TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
GENOOMIKA INSTITUUT

Maret Vilbaste

Emakakaelavähiga seotud fenotüüpide definitsiooni mõju geneetilise seoseanalüüsi tulemusele

Magistritöö

30 EAP

Juhendajad
Kaasprofessor Triin Laisk
Teadur Mariann Võsa
Nooremteadur Jelisaveta Džigurski

TARTU 2025

Infoleht

Emakakaelavähiga seotud fenotüüpide definitsiooni mõju geneetilise seoseanalüüsi tulemustele

Kokkuvõte. Emakakaelavähk on üks levinumaid günekoloogilisi pahaloomulisi kasvajaid maailmas. Viimaste aastate jooksul on ülegenoomsed seoseuuringud (GWAS) võimaldanud tuvastada emakakaelavähi ja selle eelastmetega seotud geneetilisi riskilookuseid. Selliste uuringute tulemused võivad sõltuda sellest, kuidas on uuritav fenotüüp defineeritud. Käesolevas magistritöös uuriti, kuidas erinevad fenotüübi definitsioonid mõjutavad emakakaelavähi geneetiliste seoseanalüüside tulemusi, kasutades Tartu Ülikooli Eesti Geenivaramu andmeid. Töös viidi läbi mitme fenotüübidefinitsiooni alusel GWAS analüüs ja selle interpretatsioon, hinnati geneetiliste variantide efektisuurusi ning tuvastati spetsiifilisi ja kattuvaid geneetilisi signaale. Leiti, et fenotüübi määratlused mõjutavad nii tuvastatud geneetilisi signaale kui geneetiliste seoste tugevust ja usaldusväärsust, mistõttu on oluline täpne fenotüübi määratlus tulevaste geneetiliste uuringute kavandamisel.

Märksõnad: emakakaelavähk, düsplaasia, GWAS
CERCS kood: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika; B790 Kliiniline geneetika/B200 Tsütoloogia, onkoloogia, kantseroloogia

The impact of phenotype definition related to cervical cancer on the results of genetic association analysis

Abstract. Cervical cancer is one of the most common gynecological malignancies globally. Despite effective measures like HPV vaccination and screening programs, incidence and mortality remain high in several countries, including Estonia. Genome-wide association studies (GWAS) have identified genetic loci linked to cervical cancer and its precursor lesions. However, GWAS results can vary depending on how phenotypes are defined. In this master's thesis, we investigated how different phenotype definitions affect the outcomes of genetic association studies using Estonian Biobank data. Several GWAS analyses were conducted based on varying ICD-10 code combinations representing cervical cancer, carcinoma in situ and dysplasia. The effect sizes of selected genetic variants were assessed and compared across phenotype definitions. Results show that phenotype definitions influence the strength and consistency of detected genetic signals, highlighting the need for precise phenotype classification in future research.

Keywords: cervical cancer, dysplasia, GWAS
CERCS code: B110 Bioinformatics, medical informatics, biomathematics, biometrics;
B200 Cytology, oncology, cancerology

Sisukord

Infoleht 1
Sisukord 3
Kasutatud lühendid 5
Sissejuhatus 7
1. Kirjanduse ülevaade 8
1.1 Emakakaela ehitus ja funktsioon 8
1.1.1 Emakakaela arenguline päritolu ja histoloogiline eripära 8
1.1.2 Anatoomiline ja morfoloogiline struktuur 8
1.1.3 Emakakaela funktsioonid 10
1.2 Peamised emakakaela patoloogiad 11
1.2.1 Inimese papilloomiviiruse infektsioon 11
1.2.2 Emakakaela düsplaasia 14
1.2.3 Emakakaelavähk 17
1.3 Ülegenoomne seoseuuring 21
1.4 Emakakaelavähi senised ülegenoomsed seoseuuringud 24
2. Eksperimentaalosa 29
2.1 Töö eesmärk 29
2.2 Materjal ja metoodika 29
2.2.1 Valim ja uuritavad tunnused 29
2.2.2 Genotüpiseerimine, kvaliteedikontroll, imputatsioon 31
2.2.3 Assotsiatsioonianalüüs TÜ Eesti Geenivaramu andmetel 32
2.2.4 GWAS tulemuste annoteerimine ja visualiseerimine FUMA veebiplatvormi abil 32
2.2.5 Fenotüübi definitsiooni mõju hindamine 33
2.3.1 Seoseanalüüside tulemused ja kandidaatgeenide prioritiseerimine 35
2.3.1.1 Emakakaela pahaloomuline kasvaja 39
2.3.1.2 Emakakaela pahaloomuline kasvaja ja in situ kartsinoom 41
2.3.1.3 Emakakaela in situ kartsinoom 44
2.3.1.4 Düsplaasia ja kombineeritud fenotüübid 45
2.3.2 Emakakaelavähi teadaolevate riskilookuste võrdlus 48
2.4 Arutelu 50
Kokkuvõte 52
Summary 53

Kasutatud kirjandus 54
Kasutatud veebiaadressid 62
LISAD 63
Lisa 1. 63
Lisa 2. 64

Kasutatud lühendid

AA	alternatiivne alleel (ingl *alternative allele*)
C53	emakakaela pahaloomuline kasvaja RHK-10 alusel (ingl *malignant neoplasm of cervix uteri*)
CI	usaldusintervall (ingl *confidence interval*)
CIN	emakakaela intraepiteliaalne neoplaasia (ingl *cervical intraepithelial neoplasia*)
D06	emakakaela kartsinoom *in situ* RHK-10 alusel (ingl *carcinoma in situ of cervix uteri*)
EA	efektialleel (ingl *effect allele*)
eQTL	ekspressiooni taset mõjutav lookus (ingl *expression quantitative trait loci*)
FUMA	funktsionaalne kaardistamine ja annoteerimine (ingl *Functional Mapping and Annotation*)
GWAS	ülegenoomne seoseuuring (ingl *genome-wide association study*)
HLA	inimese leukotsüüdi antigeen (ingl *human leukocyte antigen*)
HSIL	emakakaela rakumuutuste raske aste (ingl *high-grade squamous intraepithelial lesion*)
HPV	inimese papilloomiviirus (ingl *human papillomavirus*)
hrHPV	kõrge riskiga inimese papilloomiviirus (ingl *high risk human papillomavirus*)
LD	ahelduse tasakaalutus (ingl *linkage disequilibrium*)
LSIL	emakakaela rakumuutuste kerge aste (ingl *low-grade squamous intraepithelial lesion*)
MAGMA	geenipõhise analüüsi meetod (ingl *Multi-marker Analysis of Genomic Annotation*)
MAF	harvema alleeli sagedus (ingl *minor allele frequency*)
N87	emakakaela düsplaasia RHK-10 alusel (ingl *dysplasia of cervix uteri*)
N87.2	mujal klassifitseerimata raske emakakaela düsplaasia RHK-10 alusel (ingl *severe cervical dysplasia, not elsewhere classified*)
NGS	järgmise põlvkonna sekveneerimine (ingl *next generation sequencing*)
OR	šansside suhe (ingl *odds ratio*)
Pap-test	Papanicolaou test (ingl *Papanicolaou-smear*)
RHK-10	Rahvusvaheline Haiguste Klassifikatsioon, 10. väljaanne (ingl *International Classification of Diseases 10th Revision*)

SD	standardhälve (ingl *standard deviation*)
SJC	lame- ja silinderepiteeli ühendusjoon (ingl *squamocolumnar junction*)
SIL	lamerakuline intraepiteliaalne kahjustus (ingl *squamous intraepithelial lesion*)
SNP	üksiknukleotiidne polümorfism (ingl *single nucleotide polymorphism*)
TÜ EGV	Tartu Ülikooli Eesti Geenivaramu

Sissejuhatus

Emakakaelavähk on üks levinumaid günekoloogilisi pahaloomulisi kasvajaid maailmas ning WHO (*World Health Organization*) andmetel on see globaalselt levimuselt neljas vähivorm naiste seas. Vaatamata tõhusatele ennetusmeetmetele, nagu HPV vastu vaktsineerimine ja sõeluuringuprogrammid, püsib emakakaelavähi haigestumus ja suremus mitmetes riikides, sealhulgas Eestis, endiselt kõrge. Haiguse peamiseks etioloogiliseks teguris on kõrge riskiga HPV infektsioon, eriti 16 ja 18 tüüpidega. Samas ei kujune kõigil nakatunutel välja vähieelseid muutusi ega vähki. Viirusinfektsiooni ja vähi progresseerumise vahelist protsessi mõjutavad ka muud tegurid, sealhulgas geneetiline eelsoodumus.

Viimaste aastate jooksul on ülegenoomsed assotsiatsiooniuuringud (GWAS) võimaldanud tuvastada emakakaelavähi ja selle eelastmetega seotud geneetilisi lookuseid. GWAS tulemused võivad sõltuda oluliselt sellest, kuidas on defineeritud uuritav fenotüüp – kas on kaasatud vaid invasiivsed juhud või ka düsplaasia ja *in situ* kartsinoom. Seetõttu on oluline mõista, kuidas erinevad fenotüübi määratlused mõjutavad seoseanalüüsi tulemusi ning milline on nende mõju tuvastatud geneetiliste signaalide tugevusele ja usaldusväärsusele.

Käesolevas töös uuriti erinevate emakakaelavähiga seotud fenotüüpide definitsioonide mõju GWAS tulemustele, kasutades Tartu Ülikooli Eesti Geenivaramu andmeid. Töö raames viidi läbi erinevate emakakaelavähiga seotud fenotüüpide GWAS tulemuste annoteerimine ning võrdlus. Antud magistritöö on valminud Tartu Ülikooli Molekulaar- ja rakubioloogia instituudi molekulaarsete bioteaduste õppekava raames. Töö eksperimentaalne osa viidi läbi Tartu Ülikooli Genoomika instituudis.

1. Kirjanduse ülevaade

1.1 Emakakaela ehitus ja funktsioon

1.1.1 Emakakaela arenguline päritolu ja histoloogiline eripära

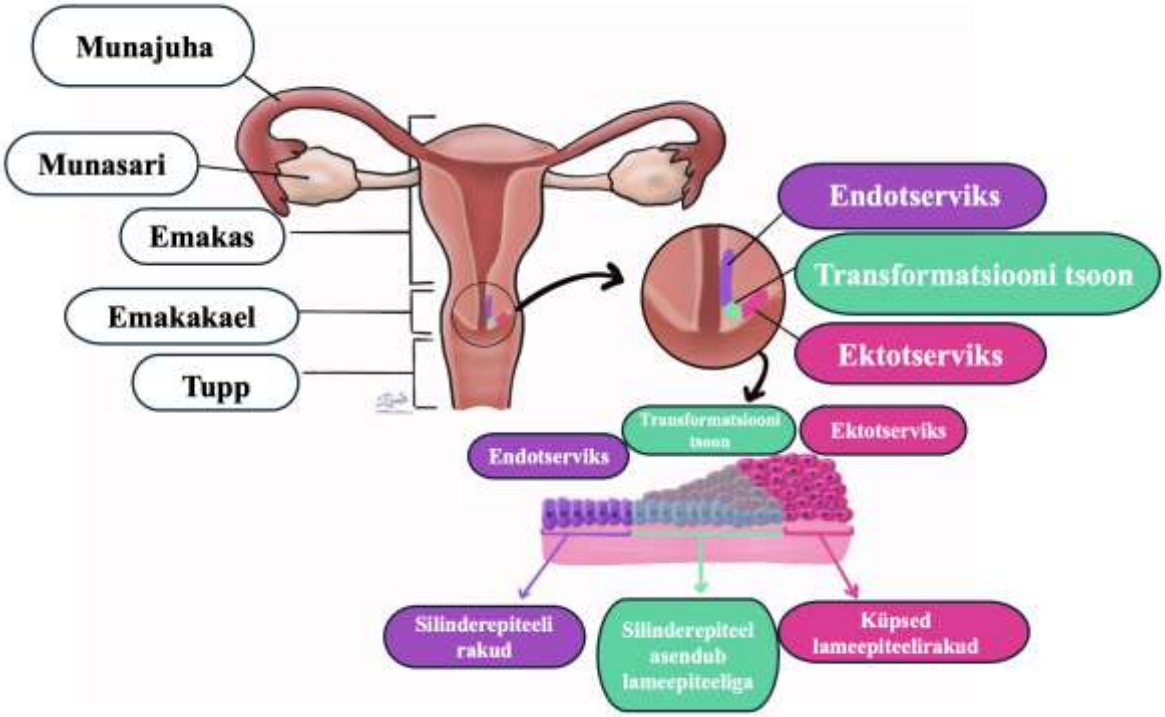
Emakakael kujuneb embrüonaalses arengus paramesonefrilistest ehk Mülleri juhadest. Nende juhade distaalsed osad ühinevad 9.–10. rasedusnädalal, moodustades emaka, emakakaela ja tupe ülemise osa (Habiba *et al*., 2021). Ühinenud osade keskosast areneb emakas, sulandunud kraniaalsed otsad aga kujunevad munajuhadeks, mille lehtrikujuline fimbriaalne osa jääb avatuks ning paikneb munasarjade lähedal (Moncada-Madrazo & Rodríguez Valero, 2025). Kaudaalne osa moodustab tupe ülemise kolmandiku (Moncada-Madrazo & Rodríguez Valero, 2025). Selles staadiumis tekib ka algselt keskjoonele vahesein, mis emakaõõnes tavaliselt resorbeerub täielikult umbes 20. rasedusnädalal, kui resorbeerumine ebaõnnestub võib tekkida vaheseinaga emakas (ingl *separate uterus*) (Moncada-Madrazo & Rodríguez Valero, 2025).

Embrüo arengu käigus toimub emakakaela epiteeli järkjärguline diferentseerumine, mille käigus kujunevad välja silindriline näärmeepiteel endotserviksis ning lameepiteel ektotserviksi ja tupe piirkonnas (Habiba *et al*., 2021). Endotserviks on tihedalt innerveeritud, peamiselt sümpaatiliste ja parasümpaatiliste närvide kaudu, mis pärinevad sakraalnärvidest (S2–S3) ning kulgevad läbi uterosakraalsete sidemete ristluu suunas. See on põhjuseks, miks emakakaela invasiivsete piirkondadega seotud protseduurid, nagu endotserviksi kuretaaž, võivad tekitada tugevat ebamugavust ja harvematel juhtudel ka vasovagaalset refleksi. Seevastu ektotserviksi piirkond on närvilõpmetega vähem varustatud, mistõttu on pindmiste protseduuride, nagu biopsia, krüoteraapia või elektrokoagulatsioon, naistele hästi talutav ka ilma lokaalanesteesiata (Mehta & Sachdeva, 2017).

1.1.2 Anatoomiline ja morfoloogiline struktuur

Emakakael on silindrilise struktuuriga fibromuskulaarne organ (Nott *et al*., 2016), mis asub emaka alumises osas ning ühendab emakat tupega (Joonis 1) (Prendiville & Sankaranarayanan, 2017). Täiskasvanud, mitteraseda naise emakakaela normaalne pikkus on ligikaudu 25 millimeetrit, anteroposterioorne läbimõõt jääb vahemikku 20–25 millimeetrit ning ristiläbimõõt 25–30 millimeetrit (Nott *et al*., 2016). Mõõtmed varieeruvad sõltuvalt naise vanusest, sünnituste arvust ja menstruaaltsükli faasist (Nott *et al*., 2016). Emakakaela stabiliseerivad

mõlemal pool paiknevad sidemed – uterosakraalsed ja kardinaalsed (ehk ristiemakakaela) sidemed, millest eriti uterosakraalsetel on oluline roll emaka hoidmisel selle tavapärases ettepoole kallutatud asendis (Nott *et al*., 2016). Emakakaela ülemine osa paikneb tupe kohal, kõhuõõnes ning alumine ehk intravaginaalne osa paikneb tupe ülemises osas ning on ümbritsetud tupevõlvide poolt – need on piirkonnad, kus tupe limaskest kaardub ringjalt ümber emakakaela (Prendiville & Sankaranarayanan, 2017). Need kaks osa on ligikaudu võrdse suurusega. Emakakaelal eristatakse kaht peamist anatoomilist piirkonda – ektotserviks (ingl *ectocervix*) ja endotserviks (ingl *endocervix*) (Mehta & Sachdeva, 2017). Ektotserviks on emakakaela välimine osa, mida näeb günekoloogilisel läbivaatusel ning see on kaetud lamerakulise epiteeliga (Prendiville & Sankaranarayanan, 2017). Endotserviks on emakakaela sisemine osa, mis moodustab kanali ja ühendab tuppe emakaga. See on kaetud silindrikujuliste näärmerakkudega, mis toodavad lima ning nääreepiteel pole tavaliselt kolposkoopial nähtav (Prendiville & Sankaranarayanan, 2017). Ekto- ja endotserviksi üleminekualal liituvad lame- ja näärmeepiteel ja seda piirkonda nimetatakse transformatsioonitsooniks või ingliskeelses kirjanduses ka *squamocolumnar junction*'iks (SCJ) (Joonis 1) (Prendiville & Sankaranarayanan, 2017). Anatoomiliselt on transformatsioonitsoon vastuvõtlik pahaloomulistele muutustele (Nott *et al*., 2016). SCJ on dünaamiline ja võib liikuda varases noorukieas kui ka esimese raseduse ajal (Prendiville & Sankaranarayanan, 2017). Transformatsioonitsooni suurus võib erineda, naise reproduktiivsetel aastatel on rohkem ektotserviksi pool ning naaseb pärast menopausi endotservikaalsesse piirkonda (Prendiville & Sankaranarayanan, 2017).



Joonis 1. Emakakaela anatoomia ja histoloogiline struktuur (Modifitseeritud Wasserman & Gorski, 2025 järgi).

1.1.3 Emakakaela funktsioonid

Emakakaelal on oluline roll viljastumisel, raseduse säilitamisel ning sünnituse kulgemisel. Kogu raseduse vältel peab emakakael olema suletud, vaatamata mehhaanilistele, sh kasvava loote ja lootekoti raskus, ning hormonaalsetele mõjudele (Myers *et al*., 2015). See võimaldab sel toimida barjäärina, takistades tupe mikrofloora sattumist emakasse ning hoida kasvavat loodet emakaõõnes. Raseduse lõpus ja sünnituse ajal peab emakakael pehmenema, lühenema ja avanema, et võimaldada lapse väljumine emakast läbi sünnitustee. Emakakaela kaasasündinud või omandatud puudulikkus võib häirida selle funktsiooni ja suurendada enneaegse sünnituse riski (Nott *et al*., 2016). Emakakaelal on mitmekülgne roll reproduktiivprotsessides, mistõttu võib selle struktuurne või funktsionaalne kahjustus avaldada mõju naise viljakusele ja üldisele reproduktiivtervisele. Seetõttu on emakakaela patoloogiate uurimine oluline mitte ainult vähi, vaid ka viljatuse ja raseduskomplikatsioonide kontekstis.

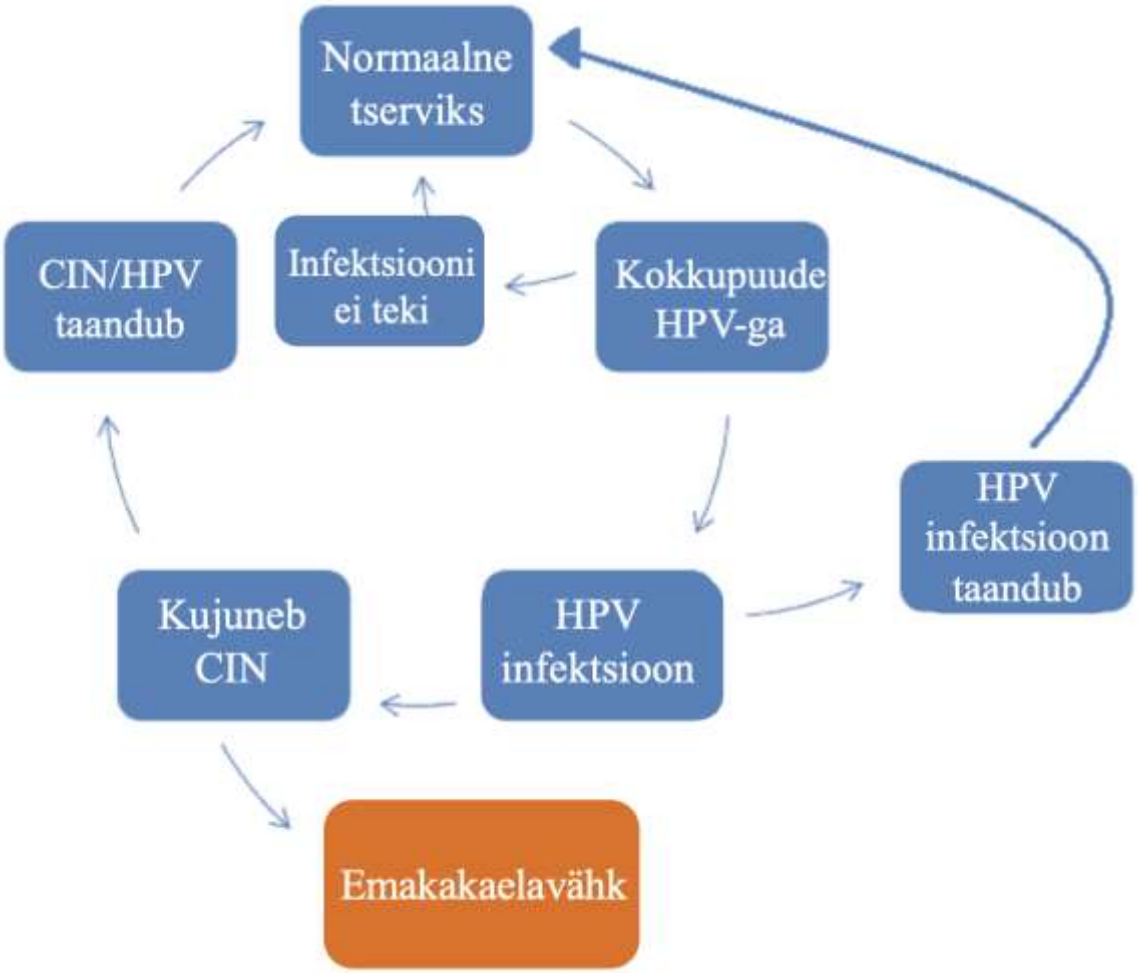
1.2 Peamised emakakaela patoloogiad

Emakakaelaga seotud patoloogiad võivad olla väga erineva iseloomuga, ulatudes põletikest ja healoomulistest muutustest (näiteks erosioonid, polüübid, tsüstid) kuni emakakaela düsplaasia

ja pahaloomuliste kasvajateni. Levinumateks haiguslikeks seisunditeks on emakakaela põletikud, emakakaela stenoos, düsplastilised muutused ning emakakaelavähk. Kuna käesoleva töö fookuses on emakakaelavähk ja selle eelnevad astmed, siis keskendutakse järgnevates alapeatükkides just nende patoloogiate üksikasjalikumale käsitlemisele.

1.2.1 Inimese papilloomiviiruse infektsioon

Inimese papilloomiviirus (HPV, ingl *human papilloma virus*) on üks levinumaid sugulisel teel levivaid viirusi maailmas – hinnanguliselt nakatub elu jooksul HPV-ga 50–80% naistest (Stanley, 2010). Nakkus võib levida igas vanuses seksuaalselt aktiivsetel naistel, kuid selle levimus on kõrgeim noorukitel ja kuni 30-aastastel naistel, eriti vanuserühmas 20–24 eluaastat (Mello & Sundstrom, 2023). Enamasti paranetakse HPV-nakkusest spontaanselt, kuid viirus võib püsida organismis ka latentsena või avalduda genitaaltüügastena ehk kondüloomidena, mis on enamasti põhjustatud madala riski HPV tüvedest – 6 ja 11 (Luria & Cardoza-Favarato, 2025). Kui organismi immuunsüsteem ei suuda takistada viirusel kahjustuste tekitamist, võib see aastateks organismi püsima jääda ning see on emakakaela rakkude vähirakkudeks muutumise protsessi üks kriitilisi tegureid (Mello & Sundstrom, 2023). Üle 30-aastastel naistel, kellel HPV jääb organismi püsima, on suurem risk infektsiooni ja sellega seotud emakakaela intraepiteliaalse neoplaasia (ingl *cervical intraepithelial neoplasia*, CIN) tekkeks (Joonis 2) (Mello & Sundstrom, 2023). Tavaliselt sellised muutused normaliseeruvad, just eriti noortel naistel, tänu immuunvastusele ja emakakaela rakkude kiirele uuenemisele (Mello & Sundstrom, 2023). Ligikaudu 60% kerge astme düsplaasiatest normaliseeruvad aasta jooksul (Mello & Sundstrom, 2023).



Joonis 2. HPV nakkuse kulgemisteed ja võimalik progresseerumine emakakaelavähiks (Modifitseeritud Leach, 2021 järgi).

Harald zur Hausen isoleeris ja tuvastas inimese papilloomiviiruse tüüpide 16 ja 18 DNA emakakaelavähi biopsiatest, näidates et HPV tüüpide 16 ja 18 DNA on aktiivselt transkribeeritud emakakaelavähi rakkudes ja kartsinoomides (Stanley, 2010). Hilisemad uuringud on kinnitanud, et enamikul emakakaelavähki juhtudest on tuvastatud hrHPV (kõrge riskiga HPV, *high risk HPV*) DNA järjestusi (Stanley, 2010). HPV 16 tüüpi leitakse 50–70% juhtudest ja HPV 18 tüüpi 7–20% juhtudest (Stanley, 2010). Pirog *et al*., (2014) uuringus leiti, et emakakaela adenokartsinoomides domineerisid kolm hrHPV genotüüpi – 16, 18 ja 45, mis moodustasid 94.1% HPV-positiivsetest kasvajatest. Samuti leiti, et HPV-18 genotüüp on Euroopas, Aasias ja Põhja-Ameerikas sagedasem kui HPV-16. Siiski tuleks arvestada, et tulemus on kallutatud, sest metaanalüüs hõlmas ka adenokartsinoomi juhtumeid, mis on sagedamini seotud just HPV-18 genotüübiga (Pirog *et al*., 2014). HPV-16 on enim levinud invasiivses emakakaelavähis – 55%, millele järgneb HPV-18 12.8% ning need kaks tüüpi

põhjustavad 70% või isegi rohkem emakakaelavähi juhtudest olenemata geograafilisest asukohast (Stanley, 2010). Esmaseks ennetusmeetmeks on HPV vastu vaktsineerimine, mis on kõige tõhusam juhul, kui seda tehakse enne viirusega kokkupuutumist (Gupta *et al*., 2023). Kuna hrHPV-tüübid on peamiseks etioloogiliseks teguriks emakakaela düsplaasia, emakakaelavähi ning ka teiste anogenitaal- ja pea-kaelapiirkonna vähkide puhul (Stanley, 2010), siis HPV vastu vaktsineerimine aitab ennetada nende teket (Gupta *et al*., 2023). Viimastel aastatel on laboratoorses diagnostikas, teadusuuringutes ja rahvastikupõhises epidemioloogias hakatud järjest enam rakendama NGS (järgmise põlvkonna sekveneerimine, ingl *next generation sequencing*) tehnoloogiaid. NGS-tehnoloogia on näidanud, et selle abil saab täpselt hinnata emakakaelavähi HPV infektsioonist tingitud osakaalu võrreldes teiste meetoditega (Arroyo Mühr *et al*., 2021). HPV molekulaarset testimist soovitatakse ka esmase sõeluuringu meetodina, sest selle tundlikkus on oluliselt kõrgem võrreldes tsütoloogilise uuringuga ning võimaldab tuvastata 60–70% rohkem invasiivseid emakakaelavähi juhte (Ronco et al., 2014). Tundlikemate testidega vähendatakse oluliselt ka valenegatiivsete tulemuste riski ning on leitud, et molekulaarne testimine on efektiivsem ka emakakaela adenokartsinoomide eelastmete tuvastamisel (Ronco et al., 2014).

Nakatumine hrHPV tüvedega eelneb raske astme düsplastiliste kahjustuste kujunemisele. Viiruse DNA on sageli tuvastatav juba enne düsplaasia esmast tsütoloogilist diagnoosimist ning eriti HPV-16 on tugevalt seotud kõrgema astme kahjustuste tekkega (Stanley, 2010). HPV nakatab ja paljuneb ainult täielikult diferentseerunud lameepiteelirakkudes (Stanley, 2010). Viirus nakatab esmalt keratinotsüüte emakakaela epiteeli basaalkihis, tekitades rakkudele mikrotrauma – epiteeli abrasiooni tagajärjel paljastuvad basaalkihi rakud, kus viirus integreerib oma geneetilise materjali peremeesraku DNA-sse (Gupta *et al*., 2023; Stanley, 2010). Kõrge riskiga HPV tüübid toodavad onkoproteiine E6 ja E7, mis kahjustavad rakkude regulatsioonimehhanisme (Gupta *et al*., 2023). E6 seondub kasvaja supressorvalgu p53-ga ja soodustab selle lagunemist, takistades rakul DNA kahjustuse korral tsükli peatamist või apoptoosi (Gupta *et al*., 2023). E7 pärsib aga retinoblastoomi valgu pRb funktsiooni, mistõttu on rakkudel võimalik kontrollimatult jaguneda (Gupta *et al*., 2023). Nende kahe onkoproteiini mehhanismide kaudu on rakkude transformatsioon ja prekantserogeensete muutuste teke soodustatud (Gupta *et al*., 2023). Pikaajaline viirusvalkude ekspressioon viib nakatunud rakkudes geneetiliste muutusteni, sealhulgas mutatsioonid ja kromosomaalsed ebastabiilsused, mis häirivad veel omakorda normaalseid rakkude regulatoorseid mehhanisme, tekitades nii soodsa pinnase emakakaela düsplaasia progresseerumiseks (Gupta *et al*., 2023).

1.2.2 Emakakaela düsplaasia

Düsplaasia terminit kasutatakse emakakaela pinnal ebanormaalsete muutustega rakkude kirjeldamiseks (Martinez *et al*., 2019). Mõiste viitab rakulise küpsemise ja diferentseerumise häirele, mille korral suureneb ebaküpsete ja ebatüüpiliste rakkude hulk, ent muutused piirduvad epiteelkoega ning ei ole levinud läbi basaalmembraani sügavamatesse kudedesse (Mehta & Sachdeva, 2017). Seega düsplaasia ise ei ole emakakaelavähk, kuid võib ravimata jätmisel progresseeruda vähiks ning ka metastaseeruda (Martinez *et al*., 2019). Emakakaela intraepiteliaalne neoplaasia kujuneb välja HPV nakkuse tagajärjel emakakaela epiteelirakkudes, mis leitakse tavaliselt Pap-testi (Papanicolaou test) või emakakaela biopsia tegemisel (Mello & Sundstrom, 2023). Eestis on kasutusel konventsionaalne ehk Pap-test ja vedelikul baseeruv günekotsütoloogiline uuring (ingl *liquid-based cytology*).

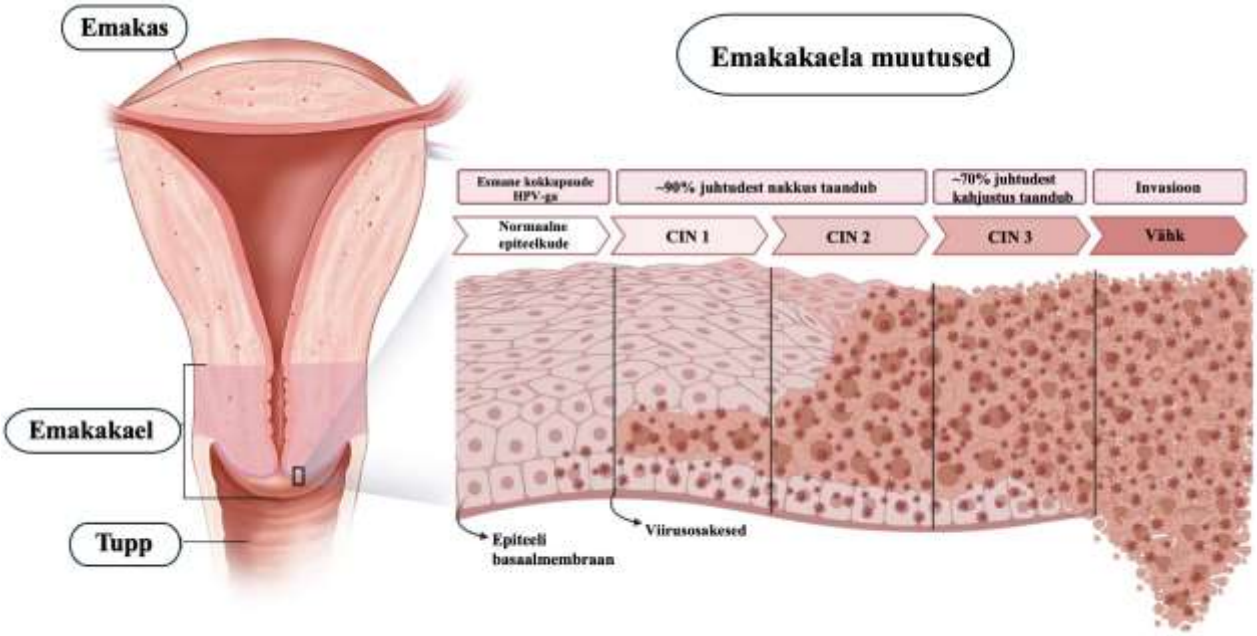
Düsplaasia jagatakse kergeks, mõõdukaks ja raskeks, mida klassifitseeritakse vastavalt CIN 1, CIN 2 ja CIN 3 (Gupta *et al*., 2023). Kliinilises praktikas kasutatakse emakakaela düsplaasia kirjeldamiseks RHK-10[1] (Rahvusvaheline Haiguste Klassifikatsioon, 10. väljaanne) koode, mis võimaldavad haigusjuhtude statistilist ja administratiivset jälgimist. Javanbakht et al., (2023) uuringu kohaselt on emakakaela raskemate astmete düsplaasia levimus maailmas 4,3% (95% usaldusvahemik: 1.8–6.3%) ning madalama astme düsplaasia levimus 6.2% (95% usaldusvahemik: 3.8–9.8%). CIN klassifikatsioon skaalal 1 kuni 3 põhineb rakkude morfoloogial mikroskoobis ning selle, kui suur osa emakakaela epiteelist on mõjutatud (Gupta *et al*., 2023) (Joonis 3):

- CIN 1 ehk kerge düsplaasia (N87.0), puhul on tegemist kergete rakuliste kõrvalekallete esinemisega emakakaela epiteelis. Emakakaela epiteelist mõjutatud üks kolmandik;
- CIN 2 ehk mõõdukas düsplaasia (N87.1), puhul on mõjutatud kolmandik kuni kaks kolmandikku emakakaela epiteelist ja näitab selgemaid rakulisi kõrvalekaldeid;
- CIN 3 ehk raske düsplaasia (N87.2), puhul on tegu ebanormaalsete rakkudega, mis mõjutavad kogu emakakaela epiteeli. Muutused ei ole veel tunginud edasi sidekoesse (Gupta *et al*., 2023). CIN 3 peetakse kartsinoom *in situ*'ks (nimetatakse ka CIS, *carcinoma in situ*) või ka 0 staadiumi emakakaelavähiks (Chen *et al*., 2013), millele vastab RHK-10 kood D06.

[1] https://rhk.sm.ee/

CIN astme määramisel kasutatakse histopatoloogilisi parameetreid nagu tuuma-tsütoplasma suhe, tuuma laienemine, hüperkromaasia, ebaregulaarne kromatiini jaotus ning mitootiliste näitajate esinemine (Waghe & Acharya, 2024). CIN 1 arenemine vähiks on pigem harv: umbes 57% juhtudest regresseerub iseeneslikult, 32% jääb püsima ning ainult 1% juhtudest areneb emakakaelavähiks (Waghe & Acharya, 2024). Seevastu kõrgema astme kahjustustega nagu CIN 2 ja CIN 3, kaasneb märkimisväärselt suurem risk invasiivseks emakakaelavähiks progresseerumiseks (Mello & Sundstrom, 2023). Näiteks CIN 2 puhul regresseerub 43% juhtudest, 35% püsib, 22% progresseerub CIN 3'ks ja 1.5% juhtudest kujuneb emakakaelavähk. CIN 3 puhul on regressiooni tõenäosus vaid 32%, püsima jääb 56% juhtudest ja vähi tekkimise risk ulatub 12%-ni (Sellors & Sankaranarayanan, 2003). Kuna haiguste progresseerumise aeg on mitu aastat, vajavad nii CIN 2 ja CIN 3 suurema tõenäosusega ravi vähi ennetamiseks (Waghe & Acharya, 2024).

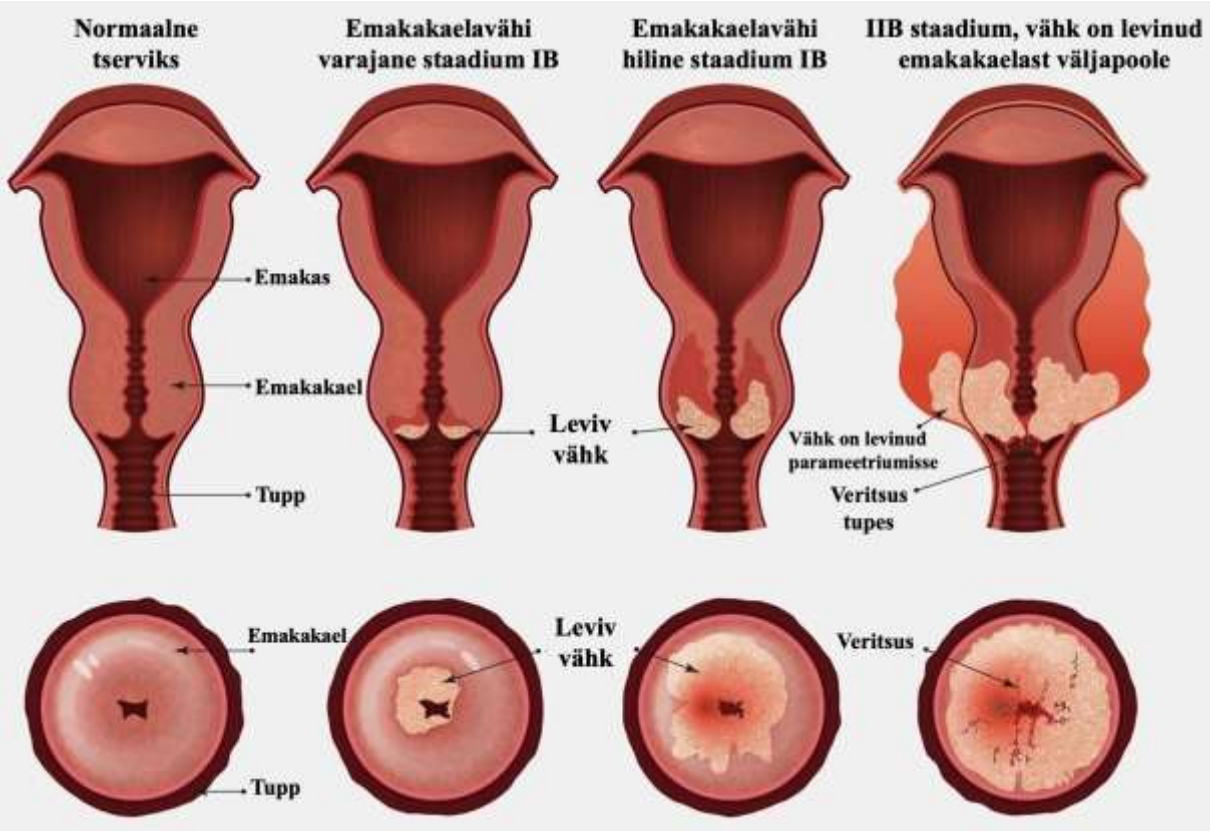
Kerge düsplaasia ravil on esmavalikuks jälgimine, sest enamik CIN 1 juhtudest taandub iseenesest. Siiski soovitatakse kordusanalüüse 6–12 kuu järel. Mõõduka düsplaasia puhul võib noorte (< 25 aastat) ja fertiilses eas naistel kaaluda aktiivset jälgimist, sest ka see võib spontaanselt kaduda. Standardne ravi raskemate düsplaasia astmete korral on kahjustatud koe kirurgiline eemaldamine, sest invasiivseks vähiks progresseerumise risk on oluliselt suurem. Kõigi CIN astmete korral on vajalik regulaarne jälgimine ja HPV testimine pärast ravi, mida tehakse üldjuhul 25–65 aastaste naiste seas. Üle 65-aastastel tekkinud CIN 3 puhul ravi ei alustata, sest on ebatõenäoline, et see jõuaks progresseeruda vähiks (Aavik et al., 2021). Harvadel juhtudel, kui naine ei soovi fertiilsust säilitada, kaalutakse hüsterektoomiat. Relapsi või püsiva infektsiooni korral on vajalik ravi efektiivsuse hindamine ning vajadusel ravi kordamine (Mello & Sundstrom, 2023).



Joonis 3. Emakakaela muutused normaalsest epiteelkoest emakakaelavähini (Modifitseeritud Causin *et al*., 2021 ja *HPV and Pap Test Results*, 2022 järgi).

1.2.3 Emakakaelavähk

Emakakaelavähil on pikk prekliiniline faas, mis võib kesta aastakümneid ilma sümptomiteta (Vu *et al*., 2018). Vähk areneb tavaliselt aeglaselt ning enne vähi ilmnemist läbivad emakakaela rakud düsplastilised muutused (vt eespool) ning sealt edasi hakkavad emakakaela kudedesse ilmuma ebanormaalsed rakud. Kui neid rakke ei hävitata või eemaldata, võivad muutustega rakud hakata arenema vähirakkudeks, hakata kasvama ja levima sügavamatesse kudedesse ja piirkondadesse (Joonis 4) (Waghe & Acharya, 2024).



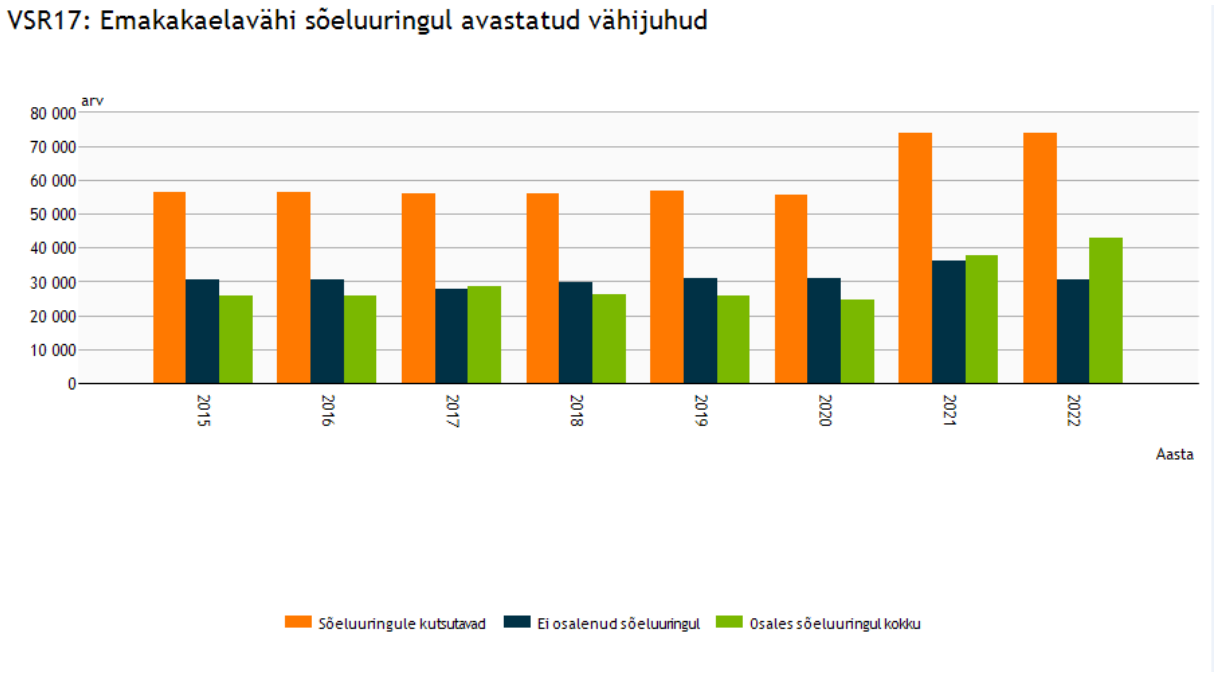
Joonis 4. Emakakaelavähi kliiniline progressioon (Modifitseeritud Mandal, 2009 järgi).

Emakakaelavähi diagnoosimisel kasutatakse kliinilises praktikas rahvusvahelist haiguste klassifikatsiooni. Kõige sagedamini on kasutusel kood C53, mis tähistab emakakaela pahaloomulist kasvajat. Lisaks on kasutusel ka D06 kood, mis vastab CIN 3 staadiumile ja harvem kasutatakse N87.2 koodi, mille puhul on tegu raskete düsplastiliste muutustega. Kaks peamist emakakaelavähi histoloogilist alatüüpi on lamerakk-kartsinoom ja adenokartsinoom (Ngoma & Autier, 2019). Lamerakk-kartsinoom on kõige sagedasem, moodustades kuni 80–90% juhtudest (Oliveira *et al*., 2023), ning on sageli seotud HPV-16 tüübiga (Tewari, 2025). See vähivorm areneb tavaliselt ektotserviksi piirkonnas paiknevatest lamerakkudest ning saab

alguse transformatsioonitsoonist – emakakaela alast, kus lamerakuline ja näärmeline epiteel kohtuvad (Prendiville & Sankaranarayanan, 2017). Adenokartsinoomid arenevad endotserviksi näärmerakkudest ja moodustavad ligikaudu 20–25% juhtudest (Tewari, 2025). Need võivad samuti pärineda transformatsioonitsoonist või selle juures paikevast näärmeepiteelist ning on sagedamini seotud HPV-18 infektsiooniga (Prendiville & Sankaranarayanan, 2017; Tewari, 2025). Nende eellasteks peetakse adenokartsinoom *in situ*, mis võib tsütoloogilise uuringu põhjal avastamata jääda (Tewari, 2025). Mõnikord esineb emakakaelavähk segatüübilisena, omades mõlema kartsinoomi tunnuseid ning sel juhul nimetatakse seda segakartsinoomiks ehk *adenosquamous*-kartsinoomiks, kuid need on pigem haruldased ja moodustavad umbes 3% juhtudest (Tewari, 2025).

Ülemaailmselt on emakakaelavähk reproduktiivses eas naiste seas levimuselt, haigestumuselt ja suremuselt teisel kohal (Wu *et al*., 2025). 2024. aastal prognoositi Ameerika Ühendriikides kokku 13 820 uut emakakaelavähi juhtu ja 4360 haigusega seotud surmajuhtu. Euroopa Liidus samal aastal aga 58 169 uut juhtu, millest 56% pärinevad Kesk- ja Ida-Euroopa riikidest, ning hinnanguliselt 22 989 surmajuhtumit (Tewari, 2025). Eriti suur on haiguse levimus madala inimarengu indeksiga riikides, kus ennetus- ja ravivõimalused on piiratud (Wu *et al*., 2025). Emakakaelavähi diagnoosimise mediaanvanus on 50 eluaastat (Tewari, 2025). Ülemaailmse koormuse vähendamiseks käivitas Maailma Terviseorganisatsioon (WHO) 2020. aastal ülemaailmse strateegia haiguse likvideerimise kiirendamiseks – eesmärgiks on, et 2030. aastaks peaks olema 90% tüdrukutest HPV vastu vaktsineeritud enne 15. eluaastat, vähemalt 70% naistest vanuses 35 ja 45 eluaastat peaks olema kaasatud sõeluuringutesse ning 90% naistest, kellel diagnoositakse vähieelsed kahjustused või emakakaelavähk, peaks saama asjakohast ravi (Wu *et al*., 2025). Sõeluuringute abil saab emakakaela kahjustusi varakult avastada (Vu *et al*., 2018). Emakakaelavähi ravi sõltub haiguse staadiumist ja patsiendi üldisest tervislikust seisundist. Eestis on kasutusel kirurgiline ravi, mis viiakse läbi varajasest staadiumis. Võimalikeks lähenemisteks on anesteesias tehtud emakakaela konisatsioon (eemaldatakse osa emakakaelast), hüsterektoomia (eemaldatakse emakas ja emakakael täielikult) või trahhelektoomia (eemaldatakse vähirakkudega osa emakakaelast, tupe ülaosa, lümfisõlmed ja ümbritsev kude) (Small Jr *et al*., 2017). Kaugemale arenenud staadiumis kasutatakse patsiendi ravimiseks ka keemia- ja kiiritusravi, radiokemoteraapiat ja sihtmärkravi (Small Jr *et al*., 2017). Keemiaravi on kasutusel peamiselt sümptomite leevendamiseks ja nende patsientide eluea pikendamiseks, kellel hilise staadiumi tõttu ei ole enam võimalik kasvajast terveneda (Small Jr *et al*., 2017).

