

Janika Sirjala

# Yksilöntunnistus suoraan verestä tehtävällä polymeerasiketjureaktiolla genotyypin määrityksen laadunvarmennukseen

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Bio- ja kemiantekniikka

Insinöörityö

3.5.2017

<p>Tekijä Otsikko</p> <p>Sivumäärä Aika</p>	<p>Janika Sirjala Yksilöntunnistus suoraan verestä tehtävällä polymeeraasiketjureaktiolla genotyyppimäärityksen laadunvarmennukseen</p> <p>36 sivua + 2 liitettä 3.5.2017</p>
Tutkinto	Insinööri (AMK)
Koulutusohjelma	Bio- ja kemiantekniikka
Suuntautumisvaihtoehto	Bio- ja elintarviketekniikka
Ohjaajat	Vanhempi tutkija Päivi Lahermo Tutkintovastaava Jarmo Palm
<p>Yksilöntunnistusta käytetään esimerkiksi rikostutkinnassa ja isyystesteissä sekä elinsiirtoelinten luovuttajia etsiessä. Tässä työssä oli tarkoitus optimoida rutiinitoimeksi tulevaa yksilöntunnistusmenetelmää yksilön ja tästä genotyyppimääritykseen saapuvan näytteen yhteneväisyyden varmentamiseen. Kehitettävän yksilöntunnistusmenetelmän on tarkoitus toimia genotyyppitulosten laaduntarkkailun välineenä, kun muita luotettavia varmennuskeinoja ei ole. Näytteenkäsittelyssä on aina riski näytteen tai näytekirjausten sekaantumiseen, jolloin yksilön genotyyppi ei vastaa todellisuudessa yksilöä. Tämä saattaa johtaa väärin tulkintoihin henkilön terveydestä tai häiritä tuloksiin perustuvien tutkimusten onnistumista.</p> <p>Polymeeraasiketjureaktio (PCR) tehdään yleensä eristetylle DNA:lle. Reaktio vaatii tarkkaa optimointia muun muassa monistettavan DNA-jakson pituuden, alukkeiden määrän sekä kiinnittymiskohdan suhteen. Myös mahdolliset epäpuhtaudet häiritsevät tai voivat estää reaktiota tapahtumasta. Suoraan verestä tehtävän PCR:n haasteita ovat arvaamattomat reaktio-olosuhteet veren sisältämien soluaineiden, valkuaisaineiden ja muiden epäpuhtauksien takia. Suoraan verinäytteestä puskuriin liuotetun näytteen DNA-konsentraatiota ei myöskään tunneta.</p> <p>Tämän insinööriyön tarkoituksena oli kehittää Suomen molekyyliääketieteen instituutin (FIMM) teknologiakeskukselle genotyyppityslaboratorioon yksilöntunnistusmenetelmää suoraan verestä tehtävän PCR:n avulla. Menetelmän kehitys on aloitettu aikaisemmin tutkija Päivi Lahermon johdolla. Tässä työssä oli tarkoituksena optimoida PCR-reaktioon tarvittavan näytteen määrää ja käsittelytapoja sekä käytettyjen markkerien toimivuutta. Lisäksi arvioidaan tilastollisesti yksilöntunnistamiseen vaadittavien SNP (Single Nucleotide Polymorphism)-markkerien lukumäärää.</p> <p>Tässä työssä käytettiin 27 SNP-markkeria yhteensä 46 näytteelle kolmessa eri ajossa. Näytetilavuutta saatiin optimoinnilla parempaan suuntaan ja näytesäilytykseen kehitettiin uusi tapa. Yksilöntunnistusmenetelmää saatiin kehitettyä toivottuun suuntaan ja menetelmää voidaan jatkossa käyttää yksilöntunnistamiseen insinööriyössä testatuilla markkerien lukumäärillä.</p>	
Avainsanat	Genotyyppitys, SNP, yksilöntunnistus, PCR suoraan verestä

Author Title	Janika Sirjala Individual identification by direct whole blood polymerase chain reaction for genotyping quality control
Number of Pages Date	36 pages + 2 appendices 3 May 2017
Degree	Bachelor of Engineering
Degree Programme	Biotechnology and Chemical Engineering
Specialisation option	Biotechnology and Food Engineering
Instructors	Päivi Lahermo, Senior Researcher Jarmo Palm, Head of Degree Programme
<p>Individual identification is used, for instance, in criminal investigation, in paternity tests and in the search for an organ transplant donor. In this thesis, the purpose was to optimize a method for verifying the identity of project samples by direct whole blood polymerase chain reaction. The developed method is supposed to become a quality-control tool for a genotyping laboratory when there are no other suitable verification methods available. It is always a possibility that samples get mixed up during sample handling or genotyping and that an individual's records and genotypes, therefore, do not match. This leads to incorrect results and to false interpretations of an individual's health or can interfere with the success of the research based on the results.</p> <p>Polymerase chain reaction (PCR) is usually made to an isolated DNA sample. The reaction requires careful optimization of, for example, the length of an amplified DNA fragment and concentrations or sequences of the primers. The possible contaminations in the sample can also prevent the reaction from occurring. Additional challenges of direct whole-blood PCR are unpredictable reaction circumstances due to cell material, proteins and other contaminants contained in the blood sample. DNA concentration of unprocessed blood samples dissolved in buffer are also not known.</p> <p>The aim of this thesis was to develop an individual identification method that uses whole blood PCR for the Genotyping Unit in Technology Centre of the Institute for Molecular Medicine Finland (FIMM). The development of this method was originally started under the leadership of Senior Researcher Päivi Lahermo. This thesis aimed to optimize the sample volume needed for PCR, the treatment methods of the samples and to test the reliable functioning of the used markers. In addition, the number of SNP (Single Nucleotide Polymorphism) markers required for identifying an individual were statistically estimated.</p> <p>In this work 27 SNP markers were used to genotype 46 individuals in 3 different runs. The sample volume was optimized, and a new method for storing the samples was devised. The identification method was developed into the desired direction by this thesis, and the method can be used for identification of the genotyped individuals with the tested numbers of markers.</p>	
Keywords	Genotyping, SNP, individual identification, whole blood direct PCR

# Sisällys

## Lyhenteet

1	Johdanto	1
2	Genomitiedon käyttö yksilöntunnistuksessa	2
2.1	Yksilöllinen geeniperimä	3
2.2	Yhden nukleotidin monimuotoisuus	5
2.3	Genotyyppitys ja genomin variaatiot	7
2.4	Genomitutkimukset	8
3	Näytteet	10
3.1	GeneRisk	10
3.2	Kokoverinäytteet	11
3.3	Genotyyppimäärityksen laadunvalvonta	12
4	Menetelmät	13
4.1	Markkerien valinta	13
4.2	Usean alukkeen SNP-genotyyppitys	14
4.3	Suoraan verestä tehtävä PCR	16
4.4	MALDI-TOF MS	17
4.5	PCR ja ekstensio	18
5	Työn toteutus	20
5.1	Näytteenkäsittely	20
5.2	Työn kulku	21
5.3	Tulosten analysointi ja laaduntarkkailu	22
5.4	Tilastolliset analyysit	26
6	Optimointi	27
6.1	Näytteensäilytysmenetelmät	27
6.2	Näytemäärä ja näytekonsentraatio	28
6.3	Toistettavuus ja näytteen säilyvyys	29
7	Tulokset ja tulosten tarkastelu	29

8	Yhteenveto	32
	Lähteet	35
	Liitteet	
	Liite 1. PCR-alukkeet	
	Liite 2. Ekstensioalukkeet	

## Lyhenteet

**Alleeli** Allele. Vastingeeni, vaihtoehtoiset geenin muodot samassa kromosomin kohdassa.

**Aлке** Primer. PCR-reaktiossa käytettävä oligonukleotidi, joka tarttuu denaturoituun yksijuosteiseen DNA:han. DNA-monistuksen aloituskohta.

## MALDI-TOF MS

Matrix Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry. Matriisiavusteinen laser-desorptio/-ionisaatiolentoaikamassaspektrometri.

**PCR** Polymerase chain reaction. Polymeraasiketjureaktio.

**SNP** Single Nucleotide Polymorphism. Yhden nukleotidin polymorfia. Puhekielessä vakiintunut termiksi "snippi".

## 1 Johdanto

Yksilöntunnistusta käytetään näytteen ja yksilön välisen yhteyden selvittämiseen. Tässä työssä verrataan tutkittavan yksilön genotyypityslaboratorioon saapuvan DNA-näytteen genotyyppimääritystulosta suoraan verinäytteestä polymeraasiketjureaktion (PCR, Polymerase Chain Reaction) avulla saatavaan tulokseen. Genotyyppimääritykset voidaan toteuttaa erilaisin menetelmin ja tässä työssä verrataan yhden emäksen polymorfioita (SNP, Single Nucleotide Polymorphism) kahden eri laitevalmistajan kehittämällä menetelmillä. DNA-näytteistä on määritetty noin 551 000 markkeria koko genomin laajuudelta ja yksilöntunnistamiseen käytettävästä verinäytteestä määritetään näistä 27 SNP-markkeria.

Näytteen kulku tutkittavilta henkilöiltä tulosten saapumiseen tutkijalle on moniportainen ketju. Henkilöistä otetut näytteet, usein verinäyte, kirjataan ja lähetetään laboratorioon, jossa näytteestä eristetään DNA. Tämän ketjun jokaisessa vaiheessa useat henkilöt käsittelevät näytteitä, jolloin on aina olemassa riski näytteiden sekaantumiselle tai kontaminaatiolle. Myös genotyyppimäärittelyn aikana on riski näytteiden sekaantumiselle, genotyyppitulosten virheille ja virheelliseen tulosten analysointiin. Tähän kaivataan luotettavaa, yksinkertaista ja edullista laadunvarmistusmenetelmää varmentamaan näytteen ja yksilön yhteneväisyys. Menetelmä on tarkoitus saada rutiinitoimeksi laadunvalvontaan genotyyppien määrittelyn yhteyteen.

Normaaliin eristettyyn DNA-näytteeseen verrattuna suoraan verinäytteestä tehtävän polymeraasiketjureaktion (PCR) haasteena on näytteiden sisältämät epäpuhtaudet. Verinäyte sisältää huomattavia määriä erilaisia komponentteja, muun muassa soluainesta ja valkuaisaineita, jotka voivat inhiboida tai estää kokonaan PCR-reaktion onnistumisen [1]. Lisähaasteena käytettävässä menetelmässä on myös usean alukkeen PCR:n toteuttaminen yhtäaikaaisesti samalle näytteelle. PCR:ssä monistetaan SNP-markkereita ympäröiviä sekvenssialueita valituista kohdista genomissa ja SNP:t tunnistetaan käyttäen spesifisiä ekstensioalukkeita. Ekstensiotuotteet erotellaan toisistaan kokonsa perusteella matriisiavusteisella laser-desorptio/-ionisaatiolentoaikamassaspektrometrilla (Matrix Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry, MALDI-TOF MS).

Menetelmän testausta on aloitettu vanhempi tutkija Päivi Lahermon johdolla FIMM:in teknologiakeskuksen genotyypitysyksikössä onnistuneesti Agena Bioscience (San

Diego, CA) Single Nucleotide Polymorphism Detection with iPLEX® ja MassARRAY® -menetelmän avulla. Menetelmään on ennen tämän työn alkamista etsitty sopivia SNP-markkereita ja testattu niiden eri lukumäärien vaikutusta. Aikaisemmissa testeissä oli kehitetty myös erilaisia entsyymejä sekä näytetilavuuksia. Menetelmään on saatu lupaavia tuloksia, ja sitä halutaan kehittää edelleen tulosten luotettavuuden ja työskentelyn kannalta optimaalisempaan suuntaan.

Insinööriä suoritettiin Helsingin yliopiston Suomen molekyyli lääketieteen instituutin, Institute of Molecular Medicine Finland (FIMM), Teknologiakeskukselle. Ohjaajina toimivat vanhempi tutkija Päivi Lahermo sekä tutkintovastaava Jarmo Palm.

## **2 Genomitiedon käyttö yksilöntunnistuksessa**

Ihmisen perimä eli genomi on sukupolvelta toiselle periytyvää informaatiota pakattuna solun tumaan kromosomeiksi. Kromosomit muodostuvat tiiviisti pakatusta nukleiinihappopoketjasta sekä siihen kiinnittyneistä proteiineista ja muista molekyyleistä. Nukleiinihappopoketjun sokeri-fosfaattirankaan kiinnittyneiden emäksisten sivuhaarojen järjestys on jokaisella yksilöllä uniikki. Tämä emäsjärjestys määrää yksilön perinnölliset ilmenevät ominaisuudet sekä piilevät ominaisuudet. Emäsjärjestyksen ja siten perimän muutoksia kutsutaan mutaatioksi. Mutaatioita on montaa eri tyyppiä, ja niiden vaikutukset ihmiseen riippuvat mutaation laadusta sekä sen kohdasta genomissa ja nukleiinihappopoketjussa. Monilla mutaatioilla ei ole välttämättä suoraa yhteyttä ihmisen terveyteen, kun taas joillain on osoitettu olevan välitön vaikutus sairastumiseen. [2, s. 44–45; 3, s. 22, 57–59.]

Geenitutkimukset ovat edenneet viime vuosina huimaa vauhtia, ja ihmisen genomista voidaan nykyään etsiä muun muassa tiettyihin sairauksiin altistavia geenejä ja mutaatioita, selvittää monitekijäisten tautien syntyä genomilaajuisten tutkimusten avulla ja tutkia henkilöiden lääkeainemetaboliaa sekä väestöjen historiaa ja rakennetta [3, s.13]. Yksilöiden genomit ovat huomattavan samanlaisia, jopa noin 99,9-prosenttisesti, mutta koko genomien laajuuden takia yksilöllinen ero DNA-sekvenssissä on huomattavaa [4; 5]. Tästä kuitenkin poikkeavat identtiset kaksoset, joiden lähtökohtana on yksi ja sama munasolu ja identtinen genomi. Uusimpien tutkimusten mukaan myös identtisten kaksosten perimä poikkeaa hieman toisistaan. Eroavuudet voivat syntyä alkionkehityksen aikana ja syntymän jälkeen tapahtuvien mutaatioiden tai epigeneettisen muuntelun ansiosta.

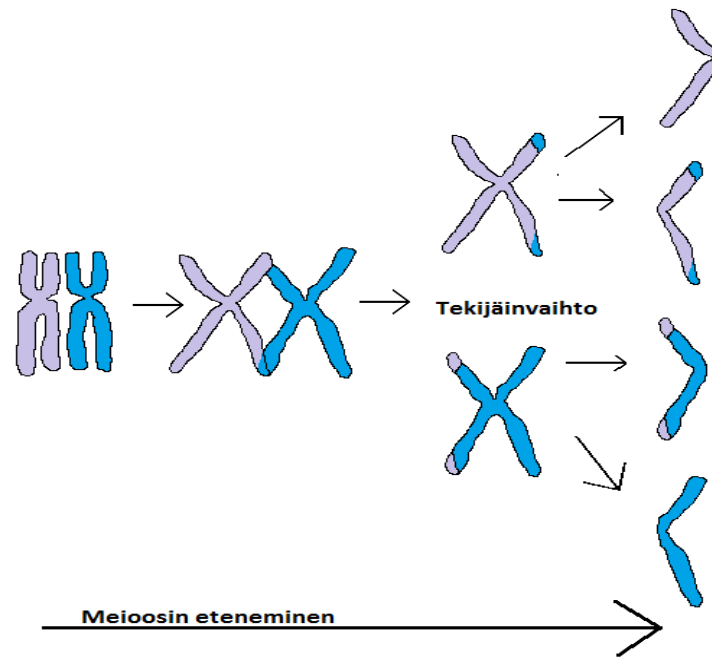


Epigeneettinen muuntelu ei vaikuta DNA:n rakenteeseen, mutta vaikuttaa muun muassa geenien säätelyyn sekä toimintaan. [3, s. 81; 6.]

