

TARTU ÜLIKOOL
BIOLOOGIA-GEOGRAAFIATEADUSKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOTEHNOLOOGIA ÕPPETOOL

TARMO PUURAND

Rekombinatsioonipunktide jaotus
inimese 19. ja 22. kromosoomis

Magistritöö

Juhendaja: Erak. prof. Maris Laan, Ph.D.

TARTU
2004

Sisukord

Sisukord.....	2
Lühendid ja mõisted	3
Sissejuhatus	4
1 Kirjanduse ülevaade	5
1.1 Inimese genoomi ehitus	5
1.1.1 Kromosoomide arhitektuur	5
1.1.2 Inimese genoomi elemendid	7
1.2 Kromosomaalse materjali vahetus – meiootiline rekombinatsioon	9
1.2.1 Meioosi käik	9
1.2.2 Rekombinatsiooni protsess ja jaotus	12
1.2.3 Rekombinatsioonis osalevad valgud	13
1.3 Ristsiirde jaotuvus genoomis ja geneetilised kaardid.	14
1.3.1 Polümorfseid markerid inimese genoomis	14
1.3.2 Inimese genoomi kaardid	15
1.3.3 Otsesed meioosi uuringud	19
1.4 Haplotüübi blokid inimese genoomis	20
1.4.1 <i>Linkage disequilibrium</i> ehk lookuste alleelne aheldatus	20
1.4.2 Haplotüübid ja -blokid	22
1.4.3 Teooriaid haploblokkide tekkest	24
2 Töö eesmärgid	26
3 Materjal ja metoodika	27
3.1 Kromosoom 22 LD kaardi koostamine populatsiooni genotüpiseerimise andmetega	27
3.1.1 Kromosoom 22 chip <i>in silico</i>	27
3.1.2 Genotüpiseerimine	27
3.1.3 LD mustrite analüüs SNP Assistant programmiga	29
3.2 CEPH perekondade kromosoom 22 ja 19 rekombinatsioonipunktide kirjeldus	30
3.2.1 CEPH perekondade genotüpiseerimisandmed	30
3.2.2 Ensembl andmebaasid	32
3.2.3 Rekombinatsioonipunktide lokaliseerimine CRI-MAP programmiga	32
4 Tulemused	35
4.1 Inimese kromosoom 22 LD kaart	35
4.2 Kromosoomide 22 ja 19 rekombinatsiooni aktiivsuse iseloomustus ja rekombinatsioonipunktide jaotuvus	38
4.3 Rekombinatsioonipunktide analüüs genoomsete näitajate vastu	43
4.4 <i>Hotspot</i> -ide ja <i>coldspot</i> -ide genoomse järjestuse ja struktuuri analüüs	46
5 Arutelu	49
5.1 Eesti populatsiooni, Saksa triode ja CEPH perekondade LD mustri võrdlus inimese 22. kromosoomi põhjal	49
5.2 Rekombinatsioonipunktide jaotus 22. ja 19. kromosoomis	50
5.3 Ristsiirete <i>hotspot</i> -ide kirjeldus	51
5.4 Järeldused	53
Kokkuvõte	54
Summary	55
Tänuavaldused	56
Kasutatud kirjandus	57
Lisad	64
Lisa 1. Rekombinatsioonipunktide asukohad inimese 22. kromosoomil	65
Lisa 2. Rekombinatsioonipunktide asukohad inimese 19. kromosoomil	67
Lisa 3. Geeniala % ja rekombinatsioonide arvu võrdlus 250 kb-s aknas	70
Lisa 4. Alu korduste ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.	71
Lisa 5. GC% ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.	72

Töö elektrooniline versioon sisaldab linke (tekst sinine ja allajoonitud) uurimistööga seotud lisamaterjalidele WWW-s.

Lühendid ja mõisted

Alu	- Alu korduse element, kuulub SINE alla
APEX	- <i>Arrayed Primer Extension</i>
bp	- <i>base pair</i> , aluspaar
CEPH	- <i>Centre d'Etude du Polymorphisme Humain</i>
cM	- centi Morgan, 1cM = 1% rekombinante uuritavate lookuste vahel ühes põlvkonnas
DBS	- <i>double strand break</i> , kaheaahelaline katkestus DNA-l
HWE	- <i>Hardy Weinberg Equilibrium</i> , Hardy-Weinbergi tasakaal, $q^2+2qp+p^2=1$
INS/DEL	- insertioon või deletsioon DNA ahelas
kb	- 1 000 aluspaari
LD	- <i>Linkage Disequilibrium</i> , tasakaalustamata aheldatus
Mb	- 1 000 000 aluspaari
MHC	- <i>Major Histocompatibility Complex</i> , põhiline immuunsobivuskompleks
LINE	- <i>long interspersed nuclear element</i> , korduste klass inimese genoomis
PCR	- polümeraasi ahelreaktsioon
sAP	- kreveti aluseline fosfataas
SINE	- <i>short interspersed nuclear element</i> , korduste klass inimese genoomis
SNP	- <i>single nucleotide polymorphism</i> , ühe aluspaariline polümorfism
U	- <i>unit</i> , ensüümi ühik
UNG	- uratsiil-N-glükosülaas
Taq	- <i>Thermus acuaticus</i> DNA polümeraas

