TARTU ÜLIKOOL

ARVUTITEADUSE INSTITUUT

INFORMAATIKA ÕPPEKAVA

Promootorite kasutust mõjutavate geneetiliste variantide leidmine CAGE tehnoloogia abil

Bakalaureusetöö (9 EAP)

Andreas Vija

Juhendaja: Kaur Alasoo

TARTU 2020

Infoleht

Promootorite kasutust mõjutavate geneetiliste variantide leidmine CAGE tehnoloogia abil

Lühikokkuvõte:

Geenide avaldumist mõjutavad palju geeni alguse lähedal asuvad DNA lõigud, mida kutsutakse promootoriteks. Ühel geenil võib olla mitu erinevat promootorit, millest võivad saada alguse erinevad transkriptid. Promootorite kasutuse erinevused inimeste vahel on seotud mitmete haiguste tekkeriskidega. Levinuim viis transkriptide mõõtmiseks on RNA sekveneerimine (RNA-seq), aga selle meetodi signaal on transkriptide alguses nõrk. CAGE (*cap analysis of gene expression*) on meetod, mis keskendub transkriptide algustele ja suudab seega promootorite kasutust paremini määrata. Töö eesmärgiks oli uurida, kas CAGE tehnoloogia abil on võimalik paremini määrata geneetiliste variantide mõju promootorite kasutusele kui palju levinuma RNA-seq meetodi abil. Töö käigus leiti, et CAGE on selle eesmärgi jaoks tõenäoliselt veidi parem. Veel koostati CAGE-i abil loodud promootorite annotatsioonide põhjal täiendavaid RNA-seq transkriptide annotatsioone. Lisatud annotatsioonide abil paranes RNA-seq võime üles leida promootorite kasutust mõjutavaid geneetilisi variante, aga annotatsioonide loomise algoritm vajab veel täiendamist, sest põhjustab ka väärpositiivseid tulemusi.

Võtmesõnad: Promootorid, transkriptoomika, CAGE, eQTL analüüs

CERCS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

Using CAGE technology to find genetic variants affecting promoter usage

Abstract:

Gene expression is heavily influenced by promoters, which are special DNA regions near the starts of genes. Differential promoter usage has been associated with multiple complex diseases. Among other things, promoters dictate the transcripts that are created based on DNA. Although RNA sequencing (RNA-seq) is commonly used to measure transcription, its signal is relatively weak at the start of transcripts. Cap analysis of gene expression (CAGE) is a method that focuses on the beginnings of transcripts and can thus better measure promoter

usage. The goal of this work was to investigate whether CAGE technology is superior to RNA-seq in detecting genetic variants that influence promoter usage. It was found that CAGE is likely somewhat better for this task. Additional RNA–seq transcript annotations were also created based on promoter annotations made using CAGE. These new annotations improved the ability of RNA-seq to detect genetic effects on promoter usage, but the annotations need to be still revised to reduce the number of false positives.

Keywords: Promoters, transcriptomics, CAGE, eQTL analysis

CERCS: B110 Bioinformatics, medical informatics, biomathematics, biometrics

Sisukord

1 Sissejuhatus

Kõikide organismide ülesehituse ja talitluse määrab nende pärilik informatsioon, mis paikneb rakkudes DNA sees [1]. DNA on lineaarne molekul, mis koosneb diskreetsetest ühikutest nukleotiididest [1,2]. Ühe liigi erinevatel isenditel ei ole identne DNA, näiteks kahe inimese DNA erineb umbes 0,1% ulatuses [2,3]. Erinevuste piirkondi nimetatakse geneetilisteks variantideks ning nagu näha inimkonna mitmekesisusest, on nende mõju suur [3,4].

DNA sees paiknevad piirkonnad, mida nimetatakse geenideks ja mille põhjal pannakse kokku organismile hädavajalikke molekule valke [1,2]. Üks osa valkude loomise protsessist on transkriptsioon, mille käigus sünteesitakse DNA lõigu järgi sellele vastav sarnase ehitusega RNA molekuli lõik transkript [1,2]. Transkript määrab toodetava valgu [1,2]. Ühe geeni põhjal on sageli võimalik luua erinevaid transkripte erinevates vahekordades [1,2].

Transkriptsiooni kontrollivad promootorid – kindlad DNA järjestused, mis paiknevad transkriptsiooni alguskohtade lähedal [1,2]. Promootoritel on organismidele suur mõju, sest need määravad muuhulgas transkripti ning kui kiiresti valku toodetakse [1,2,5,6]. Inimeste-vahelisi erinevusi promootorite kasutuses (näiteks promootorite osakaaludes valkude sünteesil) on seostatud ka mitme haiguse tekkeriskiga [7,8]. Seega on promootorite kasutus oluline uurimisobjekt.

Levinuim viis transkriptide mõõtmiseks on RNA sekveneerimine ehk RNA-seq [9,10]. RNA-seq ei ole aga võimeline suure täpsusega määrama, millisest nukleotiidist transkriptsioon algab [11]. CAGE (*cap analysis of gene expression*) on meetod, mis keskendub spetsiifiliselt transkriptsiooni alguskohtade määramisele [12,13]. Seega on see potentsiaalselt võimas tööriist promootorite kasutuse määramiseks.

Töö eesmärgiks on uurida, kas CAGE tehnoloogia abil on võimalik promootorite kasutust mõjutavaid geneetilisi variante leida paremini kui palju levinuma RNA-seq meetodi abil. See teadmine aitaks otsustada, kas geneetiliste variantide transkriptsioonilise mõju paremaks tuvastamiseks oleks vaja erinevate rakutüüpide laiaulatuslikku sekveneerimist CAGE-i abil või piisab ka põhjalikumast RNA-seq andmete analüüsist.

Töö käigus leitakse 154 Kesk-Euroopa päritolu inimese CAGE andmete põhjal hinnang inimese geenide arvule, mille promootorite kasutust mõjutavad geneetilised variandid. Tulemust

võrreldakse Garieri jt artikliga [8], kus leiti sarnane hinnang sama CAGE andmestiku abil ning Kerimovi jt artikliga [14], mille käigus leiti sama hinnang RNA-seq abil. Seejärel kasutatakse CAGE andmete põhjal koostatud FANTOM5 promootorite andmestikku [15] ja täiendatakse Kerimovi jt protsessi transkriptide annotatsioone. Selle eesmärk on uurida, kas CAGE andmeid saab kasutada ka olemasolevate RNA-seq andmetega tehtud analüüside täiendamiseks.

Peatükis „Kirjanduse ülevaade“ selgitatakse geeniekspressiooni ning promootorite olulisust selles, võrreldakse tavapärast RNA-seq meetodit CAGE-iga, kirjeldatakse kasutusele võetavaid uurimismeetodeid ning antakse ülevaade võrdluseks võetavatest teadusartiklitest. Peatükis „Mõjusate variantide leidmine“ leitakse hinnang selliste geenide arvule, mille promootorite kasutust mõjutavad geneetilised variandid ning peatükis „Uued transkriptide annotatsioonid“ koostatakse kunstlikud transkriptide annotatsioonid ja uuritakse nende mõju Kerimovi jt [14] protsessile.

Töö viidi osaliselt läbi Tartu Ülikooli teadusarvutuste keskuses. Töö käigus kirjutatud R kood, Linux shell skriptid, Snakemake töövood ja mõned algandmed on saadaval GitHubi repositooriumis andreasvija/baka-kood [16].

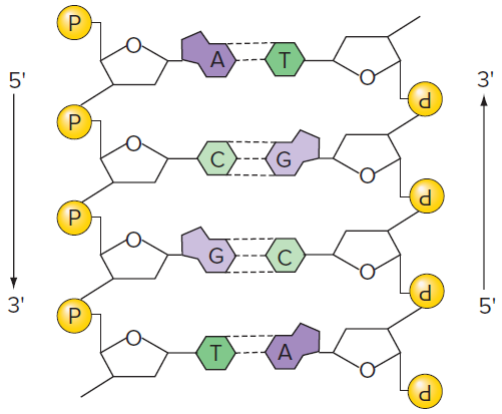
2 Kirjanduse ülevaade

2.1 Geenieksressioon

Kõik organismid – sealhulgas bakterid, taimed ja inimesed – sisaldavad informatsiooni, mis kirjeldab nende ülesehitust ja talitlust [1]. Seda informatsiooni salvestatakse desoksüribonukleiinhappe (*deoxyribonucleic acid* – DNA) molekulides ning see päritakse vanematelt [1]. DNA on lineaarne molekul, mis salvestab informatsiooni nukleotiidideks nimetatavate ühikute jadana, kus üks nukleotiid omab üht neljast erinevast diskreetset väärtusest ehk alusest, mida tähistatakse tähtedega A, T, G ja C. DNA molekul koosneb kahest kõrvuti asetsevast komplementaarsest (A moodustab sideme T-ga, G C-ga) nukleotiidide ahelast [1,2].