Genomin ja nukleiinihappojen emäsjärjestyksen avulla voidaan tunnistaa yksilö ja erottaa toisesta yksilöstä suurella varmuudella. Yksilöntunnistukseen tarvitaan aina henkilöstä otettu näyte sekä näyte, johon henkilöä verrataan. Näytteiden vertaamiseen on useita erilaisia menetelmiä genetiikan keinoin. Tässä työssä vertaillaan genotyyppimääritykseen saapuvien tutkimusnäytteiden genotyyppitulosta saman yksilön verinäytteestä saatuun tulokseen. Genotyyppitys perustuu SNP (Single Nucleotide Polymorphism, yhden nukleotidin polymorfia) -merkkeihin genomissa eli markkereihin ja niiden avulla yksilön profilointiin. Näillä markkerilla tarkoitetaan eroja DNA:n välillä tietyn emäksen kohdalla ja niitä käytetään tutkiessa genomin eroavaisuuksia yksilöillä ja populaatioilla. [2, s. 70; 4; 5.]

## 2.1 Yksilöllinen geeniperimä

Periytyvyys noudattaa niin sanottuja Mendelin periytymissääntöjä, vaikka periytyvyys onkin genetiikan avulla todistettu näitä lainalaisuuksia huomattavasti monimutkaisemmiksi. Ihmisellä on 23 kromosomiparia, yhteensä 46 kromosomia, joissa molemmilta vanhemmilta on peritty parin toinen haploidinen kromosomi. Diploidisten kromosomiparien välillä tapahtuu meioosissa uusia kromosomiyhdistelmiä, jossa kromosomiparit vaihtavat geneettistä materiaalia tekijäinvaihdossa (kuva 1). Kromosomipareissa tapahtuu keskimäärin 1–2 tekijäinvaihtoa solun jakautuessa. Jotta tekijäinvaihto tai muu muutos genomissa periytyy jälkeläisille, on sen tapahduttava ituradan soluissa, jotka lopulta päätyvät hedelmöittyneeksi sukusoluksi. [2, s. 48, 73; 3, s. 26.]



Kuva 1. Tekijäinvaihto kuvattuna diploidisen kromosomin vastinkappaleiden kesken meioosissa. [7.]

DNA-sekvenssi on ihmisen koko kromosomistossa yhteensä noin 3,3 miljardia emästä kattava ketju. Proteiineja koodaavia geenejä on noin 23 000 ja yhden geenin pituus vaihtelee yleensä 1 000–2 000 emäksen välillä. Genomin kaikkien emästen lukumäärä on niin laaja, että sen ääneen lukemiseen emäs sekunnissa kestäisi yli 100 vuotta. Tämän takia DNA-sekvenssin määrittäminen ja tulkinta vaativat valtavaa tehoa tietokoneilta ja kehittynyt teknologia on auttanut huomattavasti koko genomin laajuisissa tutkimuksissa. [3, s. 20; 4.]

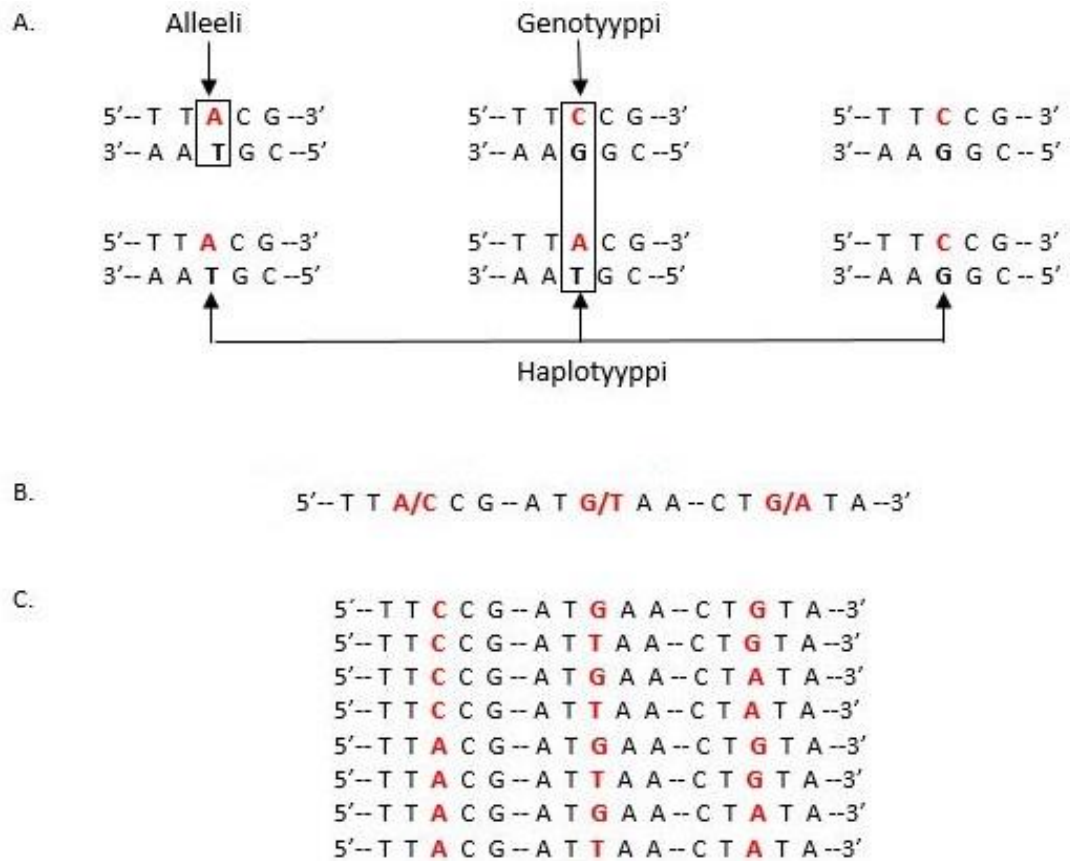
Yksilöntunnistus DNA:ta käyttäen tapahtuu vertaamalla tunnistettavan näytteen DNA-sekvenssiä yksilöstä otetun näytteen sekvenssiin. Koko genomin DNA:n sekvensointi ja niiden vertaaminen toisiinsa on haastavaa ja aikaa vievää. Tämän vuoksi yksilöntunnistamisessa käytetään vain osia sekvenssistä tai tiettyjä kohtia DNA:sta. DNA eristetään yleensä näytteestä esimerkiksi erilaisista kudoksenäytteistä, hiusjuuresta, syljestä tai verestä. DNA voidaan eristää lähes mistä tahansa eläviä tai kuolleita soluja sisältävästä näytteestä. Näytettä riittää yksilöntunnistamiseen hyvin pieni määrä, sillä DNA:ta voidaan monistaa huomattavia määriä PCR:n avulla. Yksilöntunnistusta DNA-sekvenssin avulla käytetään muun muassa rikostutkimuksissa, isyystesteissä ja elinsiirtoluovuttajia etsiessä. [2, s. 70–71; 8; 9.]

## 2.2 Yhden nukleotidin monimuotoisuus

DNA:sta huomattava osa on proteiineja koodaamatonta aluetta, jonka osuus voi olla jopa 98,5 prosenttiyksikköä. Koodaamaton alue sisältää pseudogeenejä, introneita, geenien säätelyjaksoja ja erilaisia toistojaksoja. Pseudogeenit ovat geenien toimimattomia epätäydellisiä kopioita, jotka sijaitsevat eri lokuksissa kuin toimivat geenit. Intronit ovat geenin sisältämiä jaksoja, jotka eivät osallistu proteiinisynteesiin. Säätelyjaksot voimistavat tai vaimentavat geenien toimintaa ja toistojaksoissa emäsjärjestys toistuu tuhansia tai jopa miljoonia kertoja lähes samanlaisena perimässä. Vain murto-osa genomista on aktiivisia geenejä ja RNA:ta koodaavaa aluetta. SNP-markkereita löytyy koko genominalueelta. [3, s.100–102.]

SNP-mutaatiossa yksittäinen emäs DNA-juosteessa vaihtuu toiseen emäkseen (kuva 2). Tämä on ihmisen perimän yleisin geneettisen variaation muoto ja genomissa on karkeasti arvioituna noin 10 miljoonaa SNP-markkeria, keskimäärin 300 nukleotidin välein. SNP:t esiintyvät koko genomissa, minkä ansiosta ne ovat suosittuja geneettisiä markkereita ja niitä käytetään yleisesti genotyyppien määrittämisessä eli genotyyppityksessä. SNP-profiili soveltuu myös yksilön tunnistamiseen normaalin genotyyppityksen yhteydessä vastaavilla ja riittävän useilla SNP-markkereilla. [2, s. 70; 3, s. 100; 4.]

SNP:t periytyvät periytymisen lainalaisuksien mukaan, ja niiden alleelit löytyvät yksilön kromosomistosta molemmista kromosomiparista. Yleensä SNP-alleelit muodostuvat kahden emäksen variaatioista eli ne ovat bialleelisia (Kuva 2). Tämä johtuu mutaatioiden sattumanvaraisuudesta, ja kahden mutaation todennäköisyys saman emäksen kohdalla on hyvin pieni. SNP:ksi voidaan kutsua mutaatiota, jossa emäsvariaatiosta harvinaisempaa esiintyy vähintään yhdellä prosentilla väestöstä. Yhden emäksen polymorfiat ovat yleensä transitiomutaatioita, joissa puriiniemäs vaihtuu toiseen puriiniemäkseen ja pyrimidiiniemäs vaihtuu toiseen. Mutaatiot puriini- ja pyrimidiiniemäksen välillä eli transversiomutaatiot ovat hieman harvinaisempia. [3, s. 100; 4; 5; 8.]



Kuva 2. Kohdassa A. on kuvattuna kolme yhden emäksen polymorfiaa eli SNP (Single Nucleotide Polymorphism)-markkeria komplementaarisessa DNA-juosteessa diploidisessa kromosomissa. DNA-juosteet vastaavat muuten toisiaan, vain SNP:n kohdalla on poikkeavuus emäsjärjestyksessä. Kuvassa on esitetty yhden pistemutaation esimerkivariaatiot yhdellä yksilöllä. Variaatiot ovat ylemmän juosteen 5'-päästä luettuna punaisella merkitty homotsygootti A/A, heterotsygootti C/A ja homotsygootti C/C. Kuvassa esitetään alleelin, genotyypin ja haplotyyppin erot. Alleeli on haploidisen kromosomin valittu tarkasteltava kohta, ja genotyyppi kuvaa tämän kohdan variaatioita diploidisessa kromosomistossa. Haplotyyppi on useat määritetyt alleelivariaatiot yksilöstä. [3, s. 100; 10.]

Kohdassa B. kuvataan kolmen SNP:n variaatiot yksijuosteisessa DNA:ssa. Variaatiot ovat yleensä bialleelisia, eli DNA-juosteen yhdellä emäksellä on kaksi mahdollista muotoa. Nämä variaatiot periytyvät molemmilta vanhemmilta jälkeläisille. [10.]

Kohdassa C. on kuvattu kolmen SNP:n ja kahden alleelin variaatiot. Variaatiot ovat kuvattuna haploidisina eli yhdestä kromosomista DNA:n yhdestä juosteesta. Kolmea bialleelista SNP-markkeria tarkasteltaessa voi siis esiintyä kahdeksaa eri haplotyyppiä. [10.]

Yhden nukleotidin variaatiot genomissa voivat olla täysin merkityksettömiä tai vaikuttaa huomattavasti muun muassa geenien toimintaan, säätelyyn ja niiden ilmiäsuun. SNP:n paikka genomissa määrää pitkälti sen vaikutuksia. Geenialueella ja erityisesti eksoneissa variaatiot voivat vaikuttaa geenin proteiinisynteesiin ja näin valkuaisaineiden

toimintaan, toimimattomuuteen tai niiden määrään. Translaatiossa yhden emäksen muutos kodonissa voi johtaa väärään aminohapposekvenssiin tai proteiinin virheelliseen las-  
kostumiseen. Proteiini voi toimia mutaation jälkeen eri tavoin, kuin alkuperäinen proteiini  
ja sen toiminta voi häiriintyä tai estyä kokonaan. Translaatio voi myös estyä emäsjärjes-  
tyksen muuttuessa. Yhden emäksen mutaatiot voivat siis vaikuttaa myös geenin toimin-  
taan tai sen hiljentämiseen. [3, s. 101; 10.]

Alleelit ovat molemmista kromosomipareista samasta lokuksesta löytyvät variaatiot,  
jotka vaikuttavat samaan perinnölliseen ominaisuuteen (kuva 2). Yksilöillä on geenistä  
tai lokuksesta aina kaksi alleelia, kummaltakin vanhemmalta periytyy yksi alleeli [3, s.  
30]. SNP:n kohdalla saman nukleotidin alleeleja kutsutaan homotsygooteiksi ja vastaa-  
vasti kahden eri nukleotidin alleeleja kutsutaan heterotsygootiksi (kuva 2). Alleelifrek-  
venssi kuvaa alleelien osuuksia tarkasteltavasta populaatiosta, ja ne ovat siitä vahvasti  
riippuvaisia. Tietyn alleelin esiintyvyys voi olla hyvin paikallista maantieteellisesti ja erota  
eri populaatioiden välillä. Toisiaan lähellä olevat SNP:t voivat periytyä jälkeläisille yh-  
dessä toisiinsa kytkeytyneenä [2, s. 73]. Toisiinsa kytkeytyneet SNP:t voivat löytyä sa-  
manlaisina variaatioina väestötasolla, eivätkä kytkeytyneet markkerit sovellu yksilöntun-  
nistukseen, sillä toisen SNP:n alleelista voidaan päätellä toisen alleeli. [3, s. 30, 99, 323.]

SNP-markkerien esiintyvyys ja niiden variaatiot ovat niin huomattavia, että kahden sa-  
manlaisen geneettisen perimän löytymisen on erittäin epätodennäköistä. Vaikka on ky-  
seessä vain yhden emäksen vaihtuminen toiseen, kompensoi markkerien genomilaa-  
juinen lukumäärä niiden merkitystä yksilöiden ja populaatioiden välillä. Yli 10 miljoonan  
SNP:n löytymisen tarkoittaa, että kaksi yksilöä eroaa toisistaan n. miljoonan SNP:n koh-  
dalla. Yhden emäksen variaatioita on huomattavasti helpompi ja edullisempi tutkia, kuin  
koko genomien laajuista DNA-sekvenssiä. Tuotetun datan määrä on myös pienempi, jol-  
loin aineiston käsittely ja tulkinta eivät ole yhtä haastavia toteuttaa. [3, s. 14–15; 4; 10.]

### 2.3 Genotyyppitys ja genomien variaatiot

Yksilön vanhemmilta perityt geenivariaatiot ja genotyyppi muodostavat genomien (kuva  
2). Genotyyppityksellä tarkoitetaan genomien variaatioiden määrittämistä ja vertaamista  
sitä muihin genotyyppisiin tai referenssigenotyyppiin. Yksilön genotyyppi määritetään  
siis alleeleista ja niiden variaatioista. Alleelivariaatioiden määrittäminen voi olla yksittäisten  
emäksien polymorfoiden (SNP) tunnistamista, kuten tässä työssä tai mikrosatelliittien

(STR, Short Tandem Repeat tai SSR, Simple sequence repeat) tai kopiolukuvariaatioiden (CNV, Copy Number Variation) määrittämistä. Genotyyppitys voidaan tehdä lähes koko genomilaajuudelta tietyiltä kohdilta tai valita vain muutama tietty kiinnostuksen kohde genomista. [3, s. 34, 105; 11; 12.]

SNP:t ovat hyvin suosittuja markkereita genotyyppimäärityksissä, sillä niitä esiintyy huomattavasti koko genomien laajuudelta ja niitä löytyy hyvin todennäköisesti useita esimerkiksi tutkittavan geenin alueelta ja ympäriltä. Toinen yleisesti käytetty geneettinen markkeri on mikrosatelliittimarkkeri. Mikrosatelliiteissa ja muissa isommissa toistojaksoissa toistuu tietty emäsjärjestys useita kertoja DNA-juosteessa. Monimutkaisimmissa satelliitti- ja minisatelliittimarkkereissa toistojakso voi olla kymmenistä jopa satoihin nukleotideihin. Yleinen mikrosatelliittien toistojakson pituus on noin 1–10 emästä, joka toistuu samanlaisena noin 10–150 kertaa. Yleisin toistojakso perimässä muodostuu kahdesta emäksestä, usein adeniinin ja sytosiinin (AC) muodostamasta parista. Toistojen lukumäärä vaihtelee yksilöillä ja populaatioilla. Mikrosatelliitteja käytetään muun muassa yksilöntunnistuksessa ja kytöntäanalyysissä. [3, s.102–103; 12.]

Kopioluvun muutos (CNV) syntyy usein solun jakaantumisen aikana epätasapainoisessa tekijäinvaihdossa johtuen muun muassa virheellisestä rekombinaatiosta. Tämä jakso voi sisältää useita geenejä ja olla jopa miljoonien emästen mittainen. Kromosomissa kopioluvun muutos voi esiintyä erilaisina variaatioina. Yksi esimerkki on tandemtyylinen tekijäinvaihto, jossa vastinkromosomista DNA-jakso kopioituu kahdesti peräkkäin toiseen kromosomiin. Myös deleetioita ja duplikaatioita voi tapahtua DNA-jaksoille tai jakson lukusuunta voi kääntyä inversiossa. Muutos voi olla myös toiseen kromosomiin siirtynyt DNA-jakso tai kromosomien välillä vaihtuneet jaksot. [3, s.105–106.]

## 2.4 Genomitutkimukset

Koko genomien kattavat laajat tutkimukset ovat edistäneet valtavasti genomituntemusta ja aineistojen koko on kasvanut huomattavasti viime vuosina. Merkittäviä genomitutkimuksia aloitettiin vuosina 1990–2008 silloin suurille aineistoille. NykYTEKNIKOILLA pystytään käsittelemään vielä huomattavasti suurempia aineistoja ja genomitiedon tulkinta on kehittynyt valtavasti. Vuonna 1990 aloitettiin Human Genome Project, jonka yhtenä suurimpana tuloksena vuonna 2003 saatiin ihmisen koko genomien emäsjärjestys eli sekvenssi selvitettyä. Tätä aikansa todennäköisesti suurinta biologista yhteistyöprojektia on

seurannut useat muut genomitutkimukset ja projektit kuten HapMap ja 1000 Genomes -projektit. [2, s. 15; 3, s. 5, 13, 125; 11.]

Näissä laajamittaisissa viime vuosituhanella alkaneissa tutkimuksissa selvisi huomattava yksilöllisten variaatioiden määrä ihmisen genomissa, mikä on edesauttanut tutkijoita selvittämään variaatioiden merkityksiä ihmisen fenotyyppiin eli ilmiösuun. Pyrkimyksenä näissä tutkimuksissa on muodostaa geneettisen variaation ja sen ilmenemisen välille tilastollinen yhteys ja tutkia geneettisten tekijöiden vaikutusta fenotyyppiin. Merkitysten selvitys on kuitenkin haastavaa, sillä monet geenivariaatiot vaikuttavat monitekijäisesti [3, s. 124]. Geeniperimän ja ympäristötekijöiden yhteisvaikutukset täytyy myös ottaa huomioon tutkittaessa geneettistä sairastumisalttiutta tai muuta geneettistä ilmentymistä. Näiden yhteyksien löytämiseen käytetään geneettisiä markkereita, jolloin genomilaajuisen tieto saadaan tiivistettyä haluttuihin kohtiin kromosomistossa ja DNA-sekvenssissä. [3, s.13–15; 11.]

Suurien aineistojen genomilaajuiset assosiaatiotutkimukset (Genome Wide Association Study, GWAS) pyrkivät selvittämään geeniperimän ja fenotyypin välistä merkitystä yksilötasolla. Geneettisistä assosiaatiotutkimuksista saadaan tietoa genomien vaikutuksesta yksilöiden ominaispiirteisiin tai geenin ilmenemismuotoon. Näissä tutkimuksissa kerätään tietoa muun muassa erilaisiin sairauksiin vaikuttavien geenien ja ympäristötekijöiden välisistä yhteyksistä sekä geenien toiminnan yhteisvaikutuksista. Ominaispiirre ja sen ilmentyminen voi tarkoittaa muun muassa sairastumisriskiä, perinnöllistä sairautta tai jopa oppimisvaikeuksia. Assosiaatioanalyysit tehdään väestöaineistoista ja analyysissä etsitään poikkeamia tai ylliedustusta sairaisissa tai mielenkiinnon kohteena olevissa yksilöissä. [3, s. 5, 35; 4.]

KytKentäanalyysit ovat genomitutkimuksia, jotka tehdään sukupuuaineistoista. Aineistoista seurataan mielenkiinnon kohteena olevaa ominaisuutta tai sairautta ja mahdollisten poikkeavien alleelien ylliedustusta näillä yksilöillä ja sen periytyvyyttä perheen sisällä. KytKentäanalyysien tarkoituksena on osoittaa kytKentä ominaisuuden ja alleelien välillä sekä sulkea pois alueita, joissa ei voida vahvistaa kytKentää. [2, s. 73–74; 3, s. 42–43.]

### 3 Näytteet

#### 3.1 GeneRisk

Tähän työhön valittiin näytteet GeneRisk-projektista, jossa tutkitaan suomalaisia perinnöllisiä kansantauteja. Tutkimuksessa keskitytään erityisesti sydän- ja verisuonitauteihin ja niihin liittyviin tietoihin perimästä genotyyppimäärityksen avulla. Pää tavoitteena projektissa on pyrkiä ehkäisemään sairauksien kehittymistä sekä tutkia, miten tietoa voidaan hyödyntää taudin ehkäisyssä. Tutkimuksessa hyödynnetään tautien syntyyn vaikuttavaa genomitietoa ja kehitetään menetelmiä tiedon analysointiin ja sen viemiseen perusterveydenhuoltoon, työterveyshuoltoon ja erikoissairaanhoidon. Myös genomitiedon vaikutusta yksilön terveysvalintoihin ja -käyttäytymiseen arvioidaan projektin aikana. Tutkimuksen tuloksia voidaan käyttää jatkossa myös laajemmin kansantauteihin liittyvissä tutkimuksissa, sillä tutkimustiedot tallennetaan Terveiden ja hyvinvoinnin laitoksen (THL) biopankkiin ja tutkimuksen kokonaiskesto on 20 vuotta. [13.]

Yksilöllisen tautidiagnostiikan, sairastumisen ehkäisyn, hoidon sekä seurannan toivotaan lisääntyvän arkipäivän terveydenhuollossa hyödyntämällä yksilön genomitietoa. Kansantauteihin sairastumisen riski pyritään tunnistamaan geeniprofiilista ja yksilölle voidaan muokata henkilökohtainen kohdennettu hoitomuoto. Tutkimus pyrkii saavuttamaan kustannustehokkuutta ja parempaa hoitoa yhteiskunnalle kalliissa sairastumistapauksissa. Tauteja pyritään ennaltaehkäisemään tunnistamalla sairaudelle altistavat geenit ja toteuttamalla yksilöille henkilökohtainen hoitomuoto geenialttiuden perusteella. Myös yksilön aktiivisuutta halutaan lisätä terveyden huoltamisessa luovuttamalla henkilölle tieto mahdollisesta sairastumisriskistä ja elämäntapojen vaikutuksista riskin alentamiseen. Arviointi riskin suuruudesta perustuu osallistujan terveystarkastuksen, kyselylomakkeessa annettujen tietojen ja genomitiedon yhteisanalyysiin. [13.]

Tutkimukseen osallistuvat ovat asiakkaita Mehiläisen työterveyspalvelusta ja Kymenlaakson sairaanhoito- ja sosiaalipalvelujen kuntayhtymästä Careasta sekä Suomen Punaisen Ristin Veripalvelusta verenluovuttajia. Tutkimukseen valitut henkilöt ovat 45–65-vuotiaita eteläsuomalaisia. Tutkimus on osa SalWen koordinoimaa Yksilöllistetty diagnostiikka ja hoito (GET IT DONE) -ohjelmaa, ja FIMM toimii tutkimuksen yhtenä rahoittajana. [13.]



GeneRisk-projektin näytteet genotyyppitettiin Illumina Inc. (San Diego, Ca) Infinium -menetelmällä HTS CoreExome-24 v.1.1 BeadChip -sirulla ja ne luettiin Illuminan iScan-skannerilla. CoreExome-sirulla on yli 551 000 markkeria näytettä kohden.

### 3.2 Kokoverinäytteet

Kokoveri sisältää runsaasti erilaisia komponentteja, jotka voivat häiritä tai estää PCR:ää tapahtumasta [1]. Veri koostuu muun muassa solunväliaineesta plasmasta, puna- ja valkosoluista ja verihiutaleista sekä monista ravintoaineista ja entsyymeistä. Aikuisen ihmisen veri sisältää noin puolet punasoluja ja puolet plasmaa sekä hiukan muuta ainesta. Veri kuljettaa happea, ravinteita ja lämpöä kehon eri osiin kantaen mukanaan PCR-reaktion kannalta huomattavia määriä ei-haluttuja komponentteja. [14.]

Verinäytteestä DNA saadaan eristettyä suurista tumallisista leukosyyteistä eli valkosoluista, joita on veressä suhteellisen vähän. Valkosolujen tehtävä on toimia immuunivasteena tuottamalla vasta-aineita bakteereita ja viruksia vastaan. Valkosoluista eristettävä DNA on suurimmaksi osaksi genomista DNA:ta tumasta ja lisäksi mitokondriaalista DNA:ta. [14; 15.]

Punasolut ovat litteitä soluja, jotka kuljettavat happea ja hiilidioksidia. Veren punainen väri tulee punasolujen sisältämästä hemoglobiinista. Hemoglobiini on happea ja hiilidioksidia sitova proteiini, jota on noin kolmasosa punasoluista eli noin kuudesosa veren tilavuudesta. Punasoluissa ei ole lainkaan tumaa, joten niistä ei saada eristettyä lainkaan genomista DNA:ta. [14; 15.]

Veren plasma on läpikuultavan kellertävää nestettä, josta noin 90 % on vettä ja loppuosa koostuu veren kuljettamista ainesosista kuten happi, hiilidioksidi, kuona-aineet ja plasmaproteiinit. Noin 8 % plasman painosta on proteiineja. Plasman proteiineista suurin osa, noin 60 %, on albumiineja. Muita proteiineja plasmassa ovat globuliinit ja fibrinogeenit, jotka ovat tärkeässä osassa veren hyytymisessä. Plasma sisältää myös entsyymejä, hormoneja ja muita proteiineja sekä orgaanisia ravintoaineita. [14; 15.]

Tämän työn verinäytteisiin on lisätty näytteenottovaiheessa etyleenidiamiinitetraetikkahappoa (EDTA,  $C_{10}H_{16}N_2O_8$ ), joka toimii lisäaineena estämässä veren hyytymistä eli antikoagulanttina. EDTA on hyvin yleinen verinäytteiden lisäaine, ja sen on todettu toimivan

myös PCR:ssä. Veren hyytymisen estoon on käytössä muitakin lisäaineita, esimerkiksi hepariinia, mutta niiden käyttäytyminen reaktiossa voi olla hyvin erilaista ja vaikuttaa oleellisesti työn onnistumiseen. [1.]

### 3.3 Genotyyppimäärityksen laadunvalvonta

Yliopistojen ja sairaaloiden tutkimusryhmät ja eri yritykset tekevät monenlaisia tutkimuksia joissa käytetään genotyyppimääritystuloksia. Tutkimusryhmät keräävät aineistoa yksilöistä ja on erityisen tärkeää, että laboratoriossa saadut tulokset vastaavat todellisuudessa kyseistä yksilöä. [16.]

Näytteet yksilöistä kerätään, kirjataan ja lähetetään DNA-eristykseen, josta DNA-näyte saapuu genotyyppimääritykseen ja tulokset palautetaan tutkimusryhmälle analysoitavaksi. Moniportaisen näytteenkuljetusketjun ja genotyyppimäärityksen näytekäsitteilyn aikana on riski näytetietojen tai näytteiden sekaantumiselle, jolloin lopullinen genotyyppitulos ei vastaa todellisuudessa yksilön genotyyppiä. Tällä hetkellä verinäyte saadaan genotyyppityslaboratorioon DNA:n eristysvaiheessa, jolloin on mahdotonta tietää, onko näytteitä sekoittunut jo ennen tätä. On myös mahdollista, että genotyyppimäärityksessä saadaan virheellisiä tuloksia. Tämä voi johtaa vakaviin väärinkäsityksiin ja diagnooseihin tai väärrään riskiarvioon yksilön terveydentilasta sekä voi haitata tutkimusta esimerkiksi geenejä etsittäessä. Sekoittuneet näytteet voivat siis aiheuttaa huomattavia haittoja yksilölle sekä genotyyppituloksiin pohjautuville tutkimuksille sekä lisätä huomattavasti työtä ja kuluja, kun analyysit on tehtävä uudelleen.

Saapuvista näytteistä saadaan jonkin verran esitietoja, joiden perusteella voidaan tarkastaa yksilön ja näytteen välillä olevia yhteyksiä, mutta usein yksilöistä on tiedossa vain sukupuoli tai ei edes sitä. Laadunvarmennusmenetelmänä näytesarjaan voidaan lisätä duplikaatti jostakin näytteestä, jolloin määrityksen onnistuminen tarkastetaan duplikaattien yhteneväisyydestä. Myös duplikaatin sijainnin avulla voidaan tarkastaa, että näytteet eivät ole sekoittuneet. Duplikaatti sijoitetaan satunnaiseen tunnettuun kohtaan näytelevyllä, jolloin tulosten analysoinnissa duplikaatin tulisi löytyä edelleen samasta kohdasta.

Nämä tiedot eivät yksinään riitä näytteiden yhteneväisyyden tarkistukseen ja yksilöntunnistukseen. Näytteiden sekaantumista tapahtuu ja se voidaan huomata, jos mahdolliset

sukulaistiedot eivät vastaa toisiaan, sukupuolitiedot ovat väärin tai näytteistä sadut tulokset eivät vastaa esitietoja ja odotettuja tuloksia. Tässä vaiheessa sekaannuksen ilmetessä on hyvin vaikeaa tai mahdotonta selvittää, milloin ja kuinka monet näytteistä ovat sekoittuneet.

## 4 Menetelmät

FIMM:in teknologiakeskuksen genotyyppitysyksikössä on käytössä useita erilaisia menetelmiä näytteiden genotyyppien määrittämiseen. Tässä työssä verinäytteiden analyysissä on käytetty Agena Bioscience (San Diego, CA) iPLEX -kemialla ja MassARRAY Analyzer Compact -massaspektrometriä, sekä genotyyppien lukuun ja tarkastukseen MassARRAY Typer 4.0 -ohjelmistoa. Tähän työhön on käytetty 27 alukkeen ID-PANELia, joka on tämän työn SNP-markkerikokonaisuudelle annettu jatkossa käytettävä nimi.

Yksilöntunnistus tapahtuu vertaamalla kahta genotyyppitettävää näytettä. Varsinainen näyte on normaaliin genotyyppimäärittämiseen saapuva näyte, jonka SNP-profiilia verrataan verinäytteestä saatuun genotyyppitulokseen. Normaalisissa genotyyppimäärittämisestä löytyy samat 27 SNP-markkeria, kuin verinäytteiden genotyyppityksessä käytetyssä ID-PANELissa. Tällä varmistetaan, että DNA-näytteestä saadut genotyypit ovat oikeasta tutkimushenkilöstä eivätkä näytteet ole sekaantuneet syystä tai toisesta DNA:n eristys- tai genotyyppityksen aikana. Lisäksi GeneRisk-projektin DNA-näytteet genotyyppimääritetään ID-PANELilla ja tuloksia verrataan verinäytteiden tuloksiin ja alkuperäisiin GeneRisk-tuloksiin.

### 4.1 Markkerien valinta

Tähän työhön on valittu SNP-markkerit monin eri kriteerein. Yksilöntunnistamiseen valittujen markkereiden riittävän lukumäärän lisäksi halutaan, että markkereista on aiemmin saatu toimivia vertailutuloksia laboratorion eksomisekvensointiprojekteissa, jotta ID-PANEL olisi käyttökelpoinen myös DNA-sekvensoinnin laadunvarmistukseen. Markkerien tulee myös löytyä laboratoriossa yleisimmin käytetyiltä koko genomien kattavilta genotyyppityssiruilla. Lisäksi markkereiden halutaan olevan tehokkaita, eli niiden on aikaisemmissa genotyyppitysprojekteissa todettu toimivan hyvin koko genomien siruilla, jolloin vertailumateriaalia voidaan pitää luotettavana. SNP-markkerien ja niiden alukkeiden tulee

myös toimia toisistaan riippumatta, eivätkä ne saa sijaita kromosomeissa liian lähellä toisiaan. Lähellä toisiaan sijaitsevat SNP:t voivat periytyä yhdessä toisiinsa kytkeytyneenä kaikille jälkeläisille ja voivat löytyä samanlaisena variaationa lähes koko populaatiosta, jolloin näistä kahdesta SNP:stä ei voida erottaa yksilöitä toisistaan [2, s. 73–74].

Markkereilla pyritään kattamaan koko kromosomisto niin, että jokaisesta kromosomiparista valitaan vähintään yksi SNP-markkeri. Yksilöntunnistuksessa käytettävien SNP-mutaatioiden harvinaisempi alleeli tulisi olla mahdollisimman yleinen. Harvinaisemman alleelin tulisi esiintyä vähintään 30 prosentilla populaatiosta, mutta mielellään jopa 40 prosentilla, jotta saadaan mahdollisimman hyvä varioiva tulos. Harvemmin esiintyvien alleelien variaatioista voidaan saada hyvin samanlainen genotyyppitulos useimmille yksilöille, mikä voi estää yksilöntunnistamisen.

Alukesekvenssin alle ei saisi jäädä toista SNP-mutaatiota, ettei tämän toisen alleelin polymorfia häiritse alukkeen kiinnittymistä, sillä alukkeet suunnitellaan sekvenssispesifisti ja yhdenkin nukleotidin variaatio voi haitata alukkeen kiinnittymistä. Tämä voi johtaa huonolaatuisiin genotyyppituloksiin tai jopa vaikuttaa huomattavasti alleelituloksiin, sillä monistuminen voi tällöin suosia vain jommankumman alleelin monistumista. Toisen alukkeen jäädessä monistumatta voidaan saada virheellisiä homotsygoottituloksia heterotsygooteista tai toinen homotsygoottigenotyyppi voi jäädä kokonaan monistumatta.

Tietoa markkereista saadaan esimerkiksi dbSNP-tietokannasta ([www.ncbi.nlm.nih.gov/snp](http://www.ncbi.nlm.nih.gov/snp)), joka on NCBI:n (National Center for Biotechnology Information, USA) ylläpitämä biotietokanta. Tietokannasta löytyvät tiedot SNP-markkereiden ympäriltä löytyvistä sekvensseistä, alleeleista, kromosomista, jossa SNP sijaitsee, SNP:n lokus kromosomissa sekä alleelifrekvensseistä maantieteellisesti ja eri tutkimuksien mukaan.

#### 4.2 Usean alukkeen SNP-genotyyppitys

PCR:ää voidaan käyttää välineenä SNP-variaatioiden määrittämiseen useilla spesifisillä alukkeilla samanaikaisessa reaktiossa. Usean alukkeen PCR:ssä tunnistetaan alleelit genomista ekstensioalukkeiden 3'-päästä seuraavan nukleotidin kohdalta. Näin saadaan analysoitua tehokkaasti kerralla riittävä määrä variaatiota yksilöntunnistamiseen ja poissuljettua täysin identtisten alleelivariaatioiden yhdistelmien esiintyminen yksilöiden

välillä. Yhden reaktion onnistumiseen vaaditaan tarkkaa optimointia, sillä alukkeet ja niiden konsentraatiot vaativat hyvät olosuhteet reaktion onnistumiseksi. Alukkeiden on oltava myös sekvenssiltään spesifisesti yhteensopivia DNA-juosteen kanssa, jotta vältetään muun muassa epäspesifisten sekvenssialueiden monistamiselta, DNA-juosteen muodostamilta silmukkarakenteilta sekä alukkeiden kiinnittymiseltä toisiinsa. PCR-reaktiossa myös muiden reagenssien tulee olla sopivassa suhteessa optimaalisen tuloksen takaamiseksi.

SNP-markkerit valitaan edellisen kappaleen kriteerien mukaisesti, mutta niiden toimivuus analyysissä tulee arvioida ennen niiden testaamista käytännössä. Tähän käytetään Agena Bioscience Assay Design Suite v2.0 ([mysequenom.com/Tools](http://mysequenom.com/Tools)) -ohjelmistoa. Ohjelmaan syötetään valitut SNP-markkerit rs-numeroina tai FASTA-sekvenssinä ja määritetään parametrit, joita ohjelma käyttää PCR- ja ekstensioalukkeiden suunnittelussa. Ohjelmalle annetaan markkereiden lukumäärä, jota halutaan käyttää yhteen analyysiin ja valitaan myös markkeria ympäröivän sekvenssin maksimipituus, jota ohjelma tarkastelee sekvenssitietokannasta ja käyttää hyväksi oligojen suunnittelussa.

Genotyyпитettävän SNP:n ympärillä olevat muut SNP:t, joilla voi olla merkitystä työn kannalta, halutaan kartoittaa valitsemalla tarkasteltavan populaation määritetyt SNP:t. Eurooppalaista väestöä tarkasteltaessa esimerkiksi aasialaisesta väestöstä löytyvät SNP:t eivät todennäköisesti häiritse alukkeita, vaikka kyseinen SNP jäisikin alukkeen alle DNA-juosteessa. Ohjelmassa voidaan jättää myös suunnittelussa huomiotta SNP:t, joiden status ja populaatiotiedot eivät ole tunnettuja, sekä valitun markkerin ympärillä olevat SNP:t, joiden alleelifrekvenssit ovat alle halutun rajan, jolloin niiden esiintyvyys kyseisessä populaatiossa jää hyvin pieneksi.

Alukkeiden täsmävyyttä sekvenssiin voidaan muokata valitsemalla esimerkiksi, kuinka monta emästä on vastattava sekvenssiä alukkeen 3'päästä ja kuinka monta emästä sallitaan poikkeavan alukkeen ja sen alle jäävästä sekvenssistä. Ohjelmaan voi muokata myös painoarvoja sille, mitä ominaisuuksia erityisesti halutaan ottaa huomioon suunnittelussa. Esimerkiksi voidaan haluta välttää toisiinsa kiinnittyviä alukkeita. Toisiinsa kiinnittyvät alukkeet eivät ole enää käytössä halutun juosteen monistamiseen, ja väärään kohtaan kiinnittyvät alukkeet voivat monistaa ei toivottua sekvenssiä tai muodostaa silmukoita DNA-juosteeseen.

Usean alukkeen suunnittelussa voidaan muokata monia parametrejä, mutta suuret muokkaukset voivat vaikuttaa huomattavasti ja ennalta-arvaamattomasti reaktion onnistumiseen. Ohjelma suunnittelee syötettyjen tietojen mukaisen usean alukkeen kokoelman (multiplex) ja antaa jokaiselle alukkeelle toimimisarvion prosentteina. Tämä on kuitenkin hyvin suuntaa antava arvio, ja alukkeet tulee testata myös käytännössä. Reaktioon voidaan vaikuttaa myös PCR:n optimoinnilla.

#### 4.3 Suoraan verestä tehtävä PCR

Yleensä PCR:ssä monistus tehdään eristetyistä, mahdollisimman puhtaasta ja ehjistä DNA:sta. DNA:n eristys käsittää useita vaiheita, jolloin riski kontaminaatiolle ja näyteselekaannuksille kasvaa. Tässä työssä DNA:ta ei eristetä verinäytteestä, vaan veri liuotetaan puskuriliuokseen, inkuboidaan lämpökaapissa ja liuos sentrifugoidaan. Liuoksesta käytetään supernatantti jatkoon. Näyte kuivataan PCR-levyn pohjaan ennen muiden PCR-reagenssien lisäämistä. Verinäytteestä PCR:n toteuttaminen on kuitenkin haastavaa veren sisältämien soluainekseen ja molekyylien takia, sillä jotkut näistä aineksista saattavat vaikuttaa DNA-polymeraasin aktivaatioon tai muuten inhiboida reaktiota. Myös vereen lisätyt hyytymisenestoaineet voivat häiritä PCR:n onnistumista. Suurimpina inhibiittoreina on todettu olevan punasolujen hemoglobiini ja laktoferriini sekä immunoglobuliini G plasmassa. Inhibitio tapahtuu näiden molekyylien estäessä DNA-polymeraasin kiinnittymisen genomiseen DNA:han. [1.]

Suoraan verinäytteestä tehtävää PCR:ää on testattu erilaisin menetelmin ja erilaisin veren käsittelykeinoin aikaisemminkin. Yhtenä apuvälineenä on käytetty korkeaa puskurin pH:ta, joka vaimentaa inhibitiota pienentämällä genomisen DNA:n ja inhibiittoreiden välistä sähköstaattisuutta. Korkea pH (9,1–9,6) vaikuttaa proteiinien nettovaraukseen negatiivisesti ja ehkäisee niiden kiinnittymisen negatiivisesti varautuneeseen genomiseen DNA:han. [1.]

Eräänä menetelmänä on käytetty monivaiheista PCR:ää. Kokoverta käytetään 1–2 µl 100 µl:an reaktiilavuuteen. DNA:n denaturaatiovaihe tehdään kolme kertaa nostamalla lämpötilaa ensin kolmeksi minuutiksi 94 °C:seen ja jäähdyttämällä se joka välissä kolmeksi minuutiksi 55 °C:seen. Lämpötilan vaiheittaisen nostamisen ja laskemisen useaan kertaan todettiin olevan kriittinen PCR:n onnistumisen kannalta. Tämä käsittely todennäköisesti denaturoi inhiboivia proteiineja ja edesauttaa reaktion onnistumista. [17.]

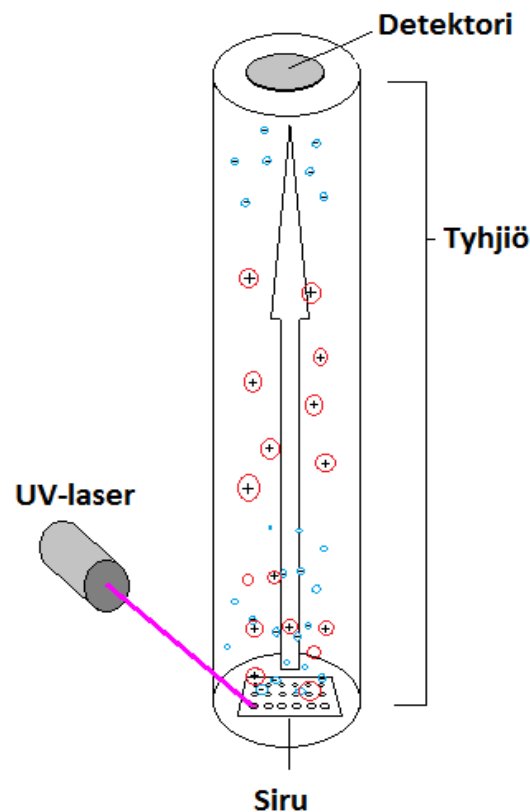
McCusker ryhmineen sai kasvatettua kokoveren näytetilavuutta edellisestä 45 µl:aan samassa reaktiutilavuudessa. Verinäytteisiin (1–50 µl) lisättiin vettä 50 µl:n kokonaistilavuuteen. Ennen PCR-seoksen lisäystä ja itse reaktion tekoa näytteet kuumennetaan 15 minuutiksi 95 °C:seen. Verinäytteiden kuumennus tehdään solujen lyysaamiseksi ja inhiboivien proteiinien inaktivoimiseksi. [18.]

Tämän opinnäytetyön erityisenä haasteena on useiden alukeparien käyttö PCR- ja ekstensioreaktiossa samanaikaisesti. Näytteet kuumennetaan tässäkin työssä (70 °C o\`n), jonka aikana veren proteiineja denaturoituu ja solukalvoja hajoaa. Lisäksi näytteet kuivataan levyille ennen PCR-seoksen lisäystä.

#### 4.4 MALDI-TOF MS

Tässä työssä käytetään genotyyppien määrittämiseen matriisiavusteista laser-desorptio/ionisaatiolentoaikamassaspektrometria (MALDI-TOF MS, Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry). MALDI-TOF MS on pitkään käytetty menetelmä DNA:n analysointiin kvantitatiivisesti ja kvalitatiivisesti. Menetelmää käytetään tunnistamaan erilaisia molekyylejä massansa perusteella sen tehokkuuden ja luotettavuuden ansiosta. [19.]

MALDI-TOF-massaspektrometria käytetään tässä työssä kehitettävässä menetelmässä ekstensiotuotteiden määrittämiseen. Massaspektrometrin avulla saadaan tieto molekyylien massasta ionisoimalla näytteet laserilla ja mittaamalla niiden lentoaika detektorille (kuva 3). Massaspektrometrin detektori tunnistaa vain ionisoidut molekyylit, ei neutraaleja molekyylejä tai sen osia. Molekyylit ionisoituvat joko luovuttamalla tai vastaanottamalla elektronin tai protonin. Elektronin siirrossa muodostuu radikaalianioneja tai -kationeja, joilla on pariton määrä elektroneja. Ionisoituminen voi tapahtua myös protonin siirrossa, jossa molekyyli luovuttaa protonin eli deprotonoituu tai vastaanottaa protonin eli protonoituu, mutta elektroneja on parillinen määrä. [8; 20, s. 17, 212.]



Kuva 3. MALDI-TOF -massaspektrometri ja sen toimintaperiaate. [21.]

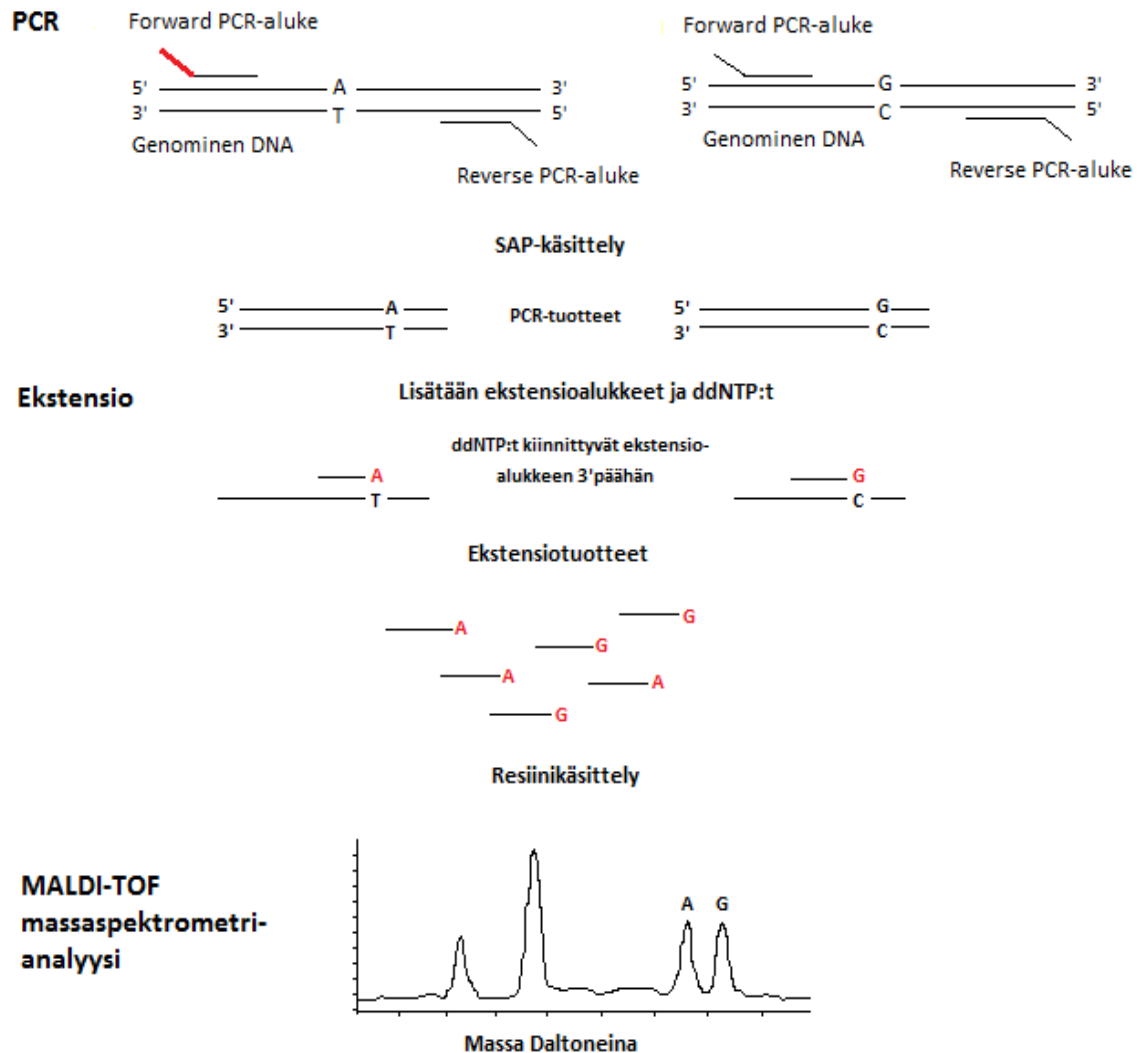
Näytteet siirretään SpectroCHIP-sirulle, jossa ekstensio-oligot kiinnittyvät matriisiin. Siru syötetään massaspektrometriin, joka vie sirun tyhjiöön. Näytteet säteilytetään laserilla, joka ionisoi näytteet, ja ionisoidut molekyylit erotetaan toisistaan niiden lentoajan perusteella. Positiivisesti varautuneet molekyylit lentävät sähkö- ja magneettikenttien avulla detektorille, joka tunnistaa ne massa-varaussuhteen perusteella. Detektori mittaa ionien lentoaikaan suhteutetun massa-varaussuhteen ja intensiteetin, jonka avulla se piirtää tunnistetuille molekyyleille spektrin. [8; 20, s. 39–42; 21.]

#### 4.5 PCR ja ekstensio

Genomisesta DNA:sta valitaan tutkittavat SNP:t ja niitä monistetaan PCR:n avulla (kuva 4). PCR-alukkeet suunnitellaan SNP:n molemmiin puolin DNA:n molempiin juosteisiin, alukkeet ovat esitettyinä liitteessä 1. PCR-alukkeisiin liitetään 5'-päästä 10 emästä, jotka eivät kiinnity DNA-juosteeseen, nämä ovat kuvattuna punaisella kuvan 4 PCR-kohdassa.



Tällä ehkäistään ylimääräisten PCR-alkukkeiden sekaantuminen lopullisiin ekstensiotuotteisiin massaspektrissä, sillä niiden massa on huomattavasti suurempi tämän sekvenssijakson ansiosta. [22.]



Kuva 4. Tässä työssä käytetty PCR, Ekstensio ja massaspektri kuvattuna Agenan iPLEX -menetelmässä [22].

Samassa reaktiossa tuotettavat usean alukkeen reaktiotuotteet täytyy erottaa jollain menetelmällä toisistaan. Tähän käytetään eripituisiksi suunniteltuja ekstensioalukkeita, joita käytettäessä ekstensiotuotteet eroavat toisistaan vähintään 16 Da:n verran. Ekstensioalukkeet ovat listattuna liitteessä 2. Joihinkin alukkeisiin on lisätty 5'-päähän häntä, jos niillä ei ole riittävästi komplementaarista sopivaa sekvenssiä. Tämän kaltaisissa tilanteissa on parempi käyttää juosteeseen kiinnittymätöntä häntää (liite 2), kuin todellista

sekvenssiä, jonka kanssa reaktio voisi toimia huonommin. Häntä lisätään, jotta se erotuisi riittävästi massaltaan muista alukkeista.

SNP:t tunnistetaan ekstensioalukkeiden pituuksista (liite 2) ja alleelit erotetaan toisistaan sen perusteella, mikä emäs ekstensioalukkeen 3'-päähän tarttuu. Massan mukaan järjestyneisiin ekstensioalukkeisiin kiinnittyy ekstensioreaktiossa SNP-markkeria vastaava ddNTP, eli dideoksinukleotidi. Reaktio päättyy ddNTP:n kiinnittymiseen, sillä nukleotidin 3'-päässä normaalin OH-ryhmän tilalle on nukleotidiä muokkaamalla jätetty vain vety. Tämän nukleotidin ja ekstensiotuotteen oletetun massan perusteella tunnistetaan allelivariaatiot näytteistä massaspektrometrillä. Ekstensio-oligojen sekvenssit on esitettyinä liitteessä 2. Niissä pienillä kirjaimilla on merkitty hännät ja isoilla kirjaimilla sekvenssi, joka kiinnittyy DNA-juosteeseen. [22.]

## 5 Työn toteutus

### 5.1 Näytteenkäsittely

EDTA-verinäyte kuivataan näyteliuskoille (EMD Millipore Corp., Merck KGaA, MQuant™ Blank Strips, Cat: 1.11860.0001) ja kuivattu näyte liuotetaan TE-puskuriin (10 mM TRIS + 1 mM EDTA pH 8,0, Cat: T00031). Näyteliuskan tyynyosa kastetaan verinäytteeseen ja annetaan kuivua noin kaksi tuntia ennen säilömistä. Näytteitä säilytetään kuivana ilmatiiviisti putkissa tai suljettavissa muovipusseissa. Putket ja muovipussit säilytetään kuivassa ja viileässä valolta suojattuna.

Liuskan tyynyosa leikataan kahdessa osassa pyöreäpohjaiseen laimennosputkeen ja pipetoidaan puskuriliuos päälle ja pyöräytetään mikrosentrifuugissa (Wealtec, E-Centrifuge, S/N ECW08024513), jotta liuskat kastuvat varmasti puskuriin. Näytteet liukenevat puskuriliuokseen yön yli 70 °C:ssa (Illumina Hybridization oven, S/N 061601464). Tämän jälkeen putket sentrifugoidaan (Eppendorf Centrifuge 5424, S/N 5424CL451204) 1 min 10 000 rpm ja pipetoidaan varovasti noin 50 µl supernatanttia välttämällä helposti hajoavaa pellettiä putken pohjalla. Leikattu tyynyosa on edelleen putkessa, ja optimaalisessa tilanteessa kiintoainees jää liuskan alle helpottaen pipetoimista. Supernatantti siirretään 96-kuoppalevylle säilytykseen jatkokäsittelyä varten. Levy voidaan säilöä lyhyen aikaa 4 °C:ssa, pidempiaikaiseen säilytykseen levy pakastetaan -20 asteeseen.

## 5.2 Työn kulku

Tämän osuuden reagenssit ovat Agenan iPlex -menetelmään suunniteltuja valmiita kittejä (Agena Bioscience Complete iPlex Gold Genotyping Reagent Sets). Työn alkuosa tehdään laboratorion pre-PCR-puolella DNA-kontaminaatioiden välttämiseksi. TE-puskuriin liuotetut näytteet pipetoidaan 384-kuoppalevylle, sentrifugoidaan (Eppendorf, Centrifuge 5810R, S/N 5811DM573326) 1 min 1000 rpm ja kuivataan 37 °C:ssa yön yli. DNA:ta tarvitaan normaalisti noin 20 ng/näyte, mutta verinäytteiden kohdalla DNA-konsentraatio ei ole tiedossa, joten näytetilavuutena käytetään 0,5 ja 1 µl supernatanttia.

Kuivatuksen jälkeen valmistetaan alukeseos, jossa sekoitetaan jokaisen SNP-markkerin kohdalta forward- ja reverse-alue ja PCR-seos, joka pipetoidaan näytteiden päälle (5 µl/näyte), ennen PCR-laitteeseen viemistä (Applied Biosystems, Veriti 384 well Thermo Cycler, S/N 299030331). Taulukossa 1. on esitetty tässä työssä käytetty PCR-ohjelma ja reagenssit. Tämän jälkeen työskennellään post-PCR-puolella. PCR:n onnistuminen tarkastetaan satunnaisista näytteistä agarosigeelielektroforeesilla (Lonza, Flash Gel 2,2 %).

Taulukko 1. PCR-ohjelma ja iPlexGoldPCR-reagenssit

95 °C	2 min		Accugene water	330,3 µl
95 °C	30 s	45 cycles	10xPCR buffer	58,6 µl
56 °C	30 s		MgCl <sup>2</sup>	46,8 µl
72 °C	1 min		dNTPmix	11,7 µl
72 °C	5 min		PCR primermix (0,5 µM)	117,1 µl
4 °C	∞		PCR enzyme (5U/µl)	21,1 µl

PCR-tuotteet puhdistetaan Shrimp Alkaline Phosphatase (SAP)-entsyymikäsittelyllä. Entsyymi defosforyloi vapaat fosfaattiryhmät poistaen ylimääräiset dNTP:t haittaamasta ekstensioreaktiota. Taulukossa 2. on SAP-ohjelma, joka toteutetaan PCR-laitteella sekä käytetyt reagenssit. SAP-reagenssiseosta käytetään jokaiselle näytteelle 2 µl.

Taulukko 2. SAP-ohjelma ja iPlexGoldSAP-reagenssit

37 °C	40 min	Accugene water	196 µl
85 °C	5 min	10xSAP buffer	22 µl
10 °C	∞	SAP enzyme (1U/µl)	38 µl

Ekstensioon valmistetaan ekstensioalukeseos. Alukkeet on jaettu viiteen massaryhmään, joista kaikista pipetoidaan massaryhmälle laskettu tilavuus aluketta. Massaryhmään jakamisella halutaan varmentaa myös pidempien ekstensio-oligojen toimiminen reaktiossa ja niitä laitetaan seokseen suhteessa enemmän. Tämän jälkeen valmistetaan ekstensioseos, joka pipetoidaan levyille (2 µl) ja ajetaan levy PCR-laitteella ekstensio-ohjelmalla (taulukko 4).

Taulukko 3. Ekstensio-ohjelma ja iPlexGoldExtension-reagenssit

94 °C	30 s		
94 °C	5 s		
52 °C	5 s	5 cycles	40 cycles
80 °C	5 s		
72 °C	3 min		
10 °C	∞		

Accugene water	85 µl
10xEXT buffer	27,5 µl
ddNTPmix	27,5 µl
EXTprimermix (7 µl: 14 µl)	129 µl
Enzyme	5,6 µl

Ekstension jälkeen tuotteet puhdistetaan muun muassa suoloista resiinikäsittelyllä. Resiini on kosteaa hiekkamaista ainetta, joka viedään muotin avulla jokaiseen näytekuppaan. Liuoksessa olevat ylimääräiset ionit kiinnittyvät resiiniin, eivätkä näin häiritse reaktiota. Levy sentrifugoidaan ilmakuplien poistamiseksi, jonka jälkeen levyä pyöritetään 30 minuuttia 360° resiinin vaikuttamiseksi. Pyörytyksen jälkeen levy sentrifugoidaan 5 min 3000 rpm, jotta resiini saadaan varmasti kuoppien pohjalle.

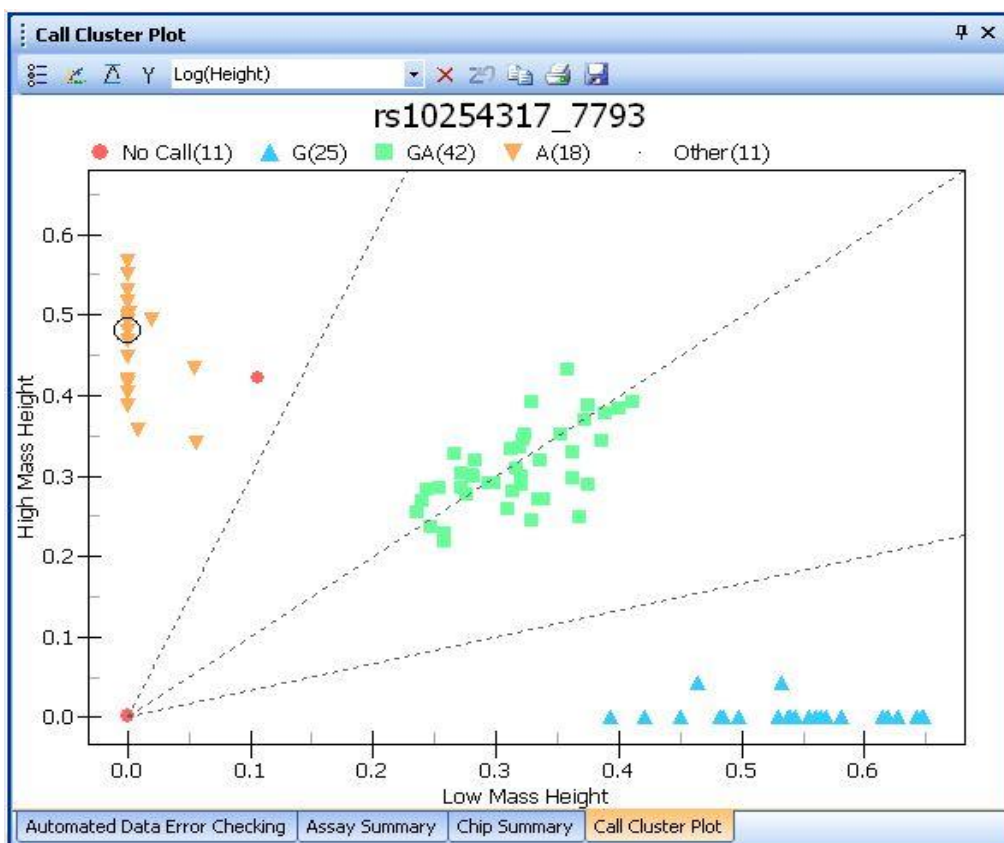
Ekstensiotuotteet siirretään pipetointirobotilla (Agena Bioscience MassARRAY Nano-dispenser RS1000, S/N 08P0627RS-01) sirulle (SpectroCHIP II-G384, lot 0000022427), jossa ne kristallisoituvat ja kiinnittyvät piilevyllä olevaan matriisiin. Näytettä pipetoidaan pipetointirobotilla sirulle hyvin pieni määrä, joka mitataan nanolitroissa.

### 5.3 Tulosten analysointi ja laaduntarkkailu

Massaspektrimestä (Agena Bioscience, S/N BA 005405) saatu data on suoraan luettavissa MassARRAY Typer 4.0 -ohjelmalla, jonka avulla näytteiden ja markkereiden onnistumiset tarkastetaan manuaalisesti. Tulosten luku ja tarkastus on tässä vaiheessa

suhteellisen subjektiivista ja tuloksiin tehdään tarvittaessa muutoksia. Ohjelman hylkäämiä näytteitä voidaan hyväksyä ja hyväksytyjä hylätä. Tulosten tarkastamiseen harjaantuu sitä tehdessä ja aloittelijaa ohjeistetaan analysoinnissa.

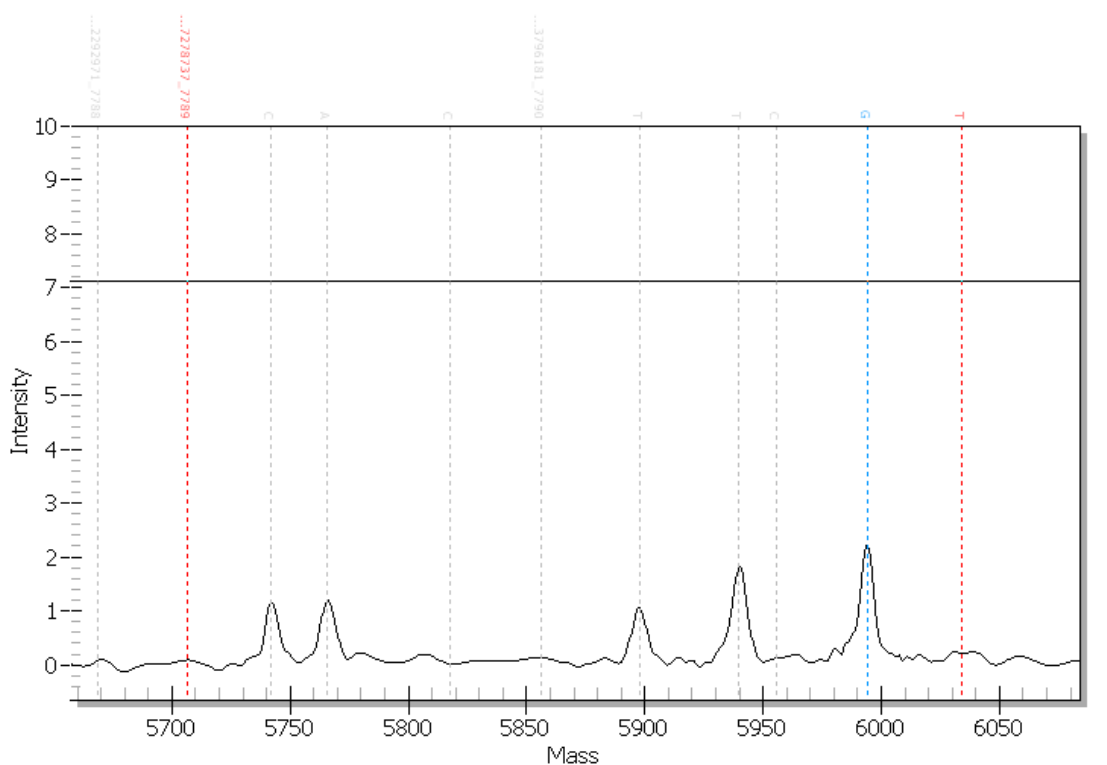
Näytteet klusteroidaan genotyypinsä mukaan homo- ja heterotsygooteiksi jokaisen markkerin kohdalla (kuva 5). Klustereiden tasaisuus tarkistetaan ja mahdolliset klustereista erillään olevat näytteet tarkastetaan manuaalisesti.



Kuva 5. Klusterit yhden markkerin kohdalla kaikista näytteistä. Vihreät merkit ovat heterotsygootit ja oranssit ja siniset merkit ovat eri homotsygootit. Punaiset merkit kuvaavat epäonnistuneita näytteitä. Klustereista eroavien näytteiden onnistumiset tarkastetaan manuaalisesti.

Yksittäisen näytteen onnistumista tarkastellaan massaspektristä. Spektrissä näkyy jokaisen alleelin ja ekstensioalukkeiden kohdat ja niiden muodostamat piikit käyrässä (kuva 6), tarkasteltavan näytteen piikit näkyvät värillisillä katkoviivoilla. Homotsygooteille muodostuu yksi piikki ja heterotsygooteille kaksi piikkiä. Mikäli heterotsygooteilla alleelipiikit ovat selvästi erikorkuisia, hylkää ohjelma tuloksen herkästi, mikä korjataan selkeissä tapauksissa käsin. Heterotsygoottipiikit ovat keskimäärin matalampia, kuin homotsygoottipiikit ja niitä saatetaan joutua hylkäämään helpommin, sillä niiden massapiikit saadaan

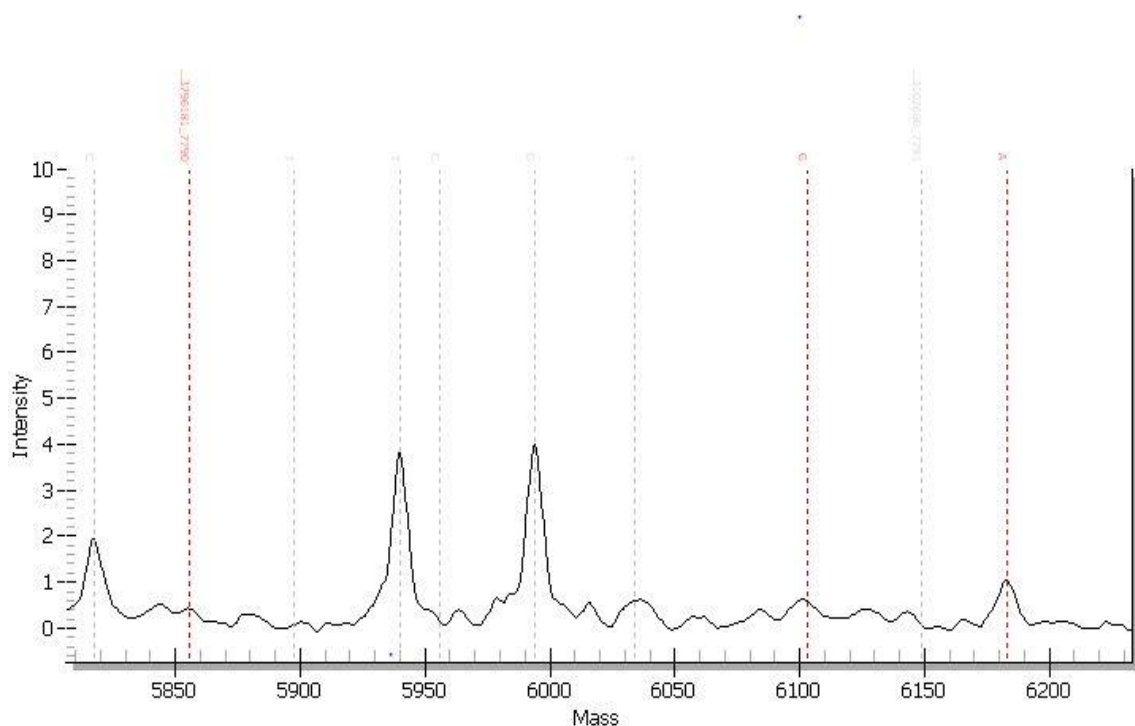
vain yhdestä alleelikopiosta. Heterotsygoottipiikeissä on kopioita lukumääräisesti vähemmän ja siitä syystä niillä on heikompi intensiteetti. Joskus piikkien korkeuserot voivat johtua myös kopiokuvavariaatiosta. Homotsygooteilla massapiikki saadaan kahden samanlaisen ekstesiotuotteen määrästä, joten niistä saadut piikit ovat yleensä huomattavasti korkeampia kuin heterotsygooteilla.



Kuva 6. Kuvassa on esitetty pieni osa massaspektriä yhden näytteen osalta. Värillisellä katkoviivalla on merkitty kohdat, johon tarkasteltavan markkerin massapiikkien tulisi muodostua. Kuvassa on vasemmassa reunassa punainen katkoviiva tarkasteltavan markkerin kohdalla ja korkea homotsygoottipiikki on merkitty sinisellä, johon on muodostunut molempien G-alleelien piikit. Sinisen piikin viereen, punaisen katkoviivan kohdalle, muodostuisi heterotsygooteilla myös T-alleelipiikki. Harmaalla katkoviivalla näkyvät piikit ovat muiden markkereiden piikkejä samalla näytteellä.

Resiinikäsittelystä huolimatta saattaa näytteeseen joskus jäädä suoloja tai muuta ainesta, jotka aiheuttavat taustapiikkejä massaspektriin. Suurista taustapiikeistä ei välttämättä pysty varmasti erottamaan alleelipiikkiä, ja tällöin täytyy tarkastaa koko markkerin onnistumisen taso yleisesti. Suolapiikit saattavat myös osua juuri massapiikin viereen tai sen kohdalle ja aiheuttaa kaksikärkisen tai leveän alleelipiikin. Jos kyseinen markkeri toimii muilla näytteillä hyvin ilman turhia hylkäämisiä ja suuria taustapiikkejä, voidaan tarkastaa klusteroinnissa lähimmät näytteet ja verrata piikkejä toisiinsa. Tämän jälkeen

tehdään näytteen hyväksyminen tai hylkäämispäätös. Jos markkeri ei toimi muilla näytteillä hyvin, verrataan sen tuloksia muihin massa-ajoihin ja tarvittaessa optimoidaan alukkeiden konsentraatiota PCR-reaktiossa. Reaktioon jäänyt suola tai muu aines voi aiheuttaa myös epätasaisen taustan massaspektriin, mikä hankaloittaa piikkien erottamista taustasta (kuva 7).



Kuva 7. Esimerkki heikosta massapiikistä ja huonosta taustasta. Tämän näytteen kohdalla ei voida varmasti sanoa, onko näyte hetero- vai homotsygootti, sillä toisen alleelin kohdalla on pieni ja leveä piikki ja taustapiikit ovat sen kanssa lähes yhtä korkeita. Kaksi korkeaa piikkiä ovat toisen markkerin hyvin onnistuneet massapiikit.

Näytteen onnistumista tarkastellaan myös kaikkien markkerien suhteen jokaisen näytteen osalta. Jos näytteestä on saatu jatkuvasti huonoja tuloksia, voidaan mahdollisesti koko näytteen tulokset poistaa, sillä tulosten luotettavuutta ei voida varmentaa. Jos huonoja tuloksia näytteestä on vain muutama tai ainut laatuaan, täytyy näytettä tarkastella muiden edellä mainittujen parametrien avulla. Epävarmoissa tilanteissa on aina tehtävä näytteen hylkäämispäätös, jotta tulokset eivät olisi virheellisiä. Näyte hylätään kokonaan, jos ainakin puolet sen genotyyppituloksista on epäonnistuneita tai hylättyjä.

#### 5.4 Tilastolliset analyysit

Mahdollisimman usean näytetilavuus-, -säilyvyys ja konsentraatiovariaatioiden luomiseksi ja ajoformaatista johtuen käytettiin jokaisessa ajossa yksi 96-kuoppalevy. Näin tekemällä ei saatu yhtä suuria testejä kaikista muuttujista ja otanta jäi pieneksi. Tästä syystä tulokset laskettiin onnistumisprosentteina ja verrattiin onnistumisia keskenään ilman laajempia tilastollisia analyyseja. Käytetty Agenan protokolla on genotyyppitysyksikössä rutiinikäytössä ja toimivaksi todettu, joten protokollan ja tekijän aiheuttamaa tulosten vaihtelua ei otettu huomioon.

SNP-markkerien tulosten riittävyttä yksilöntunnistamiseen arvioitiin tilastollisesti. SNP-variaatiot ovat yleensä biallelisiä, joten populaatiossa esiintyy joko genotyyppiä *aa*, *ab* tai *bb*. Näistä lasketaan alleelifrekvenssit, joka on *a*:n ja *b*:n esiintyminen tarkasteltavassa populaatiossa, esimerkki kaavassa 1.

$$\begin{aligned} a &= p = 0,7 \\ b &= q = 0,3 \end{aligned} \tag{1}$$

Genotyyppifrekvenssit lasketaan hyödyntämällä binomin neliöyhtälöä (kaava 2).

$$\begin{aligned} (p + q)^2 &= p^2 + 2pq + q^2 \\ f(aa) &= p^2 \\ f(ab) &= 2pq \\ f(bb) &= q^2 \end{aligned} \tag{2}$$

jossa  $p^2$  ja  $q^2$  ovat homotsygoottien genotyyppifrekvenssit ja  $2pq$  on heterotsygoottien genotyyppifrekvenssi. *aa* ja *bb* ovat genotyyppiltään homotsygootteja ja *ab* vastaavasti heterotsygootti.

Yksilöntunnistuksen todennäköisyys laskettiin kertomalla saadut frekvenssit toisillaan, josta saadaan todennäköisyys kahden täysin genotyyppiltään identtisen yksilön löytymiseen kyseisen populaation sisällä valituilla markkereilla. Tämä laskukaava pohjautuu Hardy-Weinbergin lakiin, joka perustuu populaation ideaaliseen tasapainotilaan, jossa populaatiokoko on loputon, yksilöt eivät siirry tai pariudu populaatioiden välillä ja parituminen on täysin satunnaista eikä mutaatioita synny lisää. [23; 24.]



## 6 Optimointi

Testeissä käytettiin yhteensä 46 näytettä, joista tehtiin genotyyppimääritys ID-PANEL 27-plexillä veriliuskoista 100 ja 150 µl:n tilavuuteen liuotettuna ja näistä 0,5 ja 1 µl:n näytetilavuutta käyttäen. Osa näytteistä genotyyppitettiin uudelleen näytesäilyvyyden testaimiseksi. Säilyvyydestä tehtiin pakastetusta TE-puskuriin liuotetusta ja sentrifugoidusta supernatantista.

### 6.1 Näytteensäilytysmenetelmät

Tämän työn alkaessa GeneRisk-verinäytteet olivat valmiiksi kuivattuna Blank Strip (Merck Millipore, Darmstadt, Germany) -näyteliuskoille. Vertailunäytteitä kuivattiin lisää samanlaisille liuskoille työn aikana. Nämä yleiskäyttöiset testiliuskat eivät ole varsinaisesti tarkoitettu veren säilytykseen, mutta ne soveltuivat käyttöön edullisuutensa ansiosta ja niistä oli jo käyttökokemusta toisessa laboratoriossa vastaavanlaisessa menetelmässä. Liuskojen vaihtamista verinäytteiden säilytykseen paremmin soveltuviin vaihtoehtoihin pohdittiin, sillä käsittelyssä liuskat muodostuivat haasteelliseksi sähköisyyden ja leviävän veripölyn vuoksi. Tämä korostui etenkin liuskoja leikattaessa, vaikka liuskat pyrittiin pitämään käsittelyvaiheessa näyteputken sisällä. Veripölyn leviäminen aiheuttaa työskentelyyn ja työntekijälle turhia riskejä, eikä liuskojen leikkaaminen onnistu laimenosputken sisällä sujuvasti työskentelyn kannalta. Näytteet viedään näyteliuskoille yksitellen upottamalla liuska verinäyteputkeen tai pipetoimalla näytettä liuskalle. Liuskat säilytetään näytekohtaisissa putkissa tai muovipusseissa.

Edellä mainituista syistä testiin valittiin myös FTA-kortti (GE Healthcare Life Sciences), joka on tarkoitettu biologisten näytteiden, myös veren, säilytykseen ennen PCR:n tai amplifikaation tekoa. Kortissa on matriisi, joka sitoo nukleiinihappoja ja säilyttää näytteen hyvälaatuisena huoneenlämmössä. Valitussa kortissa on 96 näytepaikkaa, joihin pipetoidaan ja kuivataan näytteet. Kortista leikataan irti leikkurilla (Harris Micro Punches, 2mm) pyöreä pala, jota käytetään jatkokäsittelyyn. Kortin ohjeen mukaan pala huuhdellaan valmistajan omalla pesuliuksella kolmesti ja TE-puskurilla kahdesti, jotta näytteen mukana kuivattava muu aines saataisiin pestyä pois PCR-reaktiota häiritsemästä. Pesu ja eristys voidaan tehdä yhdessä putkessa huoneenlämmössä, ja DNA säilyy matriisiin immobilisoituna. Näitä pesuvaiheita ei kuitenkaan tehty tässä kehitettävässä ID-PANEL-

menetelmän yhteydessä, sillä tarkoitus on pyrkiä mahdollisimman yksinkertaiseen menetelmään ja tehdä vertailu TE-puskuriin liuotetusta verinäytteestä, kuten liuskoille oli tehty. Kortista leikatut näytettä sisältävät palat käsitellään samalla tavalla kuin näyteliuskat PCR:ään. Näytteitä voidaan säilyttää kortilla jopa yli 17 vuotta PCR-monistukseen toimivana ja niitä voidaan poimia kortilta tarpeen mukaan yksittäin vahingoittamatta muita näytteitä. [25.]

## 6.2 Näytemäärä ja näytekonsentraatio

TE-puskuriin liuotettua ja sentrifugoitua näytettä pipetoitiin 384-kuoppalevyille eri määriä PCR-reaktioon. Näytemäärän testausta oli jo aloitettu ennen tätä opinnäytetyötä ja näistä testeistä saadut tulokset otettiin huomioon näytemäärän optimoinnissa. Näytettä pipetoitiin 0,5 ja 1 µl ja vertailtiin niistä saatujen tulosten onnistumista. Aikaisemmissa testeissä oli kokeiltu 2 µl:n näytemäärää. Tähän PCR-reaktioon tarvitaan normaalisti noin 20 ng eristettyä DNA:ta, mutta verinäytteiden DNA-konsentraatiota ei tunneta ennuudestaan. Verinäytteen DNA-pitoisuutta ei ole mahdollista tässä menetelmässä määrittää, sillä se toisi työhön ylimääräisen työskentelyä hankaloittavan vaiheen ja huomattavasti lisäkustannuksia.

Pienen määrän (0,5 µl) pipetoiminen on haastavaa, ja pipetointi on suoritettava erittäin tarkasti. Pipetointi tehdään 8-kanavaisella pipetillä, ja näytemäärä on tarkistettava silmämääräisesti jokaisella pipetointikerralla. Monikanavapipetoinnissa näytteen määrää pipetinkärjissä verrataan toisiinsa ja luotetaan pipettiin, jos määrät näyttävät vastaavan toisiaan. Pipetoinnin helpottamiseksi näytekonsentraatiota muutettiin lisäämällä veriliuskan päälle näyteputken 50 µl enemmän puskuria, eli yhteensä 150 µl ajatuksena kasvattaa pipetointitilavuutta. Myös näistä laimennoksista siirrettiin kuoppalevyille vertailuun 0,5 ja 1 µl näytettä.

Kokoveren määrää oli myös testattu pipetoimalla liuskoille eri tilavuuksia näytettä, mutta tätä osuutta ei jatkettu enää tämän työn osalta näyteliuskoja käytettäessä. Uudelle säilytyskortille testattiin optimaalista pipetointimäärää ja -tapoja korttien käyttöönoton yhteydessä. Näytekorttien ohjeessa kerrotaan maksimi pipetointitilavuuden olevan 5 µl jokaiseen 96 näyterenkaaseen, mutta verinäytteiden säilytykseen haluttiin näytettä mahdollisimman laajalle alueelle näyterenkaaseen, jotta yhdestä näytteestä olisi mahdollista

leikata tarvittaessa ainakin kaksi näyterinkulaa. Tästä syystä näytekortille pipetoitiin noin 15 µl verta näyterenkaisiin.

Näytekortilta testattiin yhden ja kahden 2 mm palan käyttöä liuotukseen ja pala tai palat laimennettiin 100 ja 150 µl:aan TE-puskuria. Näistä laimennoksista pipetoitiin supernatanttia vielä 0,5 ja 1 µl PCR:ää varten. Tässä testin osuudessa käytettiin yhden henkilön tuoretta verinäytettä ja vuoden verran +4 °C:ssa säilytettyä vanhaa verinäytettä sekä tunnettua vertailunäytettä.

### 6.3 Toistettavuus ja näytteen säilyvyys

Osa näytteistä genotyyppitettiin useita kertoja näytesäilyvyyden testaamiseksi. Jos syystä tai toisesta näytteestä ei yksilöntunnistuksessa saada luotettavaa tulosta tai näyte täytyy myöhemmässä vaiheessa tarkastaa yksilöä vastaavaksi, voidaan kyseinen näyte analysoida uudestaan. Tämä voidaan toteuttaa vain, jos näytteen säilyvyydestä voidaan olla varmoja. Näytteen säilyvyydestä tehtiin tässä työssä liuotetuista ja sentrifigoidusta verinäytteistä, joita oli säilytetty -20 °C:ssa. Osa näytteistä sulatettiin ja pakastettiin toistamiseen. Näytteiden onnistumista verrattiin samaan näytteeseen, josta oli pipetoitu saman verran näytettä kuoppalevylle ja joka oli liuotettu samaan tilavuuteen. Alleelitulosten vastaavuus tarkistettiin myös kaikista samoista näytteistä huomioimatta testimuuttujaa.

## 7 Tulokset ja tulosten tarkastelu

Tässä työssä vertaillaan kokonaisuudessaan 44:ää GeneRisk näytettä. Nämä näytteet on aikaisemmin genotyyppimääritetty Illuminan Infinium -menetelmällä CoreExome v1.1 sirulla. Tässä työssä nämä samat näytteet genotyyppimääritettiin suoraan verinäytteestä sekä eristetystä DNA:sta Agenan iPLEX -menetelmällä ja kehitettävällä ID-PANEL SNP-setillä. Testejä tehtiin myös käyttäen FIMM:n omaa kontrollinäytettä sekä vapaaehtoisen henkilön tuoretta ja vanhaa verinäytettä. Näytteet ajettiin kolmessa ajossa 96-formaatissa. GeneRisk näytteitä otettiin näyteliuskoilta käyttöön myös työn aikana, joten kaikki 44 näytettä eivät sisälly kaikkiin kolmeen ajokertaan. Näyteajojen tuloksia arvioitiin ajokohtaisesti sekä kaikkien ajojen tuloksia yhteisesti. Ajojen tuloksia arvioitiin SNP-

markkerien onnistumisien mukaan, näytteiden toimivuuden sekä alleelitulosten onnistumisien mukaan. Ajoihin pyrittiin saamaan mahdollisimman paljon vaihtelua näytemäärän, näytekonsentraation tai näytteen toistettavuuden ja säilyvyyden kannalta.

Kaikista näytteistä ei saatu tuloksia kaikissa ajoissa massaspektrometrillä. Tämä on kuitenkin harvinaisena menetelmälle ominaista osin tuntemattomasta syystä. Ajosta ja näytteistä riippumatta saattavat yksittäisen näytteen tulokset joskus epäonnistua esimerkiksi sirun matriisin laatuongelmien vuoksi. Kuitenkin suurin osa näistä epäonnistuneista tuloksista viittasivat selkeästi näytteestä johtuviin syihin ja nimenomaan silloin, kun näytekonsentraatio oli suuri. Tämä johtuu todennäköisemmin veren sisältämän muun aineksen reaktiota häiritsevistä ominaisuuksista. Ennen tämän työn aloitusta saaduista tuloksista huomattiin suuren määrän verinäytettä häiritsevän huomattavasti analyysiä. Ensimmäisissä testeissä oli käytetty 2 µl puskuriin liotettua verinäytettä ja näistä vain neljä kaikista 36 testatusta näytteestä onnistui. Kuitenkin näistä muutamasta onnistuneesta näytteestä suurin osa SNP-markkereista toimivat, mistä voitiin päätellä markkerien toimivuus, mutta näytetilavuutta tuli pienentää.

GeneRisk-verinäytteiden osalta 1 ja 0,5 µl näytetilavuudesta saatiin molemmista lupaa via tuloksia ja todettiin pienemmän tilavuuden toimivan keskimäärin paremmin. Pipetointitulavuuden kasvattamiseksi testattiin näytteen liotusta suurempaan puskuritulavuuteen (150 µl), jolloin konsentraatio DNA:n ja muiden veressä olevien komponenttien suhteen pienenee. Jälleen pienempi näytetilavuus ja -konsentraatio osoittautui optimaaliseksi suunnaksi näytemäärän suhteen. Testeissä joissa käytettiin 0,5 µl näytettä, huolimatta näytekonsentraatiosta, epäonnistuneita näytteitä oli vain 5 kappaletta 125 näytteestä. Yhtä mikrolitraa käytettäessä epäonnistuneita näytteitä oli 9 kappaletta 66 näytteestä. Testien osoittaessa pienemmän näytekonsentraation ja näytetilavuuden toimivammaksi voidaan olettaa veren sisältämien inhibiittoreiden ja muiden ainesten häiritsevän näytekohtaista onnistumista huomattavasti.

ID-PANELin tulosten varmentamiseksi GeneRisk-näytteistä genotyyppitettiin eristetystä DNA:sta vertailunäytteet. Nämä onnistuivat ajossa odotetusti hyvin, yhden markkerin kohdalla viisi näytettä jäi ilman genotyyppitulosta ja toisesta markkerista jäi yksi genotyyppitulos saamatta. Muut markkerit sekä näytteet toimivat todella hyvin ja saivat hyvät tulokset. Illuminan protokollalla saaduista tuloksista valittiin vastaavat näytteet ja markkerit sekä niiden määritetyt genotyypit. Vertailtaessa samoja näytteitä kahdella protokollalla yhden markkerin kohdalla viisi alleelitulosta poikkesivat kuitenkin toisistaan. Nämä

viisi tulosta olivat ID-PANELissa jokaisessa viidessä saadussa tuloksessa homotsygootti ja Illuminan tuloksissa heterotsygootti. Illuminan tulokset olivat kuitenkin hyvät ja tarkastuksessa laadultaan erinomaiset. ID-PANELia tarkasteltaessa tämä markkeri antaa näille näytteille alukepiikin sekä homotsygoottipiikin. Tämä voi tarkoittaa, että ekstensioaluke ei ole monistanut toista alleelia ollenkaan, minkä johdosta tämä markkeri hylättiin menetelmästä. Myös toinen markkeri hylättiin, sillä siitä saatiin tulokset vain keskimäärin puoleen näytteistä ajoissa.

Näytteiden säilyvyyttä puskuriin liuotettuina testattiin käyttämällä pakastettua näyteliuosta uudessa testiajossa. Osa näytteistä ajettiin kolme kertaa ja näytteet pakastettiin jokaisen ajon välissä. Alustavien tulosten mukaan pakastuskerrat eivät vaikuttaneet näillä testeillä, vain näytekonsentraatio vaikutti tässäkin eniten tuloksiin. Säilyvyydestä ei ole kokemusta näytettä säilyttäessä näyteliuskoilla, mutta kuivana ja valolta suojattuna verinäytteen voidaan olettaa säilyvän liuskalla hyvin. Näytelevyjen ohjeissa luvataan jopa 17,5 vuoden säilyvyys, joten näitä menetelmiä käytettäessä kuivana säilytetystä näytteen säilyvyydestä ei ole mielekäästä tehdä erillisiä testejä.

Näytekorttia käytettäessä testissä oli mukana kaksi näytettä, FIMM:n kontrollinäyte ja tuore ja vanha verinäyte vapaaehtoiselta. Vanha verinäyte oli säilytetty noin vuoden jääkaappilämpötilassa, joten se oli koaguloitunut ja oletettavasti huonontunut näyte. Näytteet pipetoitiin kortille, josta leikattiin 1 ja 2 palaa käyttöön. Nämä liuotettiin kahteen eri puskuritilavuuteen ja käytettiin näytettä 0,5 ja 1 µl. Kahden palan käyttö suurella näytekonsentraatiolla ei toiminut ja niistä tuloksia saatiin vain yhdestä näytteestä tuoretta verinäytettä käyttäessä. Kuitenkin vanhan verinäytteen kohdalla tästä saatiin paremmat tulokset. Yhtä rinkulaa käytettäessä saatiin edellisiä testejä vastaava tulos, pienempi näytekonsentraatio antaa varmemmin paremmat tulokset. Tämä tulos saatiin sekä vanhaa että tuoretta näytettä käyttämällä. Kokonaisuudessaan näytekortilta saatiin hieman onnistuneempia tuloksia, kuin liuskoja käytettäessä.

GeneRisk-näytteistä (44 kpl) laskettiin alleeli- ja genotyyppifrekvenssit, samoin GeneRisk-projektin näytteistä, jotka tähän mennessä on genotyyppimääritetty FIMM:n laboratoriossa, yhteensä noin 4 830 kappaletta. Näistä aineistoista laskettiin vertailuallleelifrekvenssit kyseisessä populaatiossa. Tämän perusteella laskettiin todennäköisyys yksittäisen näytteen genotyyppiprofiilin esiintymiselle tutkittavassa populaatiossa ja sitä kautta todennäköisyys tunnistaa yksilö luotettavasti tällä SNP-markkeriyhdistelmällä. Vertailusta jätettiin pois hylättyjen markkereiden lisäksi yksi sukupuolimarkkeri, jota voidaan

kuitenkin käyttää tarkasteltaessa vastaako tutkittavan näytteen sukupuoli odotettua. Kymmenen markkerin genotyyppejä tarkasteltaessa todennäköisyys kahden identtisen yksilön löytymiseen on noin 0,08 %. Markkerien määrää kasvatettaessa viiteentoista todennäköisyys on enää noin 0,002 %. Käytetyillä 24 markkerilla todennäköisyys pienenee ollen noin 0,000004 %. Todennäköisyys kahden toisiaan vastaavan yksilön löytymiselle on näitä markkereita käytettäessä mitättömän pieni. Tiettyä yksilöä verratessa kyseiseen aineistoon todennäköisyys vastaavaan genotyyppiin pieneni vielä edellisestä. Tämän perusteella yksilön tunnistus on toimiva näillä markkereilla ja onnistuu myös pienillä aineistoilla, vaikka kaikki markkerit eivät toimisikaan jokaisessa analyysissä. Nämä todennäköisyydet koskevat vertailuja tutkimusnäytteiden tietyllä sirutyypillä saatuja tuloksia tarkasteltaessa. Muita sirutyyppejä käytettäessä on mahdollisesti vaihdettava markkereita ja valittava niitä lisää.

## 8 Yhteenveto

Tämän opinnäytetyön tarkoituksena oli optimoida yksilöntunnistusmenetelmää, jotta saataisiin luotettava ja edullinen laadunvalvontaväline genotyyppityöyksikön rutiinikäyttöön. Menetelmänä käytettiin yksilöntunnistusta suoraan verinäytteestä tehtävällä PCR:llä ja vertaamalla niistä genotyyppituloksia aikaisemmin saatuihin GeneRisk-projektin genotyyppitystuloksiin. Tätä menetelmää oli jo testattu lupaavin tuloksin, joten työtä jatkettiin aiheen teoriaan perehtymisellä sekä menetelmän optimoinnilla. Työn suunnitteluvaiheessa oli haasteita, sillä näyteajoja suoritettiin suhteellisen vähän ja testattavia muuttujia oli paljon. Otannan pienen koon takia kaikkia muuttujia ei ollut mielekästä analysoida tilastollisesti kovinkaan tarkasti. Osa tuloksista on näin ollen vielä suuntaa antavia ja menetelmän käyttöönoton yhteydessä voidaan tarvittaessa jatkaa optimointia. Jokaisessa ajossa analysoitiin kuitenkin sama määrä markkereita ja niiden avulla tarkasteltiin jokaista muuttujaa kaikista ajoista. Jotkut näytteet ajettiin useaan kertaan ja niiden tuloksia vertailtiin toisiinsa.

Veriliuskojen käsittelyn haasteellisuudesta johtuen näytteiden säilytykseen haluttiin kokeilla oletettavasti käytännöllisempää vaihtoehtoa. Näytesäilytys liuskoilta kortille lisää menetelmän kustannuksia hieman, mutta vähentää huomattavasti työaikaa. Korttien käsittely on myös turvallisempaa työntekijän kannalta, kun välttyään helposti leviävältä veripölyltä. Työergonomian kannalta näytekorttia on myös helpompi käsitellä, kuin yksittäin pakattuja ja pilkottavia liuskoja. Tämä korostuu erityisesti, kun näytemäärä on suuri.

Yksilöntunnistamisen lähtökohtana on varmentaa kahden näytteen välinen yhteys. Jotta voidaan tehdä luotettavaa yksilöntunnistamista, täytyy vertailunäytteen olla varmasti yksilöstä otettu näyte. DNA-tutkimusnäytteet voivat sekoittua eristysvaiheessa tai muussa näytteenkäsittelyvaiheessa ja tätä ei saa tapahtua vertailunäytteelle, tässä työssä verinäytteelle. Tämän näytesekaannuksen välttämiseksi tulisi verinäyte yksilöntunnistamiseen ottaa henkilöstä samanaikaisesti kuin tutkimusnäyte ja lähettää se suoraan genotyyppityslaboratorioon. Näin ehkäistään väärin tulosten saaminen verinäytteiden sekaantumisen johdosta.

Näytetilavuutta ja -konsentraatiota pienentämällä päästiin parempiin tuloksiin. Tämä kuitenkin tarkoittaa pipetointitulavuuden pienentämistä hyvin pieneksi, mikä on työskentelyn kannalta haastavaa. Näytteen liuotusta voisi jatkossa kokeilla vielä suurempaan puskuritulavuuteen, jolloin pipetointitulavuuden kasvattaminen voisi olla mahdollista ja pipetointivirheistä johtuvaa epävarmuutta pienennettyä. DNA-pitoisuuden mittaamiseen näytteistä ei olisi tässä työssä käytännössä mahdollisuutta, ja se hidastaisi ja hankaloittaisi työtä huomattavasti näytemäärien ollessa suuria sekä nostaisi menetelmän kustannuksia liikaa. Menetelmän kustannukset arvioidaan jäävän alle 5 euron tarkasteltavalta näytteeltä. Tämä on kohtuullinen hinta verrattuna näytteen koko genomin analyysin uusia epävarmoissa tapauksissa, jolloin lisäkulut kohoavat useisiin kymmeniin euroihin tutkittavalta näytteeltä.

Osa markkereista hylättiin ID-PANELista niiden toimimattomuuden takia ja uusien markkereiden lisäämistä voidaan jatkoa varten harkita tarpeen mukaan. Vertailunäytteiden genotyyppitysmenetelmän muuttuessa on aina hyvä tarkastaa markkerien toimivuus ja vastaavuus menetelmissä. Tässä työssä käytetyillä onnistuneiden markkerien lukumäärällä saadaan jo hyvin toimiva ja käytännön vaatimuksia vastaava ID-PANEL-menetelmä käyttöön.

Verinäytteitä säilytetään ja niistä voidaan valmistaa uusi näyte tarvittaessa, mutta tämä on aikaa vievää ja työn ja tulosten saamisen nopeuttamiseksi testattiin jo liuotetun näytteen säilyvyyttä. Näytesäilyvyys todettiin hyväksi, ja epävarman tuloksen sattuessa voidaan käyttää jo liuotettua näytettä uudelleen. Protokollasta johtuvia epäonnistuneita tuloksia ei saada kokonaan poistettua ja näissä tapauksissa käytetään aina muitakin laadunvarmennusmenetelmiä.

Kokonaisuutena työ onnistui hyvin ja menetelmää saatiin kehitettyä eteenpäin tämän työn optimoinnilla. Näytesäilytykseen saatiin uusi menetelmä, joka on kokonaisuuden kannalta huomattavasti parempi vaihtoehto. Tuloksista saatiin luotettavampia pienentämällä näytekonsentraatiota ja kartoittamalla SNP-markkerien toimivuutta. Näiden tulosten pohjalta ID-PANEL-menetelmä voidaan ottaa käyttöön. Menetelmä on jo osoittautunut toimivaksi testattaessa yksittäisten ongelmanäytteiden vastaavuutta oletettuun näytteenantajaan.



## Lähteet

- 1 Bu, Y., Huang, H. & Zhou, G. 2007. Direct polymerase chain reaction (PCR) from human whole blood and filter-paper-dried blood by using a PCR buffer with a higher pH. *Analytical Biochemistry*. Vol. 375, s. 370–372.
- 2 Aula, P., Kääriäinen H. & Leisti J. (toim.). 2002. Perinnöllisyyslääketiede. 2. uudistettu painos. Helsinki: Kustannus Oy Duodecim.
- 3 Aittomäki, K., Moilanen, J., Perola, M. (toim.). 2016. Lääketieteellinen genetiikka. Helsinki: Kustannus Oy Duodecim.
- 4 Portin, Petter. 2006. Ihmisen erilaisuuden geneettinen perusta. *Tieteessä tapahtuu*. Vol. 3, s. 23–27.
- 5 Genome variations. 2003. Verkkodokumentti. Genome News Network. <[www.genomenewsnetwork.org/resources/whats\\_a\\_genome/Chp4\\_1.shtml](http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp4_1.shtml)>. Päivitetty 15.1.2003. Luettu 13.3.2017.
- 6 Bell, J., Spector, T. 2011. A twin approach to unraveling epigenetics. *Trends in Genetics*. Vol. 3, s. 116–125.
- 7 Tekijäinvaihto. 2006. Verkkodokumentti. Solunetti. <[www.solunetti.fi/fi/kehitysbiologia/suvullinen\\_periytyminen/](http://www.solunetti.fi/fi/kehitysbiologia/suvullinen_periytyminen/)>. Luettu 12.1.2017.
- 8 Ross, P., Hall, L., Smirnov, I. & Haff, L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature biotechnology*. Vol. 16, s. 1347–1351.
- 9 Parker, L., London, A. & Aronson, J. 2013. Incidental Findings in the Use of DNA to Identify Human Remains: An Ethical Assessment. *Forensic Sci Int Genet*. Vol. 7, s. 221–229.
- 10 Making SNPs Make Sense. Verkkodokumentti. Learn. Genetics, Genetics science learning center. <[learn.genetics.utah.edu/content/precision/snips/](http://learn.genetics.utah.edu/content/precision/snips/)> Luettu 18.1.2017.
- 11 Lewis, Cathryn M. & Knight Jo. 2012. Introduction to Genetic Association Studies. Cold Spring Harbor Laboratory Press. Vol. 10, s. 297–306.
- 12 Vieira, M., Santini, L., Diniz, A. & Munhoz, C. 2016. Microsatellite markers: what they mean and why they are so useful. *Gene Mol Biol*. Vol. 39, s. 312–328.
- 13 GeneRisk. Verkkodokumentti <[www.generisk.fi/content/generisk-tutkimuksesta](http://www.generisk.fi/content/generisk-tutkimuksesta)>. Luettu 20.1.2017.

- 14 Tietoa verestä. 2017. Verkkodokumentti. Suomen Punainen Risti, Veripalvelu. <[www.veripalvelu.fi/verenluovutus/veren-matka/tietoa-veresta](http://www.veripalvelu.fi/verenluovutus/veren-matka/tietoa-veresta)>. Luettu 10.2.2017.
- 15 Blood Components, Learn More About Each Blood Component. 2017. Verkkodokumentti. American Red Cross. <[www.redcrossblood.org/learn-about-blood/blood-components](http://www.redcrossblood.org/learn-about-blood/blood-components)>. Luettu 12.3.2017.
- 16 Lahermo, Päivi. 2017. Vanhempi tutkija, Suomen Molekyyli lääketieteen Instituutti (FIMM). Helsinki. Keskustelu. 20.1.2017.
- 17 Mercier, B., Gaucher, C., Feugeas, O. & Mazurier, C. 1990. Direct PCR from whole blood, without DNA extraction. Oxford University Press. Nucleic Acids Research. Vol. 18, No. 19.
- 18 McCusker, J., Dawson, M., Noone, D., Gannon, F. & Smith, T. 1992. Improved method for direct PCR amplification from whole blood. Oxford University Press. Nucleic Acids Research. Vol. 20, No. 24.
- 19 Ross, P., Hall, L. & Haff, L. 2000. Quantitative Approach to Single-Nucleotide Polymorphism Analysis Using MALDI-TOF Mass Spectrometry. Bio Techniques. Vol. 29, No. 3.
- 20 Ketola, R., Kostianen, R., Kotiaho, T. & Vainiotalo, P. 2010. Massa-spektrometrian perusteet. Klaukkala: Suomen Massaspektrometrian Seura.
- 21 MassARRAY System Brochure. Agena Bioscience, Inc. San Diego, CA. Luettu 17.1.2017
- 22 Single Nucleotide Polymorphism Detection with the iPLEX Assay and the MassARRAY System. Application note. Agena Bioscience, Inc. San Diego, CA. Luettu 17.1.2017
- 23 Brenner, Charles H. Forensic mathematics of DNA matching. Verkkodokumentti. Forensic mathematics. <[dna-view.com/profile.htm](http://dna-view.com/profile.htm)>. Luettu 19.1.2017
- 24 Hardy-Weinberg equilibrium. 2014. Verkkodokumentti. Scitable by nature education. <[www.nature.com/scitable/definition/hardy-weinberg-equilibrium-122](http://www.nature.com/scitable/definition/hardy-weinberg-equilibrium-122)>. Luettu 19.1.2017
- 25 FTA cards. 2011. FTA technology. GE Healthcare Life Sciences. UK.

**PCR-alukkeet**

Taulukko 1. ID-PANEL forward PCR-alukkeet.

Oligo	Sekvenssi (5'-3') forward
rs6949587	ACGTTGGATGGTTAAAGCTGCCTGCCTTAC
rs10456324	ACGTTGGATGTGGCAGAGATGAAAGCCAAG
rs2280205	ACGTTGGATGGTACTCAAGGTGACGTATGG
rs798981	ACGTTGGATGCATTCAAACCATGGAAGAGC
rs2288958	ACGTTGGATGAACCCATTAAACGCCATCCG
rs4880433	ACGTTGGATGAGGTTGCCGATAAGATAGTC
rs2292971	ACGTTGGATGTGTCCATGACACTGAGTCCA
rs7278737	ACGTTGGATGTTCTCAGACCACCACTTTGC
rs3796181	ACGTTGGATGAGGTTAAGGCCAACATGCAC
rs3102680	ACGTTGGATGGCTAAAAGAAAGGCATATCG
rs7699687	ACGTTGGATGTCCAGAATTACTGTATCAGC
rs10254317	ACGTTGGATGGTTATGCATTAGCCATGCC
rs4779794	ACGTTGGATGGGAAGAGTGCATTCTGAAC
rs1035411	ACGTTGGATGAGCGGAATCACGCTCAGGAT
rs4705	ACGTTGGATGGGTAGCTACTGTGGTCTGTC
rs4927635	ACGTTGGATGGGAAAACGAGTAAAGCTG
rs7145318	ACGTTGGATGCTTATCCATGCCTACTGTCC
rs10901333	ACGTTGGATGAGACTGTCAAAGCCTTCTTC
rs1053985	ACGTTGGATGTTTAGAGTAGCAAGGTTTAC
rs1210894	ACGTTGGATGCTTCAATCGTCCAGATGCAG
rs7190823	ACGTTGGATGTGGGCTGCAGTGCAATTAAC
rs1133076	ACGTTGGATGTGGACCTTTCAGAATCCAAC
rs7904014	ACGTTGGATGTCAATATCTGGGCCGAGTTG
rs892336	ACGTTGGATGAACAGGACAGATGGGCAATG
rs735943	ACGTTGGATGCTTTTCTTTTCTAGCCTGC
rs6756597	ACGTTGGATGAGGATGAGATGGAATGGCAG
rs1132306	ACGTTGGATGATGATACTGACCACATCTGC

Taulukko 2. ID-PANEL reverse PCR-alukkeet.

Oligo	Sekvenssi (5'-3') reverse
rs6949587	ACGTTGGATGCTCTTCATCTGGTTCTCCTC
rs10456324	ACGTTGGATGAAAAATAGGTCCACCCTTAC
rs2280205	ACGTTGGATGGGTTCTATACCAACAGCATC
rs798981	ACGTTGGATGTGTCACAATGCTGGACGATG
rs2288958	ACGTTGGATGAACATCCTCGTTACCTCTCC
rs4880433	ACGTTGGATGCCTTCATCGAACACAGAGAG
rs2292971	ACGTTGGATGCAAGGCGTAAGCAACTTCTG
rs7278737	ACGTTGGATGGGAAGAAGGCATCTTATGTG
rs3796181	ACGTTGGATGGGGATGATCTATACTCTGTG
rs3102680	ACGTTGGATGCATATCTCTTTCAAAGGTC
rs7699687	ACGTTGGATGTACACCCTTCAGAAAATTC
rs10254317	ACGTTGGATGGGGTCATTCGTCCAGCTTG
rs4779794	ACGTTGGATGGCTGATTTCTCACATTCCCG
rs1035411	ACGTTGGATGCTGAAAGTTCCAAGTTTGCG
rs4705	ACGTTGGATGCTCCCAGAGTGTCATTACAG
rs4927635	ACGTTGGATGTGGCAGAATAACCCATGCCC
rs7145318	ACGTTGGATGTGGGAACCAAAGCTATGTGC
rs10901333	ACGTTGGATGATAGCTGAACCGCTGGTCTC
rs1053985	ACGTTGGATGAATATGCAAGCAGCCTTTGG
rs1210894	ACGTTGGATGTTCTGGATCAGCTCCTTGTG
rs7190823	ACGTTGGATGTTACCAAGATGGTAACTCAC
rs1133076	ACGTTGGATGAGGCTGTTTTGCTACTGGTC
rs7904014	ACGTTGGATGAAGAAAGAGCTCTTGGAGTT
rs892336	ACGTTGGATGCTTGAGCACCTGCACTATTG
rs735943	ACGTTGGATGTGCTGCTAACATTAGCTGAC
rs6756597	ACGTTGGATGCAGAGAGTATTAGTAGCCCC
rs1132306	ACGTTGGATGGGGTGCTGGTCAAGATAAAC

## Ekstensioalukkeet

Taulukko 1. Ekstensiossa käytetyt alukkeet massan mukaan järjestettynä, hännät esitetty pienillä ja varsinainen sekvenssi isoilla kirjaimilla.

Oligo	Sekvenssi (5'-3')
rs6949587	TAGCCGCGTCTTCAC
rs10456324	CCACCCTTACCTTCTG
rs2280205	cAAGCTGGGATCCCTC
rs798981	gGATGGGGTTCACGGG
rs2288958	CTCACACAGCAGAGAAAA
rs4880433	gCAGGAGATGTCTGTCTG
rs2292971	cctAATCGACCCATGCCTC
rs7278737	TAAGAAACACTTCCTCTCT
rs3796181	TGTGTGTCTATGTGCTTAG
rs3102680	CTTCAAAGGTCAAAGAGTA
rs7699687	cccgTCTACCTACTGCACCTC
rs10254317	ttccCCAAATCGAAACCCTGC
rs4779794	TCGGCTTCCATTATTTTACTT
rs1035411	CAAGTTTGCGTATAAGACAAT
rs4705	tttTCATTACAGTGAAGACTT
rs4927635	cccttCCCATGCCGCATACCTTC
rs7145318	ggggAAAGCTATGTGCTGCACAG
rs10901333	gacaGCTGGTCTCCCAGGAAGCTTC
rs1053985	ggggAAGCAGCCTTTGGAAAATCT
rs1210894	cctaCCCAAGTTACCCTGAAAACAC
rs7190823	cacgCACATTCTATTTTGACAGGTC
rs1133076	ggcgGTTTTGCTACTGGTCCGGCCT
rs7904014	caAAAGAGCTCTTGGAGTTTGCTCA
rs892336	tctccCCTGCACTATTGCCAGATCAA
rs735943	TTTTGACATGTTTTGTCATCCCAACTA
rs6756597	tcAGGGTACCTGATAGTCCAAACGTGG
rs1132306	gGCTGGTCAAGATAAACTAGGAATCTA