Sissejuhatus

Inimese Genoomi Projekti raames sekveneeriti inimese genoom ja identifitseeriti ca 32000 geeni ning 1,42 miljonit SNP-d. Inimese Genoomi projektist on välja kasvanud mitmed järgmised sihid, millest üks on *Human Haplotype Project* (HapMap), kus püütakse ühenukleotiidsete polümorfismide (SNP) abil kirjeldada inimese genoomi alleelse aheldatuse jämedat struktuuri. Inimese geneetikute hulgas on käimas praegu diskusioon: mis on need põhifaktorid, mis kujundavad inimese genoomi alleelse aheldatuse struktuuri. Siiani peeti põhifaktoriks meioosis toimuvat rekombinatsiooni ehk ristsiiret. Selle teooria põhjal peegeldab praegune LD muster ajaloolist rekombinatsioonilist aktiivsust. Nüüd on arvuti simulatsioonide ja ka empiiriliste andmete põhjal selge ka populatsiooni ajaloo otsustav roll LD mustrite kujunemisel.

Inimese genoomis olevate rekombinatsioonipunktide detekteerimine on põhiliselt toimunud geneetiliste kaartide loomise käigus, kui on genotüpiseeritud suuri perekondi mikrosatelliitsete markeritega, mis asuvad üksteisest keskmiselt 0,5 Mb vahemaadega. SNP-d seevastu asuvad inimese genoomis ca iga 300 bp järel. SNP-de kasutamine on andnud võimaluse detailsemalt uurida ristsiirete alteid genoomseid segmente ja võimalikku molekulaarset mehhanismi. Inimese genoomis ei ole identifitseeritud rekombinatsiooni *hotspot*-i spetsiifilisi DNA konsensusjärjestusi. Kõrgtihedalt paiknevate SNP-de kasutamisega on suurem tõenäosus see küsimus lahendada, kui need konsensusjärjestused eksisteerivad.

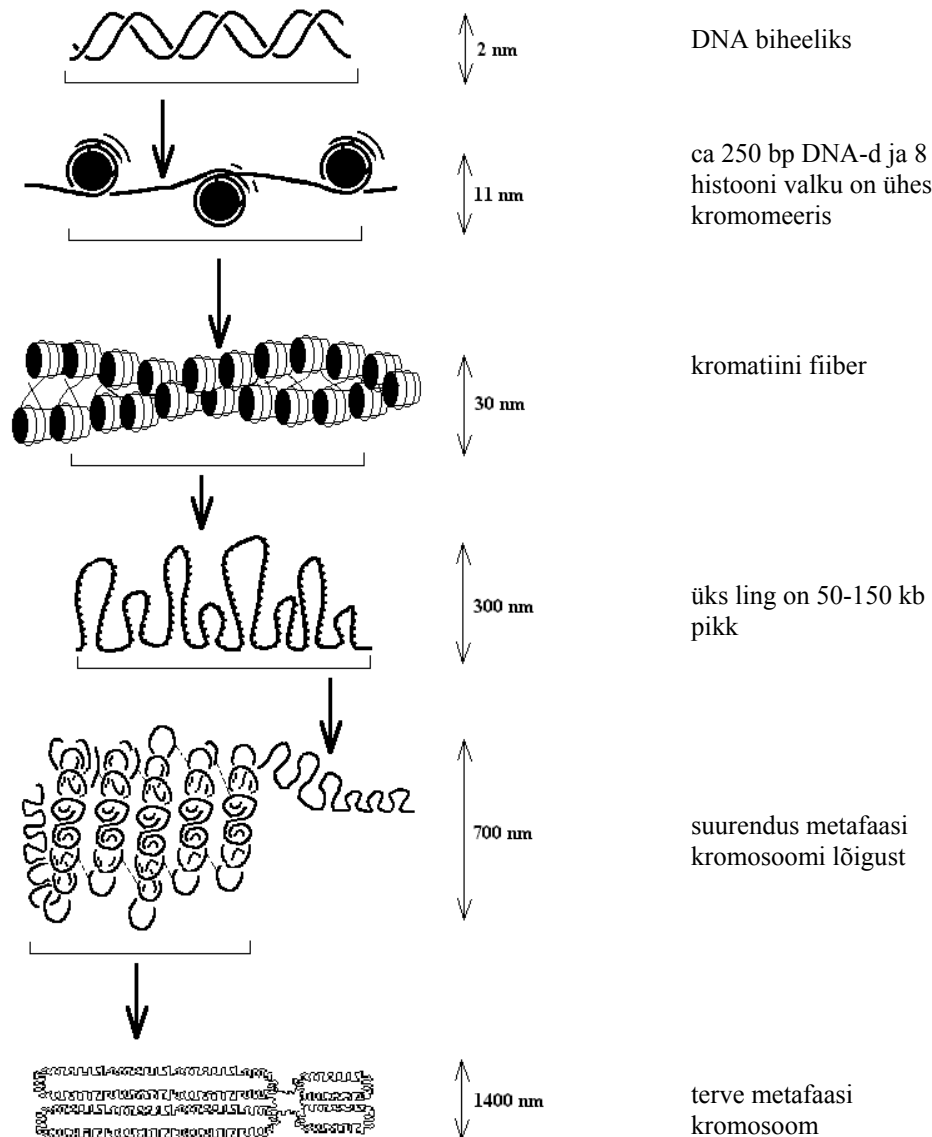
Käesoleva töö eesmärk on SNP markereid kasutades uurida ristsiirde punktide jaotust inimese genoomis ja korrelatsiooni struktuursete genoomielementidega (Alu kordused, kodeeriv ala, vöödistus, GC%) inimese 19. ja 22. kromosoomis, mis omakorda aitab paremini mõista rekombinatsiooni aktiivsust erinevates kromosoomi piirkondades, eri populatsioonides ja indiviidides.

