

Pia Pohjola

Minna Toivonen

Johanna Schleutker

Eero Vesterinen

03.05.2023

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

Pääaine: Biologia, Fysiologia ja genetiikka

Tekijä(t): Ville Raappana

Otsikko: DigitalMLPA:n käytön validointi kliinisessä työssä perinnöllisten syöpien tunnistuksessa

Ohjaaja(t): Pia Pohjola, Minna Toivonen, Johanna Schleutker & Eero Vesterinen

Sivumäärä: 52 sivua + liitteet 36 sivua

Päivämäärä: 03.05.2023

Syöpä on genomien sairaus, joka on seurausta patogeenisten mutaatioiden kertymisestä yksilön genomiin. Syöpä voi johtua ympäristötekijöistä ja perinnöllisistä tekijöistä. Näistä jälkimmäinen voidaan yhdistää perinnölliseen syöpäalttiuteen, joka on yleensä seurausta kopiolumuutoksista ituradan solujen kasvunrajoitegeneeissä. Kopiolumuutokset ovat luonteeltaan geenialueiden eksonien kahdentumia tai häviämiä, ilmeten täten mahdollisesti joko geenituotteiden yli- tai alimääränä. Yksilön periessä toiselta vanhemmaltaan kopiolumuutoksen omaavan alleelin hän omaa kasvaneen syöpäriskin riippuen geenistä. Diploidisen genomien turvin viiallisen alleelin kantaja ei täten ilmennä useimmiten heti syövän oireita. Kun yksilön toimiva alleeli mutatoituu toimimattomaksi, mutaatiota kantavat solut ajautuvat lähemmäksi syöpäsoluiksi muuntumista. Ituradan solujen geenien kopiolumuutosten havaitsemista varten on kehitetty molekyyli-genetiikan menetelmiä, joissa hyödynnetään homologista pariutumista tunnetuille geenialueille valmistetuilla koettimilla DNA:n denaturaation myötä. Multipleksiligaatiosta riippuva koettimien monistus (MLPA) on yksi tällainen menetelmä, joka on laajasti käytössä. Menetelmää varten on suunniteltu useita eri koetinsarjoja perustuen eri syöville altistavien kopiolumuutosten tunnistamiseen. Koettimien kanssa hybridisoituneet ja monistetut geenialueet tunnistetaan kapillaarisekvensoinnilla. Menetelmä on hyvä kopiolumuutosten tunnistuksessa, mutta siitä on kehitetty massasekvensointia hyödyntävä tehokkaampi versio, digitaMLPA. Massasekvensoinnin kautta sekvensoitavien geenialueiden määrä kasvaa huomattavasti, joka tarkoittaa mahdollisuutta sisältää enemmän kohdennettuja koettimia yksittäiseen reaktioon. Tämä johtaa kattavampiin syöville altistavien kopiolumuutosten seulontaan vähemmällä työmäärällä kuin MLPA:lla. Tämän työn tarkoituksena oli selvittää digitaMLPA:n luotettavuus tunnistaa TYKS Genomiikan vastualueella ennalta tutkituista potilaiden DNA-näytteistä syöville altistavat kopiolumuutokset kaupallisilla digitaMLPA koetinseoksilla. Lisäksi testauksen kohteena olivat ennalta tutkitut potilasnäytteet, joista ei ollut löydetty kopiolumuutoksia koetinseoksilla havaittavilta geenialueilta. Menetelmällä onnistuttiin havaitsemaan kaikki kopiolumuutokset sekä

varmistamaan pseudogeeni yhdestä potilasnäytteestä. Normaalin kopiolukumuutoksen omaavista näytteistä kaikki todettiin normaaleiksi. Saadut tulokset tukevat menetelmän luotettavuutta yhdessä muiden yksittäisten tutkimusryhmien tekemien digitalMLPA kokeiden kanssa.

Avainsanat: digitalMLPA, MLPA, massasekvensointi, kopiolukumuutos, syöpäalittius, iturata, perinnöllinen

Master's thesis

Subject: Biology, Physiology and genetics

Author(s): Ville Raappana

Title: Validation of the use of digitalMLPA in a clinical setting to identify hereditary cancers

Supervisor(s): Pia Pohjola, Minna Toivonen, Johanna Schleutker & Eero Vesterinen

Number of pages: 88 pages

Date: 03.05.2023

Cancer is a disease of the genome that results from the accumulation of pathogenic mutations in an individual's genome. Cancer can be caused by environmental factors and hereditary factors. The latter of these can be linked to an inherited susceptibility to cancer, which is usually the result of copy number variation (CNV) in germline cells' cancer suppressor genes. CNVs are described as duplications or deletions of exons in gene regions, thus potentially manifesting as either excess or deficiency of gene products. When an individual inherits an allele with a CNV from one parent, they have an increased risk of cancer depending on the gene. Due to the diploid genome, the carrier of the defective allele does not usually show symptoms of cancer immediately. When individual's functional allele of the same gene is mutated to non-functional variant, the cells carrying the mutation start drifting into cancer cells. Molecular genetic methods have been developed for the detection of changes in the copy number of genes in germline cells, which utilize homologous pairing with probes prepared for known gene regions after DNA denaturation. Multiplex ligation-dependent probe amplification (MLPA) is one such method that is widely used. Several different sets of probes have been designed for the method, based on the identification of copy number changes predisposing to different cancers. The gene regions hybridized with the probes are amplified and identified by capillary sequencing. The method is good for detecting copy number changes, but a more efficient version, digitalMLPA, has been developed that utilizes next-generation sequencing (NGS). Through NGS, the number of gene regions that can be sequenced increases considerably, which means the possibility of including more targeted probes in a single reaction. This leads to more comprehensive screening of cancer-predisposing copy number changes with less work compared to MLPA. The purpose of this work was to determine the reliability of digitalMLPA to identify cancer predisposing CNVs from DNA samples of patients previously examined in TYKS Genomics laboratory with commercial digitalMLPA probe mixtures. In addition, the subject of the testing were pre-examined patient samples in which no CNVs had been found in the gene regions detectable with the probe mixtures. The method succeeded in detecting all CNVs

and confirmed a pseudogene from one patient sample. Of the samples with a normal copy number change, all were found to be normal. The obtained results support the reliability of the method together with other digitalMLPA tests performed by individual research groups.

Keywords: digitalMLPA, MLPA, next-generation sequencing, copy number variance, cancer vulnerability, germline, hereditary

Sisällysluettelo

1. Johdanto	1
1.1 Ituradan solujen mutaatiot	2
1.2 Kopiolukumuutokset	3
1.2.1 Ei-alleelinen homologinen rekombinaatio	4
1.2.2 Ei-homologinen päätyliitos	6
1.2.3 Replikaatiohaarukan pysähtyminen ja templaatin vaihto	6
1.3 Massasekvensointi (NGS)	7
1.3.1 Illumina sekvensointi synteesillä	8
1.4 Geenikohdennettu häviämä- ja kahdentuma-analyysi	9
1.4.1 MLPA	10
1.5 DigitalMLPA	11
2. Aineisto ja menetelmät	14
2.1 DNA-näytteet	14
2.2 DNA-näytteiden esivalmistelu ennen digitalMLPA reaktioiden valmistamista ...	16
2.3 DigitalMLPA-reaktiot	16
2.4 DNA-kirjastojen valmistus	22
2.5 Massasekvensointi (NGS)	22
2.6 Coffalyzer digitalMLPA™ (ver. 220221.1522)	23
2.7 Koetinarvojen laskenta	24
3. Tulokset	26
3.1 Syöpänäytteet	26
3.2 Normaalinäytteet	28
3.3 χ^2 -testi	30
3.4 Datan normalisointi	30
3.5 Massasekvensointi (NGS)	32
3.6 Coffalyzer digitalMLPA™ laaturaportit	35
4. Pohdinta	36

4.1 Syöpä- ja normaalinäytteet	36
4.2 Massasekvensointi (NGS)	40
4.3 Coffalyzer digitalMLPA™ laaturaportit	41
4.4 Johtopäätökset	42
Kiitokset	45
Lähdeluettelo	46
Liitteet	53

1. Johdanto

Syöpä on genomien sairaus, jossa yksilön solujen genomiin on kertynyt patogeenisia variantteja geneista. Nämä altistavat solut onkogeenisille muutoksille ilmeten syövälle ominaisina tunnusmerkkeinä solujen fenotyypissä (Hanahan & Weinberg 2011). Tunnusmerkkeihin kuuluvat genomissa ilmenevät mutaatiot sekä usein näiden kautta ilmenevä kyky ylläpitää proliferaatiota indusoivia signaaleja samalla kiertäen proliferaatiota estävät tekijät. Näiden kahden tunnusmerkin mahdollistavat patogeeniset variantit kasvunrajoitegeneissa ja genomien ylläpitogeneissa sekä proto-onkogeneissa. Kasvunrajoitegeenit toimivat portinvartijoina solusyklin eri vaiheissa, proto-onkogeenien viedessä sykliä eteenpäin ja ylläpitogeenien estäessä mutaatioiden kertymisen solun genomiin (Weinberg 2014). Patogeenisten varianttien synnyn taustalla esiintyvät erilaiset mutaatiot ja varianttien periytyvyys riippuu missä soluissa mutaatiot ovat tapahtuneet (Deininger 1999).

Kasvunrajoitegeenien ja genomien ylläpitogeenien mutaatioiden taustalla esiintyvät usein heterotsygotian menetysmutaatiot (engl. loss of heterozygosis, LOH) sekä kopiolukumuutokset (engl. copy number variance, CNV) ilmeten ”loss-of-function” -mutaatioina, jolloin geenien aktiivisuus heikkenee tai häviää täysin (Deininger 1999; Weinberg 2014). LOH-tyypin mutaatioissa toinen kyseisen geenin alleeleista häviää, joka mahdollisesti hävittää samalla alleelin viereiset geenialueet. Tästä eroten kopiolukumuutokset ovat luonteeltaan geenin häviämiä tai kahdentumia, johtuen joko geenin ali- tai ylimäärään genomissa. Proto-onkogeneissa ilmenevät mutaatiot ovat usein geenien kahdentumia, ”gain-of-function” -mutaatioita, jotka johtavat niiden geenituotteiden yliaktiivisuuteen soluissa (Weinberg 2014).

Mutaatiot voivat tapahtua sekä yksilön somaattisissa soluissa, että ituradansoluissa, jälkimmäisen mahdollistaen tietyille syövälle altistumisen periytyminen yksilön jälkipolvelle (Weinberg 2014). Onkogeenien ja niissä ilmenevien ”gain-of-function” -mutaatioiden periytyminen on harvinaista, koska ne johtavat yleensä alkion kehityksen häiriintymiseen ja tätä kautta keskenmenoon. Patogeenisten kasvunrajoite- ja ylläpitogeenien ”loss-of function” -varianttien periytyminen taas on mahdollista. Kun häviämä ilmenee solun toisessa geenin alleelista, Knudsonin hypoteesin eli ”two-hit hypothesis” mukaan, solu kestää patogeenisen variantin solun diploidisen genomien turvin. Yksilö syntyy tällöin oman syövän kehittymiselle alistavan kopiolukumuutoksen tietystä geenistä, mikä voi heikentää geenin ilmentymistä. Vasta

molempien alleelien lakatessa toimimasta esimerkiksi kopiolumuutoksen takia, geenin luenta häiriintyy täysin johtaen onkogeeneihin muutoksiin solussa (Weinberg 2014).

Kliininen diagnostiikka, jossa selvitetään yksilön mahdollinen perinnöllinen alttius tietyn syövän puhkeamiselle, perustuu näiden ituradan soluista löytyvien patogeenisten kopiolumuutosten tunnistukseen. Muita geneettisiä muutoksia, jotka ovat selvityksen kohteena kopiolumuutosten lisäksi selvittäessä genotyypistä löytyviä muutoksia tautifenytyypin selitykseksi, ovat kromosomaaliset rakennemuutokset ja patogeeniset geenivariantit (MedlinePlus 2023). Näiden lisäksi myös epigeneettiset muutokset, kuten DNA:n metylaatio on selvityksen kohteena (Beltrán-García ym. 2019). Edellä mainittujen genomiin kohdistuvien muutosten tunnistuksessa käytetään hyväksi molekyyli-genetiikan menetelmiä, joiden kautta pystytään kartoittamaan yksilön genomista patogeeniset muutokset avustaen oikean diagnoosin tekemisessä.

1.1 Ituradan solujen mutaatiot

Ituradansolujen mutaatiot ilmenevät somaattisissa soluissa niiden kasvaneena herkkyytensä muuntautua syöpäsoluiksi. Esimerkkejä tällaisista tilanteista ovat perinnölliset mutaatiot DNA:n virheen korjausgeeneissä (engl. mismatch repair, MMR) ja homologisissa korjausgeeneissä (engl. homologous repair, HR) (Vali-Pour ym. 2022). MMR pyrkii korjaamaan DNA:n replikaation jälkeisiä virheitä, kuten spontaanit väärät emäspariutumiset ja insertiot (Pećina-Šlaus ym. 2020). Näihin geeneihin lukeutuvat eukaryooteissa *MSH2*, *MSH6*, *PMS2* sekä *MLH1*, joiden lopputuotteet toimivat heterodimeereinä luoden keskenään proteiinikomplekseja virheellisten emästen tunnistuksessa. Esimerkiksi *MSH2*- ja *MSH6*-proteiinit luovat DNA-sekvenssiä pitkin kulkevan ”sliding clamp”-rakenteen, joka tunnistaa replikaatiohaarukassa syntyneitä virheitä. Heterodimeerin osuessa virheellisen emäsparin kohdalle siihen liittyy *PMS2/MLH1*-heterodimeeri ja muodostunut proteiinikompleksi irrottaa virheellisen emäksen juosteesta (Pećina-Šlaus ym. 2020). Emäksen irtoamista seuraa DNA-synteesin uudelleen käynnistyminen kyseisessä kohdassa juostetta.

DNA:n virheen korjaus geenien *MSH2*, *MSH6* tai *PMS2* ituradansoluissa ilmenevien heterotsygoottisten häviämien periytyminen johtaa kehittyvän sikiön sairastuvan Lynch-syndroomaan (Idos & Valle 2021). Syndrooman myötä edellä kerrottu MMR proteiinien välisiin heterodimeereihin perustuva mekanismi häiriintyy. Mekanismin häiriintyminen

johtaa yksilön omaavan korkeamman alttiuden sairastua suolistosyöpään sekä moniin muihin syöpiin, joihin lukeutuvat muun muassa kohdunkaulan- munasarjan-, eturauhasen- ja vatsansyöpä.

Vastaavasti homologisiin rekombinaatiomekanismeihin lukeutuvia geenejä ovat: *BRCA1*, *BRCA2*, *RAD51C* ja *PALB2* (Vali-Pour ym. 2022). Näiden geenien tuotteet korjaavat rikkoumia DNA:n kaksoisjuosterakenteessa homologisen rekombinaation kautta, joiden toimimattomuus usein johtaa niitä korvaavien korjausmekanismien aktivaatioon solussa. Korvaavat mekanismit ovat yleisesti paljon alttiimpia tuottamaan mutaatioita DNA:han, niiden vaatiessa ainoastaan mikrohomologian yhdistettävien juosteiden välillä. Homologista rekombinaatiota korvaaviin mekanismeihin kuuluvat muun muassa ei-homologinen päätyliitoskohdan korjaus (engl. non-homologous end-joining, NHEJ).

Esimerkkinä homologisiin rekombinaatiogeeneihin lukeutuvista periytyvistä geenimutaatioista ovat autosomaalisesti dominantisti periytyvä *BRCA1*- ja *BRCA2*-geeneihin liittyvät rinta- ja munasarjasyöpä (Petrucci ym. 2022). Suurin osa mutaatioista, jotka toimivat patogeenisinä variantteina molemmista geneistä, ovat lukukehysmutaatioita johtuen geneissä tapahtuneista häviämistä (Petrucci ym. 2010). Periytyviin *BRCA1* ja *BRCA2* geeneihin linkittyy lisäksi alttius sairastua muihin syöpiin, riippuen kumman patogeenisen variantin yksilö perii vanhemmiltaan. Esimerkiksi *BRCA2*-variantteihin linkittyy kasvanut riski sairastua melanoomaan (Petrucci ym. 2022).

1.2 Kopolukumuutokset

Kopolukumuutosten on tunnistettu olevan laajasti levinneitä ihmisissä, luoden suuren osan geneettisestä variaatiosta. Arviolta noin 12 % ihmisen genomista on altis kopolukumuutoksille ja ne voivat olla joko meiotista tai somaattista alkuperää (Hastings ym. 2009). Riippuen mikä geeni on kyseessä, CNV voi ilmetä neutraalina, positiivisena tai negatiivisena muutoksena yksilön fenotyypissä. Yksi positiivinen muutos CNV pohjalta esiintyy *CCL3L*-geenin ylimääränä, joka ilmenee yksilön kasvaneena vastustuskykynä HIV-virusta vastaan. Suurin osa variaatiosta CNV:n pohjalta on kuitenkin negatiivista, joka ilmenee erilaisina genomisina häiriöinä johtaen muun muassa yksilön kasvaneeseen alttiuteen sairastua syöpään.

Kopiolukumuutos vaatii muodostuakseen muutoksen kromosomin rakenteessa, jossa kaksi aikaisemmin toisistaan erillään olevaa DNA-juostetta yhdistyvät (Hastings ym. 2009). Tähän liittyen CNV:n syntymekanismeiksi on ehdotettu kolmea eri mekanismia: ei-alleelinen homologinen rekombinaatio (NAHR), ei-homologinen päätyliitos sekä replikaatiohaarukan pysähtyminen ja templaatin vaihto (FosTes) (Zhang ym. 2009). Näistä NAHR ja NHEJ ovat ei-replikatiivisia rekombinaation muotoja eli ne eivät tapahdu DNA:n replikaation aikana, kun taas FosTes tapahtuu DNA:n replikaation aikana.

1.2.1 Ei-alleelinen homologinen rekombinaatio

Ei-alleelisessa homologisessa rekombinaatiossa tekijäinvaihto tapahtuu mitoosissa tai meioosissa kahden ei-alleelisen DNA-sekvenssin välillä, jotka omaavat korkean homologian keskenään (Hastings ym. 2009). Mitoottiset NAHR:t tuottavat mosaiikkisolupopulaatioita, joissa yksilön somaattiset solut omaavat eri kopiolukumuutoksia (Zhang ym. 2009). Meioottiset NAHR:t johtavat epätasaiseen tekijäinvaihtoon kromosomien välillä sukusoluissa, joka johtaa genomien uudelleen järjestäytymiseen. Kopiolukumuutokset, jotka ovat seurausta meioottisesta NAHR:sta voivat periytyä kantajan jälkipolvelle johtaen mahdollisiin perinnöllisiin genomisiin häiriöihin.

Korkean homologian takia ei-alleelisten homologisten rekombinaatioiden mahdollistajana esiintyvät usein suuret yli 10 kb kokoiset matalat toistot (engl. low-copy repeats, LCR) (Cardoso ym. 2016). Matalia toistoja yleisesti luonnehditaan ≥ 1 kb DNA-fragmenteiksi, jotka omaavat korkean homologian ($>97\%$) keskenään ilmeten useina kopioina genomissa (Harel ym. 2015; Cardoso ym. 2016). Arviolta noin 5 % ihmisen genomista koostuu matalista toistoista (Harel ym. 2015). Matalat toistot pystyvät indusoimaan rakenteellisia muutoksia kromosomeissa, kun kaksi paralogista matalaa toistoa sijaitsee alle 10 Mb päässä toisistaan.

Riippuen kuinka paralogiset matalat toistot ovat sijoittuneet toisiinsa nähden NAHR aikaisessa tekijäinvaihdossa, ne indusoivat erilaisia kopiolukumuutoksia genomiin (Harel ym. 2015). Tekijäinvaihto homologisten matalien toistojen välillä, jotka sijaitsevat eri kromosomeissa tai vastaavasti saman kromosomin sisarkromatideissa, johtavat sekä geenin häviämään että kahdentumaan. Jos tekijäinvaihtoon osallistuu kääntyneet

homologiset matalat toistojaksot sisarkromatideissa, tuloksena voi syntyä isokromosomi rakenne. Tekijäinvaihdon tapahtuessa samassa kromatidissa kahden homologisen toistojakson välillä, johtaa geenin häviämään rengasrakenteen muodostumisen myötä. Muodostuneesta rengasrakenteesta uupuu sentromeeri, jolloin rakenne ei kykene osallistumaan seuraavaan solun jakautumista edeltävään mitooseen (Turner ym. 2008).

Matalien toistojen lisäksi tutkimustulokset viittaavat, että korkeisiin toistoihin (engl. high-copy repeats) lukeutuvat retrotransposonit olisivat osatekijöinä ei-alleelisessa homologisessa rekombinaatiossa patologisten toistojaksojen muodostuksessa (Cardoso ym. 2016). Retrotransposonit ovat ihmisgenomissa esiintyviä siirtyviä elementtejä, jotka kykenevät monistumaan ja liittämään monistamiaan alueitaan takaisin genomiin RNA-välitteisesti (Krebs E. ym. 2013). Niihin lukeutuvat ”short interspersed elements” (SINEs) sekä ”long interspersed elements” (LINEs), jotka yhdessä kattavat noin 34 % ihmisgenomista. Yleisesti molemmat LINE:t ja SINE:t kykenevät häiritsemään geenien luentaa, joka johtaa esimerkiksi geenin luennan ennenaikaiseen päättymiseen RNA välitteisillä insertioilla.

Kahdesta siirtyvästä elementistä SINE:t ovat noin 300 bp mittaisia ei-autonomisia retrotransposoneja, joiden runsain ja tunnetuin ryhmä ovat *Alu*-elementit (Krebs E. ym. 2013). *Alu*-elementtien välisten NAHR:n on tunnistettu aiheuttavan *BRCA1* ja *STK11* geneissa patogeenisia syöville altistavia kopiolumuutoksia (Cardoso ym. 2016). Arvioituna noin 10–15 % ituradan soluissa esiintyvistä kopiolumuutoksista *BRCA1*-geenissä johtuvat isoista genomisista rakenteellisista muutoksista, joissa *Alu*-elementit ovat osallisena (Peixoto ym. 2013). Syynä *Alu*-elementtien rooliin kopiolumuutosten synnyssä potentiaalisesti johtuu niiden korkeasta lukumäärästä *BRCA1*-geenissä. Yksittäisiä *Alu*-elementtejä *BRCA1*-geenissä on tunnistettu olevan 138 (Ewald ym. 2009). Esimerkkitapaus *Alu-Alu* välisestä NAHR:sta ituradan soluissa, joka johtaa HBOC altistavaan fenotyypin, ilmenee saksalaisissa perheissä. Näissä perheissä esiintyy *BRCA1* eksonien 5–7 häviämä, joka johtuu NAHR:sta *AluSx* introni 3 ja *AluSc* introni 7 välillä (Ewald ym. 2009).

BRCA1-geenin tavoin *STK11*-geenissä esiintyy useita yksittäisiä *Alu*-elementtejä, jotka kattavat 19 % koko geenistä (Borun ym. 2015). *STK11*-geeni koostuu 10 eksonista, joista 9 ovat proteiineja koodaavia ja näistä eksonien 2 ja 3 häviämät on yhdistetty *Alu-Alu* välisiin NAHR:n geenistä löytyvien *AluY*-sekvenssien välillä. Eksonien 2 ja 3 häviämät johtavat henkilön sairastuvan Peutz-Jeghersin syndroomaan, jota ilmenee globaalisti

saman syyn takia. Syndrooma ilmenee suolistosta löytyvinä polyyppeinä, limakalvojen tummentumina sekä kasvaneena alttiutena epiteelikudoksen pahanlaatuisiin kasvaimiin (McGarrity ym. 2021).

1.2.2 Ei-homologinen päätyliitos

Ei-alleelisesta homologisesta rekombinaatiosta poiketen ei-homologinen päätyliitos ei vaadi homologisia sekvenssejä, kuten matalia toistojaksoja DNA:n kaksoisjuosteen korjaukseen (Harlen ym. 2015). NHEJ on joustavampi korjauksen käynnistyksessä, mutta samalla alttiimpi tuottamaan mutaatioita yksilön genomiin kuin NAHR. Korjauksen käynnistää Ku70/80-heterodimeerin liittyminen katkenneeseen DNA-juosteeseen, luoden Ku/DNA-kompleksin, joka toimii alustana korjaukseen vaativille molekyyleille: nukleaasille, polymeerasille sekä ligaasille (Lieber 2010). Molekyyleillä ei ole ennalta määrättyä järjestystä, jossa ne liittyvät kompleksiin ja tämän myötä DNA:n korjauksen lopputulemat vaihtelevat ilmeten muun muassa kopiolumuutosten syntynä (Lieber 2010). Kopiolumuutokset NHEJ pohjalta vaihtelevat 1–4 bp mittaisten häviämien sekä insertioiden välillä homologian uupuessa DNA:n liitoskohdissa (Hastings ym. 2009; Chen ym. 2014). Insertiot NHEJ seurauksena ovat peräisin solusta löytyvästä vapaasta DNA:sta, kuten mitokondrionaalisesta tai retrotransponisesta DNA:sta, jotka ajautuvat osaksi Ku/DNA-kompleksia (Hastings ym. 2009).

1.2.3 Replikaatiohaarukan pysähtyminen ja templaatin vaihto

NHEJ ja NAHR ovat ei-replikatiivisia rekombinaation muotoja, mutta replikaatiohaarukan pysähtyminen ja templaatin vaihto tapahtuu DNA:n replikaation aikana (Hastings ym. 2009). FosTes:ssa jälkimmäinen juoste (engl. lagging strand) irtoaa paikallisesti polymeerasista ja irrottautuu alkuperäisestä mallijuosteesta. Irronnut juoste ajautuu kiinnittymään toiseen polymeerasi/DNA-kompleksiin, jossa oleva mallijuoste omaa mikrohologian (4–15 bp) sen 3'-pään kanssa. kääntyneet toistojaksot (IR) sekä replikaation aikainen rasite voivat johtaa FosTes:n (Arlt ym. 2012; Chen ym. 2014).

Kääntyneet toistojaksot, jotka omaavat korkean homologian, voivat muodostaa sekundäärisiä DNA-rakenteita keskenään ilmeten esimerkiksi hiuspinnirakenteina DNA-juosteessa (Chen ym. 2014). Hiuspinnirakenteet toimivat juosteessa esteinä

polymeraasille pysäyttäen ja käynnistäen uudestaan sen kulun, jolloin IR:n väliset alueet häviävät replikaatiossa. Templaatin vaihdon tapahtuessa replikaatiohaarukan pysähtymisen yhteydessä häviämän lisäksi syntyy myös kahdentuma siitä alueesta, joka syntetisoidaan toisessa replikaatiohaarukassa. Riippuen syntetisoidun alueen laajuudesta irronneen jälkimmäisen juosteen pohjalta sekä kuinka monta kertaa templaatin vaihto tapahtuu saman juosteen pohjalta, IR:n kautta solun genomiin voi muodostua hyvin laajoja kopiolumuutoksia.

Replikaation aikaisessa rasitteessa replikaatiohaarukan normaali toiminta estyy sitä inhiboivien tekijöiden takia (Arlt ym. 2012). Tällaisia tekijöitä ovat muun muassa apidikoliini (APH) sekä hydroksiurea (HU). APH toimii DNA-polymeraasin inhibiittorina ja HU inhibiittorina ribonukleotidireduktaasille sekä häiritsevänä tekijänä solun vapaille nukleotideille. Arlt ym. (2012) testasivat HU:n kykyä tuottaa kopiolumuutoksia verrattaessa APH:lla ja HU:lla käsiteltyjen ihmisperäisten fibroblastien genomisekvensseistä löytyviä kopiolumuutoksia. Heidän tuloksensa viittasivat kahden käsittelyn olevan lähes yhdenvertaiset. Tulokset varmensivat HU:n toimivan mahdollisena mutageeninä, mutta lisäksi kuinka replikaatioon kohdistettu rasite voi tuottaa kopiolumuutoksia soluissa.

1.3 Massasekvensointi (NGS)

Seuraavan sukupolven sekvensointimenetelmät (engl. Next-generation sequencing, NGS) ovat kasvavissa määrin yleistymässä biotieteissä, tarjoten tehokkaampia ja edullisempia vaihtoehtoja sekvensointiin. Tehokkuuteen liittyen NGS-menetelmillä mahdollistuu usean eri DNA-näytteen luenta samanaikaisesti, PCR-pohjaisia menetelmiä hyödyntäen verrattaessa ensimmäisen sukupolven Sanger-sekvensointiin, jonka käyttö rajoittuu yhden DNA-sekvenssin luentaan kapillaarisekvenssoinnissa (Illumina 2022a). Massaluennan avulla NGS-menetelmät tarjoavat erinomaisen soveltuvuuden kliiniseen työhön, jossa potilasnäytteiden diagnostiikka kohdistetaan useammalle DNA-näytteelle kerrallaan, esimerkiksi syövälle altistavien kopiolumuutosten kartoituksessa sekvensointianalyysissä.

Yleisesti massasekvensoinnin työnkulku DNA-näytteillä voidaan jakaa neljään vaiheeseen: DNA:n pilkkomiseen, DNA-kirjaston valmistukseen, sekvensointiin ja datan tulkintaan (Qin D. 2019). DNA:n pilkonnassa tutkittava DNA pilkotaan pienempiin osiin

joko mekaanisesti esim. sonikaatiolla tai entsyymaattisesti spesifeillä restriktioentsyymeillä. Työn kannalta merkittävät DNA-fragmentit voidaan tunnistaa ja kerätä käyttämällä hybridisaatiokaappauskoetta (engl. hybridisation capture assay) tai vaihtoehtoisesti monistaa amplikonikokeen kautta (engl. amplicon assay). Hybridisaatiokaappauksessa käytetään hyväksi vastinalukkeita, jotka hybridisoituvat haluttuihin DNA-fragmenteihin eristäen ne muista ei-halutuista fragmenteista. Jälkimmäisessä, amplikonikokeessa, käytetään myös vastinalukkeita, mutta alukkeiden kanssa hybridisoituneet alueet monistetaan polymeerasiketjureaktion avulla. Molemmat metodit vaativat DNA:n denaturaation eli DNA:n kaksoiskierrerakenteen purkautumisen toimiakseen, jotta alukkeet pystyvät hybridisoitumaan kohde sekvensseihin.

Pilkkomista ja haluttujen DNA-alueiden keräämistä seuraa DNA-kirjaston laatiminen, jossa alukkeiden kanssa hybridisoituneet sekvenssit merkitään sekvensointia varten (Qin D. 2019). Merkintä voidaan suorittaa joko indeksi- tai viivaakoodioligonukleotideilla, jotka joko sisältyvät tai ne liitetään osaksi kohde DNA-sekvensseihin hybridisoituneita vastinalukkeita. Näitä tunnistussekvenssejä vaaditaan kirjaston luomisen jälkeisessä sekvensoinnissa, jotta yksittäisen näytteen tunnistus mahdollistuu. Esimerkkinä tällä hetkellä yleisimmällä sekvensointialustalla, Illuminan sekvensointisynteesillä (engl. sequencing by synthesis, SBS), sekvensointi suoritetaan fluoresoivilla nukleotideilla, joiden pohjalta rakennetaan kopioita kohde DNA:n johtavasta, että käänteisestä juosteesta (Illumina 2022b). Sekvensoinnin aikana fluoresoivien nukleotidien liittyessä emäsjärjestyksen mukaisesti kohdesekvenssiin polymeerasientsyymien toimesta, saadaan listattua tarkasti sen emäsjärjestys lopullista analyysia varten. Soveltuvia analyysimenetelmiä hyväksi käyttäen pystytään esimerkiksi vertaamaan sekvensoitua DNA:ta genomitietokantoihin mahdollista tunnistusta tai poikkeavuuksien löytämistä varten (Qin D. 2019).

1.3.1 Illumina sekvensointi synteesillä

Illuminan SBS-teknologia perustuu virtauskyvettien (engl. flow cell) käyttöön, jotka sisältävät oligonukleotideilla päällystettyjä kaivoja, joihin yksijuosteiset näytteet hybridisoituvat adaptorisekvenssien kautta (Illumina, 2022b). Hybridisaatiota seuraa siltarakenteen muodostuminen yksijuosteisen sekvenssien vapaiden päiden liittyessä virtauskennoilla oleviin adaptorisekvensseihin, jolloin käynnistyy komplementaarisen

juosteen muodostus polymeraasientsyymien toimesta. Kaksijuosteisen rakenteen muodostuksen myötä juosteet erkanevat toisistaan ja sama edellä mainittu siltarakenteen sekä komplementaarisen juosteen muodostuminen toistuvat useasti luoden useita 5'-3' sekä 3'-5' suuntaisia yksijuosteisia kopioita alkuperäisestä näytteestä.

Valmistettujen kopioiden pohjalta suoritetaan näytteen sekvensointi eli sen emäsjärjestyksen tunnistaminen fluoresoivien deoksinukleosiditriposfaattien (dNTPs) avulla (Illumina, 2022b). Sekvensointi alkaa 3'-5' suuntaisten käänteistenjuosteiden irtaantumisen virtauskennolta jättäen kennolle ainoastaan 5'-3' suuntaiset johtavajusteet. Jäljelle jääneisiin juosteisiin kiinnittyy sekvensointialue 1 estäen ei-halutun hybridisaation virtauskennolla, samalla mahdollistaen polymeraasientsyymien kiinnittymisen juosteeseen. Alukkeen liittymisen myötä polymeraasientsyymi alkaa liittämään fluoresoivia dNTP:tä juosteeseen emäsjärjestyksen mukaisesti. Kiinnittyminen pystytään havaitsemaan sekvensointilaitteistolla fluoresoivan merkkiaineen vapautuessa dNTP:stä. Vasta kun merkkiaine on vapautunut, polymeraasientsyymi pystyy liittämään seuraavan dNTP:n osaksi valmistuvaa ketjua.

Ensimmäisten lukutuotteiden valmistuttua johtavien juosteiden pohjalta, ne irtoavat juosteesta mahdollistaen yksijuosteisen siltarakenteen ja edelleen kaksijuosteisen rakenteen uudelleen muodostumisen (Illumina, 2022b). Sekvensoinnin seuraavassa vaiheessa kaksoisjuoste purkautuu jälleen, jolloin käänteiseen juosteeseen kiinnittyy sekvensointialue 2 ja fluoresoivien dNTP:n pohjalta suoritetaan käänteisenjuosteen mukainen komplementaarisen juosteen muodostus. Molemmat sekvensoinnin vaiheet tapahtuvat useita kertoja luoden tarkan kartan näyte DNA:n emäsjärjestyksestä. Riippuen suoritettavasta työstä, sekvensointi voidaan myös suorittaa ainoastaan johtavan juosteen pohjalta.

1.4 Geenikohdennettu häviämä- ja kahdentuma-analyysi

Kopiolukumuutosten ollessa mahdollisesti suuria (>50 bp) tai koko geenin kattavia häviämiä tai kahdentumia, niiden tunnistus sekvensointianalyysimenetelmillä on haastavaa (Kerkhof ym. 2017). Tunnistus on mahdollista, jos kopiolukumuutoksen leikkauskohdat sisältyvät monistettavan alueen PCR-alukkeiden välille kapillaarisekvensoinnissa tai vastaavasti massasekvensoinnissa monistuksen kohteena olevaan alueeseen (GeneReviews 2022). Haasteellisuuden takia kopiolukumuutosten

tunnistusta varten on kehitetty niille tarkoitettut geenikohdennetut häviämä- ja kahdentuma- analyysimenetelmät.

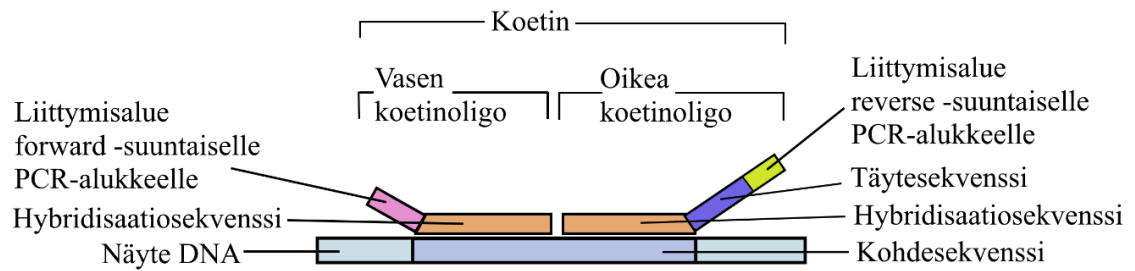
Geenikohdennettu häviämä- ja kahdentuma-analyysi on suotavaa tehdä, jos alkuoletuksena henkilön sen hetkinen tila on potentiaalisesti seurausta kopiolumuutuksesta (Luukkonen 2020). Analyysi on myös hyödyllinen, kun halutaan selvittää peittyvästi periytyviä sairauksia aiheuttavia alleeleja, jotka ovat seurausta kopiolumuutuksesta jo ennen sairauden puhkeamista. Näiden lisäksi, kun sekvensointianalyysin tulos on negatiivinen, geenikohdennettu häviämä ja kahdentuma-analyysi on hyvä sisällyttää jatkotutkimuksiin (Luukkonen 2020). Negatiivinen tulos voi olla seurausta jo mainitusta sekvensointianalyysille epäsoveltuvasta liian suuresta kopiolumuutuksesta. Geenikohdennettuun häviämä- ja kahdentuma-analyysimenetelmiin kuuluvat muun muassa jo laajasti käytössä oleva multiplexi ligaatiosta-riippuva koettimien monistus (engl. Multiplex ligand-dependent amplification, MLPA) sekä tästä edelleen kehitetty digitalMLPA.

1.4.1 MLPA

MRC Hollandin kehittämä MLPA on yksi kliinisessä työssä jo laajasti käytetty geenikohdennettu häviämä ja kahdentuma analyysimenetelmänä. MLPA perustuu DNA-näytteiden kapillaarisekvensointiin, joiden tunnistuksessa sekä erittelyssä käytetään hyväksi tietyille alleeleille valmistettuja vastinkoettimia (MRC Holland 2022a). Vastinkoettimet ovat kaksiosaisia, jonka myötä tunnistusta varten molempien osien pitää hybridisoitua kohdesekvenssiin sekä yhdistyä keskenään ligaatioreaktiossa (Kuva 1.) MLPA:n avulla pystytään diagnosoimaan potilaan DNA-näytteestä mahdolliset geenien kahdentumat hybridisoituneiden koettimien ylimääränä tai häviämät koettimien hybridisaation epäonnistuessa, joka ilmenee koettimien alimääränä lopullisessa analyysissä.

MLPA:n työnkulku alkaa kaksiosaisien koettimien hybridisaatioreaktiolla tutkittaviin DNA-näytteisiin (MRC Holland 2022a). Koettimien hybridisaation jälkeen suoritetaan koettimien monistus PCR-reaktiolla, jossa hyödynnetään fluoresoivaa forward -suuntaista PCR-aluketta. MLPA:ssa ei siis monisteta koko kohde DNA:ta vaan ainoastaan koettimien kanssa hybridisoituneet alueet. Fluoresoivien alukkeiden avulla

onnistuneesti hybridisoituneet ja edelleen monistuneet alueet tunnistetaan kapillaarisekvensoinnin jälkeisessä analyysissä.



Kuva 1: MLPA-reaktioissa käytettävä koetin, joka koostuu kahdesta osasta: vasen koetinoligo sekä oikea koetinoligo (MRC Holland 2022a). Koettimen osat kiinnittyvät kohdesekvenssiin hybridisaatiosekvenssien avulla. Muita tärkeitä koettimessa olevia alueita ovat PCR-alukkeiden kiinnityskohdat sekä täytesekvenssi oikeassa koetinoligossa, joka antaa jokaiselle koettimelle uniikin koon (MRC Holland 2022a). Forward -suuntainen PCR-alue omaa fluoresoivan merkkiaineen, jota käytetään hyväksi kapillaarisekvensoinnissa. Kuva pohjautuu MRC Holland. (2022a). Principle of MLPA. <<https://www.mrcholland.com/technology/mlpa/technique>> [Luettu: 4.4.2022].

1.5 DigitalMLPA

MLPA on hyvä työkalu geenien kopiolumuutosten tunnistuksessa, mutta MRC Holland on kehittänyt siitä tehokkaamman NGS-teknologiaa hyödyntävän version, digitalMLPA. Tämä uudempi versio perustuu samaan konseptiin kuin MLPA eli kaksiosaisten koettimien hybridisaatioon, niiden väliseen ligaatioon sekä alukkeiden kanssa hybridisoituneiden DNA-alueiden monistukseen (MRC Holland 2022b). Poiketen MLPA:sta, sekvensointi suoritetaan Illumina-pohjaisella sekvensointialustalla, joka mahdollistaa usean näytteen rinnakkaisen sekvensoinnin. Sekvensoinnin jälkeisessä näytetunnistuksessa käytetään fluoresoivien PCR-alukkeiden sijaan koettimiin liittyviä uniikkeja 10 nukleotidin mittaisia viivakoodioligoja (Kuva 2.). Jokaiselle viivakoodille on määritelty oma spesifi ID-arvo, joiden pohjalta sekvensoinnin jälkeisessä analyysissä käytettävällä ohjelmalla, Coffalyzer digitalMLPA suoritetaan näytteiden tunnistus (MRC Holland 2022b).

Ohjelma lajittelee sekvensoidut näytteet ryhmiin perustuen geneihin hybridisoituneisiin koettimien omaaviin viivakoodioligoihin. Kopiolumuutosten todentaminen ilmenee sen mukaan, kuinka monta tiettyä koetinta on liittynyt yhteen geeniin (MRC Holland

2021). Normaalissa diploidissa tilassa ainoastaan kaksi koetinta liittyy geenien eksoneihin, jolloin ohjelman antaa koetinarvon, joka on lähellä arvoa yksi (Taulukko 1.). Kahdentumasta johtuneessa triploidisessa tilassa kyseinen arvo olisi noin 1,5 tai häviämisen myötä haploidisessa tilanteessa arvo olisi noin 0,5. Homotsygoottisessa häviämässä eli molempien alleelien uupuessa, koetinarvo on nolla.

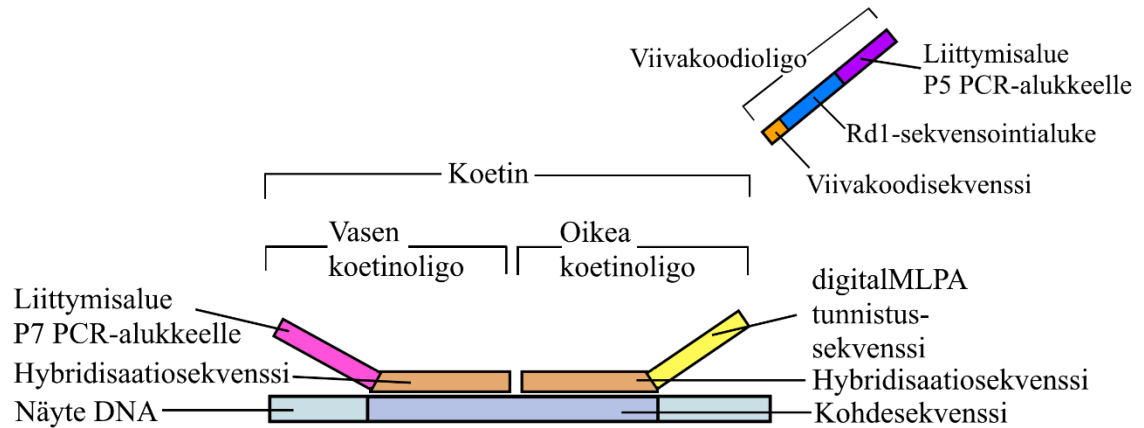
DigitalMLPA:ta verrattaessa MLPA:han, reaktioiden valmistukseen vaaditaan ≥ 20 ng DNA:ta sekä yksittäiseen reaktioon voidaan sisällyttää jopa 1000 koetinta (Benard-Slagter ym. 2017). MLPA vastaavasti vaatisi vähintään 50 ng kohde DNA:ta sekä yksittäinen reaktio on rajoitettu 60 koettimeen. Kliinisessä asetelmassa tämä tarkoittaisi työmäärän vähentymistä reaktioiden sarjakoona kasvaessa samalla säästäten vaaditun potilas DNA:n määrässä.

Menetelmä on tarkoitettu ainoastaan kopiolumuutosten tunnistukseen DNA-näytteistä, luoden käyttäjäturvallisen alustan tätä varten. Se omaa tiettyjä rajoituksia. Selvimpänä rajoittavana tekijänä ovat neutraalit kopiolumuutokset, jotka eivät muuta geenin lukukehystä sekä mutaatiot, jotka jäävät koettimien kohdesekvenssien ulkopuolelle. MRC Hollandin (2021) listaamia geenien lukukehystä muuttamattomia mutaatioita ovat geenien neutraalit kääntymät ja siirtymät.

Neutraalien kopiolumuutosten lisäksi menetelmällä ei pystytä havaitsemaan tuntemattomia pistemutaatioita (MRC Holland 2021). Tämä johtuu siitä, että koettimet ovat suunniteltu hybridisoitumaan tunnettuihin DNA-sekvensseihin, jonka yksittäinen emäksen muutos voi estää. Coffalyzer-ohjelman antamassa tulosteessa tuntematon pistemutaatio ilmenee täten virheellisenä geenin häviämänä.

Liittyy menetelmän rajoituksiin, mutta ei molekyyli-genetiikan teorian osalta, menetelmän valmistaja, MRC Holland ei ole vielä todennut menetelmän soveltuvan kliiniseen työhön. Menetelmänä digitalMLPA on valmistajan toimesta validi ainoastaan tutkimustyötä varten eli menetelmän kautta saaduilla tuloksilla ei saa pohjustaa potilasnäytteisiin perustuvia lausuntoja (MRC Holland 2022c). Tästä huolimatta yksittäiset toimijat ovat selvittäneet menetelmän luotettavuutta käyttämällä potilasnäytteitä tutkimusasetelmassa, joiden on todettu omaavan patogeenisille muutoksille omaavia kopiolumuutoksia. Alkuperäinen todentaminen patogeenisistä kopiolumuutoksista on suoritettu pohjautuen kliinisessä työssä hyväksi luettuja menetelmiä käyttäen, kuten MLPA:lla.

Tämän työn tarkoituksena on selvittää digitalMLPA:n luotettavuus kliinisessä työssä potilasnäytteiden diagnostiikassa syövälle altistavien kopiolumuutosten tunnistuksessa Turun yliopistollisen keskussairaalan (TYKS) Genomiikan vastuualueella. Jos menetelmä todetaan luotettavaksi, sitä varten laaditaan käyttöohjeet ja validointiraportti Genomiikan vastuualueelle.



Kuva 2: DigitalMLPA-reaktiossa käytettävä koetin, joka koostuu kahdesta osasta, kuten MLPA-reaktiossa käytettävä koetin (MRC Holland 2022b). Koetin koostuu muuten samoista alueista, mutta sen oikeassa koetinoligossa on täytesekvenssin tilalla digitalMLPA tunnistussekvenssi. Sen avulla tunnistetaan digitalMLPA luennat FASTQ-tiedostossa. Reaktiossa käytetty viivakoodioligo liittyy osaksi koettimen oikeaa koetinoligoa digitalMLPA tunnistussekvenssin jatkoksi. Viivakoodioligossa oleva Rd1-sekvenssi vaaditaan luennan käynnistymiseksi sekvensoinnissa (MRC Holland. 2022b). Kuva pohjautuu MRC Holland. (2022b). Principle of digitalMLPA. <<https://www.mrcholland.com/technology/digitalmlpa/technique>> [Luettu: 4.4.2022].

Taulukko 1: Eri kopiolumun tiloja vastaavat koetinarvot, jotka ovat määriteltä digitalMLPA menetelmälle (MRC Holland. 2021.).

Kopiolumun tila	Koetinarvo
Normaali	0,80 < arvo < 1,20
Homotsygootti häviämä	arvo = 0
Heterotsygootti häviämä	0,40 < arvo < 0,65
Heterotsygootti kahdentuma	1,30 < arvo < 1,65
Heterotsygootti kolminkertaistuma / Homotsygootti kahdentuma	1,75 < arvo < 2,15
Epäselvä kopiolumu	Kaikki muut arvot

2. Aineisto ja menetelmät

Pro gradu työssäni noudatin MRC Holland digitalMLPA™ yleistä työohjetta reaktioiden ja DNA-kirjastojen valmistuksessa, TYKS Laboratoriotuotteen, Genomiikan Illumina MiniSeq System -Laitteohjetta massasekvensoinnissa sekä Coffalyzer digitalMLPA™ käyttöohjetta (versioille 211020.1245 ja 220221.1522) sekvensointidatan käsittelyssä koetinarvojen laskemiseksi. Saadun datan taulukoinnissa käytin Microsoft Office 365 Excel -ohjelmaa ja taulukoiden sekä kuvien muokkauksessa Paint.net -ohjelmaa. Saatujen tulosten pohjalta suoritin χ^2 -testin selvitykseksi poikkeavatko saadut tulokset potilastiedoista, jonka nollahypoteesi oli, että tulokset vastaavat potilastietoja. Testi tehtiin Microsoft Office 365 Excel -ohjelmalla.

2.1 DNA-näytteet

Työn aineistona käytin TYKS:n Genomiikan vastuualueen arkiston 50 potilas DNA-näytettä, joista 12 omasi työssä käytettävän MRC Holland digitalMLPA perinnöllisellä syöpäpaneeli 1 (engl. hereditary cancer panel 1) havaittavissa olevia kopiolumuutoksia. Näytteiden kopiolumuutokset oli selvitetty ennakkoon ennen pro gradutyön aloittamista TYKS:n Genomiikan vastuualueella jo käytössä oleville menetelmillä, kuten MLPA:lla ja massasekvensointianalyysillä. Kaikkien näytteiden DNA oli eristetty kokoverinäytteistä käyttämällä QIASymphony DSP DNA Midi Kit ja QIASymphony SP-laitetta. Kopiolumuutokset, joita 12 näytteessä oli, olivat geneeissä: *MSH2*, *EPCAM*, *ATM*, *RAD51C*, *BRCA1*, *SMAD4* ja *MLH1* (Taulukko 2.). Loput 38 näytettä olivat normaaleja syöpäpaneelilla tunnistettavien geenialueiden eksonien kopiolumuutosten osalta. Normaalien näytteiden joukosta valitsin satunnaisesti 20 näytettä, joita käytin eri reaktiosarjojen valmistuksessa (Taulukko 2.). Kaikki työssä käyttämäni näytteet olivat nimikoodattuja. Niitä ei pystytty yhdistämään potilaisiin. Säilytin näytteet -20 °C työn ajan.

Taulukko 2: Pro gradutyössä käytetyt kaikki potilasnäytteet. Taulukon alussa on listattu näytteet, jotka omasivat patogeenisinä kopiolumuutoksia. Näitä näytteitä kutsutaan tekstissä syöpänäytteiksi. LERA-näytteestä oli oletuksena mahdollinen pseudogeeni kopiolumuutoksen takana. Työssä käytettyjen normaaleiden potilasnäytteiden kohtiin “Geeni(t)” ja “Eksoni(t)” (Kopiolumuutos) on merkitty “-“.

Näyte (Sukupuoli)	Geeni(t)	Eksoni(t) (Kopiolumuutos)
HAAN (N)	<i>EPCAM</i>	7-9 (Heterotsygootti kahdentuma)
	<i>MSH2</i>	1-16 (Heterotsygootti kahdentuma)
HEJA (N)	<i>ATM</i>	62 & 63 (Heterotsygootti kahdentuma)
INMA (N)	<i>RAD51C</i>	1-7 (Heterotsygootti kahdentuma)
KOEI (N)	<i>ATM</i>	62 & 63 (Heterotsygootti kahdentuma)
KUJA (N)	<i>BRC1</i>	13 (Heterotsygootti kahdentuma)
KUMI (N)	<i>BRC1</i>	13 (Heterotsygootti kahdentuma)
LERA (N)	<i>SMAD4</i>	1-12 (Heterotsygootti kahdentuma, pseudogeeni)
LUEL (N)	<i>BRC1</i>	13 (Heterotsygootti kahdentuma)
MAAR (M)	<i>RAD51C</i>	1-7 (Heterotsygootti kahdentuma)
PEGA (N)	<i>MLH1</i>	16 (Heterotsygootti häviämä)
TOLI (N)	<i>MSH2</i>	9-11 (Heterotsygootti häviämä)
ZYVA (N)	<i>MLH1</i>	16 (Heterotsygootti häviämä)
AHKA (N)	-	-
ALPI (N)	-	-
ANLI (N)	-	-
ANTE (M)	-	-
ASRE (N)	-	-
FRJA (N)	-	-
HATI (N)	-	-
HESA (N)	-	-
HIPI (N)	-	-
JADA (M)	-	-
KARA (M)	-	-
KAAN (N)	-	-
KASI (N)	-	-
KUPE (M)	-	-
LAAN (N)	-	-
LASU (N)	-	-
LEEI (N)	-	-
NITI (M)	-	-
OJRI (N)	-	-
PRLA (N)	-	-

2.2 DNA-näytteiden esivalmistelu ennen digitalMLPA reaktioiden valmistamista

Ennen digitalMLPA-reaktioiden valmistusta, laimensin reaktioissa käytettävät DNA-näytteet 0,1 X TE-puskuriin. DigitalMLPA™ yleisen työohjeen ohjeistama optimaalinen konsentraatio DNA:ta reaktioiden valmistuksessa oli 40–100 ng, kun yksittäiseen reaktioon mitataan 4 µl DNA-näytettä. Optimia mukaillen, ensimmäistä reaktiosarjaa varten mitoitin DNA-näytteiden konsentraatioiksi optimirajan ylärajan noin 15–20 ng/µl. DNA:n määrä reaktiossa oli tällöin noin 60–100 ng per reaktio. Käytin toisessa reaktiosarjassa samoja näytteitä samoilla konsentraatioilla. Kolmatta ja neljättä reaktiosarjaa varten laimensin DNA-näytteistä noin 10 ng/µl liuokset, jolloin näytteisiin tuli noin 40 ng DNA:ta. Säilytin valmiit DNA-laimennokset -20 °C.

Tarkistin, että mitatut konsentraatiot vastasivat tähdättyjä pitoisuuksia Qubit® 2.0 fluorometrillä Qubit™ dsDNA HS määrityssarjalla Life Technologies™ (2011) työohjetta noudattaen. Valmistin näytteistä kahdet rinnakkaiset näytteet, joiden pitoisuuksista laskin konsentraatioiden keskiarvot. Suoritin mittaukset näytekohtaisesti niin että yksi mittauskerta sisälsi kahdet rinnakkaiset näytteet ja Qubit™ dsDNA HS määrityssarjan kontrollit. Jokaista mittausta varten valmistin Qubit® työliuosta 2 ml Eppendorf -putkeen, joka koostui 796 µl Qubit® dsDNA HS puskurista sekä 4 µl Qubit® merkkiaineesta Qubit™ dsDNA HS määrityssarjasta.

Liuotin mittauksissa 10 µl molempia kontrolleja 190 µl työliuosta ja 1 µl DNA-näytettä 199 µl työliuosta Axygen® PCR-05-C putkiin, joilla suoritin mittaukset. Fluorimetrin asetuksiksi määritin dsDNA HS ja käytetyksi näytetilavuudeksi 1 µl.

2.3 DigitalMLPA-reaktiot

Reaktioiden valmistuksessa käytin MRC Holland SALSA® digitalMLPA perinnöllisen syöpäpaneeli 1 koetinseoksia D001-B1 ja D001-C1 yhdistettynä digitalMLPA reagenssisarjan DRK01-IL ja SALSA viivakoodiliuoslevyn BP01 kanssa. Ensimmäisen ja toisen reaktiosarjan valmistuksessa käytin koetinseosta D001-B1, kun taas kolmannen ja neljännen reaktiosarjan valmistuksessa D001-C1. Reaktioiden valmistuksessa noudatin digitalMLPA™ yleisen työohjeen sivuja 8–9 alkaen kohdasta 7.1 ”Thermocycler program for digitalMLPA reactions” päättyen kohtaan 7.5 ”PCR reaction (Day 2)”.

Reaktioiden valmistuksessa käytin Applied Biosystems™ Veriti Dx 96-kaivoista PCR-laitetta.

Valmistin reaktioita yhteensä 46, joiden valmistuksen jaoin neljälle eri reaktiosarjalle (Taulukko 3.). Kahdessa ensimmäisessä reaktiosarjassa valmistin kahdeksan reaktiota koostuen neljästä syöpänäytteestä, kolmesta normaalista vertaisnäytteestä ja negatiivisesta kontrollinäytteestä. Negatiivisena kontrollina kaikissa reaktiosarjoissa käytin 4 µl 0,1 X TE-puskuria. Kolmannessa sarjassa reaktiot koostuivat 16 normaalinäytteestä, joista ALPI-näyte toimi vertaisnäytteenä kahdella ensimmäisellä kerralla ja negatiivisesta kontrollinäytteestä. Neljässarja sisälsi loput seitsemän syöpänäytettä sekä neljä normaalinäytettä, joista näytteet AHKA ja ANLI toimivat vertaisnäytteinä reaktiosarjoissa 1 ja 2.

Jaoin jokaisen reaktiosarjan valmistuksen kahdelle päivälle. Aloitin reaktioiden valmistuksen mittaamalla 4 µl DNA-näyteitä ja 2 µl uniikkia viivakoodiseosta SALSA viivakoodiliuoslevyltä BP01 0,2 ml PCR-putkiin kahdeksan kärkisellä haravapipetillä (Kuva 3.). Ennen viivakoodiliuoslevyn mittausta sentrifuugasin viivakoodiliuoslevyn lyhyesti 1000 rpm, jota seurasi sen lyhyt sekoitus kuoppalevysekoittimessa ja uudelleen lyhyt sentrifugointi 1000 rpm.

Sentrifuugasin viivakoodiliuoslevyn Eppendorf Centrifuugi 5804, jonka käyttöä varten valmistin omat vastapainot eri reaktiosarjojen valmistuksen ohessa. Valmistin vastapainot mittaamalla mQ H₂O 20 µl 96 kaivoiseen kuoppalevyyn vastaamaan sen hetkisen viivakoodilevyn kaivojen tilavuuksia. Ensimmäisen reaktiosarjan valmistuksessa mQ H₂O mittasin 96 x 20 µl ja seuraavissa vähensin edellisiä reaktiosarjoja varten mitatut tilavuudet uusissa vastapainoissa. Peitin kuoppalevyn lopuksi 4titude® läpinäkyvällä PCR-tarrakalvolla.

BP01	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

Kuva 3. SALSA Viivakoodiliuoslevy BP01 koostui 96 kaivosta, joista jokainen sisälsi 20 µl värjättyä viivakoodiseosta (MRC Holland, 2022d) Jokainen viivakoodisekvenssi on kaivokohtainen. Väri on tarkoitettu helpottamaan tunnistusta, mitä riviä on käytetty reaktioiden valmistuksessa. Reaktiosarjojen valmistuksessa ensimmäisessä käytettiin riviä 2, toisessa riviä 3, kolmannessa rivejä 4, 5 ja riviltä 6 kaivoa A6. Neljännessä reaktiosarjassa käytettiin riviä 9 ja riviltä 10 kaivoja A10-E10. Kuva pohjautuu MRC Holland (2022d) digitalMLPA™ general protocol for use with Illumina sequencing platforms.

Näytteiden ja viivakoodiliuosten pipetointia seurasi reaktioiden 10 minuutin denaturaatio 98 °C, jonka aikana valmistin hybridisaatioreaktioseoksen (Kuva 4. & Taulukko 4.). Kun näytteet olivat 25 °C, pipetoin reaktioihin huoneenlämmössä, niiden ollessa ulkona PCR-laitteesta 3 µl hybridisaatioreaktioseosta. Tämän jälkeen spinnasin näytteet lyhyesti minisentriifugissa ennen kuin siirsin ne takaisin PCR-laitteeseen, josta käynnistin heti hybridisaatioreaktio-ohjelman (Kuva 4.). Reaktion aikana koettimien osat liittyivät niiden kohdesekvensseihin DNA:ssa. Anoin hybridisaatioreaktion kestää noin 18 tuntia jokaisella reaktiosarjan valmistuskerralla, jolloin reaktioiden valmistukset jakautuivat kahdelle päivälle.

Reaktioiden valmistuksen seuraavana päivänä valmistin ligaatioreaktioseoksen juuri ennen kuin jatkoin hybridisaatioreaktio-ohjelman kolmanteen vaiheeseen (Kuva 4. ja Taulukko 4.). Pipetoin ligaatioreaktioseosta 32 µl per reaktio niiden ollessa PCR-laitteessa 48 °C. Pipetoinnin ajaksi PCR-laite oli asetettu kestämään äärettömän ajan (∞) 48 °C. Kahden ensimmäisen reaktiosarjan valmistuksessa siirsin näytteet tämän jälkeen nopeasti toiseen samanlaiseen PCR-laitteeseen, johon olin valmiiksi määritellyt ligaatioreaktio-ohjelman alkamaan sen ensimmäisestä vaiheesta (Kuva 4.). Eli toinen PCR-laite oli valmiiksi 48 °C kun aloitin ligaatioreaktioseoksen pipetoinnin. Tämä johtui käyttämästäni PCR-laitteesta, jossa luodulle ohjelmalle täytyi määrittellä yksittäisen reaktion tilavuus. Kahden laitteen avulla halusin varmistaa, että näytteiden lämpötila

pysyisi mahdollisimman lähellä 48 °C PCR-ohjelman vaihdon aikana. Kiireen välttämiseksi asetin toisessa PCR-laitteessa ligaatioreaktio-ohjelman ensimmäisen vaiheen kestoksi ∞. Reaktiosarjat 3 ja 4 valmistuksessa käytin ainoastaan yhtä PCR-laitetta. Tällöin näytteiden lämpötila laski hetkellisesti 25 °C ohjelman vaihdon aikana. Aloitin ligaatioreaktio-ohjelma suoraan näytteiden lämmityksellä 48 °C 30 minuutin ajan. Kuten reaktiosarjojen 1 ja 2 valmistuksessa, pipetoin ligaatioreaktioseosta 32 µl näytteisiin, kun ne olivat PCR-laitteessa 48 °C äärettömäksi määritetyssä ajassa.

Muutamaa minuuttia ennen ligaatioreaktion loppumista valmistin polymeraasireaktioseoksen. Kun näytteet olivat 20 °C, poistin ne PCR-laitteesta ja pipetoin 10 µl polymeraasireaktioseosta jokaiseen näytteeseen huoneenlämmössä (Taulukko 4.). Pipetoinnin myötä jatkoin PCR-reaktioon, joka kesti noin 2,5 tuntia (Kuva 4.). Säilytin valmiita PCR-tuotteita 4 °C ennen DNA-kirjastojen valmistusta, jonka jälkeen säilöin tuotteet -20 °C.

Taulukko 3: Reaktiosarjoissa käytetyt näytteet ja niistä valmistettujen käyttöliuosten DNA-pitoisuudet. Reaktiosarjat poikkeavat muuten toisistaan, mutta toinen reaktiosarja koostui samoista näytteistä kuin ensimmäinen reaktiosarja.

Reaktiosarja	Näytteet (DNA-pitoisuus, ng/μl)
1	1. HAAN (13,4)
	2. HEJA (20,45)
	3. INMA (19,05)
	4. KOEI (17)
	5. AHKA (23,4)
	6. ALPI (15,1)
	7. ANLI (15,9)
	8. Negatiivinen kontrolli (-)
2	Samat kuin reaktiosarja 1
3	1. ALPI (9,1)
	2. ANTE (11,35)
	3. ASRE (9,92)
	4. FRJA (9)
	5. HATI (10,53)
	6. HESA (9,53)
	7. HIPI (9,7)
	8. JADA (10,07)
	9. KARA (11,5)
	10. KAAAN (9,49)
	11. KASI (9,4)
	12. LASU (10,3)
	13. LAAN (10,23)
	14. LEEI (9,13)
	15. NITI (9,1)
	16. PRLA (11,61)
	17. Negatiivinen kontrolli (-)
4	1. KUJA (9,78)
	2. KUMI (9,27)
	3. LERA (10,3)
	4. LUEL (9,03)
	5. PEGA (9,69)
	6. TOLI (10,04)
	7. ZYVA (9,83)
	8. AHKA (9,1)
	9. ANLI (9,39)
	10. KUPE (9)
	11. MAAR (10,45)
	12. OJRI (9,85)
	13. Negatiivinen kontrolli (-)

Taulukko 4: Reaktioiden valmistuksessa käytetyt reaktioseokset ja niiden valmistukseen vaaditut liuokset. Vaadittuja liuoksia mitattiin yhden reaktion verran enemmän mitä valmistettiin varmuudeksi, että kaikkiin reaktioihin tulee tarvittava määrä reagensseja. Reaktioseokset valmistettiin 1,5 ml Eppendorf-putkiin, joista pipetoitiin tarvittava määrä seosta reaktioihin.

Reaktioseos	Vaaditut liuokset ja tilavuudet yhtä reaktiota kohden
Hybridisaatio	1,5 µl digitalMLPA koetinseos + 1,5 µl MLPA-puskuri
Ligaatio	25 µl Nukleaasivapaa vesi + 3 µl Ligaasipuskuri A + 1 µl Ligaasi-65
Polymeraasi	7 µl Nukleaasivapaa vesi + 2 µl PCR alukesekoitus P5P7 + 1 µl Polymeraasi

DNA:n denaturaatio (6 µl)	
1. 98 °C	10 min
2. 25 °C	∞
Hybridisaatioreaktio (9 µl)	
1. 95 °C	1 min
2. 60 °C	16–20 h
3. 48 °C	∞
Ligaatioreaktio (41 µl)	
1. 48 °C	∞
2. 48 °C	30 min
3. 98 °C	5 min
4. 65 °C	20 min
5. 20 °C	∞
PCR-reaktio (51 µl)	
1. 65 °C	1 min
2. 95 °C	30 s
65 °C	40 s
72 °C	90 s
3. 15 °C	∞

} 45 sykliä

Kuva 4. PCR-laitteessa käytetyt ohjelmat mukaan lukien näytteiden tilavuudet eri ohjelmien vaiheissa suluissa ohjelman nimen vieressä. Ligaatioreaktio-ohjelman vaihe 1. jätettiin pois ohjelmasta reaktiosarjojen 3 ja 4 valmistuksessa, kun käytössä oli vain yksi PCR-laite.

2.4 DNA-kirjastojen valmistus

Valmiiden PCR-tuotteiden pohjalta valmistin DNA-kirjastot, joihin yhdistin 5 µl jokaista sekvensoitavaa PCR-reaktiotuotteita keskenään 1,5 ml LoBind Eppendorf -putkiin. Reaktiosarjoissa 1, 2 ja 3 kirjastot koostuivat sarjoissa valmistetuista näytteistä (Taulukko 2.). Reaktiosarja 4 kirjastoon lisäsin kyseisessä sarjassa valmistettujen näytteiden lisäksi näytteet: ANTE, HESA, HIPI, JADA, KARA, KAAAN, NITI ja PRLA kolmannesta reaktiosarjasta. Näytteet edellisestä sarjasta toimivat vertaisnäytteinä datan normalisointia varten sekvensoinnin jälkeisessä datan analysoinnissa.

En suorittanut DNA-kirjastojen kvantifiointia tai laadun varmistusta ennen sekvensointia digitalMLPA™ yleisten ohjeiden mukaisesti. Valmistetut kirjastot säilytettiin 4 °C ennen niiden käsittelyä sekvensointia varten. Reaktiosarjojen 1 ja 2 pohjautuvat DNA-kirjastot sekvensoin samana päivänä kuin kirjastot oli valmistettu. Reaktiosarjan 3 DNA-kirjaston sekvensoin noin kolmen päivän päästä ja reaktiosarjan 4 kirjaston seuraavana päivänä kirjaston valmistuksesta.

2.5 Massasekvensointi (NGS)

Suoritin DNA-kirjastojen sekvensoinnin erikseen kaikilla reaktiosarjoilla Illumina MiniSeq -järjestelmällä. MiniSeq:n käytössä noudatin TYKS Laboratoriotuotteen, Genomiikan Illumina MiniSeq System -Laitteohjetta pois lukien ajopohjan laatimisvaiheen. Ajopohjan tilalle MiniSeq:n Illumina Local Run Manager -ohjelmaan listasin ajopohjan sijaan Coffalyzer digitalMLPA -ohjelman käyttöä varten vaaditun selitteen suoritettavalle kokeelle. Valitsin sekvensointiasetukset ja käytettävän Illumina sarjan digitalMLPA™ yleisen ohjeen sivujen 10–11 kohdan 8.2 ”Flow cell used & instrument settings” mukaan MiniSeq -järjestelmän käyttöä varten.

Sekvensoinnissa käytetyt Illumina #FC-420-1002 MiniSeq High Output sarjan reagenssikasetit (150 sykliä) sulatettiin noin 37 °C vesihautteessa yhden tunnin ajan ja kittiin kuuluva virtauskyvetti 30 minuutin ajan huoneenlämmössä. Erikoistuva sairaalaselubiologi Jenni Tuominen TYKS Genomiikalta hoiti sulatukset reaktiosarjojen 1, 2 ja 3 osalta ennen niiden lataamista MiniSeq-järjestelmään. Neljännen reaktiosarjan kohdalla suoritin itse reagenssikasetin ja virtauskyvetin sulattamisen. Sulattamisen aikana suoritin kirjastojen laimennuksen ja denaturaation sekvensointia varten digitalMLPA™

yleisen ohjeen sivun 9 kohdan 7.6.3 ”MiniSeq instrument (Day 2)” mukaisesti. Kirjaston denaturaatiota varten valmistin aina uuden 0,1 M NaOH-liuoksen pipetoimalla 900 µl steriiliä vettä ja 100 µl valmista 1 M NaOH-liuosta.

Käytetyssä MiniSeq:n Illumina Local Run Manager-ohjelmassa ei ollut mahdollista valita DNA-kirjaston valmistelu (engl. library prep kit) asetukseksi TruSeq LT, joka oli ohjeistettuna digitalMLPA™ yleisessä ohjeessa. Tämän vuoksi valitsin asetukseksi <Custom> yhdistettynä ”Index read” 0, ”single read” ja ”read length” 110 nt asetusten kanssa.

2.6 Coffalyzer digitalMLPA™ (ver. 220221.1522)

Sekvensoinnin jälkeisessä FASTQ-tiedostojen analysoinnissa käytin MRC Hollandin Coffalyzer digitalMLPA™-ohjelmaa. Ohjelman käytössä noudatin Coffalyzer digitalMLPA™ käyttöohjeen (versioille 211020.1245 ja 220221.1522) kohtaa 3.2.1 ”Create a New Experiment”. Jokaiselle sekvensoinnista saadulle FASTQ-tiedostolle loin oma työpohjan ohjelmassa. Uuteen työpohjaan aluksi määrittelin työssä käytetyn viivakoodiliuoslevyn, jonka pohjalta valikoitiin reaktiosarjojen valmistuksessa käytetyt levyn kaivot. Kaivojen valitsemisen jälkeen määrittelin ohjelmalle reaktiosarjoissa käytetyt koetinseokset ja näytteet manuaalisesti. Näytteiden määrittelyssä numeroin ja nimesin jokaisen näytteen vastaamaan niiden selitettä reaktiosarjojen valmistuksessa (Taulukko 2.). Lisäksi määritin näytteiden sukupuolen sekä tyyppin tässä vaiheessa (Liite 1.).

Mahdolliset vaihtoehdot näytetyypeiksi olivat: testi (t), verrokki (r), ei DNA:ta (n), positiivinen (p), SD (s) ja määrittelemätön (u). Testi-näytteiksi määrittelin tutkittavat näytteet, joiden koetinarvot suhteutettiin määrittelemieni verrokinäytteiden ohjelman laskemiin koetinarvoihin. Ei DNA:ta-näyte oli työssä käytetty negatiivinen kontrolli. SD-näytteeksi merkittäisiin näyte, joka on spesifi MRC Hollandilta peräisin oleva tietyn genotyypin omaava näyte. Tällaisia näytteitä työssä ei käytetty. Näyte, jonka alkuperää ei tunneta tai, joka halutaan sulkea pois analyysistä, asetetaan määrittelemättömäksi näytteeksi.

Coffalyzer digitalMLPA™ -ohjelmasta saatujen tulosten tulokinnan apuna käytin Coffalyzer digitalMLPA™ käyttöohjeen kohtia 4. ”Coffalyzer digitalMLPA Reports” ja 5. ”Analysis of Results”.

2.7 Koetinarvojen laskenta

Coffalyzer digitalMLPA™-ohjelma laskee reaktioiden valmistuksessa käytettyjen koettimien väliset arvot perustuen kahteen eri vaiheeseen, johon vaikuttavat onko verrokinäytteitä käytetty. Kun verrokinäytteet ovat sisällytetty testiin, kuten reaktiosarjojen 1, 2 ja 4 valmistuksessa, ohjelma aluksi laskee näytekohdaiset koetin suhteet. Suhde lasketaan näyte DNA:han liittyneiden testikoettimien sekä koetinsarjaan kuuluvien verrokkikoettimien välillä. Koetinsarjoissa D001-B1 ja D001-C1 verrokkikoettimien sijaan on ennalta määrätty testikoettimet, jotka toimivat verrokkikoettimina. D001-B1 koetinsarjassa näitä koettimia on yhteensä 266 ja D001-C1 koetinsarjassa 290.

Näytekohdaisen koetinsuhteiden laskentaa seuraa jokaisen testinäytteen näytekohdaisen koetinsuhdeluvun vertaus jokaista verrokinäytteen koetinsuhdelukua kohden (Kaava 1.). Tämän pohjalta saadaan niin monta mediaaniarvoa, jokaiselle testikoettimelle kuin verrokinäytteitä on määriteltä ohjelmalle. Lopuksi ohjelma laskee koettimille keskiarvot perustuen saatuihin mediaaniarvoihin.

Ohjelmalla pystytään laskemaan koetinarvot myös ilman verrokinäytteitä, kun erillisiä verrokinäytteitä ei ole käytetty, kuten kolmannen reaktiosarjan valmistuksessa. Rajoittavana tekijänä tässä on, että DNA-näytteitä pitää olla vähintään kahdeksan ja yksilöistä, jotka eivät ole läheistä sukua keskenään. Tällöin kaikki näytteet määritellään testinäytteiksi ja ohjelma suorittaa saman laskennan näytekohdaisen koetinsuhteiden osalta kuin verrokinäytteiden kanssa. Erona laskutoimituksessa on, että ohjelma laskee koettimien mediaanien absoluuttisen keskipoikkeaman (engl. median absolute deviation, MAD) vertaamalla kaikkien näytteiden väliset näytekohdaiset koetinsuhteet keskenään. MAD-arvoja saadaan niin monta kuin tutkittuja näytteitä on käytetty kaiken kaikkiaan. Niiden pohjalta jokaiselle koettimelle lasketaan keskiarvo.

Kaava 1. Coffalyzer digitalMLPA™ -ohjelman laskukaava koettimien mediaaniarvoille, jossa Tk = testikoetin ja Vk = verrokkikoetin. Kaava pohjautuu MRC Holland (2022e) Coffalyzer digitalMLPA™ user manual for Coffalyzer digitalMLPA versions 211020.1245 and 220221.1522.

$$\text{Mediaani} \left(\frac{(\text{Tk1 näyte 1} / \text{Vk1 näyte 1})}{(\text{Tk1 verrokinäyte 1} / \text{Rp1 verrokinäyte 1})}, \dots, \frac{(\text{Tk1 näyte 1} / \text{Vk}_n \text{näyte 1})}{(\text{Tk1 verrokinäyte 1} / \text{Vk}_n \text{verrokinäyte1})} \right)$$

3. Tulokset

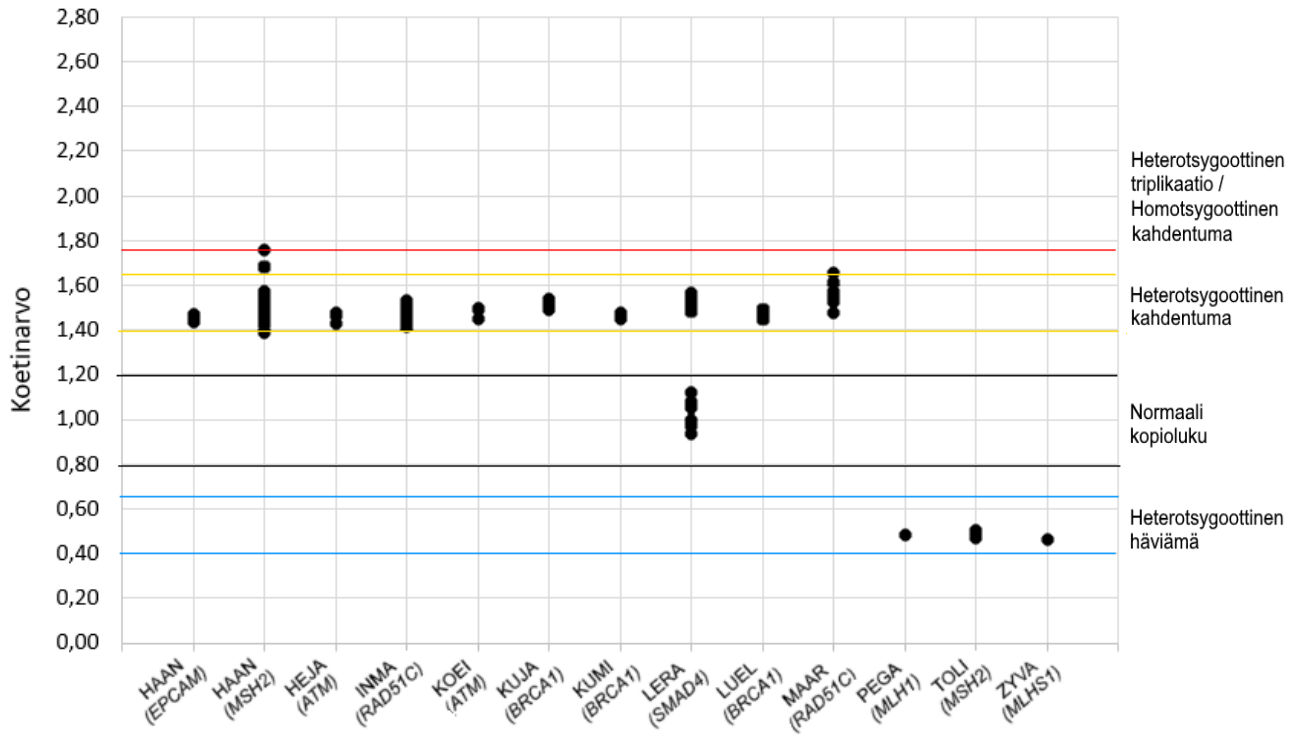
3.1 Syöpänäytteet

Kaikista työssä tutkittavista potilasnäytteistä, joissa oli eksonien kopiolumuutoksia, tunnistettiin toisen ja neljännen reaktiosarjan pohjalta digitalMLPA -menetelmällä kaikki tunnistettavana olevat eksonien kopiolumuutokset paitsi LERA-näytteen osalta. Kaikkien paitsi LERA-näytteen tutkimuksen kohteena olevien eksonien koetin arvot olivat menetelmän normaalia kopiolumuutosta kuvaavien viitearvojen ulkopuolella (Kuva 5A.). LERA-näytteessä *SMAD4*-geenialueen eksonien 2, 5, 6, 9, 10, 11 ja yksi eksonin 12 kolmesta testikoettimen arvoista, poikkesivat normaalin kopiolumun viitearvoista. Muut geenialueen eksonit olivat normaalin kopiolumun arvojen mukaiset. Näytteiden eksonien koetin arvot, jotka eivät olleet normaalin kopiolumun mukaiset, poikkesivat kaikki kahden keskihajonnan verran reaktiosarjojen vertaisnäytteiden vastaavien eksonien koetin arvoista.

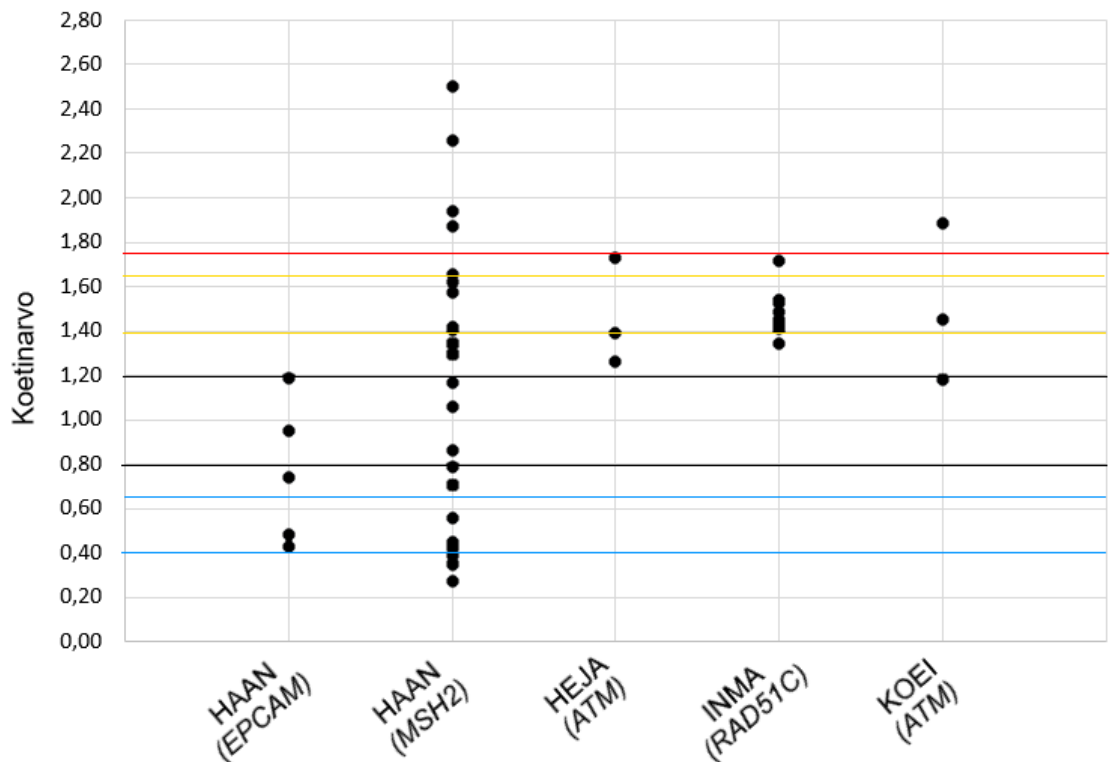
Kaikissa muissa näytteissä paitsi INMA, PEGA ja ZYVA ilmeni tutkittavana olevien eksonien lisäksi muiden geenialueiden eksoneita, joiden testikoettimien arvot olivat joko lievästi koholla tai lievästi normaalien viitearvojen alapuolella. Kyseisten koetin arvojen lukemat poikkesivat enintään yhden keskihajonnan verran näytteiden vertaisnäytteisiin nähden, jolloin ne eivät poikenneet merkitsevästi reaktiosarjojen vertaisnäytteistä.

Ensimmäisen reaktiosarjan osalta syöpänäytteiden HAAN, HEJA, INMA ja KOEI tutkittavien eksonien testikoettimien arvoissa esiintyi vaihtelua. Monet arvoista poikkesivat odotetuista kopiolumuutoksista (Kuva 5B.).