Päristuumsetes organismides (kõik organismid peale bakterite ja arhede) on parajasti mitte kasutusel olev DNA kokku pakitud [2]. Inimese DNA paikneb rakutuumas 23 erineva paari kokku pakitud DNA molekuli ehk kromosoomina, mis sisaldavad kokku umbes 6 miljardit nukleotiidi [2].

DNA kõrval kasutatakse muudeks eesmärkideks üksikust nukleotiidide ahelast koosnevat sarnaselt informatsiooni salvestavat ainet ribonukleiinhape (*ribonucleic acid* – RNA), kus nukleotiidi T asemel on U [1,2]. Mõlema molekuli puhul on iga nukleotiidide ahela kaks otsa erinevad, ühte nimetatakse 5’ ning teist 3’, kusjuures DNA komplementaarsed nukleotiidide ahelad asuvad kohakuti vastassuunaliselt [1,2]. Joonis 1 kujutab nelja nukleotiidi pika DNA lõigu struktuuri.



Joonis 1. DNA ahela struktuur ning komplementaarsete nukleotiidide ahelate suunad [2].

Paljud organismide omadused tulenevad valkudest – suurtest keerukatest molekulidest, mis on saadud paljude aminohapeteks nimetatavate molekulide ühendamisel [2,17]. Valgud moodustavad näiteks ensüüme – molekule, mis käivitavad spetsiifilisi keemilisi reaktsioone, mis muidu toimuks väga aeglaselt või ei toimuks üldse [17].

DNA ei mõjuta organismi talitlust otse, vaid läbi protsessi, mida kutsutakse geenide avaldumiseks ehk geeniekspressiooniks ja mille tulemuseks on valgud. Ekspressiooni keskmes on kaks alamprotsessi: transkriptsioon ja translatsioon. Transkriptsiooni käigus luuakse 5' suunast 3' suunda liikudes ühe DNA nukleotiidide ahela teatud lõigule vastav lõik RNA-d ehk transkript [1,2]. Transkriptsiooni abil saadakse mitut erinevat tüüpi RNA-d, neist mRNA (*messenger* RNA) nukleotiidide järjestuse põhjal seatakse translatsiooni käigus ritta aminohappeid, et toota mRNA poolt kirjeldatud valk [1,2]. Siinses töös mõeldakse edaspidi transkriptide all just mRNA molekule.

DNA piirkondi, mis kirjeldavad valkude ülesehitust, nimetatakse geenideks, mistõttu kutsutakse DNA-s sisalduvat informatsiooni muuhulgas geneetiliseks informatsiooniks või geneetiliseks materjaliks ning kogu organismi DNA-s sisalduvat infot genoomiks [2]. Kogu genoomist moodustavad geenid vähem kui 2% [2].

Inimestevahelist geneetilist variatsiooni põhjustavatest erinevustest enamiku moodustavad üksiku nukleotiidi erinevused (*single nucleotide polymorphism* – SNP) [2–4]. Piirkondi, kus genoomid omavahel erinevad, kutsutakse geneetilisteks variantideks [1]. Kuna inimesel on igat 1-22 kromosoomi kaks tükki, on inimesel ka igat geneetilist varianti kaks korda. Neid variandi esinemiskordi nimetatakse alleelideks, kusjuures need võivad olla erinevad [1].

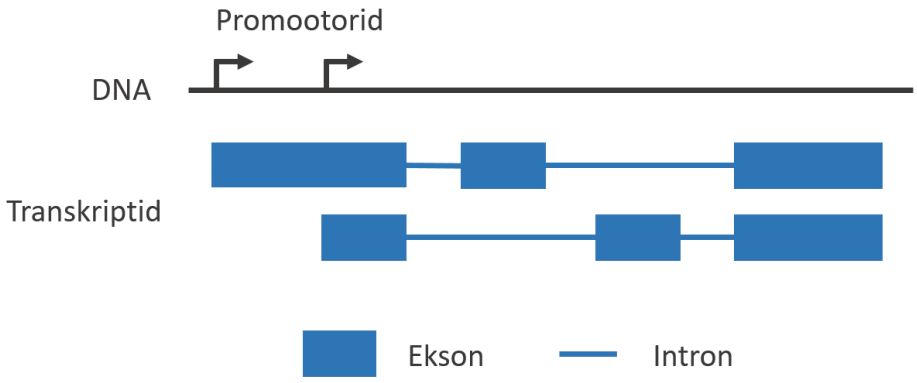
Kahe inimese DNA erineb umbes 0,1% ehk mõne miljoni nukleotiidi muutuse ulatuses [2,3]. Kuna genoomi valke kodeeriv osa on väike ning iga muutus ei pea DNA poolt kirjeldatavate valkude ülesehitust mõjutama, on inimestevaheliste mingeid tunnuseid mõjutavate erinevuste arv loetletav siiski tuhandetes, mitte miljonites [2].

2.2 Ekspressiooni regulatsioon ja promootorid

Transkriptsiooni alguskoha (*transcription start site* – TSS) ehk transkripti 5' otsale vastava koha DNA-l määrab alguskoha lähistel asuv spetsiifiline DNA järjestus promootor, mille külge kinnitub transkriptsiooni läbi viiv ensüüm [1,2]. Päristuumsetes organismides on paljudel

geenidel mitu erinevat promootorit, mida kasutatakse erinevates tingimustes, näiteks erinevates rakutüüpides või organismi arenguastmetes, erinevate osakaaludega [1,5]. Järelikult võib üks geen korraga kirjeldada mitut erinevat transkripti ja seeläbi mitut erinevat valku.

Päristuumsetes organismides toimub peale transkriptsiooni RNA töötlemine. Üks osa töötlemisest on splaissimine, mille käigus lõigatakse RNA seast välja teatud lõigud intronid ning allesjäävad lõigud eksonid liidetakse taas üheks RNA molekuliks [1,2]. Rohkem kui pooli inimese geene splaisitakse mitmel erineval viisil ning splaissimist võivad mõjutada erinevad promootorid ja rakus valitsevad tingimused [6]. Ühe geeni ekspressiooni mõjutamine erinevate splaissimisviisidega võimaldab näiteks luua väikese geenide arvuga palju rohkem erinevaid spetsialiseeritud valke [6].



Joonis 2. Promootorite mõju lõplikele transkriptidele erinevate TSS-ide ning splaissimiste kaudu.

Kuna ei transkriptsioon ega ka sellele järgnevad protsessid ei ole ühe geeni puhul alati samad, on selge, et DNA põhjal ei ole võimalik organismi talitluse kohta kõike öelda. Seega on kasulik teada, millist RNA-d on DNA põhjal sünteesitud. Promootorite mõju transkriptidele kujutab joonis 2. Erinevad promootorid võivad mõjutada mingi geeni ekspressiooni palju, näiteks kiirendada selle geeni valgu tootmist või toota erinevates olukordades selle valgu erinevaid vorme [5]. Paljusid geneetilisi variante, mis mõjutavad promootorite kasutust (näiteks promootorite osakaalusid transkriptsioonis), on seostatud ka mõne haiguse suurema riskiga [7,8]. Seega on promootorite kasutus oluline uurimisobjekt.

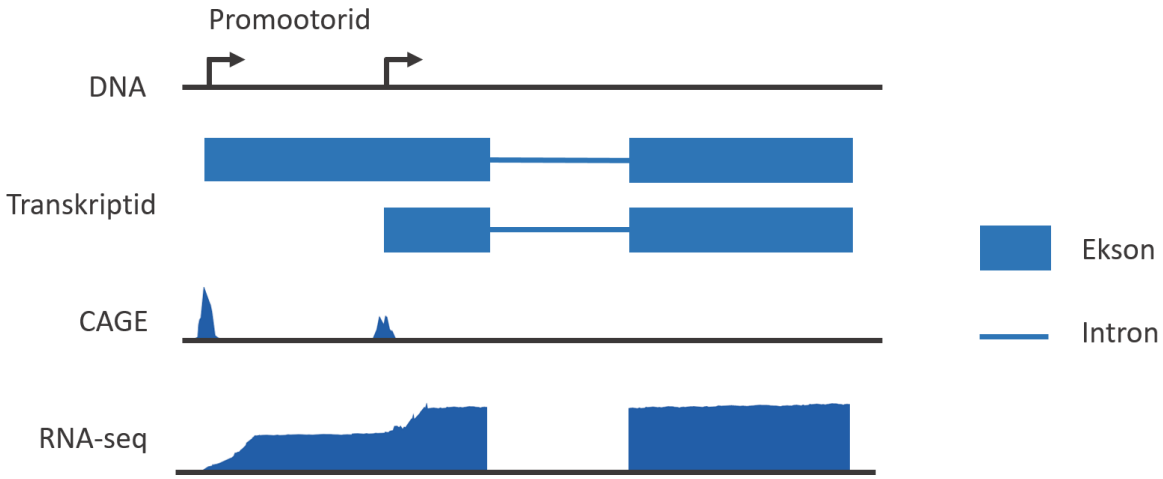
2.3 RNA sekveneerimine

RNA molekulidest nukleotiidide järjestuse lugemiseks kasutatakse protsesse, mille üldnimetus on RNA sekveneerimine (*RNA sequencing* – RNA-seq). RNA-seq protsessi käigus

sünteesitakse RNA molekulide põhjal neile vastavad DNA lõigud, mis seejärel fragmenteeritakse juhuslikest kohtadest teatud pikkusega (enamasti kuni 200 nukleotiidi) lõikudeks [9,10]. Fragmentide ehk lugemite nukleotiidijärjestused loetakse ning joondatakse viitegenoomi (näide liigi isendi tervest genoomist, tavaliselt mitme isendi põhjal loodud kunstlik genoom [2]) abil algseid RNA molekule kirjeldanuid DNA lõikudele [9,10]. Viitegenoomis asuvat geneetilise variandi versiooni nimetatakse viitealleeliks ning ülejäänuid alternatiivseteks alleelideks [18].

Joonis 3. Lugemite ühtlase pikkuse ja jaotuse korral on RNA lõigu otsad halvasti esindatud.

Promootori mõju TSS-ile tähendab, et TSS-i või transkripti põhjal saab järeldada, millist promootorit kasutati. Kuna liiga lühikesed lugemid sobivad juhuslikult mitmesse kohta hiiglaslikus genoomis (ja tuleb seetõttu joondamisest kõrvale jätta), on lihtne näha, et RNA ahela esimesed nukleotiidid esinevad kaugematega võrreldes palju vähemates sobiva suurusega lugemites. Seda kujutab joonis 3.



Joonis 4. Iga nukleotiidi esindatus joondatud RNA-seq ja CAGE lugemites kahe transkripti korral.

Seetõttu pole üllatav, et RNA-seq puhul on sageli raske kindlalt määrata rohkem kui ühte TSS-i geeni kohta [11]. CAGE (*cap analysis of gene expression*) on sekveneerimismeetod, mis juhuslike päriliku materjali fragmentide asemel loeb mRNA molekulide järjestuse 5’ otsi

[12,13]. On näidatud, et CAGE-i abil on võimalik mRNA 5’ otsi täpselt määrata ning seda on kasutatud uute TSS-ide avastamiseks [11]. Joonisel 4 on visualiseeritud neid kahte protsessi.

Kui on olemas informatsioon, millised transkriptid leiduvad, ehk transkriptide annotatsioonid, saab RNA-seq andmete abil hinnata transkriptide ekspressiooni ehk millised annoteeritud transkriptide ekspressioonivääртused seletavad ära nähtava RNA-seq signaali [19]. Selle protsessi teevad keeruliseks mitmed probleemid. Näiteks kattuvad transkriptid sageli piisavalt palju, et lugemeid ei ole võimalik üheselt kindlale transkriptile määrata [19]. Kui olemasolevad transkriptide annotatsioonid on ebatäielikud, võib sekveneerimisandmete põhjal teha transkriptide kasutuse kohta täiesti valesid järeldusi [7,20]. Lisaks, kui annoteeritud on ainult üks transkript, siis on selle osakaal alati 1, sõltumata sellest, milline RNA-seq signaal päriselt on. Nagu näha jooniselt 4, on promootorite kasutuse hindamiseks CAGE andmete analüüs lihtsam – tuleb lihtsalt kokku lugeda iga promootori algusega ülekattes olevad lugemid.

2.4 Uurimismeetodid

Paljud inimese tunnused, näiteks pikkus või eelsoodumus südamehaigusteks, ei ole määratud ühe geeni, vaid paljude erinevate geenide poolt. Seega mõjutavad neid tunnuseid korraga ka väga paljud geneetilised variandid. Levinud meetod nende variantide üles leidmiseks on ülegenoomne assotsiatsiooniuuring (*genome wide association study* – GWAS). GWAS-i käigus uuritakse statistilist seost huvipakkuva tunnuse ja väga paljude erinevate geneetiliste variantide (näiteks SNP-de) vahel üle terve genoomi paljudes isendites [21]. Selle meetodi abil on leitud kümneid tuhandeid seoseid erinevate SNP-de ning haiguste, eelsoodumuste või muude tunnuste vahel ning olemasolevate andmestike suurenedes see arv aina kasvab [21].

GWAS metoodikal on ka kitsaskohti. Näiteks on raske uurimiseks valida piisavalt mitme-külgseid populatsioone ja see ei pruugi suuta tuvastada mõjusaid variante, kui tunnust määravad üksikute variantide asemel mitme variandi kindlad kombinatsioonid [21]. Samuti takistab kindlate põhjuslike variantide leidmist aheldustasakaalutus (*linkage disequilibrium* – LD) ehk fakt, et lähestikku asetsevad geneetilised variandid päranduvad vanemalt lapsele peaaegu alati koos, mistõttu on need omavahel tugevalt korreleeritud [2]. Seega tähendab ühe variandi korrelatsioon tunnusega ka mitme lähedalasuva variandi korrelatsiooni tunnusega [21]. Tundmatu põhjusliku variandi asemel saab vaadelda siis statistiliselt olulisima seosega varianti ehk juhtvarianti.

Kui tunnus on kvantitatiivne, nagu näiteks pikkus, kutsutakse seda tunnust mõjutavaid genoomi piirkondi kvantitatiivse tunnuse lookusteks (*quantitative trait locus* – QTL) ning kui QTL mõjutab geenieksprессiooni, nagu näiteks mõne promootori osakaalu, siis ekspressiooni QTL-iks (*expression* QTL – eQTL) [2,22]. Kui uuritavatel tunnustel on asukohad genoomis (näiteks promootori osakaalu uurimise puhul saab rääkida promootori asukohast), võib arvutusmahu vähendamiseks ja mitmese testimise probleemi [23] vähendamiseks ilma paljusid eQTL-e välja jätmata uurida variante näiteks mõnesaja kb (*kilobase, base* – alus) laiuses alas tunnuse ümber, kuna eQTL-id kipuvad paiknema üsna neile vastava TSS-i lähedal [24,25].

GWAS-i jaoks vajalike geneetiliste variantide leidmiseks pole vaja isendite tervet DNA-d sekveneerida. Kindlate juba teadaolevate geneetiliste variantide olemasolu DNA-s saab (tuhandete või isegi miljonite kaupa) suhteliselt kiiresti ja odavalt kontrollida spetsiaalsete kiipide abil [2]. Lisaks on võimalik tänu LD-le hästi valitud geneetiliste variantide tuvastamise abil suure täpsusega järeldada ehk imputeerida mitu korda rohkem muid geneetilisi variante. [26,27].

2.5 Varasemad tulemused

Töö käigus võrreldakse saadud tulemusi kahe varasema sarnaste eesmärkidega artikli tulemustega.

Garieri jt [8] kasutasid 154 Kesk-Euroopa päritolu inimese CAGE sekveneerimise andmeid ja FANTOM5 promootorite andmestiku [15] annotatsioone ning leidsid muuhulgas 5376 geneetilist varianti, mis mõjutasid promootorite kasutust.

txrevise [7] on tööriist, mis muudab transkriptide annotatsioone, et võimaldada RNA-seq põhjal paremini hinnata promootorite kasutust. Muuhulgas pikendab txrevise pikima annotatsiooni põhjal lühemaid annotatsioone ja annab transkriptide keskel asuvad alternatiivsed eksonid kõikidele transkriptidele. Kerimov jt [14] kasutasid txrevise'i täiendatud transkriptide annotatsioone, et uuesti analüüsida paljusid eQTL andmestikke, sealhulgas leida promootori kasutust mõjutavaid eQTL-e.

3 Mõjusate variantide leidmine

3.1 Andmetöötlus

Töös kasutati andmeid samade isikute kohta, kelle andmeid analüüsisid Garieri jt [8]. Inimeste andmed pärinesid kahest erinevast projektist: 86 isikut 1000 Genomes Project'ist [28] ning 68 isikut GenCord projektist [29,30]. Iga inimese kohta olid olemas järgnevad andmed:

- CAGE meetodil saadud RNA lugemid, sekveneeritud Garieri jt poolt, ning nende metaandmed [31].
- 9,7 miljoni kahe alleeliga geneetilise variandi esinemise andmed. 1000 Genomes päritolu inimeste variantide olemasolu oli määratud nende genoomide täieliku sekveneerimise abil, GenCordi päritolu inimeste puhul olid Kerimov jt [14] imputeerinud variandid GenCord [29] projektis määratud variantide põhjal.

Lisaks kasutati CAGE meetodi abil koostatud FANTOM5 promootorite andmestikku [15], kust saadi viitegenoomile GRCh38 vastavad inimese promootorite ning nendele vastavate TSS-ide info [32]. Analüüsi lihtsustamiseks jäeti kõrvale kõik promootorid, mis olid andmestikus määratud mitmele geenile korraga (<1% promootoritest), mis jättis alles 101 522 promootorit üle 21 841 geeni.

CAGE lugemid joondati genoomile tööriistaga BWA [33], mida kasutasid ka Garieri jt [8]. Garieri jt kasutasid ka programmi Delve [34], aga siinse töö käigus ei olnud selle programmi kasutamine võimalik tarkvaravea tõttu (*segmentation fault*). Viitegenoomina kasutati uusimat inimese viitegenoomi GRCh38 [35]. Joondamisprotsess viidi läbi Snakemake töövoo-haldussüsteemi abil [36]. Igale promootorile vastavate RNA lugemite arvu loendamiseks kasutati programmi featureCounts [37]. 35,4% CAGE RNA lugemitest vastasid mõne FANTOM5 promootori TSS-ile.

Lugemite arvu põhjal jäeti analüüsist välja geenid, millele vastavaid promootoreid featureCounts ei tuvastanud. Selle tulemusena jäi alles 99 715 promootorit üle 20 771 geeni, millest 95 926 promootorit üle 19 845 geeni asusid kromosoomidel 1-22. Seejärel arvutati iga inimese jaoks välja iga promootori osakaal sellele vastavas geenis. Kui mõnel inimesel ei esinenud ühtegi mõne geeni promootorit, asendati nulliga jagamise vältimiseks promootorite osakaalud selle geeni kõigi promootorite osakaalude keskmisega. Siis normaliseeriti iga geeni

promootorite osakaalud, viies need järjestuse põhjal normaaljaotuse kvantiilideks. See on levinud meetod lineaarsete statistiliste mudelite parema kasutamise võimaldamiseks [38].

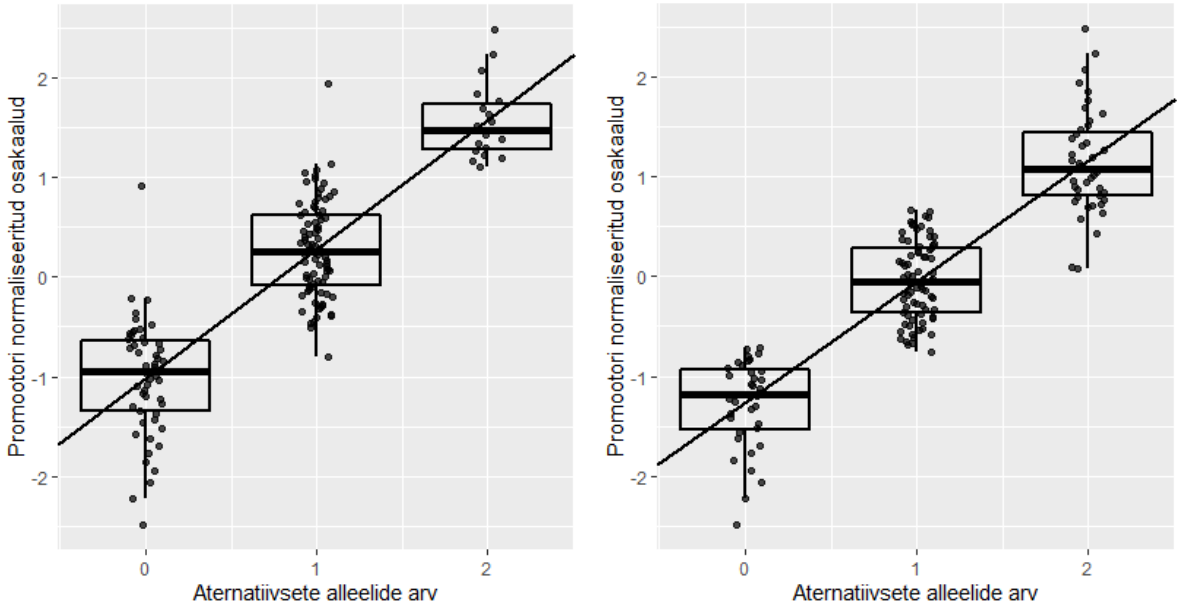
Statistiliste seoste leidmiseks geneetiliste variantide ja arvutatud promootorite osakaalude vahel kasutati töövoogu qtlmap [39], mis arendati välja Kerimovi jt tegevuse käigus [14]. Qtlmap leiab kromosoomidel 1-22 isendite tunnuste maatriksi ja kahealleeliliste geneetiliste variantide andmete vahelised seosed lineaarse regressiooni abil ning seoste statistilise olulisuse p-väärtuse permutatsioonitestide abil. Qtlmap tugineb populaarsele tööriistale QTLtools [40]. Selle töövoo väljundiks siinse töö andmete puhul on iga geeni mõne promootori kasutust kõige tõenäolisemalt mõjutav geneetiline variant ning sellele vastav testitud promootorite ja variantide arvuga arvestav p-väärtus. Kuna testiti ka paljusid geene, rakendati p-väärtustele valeavastuste määra (*false dicovery rate* – FDR) [41].

Iga geeni puhul vaadeldi geneetilisi variante 1 miljoni või 200 tuhande nukleotiidi kaugusel geeni algusest. Geenide algkoordinaadid saadi Alasoo poolt koostatud andmestikust [42]. Kui geeni infot selles andmestikus ei leidunud (<9% geenidest), kasutati geeni alguse asemel geeni FANTOM5 andmestikust saadud promootorite asukohtade keskmist.

Kerimovi jt [14] töö reprodutseerimiseks vajalikud RNA-seq lugemid [43], variantide esinemise andmed [28] ja transkriptide annotatsioonid [44] olid saadaval Tartu Ülikooli teadusarvutuste keskuse kaudu ning qtlmap-ile eelnenud töövood GitHubi kaudu [45,46].

3.2 Tulemused

Qtlmap leidis ±1Mb akna puhul 1134 ning ±200kb akna puhul 1341 geeni, millel leidus mõni promootorite kasutust mõjutav geneetiline variant FDR 5% juures. ±1Mb akna puhul leiti vähem statistiliselt oluliselt mõjutatavaid geene ning kindlaima mõjuga variantidest vähem kui 6% asus geeni algusest kaugemal kui 200kb. Seega vaadeldi edaspidi vaid ±200kb akna tulemusi. Joonis 5 kujutab qtlmapi poolt leitud seoseid.



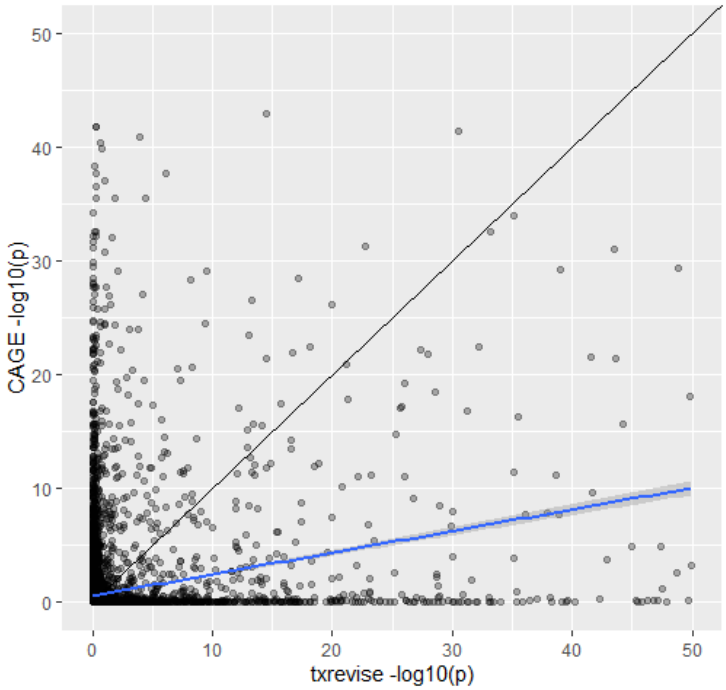
Joonis 5. Kahe kõige väiksema p-väärtusega geeni (ENSG00000082212, mille p = $2{,}3*10^{-43}$ ning ENSG00000101104, mille p = $6{,}8*10^{-43}$) puhul leitud seosed karpdiagrammil regressioonijoontega.

Garieri jt [8] leidsid samal andmestikul 5376 varianti, mis mõjutasid promootorite kasutust. Üks põhjus Garieri jt suurema arvu jaoks on see, et ühe geeni promootorite kasutust võib mõjutada mitu varianti. Veel mõõtsid Garieri jt promootorite kasutust üksiku promootori lugemite rohkuse põhjal. Nii tehes leiab seose ka näiteks siis, kui variant mõjutab terve geeni ekspressiooni kogust, aga mitte promootorite kasutuse osakaalu. Lisaks erines Garieri jt protsess veel mõnel viisil, näiteks kasutati joondamiseks BWA kõrval ka Delve'i. Seega on saadud tulem 1341 geeni oodatavas suurusjärgus.

Tabel 1. CAGE ning Kerimovi jt [14] RNA-seq + txrevise arvude võrdlus.

	CAGE	txrevise	Ühiseid
Geene	19845	20176	15137
p < 0.05 kokku	1341	1258	328
p < 0.05 ühiste geenide seas	1224	1156	

Kerimovi jt [14] protsess leidis ±200kb akna puhul 1258 geeni, millel leidus mõni promootorite kasutust mõjutav geneetiline variant FDR 5% juures. Täpsem arvude võrdlus on leitav tabelis 1. Seega oli CAGE-i abil leitud geenide arv suurem, aga ainult veidi. Kattuvus kahe meetodi leidude vahel oli samas väike – ühiseid statistiliselt olulisi geene oli vaid 328. Meetodite vähest nõustumist kujutab ka joonis 6.



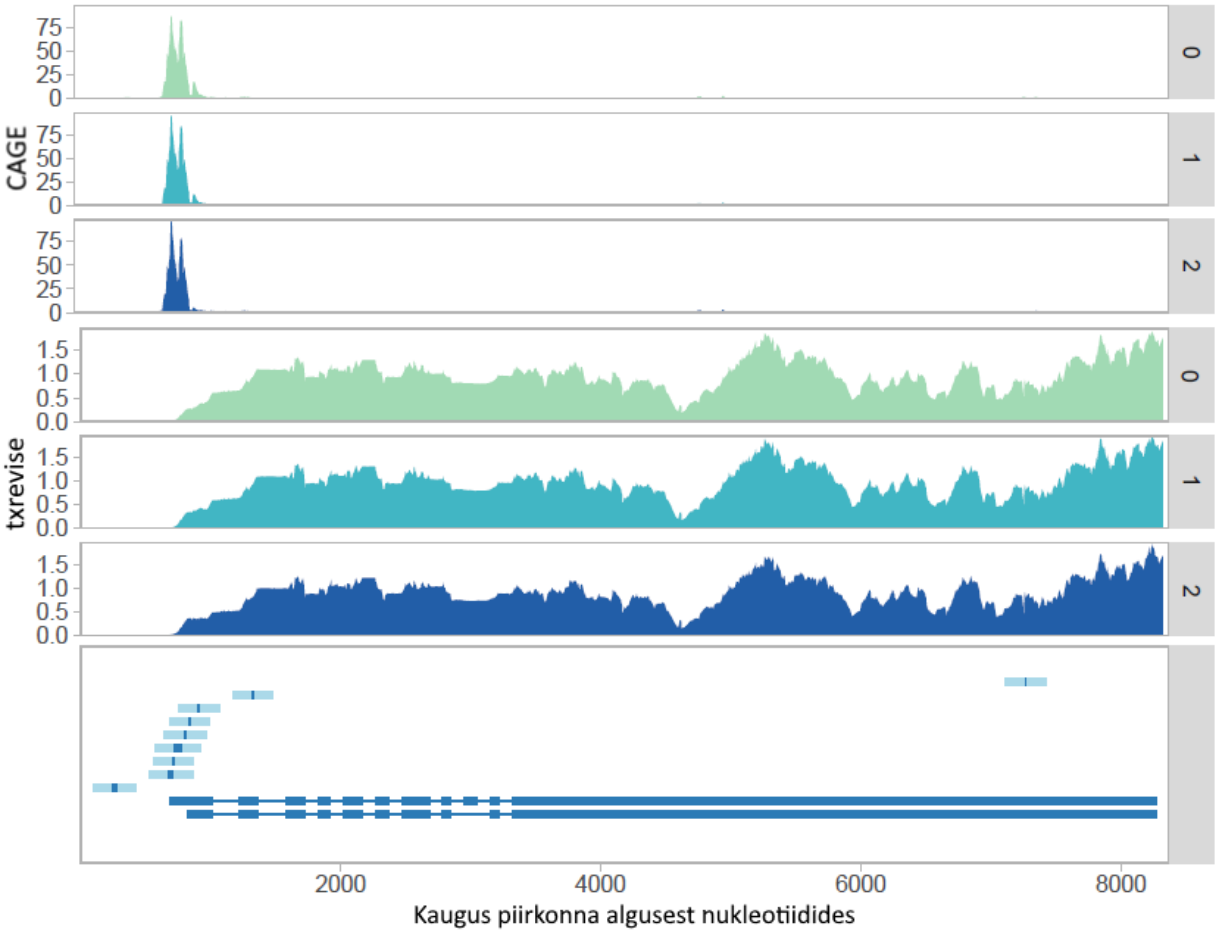
Joonis 6. CAGE-i ja txrevise'i vähimad p-väärtused iga geeni jaoks koos võrdsusjoonega (must) ja regressioonijoonega (sinine)

Ühiseid geenide statistiliselt olulisimaid geneetilisi variante oli CAGE-i ja txrevise'i vahel vaid 44. See aga ei tähenda, et kahe meetodi vahel oli vähene nõustumine ka mõjusate geneetiliste variantide osas. Vähima p-väärtusega variant ei pruugi olla põhjuslik, aga võib olla tõeliselt põhjusliku variandiga LD-s. Seega oleks variantide leidmise kattuvuse hindamiseks vaja teada mitte juhtvariantide kattuvust, vaid seda, kas need on omavahel LD-s, mis nõuab omaette analüüsi.

4 Uued transkriptide annotatsioonid

4.1 Uute annotatsioonide vajalikkus

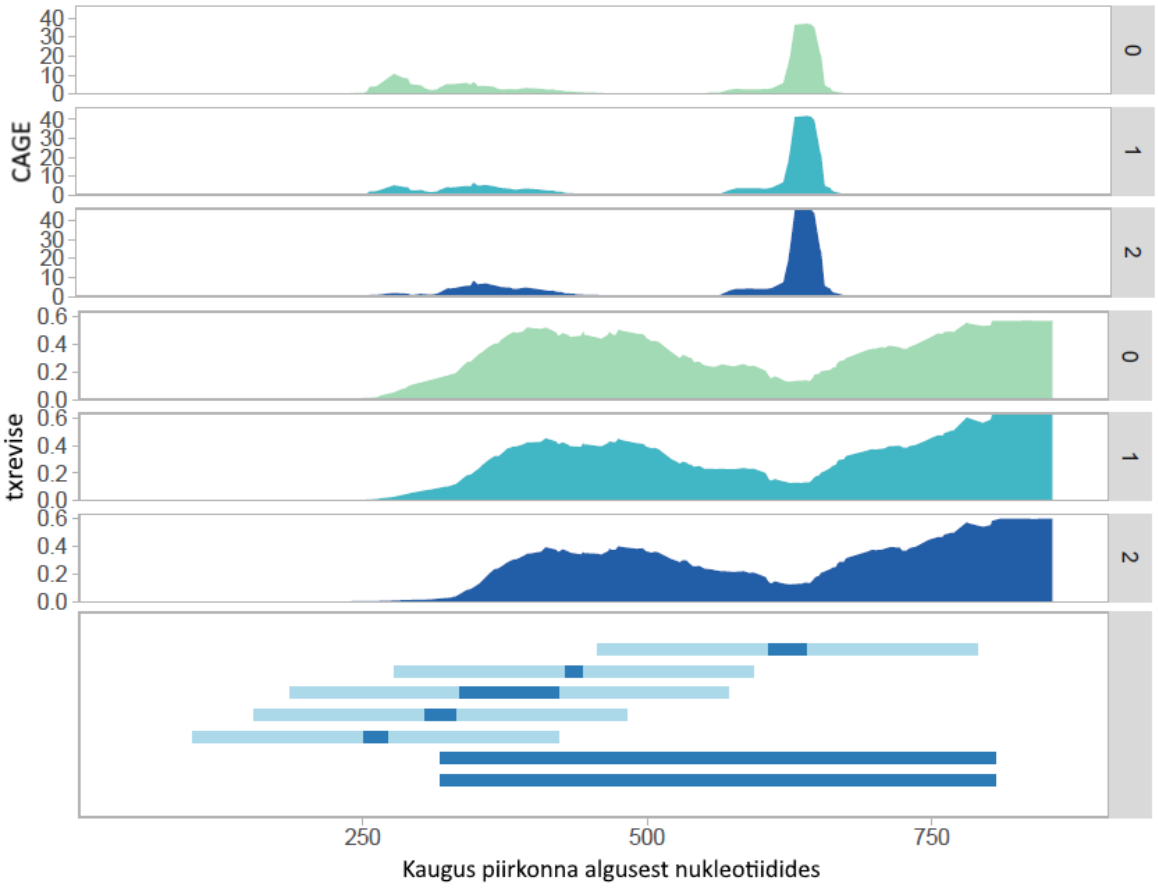
Järgmisena visualiseeriti CAGE ning RNA-seq lugemite kattuvusi annotatsioonidega ühiste geenide ja neile vastavate mõjusaimate variantide korral, kus CAGE-i põhjal leiti statistiliselt oluline seos, aga RNA-seq ja txrevise põhjal mitte. Visualiseeringute eesmärk oli mõista, kas ja kuidas oleks CAGE-i abil võimalik luua promootorite kasutust mõjutavate geneetiliste variantide leidmiseks kasulikke transkriptide annotatsioone. Joonis 7 kujutab näidet sellisest visualiseeringust.



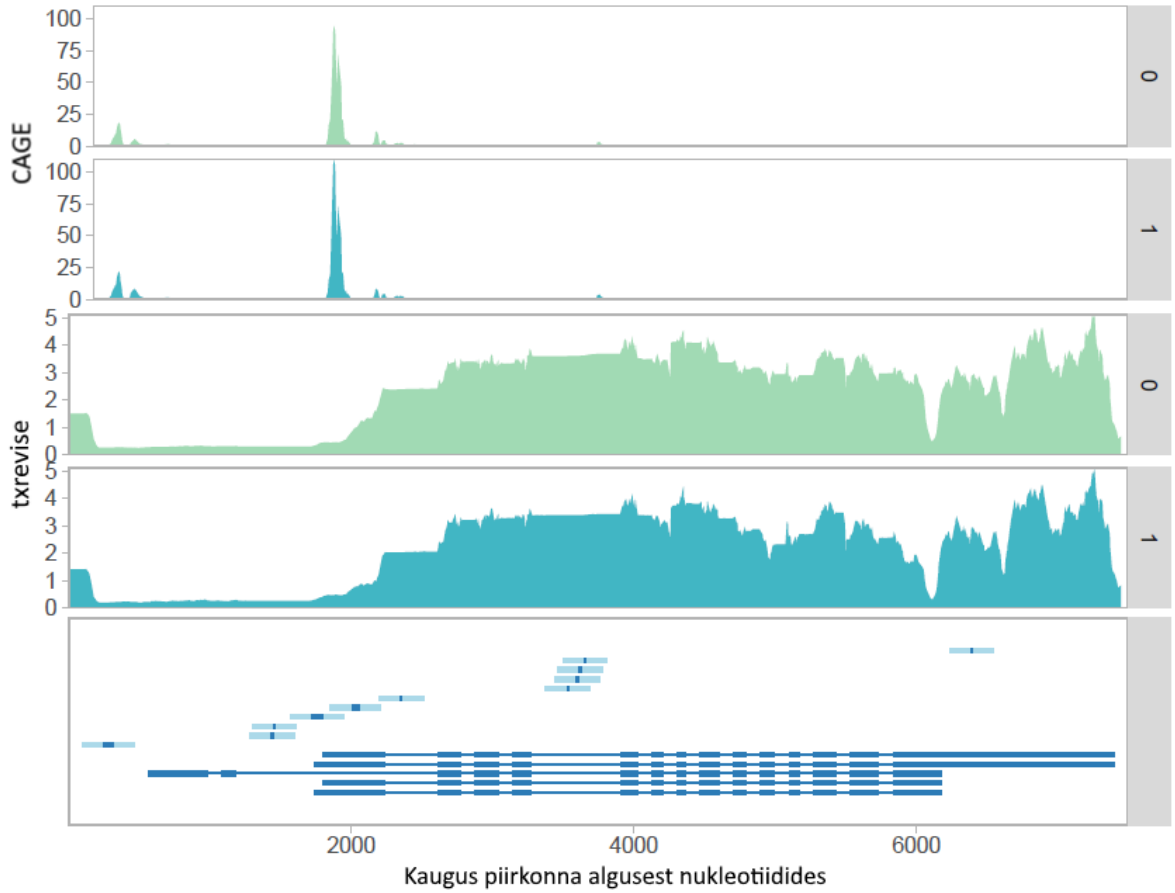
Joonis 7. Näide lugemite ja annotatsioonide kattuvusest. Välja valitud geeni ja variandi põhjal on kujutatud CAGE lugemid (üleval), RNA-seq lugemid (keskel) ning FANTOM5 promootorite ja txrevise eksonite annotatsioonid (all). Isikud on grupeeritud juhtvariandi alternatiivsete alleelide arvu alusel ning lugemid tähistavad grupi isikute keskmist lugemite arvu miljoni lugemi kohta. Annotatsioonide seas on FANTOM5 promootorid sinised, ümbritsetud helesiniste puhvritega ning txrevise eksonid sinised, ühendatud teiste sama transkripti eksonitega. X-teljel on nukleotiidide arv uuritava piirkonna algusest, aga parema nähtavuse nimel on kõik vahepealsed regioonid, kus ei asu ühtegi annotatsiooni ega puhvrit, lühendatud 100 nukleotiidi pikkuseks.

Valdava enamuse jooniste puhul puudusid uuritavale geenile vastavad txrevise annotatsioonid, RNA-seq lugemite arv oli vaadeldavas aknas täielikult või enam-vähem konstantne (tõenäoliselt intron), sageli 0, ning CAGE signaalid olid pigem nõrgad ning ei kattunud annotatsioonidega hästi. Kohati võis näha ka palju erinevaid CAGE kühme, mis ei kattunud promootorite annotatsioonidega ega RNA lugemite arvu tõusuga.

Sellest võib järeldada, et CAGE lugemite joondamine tekitas kohati müra ning et paljude geenide ekspressioon on võib-olla liiga madal bioloogiliselt oluliste järelduste tegemiseks. Müra mõju vähendamiseks tasuks tulevikus katsetada erinevaid geenide filtreerimise meetodeid, näiteks mitte kaasata analüüsidesse geene, mille ekspressioon on liiga madal. Teiseks võib proovida lühikeste CAGE lugemite jaoks paremini sobivaid joondamisalgoritme, mis võivad toota vähem joondusmüra.



Joonis 8. Üks näide geenist, mille juurest on CAGE lugemite ja promootorite annotatsioonide põhjal mõni txrevise transkripti annotatsioon puudu.



Joonis 9. Üks näide geenist, mille juurest on CAGE lugemite ja promootorite annotatsioonide põhjal mõni txrevise transkripti annotatsioon puudu.

Umbes 44 geeni puhul võis jooniselt näha, et CAGE lugemid kattusid promootori annotatsiooniga, aga samal kohal ei olnud selle geeni txrevise annotatsioonide hulgas eksoni algust. Neil juhtudel leidub tõenäoliselt sel kohal TSS ning ekson, mille kohta annotatsiooni veel ei leidu. Ilmselt ei muutu iga sellise TSS-i promootori kasutus oluliselt, aga nendele TSS-idele vastavate annotatsioonide lisamine võib aidata RNA-seq andmete põhjal leida rohkem geene, mille promootorite kasutust mõjutab geneetiline variatsioon.

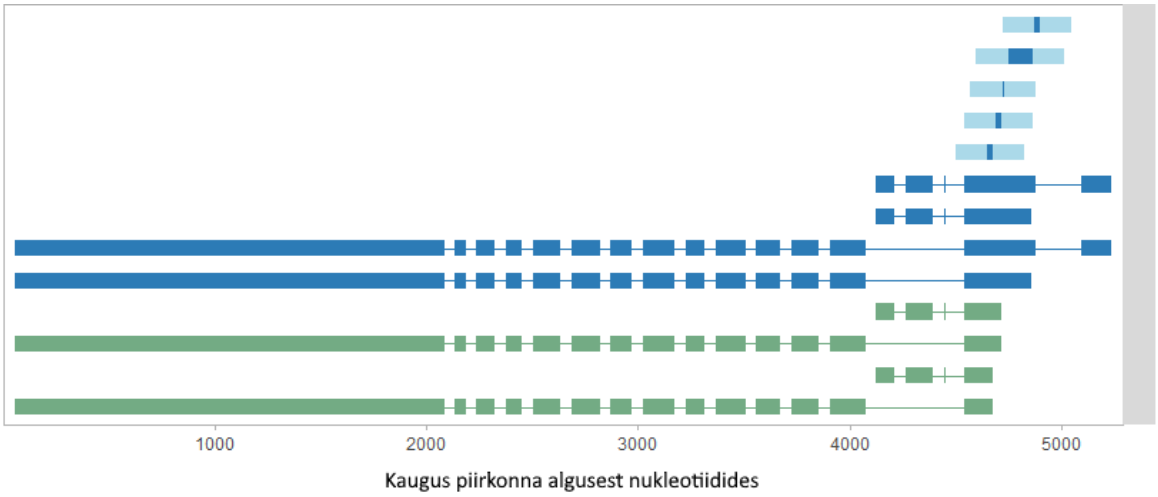
44-st sellisest geenist 40 puhul leidus potentsiaalse TSS-i lähedal mõni transkripti annotatsioon, võimaldades automaatselt genereerida kõikidel vastavatel juhtudel olemasolevate transkriptide põhjal uusi realistlikke transkripte. Selliseid olukordi kujutavad ka joonis 8 ja joonis 9. Otsustati luua uued annotatsioonid selliste olukordade jaoks. Nelja geeni puhul igasugused txrevise annotatsioonid puudusid, tehes annotatsioonide genereerimise raskemaks.

4.2 Annotatsioonide loomine

Kõikide ühiste geenide promootorite seast valiti välja promootorid, mis vastasid kõikidele järgmistele tingimustele, nii N=10 kui ka N=25 puhul:

- ei asu ühelegi olemasolevale txrevise eksoni algusele lähemal, kui N nukleotiidi,
- ei asu ühelegi juba valitud promootorile lähemal, kui N nukleotiidi
- asuvad mõne txrevise transkripti eksoni sees või sellest maksimaalselt 100 nukleotiidi kaugusel transkriptsiooni alguse suunas.

Valitud promootorite ja nendega kattuvate eksonite transkriptide abil loodi kunstlikud transkriptide annotatsioonid, mille esimene ekson oli kunstlik ekson promootori algusest kattuva eksoni lõpuni ning ülejäänud eksonid kattuva transkripti ülejäänud eksonid. Seda on kujutatud ka joonisel 10.



Joonis 10. Promootorite annotatsioonid, txrevise annotatsioonid ning nende põhjal loodud uued transkriptid (rohelised).

Selle protsessi tulemusena loodi N=10 puhul 36 582 uut transkripti üle 6980 geeni ning N=25 puhul 25 185 uut transkripti üle 6018 geeni.

4.3 Annotatsioonide mõju

Loodud transkriptid lisati Kerimovi jt [14] protsessis kasutatud transkriptide hulka Kaur Alasoo poolt [47] ning nende töövoogu jooksutati täiendatud transkriptidega uuesti. Tulemused on välja toodud tabelis 2.

Tabel 2. CAGE-i ning Kerimovi jt [14] RNA-seq + txrevise arvude võrdlus pärast annotatsioonide täiendamist.

	CAGE	txrevise N=10	txrevise N=25	Ühiseid N=10	Ühiseid N=25
Geene	19845	20176	20176	15137	15137
p < 0.05 kokku	1341	1397(+139)	1341(+83)	376(+48)	342(+14)
p < 0.05 ühiste geenide seas	1224	1295(+139)	1241(+85)		

Tabelist 2 on selgelt näha, et annotatsioonide täiendamine suurendas leitud geneetiliste variantide poolt mõjutatud promootorite kasutusega geenide arvu. Mõlemal juhul on leitud geenide arvu kasv suurem kui lugemite kattuvuse jooniste põhjal visuaalselt hinnatud puuduvate transkriptidega geenide arv 40. N=10 puhul oli isegi ka CAGE-i puhul statistiliselt oluliste leitud geenide arvu tõus 40-st veidi suurem.

Transkripte loodi kõikide, mitte ainult selgete CAGE kühmudega FANTOM5 promootorite abil. Seega pole tõenäoliselt paljusid loodud transkripte antud rakutüübis päriselt olemas. Mida rohkem välja mõeldud annotatsioone on ja mida tihedamalt need paiknevad, seda lihtsam on ilmselt teatud kohtades juhuslikult või joondusvea tõttu tekkinud lugemite koguse erinevust valesti tõlgendada transkriptide osakaalude erinevusena. Seetõttu annotatsioonide lisamise tulemusena suurenes lisaks tõeste positiivsete arvule kindlasti ka väärpositiivsete arv.

Kuna annotatsioonide lisamine aitas otsitavaid geene mingil määral paremini leida, aga muutis protsessi veaohtlikumaks, vajab annotatsioonide lisamise algoritm veel läbimõtlemist. Võrrelda tuleks annotatsioone paljude erinevate N väärtuste korral ning proovida tuleks ka teistsuguseid algoritme.

5 Kokkuvõte

Geenide avaldumist mõjutavad palju promootorid ehk geeni alguse lähedal asuvad piirkonnad, mille kasutuse erinevusi on seostatud mitmete haigustega. Tavalise RNA-seq sekveneerimismeetodi signaal on geeni alguses aga nõrk. CAGE on meetod, mis keskendub transkriptide algustele ja suudab palju paremini määrata promootorite kasutust.

Töö eesmärgiks oli uurida, kas CAGE tehnoloogia abil on võimalik promootorite kasutust mõjutavaid geneetilisi variante leida paremini kui palju levinuma RNA-seq meetodi abil. Töö käigus leiti 154 Kesk-Euroopa päritolu inimese CAGE andmete põhjal hinnang inimese selliste geenide arvule, mille promootorite kasutust mõjutavad geneetilised variandid. Seejärel uuriti joonistelt, millistes olukordades saaks RNA-seq transkriptide annotatsioone täiendada, kasutades CAGE andmete põhjal koostatud FANTOM5 promootorite andmestikku, ning loodi vastavad annotatsioonid.

Leiti 1341 geeni, millel leidus mõni promootorite kasutust mõjutav 200kb ulatuses asuv geneetiline variant FDR 5% juures. See number on oodatavas suurusjärgus ning veidi suurem kui Kerimovi jt RNA-seq-il põhineva protsessi tulemus sama akna puhul.

CAGE ja RNA-seq lugemite visualiseerimine näitas, et CAGE lugemite joondamise tulemus oli kohati väga mürane ning et paljude geenide ekspressioon oli võib-olla liiga madal bioloogiliselt oluliste järelduste tegemiseks. Müra mõju vähendamiseks tasuks tulevikus katsetada teistsuguseid joondamisalgoritme ning geenide filtreerimist.

FANTOM5 abil loodud transkriptide kasutamine suurendas kindlasti võimet RNA-seq andmete põhjal leida geene, mille promootorite kasutust mõjutavad geneetilised variandid, aga mitte palju. Samas suurendab väljamõeldud transkriptide lisamine kindlasti ka väärpositiivsete arvu. Selleks, et annotatsioonide lisamine omaks tugevamat mõju ning põhjustaks vähem väärpositiivseid, tuleks proovida erinevaid annotatsioonide genereerimise algoritme erinevate parameetritega.

6 Viidatud kirjandus

[1] Hartl D.L. Genetics: Analysis of Genes and Genomes. Fifth edition. Sudbury, Massachusetts: Jones & Bartlett Pub. 2001.

[2] Hartwell L. Genetics: From Genes to Genomes. Sixth edition. New York, NY: McGraw-Hill Education. 2018.

[3] Frazer K.A., Murray S.S., Schork N.J., Topol E.J. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, 2009, Volume 10, Issue 4, pp 241–251.

[4] Rahim N.G., Harismendy O., Topol E.J., Frazer K.A. Genetic determinants of phenotypic diversity in humans. *Genome Biology*, 2008, Volume 9, Issue 4, pp 215.

[5] Ayoubi T. Alternative Promoter Usage. *eLS*, 2005. https://doi.org/10.1038/npg.els.0005286 (03.12.2019)

[6] Kornblihtt A.R. Promoter usage and alternative splicing. *Current Opinion in Cell Biology*, 2005, Volume 17, Issue 3, pp 262–268.

[7] Alasoo K., Rodrigues J., Danesh J. et al. Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *eLife*, 2019, Volume 8. https://doi.org/10.7554/eLife.41673 (16.04.2020)

[8] Garieri M., Delaneau O., Santoni F. et al. The effect of genetic variation on promoter usage and enhancer activity. *Nature Communications*, 2017, Volume 8, Issue 1, pp 1–9.

[9] Koch C.M., Chiu S.F., Akbarpour M. et al. A Beginner's Guide to Analysis of RNA Sequencing Data. *American Journal of Respiratory Cell and Molecular Biology*, 2018, Volume 59, Issue 2, pp 145–157.

[10] Stark R., Grzelak M., Hadfield J. RNA sequencing: the teenage years. *Nature Reviews Genetics*, 2019, Volume 20, Issue 11, pp 631–656.

[11] Adiconis X., Haber A.L., Simmons S.K. et al. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nature Methods*, 2018, Volume 15, Issue 7, pp 505–511.

[12] Shiraki T., Kondo S., Katayama S. et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences*, 2003, Volume 100, Issue 26, pp 15776–15781.

[13] Kawaji H., Lizio M., Itoh M. et al. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Research*, 2014, Volume 24, Issue 4, pp 708–717.

[14] Kerimov N., Hayhurst J.D., Manning J.R. et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. *bioRxiv*, 2020. https://www.biorxiv.org/content/10.1101/2020.01.29.924266v1 (06.03.2020)

[15] A promoter-level mammalian expression atlas. *Nature*, 2014, Volume 507, Issue 7493, pp 462–470.

[16] Vija A. andreasvija/baka-kood. https://github.com/andreasvija/baka-kood (07.05.2020)

[17] Tymoczko J.L., Berg J.M., Stryer L. Biochemistry: A Short Course. Third edition. New York: W. H. Freeman. 2015.

[18] Koko M., Abdallah M.O.E., Amin M., Ibrahim M. Challenges imposed by minor reference alleles on the identification and reporting of clinical variants from exome data. *BMC Genomics*, 2018, Volume 19, Issue 1, pp 46.

[19] Garber M., Grabherr M.G., Guttman M., Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, 2011, Volume 8, Issue 6, pp 469–477.

[20] Trapnell C., Williams B.A., Pertea G. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 2010, Volume 28, Issue 5, pp 511–515.

[21] Tam V., Patel N., Turcotte M. et al. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 2019, Volume 20, Issue 8, pp 467–484.

[22] Sun W., Hu Y. eQTL Mapping Using RNA-seq Data. *Statistics in Biosciences*, 2013, Volume 5, Issue 1, pp 198–219.

[23] Chen S.-Y., Feng Z., Yi X. A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease*, 2017, Volume 9, Issue 6, pp 1725–1729.

[24] Montgomery S.B., Sammeth M., Gutierrez-Arcelus M. et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, 2010, Volume 464, Issue 7289, pp 773–777.

[25] Veyrieras J.-B., Kudaravalli S., Kim S.Y. et al. High-Resolution Mapping of Expression-QTLs Yields Insight into Human Gene Regulation. *PLOS Genetics*, 2008, Volume 4, Issue 10. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000214 (03.03.2020)

[26] Schurz H., Müller S.J., van Helden P.D. et al. Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population. *Frontiers in Genetics*, 2019, Volume 10. https://www.frontiersin.org/articles/10.3389/fgene.2019.00034/full (23.01.2020)

[27] Das S., Abecasis G.R., Browning B.L. Genotype Imputation from Large Reference Panels. *Annual Review of Genomics and Human Genetics*, 2018, Volume 19, Issue 1, pp 73–96.

[28] Data | 1000 Genomes. https://www.internationalgenome.org/data (05.03.2020)

[29] Gutierrez-Arcelus M., Lappalainen T., Montgomery S.B. et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife*, 2013, Volume 2. https://doi.org/10.7554/eLife.00523 (27.02.2020)

[30] EGAD00001000428 | European Genome-phenome Archive. https://www.ebi.ac.uk/ega/datasets/EGAD00001000428 (06.03.2020)

[31] Samples and Data < E-MTAB-5835 < Browse < ArrayExpress < EMBL-EBI. https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5835/samples/ (26.02.2020)

[32] Index of /5/datafiles/reprocessed/hg38_latest. https://fantom.gsc.riken.jp/5/datafiles/reprocessed/hg38_latest/ (27.02.2020)

[33] Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 2009, Volume 25, Issue 14, pp 1754–1760.

[34] FANTOM - Software. FANTOM. https://fantom.gsc.riken.jp/software/ (26.02.2020)

[35] GRCh38 - hg38 - Genome - Assembly - NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/ (26.02.2020)

[36] Köster J., Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 2012, Volume 28, Issue 19, pp 2520–2522.

[37] Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 2014, Volume 30, Issue 7, pp 923–930.

[38] McCaw Z.R., Lane J.M., Saxena R., Redline S., Lin X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics*, 2019. https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13214 (02.03.2020)

[39] Kerimov N., Alasoo K. kerimoff/qtlmap. https://github.com/kerimoff/qtlmap (16.02.2020)

[40] Delaneau O., Ongen H., Brown A.A. et al. A complete tool set for molecular QTL discovery and analysis. *Nature Communications*, 2017, Volume 8, Issue 1, pp 1–7.

[41] Benjamini Y., Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*, 1995, Volume 57, Issue 1, pp 289–300.

[42] Alasoo K. Metadata for various molecular traits included in the eQTL Catalogue. https://zenodo.org/record/3366011 (06.03.2020)

[43] E-GEUV-1 < Browse < ArrayExpress < EMBL-EBI. https://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-1/ (06.05.2020)

[44] Alasoo K. Transcription events constructed by txrevise. https://zenodo.org/record/3232932 (06.05.2020)

[45] eQTL-Catalogue/rnaseq. https://github.com/eQTL-Catalogue/rnaseq (06.05.2020)

[46] Kerimov N. kerimoff/qcnorm. https://github.com/kerimoff/qcnorm (06.05.2020)

[47] Alasoo K. Integrate CAGE annotations into txrevise · kauralasoo/txrevise@71993f4. https://github.com/kauralasoo/txrevise/commit/71993f40b7132a1d04b2db795e59f217d5ee2251 (04.05.2020)

7 Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, **Andreas Vija**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **„Promootorite kasutust mõjutavate geneetiliste variantide leidmine CAGE tehnoloogia abil“**, mille juhendaja on **Kaur Alasoo**, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, 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.

Andreas Vija
08.05.2020