1 Kirjanduse ülevaade

1.1 Inimese genoomi ehitus

1.1.1 Kromosoomide arhitektuur

Valdav osa inimese kromosoomide uuringutest on tehtud metafaasi kromosoomide põhjal. Metafaasis on kromosoomid kõige enam kondenseerunud (Joonis 1) ja õde kromatiidid on tsentromeeri kohalt seotud.



Joonis 1. Metafaasi kromosoomi pakkumise tasemed (modifitseeritud joonis 8-30, Alberts *et al.*, 1994).

Kromosoomi arhitektuuri uurimise varasemates töödes on metafaasi kromosoomide töödeldud erinevate DNA-ga seonduvate kemikaalidega, millest tuntuim on Giemsa värv. Värvimise tulemusena saadakse mikroskoobis nähtav vöödistus. Vöötide spetsiifika järgi on püütud iseloomustada erinevate kromosoomipiirkondade osalemist kindlates rakulistes protsessides (Tabel 1).

Tabel 1. Erinevate värvide ja kromosoomi struktuurielementide võrdlus vöötidest. Selgelt on näha GpG saarte, SINE elementide, kiasmide sageduse koondumine T-vöötidest.

Tunnus	T-vööt (telomeerides)	R-vööt (vastand G-le)	G-vööt (Giemsa värv)	C-vööt (osades hetero- kromatiinirikastes tsentromeerides)
I DNA värvid				
Quinakriin	Hall	Hall	Selge	Mõlemad
Akridiin oranž või kuumus	Selge	Selge	Hall	Hall
DAPI, DIPI (AT-paari spetsiifiline)			Enamalt	
Kloromütsiin (GC- paari spetsiifiline)		Enamalt		
II Kordusjärjestused				
Transposoon-päritolu	SINE	SINE	LINE	Harva
Satelliit DNA	Minisatelliidid	Mikrosatelliidid	Mikrosatelliidid	Satelliidid
III Primaarjärjestuse omadused				
Isohoidid	Palju H3 Mõni H1, H2	Mõni H3, Palju H1,H2	Harva H1, H2 Palju L1, L2	Mõned L1, L2
DNAaas I tundlikkus	Kõrge	Kõrge	Madal	Madal
Metüleerimata CpG saared	Väga palju	Palju	Mõned	Puudub
Atsetüleerimise tase	Kõrge	Kõrge	Madal	Madal
IV Käitumine meioosis				
Koduhoidjate ja koosspetsiifiliste geenide replikatsioon	Koduhoidjad, ½ koe- spetsiifilistest	Koduhoidjad ½ koe- spetsiifilistest	½ koe- spetsiifilistest	
Replikatsiooni aeg	S faasi I pool	S faasi I pool	S faasi II pool	S faasi II pool
Kiasmide sagedus	Väga kõrge	Kõrge	Madal	Puudub

Bernardi ja tema kolleegid avastasid 30 aastat tagasi ultratsenrifuugimise käigus laiaskaalalise GC% erinevuse kromosoomide erinevates piirkondades (Filipski *et al.*, 1973; Bernardi, 2000). Nad näitasid inimese genoomi ehitust GC% alusel 300kb blokkidena. Blokid e. isohoorid jaotatakse viide klassi, L1, L2, H1, H2 ja H3 vastavalt madala (L) või kõrge (H) GC% järgi. H3 isohoorid on geenirikkad (eriti koduhoidjad geenid) ja esinevad T-vöötides (Saccone *et al.*, 1996). G-vöödid sisaldavad valdavalt L1 ja L2 isohoore. On arvatud, et kromosoomide vöotide üleminekul on spetsiifilised järjestused, mis vahetavad GC-rikka ala GC-vaeseks alaks.