A



B



Kuva 5. (A) Kuvaaja 2 ja 4 reaktiosarjan syöpänäytteiden tutkittavien eksonien testikoettimien arvoista. Viivat rajaavat eri kopiolumuutostiloja. (B) Ensimmäisen reaktiosarjan syöpänäytteiden tutkittavien geenialueiden eksonien testikoettimien arvot. Kuvaajan rajatut alueet ovat samat kuin kuvassa 5A.

3.2 Normaalinäytteet

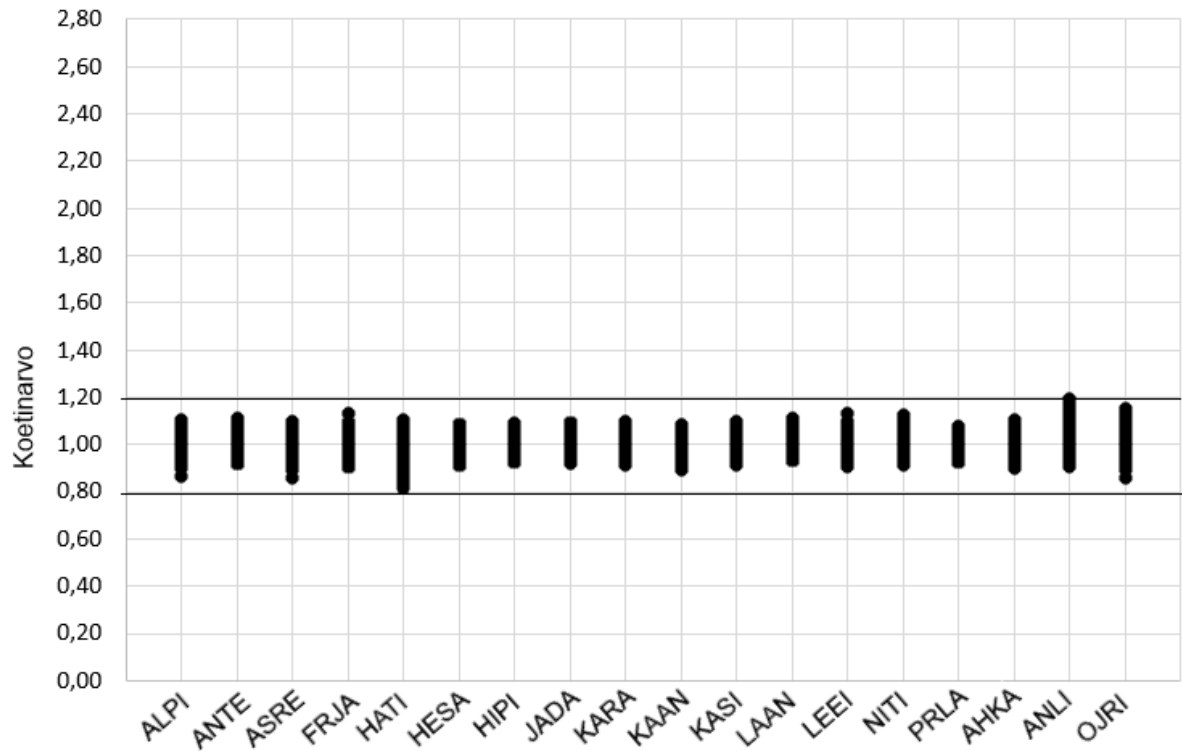
Normaalinäytteiden joukosta poissuljettiin eikä sisällytetty Coffalyzer-ohjelmalla tehtyyn analyysiin näytteitä LASU ja KUPE, johtuen eri syistä. LASU-näyte poissuljettiin potilastietosuojan nojalla jälkikäteen selvittyä, ettei henkilö ollut antanut lupaa käyttää hänen geeninäytettään anonyymissä tutkimustyössä. KUPE-näyte poissuljettiin, kun ilmeni, ettei potilasnäyte ollut peräisin terveistä ituradansoluista. Lopuista 18 reaktiosarjoissa 3 ja 4 käytetyistä näytteistä, jotka olivat määritelty normaaleiksi käytettyjen digitalMLPA-koetinsarjojen kohdegeenien eksonien osalta, ei havaittu poikkeavia kopiolumuutoksia perustuen testikoettimien arvoihin (Kuva 6A.). Kolmannen reaktiosarjan LEEI-näytteessä oli havainto *CHEK2*-geenin 1100delC mutaatiosta (rs555607708) pohjautuen mutaatiopesifiin koettiin.

Normaalinäyte ALPI:ssa, jota käytettiin verrokinäytteenä ensimmäisessä ja toisessa reaktiosarjassa oli toisen reaktiosarjan pohjalta saadun Coffalyzer-ohjelman tulosten mukaan lievästi koholla olevia koetinarvoja (Taulukko 5 & Kuva 6B.). Kun näyte testattiin uudestaan testinäytteeksi määriteltynä kolmannessa reaktiosarjassa, koettimien arvot olivat normaalin kopiolumun mukaiset. Ensimmäisessä reaktiosarjassa käytettyjen verrokinäytteiden koetinarvot vaihtelivat suuresti (Kuva 6B.).

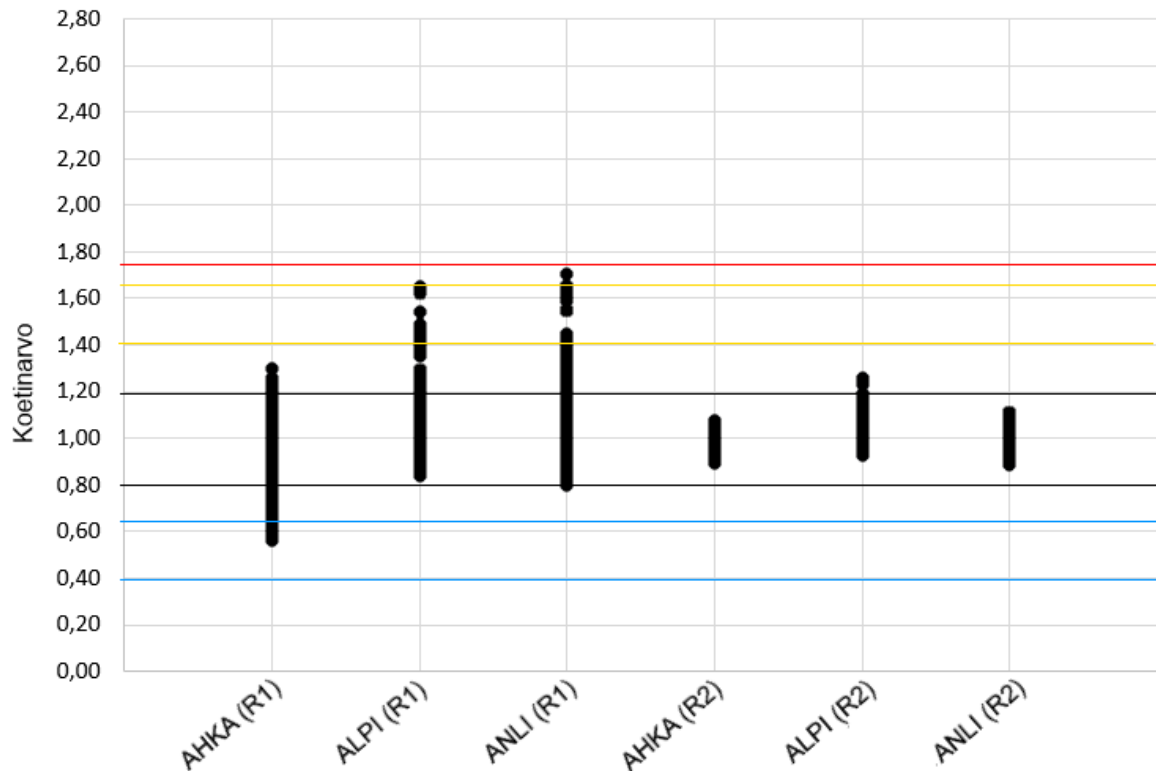
Taulukko 5: ALPI-näytteen geenialueiden eksonien koettimien arvot, jotka poikkesivat lievästi normaalin kopiolumun viitearvoista toisen reaktiosarjan pohjalta. Poikkeavat koetinarvot vaikuttivat kyseisten koetinarvojen keskihajontaan. Testattaessa näyte kolmannessa reaktiosarjassa samojen koettimien arvot olivat normaalin kopiolumun mukaiset.

Geenialue	Eksoni (Koetin ,tyyppi)	Testikoettimen arvo Reaktiosarja 2	Testikoettimen arvo Reaktiosarja 3
<i>APC</i>	18 (S017561, Testi)	1,23	1
<i>BMPRIA</i>	12 (S017494, Verrokki)	1,25	0,9
<i>SMAD4</i>	Promoottori alue A eksoni 1 yläjuoksussa (S011008, Verrokki)	1,26	1,02
	12 (S010332, Testi)	0,12	0,99

A



B



Kuva 6. (A) Reaktiosarjojen 3 ja 4 näytteiden kaikkien eksonien kopiolukua mittaavien koetin arvot. (B) Reaktiosarjojen 1 ja 2 verrokinäytteiden kaikkien eksonien kopiolukua mittaavien koetin arvot, jossa R1 = Reaktiosarja 1 ja R2 = Reaktiosarja 2. Kuvaajiin rajatut alueet ovat samat kuin kuvassa 5A määritellyt alueet.

3.3 χ^2 -testi

Onnistuneiden reaktiosarjojen 2–4 näytteiden pohjalta suoritettiin χ^2 -testi, jonka pohjalta päätettiin, voidaanko nollahypoteesi säilyttää vai hylätä. Testissä huomioitiin, että näytteet AHKA, ALPI ja ANLI testattiin kahdesti ja, että näytteet KUPE ja LASU poissuljettiin tuloksista. Lopullinen näyteotanta oli siis 12 positiivista ja 18 negatiivista näytettä. Tulokseksi testistä saatiin $\chi^2(1) = 0$, $p = 1$.

3.4 Datan normalisointi

Luotettavaa datan normalisointia ja tätä kautta luotettavien testikoettimien arvojen saantia varten reaktiosarjojen vertaisnäytteiden testikoettimien arvojen keskihajonta täytyi olla $\leq 0,10$. Toisesta ja neljännessä reaktiosarjasta neljännessä tämä vaatimus täyttyi vertaisnäytteiden osalta, mutta toisessa ei. Toisessa reaktiosarjassa niiden testikoettimien arvojen keskihajonta oli yli 0,10, joissa esiintyi lievästi kohonneita testikoettimien arvoja ALPI-näytteessä (Taulukko 6.). Lisäksi keskihajonnat olivat lievästi koholla kolmella eri geenialueella yksittäisillä testikoettimilla. Näiden eksonien osalta, niiden testikoettimien arvot vaihtelivat normaalin kopioluvun viitearvojen ylärajassa 1,18 ja 1,19 välillä ALPI-näytteessä (Taulukko 6.).

Kolmannessa reaktiosarjassa, jossa ei käytetty määritettyjä vertaisnäytteitä, vaan datan normalisointi suoritettiin kaikkien näytteiden koetinarvojen pohjalta, kaikkien testi- ja vertaiskoettimien mediaanin absoluuttiseksi poikkeamaksi (MAD) saatiin alle 0,10. MAD-arvot vaihtelivat 0,01 ja 0,07 välillä.

Ensimmäisen reaktiosarjan verrokinäytteiden testikoettimien keskihajonnassa ilmeni paljon vaihtelua, poiketen odotetusta $\leq 0,10$. Näihin lukeutui testikoettimia, jotka kohdistuivat reaktiosarjan syöpänäytteiden tutkittaviin eksoneihin (Taulukko 7.).

Taulukko 6: Toisen reaktiosarjan ALPI-näytteen eri geenialueiden eksonit joiden testikoettimien keskihajonta poikkesi odotetusta $\leq 0,10$. Taulukossa σ = keskihajonta

Geenialue	Eksoni (Koetin ,tyyppi)	Koettimen σ	Koettimen arvo
<i>APC</i>	18 (S017561, Testi)	0,15	1,23
<i>BMPRIA</i>	12 (S017494, Verrokki)	0,16	1,25
<i>PALB2</i>	8 (S010380, Verrokki)	0,12	1,18
<i>BRCA1</i>	18 (S017434, Verrokki)	0,13	1,19
<i>SMAD4</i>	Promoottori alue A eksoni 1 yläjuoksussa (S011008, Verrokki)	0,17	1,26
	12 (S010332, Testi)	0,12	1,18

Taulukko 7: Ensimmäisen reaktiosarjan verrokinäytteiden geenialueiden eksonien koettimien keskihajonnat, jotka poikkesivat odotetusta $\leq 0,10$, ja olivat tutkinnan kohteena syöpänäytteiden osalta. Taulukossa σ = keskihajonta.

Geeni	Eksoni/alue (Koetin)	Koettimien σ
<i>MSH2</i>	Geenialue noin 4,3 kb ennen eksoni 1 (S017526)	0,33
	1 (S010423)	0,17
	2 (S010424)	0,17
	8 (S010432)	0,48
	11 (S010435)	0,16
	13 (S017533)	0,13
	16 (S017534)	0,12
	16 (S010441)	0,5
	Eksonin 16 noin 3,1 kb jälkeinen alue (S010442)	0,14
<i>EPCAM</i>	7 (S018400)	0,3
	Eksoni 9 lopetuskodonin noin 50 nt päässä oleva 3' UTR (S010417)	0,15
	Eksoni 9 lopetuskodonin noin 0,2 kb päässä oleva 3' UTR (S010417)	0,15
<i>ATM</i>	63 (S010300)	0,12
<i>RAD51C</i>	6 (S017464)	0,21
	8 (S010232)	0,17

3.5 Massasekvensointi (NGS)

Illumina MiniSeq -järjestelmällä suoritettujen sekvensointiajojen klusteritiheydet vaihtelivat hieman ajojen välillä, mutta Q30-arvot pysyivät suhteellisen samoina (Taulukko 8.). Q30-arvot vaihtelivat 94,92 ja 95,40 välillä reaktiosarjoissa 2, 3 ja 4, pudoten huomattavasti ensimmäisessä reaktiosarjassa lukemaan 84. Klusteritiheyksistä korkein saatiin ensimmäisestä reaktiosarjasta, jossa se oli 210 K/mm² ja matalin toisesta reaktiosarjasta, jossa klusteritiheys oli 108 K/mm².

Sekvensoiduista digitalMLPA -kirjastoihin pohjautuvista FASTQ-tiedostoista tunnistettiin noin 700–710 uniikkia sekvenssiä per näyte (Kuvat 7A-7C.). Kaikkien tunnistettujen sekä tunnistamattomien sekvenssien määrät vaihtelivat reaktiosarjojen välillä, mutta mukailivat toisiaan reaktiosarjojen sisällä. Huomattavasti eniten sekä tunnistettuja että tunnistamattomia sekvenssejä saatiin ensimmäisestä reaktiosarjasta.

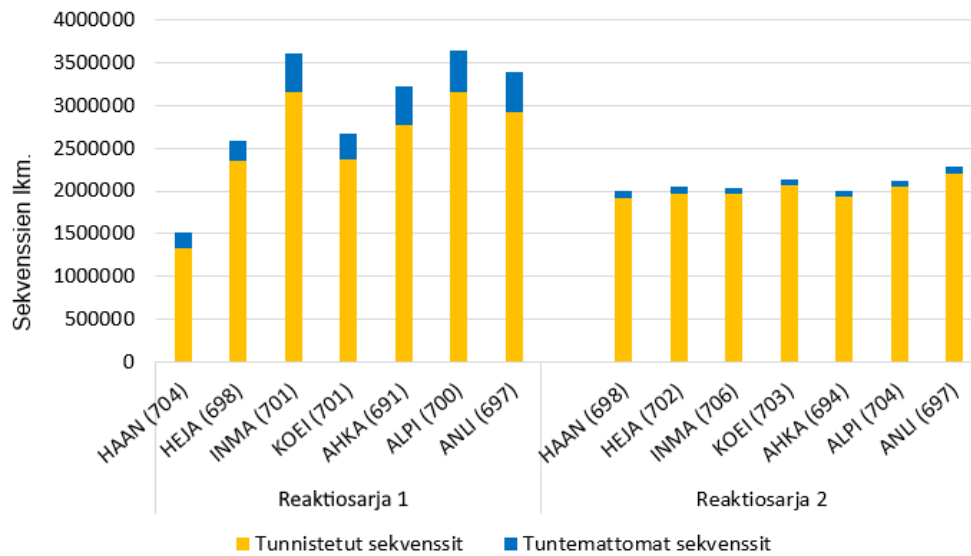
Vertaiskoettimien mediaaniarvojen lukumäärät joihin digitalMLPA™ Coffalyzer-ohjelman laskemat koetin arvot perustuvat vaihtelivat myös reaktiosarjojen välillä (Kuvat 8A-8C). Arvot mukailivat vahvasti tunnistettujen sekvenssien lukumäärää (Kuvat 7A-7C.). Kaikkien reaktiosarjojen paitsi ensimmäisen negatiivisen kontrollinäytteen pohjalta saatiin kaksi uniikkia mediaanilukemaa. Ensimmäisestä reaktiosarjasta lukemia saatiin viisi.

Kaikista näytteistä paitsi ensimmäisen reaktiosarjan HAAN-näytteestä tunnistettiin kaikki testi- ja vertaiskoettimet. Näytteestä tunnistettiin 557 testikoetinta oletetusta 561 koettimesta ja 265 vertaiskoetinta oletetusta 266.

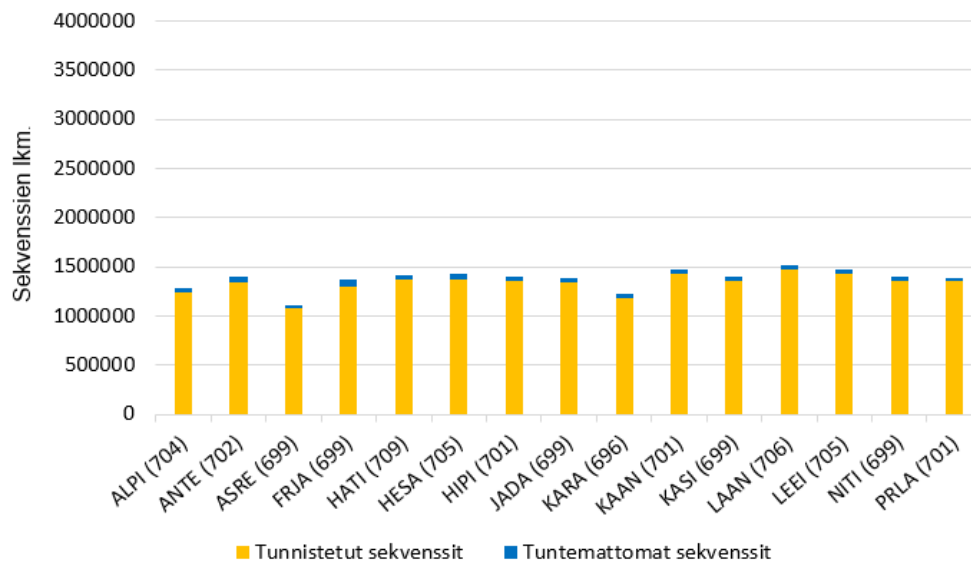
Taulukko 8: Reaktiosarjojen pohjalta valmistettujen DNA-kirjastojen sekvensoinnista saadut laatuparametrien arvot klusteritiheyden ja Q30-arvojen osalta. Q30-arvo kuvaa niiden sekvenssien osuutta, joissa noin tuhannen emäksen välein yksi emäs oli tunnistettu virheellisesti.

Reaktiosarja	Klusteritiheys (K/mm ²)	Q30 (%)
1	210	84
2	108	95
3	149	94,92
4	154	95,4

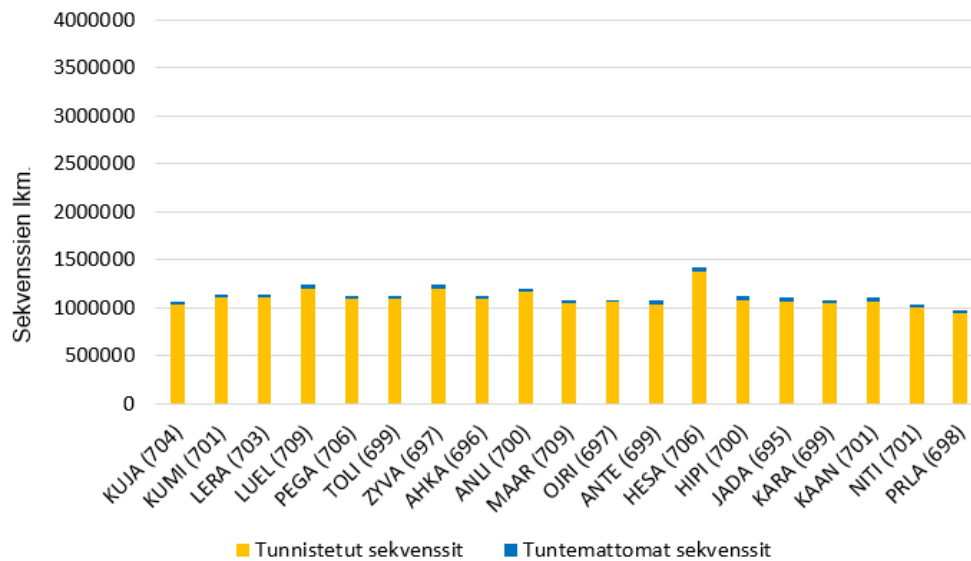
A



B

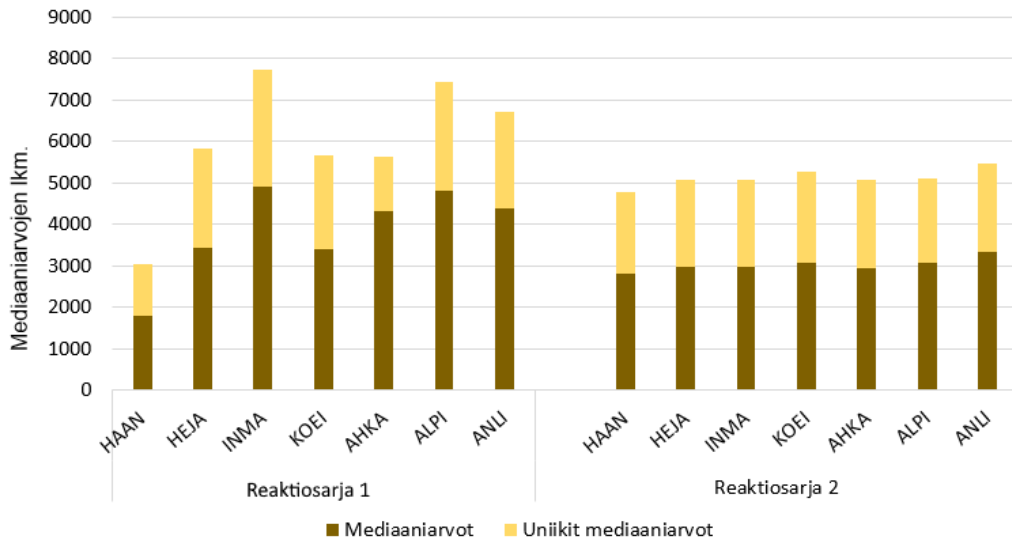


C

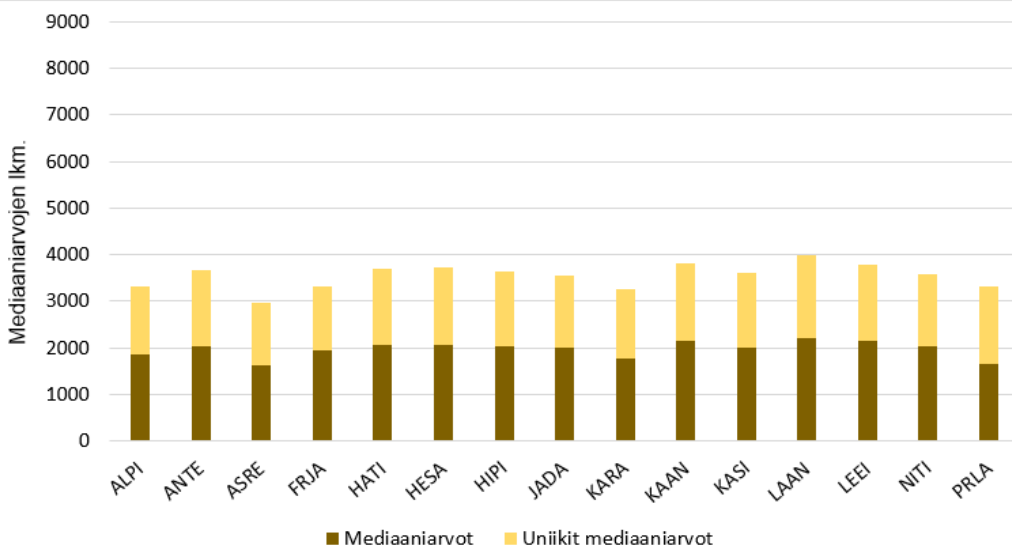


Kuva 7. (A) Reaktiosarjojen 1 ja 2, (B) 3 sekä (C) 4 Coffalyzer-ohjelmalla määriteltyjen kaikkien tunnistettujen ja tunnistamattomien sekvenssien lukumäärät. Suluissa olevat arvot kuvaavat uniikkien sekvenssien lukumäärää näytteessä.

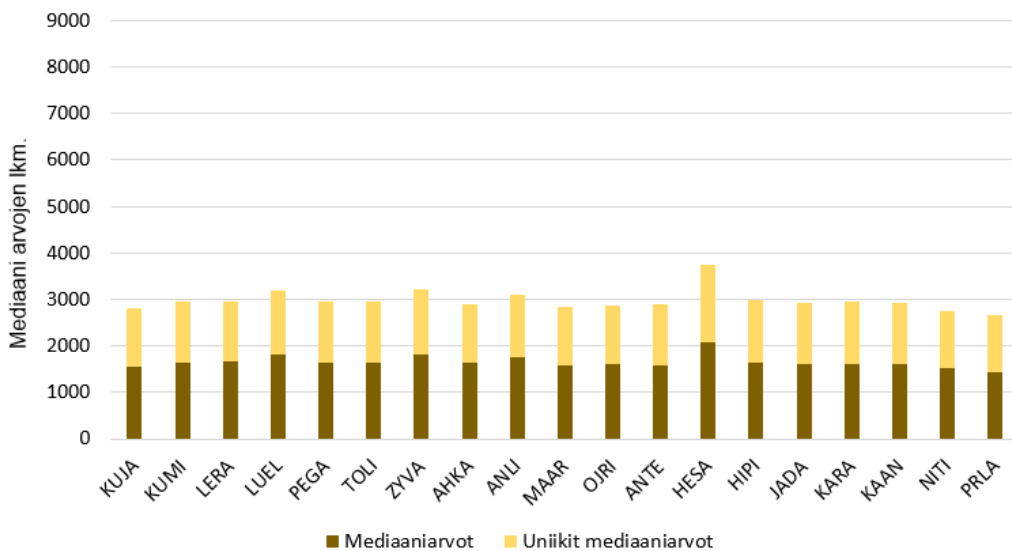
A



B



C



Kuva 8. Reaktiosarjojen (A) 1 ja 2, (B) 3 sekä (C) 4 Coffalyzer-ohjelmalla määriteltyjen kaikkien ja uniikkien mediaaniarvojen lukumäärät näytteiden verokkikoettimien osalta.

3.6 Coffalyzer digitalMLPA™ laaturaportit

Ensimmäisen reaktiosarjan näytteissä esiintyi eniten sekä varoitus, että virheilmoituksia niin reaktio- kuin komparatiivisessa analyysissä (Liite 2.). Kaikissa näytteissä toistui varoitus- tai virheilmoitus tuntemattomista sekvenssien lukumäärästä reaktioanalyysissä, sekä varoitus- tai virheilmoitukset vertaisnäytteiden ja -koettimien laadusta komparatiivisessa analyysissä. Lisäksi lähes kaikissa näytteissä oli muita varoitus- tai virheilmoituksia, johtuen hylättyyn digitalMLPA™ Coffalyzer-ohjelman laaturaporttiin.

Toisen reaktiosarjan näytteistä kaikissa esiintyi ainoastaan varoitusilmoitus käytettyjen vertaisnäytteiden laadusta komparatiivisessa analyysissä. Kaikki näytteet läpäisivät ohjelman laaduntarkistuksen. Reaktiosarjojen kolme ja neljä näytteissä ei ollut yhtään varoitus- tai virheilmoitusta, ja kaikki näytteet läpäisivät digitalMLPA™ Coffalyzer:in laaduntarkistuksen.

Lähes kaikissa eri reaktioiden näytteissä oli ohjelman mukaan käytetty yli 30 ng näyte-DNA:ta, paitsi muutamassa oli joko hieman tai huomattavasti alle sen, alittaen menetelmälle määritellyn minimimäärän (Taulukko 9.).

Taulukko 9: Eri reaktiosarjojen näytteet, joiden DNA-pitoisuudet olivat digitalMLPA™ Coffalyzer-ohjelman mukaan yhtä suuria tai alle kuin 30 ng.

Näyte (reaktiosarja)	digitalMLPA™ Coffalyzer-ohjelman ilmoittama DNA:n määrä (ng)
HAAN (1)	18
AHKA (1)	10
ALPI (1)	26
FRJA (1)	23
LERÄ (4)	29
LUEL (4)	28
PEGA (4)	29
AHKA (4)	24
OJRI (4)	30

4. Pohdinta

4.1 Syöpä- ja normaalinäytteet

Kaikista kopiolumuutosnäytteistä paitsi LERA-näytteestä onnistuttiin tunnistamaan tutkittavana olevien eksonien kopiolumuutokset. LERA-näytteestä tunnistettiin vain osa, mutta näytteestä oli tiedossa oletettu pseudogeenin olemassaolo. Tähän liittyen *SMAD4*-geenistä on tunnistettu prosessoitu pseudogeeni, joka ilmenee digitalMLPA:ssa geenin eksonien 2, 5, 6, 9, 10, 11 ja 12 kahdentumana (MRC Holland 2021). Muut geenin eksonit ilmenevät normaalina pseudogeenin myötä. Eli suoritettussa digitalMLPA-kokeessa on tunnistettu prosessoitu pseudogeeni, jonka on laskettu ilmenevän noin 0,3 % frekvenssillä, ja joka on todennäköisesti kliinisesti merkitsemätön (MRC Holland 2021). Prosessoidut pseudogeenit ovat seurausta retrotranspositiosta, jossa DNA:sta transkriptoitu lähetti-RNA kääntheistranskriptoidaan komplementaariseksi DNA:ksi ja liitetään osaksi genomia (Pink ym. 2011). Komplementaarinen DNA-juoste ei sisällä introneja, sekä se voi integroitua eri osaan genomia kuin alkuperäinen DNA-sekvenssi retrotranspositiion johdosta.

Miksi lähtökohtaisesti LERA-näytteestä haettiin koko geenin kahdentumaa, johtuu pseudogeenin vaikutuksesta näytteelle tehtyyn massasekvensointianalyysiin. Millson ym. (2015) todensivat artikkelissaan, kuinka *SMAD4*-geenin prosessoitu pseudogeeni aiheuttaa virheellisiä sekvenssilukemia massasekvensointianalyysissä. Millson ym. (2015) perustivat väitteensä MLPA:sta ja aCGH:sta saatuihin tuloksiin, jotka olivat ristiriidassa heidän massasekvensointianalyysinsä kanssa. He ehdottavat, että geenin eksonissa esiintyvät kahdentumat ovat seurausta pseudogeenien sekvensseistä, joita ei pystytä erittelemään massasekvensointianalyysissä oikeista *SMAD4*-sekvensseistä. Samanlaiseen johtopäätökseen päätyivät Mancini-Dinaro ym. (2015) työssään, jossa he pyrkivät todentamaan prosessoidun pseudogeenin olemassaolon *SMAD4* ja *NBK*-geenissä. Heidän massasekvensointi tuloksensa *SMAD4* osalta viittasivat geenin eksonien välisiin häviämiin tukien prosessoidun pseudogeenin ilmentymistä. Lisäksi heidän MLPA ja aCGH kokeet havaitsivat kohonneita *SMAD4*-geenin eksonien koetin arvoja.

Edellä mainitut seikat huomioiden, työssä saatua tulosta voidaan pitää tarkempaan kuin aikaisemmin LERA-näytteelle suoritettu massasekvensointianalyysiä. Eli menetelmällä

ei todennettu koko geenin kahdentumaa, vaan varmennettiin näytteen pseudogeenin olemassaolo, joka on johtanut alkuoletukseen mahdollisesta koko geenin kahdentumasta.

Normaalinäytteiden osalta ainoat poikkeavuudet testikoettimien arvoissa olivat Taulukossa 3 luetelluissa testikoettimisissa ALPI-näytteessä. Menetelmälle normaalia kopiolumua kuvaavaksi testikoettimen arvoiksi oli määritelty 0,8–1,2 ja heterotsygoottista kahdentumaa kuvaaviksi viitearvoiksi 1,3–1,65. ALPI-näytteen testikoettimien S017561, S017494 ja S011008 asettuvat näiden kahden arvojoukon välille, antaen epäselvän kuvan oikeasta kopiolumusta. Lisäksi testikoettimen S010332 arvo oli 0,12 kuvastaen menetelmän ohjeistuksen mukaan epäselvää kopiolumua, sen ollessa kaikkien määriteltyjen arvojoukkojen ulkopuolella. Kun näyte testattiin uudestaan ja analysoitiin uudestaan kolmannessa reaktiosarjassa, joka sisälsi 13 normaaliksi määriteltyä DNA-näytettä digitalMLPA Coffalyzer-ohjelman analyysivaiheessa, samojen neljän testikoettimien arvot olivat normaalin kopiolumun mukaiset. Koska Coffalyzer-ohjelman laatimien tulosteiden laatuparametrit eivät kertoneet ALPI-näytteen osalta mahdollisesta virheestä reaktioiden valmistukseen liittyen, suoraa selitystä miksi toisessa reaktiosarjassa näytteessä ilmeni poikkeavat testikoettimien arvot, on haastavaa esittää.

Toisessa reaktiosarjassa verrokinäytteitä käytettiin kolme, joista ALPI-näyte oli yksi, jolloin ALPI-näytteen testikoettimien arvon määrittelyssä toimivat ainoastaan näytteet AHKA ja ANLI. Käytettyjen verrokinäytteiden matala lukumäärä toisessa reaktiosarjassa on mahdollisesti johtanut epäselviin testikoettimien arvoihin ALPI-näytteen osalta. Kolmannessa reaktiosarjassa, jossa testikoettimien määrittäminen pohjautui kaikkiin käytettyihin näytteisiin, nostatti huomattavasti Coffalyzer-ohjelman laskemien mediaaniarvojen määrää yksittäistä testikoettinta kohden, tarkentaen ohjelman määrittämiä testikoettimien arvoja.

Toinen poikkeava tulos normaalinäytteiden osalta oli *CHEK2*-geenin 1100delC-mutaatio LEEI-näytteessä kolmannessa reaktiosarjassa. Löytö ei linkity kopiolumuutoksiin vaan se löydettiin koetinsarjaan sisällytettyjen mutaatioespesifisten koettimien kautta. Mutaatio ei ollut tiedossa, kun työtä varten valittiin käytettävät DNA-näytteet. 1100delC-mutaatio *CHEK2*-geenissä on lukukehysmutaatio, johtuen yhden emäksen häviämästä geenin eksonissa 10, joka johtaa ennen aikaisen stop-kodonin muodostumiseen geenin sekvenssissä (Massink ym. 2015; Hallamies ym. 2017). *CHEK2* tuotteena on CHEK2-proteiini (engl. Checkpoint kinase 2), joka osallistuu DNA:n vauriomekanismien ja solusyklin säätelyyn (Massink ym. 2015). CHEK2-proteiini aktivoituu ATM-proteiinin

fosforylaation kautta, jolloin CHEK2 homodimerisoituu. Aktiivinen CHEK2-proteiini kykenee fosforyloimaan muun muassa p53 ja BRCA1-proteiineja osallistuen DNA:n vaurioiden korjausmekanismien säätelyyn, sekä solusyklin kulkuun (Massink ym. 2015). Ennen aikaisesta stop-kodonista johtuen CHEK2-proteiinista syntesisoituu solussa lyhentynyt aminohappoketju, joka on itsessään epästabiili, mutta myös lyhentyneet lähetti-RNA:t altistuvat nonsense-välitteiselle hajoamiselle solussa (Massink ym. 2015).

Suomessa *CHEK2*-geenin 11000delC-mutaation on todettu ilmenevän noin 1,1 %-1,4 % frekvenssillä ja se on yhdistetty noin nelinkertaiseen riskiin sairastua rintasyöpään niin naisilla kuin miehillä (Hallamies ym. 2017). Yleisesti mutaation on arvioitu ilmenevän noin 5,5 % frekvenssillä kaikilla suomalaisilla rintasyöpäpotilailla, joilla on todettu positiivinen perhehistoria mutaation suhteen. Miehillä rintasyöpävän lisäksi, mutaation on havaittu nostattavan riskiä sairastua eturauhasen syöpään (Hallamies ym. 2017).

Kaiken kaikkiaan onnistuneiden reaktiosarjojen 2–4 syöpänäytteiden kopiolumuutokset vastasivat oletettuja ja normaalinäytteet olivat normaaleja kopiolumuutosten suhteen. Tulosta mukaili myös tehty χ^2 -testi.

Verrattaessa onnistumista tunnistaa eksonien kopiolumuutoksia muuhun kirjallisuuteen, digitalMLPA-tekniikalla on onnistuneesti pystytty tunnistamaan kopiolumuutoksia potilasnäytteistä. Kosztolányi ym. (2018) selvittivät digitalMLPA:n luotettavuutta tunnistaa myelooman linkitettyjen 371 lokuksen kopiolumuutoksia 56 potilaan luuydinäytteistä eristetystä DNA:sta. Potilaat oli todettu myeloomapotilaiksi Internal Myeloma Working Group laatiman ohjeistuksen pohjalta. Negatiivisiksi näytteiksi he keräsivät verinäytteitä terveiltä vapaaehtoisilta henkilöiltä sekä luuydinäytteitä potilailta, joilla ei ilmennyt pahanlaatuista sairautta. Työssään he vertasivat digitalMLPA koetinsarjalla D006 (lot: X1-0613) saatuja tuloksia MLPA:lla ja iFISH-menetelmällä saatuihin tuloksiin pohjautuen samoihin näytteisiin. Tulokseksi heidän vertailustaan he saivat eri menetelmien yhteneväisyydeksi 95 %. Kosztolányi ym. (2018) arvioivat, että laskettu ero menetelmien välillä johtuu käytetyn digitalMLPA koetinsarjan suuremmasta koetinmäärästä. Koetinsarjalla tunnistettiin 154 kopiolumuutosta 45 näytteestä, joita ei tunnistettu MLPA:lla tai iFISH:llä.

Vastaavasti Benard-Slagter ym. (2017) selvittivät digitalMLPA-tekniikan luotettavuutta, mutta tätä varten he valmistsivat oman digitalMLPA-koetinsarjansa digitalMLPA D007 ALL. Heidän valmistamansa koetinsarjansa kohdistettiin tunnistamaan akuuttiin

lymfaattiseen leukemiaan (engl. acute lymphoblastic leukemia, ALL) linkitettyjen geenialueiden eksonien kopiolukumuutoksia. Koetinsarjaa varten he yhdistivät viisi jo käytössä olevaa MLPA-koetinsarjaa sekä valmistivat itse osan uuden koetinsarjan koettimista. Materiaalina he käyttivät 67 DNA-näytettä lymfaattista leukemiaa sairastavilta potilailta, jotka he heidän itse valmistamansa koetinsarjan kanssa tutkivat rinnakkain MLPA koetinsarjojen P335 ja P327 sekä SNP arrayn kanssa. Tulokseksi he saivat digitalMLPA:n diagnostiseksi herkkyudeksi 98,9 % sekä spesifisyydeksi 97,8 %. Eli he tunnistivat omalla digitalMLPA-koetinsarjallaan lähes kaikki muilla menetelmillä positiiviksi, että negatiivisiksi todetut näytteet.

Saman tyyppisesti kuin Benard-Slagter ym. digitalMLPA-tekniikan luotettavuuden selvittämiseksi Lips ym. (2020) valmistivat oman digitalMLPA-koetinsarjansa. Heidän koetinsarjansa oli tarkoitettu havaitsemaan kopiolukumuutoksia, jotka viittaisivat *BRCA1* ja *BRCA2* kaltaiseen rintasyöpään. Yleisesti *BRCA1* ja *BRCA2* kaltaisilla syöville viitataan syöpiin, jotka ovat seurausta mutaatioista, joiden seurauksena soluissa esiintyy homologisen rekombinaatiokorjausmekanismin vajavuutta, mutta havaittavissa ei ole selviä merkkejä *BRCA1* tai *BRCA2* mutaatioista ituradan soluissa (Bodily ym., 2020). Lips ym. (2020) valmistivat digitalMLPA-koettimet jo valmiiksi suunniteltujen array-CGH markkereiden pohjalta niin, että *BRCA1*-kaltaisia mutaatioita varten koettimia valmistettiin kaiken kaikkiaan 84 ja vastaavasti 206 *BRCA2*-kaltaisia mutaatioita varten. Lisäksi he sisällyttivät koetinsarjaan erilliset digitalMLPA laaduntarkistuskoettimet. Kokeessa he testasivat yhteensä 114 jo testattua rintasyöpäpotilaan DNA-näytettä, joista he tunnistivat 91 % positiivisista *BRCA1*-kaltaisista näytteistä ja 82 % positiivisista *BRCA2*-kaltaisista näytteistä. Lips ym. (2020) perustelivat koetinsarjan tärkeyttä varsinkin *BRCA2*-kaltaisten mutaatioiden osalta, koska nykyisillä MLPA-koetinsarjoilla ei pystytä kartoittamaan kaikkia näihin mutaatioihin liitettyjä geenialueita yhdellä MLPA-testillä, joka rajoittuu noin 50 eri geenialueeseen per koetinsarja. DigitalMLPA:n mahdollistaessa noin 700 eri geenialueen samanaikaisen kartoituksen, pystytään samanaikaisesti kartoittamaan kopiolukumuutokset noin 102 eri geenialueelta, jotka on linkitetty *BRCA2*-kaltaisiin syöpiin, mutta myös noin 34 geenialuetta, jotka ovat linkitetty *BRCA1*-kaltaisiin syöpiin (Lips ym. 2020).

4.2 Massasekvensointi (NGS)

Kaikissa reaktiosarjoissa ilmeni matalaa klusterointia jäaden Illuminan suositeltujen rajojen alapuolelle, jotka ovat MiniSeq-järjestelmälle 170–220 K/mm³ (Illumina 2016). Q30-arvon omaavien sekvenssien osuus sekvenssidatasta oli korkea, mikä kuvastaa saadun datan luotettavuutta. Sekvensoinnissa käytetty Q-arvo on laatu parametri, jolla kuvataan todennäköisyyttä emäsvirheille sekvensointituloksessa (Illumina 2011). Q30-arvo kuvastaa mahdollisuutta, että noin tuhannen emäksen välein yksi emäs on tunnistettu virheellisesti. Q20-arvo vastaavasti, että noin sadan emäksen välein yksi emäs on tunnistettu virheellisesti ja Q10-arvo, että noin kymmenen emäksen välein yksi emäs on tunnistettu virheellisesti. Yksi yleisimmistä tekijöistä, joka johtaa matalien Q-arvojen saamiseen on korkea klusteritiheys sekvensoinnin aikana (Illumina 2016). Klusteritiheys on Q-arvon tavoin laatu parametri sekvensoinnin onnistumiselle, jolla pyritään kuvaamaan, kuinka useita ryppäitä sekvensointia varten monistettuja DNA-sekvenssejä on muodostunut virtauskyvetille (Illumina 2016). Suuri määrä DNA-ryppäitä virtauskyvetillä voi estää sekvensointilaitteistoa tunnistamasta sekvensoitavia emäksiä fluoresoivien markkereiden määrän ollessa liian suuri. Liian matala rypäsmäärä johtaa alhaisen sekvensointidata määrän saamiseen.

Matalat sekvensoinnin aikaiset klusteritiheydet voivat selittyä usealla eri tavalla. Käytetyssä työohjeessa oli valmiiksi määriteltynä valmistajan käyttämät liuossuhteet DNA-kirjaston laimentamista varten, joita noudatettiin työn aikana. Tämän lisäksi noudatettiin valmistajan ohjetta DNA-kirjastojen preppaamisessa eli kirjastojen PCR-jälkeistä puhdistusta eikä laaduntarkistusta suoritettu. Jos työohjeesta olisi poikettu kirjaston valmistusvaiheessa, kirjaston laimentamisvaihetta olisi voitu muokata, esimerkiksi kolmannen ja neljännen reaktiosarjan kohdalla, testaten millä liuossuhteilla olisi voitu saada Illuminan viitearvojen mukaiset klusteritiheydet, kun valmistetun kirjaston konsentraatio on tiedossa.

Toinen mahdollinen selittävä tekijä matalalle klusteritiheydelle voi olla mahdolliset epäpuhtaudet, mitä kirjastoon on jäänyt PCR-reaktiosta (Illumina 2016). Epäpuhtaudet voivat olla muun muassa PCR-reaktiossa syntyneitä vajaita DNA-juosteita tai PCR-reaktiossa käytettyjen alukkeiden muodostamia alukepareja (engl. primer dimer), jotka estävät kokonaisten haluttujen DNA-sekvenssien liittymisen virtauskyvetille. Kolmas mahdollinen selittävä tekijä voi olla kirjaston denaturaation vajaavaisuus (Illumina, 2016). Kirjastojen denaturaatiossa käytettävä 1 mM NaOH-liuos valmistettiin juuri ennen

sen lisäämistä laimennettuihin kirjastoihin eri sekvensointi kerroilla. NaOH-liuosten valmistuksessa oltiin huolellisia, mutta niiden pH:ta ei tarkistettu. Jos liuoksen pH ei ollut >12,5 eli liuoksen emäksisyys oli liian matala, se ei sisällä tarpeeksi vapaita OH-ryhmiä, jotka purkavat vetysidoksia DNA-ketjussa guaniinin ja tymiinin väliltä (Wang ym. 2014; Illumina, 2016).

Matalasta klusteroinnista huolimatta kaikkien onnistuneiden näytteiden osalta eri reaktiosarjoissa tunnistettiin huomattavasti yli optimimäärä sekvenssejä ja mediaaniarvoja verrokkikoettimille per näyte Coffalyzer digitalMLPA™ -ohjelmalla tehtävää analyysiä varten (Kuvat 7. ja 8.). Vähimmäisvaatimuksena optimaalisten tulosten saamiselle, kun on käytetty noin 600 kopiolukua mittaavaa koetinta, on noin 240 000 sekvenssilukemaa per näyte ja 400 mediaanilukemaa per verrokkikoetin MRC Holland digitalMLPA™ yleisen työohjeen mukaan.

4.3 Coffalyzer digitalMLPA™ laaturaportit

Coffalyzer digitalMLPA™ -ohjelman laatimien laaturaporttien osalta kolmannessa ja neljännessä reaktiosarjassa ei esiintynyt poikkeamia, jotka olisivat mahdollisesti voineet vaikuttaa saatuihin tuloksiin. Toisessa reaktiosarjassa poikkeamia ilmeni kaikkien näytteiden osalta komparatiivisessa analyysissä verrokkikoettimien laatuun liittyen ja ensimmäisessä reaktiosarjassa oli useita eri poikkeamia, joiden pohjalta ei pystytty todentamaan varmuudella saatujen testikoettimien arvojen todenmukaisuutta. Toisen reaktiosarjan näytteiden varoitusilmoitukset komparatiivisessa analyysissä verrokkinäytteiden laadusta todennäköisesti johtuivat siinä käytetyssä ALPI-näytteen tiettyjen verrokkikoettimien poikkeavista arvoista, joihin ohjelman tekemä datan normalisointi pohjautui (Taulukko 5.).

Ensimmäisessä reaktiosarjassa tapahtunutta virhettä sen valmistuksen osalta selvitettiin varoitus- ja virheilmoitusten kautta digitalMLPA™ yleisen ohjeen kohdan 10.3 ”Coffalyzer digitalMLPA report warning troubleshooting” osiota apuna käyttäen. Ohjeen avulla mahdolliset virheet reaktion valmistuksessa rajattiin vajaaseen pipetointiin hybridisaatioreaktioseoksen sekä polymeraasireaktioseoksen osalta ja mahdolliseen liikaa haihtumiseen yön yli aikaisessa hybridisaatioreaktiossa. Näihin virhelähteisiin päädyttiin, kun osasta näytteistä saatiin virheilmoitukset polymeraasin aktiivisuudesta, hybridisaation onnistumisesta ja hybridisaation aikaisesta T_m:sta komparatiivisessa

analyysissä (Liite 2.). Liika haihtuminen ja tämän myötä reaktioiden sisäiset DNA-pitoisuuksien vaihtelut mahdollisesti selittävät, miksi ensimmäisessä reaktiosarjassa esiintyi runsaasti vaihtelua tunnistettujen ja tunnistamattomien sekvenssien lukumäärissä.

Ongelman ratkaisua lähestyttiin testaamalla työohjeen mukaisesti käytettyjä PCR-putkia (Nippon Genetics FG-016FC) pipetoimalla 16 putkeen noin 9 µl steriiliä H₂O ja inkuboimalla niitä 60 °C yön yli noin 18 h samassa PCR-laitteessa kuin reaktioiden valmistuksessa. Putkissa havaittiin seuraavana päivänä lieviä sulamisen merkkejä ja että steriiliä H₂O oli jäljellä alle 5 µl kaikissa putkissa. Tämän myötä toista reaktiosarjaa varten vaihdettiin toisiin PCR-putkiin (Starlab I1402-8100), mutta myös kiinnitettiin erityistä huomiota reaktioseosten valmistukseen ja pipetointiin.

4.4 Johtopäätökset

MRC Holland (2022c) kuvailee digitalMLPA:ta robustina sekä tarkkana molekyyliogenetiikan menetelmänä kopiolutujen tunnistuksessa, joka sisältää itsessään kattavan tulosten laadun tarkistuksen. Tulosten tarkastelun myötä voidaan olla samaa mieltä näiden kuvausten kanssa. Aloittaen tarkkuudella, menetelmällä onnistuttiin tunnistamaan tutkittavista syöpänäytteistä patogeeniset kopiolutumuutokset sekä tarkentamaan oletetun pseudogeenin olemassaolo yhdestä näytteestä. Lisäksi yhdestä syöpänäytteestä saatiin löydös mutaatio spesifisten koettimien avulla, jota ei ollut ennen havaittu näytteestä. Normaalinäytteiden osalta reaktiosarjojen 3 ja 4 normaalinäytteistä kaikista saatiin tasaiset koetin arvot, jotka olivat normaalin kopiolutun mukaiset.

Menetelmällä saatiin johtopäätöksiä varten soveltuvia tuloksia ilman DNA-kirjaston erillistä laadun tarkistusta tai puhdistusta ennen sen sekvensointia. Tämä kuvastaa menetelmän robustia luonnetta, jota voidaan pitää isona positiivisena tekijänä. Kahden työvaiheen pois jättäminen säästää aikaa näytteiden valmistuksessa ja lopullisen analyysin saamisessa. Lisäksi saadaan vähennettyä mahdollisten erheiden määrää, joita voi tapahtua valmistusvaiheessa menetelmän ollessa perinteisesti käsin suoritettava.

Erheillä jatkaen, mitä tämän gradutyön aikanakin sattui, menetelmä on valmistettu niin, että käyttäjä saa sen kautta suoran palautteen omasta työstään, kunhan hän on sisäistänyt menetelmään liittyvän teorian. Tässä työssä Coffalyzer digitalMLPA™ -ohjelman avulla pystyttiin tehokkaasti tunnistamaan sekä todentamaan ensimmäisessä reaktiosarjassa

tapahtuneet erheet ja toistamaan sen valmistus uudestaan onnistuneesti. Selkeä ongelmanratkaisu ja laadunvarmistus Coffalyzer digitalMLPA™ -ohjelmalla mahdollistuu suurella määrällä koettimia, jotka ovat sisällytetty koetinseoksiin laadunmittausta varten.

Menetelmän positiivisena puolena edellä kuvattujen seikkojen lisäksi on, kuinka sen avulla kopiolumuutosten tunnistaminen pystytään kohdistamaan useille eri geenialueille samaan aikaan. Tässä työssä käytetyt digitalMLPA-koetinseokset olivat kohdistettu 28 eri geenialueen eksoneille yli 600 koettimella, joka on huomattavasti enemmän kuin mihin MLPA-koetinseoksella pystytään (MRC Holland 2021; MRC Holland 2022f). Eri geenialueiden eksoneiden kopiolumuutokset ovat jakautuneet usealle eri syöpätyypille. MLPA-koetinseoksella tutkinta yleensä kohdistetaan yhdelle geenialueelle noin 40 koettimen avulla (Stuppia ym. 2012). Useille geenialueille kohdistetut geenitestaukset, jotka yleensä tehdään genomien tai eksomin sekvensoinnissa, mutta nykyään mahdollisesti eri testialustoilla, johtavat helposti kasvaneeseen määrään sekundaarisissa löydöissä (Katz ym. 2020). Sekundaarisia löydöksiä geenitesteissä luonnehditaan löydöiksi, jotka eivät suoraan liity selvityksen kohteena olevaan tautifenotyyppiin, mutta voivat muuten vaikuttaa tutkittavan henkilön terveyteen (Katz ym. 2020; MedlinePlus 2022).

Merkittävistä sekundaarisista löydöksistä genomien ja eksomin sekvensoinnin osalta pidetään vuosittain päivitettyä tiivistä listaa The American College of Medical Genetics and Genomics (ACMG) toimesta (Miller ym. 2022). Yleisenä suosituksena ACMG:n puolesta on, että listassa esitetyistä geenivarianteista kerrotaan testatulle henkilölle testin suorittaneen tahon toimesta, jos niistä jokin ilmenee sekundaarisena löydöksenä. Lista on valikoitu sellaisten geenien variantteja, jotka ovat liitetty vakaviin perinnöllisiin sairauksiin, joille on nykypäivänä olemassa hoitomuotoja, joilla pystytään vähintään hoitamaan sairauksien oireita (Katz ym. 2020).

Sekundaaristen löydösten havainnoinnissa, digitalMLPA:ta rajoittavat menetelmälle jo asetetut rajoitteet. Näitä ovat pistemutaatiot, kopiolumuneutraalit kääntymät sekä siirtymät ja 55–100 nt kokoisten koettimien ulkopuolelle jäävät alueet (MRC Holland 2021). Nämä tekijät rajoittavat potentiaalisten ACMG:n määrittelemien sekundaaristen löydösten määrää, mutta eivät sulje niitä täysin pois, kun osa niistä ovat kopiolumuutoksiin sidottuja mutaatioita. Työssä käytettyjen koetinseosten kontekstissa on mahdollista, että henkilö, jota testataan yksittäisen geenialueen

kopiolukumuutoksille, paljastuukin olevan toisen tai useamman eri mutatoituneen geenialueen kantaja. Nämä potentiaaliset sekundaarilöydökset eivät välttämättä ole linkitettyinä henkilön sen hetkiseen tautifenotyyppiin.

Yhteenvetona saaduista tuloksista voidaan todeta, että digitalMLPA on geenien häviämä ja kahdentuma-analyysimenetelmänä yhtä luotettava kuin MLPA kopiolukumuutosten havaitsemisessa. Menetelmän käyttöön liittyen on hyvä huomioida käyttäjän perehdytys reaktioiden valmistukseen luotettavaa analyysiä varten sekä kuinka menetelmällä pystytään kartoittamaan useiden eri geenialueiden kopiolukumuutoksia samanaikaisesti. Tutkittavat geenialueet riippuvat käytetystä koetinsarjasta.

Onnistuneiden reaktioiden valmistusten myötä menetelmälle laadittiin ohjeet reaktioiden valmistukselle ja digitalMLPA Coffalyzer-ohjelman käytöstä (Liitteet 3 & 4.). Ohjeet pohjautuivat opinnäytetyön aineisto ja menetelmät osuuteen sekä MRC Hollandin digitalMLPA™ yleiseen ohjeeseen ja Coffalyzer digitalMLPA™ käyttöohjeeseen (versioille 211020.1245 ja 220221.1522). Lisäksi tehdystä työstä laadittiin validointiraportti TYKS Genomiikan vastuualueelle (Liite 5.).

Kiitokset

Kiitokset kaikille työni ohjaajille neuvoista ja kannustuksesta opinnäytetyön tekemisen aikana. Kiitokset myös Genetics cancer predisposition -tutkimusryhmän väelle, erityisesti Jukka Karhulle työn laboratorio osuuden tukemisesta. Lisäksi kiitokset Jenni Tuomiselle TYKS Genomiikan vastuualueelta massasekvensoinnin ohjauksesta ja tukemisesta. Lopuksi kiitokset perheelle ja ystäville jatkuvasta tuesta työn aikana.

Lähdeluettelo

Arlt M. F., Wilson T. E. & Glover T. W. (2012). Replication stress and mechanisms of CNV formation. *Current Opinion in Genetics & Development*, 22, 204–210.

<https://doi.org/10.1016/j.gde.2012.01.009>

Beltrán-García J., Osca-Verdegal R., Mena-Mollá S. & García-Giménez J. L. (2019). Epigenetic IVD Tests for Personalized Precision Medicine in Cancer. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.00621>

Benard-Slagter A., Zondervan I., de Groot K., Ghazavi F., Sarhadi V., Van Vlierberghe P., De Moerloose B., Schwab K., Vettenranta K., Harrison C. J., Knuutila S., Schouten J., Lammens T. & Savola S. (2017). Digital multiplex ligation-dependent probe amplification for detection of key copy number alterations in T- and B-cell lymphoblastic leukemia. *The Journal of Molecular Diagnostics*, 19, 659–672.

<https://doi.org/10.1016/j.jmoldx.2017.05.004>

Bodily W. R., Shirts B. H., Walsh T., Gulsuner S., King M. C., Parker A., Roosan M & Piccolo S. R. (2020). Effects of germline and somatic events in candidate BRCA-like genes on breast-tumor signatures. *PLoS One*, 15.

<https://doi.org/10.1371/journal.pone.0239197>

Borun, P., De Rosa, M., Nedoszytko, B., Walkowiak J. & Plawski A. (2015). Specific Alu elements involved in a significant percentage of copy number variations of the *STK11* gene in patients with Peutz–Jeghers syndrome. *Familial Cancer*, 14, 455–461.

<https://doi.org/10.1007/s10689-015-9800-5>

Cardoso, A. R., Oliveira, M., Amorim A. & Azevedo L. (2016). Major influence of repetitive elements on disease-associated copy number variants (CNVs). *Human Genomics*, 10. <https://doi.org/10.1186/s40246-016-0088-9>

Chen L., Zhou W., Zhang L. & Zhang F. (2014). Genome architecture and its roles in human copy number variation. *Genomics & informatics*, 12, 136–144.

<https://doi.org/10.5808/GI.2014.12.4.136>

Deininger P. (1999). Genetic instability in cancer: caretaker and gatekeeper genes. *The Ochsner Journal*, 1, 206–209. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3145442/>

Ewald I. P., Ribeiro P. L. I., Palmero E. I., Cossio S. L., Giugliani R. & Ashton-Prolla P. (2009). Genomic rearrangements in *BRCA1* and *BRCA2*: a literature review. *Genetics and Molecular Biology*, 32, 437-446. <https://doi.org/10.1590/S1415-47572009005000049>.

GeneReviews. (2022). Learn more (deletion/duplication analysis). <https://www.ncbi.nlm.nih.gov/books/NBK5191/box/further_illus-314/?report=objectonly> [Luettu: 19.7.2022]

Hallamies S., Pelttari L. M., Poikonen-Saksela P., Jekunen A., Jukkola-Vuorinen A., Auvinen P., Blomqvist C., Aittomäki K., Mattson J. & Nevanlinna H. (2017). CHEK2 c.1100delC mutation is associated with an increased risk for male breast cancer in Finnish patient population. *BMC Cancer*, 17. <https://doi.org/10.1186/s12885-017-3631-8>

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144, 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>

Harel T., Pehlivan D., Caskey C. T., Lupski J. R. (2015). Mendelian, Non-Mendelian, Multigenic Inheritance, and Epigenetics. Muokkaajat: Roger N. Rosenberg, Juan M. Pascual. Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease 5. painos. 3-27. Academic Press. <https://doi.org/10.1016/B978-0-12-410529-4.00001-2>.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, 10, 551-564. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2864001/>

Idos G. & Valle L. (2021). Lynch syndrome. *GeneReviews® [Internet]*. <https://www.ncbi.nlm.nih.gov/books/NBK1211/>

Illumina. (2022a) NGS vs. Sanger sequencing <<https://www.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html>> [Luettu: 4.4.2022]

Illumina. (2022b) Sequencing by synthesis (SBS) technology
<<https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>> [Luettu: 4.4.2022]

Illumina. (2016). Optimizing cluster density on Illumina sequencing systems.
<<https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf>>
[Luettu: 18.11.2022]

Illumina. (2011). Quality scores for next-generation sequencing.
<https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf>
[Luettu: 18.11.2022]

Katz A. E., Nussbaum R. L., Solomon H. L., Williams M. S. & Biesecker K. G. (2020). Management of secondary genomic findings. *American Journal of Human Genetics*, 107, 3–14. <https://doi.org/10.1016/j.ajhg.2020.05.002>

Kerkhof J., Schenkel L. C., Reilly J., McRobbie S., Aref-Eshghi E., Stuart A., Rupar C. A., Adams P., Hegele R. A., Lin H., Rodenhiser D., Knoll J., Ainsworth P. J. & Sadikovic B. (2017). Clinical validation of copy number variant detection from targeted next-generation sequencing panels. *The Journal of Molecular Diagnostics*, 19, 905–920. <https://doi.org/10.1016/j.jmoldx.2017.07.004>

Kosztolányi S., Kiss R., Atanesyan L., Gángó A., de Groot K., Steenkamer M., Jáksó P., Matolcsy A., Kajtár B., Pajor L., Szuhai K., Savola S., Bödör C. & Alpár D. (2018). High-throughput copy number profiling by digital multiplex ligation-dependent probe amplification in multiple myeloma. *The Journal of Molecular Diagnostics*, 20, 777-788. <https://doi.org/10.1016/j.jmoldx.2018.06.004>

Krebs J., Goldstein E., Kilpatrick S. (2013). Lewin's essential genes. 3. painos. Jones & Bartlett learning.

Lieber M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry*, 79, 181–211. <https://doi.org/10.1146/annurev.biochem.052308.093131>

Life technologies™. (2011). Qubit™ 2.0 Fluorometer
<<https://bilab.uga.edu/ggf/wp-content/uploads/sites/51/2013/12/CO16190-Qubit-Brochure.pdf>> [Luettu: 16.6.2022]

Lips E. H., Benard-Slagter A., Opdam M., Scheerman C. E., Wesseling J., Frans B. L., Linn S. C., Savola S., & Nederlof P. M. (2020). BRCAness digitalMLPA profiling predicts benefit of intensified platinum-based chemotherapy in triple-negative and luminal-type breast cancer. *Breast Cancer Research*, 22. <https://doi.org/10.1186/s13058-020-01313-7>

Luukkonen T. (2020). What are the benefits of choosing whole exome plus test that includes the analysis of del/dups (CNVs)? *Blueprint Genetics*.
<<https://blueprintgenetics.com/resources/whole-exome-plus-test-includes-the-analysis-of-del-dups-cnvs/>> [Luettu: 4.1.2023]

Mancini-Dinaro D., Judkins T., Elias M. C. Daniels C., Marshall A., Holladay J., Qian Y. R. & Roa B. (2015). Myriad Genetics Laboratories, Inc., Salt Lake City, UT. Dosage analysis by next generation sequencing and microarray CGH indicates putative processed pseudogenes in *SMAD4* and *NBN*. ACMG.
https://myriad-web.s3.amazonaws.com/ACMG2015%20Pseudogenes_Mancini-Dinardo_Presented_26MAR2015.pdf

Massink M. P. G., Kooi I. E, Martens J. W. M., Waisfisz Q & Meijers-Heijboer H. (2015). Genomic profiling of *CHEK2**1100delC-mutated breast carcinomas. *BMC Cancer*, 15. <https://doi.org/10.1186/s12885-015-1880-y>

McGarrity T. J., Amos C. I., Baker M. J. (2021), Peutz-Jeghers Syndrome. *GeneReviews*® [Internet]. <https://www.ncbi.nlm.nih.gov/books/NBK1266/>

MedlinePlus. (2022). What are secondary findings from genetic testing?
<<https://medlineplus.gov/genetics/understanding/testing/secondaryfindings/>>
[Luettu: 21.12.2022]

MedlinePlus. (2023). What are the different types of genetics tests?
< <https://medlineplus.gov/genetics/understanding/testing/types/>> [Luettu 20.2.2023]

Miller D. T., Lee K., Abduö-Husn N. S., Amendola L. M., Brothers K., Chung W. K., Gollob M. H., Gordon A. S., Harrison S. M., Hersberger R. E., Klein T. E., Richards C. S., Stewart D. R., Martin C. L. (2022). ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 24, 1407–1414. <https://doi.org/10.1016/j.gim.2022.04.006>

Millson A., Lewis T., Pesaran T., Salvador D., Gillespie K., Gau C. L., Pont-Kingdom G., Lyon E. & Bayrak-Toydemir P. (2015). Processed pseudogene confounding deletion/duplication assays for *SMAD4*. *The Journal of Molecular Diagnostics*, 17, 576–582. <https://doi.org/10.1016/j.jmoldx.2015.05.005>

MRC Holland. (2021). Product description SALSA® digitalMLPA probemix D001-B1 hereditary cancer panel 1

MRC Holland. (2022a). Principle of MLPA.
<<https://www.mrcholland.com/technology/mlpa/technique>> [Luettu: 4.4.2022]

MRC Holland. (2022b). Principle of digitalMLPA.
<<https://www.mrcholland.com/technology/digitalmlpa/technique>> [Luettu: 4.4.2022]

MRC Holland. (2022c). digitalMLPA: Where MLPA meets NGS for broad copy number certainty.
<<https://www.mrcholland.com/technology/digitalmlpa>> [Luettu 21.12.2022]

MRC Holland (2022d) digitalMLPA™ general protocol for use with Illumina sequencing platforms.

MRC Holland (2022e) Coffalyzer digitalMLPA™ user manual for Coffalyzer digitalMLPA versions 211020.1245 and 220221.1522.

MRC Holland. (2022f). Product description SALSA® digitalMLPA probemix D001-C1 hereditary cancer panel 1

Pećina-Šlaus, N., Kafka, A., Salamon, I., & Bukovac, A. (2020). Mismatch repair pathway, genome stability and cancer. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.00122>

Peixoto, A., Pinheiro, M., Massena L., Santos C., Pinto P., Rocha P., Pinto C. & Teixeira M. R. (2013). Genomic characterization of two large Alu-mediated rearrangements of the *BRCA1* gene. *Journal of Human Genetics*, 58, 78–83. <https://doi.org/10.1038/jhg.2012.137>

Petrucelli, N., Daly, M. & Feldman, G. (2010). Hereditary breast and ovarian cancer due to mutations in *BRCA1* and *BRCA2*. *Genetics in Medicine*, 12, 245–259. <https://doi.org/10.1097/GIM.0b013e3181d38f2f>

Petrucelli N, Daly M. B. & Pal T. (2022). *BRCA1*- and *BRCA2*-Associated Hereditary Breast and Ovarian Cancer. *GeneReviews® [Internet]*. <https://www.ncbi.nlm.nih.gov/books/NBK1247/>

Pink R. C., Wicks K., Caley P. C., Punch E. K., Jacobs L. & Carter D. R. F. (2011). Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA*, 17, 792–798. <https://doi.org/10.1261/rna.2658311>

Qin D. (2019). Next-generation sequencing and its clinical application. *Cancer Biology & Medicine*, 16, 4–10. <https://doi.org/10.20892/j.issn.2095-3941.2018.0055>

Stuppia L., Antonucci I., Palka G. & Gatta V. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *International Journal of Molecular Sciences*, 13, 3245–3276. <https://doi.org/10.3390/ijms13033245>

Turner, D. J., Miretti, M., Rajan, D., Fiegler, H., Carter, N. P., Blayney, M. L., Beck, S., & Hurles, M. E. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature Genetics*, 40, 90-95. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2669897/>

Vali-Pour, M., Lehner, B. & Supek, F. (2022) The impact of rare germline variants on human somatic mutation processes. *Nature Communications* 13. <https://doi.org/10.1038/s41467-022-31483-1>

Wang X., Lim H. J. & Son A. (2014). Characterization of denaturation and renaturation of DNA for DNA hybridization. *Environmental Health and Toxicology*, 29. <https://doi.org/10.5620/eh.2014.29.e2014007>

Weinberg R. (2014). *The Biology of Cancer*. 2. painos. W. W. Norton & Company.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*, 10, 451–481. <https://doi.org/10.1146/annurev.genom.9.081307.164217>

Liitteet

Liite 1. (A) Näkymä Coffalyzer digitalMLPA™- ohjelmaan määritellyistä asetuksista ensimmäisen valmistetun reaktiosarjan pohjalta. Käytetty viivakoodilevy ja levyltä pipetoidut viivakoodiliuokset ilmenevät kohdista ”barcode” ja ”position”. Sample -kohtaan määriteltiin mikä näyte on käsitelty kullakin viivakoodiseoksella ja type -kohtaan mikä on tutkittavan näytteen tyyppi. Sheet -kohta ilmaisee reaktioiden valmistuksessa käytetyn koetinsarjan. (B) Näkymä Coffalyzer digitalMLPA™ -ohjelmaan määritellyistä asetuksista toisen valmistetun reaktiosarjan pohjalta. Asetukset ovat samat kuin ensimmäisessä, mutta näytteet ovat merkattu eri viivakoodiliuoksilla. (C) Näkymä Coffalyzer digitalMLPA™ -ohjelmaan määritellyistä asetuksista kolmannen reaktiosarjan pohjalta. Kaikki näytteet ovat määritelty testinäytteiksi, jolloin sekvensoinnin jälkeinen datan normalisointi pohjautuu kaikkiin näytteisiin. Käytetty koetinsarja vaihdettiin uudempaan digitalMLPA D001-C1 koetinseokseen. (D) Näkymä Coffalyzer digitalMLPA™-ohjelmaan määritellyistä asetuksista neljännen valmistetun reaktiosarjan pohjalta. Näytteet 14–15 olivat peräisin kolmannesta reaktiosarjasta toimien verrokkinäytteinä näytteille 1–12.

A

	barcode	position	plate	sample	type	options	gender	tracker(s)	experiment	type	sheet
▶	BP01-09	A02	1	HAAN	Test	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-10	B02	1	HEJA	Test	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-11	C02	1	INMA	Test	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-12	D02	1	KOEI	Test	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-13	E02	1	AHKA	Reference	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-14	F02	1	ALPI	Reference	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-15	G02	1	ANLI	Reference	Default	Female		Validointi1	digitalMLPA	B1-1119
	BP01-16	H02	1	No DNA	No DNA	Default	Undefined		Validointi1	digitalMLPA	B1-1119

B

	barcode	position	plate	sample	type	options	gender	tracker(s)	experiment	type	sheet
▶	BP01-17	A03	1	HAAN	Test	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-18	B03	1	HEJA	Test	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-19	C03	1	INMA	Test	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-20	D03	1	KOEI	Test	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-21	E03	1	AHKA	Reference	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-22	F03	1	ALPI	Reference	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-23	G03	1	ANLI	Reference	Default	Female		Validointi2	digitalMLPA	B1-1119
	BP01-24	H03	1	No DNA	No DNA	Default	Undefined		Validointi2	digitalMLPA	B1-1119

C

	barcode	position	plate	sample	type	options	gender	tracker(s)	experiment	type	sheet
▶	BP01-25	A04	1	1. ALPI	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-26	B04	1	2. ANTE	Test	Default	Male		Validointi3	digitalMLPA	C1-1221
	BP01-27	C04	1	3. ASRE	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-28	D04	1	4. FRJA	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-29	E04	1	5. HATI	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-30	F04	1	6. HESA	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-31	G04	1	7. HIPI	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-32	H04	1	8. JADA	Test	Default	Male		Validointi3	digitalMLPA	C1-1221
	BP01-33	A05	1	9. KARA	Test	Default	Male		Validointi3	digitalMLPA	C1-1221
	BP01-34	B05	1	10. KAAK	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-35	C05	1	11. KASI	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-36	D05	1	12. LASU	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-37	E05	1	13. LAAN	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-38	F05	1	14. LEEI	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-39	G05	1	15. NITI	Test	Default	Male		Validointi3	digitalMLPA	C1-1221
	BP01-40	H05	1	16. PRLA	Test	Default	Female		Validointi3	digitalMLPA	C1-1221
	BP01-41	A06	1	17. No DNA	No DNA	Default	Undefined		Validointi3	digitalMLPA	C1-1221

D

	barcode	position	plate	sample	type	options	gender	tracker(s)	experiment	type	sheet
▶	BP01-26	B04	1	14. ANTE	Reference	Default	Male		Validointi4	digitalMLPA	C1-1221
	BP01-30	F04	1	15. HESA	Reference	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-31	G04	1	16. HIPI	Reference	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-32	H04	1	17. JADA	Reference	Default	Male		Validointi4	digitalMLPA	C1-1221
	BP01-33	A05	1	18. KARA	Reference	Default	Male		Validointi4	digitalMLPA	C1-1221
	BP01-34	B05	1	19. KAAK	Reference	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-39	G05	1	20. NITI	Reference	Default	Male		Validointi4	digitalMLPA	C1-1221
	BP01-40	H05	1	21. PRLA	Reference	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-65	A09	1	1. KUJA	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-66	B09	1	2. KUMI	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-67	C09	1	3. LERA	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-68	D09	1	4. LUEL	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-69	E09	1	5. PEGA	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-70	F09	1	6. TOLI	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-71	G09	1	7. ZYVA	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-72	H09	1	8. AHKA	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-73	A10	1	9. ANLI	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-74	B10	1	10. KUPE	Test	Default	Male		Validointi4	digitalMLPA	C1-1221
	BP01-75	C10	1	11. MAAR	Test	Default	Male		Validointi4	digitalMLPA	C1-1221
	BP01-76	D10	1	12. OJRI	Test	Default	Female		Validointi4	digitalMLPA	C1-1221
	BP01-77	E10	1	13. No DNA	No DNA	Default	Undefined		Validointi4	digitalMLPA	C1-1221

Liite 2. (A) Ensimmäisen reaktiosarjan digitalMLPA™ Coffalyzer -ohjelman tulosteen pohjalta koottu taulukko reaktiosarjan näytteiden laatuparametreistä reaktioanalyysistä. (B) Ensimmäisen reaktiosarjan digitalMLPA™ Coffalyzer -ohjelman tulosteen pohjalta koottu taulukko reaktiosarjan näytteiden laatuparametreistä komparatiivisesta analyysistä. Kuva tehtiin Microsoft Excel -ohjelmalla.

A

	HAAN	HEJA	INMA	KOEI	AHKA	ALPI	ANLI
1							
2							
3	Virheilmoitus	Varoitusilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus
4							
5	Varoitusilmoitus			Varoitusilmoitus			Virheilmoitus
6	Virheilmoitus						
7	Varoitusilmoitus						
8							
9							
10							
11	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus		

- 1 = Lukusyvyys
- 2 = Sekvenssien laatu
- 3 = Tuntemattomat sekvenssit
- 4 = Näyte DNA:n denaturaatio
- 5 = Näyte DNA:n depurinaatio
- 6 = Hybridisaation Tm
- 7 = Hybridisaation onnistuminen
- 8 = Ligaation aloitus lämpötila
- 9 = Ligaasin aktiivisuus
- 10 = Ligaasin inaktiivisuus
- 11 = Polymeeraasin aktiivisuus

Ok
Varoitusilmoitus
Virheilmoitus

B

	HAAN	HEJA	INMA	KOEI	AHKA	ALPI	ANLI
1	Virheilmoitus	Virheilmoitus		Virheilmoitus			
2	Virheilmoitus	Varoitusilmoitus		Varoitusilmoitus			
3	Virheilmoitus	Varoitusilmoitus		Varoitusilmoitus			
4	Virheilmoitus			Virheilmoitus			
5	Virheilmoitus				Varoitusilmoitus		Varoitusilmoitus
6							
7	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus		
8	Virheilmoitus	Virheilmoitus	Varoitusilmoitus	Virheilmoitus	Varoitusilmoitus	Virheilmoitus	Virheilmoitus
9	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus	Virheilmoitus

- 1 = Näyte DNA:n denaturaatio
- 2 = Näyte DNA:n depurinaatio
- 3 = Näyte DNA:n fragmentti koko
- 4 = Hybridisaatio Tm
- 5 = Hybridisaation onnistuminen
- 6 = Ligaasin aktiivisuus
- 7 = Polymeeraasin aktiivisuus
- 8 = Vertaiskoettimen laatu
- 9 = Vertaisnäytteen laatu

Ok
Varoitusilmoitus
Virheilmoitus

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

DIGITALMLPA (DIGITAL MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION) VALMISTAJAN PROTOKOLLA

YLEISTÄ MENETELMÄSTÄ

MRC Hollandin, kehittämä digitalMLPA perustuu samaan konseptiin kuin perinteinen MLPA eli kaksiosaisten koettimien hybridisaatioon, niiden väliseen ligaatioon sekä alukkeiden kanssa hybridisoituneiden DNA-alueiden monistukseen PCR:llä. Perinteisestä MLPA:sta poiketen monistettujen sekvenssien tunnistus pohjautuu spesifeihin noin 10 nt kokosiin viivakoodioligoihin, jotka liittyvät osaksi koettimia hybridisaatiovaiheessa. Valmistettujen amplikonien emäsjärjestys selvitetään Illumina pohjaisella massasekvensoinnilla, ”sequnece by synthesis”. Massasekvensoinnin ja viivakoodioligojen avulla voidaan määrittää usean eri DNA-alueen eksonien relatiivinen kopioluku yksittäisessä reaktiossa. Riippuen käytetystä koetinseoksesta tunnistuksen kohteena voi olla jopa lähes 30 eri geenialueen eksonit. Sekvensoinnin jälkeisessä analyysivaiheessa digitalMLPA Coffalyzer-ohjelmalla pystytään tunnistamaan kopiolukumuutokset hybridisoituneiden koettimien suhteellisten arvojen pohjalta.

Genomista DNA:ta käytetään reaktiossa **40 ng** (laimennos 10 ng/μl). DNA:n eristys tehdään Qiasymphonyllä.

Tarvikkeet:

Finnpipettejä tai vastaavia (0,3-3; 1-10; 5-50; 200 - 1000 μl)

Steriilejä filterillisiä pipetinkärkiä

Steriilejä 1,5 ml:n putkia

Steriilejä 0,5 ml:n putkia (Starlab I1405-8100 tai vastaava)

Steriilejä 0,2 ml:n putkia (Starlab I1402-8100 tai vastaava)

96-kuoppalevy ja kansi

Sentrifuugi (Eppendorf Centrifuge 5415C tai vastaava)

PCR-kone

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Steriilejä 1,5 ml LoBind putkia

Illumina MiniSeq System -sekvensaattori [# FC-420-1002 MiniSeq High Output Kit (150 cycle)]

Jäitä

Reagenssit:

digitalMLPA-koetinseos (MRC Holland), sisältäen tutkittaville genomien alueille spesifisen koetinseoksen.

digitalMLPA reagenssisarja, johon sisältyy: SALSA MLPA puskuri (kelt. korkki), SALSA ligaasi 65 (vihr. korkki), SALSA ligaasipuskuri A (läpinäk. korkki), SALSA ligaasipuskuri C (pink. korkki), SALSA PCR alukeseos P5P7 (rusk. korkki) ja SALSA polymeraasi (valk. korkki).

SALSA® viivakoodiliuoslevy BP01-IL-v03

0,1 X TE-puskuri

ddH₂O (kaupallinen)

Tuore 0,1 M NaOH-liuos

Kontrollinäytteet: Molempaa sukupuolta kohden täytyy olla vähintään kolme vertaisnäytettä. Kun näytteiden lukumäärä on yli 21, lisää yksi vertaisnäyte jokaista seitsemää näytettä kohden (esim. kun 28 testattavaa näytettä, väh. 4 vertaisnäytettä per sukupuoli). Varsinaisten näytteiden ja kopiolutukontrollien lisäksi tehdään aina vesikontrolli

Koetinseokset on jaettu pienempiin osiin 0,5 ml:n putkiin. Koetinseokset ovat PCR-laboratorion -20 °C:n pakastimessa. Kitteihin on merkitty päivämäärä, johon asti reagenssien pitäisi olla toimivia. Uusien kittien tilaus tehdään MRC-Hollandilta sähköpostitse (kts. tarkemmat ohjeet tilauksista osoitteesta <http://www.mlpa.com>).

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Huom! Näytteiden nimeäminen: Tutkimus, näytenumero, nimikirjaimet (Tutkimus_DN13-XXXX_NN).

Liittyvät työohjeet:

Illumina MiniSeq System Laiteohje

TYÖOHJE

PCR-kontaminaation välttämiseksi ja havaitsemiseksi noudatetaan seuraavia työtapoja:

PCR-reaktiota edeltävät vaiheet (denaturaatio, hybridisaatio, ligaatio) tehdään puhdistilassa (ks. ohje SOP-DNA-14) olevalla tähän tarkoitukseen erikseen varatulla PCR-koneella. Tätä PCR-konetta ei saa käyttää PCR-reaktioiden tekemiseen, vaan pelkästään sitä edeltävien inkubaatioiden toteuttamiseen.

PCR-reaktioseos pipetoidaan ligaatiovaiheen jälkeen putkiin puhdistilassa, jonka jälkeen suljetut putket viedään PCR-koneeseen PCR-reaktiota varten. Valmiita PCR-tuotteita sisältäviä putkia ei saa tuoda puhdistilaan.

DNA-NÄYTTEET JA KÄYTETTÄVÄT MÄÄRÄT

Taulukko 1: DNA-näytetyypit ja niitä tutkimuksessa käytettävät määrät.

Näytteen tyyppi	liuoksen DNA-konsentraatio	Käytettävä DNA, kokonais määrä ng	Käytettävä DNA-liuoksen määrä µl
Valmiiksi eristetty DNA (Fimlab, Nordlab ja HUS)	10 ng/µl	40 ng	4 µl
Qiasymphony	10 ng/µl	40 ng	4 µl

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

1. PÄIVÄ: DENATURAATIO

1. Valmista DNA-laimennokset. DNA-laimennetaan 0,1 X TE-puskuriin.
2. Sulata viivakoodiliuoslevy huoneen lämmössä, jonka jälkeen sentrifuugaa se 1000 rpm, sekoita lyhyesti levysekoittimessa ja lopuksi sentrifuugaa uudestaan 1000 rpm.
3. Pipetoi kuoppalevyllä 4 µl DNA-laimennosta ja 2 µl viivakoodiliuosta per reaktio. Jokaiseen reaktioon pipetoidaan eri viivakoodiliuosta viivakoodilevyiltä. Spinnaa putken pohjalle ja laita PCR-koneeseen.

Denaturaatio-ohjelma: 10 min 98 °C, pause 25 °C

4. Denaturaation jälkeen, kun reaktiot ovat 25 °C, ne voi poistaa koneesta.

SALSA-KOETTIMIEN HYBRIDISAATIO

1. Vorteksoi tai sekoita pipetillä MLPA puskuuri ja koetinseos ennen käyttöä.
2. Valmista hybridisaatioseos taulukon 2 mukaan denaturaatio-ohjelman aikana. Sekoita seos pipetoimalla edestakaisin tai vorteksoimalla:

Taulukko 2: Hybridisaatioseoksen koostumus. Taulukon luvut ovat mikrolitroina.

Reagenssi	1x
SALSA Probe mix	1,5
SALSA MLPA buffer	1,5
yhteensä	3

3. Kun denaturaatio-ohjelma on päättynyt, lisää hybridisaatioseosta 3 µl per reaktio, sekoita pipetoimalla edestakaisin. Jos tarpeen, spinnaa lyhyesti putken pohjalle. Kokonaistilavuus 9 µl.
4. Laita reaktiot PCR-koneeseen ja käynnistä hybridisaatio-ohjelma.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Hybridisaatio-ohjelma: 1 min 95 °C, 16–20 h 60 °C.

2. PÄIVÄ: LIGAATIO

1. Valmista ligaatioseos taulukon 3 mukaan. Vorteksoi ja spinnaa ligaasi-puskurit ennen käyttöä. Sekoita ligaasi 65 pipetoimalla sitä edestakaisin ennen käyttöä. Sekoita lopuksi valmis ligaatioseos pipetoimalla edestakaisin.

Taulukko 3: Ligaatioseoksen koostumus. Taulukon luvut ovat mikrolitroina.

Reagenssi	1x
Ligase-65 buffer A	3
Ligase-65 buffer C	3
dH ₂ O	25
Salsa Ligase-65	1
yhteensä	32

2. Käynnistä ligaatio-ohjelma suoraan hybridisaatio-ohjelman jälkeen.

Ligaatio-ohjelma: pause 48 °C, 30 min 48 °C, 5 min 98 °C, 20 min 65 °C, pause 20 °C

3. Kun blokin lämpötila on 48 °C, paina ”Pause” (ohjelma pysähtyy korkeintaan 10 minuutiksi).
4. Lisää kuhunkin reaktioon **32 µl** ligaatioseosta. Sekoita reaktiot pipetoimalla edestakaisin välttämällä ilmakuplia, sillä näytteitä ei saa poistaa PCR-koneesta spinnausta varten. Kun olet pipetoinut ligaatioseoksen kaikkiin putkiin, paina ”Resume” ja ohjelma jatkuu.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

5. Kun blokin lämpötila on 20 °C, paina ”Pause”. Reaktiot voi ottaa pois koneesta. Mikäli et jatka saman tien PCR-vaiheeseen, säilytä ligaatioseoksia 4 °C:ssa (jääkaapissa).
6. Ennen ligaatio-ohjelman loppumista valmista PCR-seos taulukon 4 mukaisesti. Vorteksoi PCR alukeseos kevyesti. Sekoita polymeraasi pipetoimalla sitä edestakaisin.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

2. PÄIVÄ PCR

Taulukko 4: PCR-seoksen koostumus. Taulukon luvut ovat mikrolitroina.

Reagenssi	1x
dH ₂ O	7
Salsa MLPA-primermix	2
P5P7	
SALSA Polymerase	1
yhteensä	10

1. Valmista PCR-seos taulukon 4 mukaan ja sekoita valmis seos pipetoimalla edestakaisin ja putkea naputtelemalla. Älä vorteksoi! Spinnaa lyhyesti.
2. Lisää **10 µl** PCR-seosta per reaktio ja sekoita pipetillä.
3. Laita reaktiot PCR-koneeseen ja aloita PCR-ohjelma (Taulukko 5).

Taulukko 5: PCR-ohjelma.

Vaihe	Lämpötila °C	Kesto	Toisto
1.	65	1 min	-
2. Denaturaatio	95	30 s	
3. Alukkeiden kiin.	65	40 s	x 45
4. Ekstensio	72	90 s	
5.	15	pause	-

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

NÄYTTEIDEN SEKVENSOINTI ILLUMINA MINISEQ SYSTEM SEKVENSAATTORILLA

1. Valmiit PCR-tuotteet voi säilyttää valolta suojattuna jääkaapissa maksimissaan yhden viikon. Pidempää säilöntää varten näytteet voi säilyttää $-25\text{ }^{\circ}\text{C}$ ja $-15\text{ }^{\circ}\text{C}$ välillä.
2. Valmistele Illumina MiniSeq sekvensaattori käyttöä varten Illumina MiniSeq System Laiteohjeen mukaan. Sekvensoinnissa käytetään Illumina # FC-420-1002 MiniSeq High Output Kit (150 cycle) virtauskyvettä ja reagenssikasettia.
3. Sekvensointia varten valmista kokoomanäyte (DNA-kirjasto) pipetoimalla **5 μl** jokaista sekvensoitavaa näytettä yhteen 1,5 ml LoBind putkeen. Yhteen kokoomanäytteeseen voi kerätä maksimissaan 64 eri PCR-tuotetta. Sekoita valmis kokoomanäyte vorteksoimalla lyhyesti.
4. Pipetoi **5 μl** kokoomanäytettä **75 μl** ddH₂O. Tämä on noin 12,5 nM laimennos kokoomanäytteestä.
5. Pipetoi **4 μL** 12,5 nM laimennoksesta **251 μl** ddH₂O. Tämä on noin 200 pM laimennos kokoomanäytteestä.
6. Pipetoi **5 μl** 200 pM laimennosta **5 μl** tuoreeseen 0,1 NaOH. Sekoita hyvin vorteksoimalla ja inkuboi huoneen lämmössä 5 min.
7. Lisää **990 μl** vasta sulatettua HT1-puskuria 5 min inkuboinnin jälkeen. Tämä on noin 1 pM denaturoitu laimennos kokoomanäytteestä. Säilytä jäällä, kunnes aloitat sekvensoinnin.
8. Pipetoi **500 μl** denaturoitua kirjastoa oransilla merkittyyn kaivoon Illumina MiniSeq reagenssikasetilla. Noudata ajon käynnistyksessä Illumina MiniSeq System Laiteohjetta. Ajopohjaa laatiessasi; valitse ”Library prep kit” kohdassa ”Custom”, aseta ”Index reads” 0, valitse ”Single read” ja aseta lopuksi ”Read

digitalMLPA, validointiraportti (Ville Raappana)

Liite 3.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

length (Read 1)" 110. Kirjoita sekvensoitavan kirjaston nimi kohdan "SampleID" alle ensimmäiselle riville. Käynnistä ajo.

Tulokset analysoidaan Coffalyzer -ohjelmalla (kts. digitalMLPA-tulosten analysointi Coffalyzer-ohjelmalla).

JÄTTEENKÄSITTELY

Sekvensoinnin jälkeisessä jätteiden käsittelyssä noudatetaan Illumina MiniSeq System Laiteohjetta.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

DIGITALMLPA TULOSTEN ANALYSOINTI DIGITALMLPA COFFALYZER-OHJELMALLA

Periaate:

Valmis digitalMLPA kokoomanäyte massasekvensoidaan Illumina MiniSeq sekvensaattorilla. Raakadata haetaan työasemalle verkon kautta (tarkentuu vasta käyttöönnoton myötä) ja analysoidaan Coffalyzer-ohjelmalla. Tarkat ohjeet ohjelman käytöstä löytyvät Coffalyzer digitalMLPA™ User Manual.

Käytettävät ohjelmistot ja tietokannat:

Coffalyzer digitalMLPA -ohjelma (v. 221020.1234 kirjoitus hetkellä)

Microsoft Excel

ANALYSOINTI

UUDEN POHJAN LUOMINEN

1. Avaa Coffalyzer-ohjelma (MRC.DataAnalysis.Client).
2. **Coffalyzer Definition File** alta valitse ”Create new digitalMLPA definition file” ja paina <Next>
3. Valitse kansio, jonne tallennat uuden pohjan. Tiedosto on muotoa .cfd.
4. **Data Conversion Definition** -ikkuna avautuu, jossa pystyt asettamaan uuden pohjan tarkemmat tiedot.
 - ”analysis settings” -> valitse ”Default profile”
 - ”barcode lot” -> valitse reaktioiden valmistuksessa käytetty viivakoodiliuoslevy (03-xxx-yymmdd ; Illumina standard)

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

5. Valitse näiden kahden asetuksen jälkeen <Configure>, josta aukeaa **Samples** -ikkuna. Tässä ikkunassa pystyt määrittelemään mitä viivakoodioligoita on käytetty eri reaktioissa.
 - Yli maalaa ne kaivot, joita on käytetty reaktioiden valmistuksessa (Voit käyttää hyväksi taulukon vasemmassa reunassa olevaa nuolta).
 - Paina hiiren oikeaa näppäintä maalattuasi kaikki käytetyt kaivot ja valitse “Paste Sample Details From Clipboard”. Tällä komennolla pystyt suoraan määrittelemään valmiista Excel -tiedostosta valmistetut näytteet, kun olet kopioinut ne Excelistä.
 - Kun olet kopioinut näytteiden tiedot paina hiiren oikeaa näppäintä ja valitse “Create New Experiment (Manual)”. “Experiment Overview” ikkuna avautuu.
 - Uudessa ikkunassa “experiment name” kohdan alle kirjoita (esim. SampleID sekvensoinnista). “experiment type” kohdassa valitse digitalMLPA. “product” ja “sheet” voivat olla “(auto detect during analysis)”.
 - Valitse <OK> ja ikkuna sulkeutuu ja palaat takaisin **Samples** -ikkunaan. Valitse <OK> ja ikkuna sulkeutuu ja palaat takaisin **Data Conversion Definition** -ikkunaan.
6. “undefined samples” kohdassa valitse “Ignore and exclude”.
7. Valitse kansio, jonne ohjelma tallentaa tulokset kohdasta “output directory”.
8. Lopuksi valitse <Add File> ja valitse se sekvensoinnista saatu FASTQ tiedosto (muotoa .fastq.gz) ja paina <OK>.
9. Ohjelma aloittaa analyysin, jonka kesto vaihtelee näytemäärän mukaan.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

10. Kun ohjelma on valmis paina <Open Results Folder>, josta voit tarkastella saatuja tuloksia

Huomioitavaa:

- Molempaa sukupuolta kohden täytyy olla vähintään kolme vertaisnäytettä. Kun näytteiden lukumäärä on yli 21, jokaista seitsemää näytettä kohden täytyy olla yksi vertaisnäyte.

UUDELLEEN ANALYSOINTI VALMIILLA POHJALLA

1. Avaa Coffalyzer-ohjelma (MRC.DataAnalysis.Client).
2. **Coffalyzer Definition File** alta valitse "Use existing definition file" ja paina <Next>
3. Valitse käytettävä Coffalyzer definition tiedosto (muotoa .cdf). **Data Conversion Definition** -ikkuna aukeaa.
4. Valitse <configure> muokataksesi tiedostoa. Tämä avaa **Samples** -ikkunan. Voit lisätä tai poistaa näytteitä tässä ikkunassa.
5. Voit lisätä näytteitä kopioimalla tiedot valmiista Excel-taulukosta tai manuaalisesti.
 - Manuaalisesti lisätessäsi näytteen kirjoita näytteen nimi "sample" kohdan alle ja valitse näytteen tyyppi, "type". Jos kyseessä on analysoitava näyte valitse "Test", "Reference" jos kyseessä on vertaisnäyte tai "no DNA" jos kyseessä on vesikontrolli. Lopuksi valitse näytteen sukupuoli.
6. Kun olet kopioinut tai lisännyt näytteet, yli maalaa ne ja paina hiiren oikeaa näppäintä. Valitse valikosta "Add to experiment" ja edelleen se koe, johon lisää näytteet. Valittujen näytteiden pitäisi näkyä vihreinä tämän jälkeen.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

7. Paina <OK> niin palaat takaisin **Data Conversion Definition** -ikkunaan.
8. Aseta undefined → Ignore and exclude.
9. Jos tarvitsee voit muuttaa ”output directory” tässä vaiheessa.
10. Tarvittaessa pystyt vaihtamaan analysoitavan FASTQ-tiedoston. Poista nykyinen FASTQ-tiedosto ensin valitsemalla se ”input files” laatikosta ja paina <Remove File>. Tämän jälkeen valitse uusi FASTQ-tiedosto <Add File> komennolla.
11. Kun olet valmis paina <OK> ja ohjelma käynnistää analyysin.
12. Ohjelman lopetettua analyysin avaa tulokset valitsemalla <Open results folder>.

UUELLEEN ANALYSOINTI COFFA-TIEDOSTOJEN AVULLA

Voit analysoida näytteitä uudestaan valmiiden Coffa-tiedostojen avulla, jonka Coffalyzer-ohjelma luo jokaiselle näytteelle analysoinnin jälkeen. Coffa-tiedostot löytyvät reaktiosarjojen tulokansiosta ja ovat muotoa *.coffa. Tätä kautta pystyt esimerkiksi valitsemaan jo analysoiduille näytteille uudet vertaisnäytteet. Käytettävät coffa-tiedostot täytyy koota yhteen kansioon **(muista sisällyttää vesikontrolli)**.

Jotta analysointi on mahdollinen, käytettävät coffa-tiedostot pitää olla tuotettu samalla Coffalyzer-ohjelman versiolla ja reaktioiden valmistuksessa on käytetty samaa koetinseosta. Ohjelma antaa virheilmoituksen, ”Error -failed to retrieve the search graph blob (xxx) specified in the first encounterd value (xxx)”, jos tiedostot ovat tuotettu ohjelman eri versioilla.

1. Avaa Coffalyzer-ohjelma (MRC.DataAnalysis.Client).
2. **Coffa files** kohdan alta valitse Analyze ja paina <Next>
3. Valitse kansio, johon käytettävät coffa-tiedostot ovat koottu. **Data Analysis Definition** ikkuna aukeaa.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

4. Aseta analysis settings → Default profile
5. Paina <Configure> avataksesi *Samples* -ikkunan, josta näet kaikki uuteen analyysiin sisällytetyt näytteet.
 - Määritä analyysin vertaisnäytteet valitsemalla käytettävien vertaisnäytteiden tyyppiä (type) Reference.
6. Kun olet valmis paina <OK> niin palaat takaisin ”Data Analysis Definiton” ikkunaan.
7. Lopuksi paina <OK> aloittaaksesi analyysin. Kun analyysi on valmis, avaa tulokansio painamalla <Open Results Folder>.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

TULOSTEN LAADUNTARKISTUS

1. Avaa kansio, jonka määrittelit ”output directory” päätteeksi tai jatka suoraan <Open Results Folder> valinnalla. Valitse analysoitu kansio. Kansiosta löytyy erilliset pdf-tiedostot jokaista näytettä kohden, sekä Overview pdf-tiedosto ja Ratios Excel-taulukko.
2. Voit aloittaa laaduntarkistuksen silmäilemällä Overview -tiedoston. Tästä tiedostosta pystyt nopeasti tarkistamaan, onko reaktioiden valmistus onnistunut. Taulukon ”Quality checks” alla näytteiden kohdalla lukee ”Passed” vihreällä pohjalla kun reaktion valmistus on onnistunut. Jos tämän sijaan taulukossa lukee ”Failed” punaisella pohjalla, reaktion valmistus on epäonnistunut eikä tulos ole luotettava. Vesikontrollin kohdalla lukee aina ”n/a”.
3. Tämän jälkeen voit edetä yksittäisten näytteiden silmäilyyn.
 - Jokaisesta näytteestä on kaksi eri pdf-tiedostoa: ”General details” ja ”Sample Results”. Molemmat tiedostot sisältävät muuten samat tiedot, mutta jälkimmäinen esittää näytteen poikkeavat koetin arvot, jos niitä on.
4. Avaa General details -tiedosto tukittavasta näytteestä tai vertaisnäytteestä.
5. Aluksi voit katsoa läpi **General Information** osion (Kuva 1.). Osiossa on hyvä kiinnittää huomiota:
 - Käytetyn DNA:n määrän, joka ei mielellään saa olla alle 20 ng
 - ”Product detection” eli onhan ohjelma tunnistanut käytetyn koetinseoksen
 - Vastaako tunnistettujen (Detected) koettimien lukumäärät odotettuja määriä (Expected) kohdassa **Probes**
 - ”Median total reads” on vähintään 400

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

- Vastaako tunnistettujen sekvenssien lukumäärä (Recognised -> Reads) vähimmäisvaatimusta valmistetulle reaktiolle. Esimerkiksi koetinseos D001-C1-1221 sisältää noin 725 koetinta, jolloin yksittäisessä näytteessä pitäisi olla vähintään $725 \times 400 = 290\,000$ tunnistettua sekvenssiä. Laskussa 400 on ”Median total reads”.
6. Seuraavaksi voit siirtyä **Quality Details** -taulukkoon (Kuva 1.). Taulukko koostuu kahdesta osasta:
- **Reaction analysis:** Osio pohjautuu ohjelman sisään rakennettuun laadunvertaukseen ennalta määritettyjen MRC Hollandin viitearvoihin.
 - **Comparative analysis:** Osio pohjautuu reaktiossa käytettyjen testattavien näytteiden ja vertaisnäytteiden väliseen vertailuun.

Onnistuneessa reaktiossa kaikissa taulukon kohdissa lukee Ok. Jos jokin reaktioiden valmistusvaihe on epäonnistunut, sen kohdalla lukee Failed. Jos taulukossa ilmenee Warning -ilmoitus, tulos on käytettävissä, mutta on hyvä selvittää mistä ilmoitus johtuu.

- Jos lähdet selvittämään virheiden lähdettä, käytä hyväksi digitalMLPA™ General Protocol kohta 10.3 Coffalyzer digitalMLPA report warning troubleshooting.
7. Lopuksi voit tarkistaa vesikontrollin (No DNA report). Vesikontrollista tärkeimmät kohdat ovat (Kuva 2.):
- Estimated DNA quantity: Näytteestä tunnistettu DNA:n määrä (ng). Näyte on ok, jos kohdassa lukee ”Insufficient reads for DNA quantity estimation (as expected)
 - Median distinct reads: Näytteestä laskettu uniikkien mediaanien määrä. Yleisesti, mitä pienempi sen parempi.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

- Product detection: Ilmoittaa onko reaktion valmistuksessa käytetty koetinseos tunnistettu. Jos on, kohdassa lukee ”Identity detected”.

General Details Report - BP01-75 - 11. MAAR



General Information

Software information	Executed	Program version
Configuration	n/a	2022-02-21 15.22
Data conversion	2022-08-29 14.39	2021-10-19 09.58
Fragment analysis	2022-08-29 14.40	2021-10-19 09.58
Comparative analysis	2022-08-29 14.40	2021-10-19 09.58
Extended analysis	n/a	n/a
Reporting	2022-08-29 14.40	2021-10-19 09.58

Sample

Sample name (type)	BP01-75 - 11. MAAR (Test)
Barcode - well (lot)	BP01-75 - C10 (03-009-yyymmdd)
SNP code	11021.01100.02011.12111.11112.11112.11111.0011 v1
Gender	Male
Estimated DNA quantity	> 30 ng
Sample track probes	(none)

Experiment settings

Experiment name (type)	Validointi4 (digitalMLPA)
Product name	D001-C1-1221 [default sheet]
Product detection	Identity detected
Digital signatures	All signatures valid
Probe catalogue	Default, Version 2022-02-21 15.22
Source files	F:\digitalMLPA_Validointi4_26082022-18322315\FASTQ_Generation_2022-08-26_17_19_40Z-27433540\Validointi4_L001-ds.602dedf3cad442a486b66bc7ea266fd81\Validointi4_S1_L001_R1_001.fastq.gz

Analysis details

Analysis settings	Default profile
Normalisation method	Individual reference probe normalisation [default]
Coffa file location	E:\Coffalyser Output\20220829_143645\Validointi4\BP01-75 - 11. MAAR.Coffa
Coffa file id	ef1ee51-a07d-4600-860f-7bbfb89b989b
Extended module	n/a
Read count type	Total
Reference samples	14. ANTE, 15. HESA, 16. HIPI, 17. JADA, 18. KARA, 19. KAAAN, 20. NITI, 21. PRLA

Reads	Items	Reads	Probes	Expected	Detected	Unexpected
Recognised	709	1 045 824	Test probes	579	579	0
Unrecognised	n/a	29 733	Reference probes	290	290	0
Median total reads	n/a	1 582	Y-control probes	3	3	n/a
Median distinct reads	n/a	1 264	Mutation probes	0-5	0	n/a

Quality Details

Quality Checks: Passed

Quality	Description	Classification
Reaction analysis	Read depth	Ok
	Sequence quality	Ok
	Unrecognised reads	Ok
	Sample DNA denaturation	Ok
	Hybridisation Tm	Ok
	Hybridisation completeness	Ok
	Ligation start temperature	Ok
	Ligase activity	Ok
	Ligase inactivation	Ok
	Polymerase activity	Ok
Comparative analysis	Sample DNA denaturation	Ok
	Sample DNA depurination	Ok
	Sample DNA fragment length	Ok
	Hybridisation Tm	Ok
	Hybridisation completeness	Ok
	Ligase activity	Ok
	Polymerase activity	Ok
	Reference probe quality	Ok
Reference sample quality	Ok	

Kuva 1: Esimerkki onnistuneen reaktion raportista. Kuvaan on rajattu punaisella kohdassa 5 esitetyt alueet, jotka ovat hyvä tarkistaa. Lisäksi raportista on rajattu punaisella laatutaulukko (kuvassa oik.), jossa kaikki ovat merkitty ”ok” eli reaktion valmistus on onnistunut.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

No DNA report - BP01-77 - 13. No DNA



General Information

Software information	Executed	Program version
Configuration	n/a	2022-02-21 15.22
Data conversion	2022-08-29 14.40	2021-10-19 09.58
Fragment analysis	2022-08-29 14.40	2021-10-19 09.58
Extended analysis	n/a	n/a
Reporting	2022-08-29 14.40	2021-10-19 09.58

Sample	
Sample name (type)	BP01-77 - 13. No DNA (No DNA)
Barcode - well	BP01-77 - E10
Estimated DNA quantity	Insufficient reads for DNA quantity estimation (as expected)
Sample track probes	(none)
Median distinct reads	2

Experiment settings	
Experiment name (type)	Validointi4 (digitalMLPA)
Product name	D001-C1-1221 [default sheet]
Product detection	Identity detected
Digital signatures	All signatures valid
Probe catalogue	Default, Version 2022-02-21 15.22
Source files	F:\digitalMLPA_Validointi4_26082022-18322315\FASTQ_Generation_2022-08-26_17_19_40Z-27433540\Validointi4_L001-ds.602dedf3cad442a486b66bc7ea266fd8\Validointi4_S1_L001_R1_001.fastq.gz

Analysis details	
Analysis settings	Default profile
Normalisation method	Individual reference probe normalisation [default]
Coffa file location	E:\Coffalyser Output\20220829_143645\Validointi4\BP01-77 - 13. No DNA.Coffa
Coffa file id	a8349786-7774-47ab-8b60-835c0ca313f6
Extended module	n/a
Read count type	Total

Kuva 2: Esimerkki vesikontrollin raportista. Raportista on rajattu kohdassa 7 esitetyt kohdat, jotka ovat hyvä tarkistaa kontrollin osalta. Vesikontrollista ei esitetä samanlaista laatutaulukkoa kuin testi- ja vertaisnäytteistä.

KOETINARVOJEN SELVITYS

1. Avaa tutkittavan näytteen ”Sample Results” pdf-tiedosto ja selaa tiedoston kohtaan ”Results Summary”.
2. Sivulla on kuvaaja, johon on sisällytetty kaikki käytetyn koetinseoksen eksonien kopiolumittaavien koettimien arvot (Inter ratio) (Kuva 3.). Jos näytteessä ilmenee poikkeavia koetinarvoja, sivulla on toinen kuvaaja, johon on eritelty poikkeavat koetinarvot. Katso selitykset koetinarvoille taulukosta 1.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

3. Geenialueiden kaikkien eksonien koetin arvot, joissa esiintyi poikkeavia koetin arvoja ovat taulukoitu kohdan ”Aberrant region(s) table” alle (Kuva 4.). Taulukon eri osiot ovat kerrottu taulukossa 2.
4. Koetin arvojen lisäksi tiedostossa esitetään spesifeistä mutaatioista, riippuen käytetystä koetinseoksesta kohdassa ”Mutation-specific probe(s)” (Kuva 4.). Näihin ei ole linkitetty omaa koetin arvoa, vaan ainoastaan ilmeneekö spesifi geenivariantti näytteessä.
 - Jos tarpeen, voit tarkastella kaikkien eri geenialueiden koetin arvoja ”Ratios” Excel-tiedostosta.

Result Summary

Note: Please refer to the 'Probe Information File' for additional information that may be crucial for correct result interpretation.

Chart(s)



Kuva 3: Esimerkki kuvaajat näytteen koetin arvoista. Näytteen kaikki muut koetin arvot ovat normaalin kopiolumun mukaiset, mutta geenialueen *RAD51C* kahdeksassa eksonissa ilmenee kohonneita koetin arvoja.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Taulukko 1: Eri kopioluvun tiloja kuvaavat arvojoukot.

Kopioluvun tila	Koetin arvo
Normaali	0,80 < arvo < 1,20
Homotsygootti häviämä	arvo = 0
Heterotsygootti häviämä	0,40 < arvo < 0,65
Heterotsygootti kahdentuma	1,30 < arvo < 1,65
Heterotsygootti kolminkertaistuma / Homotsygootti kahdentuma	1,75 < arvo < 2,15
Epäselvä kopioluku	Kaikki muut arvot

Mutation-specific probe(s)

Gene	Mutation	Probe	Mutation status	Additional information
MSH2	10Mbinv	S010444	Absent	MSH2 10 Mb inversion (PMID24114314); This probe consists of three oligos and has two ligation sites.
MSH2	10Mbinv	S010443	Absent	MSH2 10 Mb inversion (PMID24114314); This probe consists of three oligos and has two ligation sites.
MITF	952G>A	S010555	Absent	Mutation also known as p.E318K (rs149617956).
PMS2	SVA-ins	S010670	Absent	PMS2 SVA element insertion (PMID22461402).
CHEK2	1100delC	S010117	Absent	Mutation also known as rs555607708; This probe consists of three oligos and has two ligation sites.

Aberrant region(s) table

Region	Chromosomal band (hg38)	Region size	Region ratio	StdDev	Probes in region	Probe number	Probe function	Mapview (hg38) in kb	Normal copy number	Gene	Exon	Inter ratio	Within normal range	Different from reference population	Ambiguous
RAD51C	17q22	32,1 Kb	1,56	0,04	8	S010224	CN	17-058 692	2	TEX14	01	1,48	TooHigh	Increased2StdDev	False
						S010225	CN	17-058 693	2	RAD51C	01	1,61	TooHigh	Increased2StdDev	False
						S010226	CN	17-058 695	2	RAD51C	02	1,56	TooHigh	Increased2StdDev	False
						S017463	CN	17-058 697	2	RAD51C	03	1,55	TooHigh	Increased2StdDev	False
						S010228	CN	17-058 703	2	RAD51C	04	1,56	TooHigh	Increased2StdDev	False
						S010229	CN	17-058 710	2	RAD51C	05	1,57	TooHigh	Increased2StdDev	False
						S017464	CN	17-058 721	2	RAD51C	06	1,66	TooHigh	Increased2StdDev	False
						S010231	CN	17-058 724	2	RAD51C	07	1,52	TooHigh	Increased2StdDev	False
	17q22	2,7 Kb	0,99	0,01	3	S010232	CN	17-058 733	2	RAD51C	08	0,99	AsExpected	Equal	False
						S010233	CN	17-058 734	2	RAD51C	09	0,99	AsExpected	Equal	False
						S017465	CN	17-058 735	2	RAD51C	down	1,05	AsExpected	Equal	False

Kuva 4: Esimerkki taulukko, jossa on esitetty tarkat tiedot kuvassa 2 esitetyistä poikkeavista koetin arvoista *RAD51C* geenialueella. Koettimien joukkoon sisältyy *TEX14* geenin kopiolukua mittaava koetin. *TEX14* sijaitsee *RAD51C* geenin yläjuoksussa. Kaikki poikkeavat koetin arvot poikkeavat kahden keskihajonnan verran vertaisnäytteistä eli poikkeama on merkittävä. Lisäksi kuvassa näkyy taulukko koetinsarjan mutaatiopesifistä koettimista ja onko niitä havaittu näytteestä. Kyseisestä näytteestä spesifejä mutaatioita ei havaittu.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Taulukko 2: Selitykset "Aberrant region(s) table" eri osioille. **(1)** Mahdolliset arvot: AsExpected (0,8 - 1,2), Low (0,7 - 0,8), TooLow (<0,7), High (1,2 - 1,3) ja TooHigh (>1,3). Suluissa olevat arvot viittaavat koetinarvoihin. **(2)** Mahdolliset arvot: Equal (vertaisnäytteiden mukainen), Decreased1Stdv / Deceased2Stdv (Koetinarvo on yhden tai kahden keskihajonnan verran alempi kuin vertaisnäytteiden koettimien keskiarvo), Increased1Stdv / Increased2Stdv (Koetinarvo on yhden tai kahden keskihajonnan verran korkeampi kuin vertaisnäytteiden koettimien keskiarvo). Yleisesti aina kahden keskihajonnan poikkeamat ovat merkittäviä ja yhden eivät. **(3)** Esimerkki "epäselvästä" tuloksesta on muun muassa, kun koetinarvo selvästi poikkeaa normaalin kopiolumun viitearvoista, mutta on vertaisnäytteiden kopiolumun mukainen.

Region	Tutkittu geenialue
Chromosomal band (hg38)	Geenialueen kromosomaalinen sijainti
Region size	Geenialueen koko
Region ratio	Kaikkien koetinarvojen keskihajonta geenialueella
Stdev	Kaikkien koetinarvojen keskiarvo geenialueella
Probes in region	Geenialueelle kohdistettujen koettimien lukumäärä
Probe number	Lista geenialueen koettimista
Probe function	Koetintyyppi (CN=kopioluku, CTRL=kontrolli, MUT=mutaatio spesifi)
Mapview (hg38) in kb	Koettimen kohde eksonin sijainti geenialueella
Normal copy number	Eksonin normaali kopioluku
Gene	Geeni, johon eksoni kuuluu
Exon	Eksonin numero (voi myös olla esim. eksonia edeltävä introni)
Inter ratio	Koetinarvo
Within normal range	Ilmoittaa asettuuko koetinarvo normaalin kopiolumun viitearvoihin (1)
Different from reference population	Ilmoittaa onko koetinarvo vertaisnäytteiden mukainen vai poikkeava (2)
Ambiguous	Ilmoittaa onko saatu tulos epäselvä (3)

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Huomioitavaa

- ”Inter ratio” viittaa koetinarvoihin, jotka on laskettu jakamalla testattavan näytteen kopiolutukua mittaavan koettimen lukumäärä vastaavan koettimen vertaiskoettimella samasta testinäytteestä. Osamäärä edelleen jaetaan vertaisnäytteen saman kopiolutukua mittaavan koettimen ja vastaavan vertaiskoettimen osamäärällä. Tämä toistetaan jokaisen koettimen kohdalla ja jokainen testinäyte jaetaan, jokaisen vertaisnäytteen kanssa. Lopputuloksena saadaan niin, monta mediaani arvoja, jokaista koetinta kohden kuin vertaisnäyteitä on käytetty. Lopullinen ”Inter ratio” yksittäiselle koettimelle on sen mediaanien keskiarvo.

RATIOS EXCEL-TIEDOSTON TARKASTELU

1. Avaa ”Ratios” Excel-tiedosto tulos kansioista.
2. Tiedosto koostuu viidestä eri välilehdestä:
 - **Inter ratios:** Tämä taulukko sisältää reaktioiden koetinarvot, jokaisen näytteen kaikille kopiolutukumittaaville geenialueille sekä kontrollikoettimille
 - **Reference sample quality:** Tämä taulukko sisältää tiedot vertaisnäytteiden koetinarvoista, mutta lisäksi vertaisnäytteiden keskiarvon, mediaanin, minimin, maksimin ja keskihajonnan jokaisesta koettimesta. Jos vertaiskoettimia ei käytetty reaktioiden valmistuksessa, keskihajonnan (StDev) tilalla on esitetty mediaanien absoluuttinen keskipoikkeama (MAD), joka on verrannollinen keskihajontaan.
 - **Read counts:** Tässä taulukossa on listattu, kuinka monta lukemaa yksittäisestä koettimesta tunnistettiin analyysin aikana jokaisessa näytteessä.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 4.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

- **Intra ratios:** Taulukossa on suhteutetut koetin arvot pohjautuen koettimista laskettuihin lukemiin näytekohtaisesti (kts. Read counts).
- **Software version:** Lista analyysiin sisällytetyistä näytteistä ja millä ohjelman versiolla (Configuration) analyysi suoritettiin. Lisäksi listaukset analyysin ajankohdasta.

Huomioitavaa

- Excel tiedostossa poikkeavat arvot ovat värikoodattu helpottamaan niiden tunnistusta. Arvot ovat sinisen sävyllä, jos ne ovat normaalia kopiolukua suuremmat ja keltaisen sävyllä, jos ne ovat normaalia matalammat. Arvojoukot ovat samat mitä esitetty taulukon 2 yhteydessä kohdassa ”Different from reference population”.
- Jos arvot poikkeavat yhden keskihajonnan verran vertaisnäytteistä, arvo on **korostettu**. Jos arvo poikkeaa kahden keskihajonnan verran eli se poikkeavaa huomattavasti vertaisnäytteistä, arvo on **korostettu ja kursivoitu**

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

DIGITALMLPA MENETELMÄ YHDISTETTYNÄ SALSA® DIGITALMLPA PERINNÖLLISEN SYÖPÄPANEELI 1 KANSSA, VALIDOINTIRAPORTTI

TUTKIMUKSEN TARKOITUS:

Syöpä on genomin sairaus, joka on seurausta patogeenisten mutaatioiden kertymisestä yksilön genomiin. Syöpä voi johtua ympäristötekijöistä ja perinnöllisistä tekijöistä. Näistä jälkimmäinen voidaan yhdistää perinnölliseen syöpäalttiuteen, joka on yleensä seurausta kopiolumuutoksista ituradan solujen kasvunrajoitegeneisissä. Kopiolumuutokset ovat luonteeltaan geenialueiden eksonien kahdentumia tai häviämiä, ilmeten täten mahdollisesti joko geenituotteiden yli- tai alimääränä. Yksilön periessä toiselta vanhemmaltaan kopiolumuutoksen omaavan alleelin hän omaa kasvaneen syöpäriskin riippuen geenistä. Ituradan solujen geenien kopiolumuutosten havaitsemista varten on kehitetty molekyylogeniikan menetelmiä, joissa hyödynnetään homologista pariutumista tunnetuille geenialueille valmistetuilla koettimilla. Multipleksiligaatiosta riippuva koettimien monistus (MLPA) on yksi tällainen menetelmä, joka on laajasti käytössä. Menetelmää varten on suunniteltu useita eri koetinsarjoja perustuen eri syöville altistavien kopiolumuutosten tunnistamiseen. Koettimien kanssa hybridisoituneet ja monistetut geenialueet tunnistetaan kapillaarisekvensoinnilla. Menetelmä on hyvä kopiolumuutosten tunnistuksessa, mutta siitä on kehitetty massasekvensointia hyödyntävä tehokkaampi versio, digitalMLPA. Massasekvensoinnin kautta sekvensoitavien geenialueiden määrä kasvaa huomattavasti, joka tarkoittaa mahdollisuutta sisältää enemmän kohdennettuja koettimia yksittäiseen reaktioon. Tämä johtaa kattavampiin syöville altistavien kopiolumuutosten seulontaan vähemmällä työmäärällä kuin MLPA:lla.

Tässä validoinnissa selvitettiin digitalMLPA-menetelmän luotettavuutta patogeenisten kopiolumukujen tunnistuksessa kaupallisten SALSA® digitalMLPA perinnöllisen syöpäpaneelin 1 koetinsarjoilla D001-B1 ja D001-C1. Menetelmä ja sille suunnitellut koetinsarjat ovat validointiajankohtana RUO (Research use only). Aineistona käytettiin

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

kliinisiä näytteitä TYKS Genomiikan vastualueen arkistosta. Validoinnin odotuksena on, että menetelmällä onnistutaan tunnistamaan kaikki näytteet ennalta odotetuiksi.

VALIDOINTISUUNNITELMA

Validoinnin suunnittelu tehtiin pro gradu -opiskelijan Ville Raappanan ja sairaalageneetikkojen Pia Pohjola ja Minna Toivosen toimesta. Reaktioiden valmistuksen suoritti Ville Raappana menetelmän valmistajan ohjeiden pohjalta Genetics cancer predisposition -tutkimusryhmän tiloissa (PI: Johanna Schleutker). Reaktioiden sekvensointi suoritettiin TYKS Genomiikan vastualueen tiloissa. Ville Raappana suoritti sekvensointidatan pohjalta tulosanalyysit digitalMLPA Coffalyzer-ohjelmalla.

ULKOINEN LAADUNTARKKAILU

Toistaiseksi digitalMLPA-menetelmälle ei ole suoraan kohdennettua ulkoista laaduntarkkailua. Tilanteen kehittymistä tulee seurata ja harkinnan mukaan menetelmällä tulee osallistua sellaisiin ulkoisen laaduntarkkailun kierroksiin, joihin se soveltuu, esimerkiksi NGS-tutkimusten lisänä.

TUTKIMUSMENETELMÄN KUVAUS JA LAATUVAATIMUKSET:

Menetelmä perustuu käytettyjen koetinsarjojen alukkeiden kanssa hybridisoituneiden geenialueiden PCR-monistukseen. Amplikonien pohjalta valmistetaan DNA-kirjasto massasekvensointia varten, josta saadun FASTQ-tiedoston pohjalta suoritetaan näytteiden analyysi. Analyysi suoritetaan digitalMLPA Coffalyzer-ohjelmalla, joka on suunniteltu digitalMLPA-näytteiden analysointiin. Ohjelma laskee jokaiselle näytteen geenialueen eksonille, jotka ovat olleet tutkinnan alla oman koetinarvon, joka kuvastaa geenialueen kopioluvun tilaa.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Laatuvaatimuksina reaktioiden luotettavaa tulkintaa varten on: riittävä määrä sekvenssilukemia ja mediaani arvoja, jokaista näytettä kohden (noin 600 koettimensarjalla noin 240 000 sekvenssilukemaa ja 400 mediaanilukemaa), vesikontrollissa ei saa ilmetä huomattavaa määrää DNA:ta eikä tätä kautta koettimien mediaaniarvoja. Lisäksi luotettavaa analyysia varten kaikkien reaktiosarjojen vertaisnäytteiden testikoettimien keskijakauma täytyy olla $\leq 0,10$. Lopuksi, digitalMLPA Coffalyzer-ohjelman antamissa analyysiraporteissa ei saa ilmetä virheilmoituksia, jotka viitaisivat virheisiin reaktioiden valmistuksessa.

TULOS JA SEN TULKINTA:

Tulokset saadaan laskemalla digitalMLPA Coffalyzer-ohjelmalla näytekohtaiset koetinarvot. Tulokset tulkitaan seuraavien määritelmien pohjalta:

Kopioluvun tila:

Normaali: $0,8 < \text{arvo} < 1,20$

Homotsygootti häviämä: $\text{arvo} = 0$

Heterotsygootti häviämä: $0,4 < \text{arvo} < 0,65$

Heterotsygootti kahdentuma $1,3 < \text{arvo} < 1,65$

Heterotsygootti kolminkertaistuma / homotsygootti kahdentuma: $1,75 < \text{arvo} < 2,15$

Epäselvä koetinarvo: kaikki muut arvot

Jos koetinarvo poikkeaa kahden keskihajonnan verran vertaisnäytteiden arvoista, tulos on merkitsevä. Koetinsarjoihin sisältyy mutaatio spesifisiä koettimia, joista tulee ilmoitus, jos niitä on hybridisoitunut DNA-näytteeseen.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

VALIDOINTITESTAUKSEN TULOS:

Menetelmällä onnistuttiin todentamaan kaikista kopiolumuutosnäytteistä, paitsi LERA-näytteestä tutkinnan kohteena olevat kopiolumuutokset. LERA-näytteen osalta osa tutkittavan *SMAD4*-geenin eksonien kopiolumut olivat normaalin kopiolumun viitearvojen mukaiset. Käytetyn koetinsarjan valmistajan ohjeistuksen mukaan tilanne selittyy prosessoidun pseudogeenin kautta. Tämän myötä geenialueen eksonien 2, 5, 6, 9, 10, 11 ja 12 kopiolumut viittaavat heterotsygoottiseen kahdentumaan samalla, kun geenin muiden eksonien koetinarvot ovat normaalin viitearvojen sisällä. Prosessoitu pseudogeeni on todennäköisesti kliinisesti merkityksetön. Muu kirjallisuus viittaa myös mahdolliseen *SMAD4* prosessoituun pseudogeeniin, jonka myötä kyseiset geenialueen eksonit ovat monistuneet (Mancini-Dinaro ym. 2015; Millson ym. 2015). Prosessoidun pseudogeenin myötä näytteen massasekvensointianalyysi voi antaa kuvan koko geenin kahdentumasta. Kaikkien kopiolumuutosnäytteiden tutkittavien geenialueiden eksonien koetinarvot ovat havainnollistettu kuvassa 1A.

Normaalinäytteiden osalta, näytteiden joukosta jouduttiin pois sulkemaan näytteet LASU ja KUPE. LASU-näyte pois suljettiin, kun selvisi, ettei henkilö ollut antanut lupaa käyttää hänen geeninäytettään anonymisissa tutkimustyössä. KUPE-näyte poissuljettiin, koska työn aikana selvisi, ettei potilasnäyte ollut peräisin terveistä ituradan soluista. Lopuista 18 reaktiosarjoissa 3 ja 4 käytetyistä näytteistä, jotka olivat määritelty normaaleiksi käytettyjen digitalMLPA-koetinsarjojen kohdegeenien eksonien osalta, ei havaittu poikkeavia kopiolumuutoksia perustuen testikoettimien arvoihin. Normaalinäyte ALPI:ssa, jota käytettiin verrokinäytteenä ensimmäisessä ja toisessa reaktiosarjassa oli toisen reaktiosarjan pohjalta saadun Coffalyzer-ohjelman tulosteen mukaan lievästi koholla olevia koetinarvoja. Kun näyte testattiin uudestaan testinäytteeksi määriteltynä kolmannessa reaktiosarjassa, koettimien arvot olivat normaalin kopiolumun mukaiset. Kolmannen reaktiosarjan LEE-näytteestä löydettiin *CHEK2*-geenin 1100delC-mutaatio (rs555607708) mutaatiopesifisten koettimien kautta, joka ei ollut tiedossa ennen validointia. Kaikkien normaalinäytteiden kaikkien geenialueiden koetinarvot ovat havainnollistettu kuvassa 1B.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Luotettavaa datan normalisointia ja tätä kautta luotettavien testikoettimien arvojen saantia varten reaktiosarjojen vertaisnäytteiden testikoettimien arvojen keskihajonta täytyi olla $\leq 0,10$. Toisesta ja neljännessä reaktiosarjasta neljännessä tämä vaatimus täyttyi vertaisnäytteiden osalta, mutta toisessa ei. Toisessa reaktiosarjassa niiden testikoettimien arvojen keskihajonta oli yli $0,10$, joissa esiintyi lievästi kohonneita testikoettimien arvoja ALPI-näytteessä (Liite 2.). Lisäksi keskihajonnat olivat lievästi koholla kolmella eri geenialueella yksittäisillä testikoettimilla. Näiden eksonien osalta, niiden testikoettimien arvot vaihtelivat normaalin kopioluvun viitearvojen ylärajassa $1,18$ ja $1,19$ välillä ALPI-näytteessä (Liite 2.). Miksi ALPI-näytteen koetinarvot poikkesivat toisessa reaktiosarjassa ja olivat kolmannessa reaktiosarjassa $\leq 0,10$, johtui todennäköisesti toisen reaktiosarjan vähäisestä verrokinäytteiden lukumäärästä.

Kolmannessa reaktiosarjassa, jossa ei käytetty määritettyjä vertaisnäytteitä, vaan datan normalisointi suoritettiin kaikkien näytteiden koetinarvojen pohjalta, kaikkien testi- ja vertaiskoettimien mediaanin absoluuttiseksi poikkeamaksi (MAD) saatiin alle $0,10$. MAD-arvot vaihtelivat $0,01$ ja $0,07$ välillä.

Onnistuneiden reaktiosarjojen 2–4 pohjalta suoritettiin χ^2 -testi selvitykseksi, onko oletettujen potilastietojen ja saatujen tulosten välillä tilastollista eroa. Testin nollahypoteesi oli, että saadut tulokset vastaavat potilastietoja. Testissä huomioitiin, että näytteet AHKA, ALPI ja ANLI testattiin kahdesti ja näytteet KUPE ja LASU poissuljettiin tuloksista, eli lopullinen otanta oli 12 positiivista ja 18 negatiivista näytettä. Käytetyllä otannalla testin tulos oli $\chi^2(1) = 0$, $p = 1$, eli potilastiedot ja saadut tulokset vastasivat toisiaan.

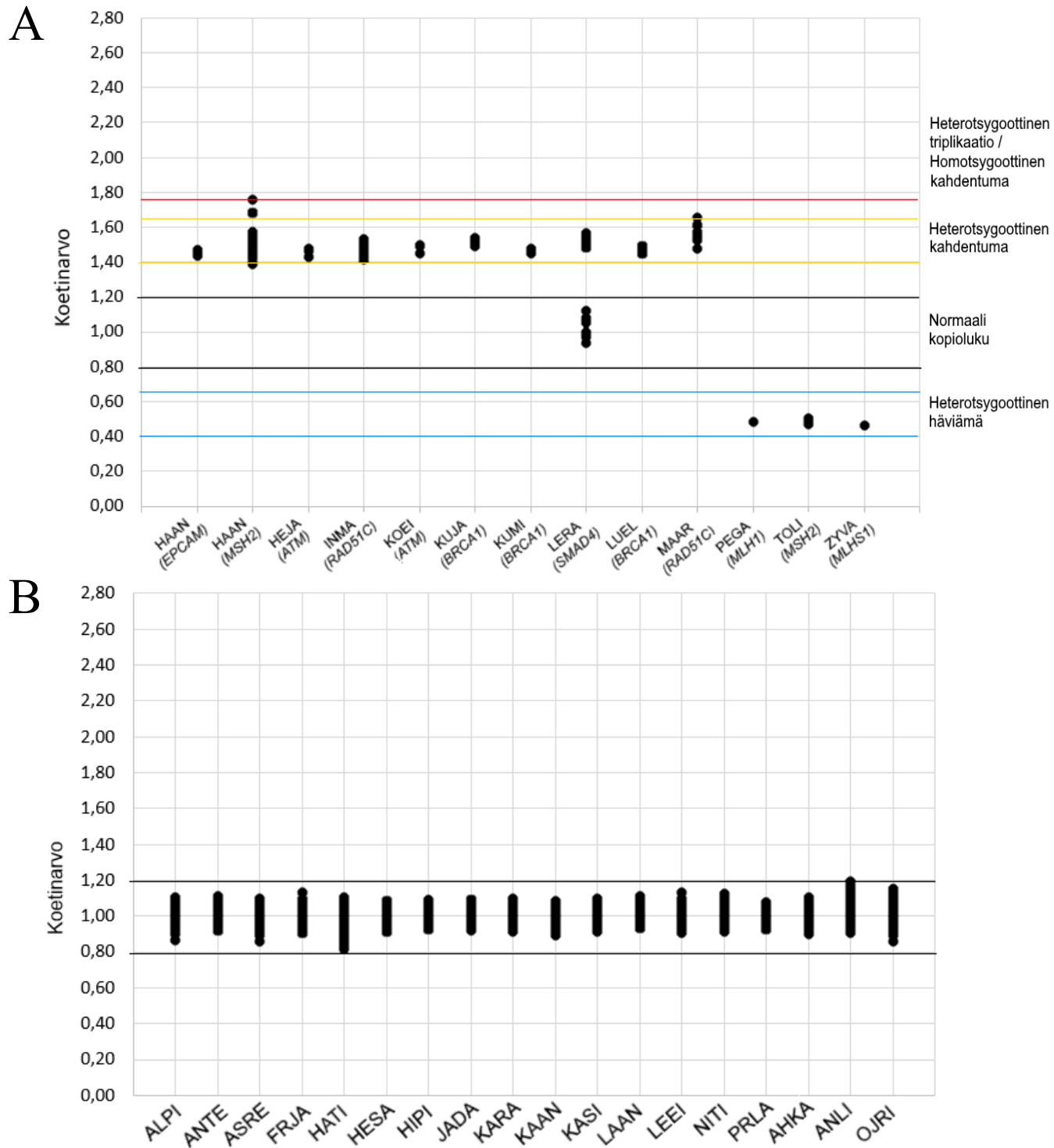
digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka



Kuva 2: (A) Kopiolumuutostenäytteiden tutkittavien eksonien koetinrvo. Kuvaajaan rajatut alueet esittävät eri kopiolumun tiloja osoittavia viitearvoja. (B) Normaalinäytteiden kaikkien geenialueiden eksonien koetinrvo. Molemmat kuvaajat tehtiin Microsoft Excel-ohjelmalla.

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

JOHTOPÄÄTÖS

Validoinnin perusteella, menetelmä yhdistettynä SALSA® digitalMLPA perinnöllisen syöpäpaneeli 1 kanssa, on yhtä luotettava kopiolumuutosten tunnistuksessa kuin ne menetelmät, joilla näytteet olivat jo analysoitu TYKS Genomiikalla. Täten voidaan todeta, että menetelmä soveltuu diagnostiseen käyttöön ja voidaan ottaa rutiinitestikäyttöön. Kun valmistajan virallinen IVD-versio tulee saataville, siirrytään käyttämään sitä.

LÄHTEET

Mancini-Dinaro D., Judkins T., Elias M. C. Daniels C., Marshall A., Holladay J., Qian Y. R. & Roa B. (2015). Myriad Genetics Laboratories, Inc., Salt Lake City, UT. Dosage analysis by next generation sequencing and microarray CGH indicates putative processed pseudogenes in *SMAD4* and *NBN*. ACMG.

https://myriad-web.s3.amazonaws.com/ACMG2015%20Pseudogenes_Mancini-Dinardo_Presented_26MAR2015.pdf

Millson A., Lewis T., Pesaran T., Salvador D., Gillespie K., Gau C. L., Pont-Kingdom G., Lyon E. & Bayrak-Toydemir P. (2015). Processed pseudogene confounding deletion/duplication assays for *SMAD4*. *The Journal of Molecular Diagnostics*, 17, 576–582. <https://doi.org/10.1016/j.jmoldx.2015.05.005>

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

LIITTEET

Liite 1: (A) Validoinnissa käytetyt näytteet. Normaalinäytteiden kohdalle kohtiin Geenit ja Eksoni(t) (Kopiolukumuutos on merkitty ”-”). (B) Valmistetut reaktiosarjat.

A

Näyte (Sukupuoli)	Geeni(t)	Eksoni(t) (Kopiolukumuutos)
HAAN (N)	<i>EPCAM</i>	7-9 (Heterotsyg. kahdentuma)
	<i>MSH2</i>	1-16 (Heterotsyg. kahdentuma)
HEJA (N)	<i>ATM</i>	62 & 63 (Heterotsyg. kahdentuma)
INMA (N)	<i>RAD51C</i>	1-7 (Heterotsyg. kahdentuma)
KOEI (N)	<i>ATM</i>	62 & 63 (Heterotsyg. kahdentuma)
KUJA (N)	<i>BRCAl</i>	13 (Heterotsyg. kahdentuma)
KUMI (N)	<i>BRCAl</i>	13 (Heterotsyg. kahdentuma)
LERÄ (N)	<i>SMAD4</i>	1-12 (Heterotsyg. kahdentuma, pseudogeeni)
LUEL (N)	<i>BRCAl</i>	13 (Heterotsyg. kahdentuma)
MAAR (M)	<i>RAD51C</i>	1-7 (Heterotsyg. kahdentuma)
PEGA (N)	<i>MLH1</i>	16 (Heterotsyg. häviämä)
TOLI (N)	<i>MSH2</i>	9-11 (Heterotsyg. häviämä)
ZYVA (N)	<i>MLH1</i>	16 (Heterotsyg. häviämä)
AHKA (N)	-	-
ALPI (N)	-	-
ANLI (N)	-	-
ANTE (M)	-	-
ASRE (N)	-	-
FRJA (N)	-	-
HATI (N)	-	-
HESA (N)	-	-
HIPI (N)	-	-
JADA (M)	-	-
KARA (M)	-	-
KAAN (N)	-	-
KASI (N)	-	-
KUPE (M)	-	-
LAAN (N)	-	-
LASU (N)	-	-
LEEI (N)	-	-
NITI (M)	-	-
OJRI (N)	-	-
PRLA (N)	-	-

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

B

Reaktiosarja	Näytteet (DNA-pitoisuus, ng/μl)
1	1. HAAN (13,4)
	2. HEJA (20,45)
	3. INMA (19,05)
	4. KOEI (17)
	5. AHKA (23,4)
	6. ALPI (15,1)
	7. ANLI (15,9)
	8. Negatiivinen kontrolli (-)
2	Samat kuin reaktiosarja 1
3	1. ALPI (9,1)
	2. ANTE (11,35)
	3. ASRE (9,92)
	4. FRJA (9)
	5. HATI (10,53)
	6. HESA (9,53)
	7. HIPI (9,7)
	8. JADA (10,07)
	9. KARA (11,5)
	10. KAAAN (9,49)
	11. KASI (9,4)
	12. LASU (10,3)
	13. LAAN (10,23)
	14. LEEI (9,13)
	15. NITI (9,1)
	16. PRLA (11,61)
	17. Negatiivinen kontrolli (-)
4	1. KUJA (9,78)
	2. KUMI (9,27)
	3. LERA (10,3)
	4. LUEL (9,03)
	5. PEGA (9,69)
	6. TOLI (10,04)
	7. ZYVA (9,83)
	8. AHKA (9,1)
	9. ANLI (9,39)
	10. KUPE (9)
	11. MAAR (10,45)
	12. OJRI (9,85)
	13. Negatiivinen kontrolli (-)

digitalMLPA, validointiraportti (Ville Raappana)

Liite 5.

TYKS Genomiikka

Lääketieteellinen genetiikka

DNA-diagnostiikka

Liite 2: Toisen reaktiosarjan ALPI-näytteen eri geenialueiden eksonit joiden testikoettimien keskihajonta poikkesi odotetusta $\leq 0,10$. Taulukossa σ = keskihajonta

Geenialue	Eksoni (Koetin ,tyyppi)	Koettimen σ	Koettimen arvo
<i>APC</i>	18 (S017561, Testi)	0,15	1,23
<i>BMPRIA</i>	12 (S017494, Verrokki)	0,16	1,25
<i>PALB2</i>	8 (S010380, Verrokki)	0,12	1,18
<i>BRCA1</i>	18 (S017434, Verrokki)	0,13	1,19
<i>SMAD4</i>	Promoottori alue A eksoni 1 yläjuoksussa (S011008, Verrokki)	0,17	1,26
	12 (S010332, Testi)	0,12	1,18