Kuigi emakakaelavähi sõeluuringuprogramm on Eestis kestnud peaaegu kakskümmend aastat, on haiguse esinemissagedus riigis endiselt Euroopa kõrgeimate seas (Rigby *et al*., 2024; Tisler *et al*., 2024). 2020. aasta andmete järgi diagnoositakse Eestis igal aastal umbes 196 uut emakakaelavähi juhtu. Emakakaelavähk on naiste seas levimuselt 5. kohal kõikide vähivormide hulgas ning 15–44-aastaste naiste seas levimuselt 2. kohal. Emakakaelavähi haigestumus Eestis ulatub 16.3 juhuni 100 000 naise kohta aastas ning surmajuhtude määr on 5.6/100 000 naist aastas, surm saabub sageli keskmisel 64. eluaastaks. Eestis on emakakaelavähi suremus Euroopa Liidus üks kõrgemaid, kuigi üle riigi on olemas nii sõeluuringud kui ka HPV vaktsineerimine (Rigby *et al*., 2024). Eestis rakendatud HPV vaktsineerimisprogrammi puhul on tegu esmase ennetusmeetmega, suunatud 12–14 aastastele, kuid vaktsineeritus on jäänud alla WHO soovitatud taseme (90%). Tervise Arengu Instituudi statistika järgi oli 2024. aastal HPV vastu vaktsineeritud 54% 12- kuni 14-aastastest tütarlastest ja 41% samas vanuseastmes poistest. Enamiku emakakaelavähi juhtude algpõhjuseks on HPV infektsioon, eeskätt kõrge riskiga tüvedega 16 ja 18. Eestis on hrHPV seotud hinnanguliselt 74% emakakaelavähi juhtudega, HPV-16 on seotud ligikaudu 52.5% ja HPV-18 11.2% emakakaelavähi juhtudest, ülejäänud ligikaudse 10% moodustavad teised hrHPV genotüübid (nt 31, 33, 45). Nende genotüüpide levik suureneb vastavalt haiguse raskusastmele: 27.1% CIN 1; 54.5% CIN 2 ja CIN 3 ning 74% invasiivse emakakaelavähi juhtudest (Human Papillomavirus and Related Diseases Report: ESTONIA, 2023). Lisaks HPV-le on teada mitmeid täiendavaid riskitegureid, mis suurendavad haiguse kujunemise tõenäosust: suitsetamine, HIV-nakkus, varajane seksuaalelu algus, suukaudsete hormonaalsete rasestumisvastaste vahendite pikaajaline kasutamine ning rohke seksuaalpartnerite arv. Nende tegurite koostoime ning HPV-nakkuse selgelt määratletav roll muudavad emakakaelavähi suurel määral ennetatavaks haiguseks (Wu *et al*., 2025). Regulaarses sõeluuringuprogrammis, mille puhul on tegu sekundaarse ennetusmeetmega, osaleb sihtrühmast vaid ligikaudu 47%, mis jääb samuti alla WHO soovitatud osalusmäära – 70% (Joonis 5) (Human Papillomavirus and Related Diseases Report: ESTONIA, 2023). Vähene osalus emakakaelavähi sõeluuringutes on tingitud mitmes tegurist, näiteks piiratud teadlikkusest, kultuurilistest uskumustest ning logilistilistest takistustest, nagu ligipääs tervishoiuteenustele või ajapuudus (Cadman *et al*., 2012; Chaw *et al*., 2022). Lisaks on täheldatud, et varasemad traumaatilised sündmused, nagu seksuaalne väärkohtlemine või eelnevad valulikud günekoloogilised uuringud, võivad viia naised selleni, et keeldutakse sõeluuringutest (Ragupathy & McMullen, 2025).



Joonis 5. Emakakaelavähi sõeluuringule kutsutute ja osalenute arv aastatel 2015–2022. Andmed pärinevad Tervise Arengu Instituudi vähi sõeluuringute registrist[2]. Näidatud on sõeluuringutele kutstutud naiste koguarv, osalenute ja mitteosalenute arv.

Geneetilised seoseuuringud on näidanud, et emakakaelavähk on osaliselt päritav. Näiteks Rashkin *et al*., (2020) uuringus leiti emakakaelavähi päritavuseks 7% (95% CI 2–12%), samas Koel *et al*., (2023) leidsid mitte-HLA piirkonna pärilikkuseks 4.75%. Varasemates uuringutes on täheldatud kõrgemaid pärilikkuse määrasid – Czene *et al*., (2002) leidsid invasiivse CC päritavuseks 22% (95% CI 14–27%) ning CIN 3 13% (95% CI 6–15%). Magnusson *et al*., (2000) uuringus hinnati CC päritavuseks koguni 27% (95% CI 26–29%). Sellised tulemused viitavad, et lisaks keskkonnateguritele, just eriti HPV nakkus, on haiguse kujunemises oluline roll peremeesorganismi geneetilisel eelsoodumusel. Tuvastatud on mitmeid geneetilisi riskilookuseid, sealhulgas ka HLA piirkonnas, mis viitavad immuunvastuse geneetilisele taustale ja selle potentsiaalsele rollile HPV elimineerimisel või püsimisel. Käesolevas töös tutvustatakse järgmistes peatükkides ülegenoomse seoseuuringu üldiseid põhimõtteid ja tuvastatud emakakaelavähi riskilookuseid.

[2] https://statistika.tai.ee/

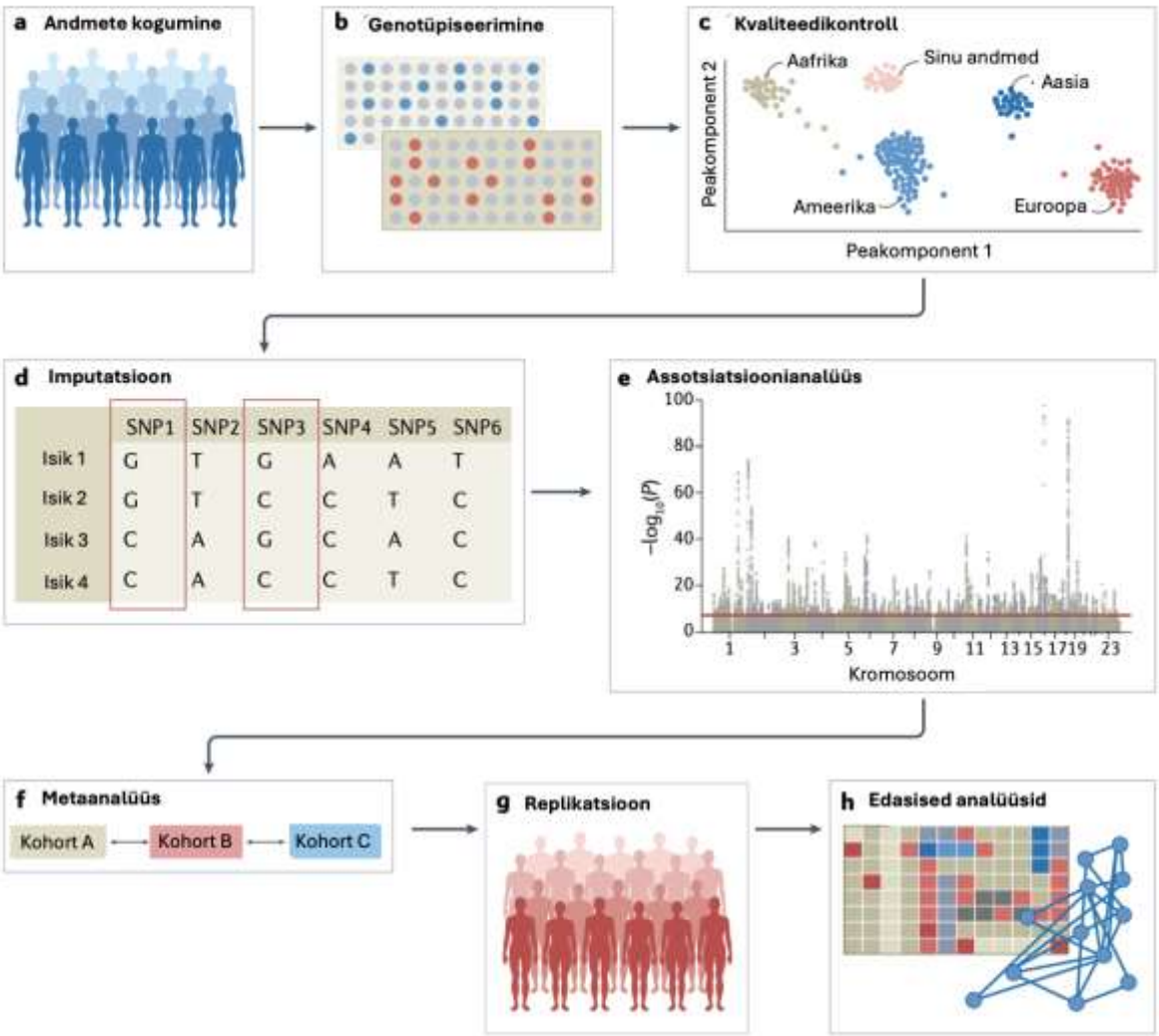
1.3 Ülegenoomne seoseuuring

Ülegenoomse seoseuuringu (*genome-wide association study*, edaspidi GWAS) puhul on tegemist uuringuga, mille käigus genotüpeeritakse sadu tuhandeid ühe nukleotiidi polümorfisme (SNP) üle kogu genoomi ning testitakse nende seost huvipakkuva fenotüübiga (Zeng *et al*., 2015). Uuringute eesmärgiks on tuvastada genotüübi ja fenotüübi vahelisi seoseid, võrreldes geneetiliste variantide alleelisagedusi fenotüübiliselt erinevate, kuid esivanemliku tausta poolest sarnaste indiviidide hulgas (Uffelmann *et al*., 2021). Tavapäraste hüpoteesipõhiste uuringute asemel, mis lähtuvad kindlatest kahtlustest konkreetse geeni osas, läheneb GWAS küsimusele laiemalt ja ilma eeldusteta – uuritakse geneetilist varieeruvust üle kogu genoomi, et leida seosed geenivariatsioonide ja tunnuste vahel (Kitsios & Zintzaras, 2009). Seetõttu nimetatakse seda ka hüpoteesivabaks või agnostiliseks meetodiks, sest enne analüüsi ei ole eelduseid, millised geenid võiksid konkreetse fenotüübiga seotud olla (Kitsios & Zintzaras, 2009). GWAS uuringud on tuvastanud tuhandeid riskilookuseid, mis on seotud paljude komplekshaiguste ja tunnustega, näiteks mitmesugused vähivormid, bipolaarne häire, pärgartertõbi, Crohni tõbi, hüpertensioon jne (Zeng *et al*., 2015). Uuringu tulemused hõlmavad tavaliselt genoomi regioonide määratlemist, kus mitmed omavahel tugevas korrelatsioonis olevad SNP-id näitavad statistiliselt olulist seost uuritava tunnusega, selliseid regioone nimetatakse genoomseteks riskilookusteks (Uffelmann *et al*., 2021). Varasematest GWAS uuringutest on täheldatud, et enamikku fenotüüpe mõjutavad tuhanded geneetilised variandid, mille individuaalne mõju on suhteliselt väike (Uffelmann *et al*., 2021).

Kaasaegsed ülegenoomsed assotsiatsiooniuuringud kasutavad ka aheldatuse tasakaalustamatust (*linkage disequilibrium*, LD), mis peegeldab DNA variantide vahelist korrelatsiooni (Visscher *et al*., 2017). LD olemasolu suurendab küll statistilist võimekust tuvastada assotsiatsioone DNA variantide ja fenotüüpide vahel, kuid samal ajal raskendab see otseste põhjuslike ja bioloogiliste mehhanismide eristamist, kuna põhjuslikud ja mittepõhjuslikud variandid on sageli omavahel korreleerunud (Uffelmann *et al*., 2021). See, kuidas erinevad tegurid mõjutavad LD-i konkreetse lookuste paari vahel või genoomses piirkonnas, sõltub lookuses olevatest rekombinatsioonimääradest (Slatkin, 2008). Seega sõltub GWAS-i tulemuste tõlgendamine ja põhjuslike variantide tuvastamine nii uuritava valimi suurusest, populatsioonis esinevate põhjuslike variantide mõjusuuruste jaotusest, nende esinemissagedusest kui ka LD-st täheldatud genotüüpide ja põhjuslike variantide vahel (Visscher *et al*., 2017).

Ülegenoomset assotsiatsiooniuuringut saab teostada, kasutades andmeid biopankadest, haiguspõhistest või elanikkonnapõhistest kohortidest, samuti ka otse tarbijale suunatud ettevõtete andmebaasidest. GWAS-i eksperimentaalne töövoog hõlmab järgmisi etappe (Joonis 6) (Uffelmann *et al*., 2021):

- DNA ja fenotüübilise teabe (näiteks vanus, sugu, haiguse staatus) kogumine indiviidide rühmalt;
- iga indiviidi genotüpiseerimine olemasolevate SNP-kiipide või genoomi sekveneerimise strateegiate abil;
- kvaliteedikontroll;
- genotüpiseerimata variantide imputatsioon, kasutades haplotüübi faasimist ja referentspaneeli;
- statistiline testimine;
- metaanalüüs (valikuline);
- sõltumatu replikatsioon mõnes teises valimis;
- tulemuste tõlgendamine erinevate jätkuanalüüsidega.



Joonis 6. Skemaatiline ülevaade GWAS analüüsi läbiviimisest (Modifitseeritud Uffelmann *et al*., 2021 järgi).

1.4 Emakakaelavähi senised ülegenoomsed seoseuuringud

Senini teostatud suuremad emakakaelavähi ülegenoomsed seoseuuringud on kokkuvõtvalt välja toodud Tabelis 1. Väljatoodud uuringute ülesehitus ning uuritud populatsioonid varieeruvad oluliselt. Lisaks uuringu ülesehitusele varieeruvad ka uuritavate fenotüüpide definitsioonid ja osalejate päritolu. Uuringutesse kaasatud erineva geograafilise ja etnilise taustaga populatsioonid, nagu Euroopa, Aasia ning Ladina-Ameerika, võimaldavad uurida tulemuste ülekantavust erineva geneetilise taustaga populatsioonides. Assotsiatsiooniuuringud on tuvastanud mitmeid lookusi, millel on statistiliselt oluline seos emakakaelavähi riskiga. Korduvalt on näidanud seost HLA piirkonna geenid (nt *HLA-DPB2*, *HLA-DRB1/DQA1*), samuti geenid nagu *PAX8*, *CLPTM1L* ja *PRKD1*. Märkimisväärseks võiks pidada ka seda, et osades uuringutes on identifitseeritud sama geneetiline marker sõltumata uuritavast populatsioonist, mis viitab nende lookuste võimalikule universaalsele olulisusele haiguse kujunemisel. Sarnaselt teiste komplekshaigustega on valimimahu kasv ning metaanalüütilised lähenemised suurendanud tuvastatud lookuste arvu.

Mõnes uuringus on keskendutud vaid emakakaelavähi juhtudele, samas kui teistes kaasati ka emakakaela düsplaasia erinevad raskusastmed (nt CIN 2 ja CIN 3). Juhtude tuvastamiseks kasutatakse üldjuhul RHK-10 koode, mis võimaldavad uurida erinevaid invasiivseid vähijuhtumeid (nt C53) ja vähieelseid seisundeid (nt D06, N87). Näiteks Takeuchi *et al*., (2019) ja Ramachandran *et al*., (2022) uuringutes käsitleti nii invasiivset vähki kui ka düsplaasiat, võimaldades hinnata geneetiliste markerite seost haiguse eri etappidega. Koel *et al*., (2023) uurisid emakakaelavähi ja emakakaela düsplaasia geneetilisi seoseid, kasutades Eesti Geenivaramu ja FinnGeni andmeid. Leiti, et mõlemad haigusjuhud jagavad mõningaid ühiseid geneetilisi lookuseid, kuid esineb ka olulisi erinevusi. Näiteks tuvastati *PAX8/PAX8-AS1* ühise riskilookusena nii düsplaasia kui vähi puhul, viidates sellele, et sellel geenil on oluline roll mõlema haigusjuhu puhul. Samas avastati emakakaelavähi puhul täiendavaid spetsiifilisi lookuseid, nagu *CDC42* ja *GSDMB*, mida düsplaasia puhul aga ei täheldatud, mis tähendab seda, et teatud geneetilised riskitegurid võivad olla seotud ainult haiguse progresseerunud vormiga. Kuigi mitmes töös on leitud kattuvaid lookuseid, mis viitavad potentsiaalsetele ühistele geneetilistele riskiteguritele, esineb ka olulisi erinevusi leitud seostes. Üheks võimalikuks põhjuseks võib olla see, kuidas on defineeritud juhud ja kontrollid – kas on kaasatud ainult invasiivsed juhtumid või ka vähieelsed seisundid ning kui täpselt on fenotüübid määratletud.

Edasiste uuringute paremaks planeerimiseks võiks standardiseerida fenotüüpide määratlemise lähenemised. Erinevates emakakaelavähi GWAS uuringutes kasutatakse varieeruvaid fenotüüpide määratlusi – osades piirdutakse invasiivsete vähijuhtudega, teistes aga kombineeritakse vähk ja selle eelastmed ühiseks fenotüübiks. Sellest tulenevalt võiks välja töötada soovitusliku käsitluse RHK-10 koodide kombineerimiseks või välistamiseks vastavalt uuritavale hüpoteesile. Standardiseerimine tagaks ka tulemuste parema võrreldavuse, integreeritavuse ja reprodutseeritavuse.

Tabel 1. Kokkuvõte varasemate ülegenoomsete emakakaelavähi seoseuuringute tulemustest. CC – emakakaelavähk (*cervical cancer*); CIN – emakakaela intraepiteliaalne neoplaasia (*cervical intraepithelial neoplasia*); EA – efektialleel (*effect allele*); AA – alternatiivne alleel (*alternative allele*); OR – šansside suhe (*odds ratio*); 95% CI – 95% usaldusintervall (*confidence interval*). Mõnede markerite puhul ei leitud täpseid OR ja p-väärtusi või olid need esitatud kombineeritult mitme uuringuetapi kohta. Paksus kirjas on välja toodud geenid, mis toodi välja mitmes uuringus.

Marker (EA)	Geen	Juhu definitsioon	OR (95% CI)	p-väärtus	Juhud/kontrollid	Viide
rs13117307 (T)	*EXOC1* (4q12)	CC + CIN	1.26 (1.16–1.36)	9.7×10^{-9}	1364/3028	(Shi et al., 2013)
rs8067378 (G)	***GSDMB* (17q12)**		1.18 (1.11–1.25)	2.0×10^{-8}		
rs4282438 (G)	***HLA-DPB2* (6p21.32)**		0.75 (0.71-0.79)	4.5×10^{-27}		
rs9277952 (A)	geenidevaheline piirkond (6p21.32)		0.85 (0.81-0.90)	2.3×10^{-9}		
rs2516448 (T)	*MICA* (6p21.3)	CC	1.42 (1.31–1.54)	1.6×10^{-18}	1075/4014	(Chen et al., 2013)
rs9272143 (C)	*HLA-DRB1/DQA1*		0.67 (0.62–0.72)	9.3×10^{-24}		
rs3117027 (A)	***HLA-DPB2***		1.25 (1.15–1.35)	4.9×10^{-8}		
rs59661306(A)	5q14 (*ADGRV1*)	CC	0.77 (0.74-0.79)	2.4×10^{-11}	4070/8007	(Takeuchi et al., 2019)
rs2973455	5q14 (*ARRDC3*)		NA	3.4×10^{-5}		
rs7457728 (C)	7p11 *(VSTM2A)*		1.19 (1.16-1.21)	1.2×10^{-8}		
rs8067378 (A)	***GSDMB* (17q12)**	CC	0.82	3.3×10^{-6}		

rs10175462 (A)	***PAX8***	CIN3+CC	0.87 (0.84–0.91)	1.0×10^{-9}	4769/145545	(Bowden et al., 2021)
rs27069 (C)	***CLPTM1L***		0.88 (0.84–0.92)	2.5×10^{-9}		
rs6938453 (A)	*MICA*		0.79 (0.75-0.83)	2.0×10^{-17}		
rs55986091 (A)	***HLA-DQB1***		0.66 (0.60-0.72)	6.4×10^{-7}		
rs9272050 (G)	***HLA-DQA1***		1.27 (1.21-1.32)	2.5×10^{-28}		
rs9266183 (C)	***HLA-B***		0.73 (0.64–0.83)	1.5×10^{-6}		
rs225902 (C)	*PRKD1*	CC	1.35 (1.18–1.55)	9.7×10^{-5}	2364/1667	(Ramachandran et al., 2022)
rs1048943 (G)	*CYP1A1*	CC	1.75 (1.49–2.05)	1.0×10^{-5}	8313/9325	(Wang et al., 2015)
rs12603332 (C)	*ORMDL3* (17q12)	CC	0.90 (0.88–0.93)	1.2×10^{-9}	8624/400573	(Koel et al., 2023)
rs2268177 (T)	*CDC42*		1.12 (1.07-1.16)	3.1×10^{-8}		
rs4849177 (C)	***PAX8***		0.87 (0.85-0.91)	9.4×10^{-15}		
rs27069 (T)	*CLPTM1L*		0.88 (0.85-0.91)	1.3×10^{-14}		
rs35508382 (G)	***HLA-DQA1***		0.67 (0.63-0.71)	1.0×10^{-39}		

rs1049137 (G)	*PSD4*	N87	0.92 (0.91–0.94)	3.9×10^{-10}	33979/193452	(Koel et al., 2023)
rs1049137 (G)			0.92 (0.91–0.94)	6.4×10^{-9}		
rs12611652 (A)	*DAPL1*		0.92 (0.91–0.94)	3.2×10^{-9}		
rs6866294 (C)	*TERT*		1.08 (1.06–1.10)	2.1×10^{-9}		
rs1053726 (G)	***HLA-B***		0.91 (0.88–0.95)	9.1×10^{-9}		
rs36214159 (G)	***HLA-DQA1***		0.79 (0.76–0.83)	1.6×10^{-22}		
rs4282438 (G)	***HLA-DPB1/DPB2***	CSCC (emakakaela lamerakk-kartsinoom)	0.67 (0.55–0.80)	1.5×10^{-5}	507/432	(Yang et al., 2018)
rs8067378 (G)	***GSDMB* (17q12)**	CC (invasiivne)	1.47 (1.12–1.94)	7.2×10^{-3}	214/288	(Miura et al., 2016)

2. Eksperimentaalosa

2.1 Töö eesmärk

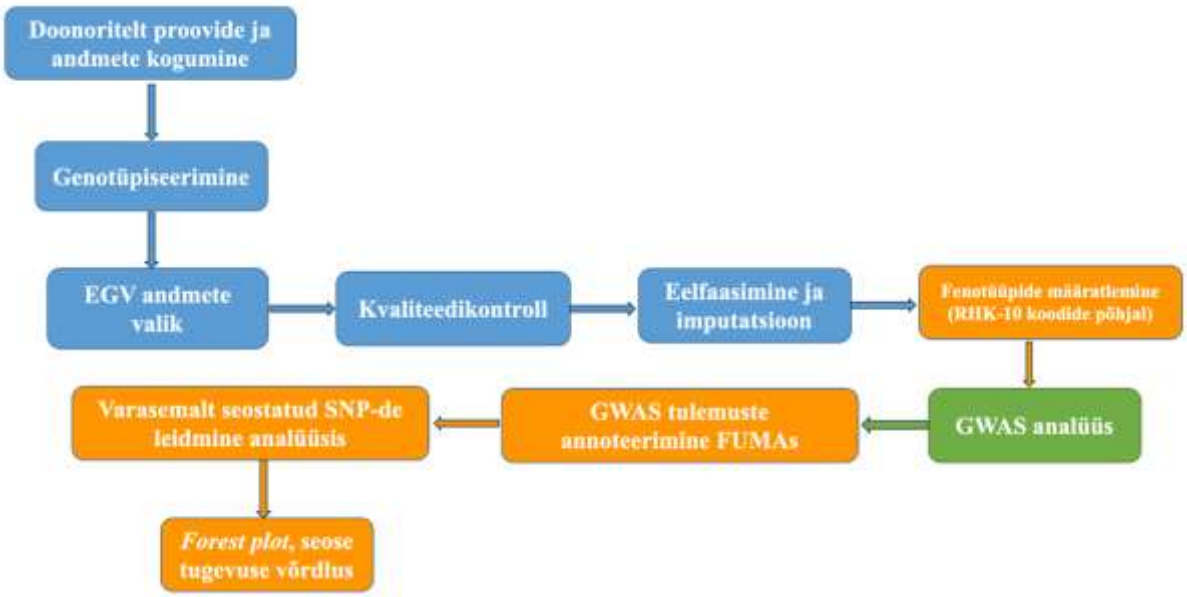
Käesoleva magistritöö eksperimentaalse osa eesmärgiks on hinnata, kuidas erinevad fenotüübi definitsioonid mõjutavad emakakaelavähi geneetiliste seoseanalüüside tulemusi, kasutades selleks Eesti Geenivaramu andmeid.

2.2 Materjal ja metoodika

2.2.1 Valim ja uuritavad tunnused

Antud uurimuse läbiviimiseks kasutati Tartu Ülikooli Eesti Geenivaramu (TÜ EGV) proove, mis on kogutud üle Eesti vabatahtlikelt geenidoonoritelt (Milani *et al*., 2025). Geenivaramule vereproove annetanud isikud allkirjastasid laialdase nõusoleku, mis võimaldab osalejate andmeid regulaarselt täiendada, saades andmeid Eesti Haigekassast ja Tervise Infosüsteemist ning teistest riiklikest terviseandmebaasidest (Milani *et al*., 2025). Osalejate DNA proovid on genotüpiseeritud ning biopanga andmeid kasutatakse palju erinevates rahvusvahelistes GWAS projektides ning 2024. aasta seisuga on EGV-s 212 000 osalejat, peamiselt Euroopa päritolu, moodustades ligikaudu 20% täiskasvanud Eesti elanikkonnast (Milani *et al*., 2025). EGV-s on üleesindatud just naissoost osalejad ning reproduktiivses eas naiste suur osakaal teeb sellest ainulaadse andmestiku, mis võimaldab põhjalikult uurida naiste reproduktiivtervist (Milani *et al*., 2025).

Uuringu töövoog on illustreeritud joonisel (Joonis 7), kus on välja toodud peamised etapid andmete kogumisest kuni analüüsitulemusteni.



Joonis 7. Analüüsietappide töövoog. Sinisega on tähistatud tööetapid mis viidi läbi tsentraalselt tuumiklaboris. Oranžiga on märgitud etapid, mille viis läbi uuringu autor. Rohelisega on tähistatud indiviiditasemel analüüsid (GWAS), mille teostas Mariann Võsa.

Uurimistöös käsitletud fenotüübid on klassifitseeritud RHK-10 koodide alusel. Selles uurimistöös keskendutakse fenotüüpidele, mis on seotud emakakaela vähi ja vähieelsete seisunditega. Need on tähistatud koodidega C53, D06, N87 ning nende erinevate kombinatsioonidega (Tabel 2 ja Lisa 1). C53 tähistab emakakaela pahaloomulist kasvajat (ehk emakakaelavähki), mis on üks levinumaid günekoloogilisi vähivorme. D06 viitab emakakaela *in situ* kartsinoomile ehk varajases staadiumis vähieelsele seisundile, kus vähirakud ei ole veel invasiooni alustanud. N87 kood tähistab emakakaela düsplaasiat, mis on rakkude ebanormaalne areng ja võib viia vähieelse seisundini ning ka vähini. Erinevad düsplaasia astmed on tähtsad haiguse progresseerumise hindamisel. Düsplaasia raskusastmed on RHK-10 klassifikatsioonis täpsustatud alakoodidega: N87.0 tähistades emakakaela kerget düsplaasiat, N87.1 mõõdukat düsplaasiat ja N87.2 raskekujulist düsplaasiat. Astmed kajastavad emakakaela muutuste tõsidust ning võimalust haiguse progresseerumiseks vähieelseks seisundiks või vähiks. Diagnoosikoodide kombinatsioonid, nagu C53_D06_N87 või C53_D06_N87.2, tähistavad uurimisgruppide juhtude kombinatsioone. Näiteks C53_D06_N87 grupi puhul on tegu laia definitsiooniga, mis hõlmab naisi, kellel on olnud kas emakakaelavähk (C53), emakakaela *in situ* kartsinoom (D06) või emakakaela düsplaasia (N87). Grupp C53_D06_N87.2 on aga kitsam, sinna kuuluvad raskemad seisundid – emakakaelavähk, *in situ* kartsinoom või raskekujuline düsplaasia (N87.2). Kõige konkreetsemad on aga ühe RHK-koodiga defineeritud grupid, näiteks C53, mis hõlmab ainult emakakaelavähi juhtumeid ning seepärast võib seda pidada selgelt määratletud fenotüübiks. Kontrollidena kasutati isikuid, kellel vastavad diagnoosikoodid puudusid. Emakakaelaga seotud haiguslike fenotüüpide ja kombineeritud

fenotüüpide uurimine on oluline, mõistmaks kuidas erinevad fenotüübidefinitsioonid mõjutavad ülegenoomsete seoseuuringute tulemusi.

Käesolevas töös esitatud analüüsid on teostatud Eesti bioeetika ja inimuuringute nõukogu poolt väljastatud eetikalubade 1.1-12/2732 ning 1.1-12/624 raames. Indiviiditasemel analüüsid teostas Mariann Võsa.

Tabel 2. Fenotüüpide klassifikatsioon ja osalejate jaotus

Fenotüüp	Fenotüübi definitsioon	Lühend	Juhud/kontrollid
C53	emakakaela pahaloomuline kasvaja	CC	904/120437
C53_D06	kombineeritud emakakaela pahaloomuline kasvaja ja emakakaela *in situ* kartsinoom	CC + CIS	1171/120213
D06	emakakaela kartsinoom *in situ*	CIS	441/120213
N87	emakakaela düsplaasia	Dys	18130/116834
C53_D06_N87	kombineeritud emakakaela pahaloomuline kasvaja, emakakaela *in situ* kartsinoom ja emakakaela düsplaasia	CC + CIS + Dys	18689/103789
C53_D06_N87.2	kombineeritud emakakaela pahaloomuline kasvaja, emakakaela *in situ* kartsinoom ja raske emakakaela düsplaasia	CC + CIS + Dys	4260/117124

2.2.2 Genotüpiseerimine, kvaliteedikontroll, imputatsioon

Käesolevas magistritöös kasutati olemasolevaid genotüübiandmeid. TÜ EGV osalejate proovide genotüpiseerimine, andmete kvaliteedikontroll ja imputeerimine viidi läbi keskselt TÜ Eesti Geenivaramu tuumiklaboris. Genotüpiseerimisel kasutati *Illumina* GSAv.1.0, GSA2.0 ja GSAv2.0_EST kiipe[3].

Genotüpiseerimise andmete kvaliteedikontrolliks kasutati järgnevaid parameetreid:

- edukalt genotüpiseeritud markerite protsentuaalne osakaal (*call rate*, CR);
- genotüübilise ja fenotüübilise soo mittevastavus;
- Hardy-Weinbergi tasakaalustatus (*Hardy-Weinberg equilibrium*, HWE);
- harvema alleeli sagedus (*minor allele frequency*, MAF).

Kvaliteedikontrolli käigus eemaldati andmestikust isikud, kelle genotüübilise ja fenotüübilise soo vahel esines ebakõla või kelle CR jäi alla 95%. Samuti jäeti välja geneetilised variandid, mille CR < 95%, HWE testi p-väärtus < 10^{-4} (autosomaalsed variandid) ning MAF < 1%. Geneetiliste variantide positsioonid määrati inimese referentsgenoomi versiooni 37 (GRCh37)

[3] https://www.illumina.com/

järgi. Eelfaasimiseks kasutati Eagle v2.3 tarkvara (Loh *et al*., 2016). Imputatsiooniks kasutati Beagle v.28Sep18.793 tarkvara (Browning & Browning, 2007) ning Eesti-spetsiifilist referentspaneeli, mis põhineb 2297 geenidoonori genoomi sekveneerimise andmetel (Mitt *et al*., 2017).

2.2.3 Assotsiatsioonianalüüs TÜ Eesti Geenivaramu andmetel

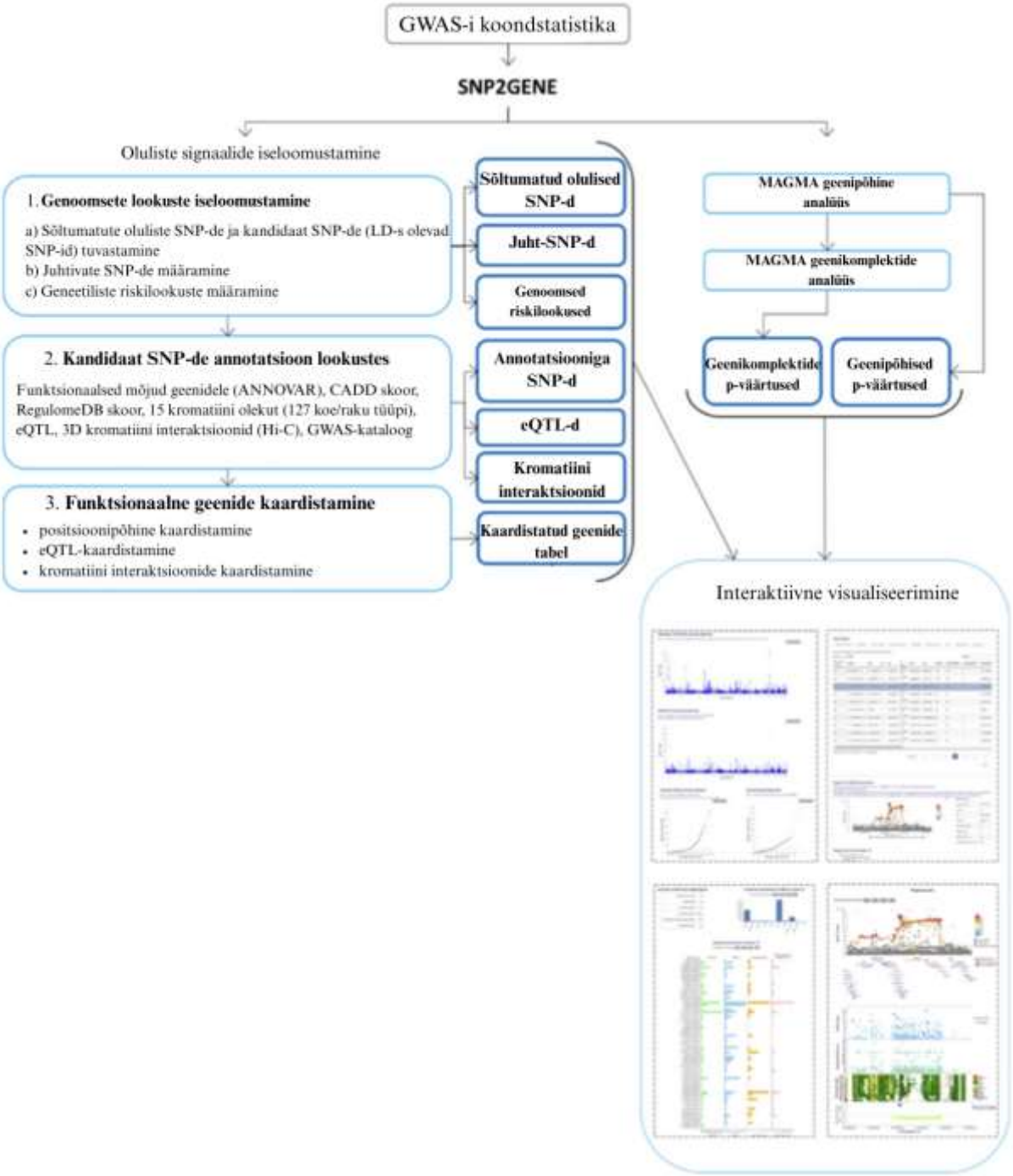
Assotsiatsioonianalüüs TÜ EGV andmetel viidi läbi kasutades tarkvarapaketti REGENIE v2.2.4 (Mbatchou *et al*., 2021). REGENIE on kiire, mälusäästlik ja võimaldab analüüsida mitut tunnust korraga ning seetõttu on eriti sobiv suurte biopankade andmestike analüüsimiseks (Mbatchou *et al*., 2021). REGENIE võimaldab arvestada populatsiooni struktuuri ja isikutevahelist sugulust (Mbatchou *et al*., 2021). Analüüsis kasutati kovariaatidena sünniaastat ja kümmet esimest geneetilist peakomponenti (PC), et kontrollida populatsiooni struktuuri ja võimalikke segavaid faktoreid. Geenivariandid, mille imputatsiooni kvaliteedinäitaja (*imputation INFO score*) oli < 0.4 ning harvema alleeli koguarv < 5, jäeti analüüsist välja. Analüüsi teostas Mariann Võsa.

2.2.4 GWAS tulemuste annoteerimine ja visualiseerimine FUMA veebiplatvormi abil

Ülegenoomse seoseuuringu väljundiks on koondstatistikute fail, kus iga testitud SNP-i kohta on toodud välja üldine informatsioon: kromosoom, positsioon, efektialleel, alternatiivne alleel, efekti suurus (beeta) ja selle suund ning assotsiatsioonitesti p-väärtus. GWAS-analüüsi tulemuste tõlgendamiseks kasutati FUMA veebiplatvormi[4], mis on loodud geneetiliste seoseanalüüsi tulemuste annotatsiooniks ja interpreteerimiseks (Watanabe *et al*., 2017). FUMA koondab mitmetest bioloogilistest andmebaasidest pärinevat teavet, et lihtsustada GWAS tulemuste tõlgendamist ning tulemuste interaktiivset visualiseerimist (Watanabe *et al*., 2017). FUMA võimaldab kandidaatgeenide kaardistamist nii positsioonipõhiselt kui ka geeniekspressiooni taset mõjutavaid lookusi (eQTL) arvestades, samuti kromatiini interaktsioonide põhjal (Watanabe *et al.*, 2017). Lisaks pakub FUMA geenipõhiseid ja koespetsiifilisi rikastusanalüüse ning seob leitud geneetilised variandid GWAS kataloogi andmetega, võimaldades uurijal hinnata, kas tuvastatud signaalid on juba varasemalt teada või potentsiaalselt uued (Watanabe *et al*., 2017). Visualiseerimiseks saab FUMA-s kasutada

[4] https://fuma.ctglab.nl/

erinevaid graafilisi väljundeid, näiteks Manhattan'i jooniseid, geenikaarte ja koespetsiifilisi ekspressioonmustreid (Joonis 8) (Pilalis *et al*., 2025).



Joonis 8. FUMA SNP2GENE töövoog GWAS-i tulemuste funktsionaalseks annoteerimiseks ja visualiseerimiseks (Modifitseeritud Watanabe et al., 2017 järgi).

2.2.5 Fenotüübi definitsiooni mõju hindamine

Fenotüübi definitsiooni mõju GWAS uuringu tulemustele hinnati kahel moel. Esmalt võrreldi erinevate fenotüübidefinitsioonide GWAS-i tulemusi, et tuvastada nii kattuvad kui ka unikaalseid signaale. Analüüsist selgusid geneetilised variandid, mis esinesid mitme määratluse

puhul, samuti need, mis olid spetsiifilised vaid teatud fenotüübile (nt D06 või N87), viidates võimalikule seosele haiguse konkreetse staadiumi või alatüübiga. Sellised võrdlused aitavad selgitada, kas fenotüübi määratlus mõjutab seoseanalüüsi tundlikkust ja avastatud signaalide bioloogilist tähendust. Kitsama definitsiooni kasutamine võib tõsta fenotüübi spetsiifilisust, kuid vähendada statistilist võimsust. Samas laiem määratlus võib suurendada võimsust, kuid viia heterogeensema grupini. Tulemused aitavad hinnata, kuivõrd on geneetilised signaalid sõltuvad valitud fenotüübi täpsusest ning mida tuleks arvesse võtta sarnaste uuringute kujundamisel või replitseerimisel teistsugustest andmestikes.

Teiseks vaadeldi teadaolevate emakakaelavähi riskilookuste efektisuurusi erinevate fenotüübi määratluste lõikes, et hinnata, kas ja kuidas muutub nende mõju sõltuvalt kasutatud diagnoosi kriteeriumitest. See võimaldab kontrollida, kui stabiilsed on senised teadaolevad assotsiatsioonid erinevate juhtude määratluse korral ja kas mõni määratlus võiks anda süstemaatiliselt tugevama või nõrgema signaali. Varasemates uuringutes, mida on käsitletud peatükis 1.4, on emakakaelavähi puhul tuvastatud mitmeid tugevaid geneetilisi riskilookuseid, eriti HLA piirkonnas (*HLA-DQA1*, *HLA-DPB2*, *HLA-DRB5*) ning geenides nagu *PAX8*, *CLPTM1L*, *SKIV2L* ja *GSDMB* (näiteks Koel *et al*., 2023; Shi *et al*., 2013; Wang *et al*., 2015). Käesolevas töös analüüsiti viit geneetilist varianti (SNP-i), mis valiti artikli (Koel *et al*., 2023) tabel 2 alusel. Valitud SNP-id otsiti käesoleva töö raames analüüsitud fenotüüpide koondstatistikutest. Selleks kasutati lokaalset arvuti *Terminal* keskkonda ja käsurea tööriista *grep*, millega leiti iga variandi vastavad kirjed uutest failidest. Iga SNP-i kohta koguti andmed: efektisuurus, standardhälve, 95% usaldusvahemik ja p-väärtused. Efektisuuruse visualiseerimiseks koostati joonis, mis näitab nii efektisuuruse arvväärtust kui ka usaldusvahemikke, nn *forest plot* (vt Joonis 16), mis genereeriti R-tarkvara[5] (versioon 4.3.2) ja ggplot2 paketi (versioon 3.5.0) abil. Graafikul kujutatakse iga markeri (SNP) kohta erinevate fenotüübi definitsioonide puhul šansside suhet (*odds ratio*, OR) ja 95% usaldusvahemikku. X-teljel on esitatud OR väärtused, telje keskele on lisatud vertikaalne joon, väärtusel 1, nullefekti tähistamiseks. Graafik võimaldab visuaalselt võrrelda efektisuurusi ja nende kattuvust eri fenotüübi definitsioonide lõikes.

[5] https://www.r-project.org/

2.3 Tulemused

2.3.1 Seoseanalüüside tulemused ja kandidaatgeenide prioritiseerimine

Selgitamaks, millised geenid võivad olla seotud emakakaelavähi tekkemehhanismidega, visualiseeriti ülegenoomse seoseanalüüsi tulemused FUMA platvormi SNP2GENE funktsiooni abil. See võimaldab siduda statistiliselt olulisi SNP-e nende potentsiaalsete kandidaatgeenidega, võttes arvesse genoomset asukohta, regulatoorseid elemente ja bioloogilist konteksti. Esmalt kaardistati üksikvariantide assotsiatsioonid uuritavate diagnoosidega. Sellekäigus genereeris FUMA veebiplatvorm GWAS tulemustest kokkuvõtte, kus on välja toodud tunnustega seotud geneetilised riskilookused ning nendes asuvad olulised (p-väärtus < 5×10^{-8}) juhtivad geneetilised variandid. Variantide genoomne asukoht (CHR:POS) on välja toodud tabelis 3 ning need põhinevad FUMA platvormi integreeritud ANNOVAR-tarkvaral.

Antud töös tuvastati seitse emakakaelavähiga seotud lookust kuues kromosoomis. Mitmed markerid, mida varasemates uuringutes on seostatud emakakaelavähi riskiga, leidsid kinnitust ka antud uuringus. Leiti veel ka unikaalseid variante, mis ilmnesid vaid kindlate fenotüüpide korral ja võivad viidata uutele geneetilise eelsoodumuse piirkondadele. Ülevaade tuvastatud riskilookuste juht-SNP-idest on esitatud tabelis 3. Vastavad assotsiatsioonisignaalid on fenotüüpide lõikes kujutatud ka joonisel 8. Täpsemad tulemused on esitatud fenotüübi kaupa alapeatükkides 2.3.1.1 kuni 2.3.1.4.

Riskilookuste kõige tõenäolisemate kandidaatgeenide kaardistamiseks kasutati FUMA platvormi olemasolevaid võimalusi ja andmekihte. Arvesse võeti järgmised tegurid:

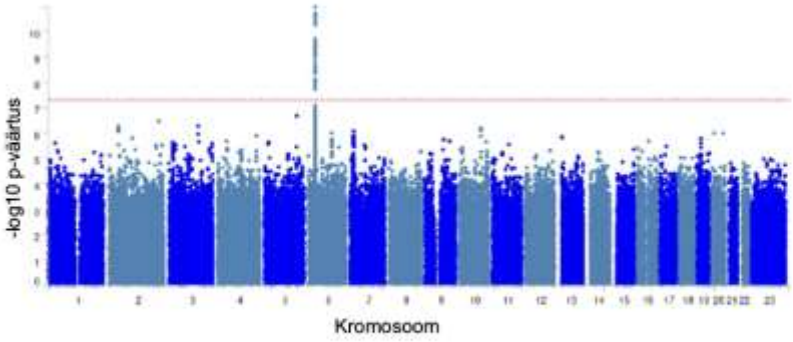
1. juhtivale SNP-ile lähimad geenid;
2. MAGMA geeni-põhise analüüsi tulemused;
3. riskilookuses olevate variantide mõju geeniekspressioonile (nn eQTL);
4. riskilookuses olevate variantide ülekate teadaolevate kromatiini interaktsioonides.

eQTL ja kromatiin interaktsioonide põhinevate geenikaardistamiste tulemused on FUMA platvormi abil kantud *circos* joonistele (vt Joonis 12, Lisa 2). Kõik 6. kromosoomil tuvastatud signaalid paiknevad *HLA* geeniklastri piirkonnas, milles on väga tihe ja keerukas LD (*linkage disequilibrium*) struktuur. LD tõttu on antud piirkonnas kandidaatgeenide ja variantide eristamine oluliselt raskendatud, sest järjestikused variandid võivad päranduda koos ja anda kattuvaid assotsiatsioonisignaale. See seab olulisi piiranguid kandidaatgeenide määramise usaldusväärsusele ja suurendab ka valepositiivsete kandidaatgeenide riski. Seetõttu antud töös piirduti kandidaatgeenide prioritiseerimisel kahe kriteeriumiga:

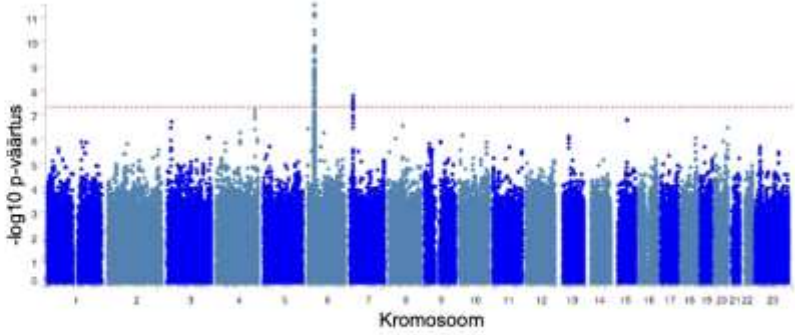
1. Keskenduti ainult nendele leidudele, mis esindavad uusi, varasemalt kirjeldamata seoseid.
2. Prioritiseeriti kandidaatgeene, mille signaalid asuvad väljaspool HLA piirkonda, et vältida LD-st tingitud määramatust ja võimaldada usaldusväärsemat geenide funktsionaalselt tõlgendust.

Tabel 3. Statistiliselt olulised tuvastatud SNP-id ja nendega seotud geenid. CHR:POS – kromosoom:positsioon; SNP-ID – SNP identifikaator; EA/MAF – efektialleel/harvema alleeli sagedus

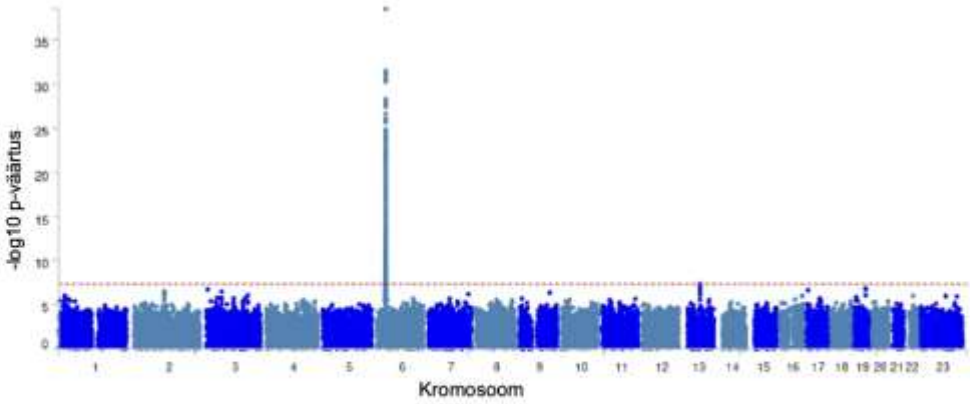
Fenotüüp	CHR:POS (GRChr37)	SNP-ID	EA/MAF	P-väärtus	Lähim geen
C53	6:31930462	rs429608	A/0.148	1.3×10^{-8}	*SKIV2L*
C53	6:32592313	rs35224566	C/0.096	1.1×10^{-11}	*HLA-DQA1*
C53_D06	6:32626432	rs28515515	T/0.291	3.2×10^{-12}	*HLA-DQB1*
C53_D06	7:16822054	rs7809828	A/0.042	1.7×10^{-8}	*AGR2*
D06	5:30948046	rs185789003	T/0.002	4.0×10^{-8}	*RPL19P11*
N87	13:65270181	rs559412963	A/0.001	1.4×10^{-8}	*LGMNP1*
N87	6:32503909	rs112401084	A/0.034	2.5×10^{-35}	*HLA-DRB5*
C53_D06_N87	6:32503909	rs112401084	A/0.034	3.3×10^{-39}	*HLA-DRB5*
C53_D06_N87.2	6:32503909	rs112401084	A/0.034	3.3×10^{-39}	*HLA-DRB5*



Joonis A.



Joonis B.



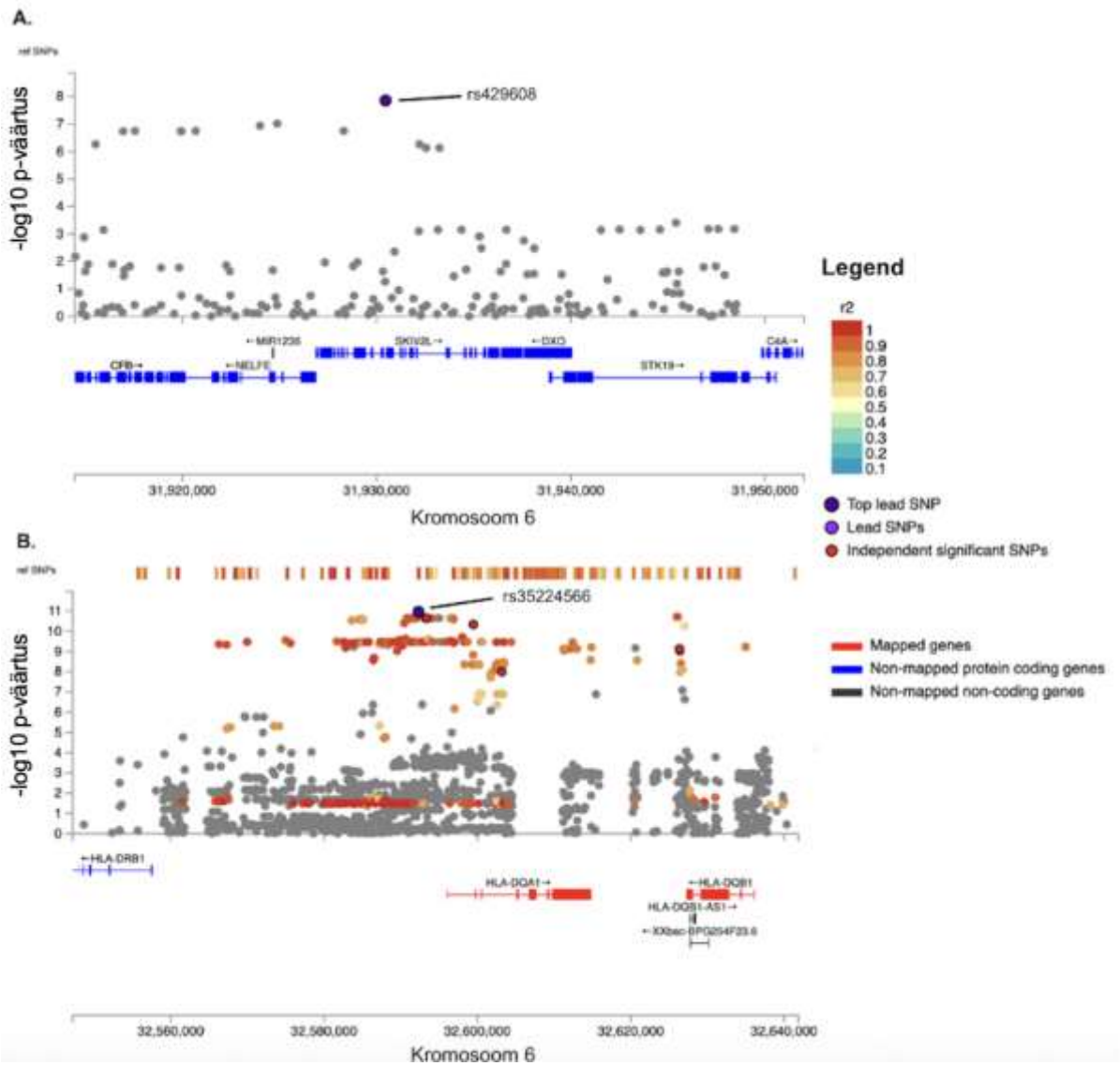
Joonis C.

Joonis 9. Manhattan'i joonised analüüsitud fenotüüpide seoseuuringute tulemustest. X-teljel on geneetiliste markerite skemaatilised asukohad genoomis. Y-teljel on vastava assotsiatsiooni p-väärtuse negatiivne kümnendlogaritm. Punase katkendjoonega on märgitud mitmesele testimisele kohandatud olulisuse piir (p-väärtus $\leq 5 \times 10^{-8}$). Joonis on tehtud FUMA veebiplatvormi abil. | **Joonisel A** on kujutatud emakakaela pahaloomulise kasvaja (C53) fenotüübi seoseuuringute tulemuste Manhattan'i joonis. | **Joonisel B** kombineeritud emakakaela pahaloomuline kasvaja ja emakakaela *in situ* kartsinoom (C53_D06) fenotüübi seoseuuringute tulemuste Manhattan'i joonis. | **Joonisel C** kombineeritud emakakaela pahaloomuline kasvaja, emakakaela *in situ* kartsinoom ja emakakaela düsplaasia (C53_D06_N87) fenotüübi seoseuuringute tulemuste Manhattan'i joonis.

2.3.1.1 Emakakaela pahaloomuline kasvaja

Fenotüübi C53 (CC) puhul kaardistati 2 riskilookust 6. kromosoomis, mis on visualiseeritud joonisel 10. Esimene lookuse juhtiv SNP on geneetiline variant rs429608 (p-väärtus 1.3×10^{-8}), mis paikneb *SKIV2L* geeni intronis. *SKIV2L* (*superkiller viralicidic activity 2-like*) kodeerib valku, mis on seotud mRNA kvaliteedikontrolli ja lagundamisega (Kögel *et al*., 2022). Antud geen on seotud ka viirusvastase immuunsusega, vahendades endogeensete immunostimuleerivate RNA molekulide lagundamist, mis tekivad raku stressivastuse käigus (Eckard *et al*., 2014). Prindle *et al*., (2025) uuringus leiti, et *SKIV2L* toimib negatiivse regulaatorina viirusvastases RIG-I-laadses (*retinoic acid-inducible gene I like*) retseptorite vahendatud immuunvastuses, piirates endogeense dsRNA kuhjumist ning vähendades kasvajate immunogeensust.

Teine lookuse juht-SNP on variant rs35224566 (p-väärtus 1.1×10^{-11}), mis paikneb HLA piirkonnas, geenidevahelises alas, umbes 3643 aluspaari kaugusel *HLA-DQA1* kodeerivast geenist. *HLA-DQA1* geen on polümorfne, omades mitmeid alleele, mis tõttu võib mõjutada vastuvõtlikkust erinevatele haigustele. Näiteks on varasemalt teada, et *HLA-DQA1* ja *HLA-DQB1* alleelide kombinatsioonid on seotud suurenenud riskiga tsöliaakia ja 1. tüüpi diabeedi tekkeks (Rajalingam *et al*., 2004). Samuti on *HLA-DQA1* geeniekspressioon seotud ka erinevate vähivormide, näiteks söögitoru lamerakulise kartsinoomi progresseerumisega (Megiorni & Pizzuti, 2012). Ramachandran *et al*., (2020) uuringus leiti, et *HLA-DQA1* transkriptsioonitase oli oluliselt kõrgem HPV-16 positiivsetes emakakaelaproovides võrreldes HPV-negatiivsetega (p-väärtus 0.019) ning samuti kõrgem ka HPV-18 positiivsetes proovides (p-väärtus 0.0005), kuigi viimase puhul valim oli väike. Sellised tulemused viitavad sellele, et antud geen on HPV-infektsiooniga seotud immuunvastuse regulatsioonis ning mõjutab seeläbi ka emakakaelavähi riski. Lisaks tuvastati samas uuringus, et *HLA-DQA1* geneetiline variant rs9272117 on seotud nii invasiivse emakakaelavähi kui ka raske astme düsplaasiaga, mis viitab geeni rollile haiguse geneetilises vastuvõtlikkuses. Antud SNP on visualiseeritud ka *circos* joonisel lisa 2, joonis 17.



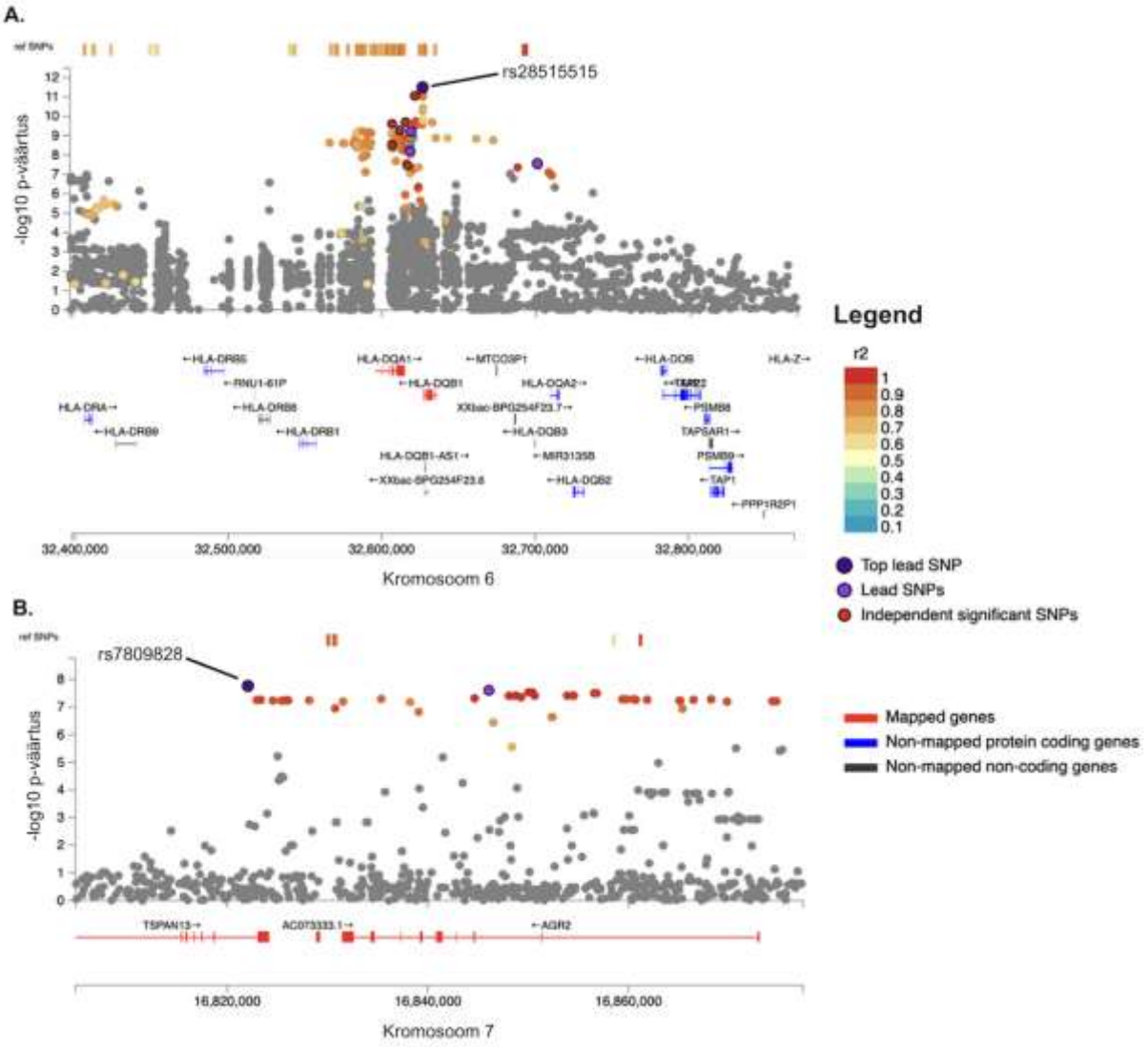
Joonis 10. *Regional plot* C53 fenotüübi puhul leitud juht-SNP-idest. X-teljel on kujutatud SNP positsiooni kromosoomil 6; Y-teljel SNP assotsiatsiooni p-väärtuse negatiivne kümnendlogaritm (-log^{10} p-väärtus). Värvikaardil on toodud LD tugevus (r^2) võrreldes lookuse madalaima p-väärtusega SNP-ga. Tumesinisega ringiga on näidatud lookuse madalaima p-väärtusega SNP. Geenipaneelil on punasega tähistatud kaardistatud geenid. Sinisega valke kodeerivad, kuid mitte kaardistatud. Halliga on toodud mitte-kodeerivad geenid. Joonise all on välja toodud piirkonnas olevate geenide asukohad ning noolega on tähistatud transkribeeritud ahel. Joonised on genereeritud FUMA veebiplatvormi abil. | **Joonisel A** on kujutatud *SKIV2L* geeni piirkonnas asuv tugevaim juhtvariant rs429608. | **Joonisel B** on kujutatud *HLA-DQA1* geeni piirkonnas asuv tugevaim juhtvariant rs35224566.

2.3.1.2 Emakakaela pahaloomuline kasvaja ja *in situ* kartsinoom

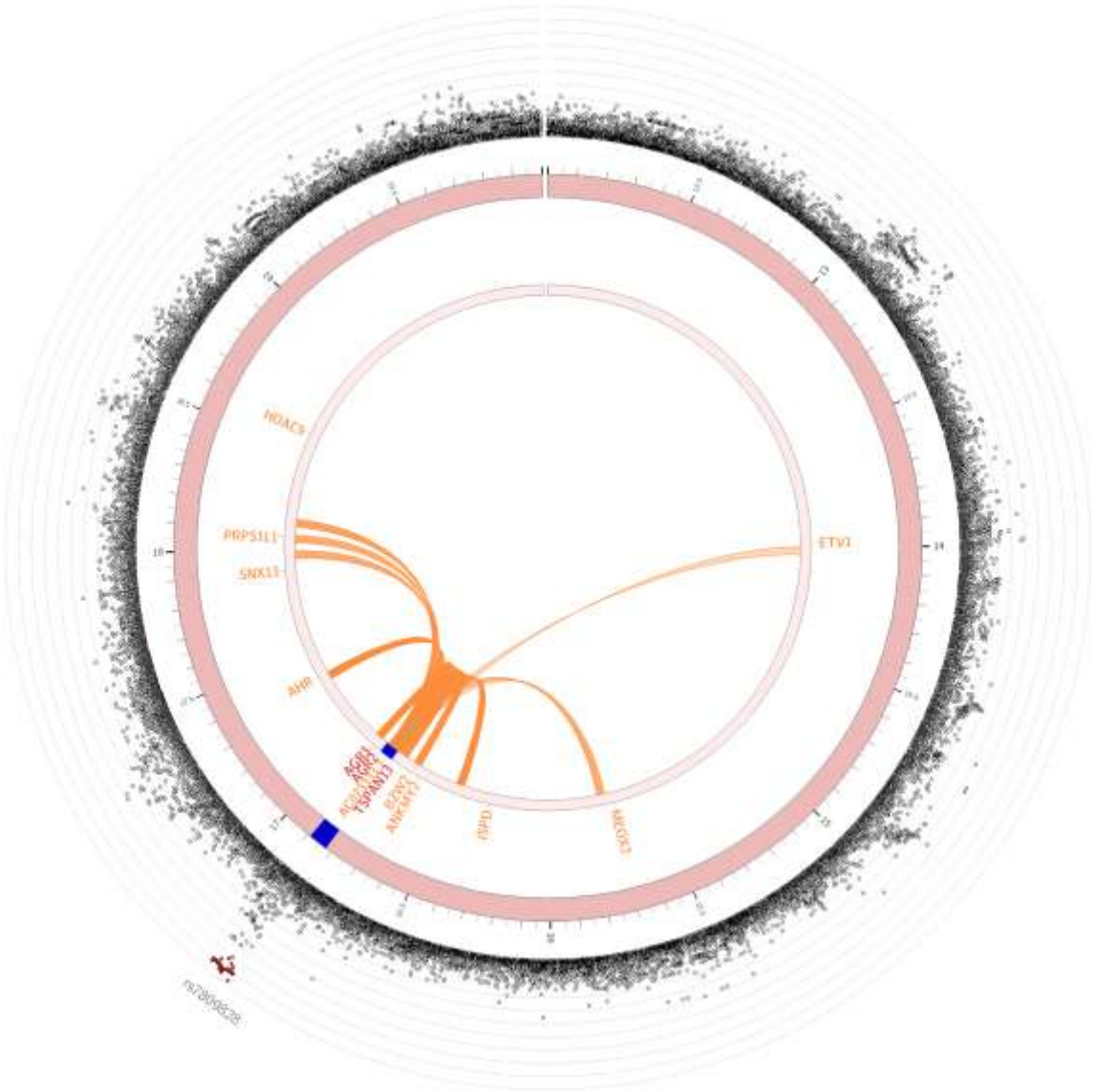
C53_D06 (CC+CIS) kombineeritud fenotüübiga tuvastati samuti 2 riskilookust, mis on välja toodud joonisel 11. Esimene tugevaim juhtiv SNP (ingl *top lead SNP*) rs28515515 (p-väärtus 3.2×10^{-12}) asub 6. kromosoomis, HLA piirkonnas, 9728 bp ülesvoolu *HLA-DQB1* kodeerivast geenist. See geen kuulub MH kompleksi (peamine koesobivuskompleks, ingl *major histocompability complex*) ning osaleb antigeeni esitlemisel T-rakkudele (Sundaramurthi et al., 2017). Selle piirkonna geneetilised variandid on seotud aga paljude autoimmuunsete ja infektsioonidega seotud haiguste, sealhulgas ka emakakaelavähi riskiga (Adebamowo et al., 2024; Gough & Simmonds, 2007; Zoodsma et al., 2005). *HLA-DQB1* teatud alleelid on seostatud kiirema haiguse progresseerumise või suurema nakatumise sagedusega (Sundaramurthi et al., 2017). Üldiselt on leitud, et HLA geenid võivad suurendada või vähendada haiguste riske ning emakakaelavähi puhul tõenäoliselt on see seotud organismi võimega tuvastada ja kõrvaldada HPV infektsioon. Selle konkreetse variandi täpsema mõju mõistmiseks on siiski vajalikud edasised funktsionaalsed uuringud. Antud SNP on visualiseeritud ka *circos* joonisel lisa 2, joonis 18.

Teine lookus asub 7. kromosoomis, tugevaim juhtiv SNP rs7809828 (p-väärtus 1.7×10^{-8}), paikneb *TSPAN13* geeni intronis. Analüüsis kaardistati võimalike kandidaatgeenidena eQTL ja kromatiini interaktsiooni andmete põhjal *TSPAN13*, *AGR2* ning *AGR3* (Joonis 12). rs7809828 SNP pole varasemalt otseselt GWAS uuringutes emakakaelavähiga seostatud, kuid *AGR2* võimalik roll kasvajate arengus on varasemalt teada (Brychtova *et al*., 2015). Seetõttu on kõige tõenäolisem kandidaatgeen selles lookuses *AGR2* (ingl *anterior gradient 2*). Antud geen kodeerib disulfiidisomeraaside perekonda kuuluvat valku ning osaleb valkude voltimises endoplasmaatilises retiikulumis (Salmans *et al*., 2013). Sellel geenil on epiteelibarjääri säilitamisel oluline roll ning valesti aktiveerituna soodustab see kasvajarakkude proliferatsiooni, invasiooni ja metastaseerumist (Brychtova *et al*., 2015). Üldiselt on leitud, et *AGR2* üleekspressioon on seotud rinnavähi ja teiste adenokartsinoomide progresseerumise ning halva prognoosiga (Salmans *et al*., 2013). Lisaks võib *AGR2* mõjutada kasvajarakkude vastupanuvõimet ravile, sealhulgas ka kiiritusravile (Brychtova *et al*., 2015). Variant rs7809828 ja selle seosed *AGR2* geeniga on visualiseeritud ka *circos* joonisel (vt Joonis 12), kus on näha lookuse kattuvus kromatiininteraktsiooni piirkonnaga ning võimalik mõju geeniekspressioonile. Erinevalt *AGR2* rollist vähi kontekstis on *AGR3* funktsiooni vähem uuritud ning olemasolevad andmed on vähesed ja vastuolulised. Xu et al., (2020) leidsid, et *AGR3* ekspressioon suureneb rinnavähi progresseerumisel ning on seotud metastaseerumise riskiga. Lisaks uuriti *AGR3* geeniekspressiooni seost vähi metastaseerumist levinumatesse

sihtorganitesse (luu, kops, maks, aju) rinnavähi patsientide seas ning leiti, et antud geeni ekspressioon korreleerus statistiliselt oluliselt luu- ja maksametastaasidega, samas kopsu- ja ajusiiretega seost ei leitud (Xu et al., 2020). *TSPAN13* on valku kodeeriv geen, mis kuulub tetraspaniinide hulka. Need valgud vahendavad signaaliülekandeid, mis reguleerivad rakkude arengut, aktivatsiooni, kasvu ja liikuvust. Enamik tetraspanini (TSPAN) perekonna valkudest soodustavad kasvajate teket, arengut ning ravivastust (Zhou et al., 2023).



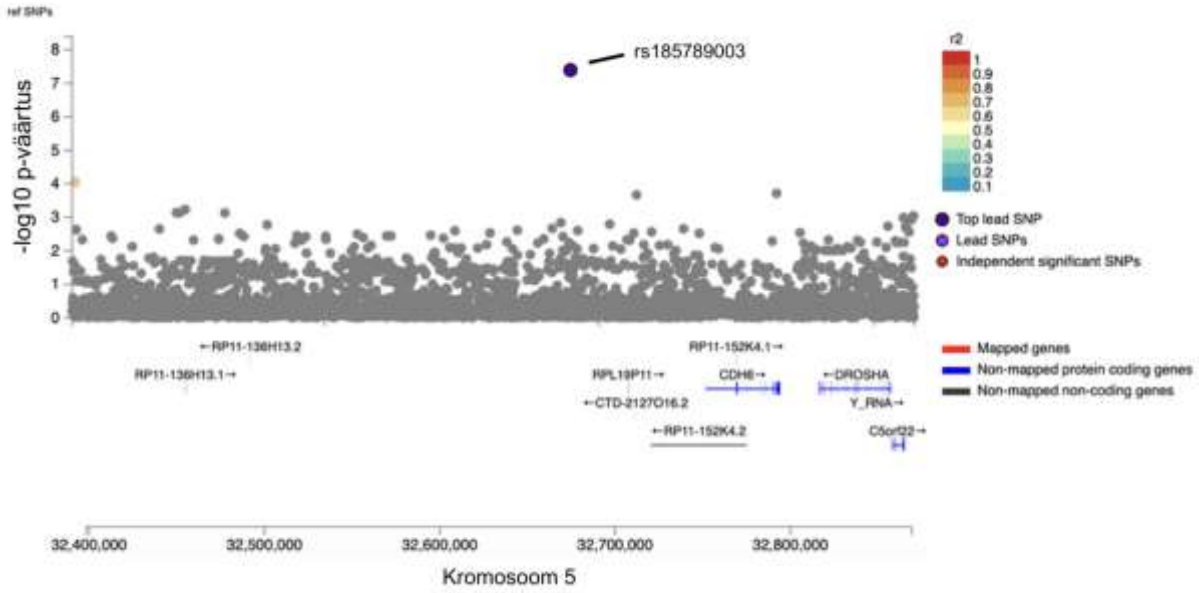
Joonis 11. *Regional plot* C53_D06 fenotüübi puhul leitud juht-SNP-idest. X-teljel on kujutatud SNP positsiooni kromosoomil 6; Y-teljel SNP assotsiatsiooni p-väärtuse negatiivne kümnendlogaritm (-$\log^{10}$ p-väärtus). Värvikaardil on toodud LD tugevus (r^2) võrreldes lookuse madalaima p-väärtusega SNP-ga. Tumesinisega ringiga on näidatud lookuse madalaima p-väärtusega SNP. Geenipaneelil on punasega tähistatud kaardistatud geenid. Sinisega valke kodeerivad, kuid mitte kaardistatud. Halliga on toodud mitte-kodeerivad geenid. Joonise all on välja toodud piirkonnas olevate geenide asukohad ning noolega on tähistatud transkribeeritud ahel. Joonised on genereeritud FUMA veebiplatvormi abil. | **Joonisel A** on kujutatud *HLA-DQB1* geeni piirkonnas asuv tugevaim juhtvariant rs28515515. | **Joonisel B** on kujutatud *TSPAN13* geeni piirkonnas asuv tugevaim juhtvariant rs7809828.



Joonis 12. Emakakaela vähi GWAS analüüsi tulemuste põhjal valmistatud C53_D06 fenotüübi 7. kromosoomi *circos* joonis. Välimisel kihil on Manhattani joonis. Joonisel on kujutatud geneetilised variandid, mille p-väärtus on ≤ 0.05. Y-teljel on vastava geneetilise variandi p-väärtuse negatiivne kümnendlogaritm. SNP-d on tähistatud vastavalt nende suurimale r^2 väärtusele ühega riskilookuses olevatest sõltumatutest olulistest SNP-dest: punane ($r^2 > 0.8$), oranž ($r^2 > 0.6$), roheline ($r^2 > 0.4$), sinine ($r^2 > 0.2$). SNP-id mis ei ole LD-s ühegi sõltumatu olulise SNP-ga on tähistatud halliga ($r^2 \leq 0.2$). Keskmisel kihil on kujutatud kromosoom koos koordinaatidega, sisemisele kihile – koordinaatideta ning sinise värviga on tähistatud riskilookus. Kandidaatgeenid on tähistatud vastavalt toetusandmete olemusele: oranž (kromatiini interaktsioonid) ja punane (eQTL ja kromatiini interaktsioon). Sisemise kihi kaared tähistavad eQTL-e (roheline) ja kromatiini interaktsioone (oranž). Joonis on tehtud FUMA veebiplatvormi abil.

2.3.1.3 Emakakaela *in situ* kartsinoom

D06 (CIS) fenotüübi puhul leiti tugevaimaks SNP-ks rs185789003 (p-väärtus 4.0×10^{-8}), mis paikneb 5. kromosoomis, geenidevahelises piirkonnas *RPL19* pseudogeeni *RPL19P11* läheduses. Pseudogeenid ei kodeeri valke, kuid *RPL19* geen kodeerib ribosoomi 60S alaühiku valku L19, mis osaleb teiste valkude sünteesis (Bee *et al*., 2006; Pink *et al*., 2011). L19 valgu üleekspressioon võib põhjustada valkude sünteesi häireid, mis soodustab kasvajate teket onkogeensete valkude üleproduktsiooni ja kasvajasupressorite pärsitud aktiivsuse kaudu (Bee *et al*., 2006). Lisaks on teada, et ribosoomi valkude geenid, sealhulgas *RPL19*, võivad olla seotud kasvajate arenguga (El Khoury & Nasr, 2021). *RPL19* üleekspressiooni on seostatud ka agressiivse fenotüübiga eesnäärmevähi puhul (Bee *et al*., 2006). Seetõttu on võimalik, et antud SNP võib mõjutada lähedal asuvate geenide regulatsiooni, mis omakorda võib mängida rolli emakakaelavähi arengus. Tegemist on väga harva esineva variandiga (MAF 0.002) ning lisaks oli antud fenotüübi puhul tegu väga väikese valimiga, mistõttu vajab seos täiendavat kinnitamist edasistes uuringutes. SNP rs185789003 on visualiseeritud joonisel 13.

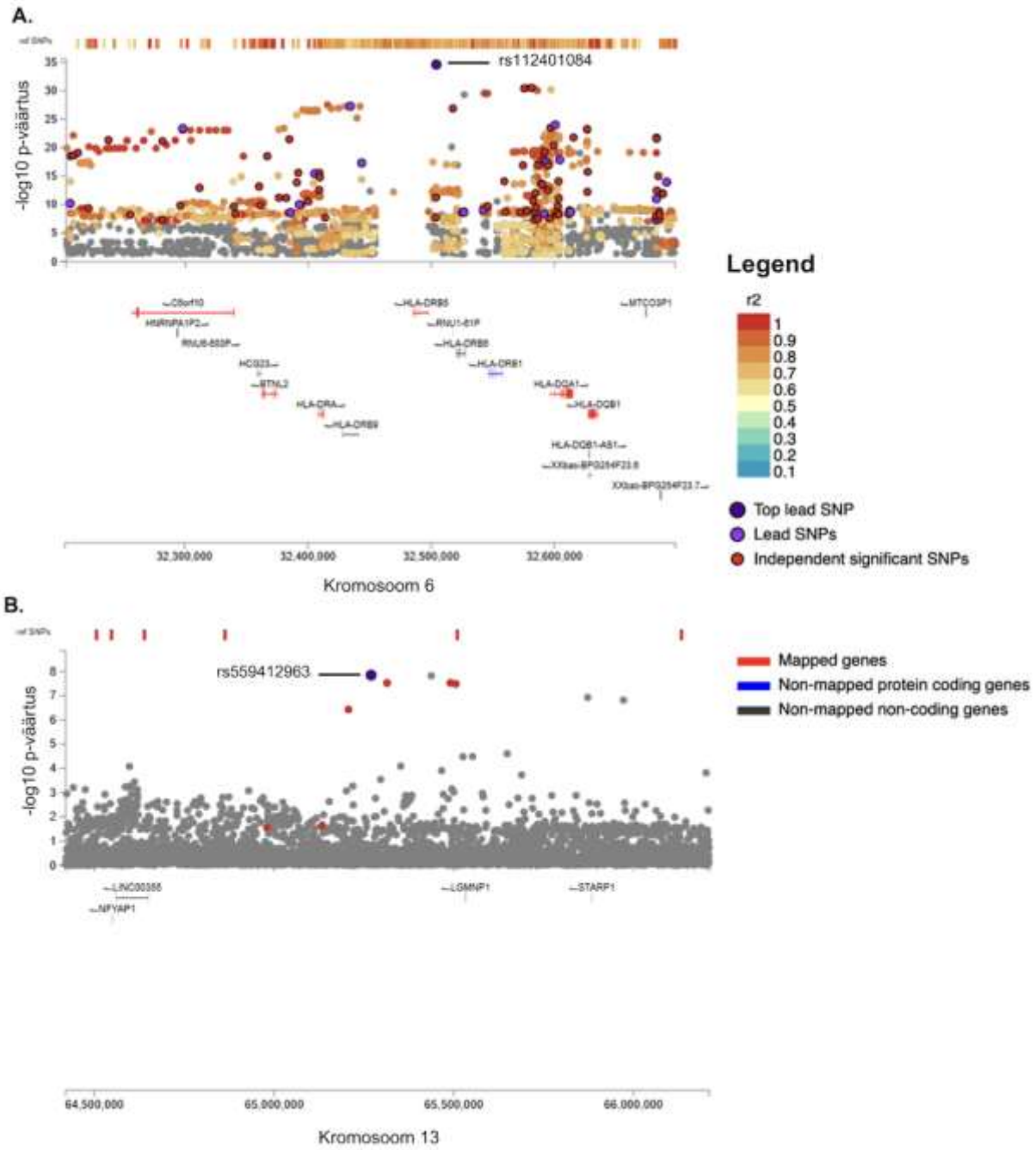


Joonis 13. *Regional plot* D06 fenotüübi puhul leitud juht-SNP. X-teljel on kujutatud SNP positsiooni kromosoomil 6; Y-teljel SNP assotsiatsiooni p-väärtuse negatiivne kümnendlogaritm ($-\log^{10}$ p-väärtus). Värvikaardil on toodud LD tugevus (r^2) võrreldes lookuse madalaima p-väärtusega SNP-ga. Tumesinisega ringiga on näidatud lookuse madalaima p-väärtusega SNP. Geenipaneelil on punasega tähistatud kaardistatud geenid. Sinisega valke kodeerivad, kuid mitte kaardistatud. Halliga on toodud mitte-kodeerivad geenid. Joonise all on välja toodud piirkonnas olevate geenide asukohad ning noolega on tähistatud transkribeeritud ahel. Joonised on genereeritud FUMA veebiplatvormi abil. | Joonisel on kujutatud *RPL19P11* pseudogeeni piirkonnas asuv tugevaim juhtvariant rs185789003.

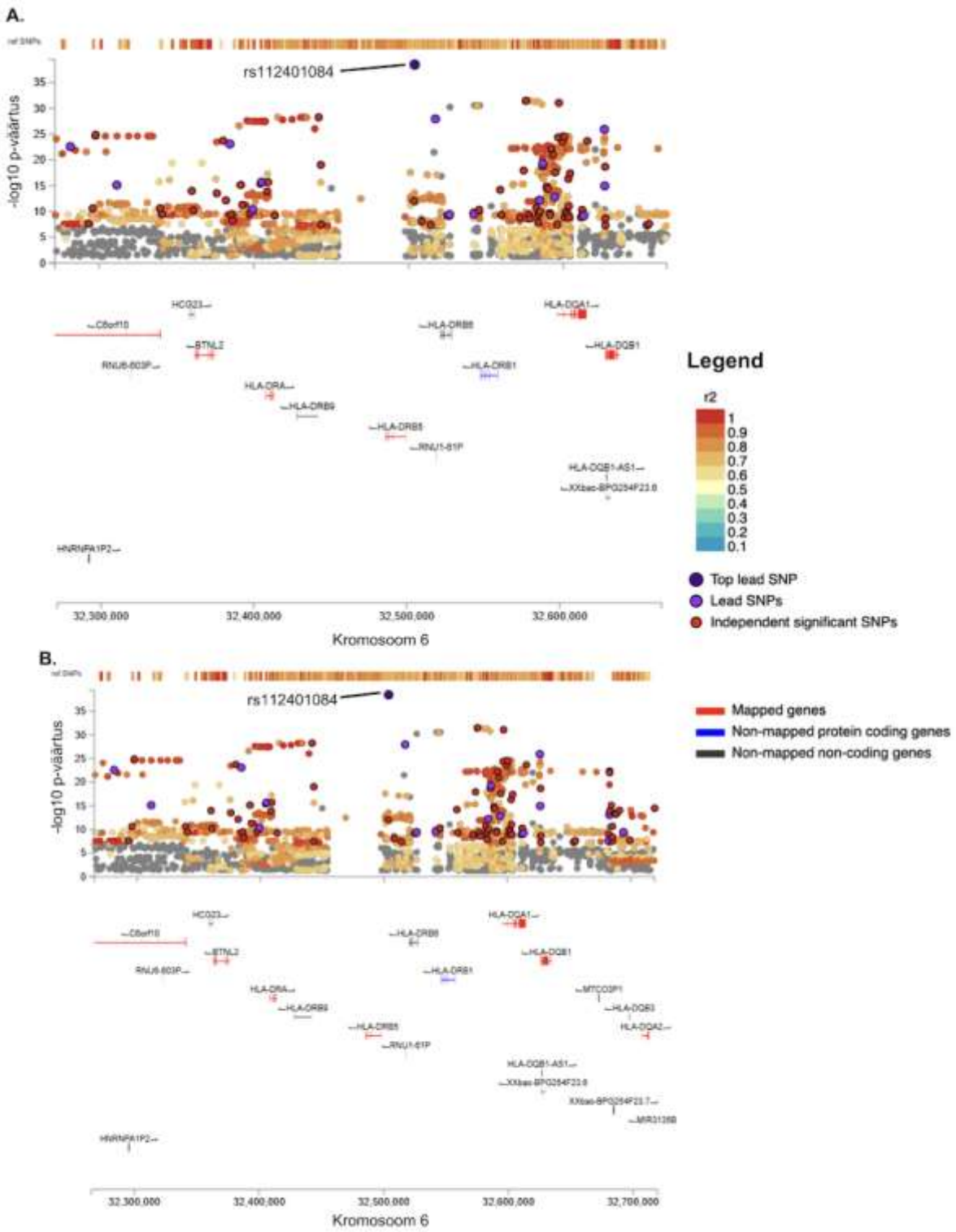
2.3.1.4 Düsplaasia ja kombineeritud fenotüübid

N87 (Dys) fenotüübiga leiti 2 statistiliselt oluliselt seotud lookust (Joonis 14), millest esimene tugevaim juhtiv SNP asub 6. kromosoomil HLA piirkonnas, geneetilise variandiga rs112401084 (p-väärtus 2.5×10^{-35}), mis asub geenidevahelises alas ning lähimaks geeniks *HLA-DRB5*. See geen on tugevas LD-s *HLA-DRB1* geeniga, millel on oluline roll antigeeni esitamisel ja immuunvastuse reguleerimisel (Odani *et al*., 2012). Sama variant (rs112401084) tuvastati ka kombineeritud fenotüüpide puhul – C53_D06_N87 (CC+CIS+Dys) ja C53_D06_N87.2 (CC+CIS+Dys3), kus täheldati veel tugevamat assotsiatsioonisignaali (p-väärtus 3.3×10^{-39}). Variant rs112401084 ja selle lookuse assotsiatsioonisignaalid on kujutatud ka joonistel 14 ja 15. Antud SNP on visualiseeritud ka *circos* joonisel lisa 2, joonis 19. Selline aga leid viitab stabiilsele ja fenotüübist sõltumatule geneetilisele eelsoodumusele HLA piirkonnas. Arvestades HLA piirkonna rolli immuunregulatsioonis ning HPV infektsiooni olulisust emakakaelavähi patogeneesis, võib variant rs112401084 mõjutada *HLA-DRB1* geeni ning seeläbi organismi võimet viirus elimineerida või selle püsima jäämist organismi, soodustades seeläbi vähieelseid muutuseid või kasvajate progressiooni.

Teine tugevaim leitud SNP rs559412963 (p-väärtus 1.4×10^{-8}) asub 13. kromosoomil, geenidevahelises piirkonnas ning lähim teadaolev geen on *LGMNP1*. Kuigi markerit rs559412963 ei ole varasemates emakakaelavähi või düsplaasia GWAS-analüüsides kirjeldatud, viitab selle asukoht võimalikule regulatiivsele mõjule, mis võib mõjutada läheduses paiknevate geenide transkriptsiooni. *LGMNP1* (*leucine rich glioma inactivated 1 neighboring pseudogene*) puhul on tegu pseudogeeniga, mida on varasemalt uuritud glioblastoomi rakuliinides, kus leiti, et antud geeni üleekspressioon vähendab DNA kahjustusi ja soodustab kiiritusravil tekkivat resistentsust (Xu et al., 2019). Vastus DNA kahjustusele ja apoptoosi regulatsioon on vähi seisukohalt väga olulised, seega võib tuvastatud seos olla funktsionaalselt oluline. Kuna tegemist on potentsiaalselt uue ja seni kirjeldamata seosega ning harvaesineva variandiga (MAF < 0.001), vajab see leid edasist funktsionaalset valideerimist ja replitseerimist.



Joonis 14. *Regional plot* N87 fenotüübi puhul leitud juht-SNP-idest. X-teljel on kujutatud SNP positsiooni kromosoomil 6; Y-teljel SNP assotsiatsiooni p-väärtuse negatiivne kümnendlogaritm (-$\log^{10}$ p-väärtus). Värvikaardil on toodud LD tugevus (r^2) võrreldes lookuse madalaima p-väärtusega SNP-ga. Tumesinisega ringiga on näidatud lookuse madalaima p-väärtusega SNP. Geenipaneelil on punasega tähistatud kaardistatud geenid. Sinisega valke kodeerivad, kuid mitte kaardistatud. Halliga on toodud mitte-kodeerivad geenid. Joonise all on välja toodud piirkonnas olevate geenide asukohad ning noolega on tähistatud transkribeeritud ahel. Joonised on genereeritud FUMA veebiplatvormi abil. | **Joonisel A** on kujutatud *HLA-DRB5* geeni piirkonnas asuv tugevaim juhtvariant rs112401084. | **Joonisel B** on kujutatud *LGMNP1* pseudogeeni piirkonnas asuv tugevaim juhtvariant rs559412963.



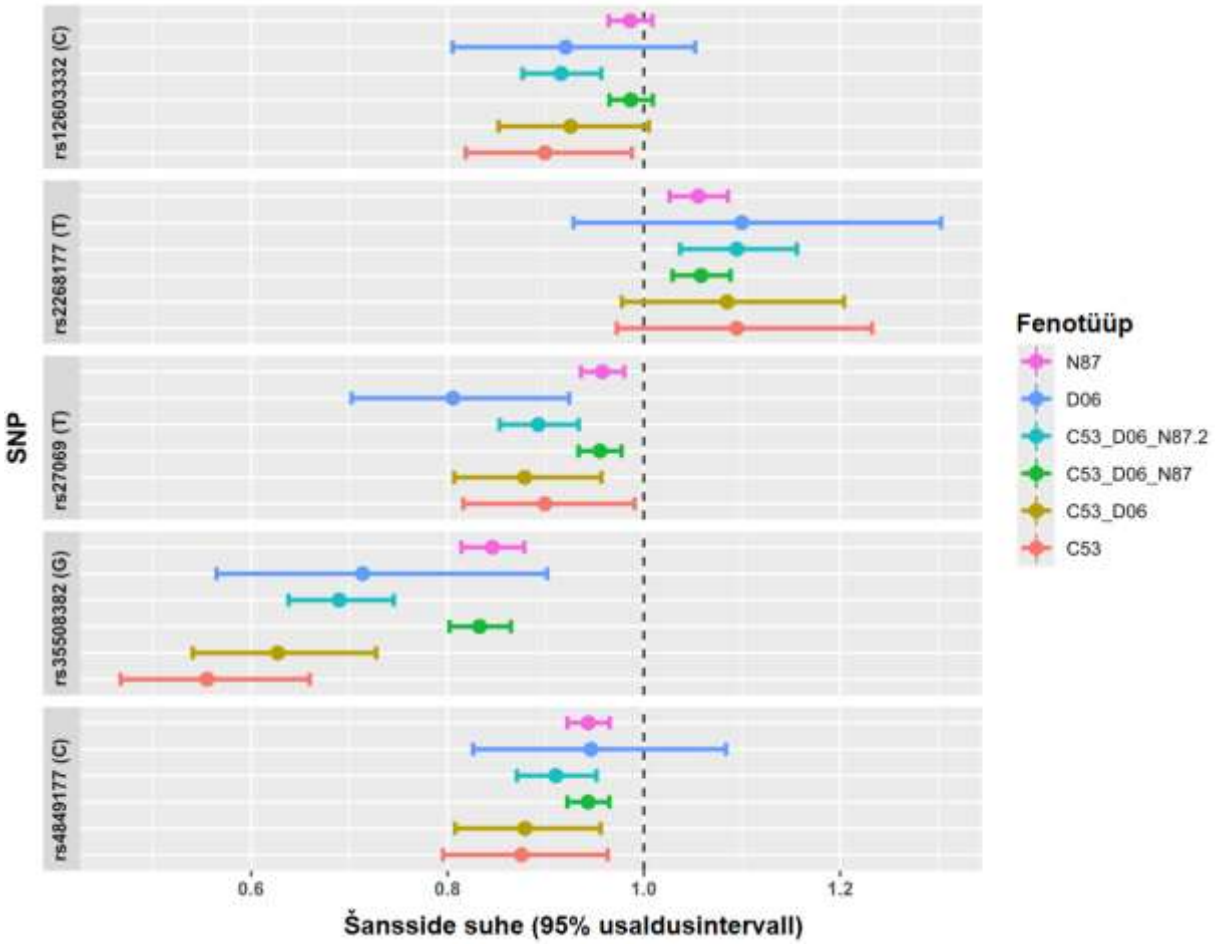
Joonis 15. *Regional plot* C53_D06_N87 ja C53_D06_N87.2 fenotüübi puhul leitud juht-SNP-idest. X-teljel on kujutatud SNP positsiooni kromosoomil 6; Y-teljel SNP assotsiatsiooni p-väärtuse negatiivne kümnendlogaritm (-$\log^{10}$ P-väärtus). Värvikaardil on toodud LD tugevus (r^2) võrreldes lookuse madalaima p-väärtusega SNP-ga. Tumesinisega ringiga on näidatud lookuse madalaima p-väärtusega SNP. Geenipaneelil on punasega tähistatud kaardistatud geenid. Sinisega valke kodeerivad, kuid mitte kaardistatud. Halliga on toodud mitte-kodeerivad geenid. Joonise all on välja toodud piirkonnas olevate geenide asukohad ning noolega on tähistatud transkribeeritud ahel. Joonised on genereeritud FUMA veebiplatvormi abil. | Nii **joonisel A** kui ka **joonisel B** on kujutatud *HLA-DRB5* geeni piirkonnas asuv tugevaim juhtvariant rs112401084.

2.3.2 Emakakaelavähi teadaolevate riskilookuste võrdlus

Võttes aluseks teadaolevad emakakaelavähi riskilookused (Koel *et al*., 2023), võrreldi nende efektisuurusi erinevate fenotüübi definitsioonide korral. Joonisel 16 on kujutatud viie varasemalt tuvastatud emakakaelavähi riskiga seotud SNP-ide šansside suhe koos 95% usaldusvahemikega, et hinnata nende mõju vastavalt erinevate fenotüübi määratluste korral. Usaldusvahemike kattumine, võib viidata selle, et erinevused ei pruugi olla statistiliselt olulised. Efektihinnangute varieeruvus, isegi kattuvate usaldusvahemike korral, võib osutada fenotüüpide heterogeensusele või erinevale bioloogilisele tähendusele.

Tulemustest on näha, et rs27069 (T) ja rs4849177 (C) puhul on tulemused suhteliselt sarnased ning efektisuuruste hinnangud valdavalt kattuvad. rs12603332 (C) puhul on aga näha, et C53_D06_N87.2 ning C53_D06_N87 fenotüüpide efektihinnangud ei kattu, mis viitab fenotüüpide vahelisele heterogeensusele. SNP-i mõju varieerub sõltuvalt kaasatud haigusjuhtudest ning tundub olevat tugevam kitsama ja raskema määratlusega fenotüübis (C53_D06_N87.2), mis viitab sellele, et rs12603332 avaldab mõju kliiniliselt raskemates või selgemates haigusjuhtudes. Erinevus võib tuleneda ka selles, et laiem fenotüüp – C53_D06_N87, hõlmab rohkem heterogeensemaid juhtumeid ja seega on ka ebatäpsemalt määratletud. Tulemustele hinnangu andmisel tuleks arvestada, et D06 fenotüübi valim oli ka kõige väiksem, mistõttu on selle usaldusintervallid oluliselt laiemad. Kui jätta fenotüüp D06 analüüsist välja, ilmneb, et mõnede SNP-ide puhul – rs12603332, rs27069 ja eriti rs35508382, on N87 ja C53_D06_N87 fenotüüpide korral täheldatavad väiksemad efektihinnangud. See viitab võimalusele, et SNP-i mõju avaldub tugevamalt täpsemate fenotüübi määratluse korral. Põhjuseks võiks olla asjaolu, et N87 fenotüüp hõlmab ka kergemaid düsplaasia juhtumeid, mis võivad viia efektihinnangut mõnevõrra madalamale. SNP rs35508382 G-alleli korral, on näha süstemaatiliselt üsna madalaid OR hinnanguidt, mis näitab, et see konkreetne alleel võib olla seotud tugeva kaitsega emakakaelavähi ja teiste sellega seotud seisundite vastu, olenemata fenotüübi definitsioonist.

Võib tähendada, et riskilookuste tulemused võivad erineda sõltuvalt kasutatud fenotüübi definitsioonidest. Fenotüübi määratluse muutmine võib mõjutada riskilookuste mõju suurust ja statistilist olulisust. Samas olid üldised tulemused erinevate fenotüübi määratluste puhul valdavalt üsna sarnased, mis viitab, et kasutatud definitsioonide varieeruvus ei avaldanud seoste tuvastamisele süstemaatilist mõju, ning erinevused sõltuvad konkreetsest lookusest.



Joonis 16. SNP-de ja fenotüüpide vahelised seosed. Kasutatud on šansside suhet (*odds ratio*, OR) koos 95%-lise usaldusintervalliga (*confidence interval*, CI). Iga punkt tähistab konkreetse SNP seost vastava fenotüübi definitsiooniga. Katkendlik must joon tähistab OR= 1.0, mis tähistab seose puudumist SNP ja uuritud fenotüübi vahel. Erinevate värvidega on tähistatud uuritud fenotüübid ning nende tähistus on välja toodud joonise parempoolses legendis.

2.4 Arutelu

Antud magistritöö eesmärgiks oli hinnata, kuidas erinevad fenotüüpide definitsioonid mõjutavad emakakaelavähi geneetiliste seoseanalüüside tulemusi, kasutades selleks Eesti Geenivaramu andmeid. Selleks kaardistati GWAS analüüsi tulemused, kirjeldati diagnoosidega seotud geneetilisi riskilookuseid ning pakuti välja kandidaatgeene. Uuringu käigus tuvastati mitmeid statistiliselt olulisi geneetilisi lookuseid, ka HLA piirkonnas, näiteks geenide *HLA-DQA1*, *HLA-DQB1* ja *HLA-DRB5* läheduses. Lisaks leiti seoseid *SKIV2L*, *AGR2* ja *RPL19P11* geenidega. Variant rs112401084 6. kromosoomil oli seotud mitme erineva fenotüübi definitsiooniga, millest võib järeldada selle laiemat rolli emakakaelavähi patogeneesi erinevates etappides.

Ilmnes ka unikaalseid leide, mis esinesid ainult spetsiifilise fenotüübi korral. Näiteks variant rs559412963, 13. kromosoomil *LGMNP1* geeni läheduses, oli statistiliselt oluline ainult N87 fenotüübi puhul ning variant rs185789003 ilmnes ainult D06 fenotüübiga. Nende SNP-ide puhul on tegu harvaesinevate variantidega ning nende funktsionaalset toimet ei ole varasemalt kirjeldatud emakakaelavähiga seotult, seetõttu vajavad edasist valideerimist. Ainult fenotüübi C53_D06 korral tuvastati seos *AGR2* geeni piirkonnas. Tuvastatud seos ei ilmnenud aga C53_D06_N87.2 kombineeritud fenotüübi puhul, kuigi sellel kombinatsioonil oli suurem valim. See viitab sellele, et *AGR2* seos võib olla spetsiifilisem haiguse raskematele staadiumitele. Arvestades geeni teadaolevat roll teiste kasvajate, sh seedetrakti ja rinnanäärmevähi arengus ning selle ekspressiooni seost emakakaelavähi progresseerumise ja halva prognoosiga (Brychtova et al., 2015; Salmans et al., 2013), oleks soovitatav edasistes töödes analüüsida põhjalikumalt selle geeni regulatsiooni- ja ekspressioonimustreid emakakaela koeproovides. Lisaks võiks hinnata selle geeni võimalikku kasutust varajase diagnostika biomarkerina või ka sihtmärkravis. Samas tuleks tulemusi tõlgendada ettevaatlikult, kuna teatud signaalide ilmnemine ainult kindlate fenotüüpide puhul võib tuleneda erinevusest juhtude arvus ja valimi suuruses.

HLA piirkond tuli esile mitme erineva fenotüübi puhul, viidates et sellel võib olla keskne roll emakakaela patogeneesis. Arvestades HPV infektsiooni olulisust haiguse tekkes, võivad HLA alleelid mõjutada organismi võimet viirusest vabaneda või ka selle püsima jäämist organismi, luues soodsad tingimused vähi arenguks. Selline korduvalt esile tulev seos HLA piirkonnas viitab sellele, et see on oluline riskihindamisel ning vajab edasist uurimist sõltumata haiguse staadiumist või täpsest fenotüübist. Töö tulemuste põhjal võib järeldada, et fenotüübi määratlus mõjutab GWAS-i tulemusi. Kuigi mitmed signaalid kordusid sõltumata definitsioonist, siis mõned seosed ilmnesid vaid teatud juhtude korral. See võib viidata, et fenotüüpide spetsiifilisus

ning kombineeritus võivad mõjutada seoseanalüüsi tundlikkust ja võimsust. Tulevastes uuringutes oleks oluline tasakaalustada nii valimi puhtus kui ka suurus: ainult invasiivsete vähijuhtude kasutamine võimaldab täpsemat fenotüübilist kirjeldust, kuid väiksem valim võib vähendada statistilist võimsust. Samas kui kombineeritud juhtude kaasamine suurendab valimi suurust, kuid võib sisse tuua heterogeensust. Leiud toetavad osaliselt Koel *et al*. (2023) järeldusi, kus autorid soovitasid haigusjuhtude C53, D06 ja N87 kombineerimist, et seoseanalüüsi võimsust tõsta, kuid rõhutati ka vajadust fenotüüpide täpse määratluse ja eristamise järele. Tulevaste uuringute kavandamisel tulekski kaaluda, millises ulatuses juhtude kombineerimine on otstarbekas ning millal oleks soovitatav keskenduda "kitsamatele" juhtudele, mis on ka homogeensemad. Näiteks võib soovitada lähenemist, kus esmased analüüsid tehakse võimalikult suure valimimahuga ning seejärel tehakse tundlikkusanalüüs väiksemates spetsiifilisema definitsiooniga gruppides. Kuigi N87 fenotüübiga uuringusse kaasatud isikute arv oli suurim, siis ei võimaldanud antud töö piisavat statistilist võimsust, et analüüsida kõiki N87 aladiagnoose (sh CIN astmed) eraldi. Edasistes uuringutes, kui andmestik on suurem kui antud töös, siis võiks nende aladiagnooside eristamine võimaldada kirjeldada geneetilisi seoseid kogu emakakaela patoloogia gradiendi ulatuses.

Antud uuringul on mitmeid piiranguid, mida tuleb tulemuste tõlgendamisel arvesse võtta. Tegu on Eesti Geenivaramu andmetel põhineva analüüsiga, mis piirab tulemuste üldistatavust teiste populatsioonide suhtes. Samuti väiksem juhtude arv kitsamate fenotüüpide korral, nagu näiteks C53, võis vähendada statistilist võimsust harva esinevate variantide tuvastamisel. Fenotüübid olid määratletud RHK-10 koodide alusel, mis sõltuvad kliinilise dokumenteerimise täpsusest ja ei pruugi alati kajastada patoloogiliselt kinnitatud diagnoose. Käesolevas töös võrreldi erinevate fenotüübi määratluste mõju GWAS analüüsi tulemustele, siis edasistes uuringutes võiks veelgi täpsustada fenotüübi määratlusi, kasutades näiteks Pap-testi leidude klassifikatsiooni (Bethesda süsteem) või andmestikke, kus on haigusjuhud histopatoloogiliselt kinnitatud. See aitaks eristada ja täpsustada, kas avastatud signaalid on seotud konkreetse haiguse staadiumiga.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli uurida, kuidas erinevad fenotüübi määratlused (RHK-10 koodide alusel) mõjutavad emakakaelavähiga seotud geneetilise seoseanalüüsi tulemusi. Uuritavateks fenotüüpideks olid emakakaelavähk (C53), emakakaela *in situ* kartsinoom (D06), emakakaela düsplaasia (düsplaasia N87; mujal täpsustamata raske düsplaasia N87.2) ning nende erinevad kombinatsioonid. Lisaks hinnati varasemalt tuvastatud riskivariantide seost erinevalt defineeritud fenotüüpidega.

GWAS analüüsi tulemused analüüsiti ja visualiseeriti FUMA veebiplatvormi abil ning võrreldi varasemalt tuvastatud riskivariantide efektisuurusi erinevate fenotüüpide lõikes. Analüüsi tulemused näitasid, et mitmed eelnevalt tuvastatud riskivariandid, eriti HLA piirkonnas, kinnitasid oma seost ka antud töös. Näiteks marker rs112401084 *HLA-DRB5* geeni läheduses andis tugeva signaali mitmete fenotüübi definitsioonide korral, viidates selle markeriga seotud geneetilisele eelsoodumusele emakakaelavähi tekkes.

Lisaks tuvastati mitu uut signaali, sealhulgas variandid geenide *SKIV2L*, *AGR2* ja *LGMNP1* piirkonnas, mille bioloogiline seos emakakaelavähi kujunemisega vajab edasist funktsionaalset valideerimist ja replitseerimist. Varasemalt tuvastatud riskivariantide võrdlus fenotüüpide lõikes näitas, et kuigi suuresti tulemused kattuvad, võivad efektisuurused varieeruda sõltuvalt kasutatud fenotüübi kriteeriumitest. Töö tulemused rõhutavad fenotüüpide täpse määratluse olulisust geneetilistes seoseuuringutes ning toetavad HLA piirkonna rolli emakakaelavähi geneetilises eelsoodumuses.

The impact of phenotype definition related to cervical cancer on the results of genetic association analysis

Maret Vilbaste

Summary

Cervical cancer is one of the most common malignancies among women, with human papillomavirus (HPV) being the primary risk factor. However, increasing evidence suggests that genetic predisposition also plays a role in the disease's development. Previous genome-wide association studies (GWAS) have identified several genetic loci associated with cervical cancer, particularly in the HLA region.

The aim of this thesis was to investigate how different phenotype definitions (based on ICD-10 codes) influence the results of genetic association analyses for cervical malignancies. The studied phenotypes were cervical cancer (C53), carcinoma *in situ* (D06), dysplasia (N87; N87.2) and their various combinations. Genotyping data from the Estonian Biobank were used and GWAS results were annotated and visualized using the FUMA web platform. In addition, previously described SNPs were selected based on literature. These variants were extracted from summary statistics, and their effect sizes were compared across various phenotype definitions.

The results showed that several associations previously described in the literature were also confirmed in our GWAS analyses, particularly the SNP rs112401084 near the *HLA-DRB5* gene. Additionally, several new potential signals were identified, warranting further functional validation and replication. The definition of phenotypes influenced the strength of the associations and the detection of new signals, highlighting the importance of precise phenotype classification in genetic studies. This work contributes to a better understanding of the genetic background of cervical cancer and supports the use of refined phenotypes in future association analyses.

Kasutatud kirjandus

Aavik, I., Padrik, L., Raud, T., Täär, K., & Veerus, P. (2021). *Emakakaela, tupe ja vulva vähieelsete muutuste diagnoosimine, jälgimine ja ravi. HPV-vastase vaktsineerimise soovitused. ENSi ravijuhend, versioon 4*. https://www.scribd.com/document/499943358/emakakaela-juhis-versioon-4-korrigeeritud-variant-ka-17

Adebamowo, S. N., Adeyemo, A., Adebayo, A., Achara, P., Alabi, B., Bakare, R. A., Famooto, A. O., Obende, K., Offiong, R., Olaniyan, O., Ologun, S., Rotimi, C., & Adebamowo, C. A. (2024). Genome, HLA and polygenic risk score analyses for prevalent and persistent cervical human papillomavirus (HPV) infections. *European Journal of Human Genetics*, *32*(6), 708–716. https://doi.org/10.1038/s41431-023-01521-7

Arroyo Mühr, L. S., Guerendiain, D., Cuschieri, K., & Sundström, K. (2021). Human Papillomavirus Detection by Whole-Genome Next-Generation Sequencing: Importance of Validation and Quality Assurance Procedures. *Viruses*, *13*(7), 1323. https://doi.org/10.3390/v13071323

Bee, A., Ke, Y., Forootan, S., Lin, K., Beesley, C., Forrest, S. E., & Foster, C. S. (2006). Ribosomal Protein L19 Is a Prognostic Marker for Human Prostate Cancer. *Clinical Cancer Research*, *12*(7), 2061–2065. https://doi.org/10.1158/1078-0432.ccr-05-2445

Browning, S. R., & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

Brychtova, V., Mohtar, A., Vojtesek, B., & Hupp, T. R. (2015). Mechanisms of anterior gradient-2 regulation and function in cancer. *Seminars in Cancer Biology*, *33*, 16–24. https://doi.org/10.1016/j.semcancer.2015.04.005

Cadman, L., Waller, J., Ashdown-Barr, L., & Szarewski, A. (2012). Barriers to cervical screening in women who have experienced sexual abuse: An exploratory study. *Journal of Family Planning and Reproductive Health Care*, *38*(4), 214–220. https://doi.org/10.1136/jfprhc-2012-100378

Causin, R. L., Freitas, A. J. A. de, Trovo Hidalgo Filho, C. M., Reis, R. dos, Reis, R. M., & Marques, M. M. C. (2021). A Systematic Review of MicroRNAs Involved in Cervical Cancer Progression. *Cells*, *10*(3), Article 3. https://doi.org/10.3390/cells10030668

Chaw, L., Lee, S. H. F., Ja'afar, N. I. H., Lim, E., & Sharbawi, R. (2022). Reasons for non-attendance to cervical cancer screening and acceptability of HPV self-sampling among Bruneian women: A cross-sectional study. *PLOS ONE*, *17*(3), e0262213. https://doi.org/10.1371/journal.pone.0262213

Chen, D., Juko-Pecirep, I., Hammer, J., Ivansson, E., Enroth, S., Gustavsson, I., Feuk, L., Magnusson, P. K. E., McKay, J. D., Wilander, E., & Gyllensten, U. (2013). Genome-wide Association Study of Susceptibility Loci for Cervical Cancer. *JNCI: Journal of the National Cancer Institute*, *105*(9), 624–633. https://doi.org/10.1093/jnci/djt051

Eckard, S. C., Rice, G. I., Fabre, A., Badens, C., Gray, E. E., Hartley, J. L., Crow, Y. J., & Stetson, D. B. (2014). The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nature immunology*, *15*(9), 839–845. https://doi.org/10.1038/ni.2948

El Khoury, W., & Nasr, Z. (2021). Deregulation of ribosomal proteins in human cancers. *Bioscience Reports*, *41*(12), BSR20211577. https://doi.org/10.1042/BSR20211577

Gough, S. C. L., & Simmonds, M. J. (2007). The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Current Genomics*, *8*(7), 453–465. https://doi.org/10.2174/138920207783591690

Gupta, S., Nagtode, N., Chandra, V., & Gomase, K. (2023). From Diagnosis to Treatment: Exploring the Latest Management Trends in Cervical Intraepithelial Neoplasia. *Cureus*. https://doi.org/10.7759/cureus.50291

Habiba, M., Heyn, R., Bianchi, P., Brosens, I., & Benagiano, G. (2021). The development of the human uterus: Morphogenesis to menarche. *Human Reproduction Update*, *27*(1), 1–26. https://doi.org/10.1093/humupd/dmaa036

Human Papillomavirus and Related Diseases Report: ESTONIA. (2023). HPV Information Centre. https://hpvcentre.net/statistics/reports/EST.pdf

Javanbakht, Z., Kamravamanesh, M., Rasulehvandi, R., Heidary, A., Haydari, M., & Kazeminia, M. (2023). Global Prevalence of Cervical Dysplasia: A Systematic Review and Meta-Analysis. *Indian Journal of Gynecologic Oncology*, *21*(3), 62. https://doi.org/10.1007/s40944-023-00741-5

Kitsios, G., & Zintzaras, E. (2009). Genome-Wide Association Studies: Hypothesis-“free” or “engaged”? *Translational research : the journal of laboratory and clinical medicine*, *154*(4), 161–164. https://doi.org/10.1016/j.trsl.2009.07.001

Koel, M., Võsa, U., Jõeloo, M., Läll, K., Gualdo, N. P., Laivuori, H., Lemmelä, S., Daly, M., Palta, P., Mägi, R., & Laisk, T. (2023). GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. *Human Molecular Genetics*, *32*(12), 2103–2116. https://doi.org/10.1093/hmg/ddad043

Kögel, A., Keidel, A., Bonneau, F., Schäfer, I. B., & Conti, E. (2022). The human SKI complex regulates channeling of ribosome-bound RNA to the exosome via an intrinsic gatekeeping mechanism. *Molecular Cell*, *82*(4), 756-769.e8. https://doi.org/10.1016/j.molcel.2022.01.009

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., & L Price, A. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, *48*(11), 1443–1448. https://doi.org/10.1038/ng.3679

Luria, L., & Cardoza-Favarato, G. (2025). Human Papillomavirus. *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK448132/

Mandal, A. (2009, november 22). *What is Cervical Cancer?* News-Medical. https://www.news-medical.net/health/What-is-Cervical-Cancer.aspx

Martinez, L., Johns, E., Dugdale, D. C., & Conaway, B. (2019, juuli 31). *Cervical dysplasia: MedlinePlus Medical Encyclopedia*. Bethesda (MD): National Library of Medicine (US). https://medlineplus.gov/ency/article/001491.htm

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O’Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell, E., & Marchini, J. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, *53*(7), 1097–1103. https://doi.org/10.1038/s41588-021-00870-7

Megiorni, F., & Pizzuti, A. (2012). HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: Practical implications of the HLA molecular typing. *Journal of Biomedical Science*, *19*(1), 88. https://doi.org/10.1186/1423-0127-19-88

Mehta, S., & Sachdeva, P. (Toim). (2017). *Colposcopy of Female Genital Tract*. Springer Singapore. https://doi.org/10.1007/978-981-10-1705-6

Mello, V., & Sundstrom, R. K. (2023). Cervical Intraepithelial Neoplasia. *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK544371/

Milani, L., Alver, M., Laur, S., Reisberg, S., Haller, T., Aasmets, O., Abner, E., Alavere, H., Allik, A., Annilo, T., Fischer, K., Hofmeister, R., Hudjashov, G., Jõeloo, M., Kals, M., Karo-Astover, L., Kasela, S., Kolde, A., Krebs, K., … Metspalu, A. (2025). The Estonian Biobank's journey from biobanking to personalized medicine. *Nature Communications*, *16*, 3270. https://doi.org/10.1038/s41467-025-58465-3

Mitt, M., Kals, M., Pärn, K., Gabriel, S. B., Lander, E. S., Palotie, A., Ripatti, S., Morris, A. P., Metspalu, A., Esko, T., Mägi, R., & Palta, P. (2017). Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *European Journal of Human Genetics*, *25*(7), 869–876. https://doi.org/10.1038/ejhg.2017.51

Moncada-Madrazo, M., & Rodríguez Valero, C. (2025). Embryology, Uterus. *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK547748/

Myers, K. M., Feltovich, H., Mazza, E., Vink, J., Bajka, M., Wapner, R. J., Hall, T. J., & House, M. (2015). The mechanical role of the cervix in pregnancy. *Journal of Biomechanics*, *48*(9), 1511–1523. https://doi.org/10.1016/j.jbiomech.2015.02.065

Ngoma, M., & Autier, P. (2019). Cancer prevention: Cervical cancer. *ecancermedicalscience*, *13*, 952. https://doi.org/10.3332/ecancer.2019.952

Nott, J. P., Bonney, E. A., Pickering, J. D., & Simpson, N. A. B. (2016). The structure and function of the cervix during pregnancy. *Translational Research in Anatomy*, *2*, 1–7. https://doi.org/10.1016/j.tria.2016.02.001

Odani, T., Yasuda, S., Ota, Y., Fujieda, Y., Kon, Y., Horita, T., Kawaguchi, Y., Atsumi, T., Yamanaka, H., & Koike, T. (2012). Up-regulated expression of HLA-DRB5 transcripts and high frequency of the HLA-DRB5*01:05 allele in scleroderma patients with interstitial lung disease. *Rheumatology*, *51*(10), 1765–1774. https://doi.org/10.1093/rheumatology/kes149

Oliveira, S. P., Montezuma, D., Moreira, A., Oliveira, D., Neto, P. C., Monteiro, A., Monteiro, J., Ribeiro, L., Gonçalves, S., Pinto, I. M., & Cardoso, J. S. (2023). A CAD system for automatic dysplasia grading on H&E cervical whole-slide images. *Scientific Reports*, *13*(1), Article 1. https://doi.org/10.1038/s41598-023-30497-z

Pilalis, E., Zisis, D., Andrinopoulou, C., Karamanidou, T., Antonara, M., Stavropoulos, T. G., & Chatziioannou, A. (2025). Genome-wide functional annotation of variants: A systematic review of state-of-the-art tools, techniques and resources. *Frontiers in Pharmacology*, *16*, 1474026. https://doi.org/10.3389/fphar.2025.1474026

Pink, R. C., Wicks, K., Caley, D. P., Punch, E. K., Jacobs, L., & Francisco Carter, D. R. (2011). Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA*, *17*(5), 792–798. https://doi.org/10.1261/rna.2658311

Pirog, E. C., Lloveras, B., Molijn, A., Tous, S., Guimerà, N., Alejo, M., Clavero, O., Klaustermeier, J., Jenkins, D., Quint, W. G., Xavier Bosch, F., Alemany, L., & de Sanjosé, S. (2014). HPV prevalence and genotypes in different histological subtypes of cervical adenocarcinoma, a worldwide analysis of 760 cases. *Modern Pathology*, *27*(12), 1559–1567. https://doi.org/10.1038/modpathol.2014.55

Prendiville, W., & Sankaranarayanan, R. (2017). Anatomy of the uterine cervix and the transformation zone. *Colposcopy and Treatment of Cervical Precancer*. International Agency for Research on Cancer. https://www.ncbi.nlm.nih.gov/books/NBK568392/

Prindle, V., Richardson, A. E., Sher, K. R., Kongpachith, S., Kentala, K., Petiwala, S., Cheng, D., Widomski, D., Le, P., Torrent, M., Chen, A., Walker, S., Palczewski, M. B., Mitra, D., Manaves, V., Shi, X., Lu, C., Sandoval, S., Dezso, Z., … Plotnik, J. P. (2025). Synthetic lethality of mRNA quality control complexes in cancer. *Nature*, *638*(8052), 1095–1103. https://doi.org/10.1038/s41586-024-08398-6

Ragupathy, K., & McMullen, W. (2025). Cervical cancer prevention – interventions to improve screening. *Scottish Medical Journal*, *70*(2), 25–26. https://doi.org/10.1177/00369330251332863

Rajalingam, R., Ge, P., & Reed, E. F. (2004). A sequencing-based typing method for HLA-DQA1 alleles. *Human Immunology*, *65*(4), 373–379. https://doi.org/10.1016/j.humimm.2004.01.012

Ramachandran, D., Dennis, J., Fachal, L., Schürmann, P., Bousset, K., Hülse, F., Mao, Q., Wang, Y., Jentschke, M., Böhmer, G., Strauß, H.-G., Hirchenhain, C., Schmidmayr, M., Müller, F., Runnebaum, I., Hein, A., Stübs, F., Koch, M., Ruebner, M., … Dörk, T. (2022). Genome-wide association study and functional follow-up identify 14q12 as a candidate risk locus for cervical cancer. *Human Molecular Genetics*, *31*(15), 2483–2497. https://doi.org/10.1093/hmg/ddac031

Ramachandran, D., Schürmann, P., Mao, Q., Wang, Y., Bretschneider, L.-M., Speith, L.-M., Hülse, F., Enßen, J., Bousset, K., Jentschke, M., Böhmer, G., Strauß, H.-G., Hirchenhain, C., Schmidmayr, M., Tarbiat, J., Runnebaum, I., Dürst, M., Hein, A., Koch, M., … Dörk, T. (2020). Association of genomic variants at the human leukocyte antigen locus with cervical cancer risk, HPV status and gene expression levels. *International Journal of Cancer*, *147*(9), 2458–2468. https://doi.org/10.1002/ijc.33171

Rigby, S., Greenley, R., Uuskula, A., Tisler, A., Suurna, M., Jesse, M., Bardou, M., & McKee, M. (2024). Expanding eligibility and improving quality of cervical cancer screening in Estonia: The 2021 reforms. *Health Policy*, *144*, 105077. https://doi.org/10.1016/j.healthpol.2024.105077

Ronco, G., Dillner, J., Elfström, K. M., Tunesi, S., Snijders, P. J. F., Arbyn, M., Kitchener, H., Segnan, N., Gilham, C., Giorgi-Rossi, P., Berkhof, J., Peto, J., & Meijer, C. J. L. M. (2014). Efficacy of HPV-based screening for prevention of invasive cervical cancer: Follow-up of four European randomised controlled trials. *The Lancet*, *383*(9916), 524–532. https://doi.org/10.1016/S0140-6736(13)62218-7

Salmans, M. L., Zhao, F., & Andersen, B. (2013). The estrogen-regulated anterior gradient 2 (AGR2) protein in breast cancer: A potential drug target and biomarker. *Breast Cancer Research*, *15*(2), 204. https://doi.org/10.1186/bcr3408

Sellors, J. W., & Sankaranarayanan, R. (2003). *Colposcopy and treatment of cervical intraepithelial neoplasia: A beginners' manual*. Intern. Agency for Research Cancer.

Shi, Y., Li, L., Hu, Z., Li, S., Wang, S., Liu, J., Wu, C., He, L., Zhou, J., Li, Z., Hu, T., Chen, Y., Jia, Y., Wang, S., Wu, L., Cheng, X., Yang, Z., Yang, R., Li, X., … Ma, D. (2013). A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nature Genetics*, *45*(8), 918–922. https://doi.org/10.1038/ng.2687

Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), 477–485. https://doi.org/10.1038/nrg2361

Small Jr, W., Bacon, M. A., Bajaj, A., Chuang, L. T., Fisher, B. J., Harkenrider, M. M., Jhingran, A., Kitchener, H. C., Mileshkin, L. R., Viswanathan, A. N., & Gaffney, D. K. (2017). Cervical cancer: A global health crisis. *Cancer*, *123*(13), 2404–2412. https://doi.org/10.1002/cncr.30667

Stanley, M. (2010). Pathology and epidemiology of HPV infection in females. *Gynecologic Oncology*, *117*(2, Supplement), S5–S10. https://doi.org/10.1016/j.ygyno.2010.01.024

Sundaramurthi, J. C., Ashokkumar, M., Swaminathan, S., & Hanna, L. E. (2017). HLA based selection of epitopes offers a potential window of opportunity for vaccine design against HIV. *Vaccine*, *35*(42), 5568–5575. https://doi.org/10.1016/j.vaccine.2017.08.070

Zeng, P., Zhao, Y., Qian, C., Zhang, L., Zhang, R., Gou, J., Liu, J., Liu, L., & Chen, F. (2015). Statistical analysis for genome-wide association study. *Journal of Biomedical Research*, *29*(4), 285–297. https://doi.org/10.7555/JBR.29.20140007

Zhou, Z., Yang, Z., Zhou, L., Yang, M., & He, S. (2023). The versatile roles of testrapanins in cancer from intracellular signaling to cell–cell communication: Cell membrane proteins without ligands. *Cell & Bioscience*, *13*(1), 59. https://doi.org/10.1186/s13578-023-00995-8

Zoodsma, M., Nolte, I., Schipper, M., Oosterom, E., van der Steege, G., de Vries, E. G. E., te Meerman, G. J., & van der Zee, A. G. J. (2005). Analysis of the entire HLA region in susceptibility for cervical cancer: A comprehensive study. *Journal of Medical Genetics*, *42*(8), e49. https://doi.org/10.1136/jmg.2005.031351

Takeuchi, F., Kukimoto, I., Li, Z., Li, S., Li, N., Hu, Z., Takahashi, A., Inoue, S., Yokoi, S., Chen, J., Hang, D., Kuroda, M., Matsuda, F., Mizuno, M., Mori, S., Wu, P., Tanaka, N., Matsuo, K., Kamatani, Y., … Shi, Y. (2019). Genome-wide association study of cervical cancer suggests a role for ARRDC3 gene in human papillomavirus infection. *Human Molecular Genetics*, *28*(2), 341–348. https://doi.org/10.1093/hmg/ddy390

Tewari, K. S. (2025). Cervical Cancer. *New England Journal of Medicine*, *392*(1), 56–71. https://doi.org/10.1056/NEJMra2404457

Tisler, A., Võrk, A., Tammemägi, M., Ojavee, S. E., Raag, M., Šavrova, A., Nygård, M., Nygård, J. F., Stankunas, M., Kivite-Urtane, A., & Uusküla, A. (2024). Nationwide study on development and validation of a risk prediction model for CIN3+ and cervical cancer in Estonia. *Scientific Reports*, *14*(1), 24589. https://doi.org/10.1038/s41598-024-75697-3

Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 59. https://doi.org/10.1038/s43586-021-00056-9

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vu, M., Yu, J., Awolude, O. A., & Chuang, L. (2018). Cervical cancer worldwide. *Current Problems in Cancer*, *42*(5), 457–465. https://doi.org/10.1016/j.currproblcancer.2018.06.003

Waghe, T., & Acharya, N. (2024). Advancements in the Management of Cervical Intraepithelial Neoplasia: A Comprehensive Review. *Cureus*. https://doi.org/10.7759/cureus.58645

Wang, S., Sun, H., Jia, Y., Tang, F., Zhou, H., Li, X., Zhou, J., Huang, K., Zhang, Q., Hu, T., Yang, R., Wang, C., Xi, L., Deng, D., Wang, H., Wang, S., Ma, D., & Li, S. (2015). Association of 42 SNPs with genetic risk for cervical cancer: An extensive meta-analysis. *BMC Medical Genetics*, *16*(1), 25. https://doi.org/10.1186/s12881-015-0168-z

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Wu, J., Jin, Q., Zhang, Y., Ji, Y., Li, J., Liu, X., Duan, H., Feng, Z., Liu, Y., Zhang, Y., Lyu, Z., Yang, L., & Huang, Y. (2025). Global burden of cervical cancer: Current estimates, temporal trend and future projections based on the GLOBOCAN 2022. *Journal of the National Cancer Center*. https://doi.org/10.1016/j.jncc.2024.11.006

Xu, H., Chen, B., Xing, J., Wei, Z., Liu, C., Qiu, Y., Lin, Y., & Ren, L. (2019). Upregulation of LGMNP1 confers radiotherapy resistance in glioblastoma. *Oncology Reports*, *41*(6), 3435–3443. https://doi.org/10.3892/or.2019.7128

Xu, Q., Shao, Y., Zhang, J., Zhang, H., Zhao, Y., Liu, X., Guo, Z., Chong, W., Gu, F., & Ma, Y. (2020). Anterior Gradient 3 Promotes Breast Cancer Development and Chemotherapy Response. *Cancer Research and Treatment : Official Journal of Korean Cancer Association*, *52*(1), 218–245. https://doi.org/10.4143/crt.2019.217

Kasutatud veebiaadressid

Almostadoctor meditsiiniline entsüklopeedia. *Cervical cancer and CIN*. https://almostadoctor.co.uk/encyclopedia/cervical-cancer-and-cin (15.01.2025)

FUMA veebiplatvorm: https://fuma.ctglab.nl/ (05.05.2025)

Illumina koduleht: https://www.illumina.com/ (23.01.2025)

MyPathologyReport.ca *High grade squamous intraepithelial lesion (HSIL) of the cervix*. https://www.mypathologyreport.ca/diagnosis-library/high-grade-squamous-intraepithelial-lesion-of-the-cervix/ (15.01.2025)

National Institutes of Health. *What Is Cervical Cancer?* https://www.cancer.gov/types/cervical (07.01.2025)

Rahvusvahelise haiguste klassifikatsiooni 10.väljaanne: https://rhk.sm.ee/ (23.10.2024)

R tarkvara: https://www.r-project.org/ (23.10.2024)

Tervise Arengu Instituudi tervisestatistika ja terviseuuringute andmebaas: https://statistika.tai.ee/ (07.04.2025)

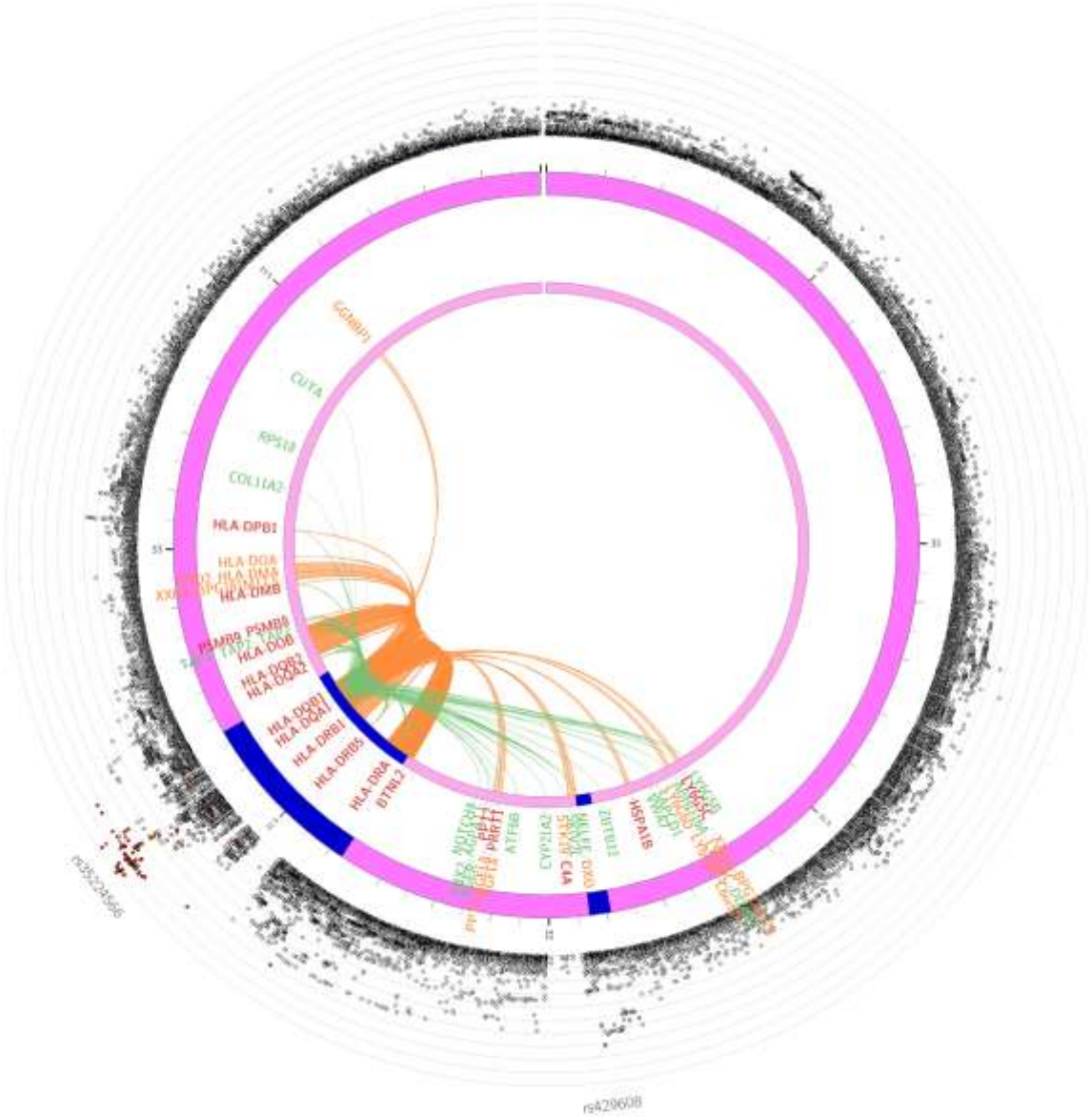
LISAD

Lisa 1.

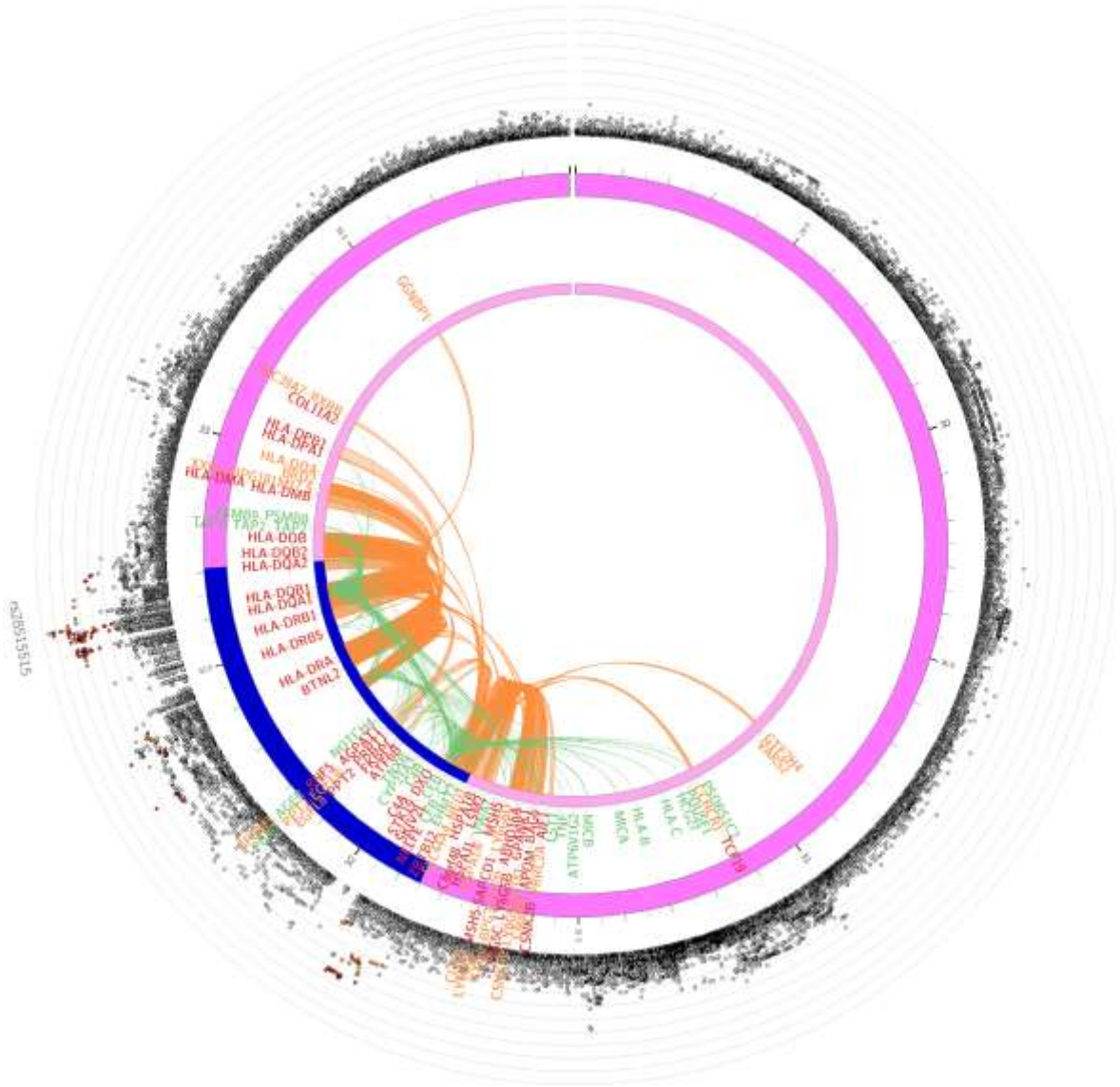
Tabel 4. RHK-10 koodid ning neile vastavad emakakaelaga seotud haigused

RHK-10 kood	**Nimetus**
C53	Emakakaela pahaloomuline kasvaja
D06	Emakakaela kartsinoom *in situ*
N87	Emakakaela düsplaasia
N87.2	Mujal klassifitseerimata raske emakakaela düsplaasia

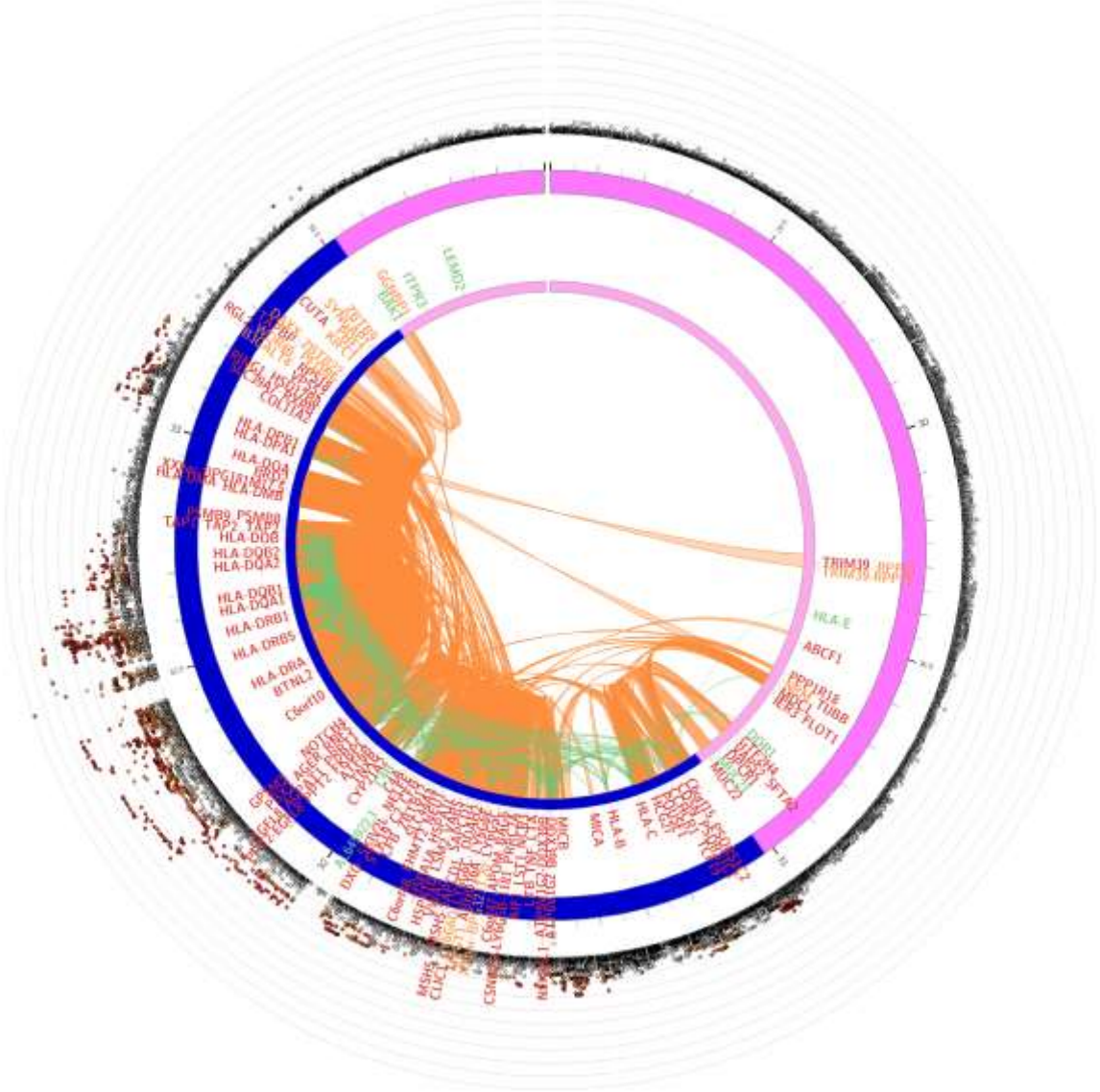
Lisa 2.



Joonis 17. Emakakaela vähi GWAS analüüsi tulemuste põhjal valmistatud C53 fenotüübi 6. kromosoomi *circos* joonis. Välimisel kihil on Manhattani joonis. Joonisel on kujutatud geneetilised variandid, mille p-väärtus on ≤ 0.05. Y-teljel on vastava geneetilise variandi p-väärtuse negatiivne kümnendlogaritm. SNP-d on tähistatud vastavalt nende suurimale r^2 väärtusele ühega riskilookuses olevatest sõltumatutest olulistest SNP-dest: punane ($r^2 > 0.8$), oranž ($r^2 > 0.6$), roheline ($r^2 > 0.4$), sinine ($r^2 > 0.2$). SNP-id mis ei ole LD-s ühegi sõltumatu olulise SNP-ga on tähistatud halliga ($r^2 \leq 0.2$). Keskmisel kihil on kujutatud kromosoom koos koordinaatidega, sisemisele kihile – koordinaatideta ning sinise värviga on tähistatud riskilookus. Kandidaatgeenid on tähistatud vastavalt toetusandmete olemusele: roheline (eQTL), oranž (kromatiini interaktsioonid) ja punane (eQTL ja kromatiin interaktsioon). Sisemise kihi kaared tähistavad eQTL-e (roheline) ja kromatiin interaktsioone (oranž). Joonis on tehtud FUMA veebiplatvormi abil.



Joonis 18. Emakakaela vähi GWAS analüüsi tulemuste põhjal valmistatud C53_D06 fenotüübi 6. kromosoomi *circos* joonis. Välimisel kihil on Manhattani joonis. Joonisel on kujutatud geneetilised variandid, mille p-väärtus on ≤ 0.05. Y-teljel on vastava geneetilise variandi p-väärtuse negatiivne kümnendlogaritm. SNP-d on tähistatud vastavalt nende suurimale r^2 väärtusele ühega riskilookuses olevatest sõltumatutest olulistest SNP-dest: punane ($r^2 > 0.8$), oranž ($r^2 > 0.6$), roheline ($r^2 > 0.4$), sinine ($r^2 > 0.2$). SNP-id mis ei ole LD-s ühegi sõltumatu olulise SNP-ga on tähistatud halliga ($r^2 \leq 0.2$). Keskmisel kihil on kujutatud kromosoom koos koordinaatidega, sisemisele kihile – koordinaatideta ning sinise värviga on tähistatud riskilookus. Kandidaatgeenid on tähistatud vastavalt toetusandmete olemusele: roheline (eQTL), oranž (kromatiini interaktsioonid) ja punane (eQTL ja kromatiin interaktsioon). Sisemise kihi kaared tähistavad eQTL-e (roheline) ja kromatiin interaktsioone (oranž). Joonis on tehtud FUMA veebiplatvormi abil.



Joonis 19. Emakakaela vähi GWAS analüüsi tulemuste põhjal valmistatud C53_D06_N87 fenotüübi 6. kromosoomi *circos* joonis. Välimisel kihil on Manhattani joonis. Joonisel on kujutatud geneetilised variandid, mille p-väärtus on ≤ 0.05. Y-teljel on vastava geneetilise variandi p-väärtuse negatiivne kümnendlogaritm. SNP-d on tähistatud vastavalt nende suurimale r^2 väärtusele ühega riskilookuses olevatest sõltumatutest olulistest SNP-dest: punane ($r^2 > 0.8$), oranž ($r^2 > 0.6$), roheline ($r^2 > 0.4$), sinine ($r^2 > 0.2$). SNP-id mis ei ole LD-s ühegi sõltumatu olulise SNP-ga on tähistatud halliga ($r^2 \leq 0.2$). Keskmisel kihil on kujutatud kromosoom koos koordinaatidega, sisemisele kihile – koordinaatideta ning sinise värviga on tähistatud riskilookus. Kandidaatgeenid on tähistatud vastavalt toetusandmete olemusele: roheline (eQTL), oranž (kromatiini interaktsioonid) ja punane (eQTL ja kromatiin interaktsioon). Sisemise kihi kaared tähistavad eQTL-e (roheline) ja kromatiin interaktsioone (oranž). Joonis on tehtud FUMA veebiplatvormi abil.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Maret Vilbaste,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

   "**Emakakaelavähiga seotud fenotüüpide definitsiooni mõju geneetilise seoseanalüüsi tulemustele**",

   mille juhendajad on kaasprofessor Triin Laisk, teadur Mariann Võsa ja nooremteadur Jelisaveta Džigurski,

   reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Maret Vilbaste
28.05.2025