

Noora Salokorpi

VARIAATIOIDEN VAIKUTUS KROMATIININ AVOIMUUTEEN ETURAUHASSYÖVÄN EDESSÄ

Lääketieteen ja terveysteknologian tiedekunta
Kandidaatintutkielma
Huhtikuu 2021

TIIVISTELMÄ

Noora Salokorpi: Variaatioiden vaikutus kromatiinin avoimuuteen eturauhassyövän edetessä
Kandidaatintutkielma
Tampereen yliopisto
Bioteknologian ja biolääketieteen tekniikan tutkinto-ohjelma
Huhtikuu 2021

Eturauhassyöpä on 2:ksi yleisimmin diagnosoitu ja 5:nneksi yleisin kuolinsyy miehillä. Eturauhassyöpä voi olla primaarinen tai kastraatioresistentti, ja tällä tyyppillä on vaikutusta syövän hoitokeinoihin sekä selviytymisennusteeseen. Syövän syntymekanismi tunnetaan osittain, mutta tietämyksessä on vielä monia puutteita. Hoitokeinoja on useita ja niitä kehitetään jatkuvasti lisää, mutta erityisesti metastoivien kastraatioresistenttien eturauhassyövän hoitoennuste on edelleen huono. Syövän lisäksi eturauhaseen voi kehittyä hyvälaatuista liikakasvua, joka diagnosoidaan lähes kaikilla miehillä viimeistään heidän täyttäessään 60 vuotta.

Tässä tutkimuksessa tutkittiin yksittäisten nukleotidien variaatioiden vaikutusta hyvälaatuisen eturauhasen liikakasvun, primaarisen eturauhassyövän sekä kastraatioresistenttien eturauhassyövän etenemiseen. Tutkimuksessa haluttiin selvittää, missä määrin genomien avoimia kohtia voidaan selittää yksittäisten nukleotidien variaatioilla. Yksittäisen nukleotidin variaatioita verrattiin ATAC-sekvensoinnilla saatuihin genomien avoimiin kohtiin ja tutkittiin, onko näiden arvojen välillä yhteyttä. Yksittäisen nukleotidin variaatioita tunnistettiin vertaamalla kunkin potilaan näytteestä saatuja sekvenssejä referenssigenomiin, minkä jälkeen tutkimukseen valittiin mukaan tietyt laadun sekä lukusyvyyden raja-arvot ylittävät muuntuneet nukleotidit. Lisäksi haluttiin tutkia, onko yksittäisen nukleotidin variaatioiden kokonaismäärällä vaikutusta avoimien kohtien määrään.

Tutkimuksessa huomattiin, että vähintään yhden yksittäisen nukleotidin variaation sisältävien avoimien kohtien määrän mediaani nousi hiukan syövän edetessä, mutta kaikkien näytteiden arvojen välillä oli suurta vaihtelua. Pelkän mediaanin perusteella ei kuitenkaan voida tehdä johtopäätöksiä arvojen yhteydestä, vaan aihe vaatii lisätutkimusta. Näytteiden välillä olevien erojen tilastollista merkitsevyyttä tutkittiin t-testeillä, joista saatujen p-arvojen perusteella tulokset olivat johdonmukaisia sen kanssa, että tilastollista merkitsevyyttä ei ole näytteiden välillä. Vertailtaessa kaikkien näytteiden yksittäisten nukleotidien variaatioiden ja avoimien kohtien määrää saatiin selville mahdollinen yhteys näiden arvojen välillä. Tutkimuksen mukaan yksittäisten nukleotidien variaatioiden määrän kasvaessa myös avoimien kohtien kokonaismäärä kasvaa. Asiaa tulisi kuitenkin tutkia lisää, jotta saadaan selville, onko kyse sattumasta vai yhteydestä.

Avainsanat: Yksittäisen nukleotidin variaatio, ATAC-sekvensointi, hyvälaatuinen eturauhasen liikakasvu, primaarinen eturauhassyöpä, kastraatioresistentti eturauhassyöpä

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

ALKUSANAT

Tämä kandidaatintutkielma on tehty Tampereen yliopiston Lääketieteen ja terveysteknologian tiedekuntaan keväällä 2021. Tutkielma oli projektityö Laskennallisen biologian tutkimusryhmälle, ja opin sitä tehdessä paljon suurien datamäärien käsittelystä ja muokkaamisesta.

Haluaisin kiittää tästä työstä erityisesti ohjaajaani Matti Nykteriä, joka antoi minulle tutkimuksen aiheen sekä on ohjannut ja auttanut minua aina tarvittaessa. Lisäksi haluaisin kiittää Tampereen yliopistoa opintojeni mahdollistamisesta.

Tampereella, 23.4.2021

Noora Salokorpi

SISÄLLYSLUETTELO

1. JOHDANTO	4
2. MATERIAALIT JA MENETELMÄT	5
2.1 Aineisto	6
2.2 Genotyyppien todennäköisyyksien laskeminen ja kutsuminen	6
2.3 Mutaatioiden suodattaminen	6
2.4 Variaatioiden vertailu genomien avoimiin kohtiin	7
2.5 T-testi	7
3. TULOKSET	8
4. TULOSTEN TARKASTELU	11
5. YHTEENVETO	12
LÄHTEET	13

1. JOHDANTO

Eturauhassyöpä on 2:ksi yleisin ja 5:nneksi tappavin syöpä miehillä (Bray ym. 2018). Eturauhassyöpä on yleisimmin diagnosoitu syöpä yli 100 maassa, erityisesti Amerikassa, Pohjois- ja Länsi-Euroopassa, Australiassa, Uudessa-Seelannissa sekä Saharan eteläpuolisessa Afrikassa (Bray ym. 2018). Eturauhassyövän löytöä edeltää usein hyvälaatuinen eturauhasen liikakasvu (BPH), joka diagnosoidaan suurimmalla osalla miehistä siihen mennessä, kun he ylittävät 60 vuoden iän (Miah ja Catto 2014). Tutkimuksen mukaan BPH on löydettävissä 83,3 %:ssa diagnosoiduista eturauhassyövistä (Miah ja Catto 2014). Primaarisella eturauhassyövällä (PC) selviytymisen todennäköisyys on korkea, kun taas metastaattisen kastroatioresistentin eturauhassyövän (mCRPC) sekä kastroatioresistentin eturauhassyövän (CRPC) hoitoennuste on heikompi (Wang ym. 2018). Monimuotoisista hoitokeinoista huolimatta mCRPC-diagnoosi tarkoittaa tällä hetkellä melko varmaa kuolemaa (Wang ym. 2018). Tutkimusmenetelmien kehittyessä kuitenkin sekä mCRPC:n että CRPC:n hoitoon kehitetään uusia menetelmiä, joiden avulla selviytymisastetta on saatu nostettua (Komura ym. 2018, Siegel ym. 2018). Uusien hoitokeinojen lisäksi käytössä on edelleen androgeenireseptoreiden vaimentaminen kirurgisella tai lääkkeiden avulla tehtävällä kastroatiolla (Komura ym. 2018).

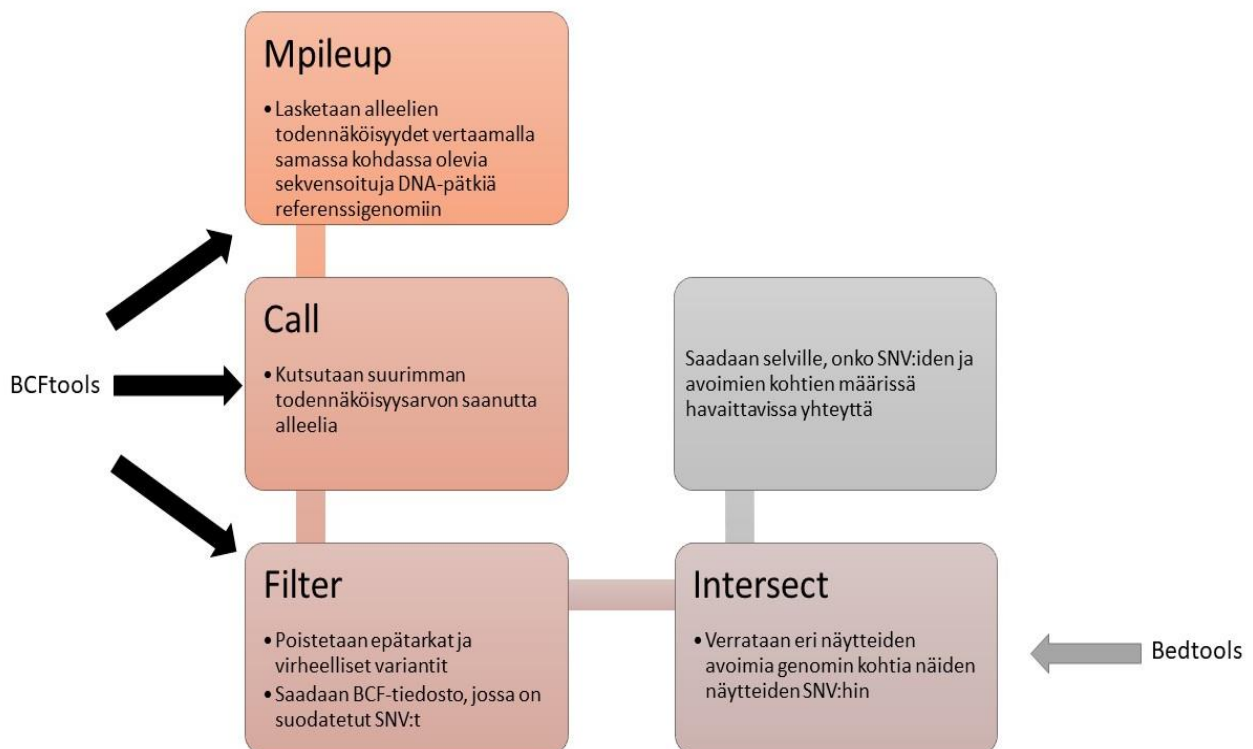
Eturauhassyövän etenemisessä on tutkittu monia erilaisia tekijöitä, jotka voisivat selittää syövän kehittymiselle. Eräs mahdollinen selittävä tekijä on mutaatiot genomissa. Yksittäisen nukleotidin variaatio (SNV) on mutaatiotyyppi, jossa yksittäinen nukleotidi on vaihtunut toiseksi. SNV:t ovat määrällisesti yleisin muutos sekvenssissä, kun vertaillaan genomia toiseen. Jos SNV on peritty sekä esiintyy vähintään 1 %:lla väestöstä, sitä kutsutaan yksittäisen nukleotidin polymorfismiksi (SNP). SNV:t voivat olla mutaatiotyypiltään siirtymiä tai kääntymiä. Siirtymä tarkoittaa sitä, että nukleotidin puriiniemäs on vaihtunut toiseksi puriiniksi tai vastaavasti pyrimidiiniemäs toiseksi pyrimidiiniksi. Kääntymässä muutos on ollut puriinista pyrimidiiniksi tai toisin päin. SNV:t voivat aiheuttaa perinnöllisiä tai somaattisia tauteja, ja sen vuoksi niitä käytetäänkin monissa tapauksissa diagnostiikkatunnuksina. (Spencer ym. 2015) Tässä tutkimuksessa on voitu vertailla sekä SNP:itä että SNV:itä, mutta selvyuden vuoksi kaikkia variantteja kutsutaan SNV:ksi.

Tässä tutkimuksessa käytetty materiaali on tuotettu ATAC-sekvensoinnilla, jonka avulla voidaan tutkia kromatiinin avoimuutta genomien laajuisesti (Buenrostro ym. 2015). ATAC-sekvensointi perustuu hyperaktiiviseen Tn5-transposaasiin, joka kiinnittää sekvenssiadaptoreita kromatiinin avoimiin kohtiin (Buenrostro ym. 2015). Menetelmää voidaan käyttää halvempaan vaihtoehtona esimerkiksi DNA-sekvensoinnille, jos olennaista on avoimien kromatiinialueiden tunnistaminen.

Sekvensointidata koostuu monista toistensa kanssa linjassa olevista sekvensoiduista DNA-pätkistä, joiden alleelit voivat olla erilaisia genomissa samassa kohdassa. Alleelien erilaisuus voi johtua esimerkiksi perinnöllisistä syistä, somaattisista muutoksista tai sekvensointivirheistä. Sekvensointidataa analysoitaessa verrataan löydettyjen nukleotidien laatua ja määrää referenssigenomiin, ja siten koostetaan tiedot jokaisen kohdan genotyyppien todennäköisyyksistä (Li 2011). Kun suurimmalla todennäköisyydellä löytyvä genotyyppi määritetään jokaisessa kohdassa, saadaan tuloksena tiedosto, joka sisältää tiedot jokaisen kohdan referenssi- sekä vaihtoheitoisalleelista (Danecek ym. 2021). Tämän tiedoston arvojen avulla SNV:iden yleisyyttä ja sijaintia genomissa on sujuvaa tutkia.

2. MATERIAALIT JA MENETELMÄT

Tämän tutkimuksen työnkulku on nähtävillä kuvassa 1. Tästä kuvasta nähdään käytetty ohjelma, annetut käskyt sekä käskyjen tiivistetty toiminta. Mpileup- sekä call-käskyt käydään kohdassa 2.2, filter-käsky kohdassa 2.3 ja intersect-käsky kohdassa 2.4.



Kuva 1. Tutkimuksen työnkulku

2.1 Aineisto

Tämän tutkimuksen aineisto on saatu Matti Nykterin Laskennallisen biologian tutkimusryhmältä. Aineisto on tuotettu ATAC-sekvensoinnilla. ATAC-sekvensoitu raaka-aineisto on karsittu Trim Galore -ohjelman versiolla 0.5.0 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; viitattu 31.3.2021). Tämän jälkeen aineisto on linjattu Bowtie 2 -ohjelman versiolla 2.3.4.1 (Langmead ja Salzberg 2012), ja linjaaminen tehtiin vertaamalla aineistoa ihmisen hg38-referenssigenomiin. Saatua aineistoa on vielä suodatettu, järjestelty ja indeksoitu käyttäen Samtools-ohjelman versiota 1.8 (Danecek ym. 2021). Lopuksi aineistosta on poistettu kaksoiskappaleet Picard MarkDuplicates -työkalulla (<https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard->; viitattu 21.4.2021), minkä jälkeen aineisto on ollut käyttövalmis tätä tutkimusta varten.

Tämä tutkimus käsittelee 11 BPH-, 16 PC- sekä 11 CRPC-näytettä.

2.2 Genotyyppien todennäköisyyksien laskeminen ja kutsuminen

Työvaiheen tarkoituksena oli löytää kaikki muuntuneet nukleotidit, kun verrattiin jokaista näytettä hg38-referenssigenomiin. Genotyyppien todennäköisyyksien laskeminen sekä kutsuminen toteutettiin BCFtools-ohjelmalla (Danecek ym. 2021). Todennäköisyyksien laskemisen suurimmaksi mahdolliseksi lukusyvyydeksi asetettiin arvo 8000, jotta jokainen genomien kohta käydään riittävän tarkasti läpi. Lukusyvyys kertoo, kuinka usea sekvensoitu DNA-pätkä on genomien tietyssä kohdassa otettu huomioon. Tässä tutkimuksessa keskityttiin vain SNV:iden etsintään liittymien ja häviämien sijaan, mikä teki tutkimuksesta suoraviivaisempaa.

Tuloksena saatiin jokaiselle näytteelle binary variant calling format -tiedosto (BCF), joka sisältää tiedot esimerkiksi referenssialleelista sekä vaihtoehtoisesta alleelista, saadun alleelin laadusta sekä sen kromosomaalisesta kohdasta. Näiden tietojen avulla mutaatioiden suodattaminen oli mahdollista.

2.3 Mutaatioiden suodattaminen

Genotyyppien todennäköisyyksien laskemisen ja kutsumisen lisäksi tutkimuksessa oli tärkeää suodattaa SNV:itä mahdollisimman tarkkoilla kriteereillä. Mutaatioiden suodattamisessa hyödynnettiin riittävän laadun sekä lukusyvyyden ominaisuuksia. Laadun laskemiseen käytettiin Phred-laaturuostoa, jossa emästen virheiden todennäköisyyksiä verrattiin logaritmisella asteikolla (Liao ym. 2017). Laskutapa on esitetty alla olevassa kaavassa, jossa Q on Phredin laaturuostotulos ja P on virheen todennäköisyys.

$$Q = -10 \log_{10} P$$

Laatu on riittävä silloin, kun nukleotidien siirtymien ja kääntymien välinen suhde on koko genomin sekvensointiaineistossa 2,0–2,2 (Baes ym. 2014). Vaikka tässä tutkimuksessa on käytetty koko genomin sekvensoinnin sijasta ATAC-sekvensointia, tätä samaa suhdetta hyödynnettiin myös tässä tutkimuksessa. Jokaisen näytteen kääntymien ja siirtymien välinen suhde tarkastettiin, jotta voitiin varmistaa, että laaduksi valittiin sopiva arvo.

Toisena suodatuskriteerinä käytettiin lukusyvyyttä. Liian matala lukusyvyys mahdollistaa sekvensoinnissa syntyvien virheiden siirtymisen tutkimusaineistoon, sillä vertailukohtien määrä on tällöin liian matala ollakseen luotettava. Tutkimuksen mukaan eri lukusyvyyksiä verratessa lukusyvyyden alaraja 10 antoi luotettavimmat tulokset (Jiang ym. 2019).

Käytettyjen suodatuskriteerien avulla SNV:istä jäi jäljelle tutkimuksen kannalta oleelliset SNV:t. Jokaiselle näytteelle laskettiin SNV:iden kokonaismäärä.

2.4 Variaatioiden vertailu genomin avoimiin kohtiin

Tutkimuksessa verrattiin saatuja SNV:itä avoimeen genomiin. Vertailu tehtiin Bedtools-ohjelmalla (Quinlan 2014).

Tutkimukseen valittiin mukaan vain avoimet genomin kohdat ja muut osat rajattiin pois. Jokaisen näytteen SNV:itä verrattiin näiden näytteiden avoimiin kohtiin, ja tuloksena saatiin lista niistä kohdista, joiden avoimissa kohdissa oli yksi tai useampi SNV sekä vastaavasti niistä kohdista, joissa ei näitä muutoksia havaittu.

2.5 T-testi

Aiemmissa vaiheissa saatujen tulosten tilastollista merkitsevyyttä haluttiin myös selvittää. Testaus toteutettiin Welchin kahden näytteen t-testillä, jossa jokaista näytteryhmää verrattiin pareittain toisiinsa.

Testin tuloksena saatiin jokaiselle parille p-arvo, joka kertoo, kuinka todennäköisesti nollihypoteesi on totta. Tässä tutkimuksessa nollihypoteesina oli, että eri näytteryhmien välisillä eroilla ei ole tilastollista merkitsevyyttä. Lisäksi t-testin tuloksena saatiin t-arvo, joka kertoo näytteiden erojen koon suhteessa näytetietojen vaihteluun sekä vapausasteluku, joka kertoo niiden riippumattomien tietokappaleiden määrän, jotka menivät saadun arvion laskemiseen.

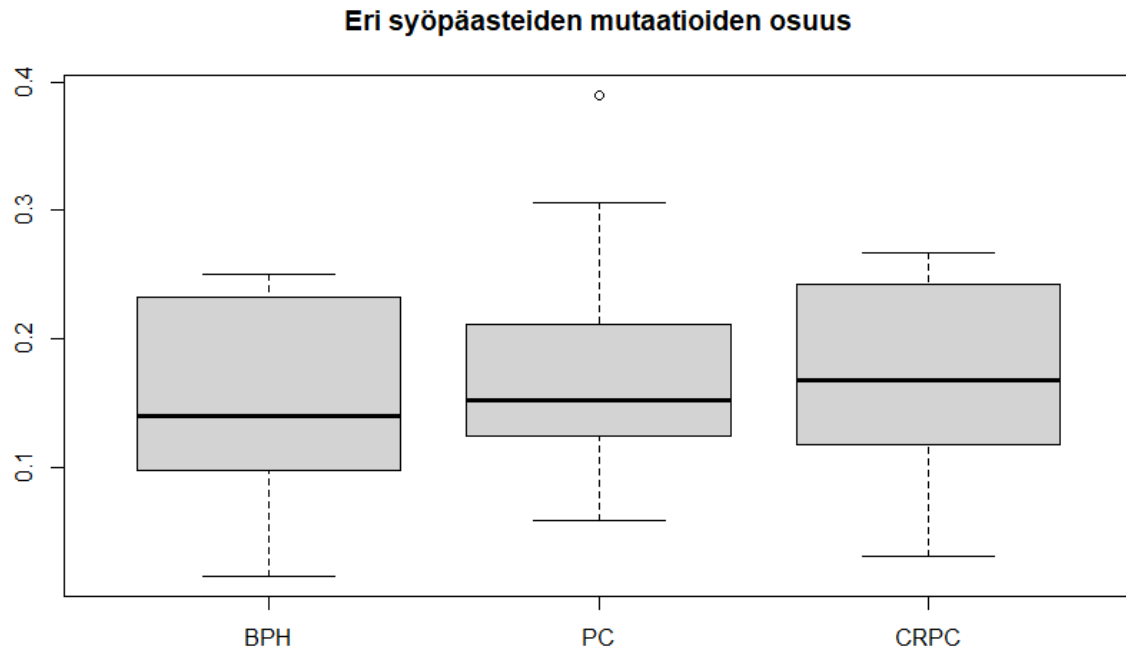
3. TULOKSET

Aineiston käsittelyvaiheessa saatiin jokaiselle näytteelle suodatettujen SNV:iden sekä avoimien kohtien määrä. BPH:n, PC:n sekä CRPC:n pienimmät ja suurimmat arvot voidaan nähdä taulukosta 1. SNV:iden sekä avoimien kohtien pienimmät määrät löytyivät BPH-näytteistä, ja suurin SNV:iden määrä CRPC-näytteistä. Suurin avoimien kohtien määrä löytyi PC-näytteistä. Sekä SNV:iden että avoimien kohtien kokonaismäärissä oli kaikkien näytteiden välillä suurta vaihtelua.

Taulukko 1. SNV:iden ja avoimien kohtien määrien vaihteluväli eri näytetyypeissä

Näytetyyppi	SNV:iden määrä	Avoimien kohtien määrä
BPH	4 187–33 763	39 507–77 018
PC	11 465–61 609	44 138–81 806
CRPC	5 606–93 310	47 686–72 859

Tutkimuksessa haluttiin tietää, kuinka monessa avoimessa kohdassa on vähintään yksi SNV. Tämän tiedon avulla voitaisiin mahdollisesti selvittää yhteys variaatioiden ja syövän etenemisen välillä. SNV:n sisältävien avoimien kohtien määrän vertailu kaikkiin avoimiin kohtiin nähdään kuvassa 2. Kuvan laatikko-janakuviot osoittavat tämän vertailun prosentuaaliset arvot BPH:n sekä eturauhassyövän eri vaiheissa. Kuvasta nähdään, että prosenttiosuudet vaihtelevat eri syöpäasteiden välillä. Kuvasta voidaan havaita, että mediaanissa tapahtuu syövän edetessä pientä kasvua, mutta muiden arvojen tulkinta ei ole yhtä selkeää suuren vaihtelun vuoksi.



Kuva 2. SNV:n sisältävien avoimien kohtien prosenttiosuus avoimesta genomista BPH-, PC- sekä CRPC-näytteissä

Kuvan 2 laatikko-janakuvion minimi-, maksimi-, mediaani-, keski- sekä 1:sen ja 3:nnen kvartaalin arvot nähdään taulukosta 2. Prosenttiosuuksien mediaani, kuten kuvasta 2 nähdään, nousee taulukon mukaan syövän edetessä, mutta keskiarvon suurin arvo, joka on 17,19 %, nähdään PC-näytteissä. Myöskin maksimi-arvo, joka on 38,99 %, on suurin PC-näytteissä. Minimiarvo on pienin BPH-näytteessä, ja tämä arvo on 1,51 %. Taulukosta 2 sekä kuvasta 2 voidaan siis nähdä, että pienin arvo löytyy BPH-näytteestä ja suurin arvo PC-näytteestä, mutta mediaania katsoessa CRPC-näytteellä on suurin arvo, joka on 16,78 %. Prosenttiosuuksien vaihtelu on suurta kaikkien näytetyyppien välillä.

Taulukko 2. Näytetyyppien laatikko-janakuviosta saadut arvot

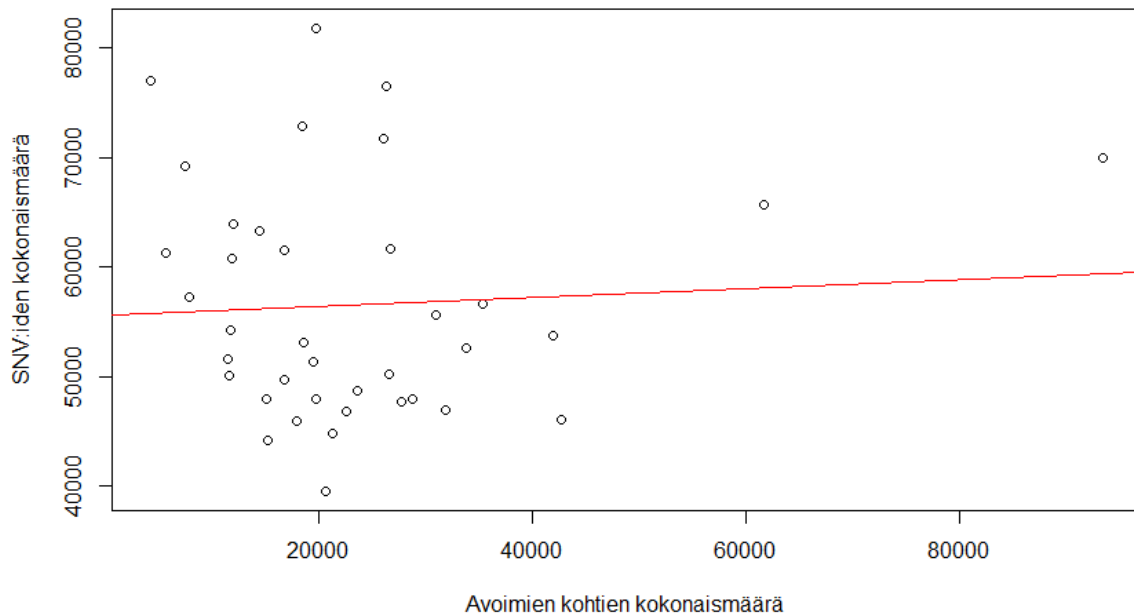
Näytetyyppi	Minimi	1. kvartaali	Mediaani	Keskiarvo	3. kvartaali	Maksimi
BPH	0,01510	0,09775	0,14040	0,15079	0,23210	0,25000
PC	0,0582	0,1249	0,1519	0,1719	0,2107	0,3899
CRPC	0,0309	0,1178	0,1678	0,1716	0,2421	0,2668

Saatujen arvojen perusteella haluttiin tutkia, onko ryhmien välillä tilastollisesti merkitsevää eroa. Saadut p-arvot nähdään taulukosta 3. Samasta taulukosta nähdään myös t-arvot sekä vapausasteluvut. P-arvon yleisenä raja-arvona pidetään arvoa 0,05, ja kaikkien näytteryhmien vertailussa p-arvo oli selkeästi yli tämän arvon. Kun p-arvo on yli 0,05, on aineisto johdonmukainen nollahypoteesin kanssa. Lisäksi kaikkien näytteiden vertailussa t-arvo oli lähellä arvoa 0. Mitä lähempänä saatu t-arvo on arvoa 0, sitä suuremmalla todennäköisyydellä tilastollista merkitsevyyttä ei ole.

Taulukko 3. Eri näytetyyppien välillä tehdyistä t-testeistä saadut arvot

Näytetyypit	P-arvo	T-arvo	Vapausasteluku
BPH ja PC	0,5314	-0,63599	21,761
PC ja CRPC	0,9915	0,010776	21,757
BPH ja CRPC	0,5711	-0,57596	20

Avoimien kohtien, joissa on vähintään yksi SNV, lisäksi haluttiin tutkia, voiko SNV:iden kokonaismäärässä ja kaikkien avoimien kohtien määrässä nähdä korrelaatiota. Kuvassa 3 on kaikkien näytteiden yhteinen hajontakuvio. X-akselilla nähdään avoimien kohtien kokonaismäärä ja Y-akselilla SNV:iden kokonaismäärä. Kuvaan on myös lisätty kuvaaja, joka osoittaa arvojen kulmakertoimen. Kuvaajasta nähdään, että se on lievästi nouseva. Tämä tulos viittaa siihen, että SNV:iden kokonaismäärän kasvaessa myös avoimien kohtien määrä kasvaa. Kuten kuvasta näkee, arvojen vaihtelu on kuitenkin suurta, ja sen vuoksi tulkinta ei ole yksiselitteinen.



Kuva 3. SNV:iden kokonaismäärän ja avoimien kohtien hajontakuviokuva kaikille näytteille

4. TULOSTEN TARKASTELU

Tutkimuksessa huomattiin, että sekä SNV:iden että avoimien kohtien määrässä oli suurta vaihtelua eri yksilöiden välillä. Koska suurin osa ihmisen genomin variaatioista on vanhemmilta perittyjä sukulinjan mutaatioita, on vaihtelun määrä tähän tietoon suhteutettuna erittäin suurta. Jotta tämän tuloksen merkitystä voidaan ymmärtää paremmin, olisi olennaista tutkia, minkä geenien säätelyalueilla mutaatiot ovat tapahtuneet ja onko muutoksilla vaikutusta geenien ilmentymiseen. Tätä asiaa voisi tutkia RNA-sekvensointidatan avulla laajemmin. Jos mutaatioita löydetään syövän kehittymiseen liittyvistä geneeistä tai niiden säätelyalueilta, voi yhteyden löytäminen olla varmempaa.

Tutkimuksessa huomattiin, että SNV:iden määrän prosenttiosuuden mediaani kasvaa hiukan syövän edetessä, mutta suurin arvo nähdään kuitenkin PC-näytteessä CRPC-näytteen sijasta. Koska SNV:iden määrässä oli vaihtelua, voidaan mediaania pitää hyvänä arvona sen vuoksi, että siihen ei vaikuta yksittäiset suuremmat tai pienemmät arvot. Näillä tiedoilla voidaan tehdä oletus, että SNV:iden määrä kasvaa eturauhassyövän edetessä, mutta yksilöiden välillä vaihtelua on. Vielä ei kuitenkaan tiedetä, onko SNV:iden määrän kasvulla vaikutusta syövän kehittymiseen, vai ovatko ne toisistaan riippumattomia tekijöitä. Saatujen tulosten varmentamiseksi tulee vielä tarkemmin tut-

ku, onko SNV:ien määrän vaihtelu biologisista vai aineiston tuotanto- tai analyysivaiheessa syntyneistä teknisistä syistä johtuvaa. Eri ohjelmilla on erilaisia laskukaavoja genotyyppien todennäköisyyksien laskemiselle sekä kutsumiselle, mikä voi johtaa erilaisiin tuloksiin samoilla aineistoilla. Olennaista olisi myös tutkia, kuinka suuri osuus löydetyistä varianteista on perinnöllisiä SNP:itä, ja mikä niiden merkitys on kromatiinin avoimuudelle ja mahdollisesti syövän etenemiselle.

Eri näytteiden välillä tehtyjen t-testien perusteella p-arvo oli suurempi kuin yleisenä raja-arvona toimiva 0,05 jokaisella näyteparilla. Tämä tulos tarkoittaa sitä, että tutkimustulos on johdonmukainen nollahypoteesin kanssa, eli että eri näyteryhmien välillä ei ole tilastollisesti merkitsevää eroa. Tulos ei kuitenkaan tarkoita, että nollahypoteesi olisi totta. Johdonmukaisuutta nollahypoteesin kanssa tukee myös jokaisen näyteparin t-arvo, joka oli kaikilla pareilla lähellä 0:aa oleva arvo. Nollahypoteesin toteaminen todeksi vaatii lisätutkimusta.

Tuloksissa vertailtiin myös kaikkien näytteiden SNV:iden kokonaismäärää avoimien kohtien määrään. Saatiin selville, että SNV:iden määrän kasvaessa myös avoimien kohtien määrät kasvavat. Pelkästään tällä tiedolla ei kuitenkaan voida vielä vetää johtopäätöksiä asiasta, sillä näin pienellä näyteotannalla kyse voi olla myös sattumasta. Saadut tulokset kuitenkin viittaavat siihen, että SNV:iden ja avoimien kohtien kokonaismäärällä voi olla yhteys, ja tätä asiaa tulisi tutkia lisää sen merkittävyyden selvittämiseksi. Asian tutkiminen olisi mahdollista yhdistämällä varianttitietoa transkriptiotekijöiden sitoutumisanalyysiin, jolloin voidaan laskennallisesti analysoida, mitkä löydetyistä varianteista tuhoavat tai tuottavat uusia transkriptiotekijöiden sitoutumispaikkoja. Tämän tutkimusmenetelmän avulla saadaan tarkempaa tietoa SNV:iden vaikutuksesta ja voidaan keskittyä analyysissä tarkemmin niihin variantteihin, joilla ennustetaan olevan suora vaikutus transkriptiotekijöiden säätelyyn.

5. YHTEENVETO

Tuloksista saatavana loppupäätelmänä voidaan todeta, että SNV:iden määrä avoimen genomin alueilla on hyvin vaihteleva eri näytteiden välillä. Vaihtelu oli odotettua suurempaa, joten sen lisätutkimus voisi tarjota uusia näkökulmia perittyjen sekä somaattisten mutaatioiden vaikutuksista. Samassa kohdassa tapahtuvia, ei-koodaavan alueen somaattisia pistemutaatioita ei ole juurikaan raportoitu kirjallisuudessa, joten lisätutkimus on myös tämän johdosta tärkeää, jotta saadaan luotettavia tulkintoja tuloksista.

SNV:iden määrän mediaani genomin avoimissa kohdissa kasvoi syövän edetessä, mutta näillä tiedoilla ei vielä voida todeta, onko kyse sattumasta vai kausaliteetista. Nollahypoteesia tukeva t-

testin tulos antaa viitteitä siihen suuntaan, että tilastollista merkitsevyyttä erojen välillä ei ole, mutta tätäkään ei voida todeta varmaksi yksittäisellä testillä. Tämän vuoksi useammat tilastolliset testit voisivat tarjota uusia tietoja ja tulkintoja. Vertailua eri näytteidien välillä voitaisiin myös toteuttaa eri ohjelmilla, jolloin voitaisiin tutkia käytetystä ohjelmasta johtuvan teknisen virheen mahdollisuutta.

Tutkimuksessa nousi kysymyksiä koskien SNV:iden kromosomaalisen sijainnin vaikutusta BPH:n ja eturauhassyövän kehittymiseen. Tätä asiaa ovat tutkineet esimerkiksi Zhang ym. (2018), jotka löysivät mutaatioita kasvunrajoitegeeneistä, kuten *PKD1*- sekä *KLK1*-geeneistä, sekä erilaisilta geenien säätelyalueilta. Näiden geenien merkitys voi olla suuri eturauhassyövän kehittymisessä (Zhang ym. 2018). Tässä tutkimuksessa löytyneistä SNV:eistä voidaan mahdollisesti löytää uusia kromosomaalisia alueita, joissa variaatiot vaikuttavat eturauhassyövän syntyyn ja etenemiseen. Tämä tutkimus voi tarjota sekä tukea vanhoille tutkimuksille että täysin uutta tietoa, joka voi mahdollistaa uudenlaisia ulottuvuuksia eturauhassyövän tutkimukseen.

LÄHTEET

Baes, C.F., Dolezal, M.A., Koltjes, J.E., ym. (2014): Evaluation of variant identification methods for whole genome sequencing data in dairy cattle. *BMC genomics* **15**(1), 948.

Bray, F., Ferlay, J., Soerjomataram, I., ym. (2018): Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**(6), 394-424.

Buenrostro, J.D., Wu, B., Chang, H.Y., ym., (2015): ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Wiley* **109**(1), 21.29.1-21.29.9.

Danecek, P., Bonfield, J.K., Liddle, J., ym., (2021): Twelve years of SAMtools and BCFtools. *GigaScience* **10**(2), 1-4.

Jiang, Y., Jiang, Y., Wang, S., ym. (2019): Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC bioinformatics* **20**(1), 556.

Komura, K., Sweeney, C.J., Inamoto, T., ym. (2018): Current treatment strategies for advanced prostate cancer. *International journal of urology* **25**(3), 220-231.

Langmead, B. & Salzberg, S.L. (2012): Fast-gapped read alignment with Bowtie 2. *Nature methods* **9**(4), 357-359.

Li, H. (2011): A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Oxford Journals* **27**(21), 2987–2993.

- Liao, P., Satten, G.A. & Hu, Y. (2017): PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genetic epidemiology* **41**(5), 375-387.
- Miah, S. & Catto, J. (2014): BPH and prostate cancer risk. *Indian J Urol.* **30**(2), 214-218.
- Quinlan, A.R. (2014): BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics* **47**(1), 11.12.1-11.12.34.
- Siegel, R.L., Miller, K.D. & Jemal, A. (2018): Cancer statistics, 2018. *CA: a cancer journal for clinicians* **68**(1), 7-30.
- Spencer, D.H., Zhang, B. & Pfeifer, J., (2015): Chapter 8 - Single Nucleotide Variant Detection Using Next Generation Sequencing. *Clinical Genomics*, Elsevier Inc, 109-127.
- Wang, G., Zhao, D., Spring, D.J., ym., (2018): Genetics and biology of prostate cancer. *Cold Spring Harbor Laboratory Press* **32**(17-18), 1105–1140.
- Zhang, Z., Wu, H., Zhou, H., ym. (2018): Identification of potential key genes and high-frequency mutant genes in prostate cancer by using RNA-Seq data. *Oncology letters* **15**(4), 4550–4556.