Inimese kromosoomide keskmine GC sisaldus on 41% sekveneeritud genoomi põhjal (Lander *et al.*, 2001). Kuna seniks sekveneerimata heterokromatiini GC% on enamuses alla 35%, siis keskmine GC% langeb 38%-ni. Sellisel juhul jaotub 20 kb pikkuste raamide arv GC% skaalal normaaljaotuse järgi. Kõrget GC% seostatakse geenialadega.

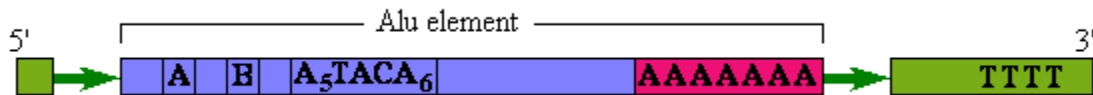
1.1.2 Inimese genoomi elemendid

Viimase dekaadi jooksul on kromosoomide visuaalsest iseloomustamisest ülemindud kromosoomide sisulisele kirjeldamisele, kus aluseks on kromosoomide täpne nukleotiidne järjestus. Inimese Genoomi Projekti käigus saadud andmete põhjal on kromosoomi bioloogiliste elementide (korduselemendid, geenid, duplikatsioonid) jaotuvus väga mitmekesine. 45% inimese kromosoomidest kirjeldatakse kui korduselemendid. Ülevaade korduselementidest on Tabelis 2.

Tabel 2. Korduselementide gruppide jaotus inimese genoomis. 45% inimese genoomist on korduselementide all (Lander *et al.*, 2001).

	Arv (1000x)	Kogupikkus (Mb)	Osalus genoomis (%)
SINE	1558	359,6	13,14
LINE	868	558,8	20,42
LTR	443	227,0	8,29
DNA transposoonid	294	77,6	2,84
Klassifitseerimata	3	3,8	0,14
Kokku:		1226,8	44,83

10,6% genoomist on enda alla võtnud kõige enam levinud primaatide spetsiifilised Alu elemendid (1,09 miljonit tk, Joonis 2), mis kuuluvad SINE-de alla ja mis on 130 bp ulatuses sarnane 7SL RNA järjestusega. On mitmeid tõendeid, et Alu elementide olemasolu genoomisegmendis tõstab selle rekombinatsioonilist aktiivsust (Dawson *et al.*, 2002).



Joonis 2. Alu elemendi skeem. A ja B blokk on RNA polümeraas III promootor. Nooled näitavad Alu-siseste korduvate järjestuste orientatsiooni.

Alu elemendid on jaotatud ca 20 alamklassiks. Alu element on ca 300 aluspaari pikk, sõltuvalt 3' oligo(dA)-rikkast alast. Keskkel on A rikas ala, otstes lühikesed pööramata kordused ja keskkel RNA-polümeraas-III promootor. Kuna Alu elementidel puudub avatud lugemisraam, siis arvatakse, et nad on vajalikud LINE elementide transkriptsiooniks. Kuigi Alu elementide järjestus on omavahel väga sarnane, on nende järjestus 55 miljoni aasta jooksul väga palju muteerunud. Uue Alu järjestuse koopia ekspansiooniks on arvatud üks iga 200 sünni kohta (Batzer *et al.*, 2002). Alu järjestusi teistel imetajatel leitud ei ole. Hiire Alu sarnane element, B1, on ca 150 bp pikk ja on esmaselt hiire genoomi sisenenud samal ajal, kui Alu primaatidele (ca 55 milj. aastat tagasi).

Kromosoomides asuvate duplikatsioonide analüüs on näidanud, et 3,3% kogu genoomist moodustavad intra- ja interkromosomaalsed duplikatsioonid. Eriti rohkelt on duplikatsioone kromosoomides 7, 9, 10, 16, 17 ja 22. Paraloogse rekombinatsiooni käigus põhjustavad ligistikku paiknevad duplikatsioonid pikkade kromosoomide lõikude deletsioone (Smith-Magenis, