

lina Koivisto

MIKROYMPÄRISTÖN MUUTOSTEN LASKENNALLINEN ANALYYSI ETURAU- HASSYÖVÄN ETENEMISESSÄ

Lääketieteen ja terveysteknologian tiedekunta
Kandidaatintutkielma
Huhtikuu 2021

TIIVISTELMÄ

lina Koivisto: Mikroympäristön muutosten laskennallinen analyysi eturauhassyövän etenemisessä

Kandidaatintutkielma

Tampereen yliopisto

Bioteknologian ja biolääketieteen tekniikan tutkinto-ohjelma

Huhtikuu 2021

Eturauhassyövällä on useita eri muotoja, ja potilaan selviytymisennuste vaihtelee suuresti sen mukaan, minkä tyyppinen eturauhassyöpä hänelle diagnosoidaan. Primääriä eturauhassyöpää pystytään hoitamaan tehokkaasti hormonihoidoilla ja siihen sairastuneella on paljon parempi selviytymisennuste kuin kastraatioresistenssiin syöpätyyppiin sairastuneella. Kastraatioresistenssi eturauhassyöpä on kehittynyt androgeeneista riippumattomaksi, minkä takia sitä ei voida hoitaa samoin kuin primääriä syöpätyyppejä. Androgeeniriippumattomuuden syntymisen mekanisme ei vielä tunneta, mutta kasvaimen kehityksen aikana tapahtuvien muutosten selvittäminen voi antaa tähän vastauksia. Tässä tutkielmassa pyrittiin kartoittamaan eroavaisuuksia hyvänlaatuisen eturauhasen liikakasvun mikroympäristön, primäärin eturauhassyöpäkasvaimen mikroympäristön ja kastraatioresistentin eturauhassyöpäkasvaimen mikroympäristön solukoostumusten välillä.

Erilaiset laskennallisen biologian menetelmät ovat kehittyneet nopeaa vauhtia viimeisten vuosikymmenien aikana ja ne tarjoavat uudenlaisia mahdollisuuksia tutkimukseen. Tässä työssä hyödynnettiin CIBERSORTx dekonvoluutio-ohjelmaa näytteiden solukoostumuksen määrittämiseen RNA-sekvensointidatasta. Tuloksista piirrettiin laatikkojanakuvaajia näyteryhmittäin ja etsittiin tilastollisten testien avulla eroavaisuuksia analyysissä käytettyjen immuunisolutyyppien suhteellisissa määrissä näyteryhmien välillä.

Tilastollisesti merkitseviä eroja löytyi seitsemän immuunisolutyyppin suhteellisissa määrissä, ja näistä makrofagien ryhmän 2 solujen ja plasmaselujen kohdalla ero löytyi primäärin eturauhassyövän ja kastraatioresistentin eturauhassyövän väliltä. Jatkotutkimukset näiden kahden immuunisolutyyppin roolista eturauhassyövässä, voi auttaa selvittämään androgeeniriippumattomuuden syntymismekanismeja. Dekonvoluutiomenetelmiä käytettäessä tulosten tarkkuuteen vaikuttaa kuitenkin useat asiat, ja tässä työssä esitettyjen tulosten varmistamiseksi olisi syytä toteuttaa analyysi isomalla näytekoolla ja käyttäen eturauhassyöpäkudoksesta kehitettyä signatuurimatriisia, sekä mahdollisesti useampia dekonvoluutio-ohjelmia.

Avainsanat: eturauhassyöpä, eturauhasen hyvänlaatuinen liikakasvu, primääri eturauhassyöpä, kastraatioresistentti eturauhassyöpä, mikroympäristö, dekonvoluutio, solukoostumus

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

ALKUSANAT

Tämä kandidaatintutkielma on tehty Tampereen yliopiston Lääketieteen ja terveysteknologian tiedekuntaan osana bioteknologian ja biolääketieteen tekniikan kandidaatin tutkintoa keväällä 2021. Kandidaatintyön tekeminen on ollut hyvin antoisaa ja opettanut paljon tutkimuksen tekemisestä.

Haluan kiittää ohjaajaani Matti Nykteriä mielenkiintoisesta aiheesta ja selkeistä ohjeistuksista sekä nopeista vastauksista aina ohjausta tarvitessani. Lisäksi erityiskiitos koko Computational Biology tutkimusryhmälle heidän datansa käytöstä ja apulaisprofessori Katri Lindforsille hänen tarjoamastaan ohjauksesta kandiseminaareissa. Haluan myös osoittaa kiitokseni perheenjäsenilleni ja opiskelukavereilleni, jotka ovat tukeneet minua tätä työtä tehdessäni.

Tampereella, 25.4.2021

Iina Koivisto

SISÄLLYSLUETTELO

1. JOHDANTO	4
2. MATERIAALIT JA MENETELMÄT	6
2.1 Näytteet ja Deseq2 normalisointi.....	6
2.2 Mikroympäristön solukoostumuksen määrittäminen	6
2.3 Dekonvoluution tulosten käsittely ja visualisoiminen	7
3. TULOKSET JA NIIDEN TARKASTELU.....	8
3.1 Dekonvoluution tulokset.....	8
3.2 Kruskal-Wallis testin tulokset	10
3.3 Mann-Whitney testin tulokset	10
4. POHDINTA	12
5. LÄHTEET	13
6. LIITE 1: R-KOODI.....	14

1. JOHDANTO

Pelkästään vuonna 2020 1 414 259 ihmistä sairastui eturauhassyöpään ja 375 304 kuoli sen seurauksena (Sung ym. 2021). Eturauhassyövän kuolleisuus onkin miehillä kaikista syöpätyypeistä neljänneksi korkein. Kuten muillakin syöpätyypeillä, myös eturauhassyövällä on monia erilaisia muotoja, ja varsinaisten kasvainten lisäksi erityisesti ikääntyvillä miehillä tavataan paljon hyvänlaatuisia eturauhasen liikakasvua. Niin eturauhasen liikakasvuun kuin myös paljon sitä vaarallisemman eturauhassyövän kehitykseen vaikuttaa androgeenit.

Androgeenit aktivoivat androgeenireseptoreita, minkä seurauksena solu saa jakautumiselle ja kasvulle välttämättömiä signaaleja. Androgeenien tuotto on tarkasti säädeltyä ja normaalitilanteessa ne stimuloivat eturauhasen kasvua lähinnä sikiön kehittyessä ja esipuberteetin aikana. Primäärissä eturauhassyövässä homeostasia on häiriintynyt, mutta kasvain on riippuvainen androgeeneistä. Tämän johdosta primääriä syöpätyyppejä voidaan hoitaa tehokkaasti hormonihoidoilla, jotka perustuvat kasvaimen androgeenien saannin estämiseen. Primääri kasvain voi kuitenkin kehittyä androgeeneistä riippumattomaksi, jolloin hormonihoitot eivät enää toimi. Tämän tyyppistä eturauhassyöpää kutsutaan kastraatioresistentiksi eturauhassyöväksi. Kastraatioresistentti eturauhassyöpä levittää yleensä myös etäpesäkkeitä, mikä vaikeuttaa sen hoitoa entisestään. Selviytymisennuste etäpesäkkeitä levittävälle kastraatioresistentille eturauhassyöpäpotilaalle on vain 15–36 kuukautta (Crawford ym. 2017).

Eturauhassyöpäkasvaimen androgeeniriippumattomuuden kehittymismekanismeja ei tiedetä tarkalleen, mutta sen selvittäminen ja kasvaimessa syövän kehityksen aikana tapahtuvien muutosten kartoittaminen, voi tulevaisuudessa mahdollistaa uusien nykyistä tehokkaampien hoitomuotojen kehittämisen. Yksi syöpätutkimuksen näkökulma on kasvaimen mikroympäristön tutkiminen.

Kasvaimen mikroympäristö eroaa normaalista kudoksesta: soluväliaineen koostumus on muuttunut ja soluille on kehittynyt uusia syöpäsoluja tukevia toimintoja. Myöskin mikroympäristön signaalintimolekyylien kirjo on erilainen, kuin normaalissa kudoksessa. Eroavaisuuksia löytyy myös useasti kasvaimen hapen osapaineessa ja muissa fysikaalisissa tekijöissä. Kaikki osatekijät mahdollistavat muun muassa kasvaimen kasvun, angiogeneesin, etäpesäkkeiden levittämisen, immuunipuolustuksen hiljentämisen ja näin ollen syövän kehittymisen.

Normaali eturauhanen koostuu epiteelikudoksesta ja sitä ympäröivästä pääosin sileitä lihassoluja ja fibroblasteja sisältävästä stroomasta. Eturauhassyöpä kehittyy yleensä rauhasepiteelistä ja valjastaa eturauhasen strooman omaan käyttöönsä (<https://www.cancer.org/cancer/prostate-can->

cer/about/what-is-prostate-cancer.html#written_by, 23.3.2021). Kasvaimen strooma voi olla tilavuudeltaan jopa 90 % kasvaimen koosta ja on hyvin heterogeeninen ryhmä eri solutyyppejä (Feig ym. 2012).

Kasvaimen mikroympäristön solukoostumusta, eli strooman koostumusta, voidaan tutkia laskennallisen dekonvoluution avulla. Dekonvoluutio tarkoittaa alkuperäisen tiedon palauttamista yhdistelemällä nykyistä tietoa. Solukoostumuksen selvittämiseen tällä menetelmällä käytetään tutkittavasta näytteestä saatua geeniekspressiodataa ja useimmiten geenisignatuurimatriisia, joka sisältää tiedon kunkin määrittämissä mukana olevan solutyypin geenisignatuureista. Näitä kahta edellä kuvattua tietoa hyödyntämällä matemaattinen ohjelma laskee näytteen solukoostumuksen. Tämän tyyppistä dekonvoluutiota voidaan kuvata yksinkertaistetulla algoritmilla

$$m = f \times B$$

jossa m on geeniekspressiovektori, f vektori, joka sisältää näytteen solukoostumuksen, ja B signatuurimatriisi. Solukoostumus selvitetään ratkaisemalla f erilaisin laskennallisin menetelmin (Chen ym. 2018).

Näytteen solukoostumuksen määrittämiseen soveltuvia dekonvoluutio-ohjelmia on julkaistu viimeisen 10 vuoden aikana kiihtyvään tahtiin. Ohjelmat pohjautuvat erilaisiin algoritmeihin ja antavat toisistaan osittain poikkeavia tuloksia. Menetelmien välillä on tehty vertailua ja pyritty selvittämään niiden tarkkuutta (Avila Cobos ym. 2020). Tarkkuuteen vaikuttaa kuitenkin useat tekijät, minkä takia on vaikea nimetä yhtä dekonvoluutiomenetelmää, joka olisi selkeästi paras.

Käytettävän datan alkuperällä ja sen käsittelyllä on vaikutusta dekonvoluution tarkkuuteen. Osalla menetelmistä on tiukemmat vaatimukset käytettävän datan tyyppille, esimerkiksi EPIC olettaa datan olevan transkripteja per miljoona (TPM) normalisoitua (Racle ym. 2017, Avila Cobos ym. 2020). Omalle datalle soveltuvan menetelmän valinnan lisäksi, tulosten tarkkuuteen vaikuttaa analyysissä käytettävä signatuurimatriisi. Hyvä signatuurimatriisi perustuu geneihin, jotka ekspressoituvat ainoastaan yhden solutyypin soluilla, mutta kaikilla sen variaatioilla (Gustafsson ym. 2020).

Tässä tutkielmassa pyritään kartoittamaan eturauhasen hyvänlaatuisen liikakasvun, primäärin eturauhassyövän ja kastroatioresistenssin eturauhassyövän mikroympäristön solukoostumuksen eroja hyödyntämällä dekonvoluutiota ja tilastollisia testejä. Erityisen kiinnostuneita ollaan primäärin ja kastroatioresistenssin syöpätyypin välisistä eroista.

2. MATERIAALIT JA MENETELMÄT

Työssä käytettiin RNA-sekvensointidataa, joka on peräisin Tampereen yliopiston Computational Biology tutkimusryhmältä. Data on myös Deseq2 normalisoitu tutkimusryhmän puolesta.

2.1 Näytteet ja Deseq2 normalisointi

Tutkielmassa käsiteltiin dataa 37:stä eri kasvaimen mikroympäristön näytteestä. Näistä näytteistä 10 on peräisin henkilöiltä, joille oli diagnosoitu hyvänlaatuinen eturauhasen liikakasvu (BPH), 16 henkilöiltä, joille oli diagnosoitu primääri eturauhassyöpä (PC) ja 11 henkilöiltä, joille oli diagnosoitu kastratioresistentti eturauhassyöpä (CRPC). Kaikista näytteistä on RNA sekvensoinnin yhteydessä luettu 18 537 proteiinia koodaavan geenin ekspressio.

Normalisoinnissa raakaa RNA sekvensointidataa käsitellään, jotta eri näytteet ovat paremmin vertailtavissa keskenään. Deseq2 metodissa normalisointi perustuu suhdelukujen keskiarvojen laskemiseen (englanniksi the median of ratios method) (Love ym. 2014). Menetelmä tasaa RNA koostumuksen aiheuttamia vääristymiä ja lukusyvyyden vaihtelun vaikutuksia. Normalisoinnin tuloksena syntyy taulukko, jossa kunkin geenin ekspressiota kuvaa suhteellinen arvo (Love ym. 2014).

2.2 Mikroympäristön solukoostumuksen määrittäminen

Näytteiden solukoostumuksen määrittäminen toteutettiin laskennallisen dekonvoluution avulla. Deseq2 normalisoidun RNA-seq datan kanssa yhteensopivia dekonvoluutio-ohjelmia on esimerkiksi pienimmän neliösumman menetelmään perustuva non-negative least squares (nnls), tukivektorin regressioon perustuva CIBERSORTx sekä vakaan regression lähestymisen menetelmään (englanniksi robust regression approaches) perustuva FARDEEP (Newman ym. 2015, Hao ym. 2019, Avila Cobos ym. 2020). Ohjelmien tarjoamat valmiit signatuurimatriisit poikkeavat toisistaan ja näin ollen vaikuttavat dekonvoluution tuloksiin. Tämä tekee eri ohjelmien välisten tulosten vertailun vaikeaksi. Koska työssä ei ollut käytössä omaa signatuurimatriisia, päädyttiin dekonvoluutio suorittamaan vain yhdellä ohjelmalla. Käytettäväksi valittiin CIBERSORTx, koska sen tarkkuudesta löytyi näyttöä eri julkaisuista ja se tarjoaa helppokäyttöisen selaimessa käytettävän alustan, sekä laajan geenisignatuurimatriisin (Newman ym. 2015, Avila Cobos ym. 2020).

Data ladattiin CIBERSORTx:n nettisivuille (<https://cibersortx.stanford.edu/index.php>, 21.4.2021) ja valittiin "Impute Cell Fractions" analyysi moduulien joukosta. Ajo suoritettiin käyttämällä ohjelman

valmiina tarjoamaa geenisignatuurimatriisia LM22. Se sisältää 547 geeniä, joiden perusteella pystytään määrittämään 22 ihmisen hematopoieettista solutyyppiä, mukaan lukien seitsemän T-solutyyppiä, naiivi -ja muisti-B-solutyyppejä, plasmakolut, luonnolliset tappajasolut ja erilaisia myeloidisen linjan leukosyyttejä (Newman ym. 2015). Solutyyppien joukosta löytyy myös kolme makrofagiryhmää, joiden erona on solujen erottelutaktiikka signatuurimatriisin valmistusvaiheessa. Ryhmän 0 makrofagit ovat erilaistettu monosyyteistä, kun taas ryhmän 1 ja 2 makrofagit on puhdistettu verestä ja ne eroavat toisistaan ainoastaan siinä, että ryhmän 2 makrofagien aktivoimiseen on käytetty interferoni gamma ($\text{IFN-}\gamma$) ja lipopolysakkaridin lisäksi interleukiini 4:sta (IL-4) (Newman ym. 2015).

CIBERSORTx tarjoaa käyttäjälle mahdollisuuden sarjankorjausmoodin käyttöön signatuurimatriisin ja ekspressiomatriisin teknisten erojen korjaamiseen. Työssä käytettiin B-moodia. Muut ajon asetukset olivat seuraavat: kvartaalien normalisointi oli pois päältä, koska sitä ei suositella RNA-sekvensointidatalle, permutaatioksi asetettiin sata (100) ja ajo suoritettiin absoluuttisessa moodissa.

Absoluuttisessa moodissa suhteelliset soluosuudet skaalataan arvoiksi, jotka esittävät kunkin solutyyppin todellista osuutta näytteessä. Arvot lasketaan signatuurimatriisin ekspressiotasojen mediaanin ja kunkin näytteen kaikkien geenien ekspressiotasojen mediaanin suhteesta. Absoluuttisen moodin arvot ovat suoraan vertailtavissa solutyyppien ja näytteiden välillä (Newman ym. 2015, Sturm ym. 2019).

2.3 Dekonvoluution tulosten käsittely ja visualisointi

Dekonvoluutiosta saadut tulokset käsiteltiin R-ohjelmalla. Näytteet ryhmiteltiin näytetyypin mukaan (BPH, PC ja CRPC) ja niistä piirrettiin laatikkojanakuvaajia. Dekonvoluution onnistumista arvioitiin tarkastelemalla analyysin p-arvoja. Eroja eri näyteryhmien solukoostumuksessa selvitetiin Kruskal-Wallis testillä, jonka nollahypoteesiksi asetettiin: "Ryhmien välillä ei ole tilastollisesti merkitseviä eroja solutyyppien suhteellisten määrien mediaanien välillä". Kruskal-Wallis testin perusteella valittiin Mann-Whitney testiin ne solutyyppit, joiden ryhmien väliltä löytyi eroja. Mann-Whitneyn testillä määritettiin tarkemmin minkä näyteryhmien välillä erot löytyvät. Testi tehtiin käyttämällä p-arvon korjaamiseen Holmin menetelmää, joka pienentää tyyppin 1 virhettä, eli tilannetta, jossa tosi nollahypoteesi kumotaan (Holm 1979). Kaikki tilastolliset testit tehtiin 0,05 merkitsevyystasolla. Käytetty R koodi löytyy liitteistä (liite 1).

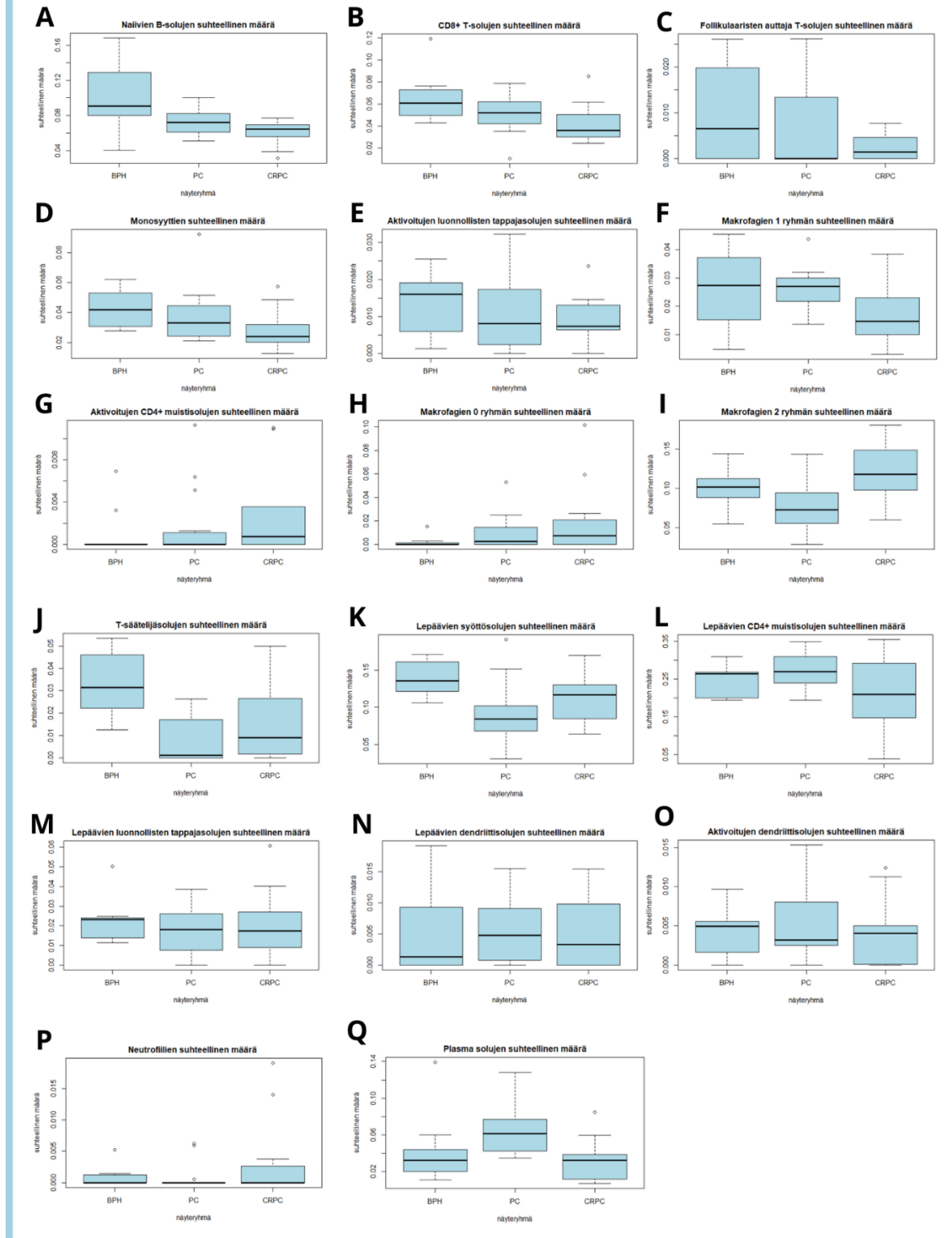
3. TULOKSET JA NIIDEN TARKASTELU

3.1 Dekonvoluution tulokset

CIBERSORTx laski kaikille näytteille p-arvoksi nollan. Kun tilastollisen merkitsevyyden rajana käytetään p-arvoa pienempi kuin 0,05, kaikki näytteet kumosivat nollahypoteesin: ”Mitään soluista, jotka sisältyvät signatuurimatriisiin, ei ole tarkastellussa näytteessä” (Newman ym. 2015). Kaikkien näytteiden dekonvoluutio on näin ollen onnistunut ainakin p-arvojen perusteella.

Eri solutyypeistä piirretyistä laatikkojanakuvaajista voidaan nähdä, että data ei ole suurimmalla osalla solutyypeistä ja näyteryhmistä normaalisti jakautunut (kuva 1). Voidaan myös havaita, että naiivien B-solujen, CD8+ T-solujen, follikulaaristen auttaja T-solujen, monosyyttien, aktivoitujen luonnollisten tappajasolujen ja makrofagien 1 ryhmän solujen suhteellinen määrä on korkein BPH:ssa, toiseksi korkein PC:ssa ja matalin CRPC:ssa (kuva 1 a–f). Aktivoitujen CD4+ muistisolujen ja makrofagien 0 ryhmän profiili taas on edellisiin solutyyppeihin verrattuna päin vastainen, eli niiden määrät ovat taas korkeimpia CRPC näytteissä ja matalimpia BPH näytteissä (kuva 1 g ja h). Muiden solutyyppeiden kohdalla selkeää nousevaa tai laskevaa kaavaa ei ole havaittavissa, kun tarkastellaan kaikkia kolmea näyteryhmää. B muistisoluja, gamma delta T-soluja, naiiveja CD4+ T-soluja, eosinofiileja ja aktivoituja syöttösoluja ei havaittu ollenkaan tai havaittiin vain parista näytteestä minimaalisia määriä, joten niitä, ei oteta mukaan jatkotutkimuksiin. Koska laatikkojanakuvaajista tehdyt analyysit ovat silmämääräisiä ja näin ollen eivät kovin luotettavia, etsittiin eroja myös tilastollisia testejä tekemällä.

Solutyypin suhteelliset määrät näytteryhmittäin



Kuva 1. Solutyypin suhteelliset määrät näytetyypiryhmittäin dekonvoluutiosta analysoidulle datalle. BPH=hyvänlaatuinen eturauhasen liikakasvu, PC= primääri eturauhassyöpä ja CRPC= kastroaatioreistentti eturauhassyöpä.

3.2 Kruskal-Wallis testin tulokset

Tarkemmat tutkimukset näytetyyppiryhmien välisten erojen kartoittamiseen suoritettiin Kruskal-Wallis testillä. Testin tulokset on koottu taulukkoon 1 ja siitä nähdään, että yhteensä 7 eri solutyyppeä kumosi nollahypoteesin ja näin ollen voidaan todeta, että niiden tasoissa on tilastollisesti merkitseviä eroja BPH, PC ja CRPC näyteryhmien välillä. Tulokset eivät kuitenkaan kerro, minkä ryhmien väliltä ero(t) löytyvät.

Taulukko 1. *Kruskal-Wallis testin tulokset. Nollahypoteesi= "Näyteryhmien välillä ei ole tilastollisesti merkitseviä eroja solutyyppeiden tasoissa". Tilastollisen merkitsevyyden rajana käytettiin p-arvoa pienempi kuin 0,05 ja taulukosta nähdään, että 7 alinta solutyyppeä alittaa sen.*

solutyyppe	Kruskal-Wallis p-arvot	ryhmissä eroja
lepäävät CD4+ T-muistisolut	0,33	ei
aktivoituneet CD4+ T-muistisolut	0,28	ei
follikulaariset auttaja T-solut	0,44	ei
lepäävät luontaiset tappajasolut	0,53	ei
aktivoituneet luontaiset tappajasolut	0,36	ei
makrofagit M0	0,22	ei
makrofagit M1	0,11	ei
lepäävät dendriittisolut	0,61	ei
aktivoituneet dendriittisolut	0,69	ei
naiivit B-solut	0,00	kyllä
plasmasolut	0,00	kyllä
CD8+ T-solut	0,02	kyllä
T-säätelijäsolut	0,00	kyllä
monosyytit	0,03	kyllä
makrofagit M2	0,00	kyllä
lepäävät syöttösolut	0,00	kyllä

3.3 Mann-Whitney testin tulokset

Tilastollisen Mann-Whitney testin perusteella voidaan todeta, että 95 % luottamustasolla minkään solutyypin kaikkien ryhmien välillä ei ole tilastollisesti merkitsevää eroa. Makrofagien 2 ryhmän ja plasmasolujen välillä oli tilastollisesti merkitseviä eroja sekä BPH ja PC että PC ja CRPC näyteryhmien välillä. Tämän lisäksi kahden ryhmän välillä eroja löytyi 5 solutyypiltä (taulukko 2). Naiivien B-solujen, CD8+ T-solujen ja monosyyttien solumäärissä oli tilastollisesti merkitseviä eroja BPH ja

CRPC näyteryhmien välillä ja säätelijä T-solujen ja lepäävien syöttösolujen määrissä oli tilastollisesti merkittävä ero BPH ja PC näytteiden välillä.

Taulukko 2. Mann-Whitney testin tulokset. Taulukossa on esitetty minkä näyteryhmäparien välillä on tilastollisesti merkitsevä ero tutkittujen solutyypin suhteellisissa määrissä. Tilastollisen merkitsevyyden raja $<0,05$.

solutyyppi	ero BPH ja PC välillä	ero BPH ja CRPC välillä	ero PC ja CRPC välillä
naiivit B-solut	ei	kyllä	ei
plasmasolut	kyllä	Ei	kyllä
CD8+ T-solut	ei	kyllä	ei
säätelijä T-solut	kyllä	Ei	ei
monosyytit	ei	kyllä	ei
makrofagit M2	kyllä	ei	kyllä
lepäävät syöttösolut	kyllä	ei	ei

BPH=hyvänlaatuinen eturauhasen liikakasvu, PC= primäärieturauhassyöpä ja CRPC= kastratioresistenssieturauhassyöpä.

Kun vertaillaan kuvan 1 laatikkojanakuvaajia ja Mann-Whitney testin tuloksia voidaan päätellä, että monosyyttien, B-solujen ja CD8+ solujen suhteelliset määrät laskevat BPH ja CRPC näyteryhmien välillä ja säätelijä-T-solujen sekä lepäävien syöttösolujen suhteelliset määrät laskevat BPH ja PC näyteryhmien välillä. Makrofagien 2 ryhmän solujen suhteellinen määrä taas nousee BPH ja CP näyteryhmien välillä ja sitten laskee CP ja CRPC ryhmiä vertailtaessa. Plasmasolujen suhteelliset määrät käyttäytyvät päin vastaisesti makrofagien 2 ryhmän kanssa, eli ensin nousevat BPH ja PC näyteryhmien välillä ja sitten laskevat CP ja BPH näyteryhmien välillä.

4. POHDINTA

Dekonvoluution ja tilastollisten testien perusteella löydettiin tilastollisesti merkitseviä eroavaisuuksia naiivien B-solujen, plasmakomponenttien, CD8+ T-solujen, säätelijä T-solujen, monosyyttien, makrofagien 2 ryhmän solujen ja lepäävien syöttösolujen suhteellisissa tasoissa hyvänlaatuisen eturauhasepiteelin liikakasvun mikroympäristön, primääriin eturauhassyövän mikroympäristön ja kastroatioresistenssin eturauhassyövän mikroympäristön näytteiden välillä. Eroja löydettiin joko yhden tai kahden näytteryhmäparin väliltä, mutta ei kaikkien kolmen. Useimpien solutyypin kohdalla löydetty eroavaisuus oli laskeva BPH:n ja PC:n tai CRPC:n välillä, mikä voisi selittyä sillä, että PC tai CRPC kasvain on kehittynyt hiljentämään immuunipuolustuksen soluja.

Tässä tutkielmassa tarkastelluista immuunisoluista ainoastaan makrofagien 2 ryhmän solujen ja plasmakomponenttien väliltä löydettiin tilastollisesti merkitseviä eroja PC:n ja CRPC:n välillä. Jatkotutkimuksissa voitaisiinkin keskittyä juuri näihin kahteen solutyypin ja niiden rooliin kastroatioresistenssin eturauhassyövän kehityksessä. Tässä työssä saatujen tulosten varmistamiseksi ja tarkkuuden parantamiseksi dekonvoluutio tulisi suorittaa uudestaan isommalla näytekoolla ja käyttäen eturauhasepiteelin ja eturauhassyöpäkasvaimen pohjalta kehitettyä signatuurimatriisia ja näin ollen pyrkiä minimoimaan virhetekijöitä. Jos käytössä olisi oma signatuurimatriisi, dekonvoluutio voitaisiin myös suorittaa useammalla kuin yhdellä ohjelmalla, jolloin päästäisiin vertailemaan dekonvoluutio-ohjelmien välisiä eroja.

5. LÄHTEET

- Avila Cobos F, Alquicira-Hernandez J, Powell JE, ym. Benchmarking of cell type deconvolution pipelines for transcriptomics data, *Nature Communications* 2020;11.
- Chen B, Khodadoust MS, Liu CL, ym. Profiling Tumor Infiltrating Immune Cells with CIBERSORT, *Cancer Systems Biology; Methods Mol Biol* 2018;1711;243-59.
- Crawford ED, Petrylak D & Sartor O Navigating the evolving therapeutic landscape in advanced prostate cancer, *Urologic Oncology: Seminars and Original Investigations* 2017;35;S1-S13.
- Feig C, Gopinathan A, Neesse A, ym. The Pancreas Cancer Microenvironment, *Clinical Cancer Research* 2012;18;4266-76.
- Gustafsson J, Held F, Robinson JL, ym. Sources of variation in cell-type RNA-Seq profiles, *PLOS ONE* 2020;15;e0239495.
- Hao Y, Yan M, Heath BR, ym. Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares, *PLoS computational biology* 2019;15;e1006976.
- Holm S A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavian Journal of Statistics* 1979;6;65-70.
- Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome biology* 2014;15;550.
- Newman AM, Liu CL, Green MR, ym. Robust enumeration of cell subsets from tissue expression profiles, *Nature Methods* 2015;12;453-7.
- Racle J, de Jonge K, Baumgaertner P, ym. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data, *eLife* 2017;6;e26476.
- Sturm G, Finotello F, Petitprez F, ym. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, *Bioinformatics* 2019;35;i436-45.
- Sung H, Ferlay J, Siegel RL, ym. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: A Cancer Journal for Clinicians* 2021.

6. LIITE 1: R-KOODI

```
#Datan ryhmittely
naytetyyppi<-c()
naytetyyppi[1:10]<- "BPH"
naytetyyppi[11:26]<- "PC"
naytetyyppi[27:37]<- "CRPC"
CIBERSORTx_data<-cbind(naytetyyppi,CIBERSORTx_data)
CIBERSORTx_data$naytetyyppi<- factor(CIBERSORTx_data$naytetyyppi ,
levels=c("BPH", "PC", "CRPC"))

#P-arvojen tutkiminen
summary(CIBERSORTx_data$`P-value` )

#Laatikkojanakuvaajien piirtäminen solutyypeittäin. Kaikkien solutyypien ku-
vaajat piirrettiin samalla tavalla kuin naiivien B-solujen.

#Naiivit B-solut
boxplot(CIBERSORTx_data$`B cells naive`~CIBERSORTx_data$naytetyyppi, main="Na-
iivien B-solujen suhteellinen määrä",xlab="näyteryhmä",ylab="suhteellinen
määrä",col="lightblue",border="black")

#Kruskal-Wallis testi. Testaan nollahypoteesia: "Ryhmiem välillä ei ole eroa
solutyypin määrässä". Tilastollisen merkitsevyyden rajana käytetään p-arvoa
0,05.

Kruskal_Wallis_p_arvot<-c()

for (solutyyppi in 3:24) {
  tulos<-kruskal.test(CIBERSORTx_data[,solutyyppi]~naytetyyppi, data = CIBER-
SORTx_data)
  Kruskal_Wallis_p_arvot<-c(Kruskal_Wallis_p_arvot,tulos$p.value)
}

solutyypit<-colnames(CIBERSORTx_data[3:24])
ryhmissa_eroja<-Kruskal_Wallis_p_arvot <=0.05
Kruskal_taulukko<-cbind(solutyypit,Kruskal_Wallis_p_arvot,ryhmissa_eroja)

#Selvitetään minkä ryhmiem välillä eroja löytyy pairwise Wilcoxon testillä.
for (solutyyppi in c(3,5,6,11,15,18,21)) {
  print(pairwise.wilcox.test(CIBERSORTx_data[,solutyyppi], CIBER-
SORTx_data$naytetyyppi,paired = FALSE))
}
```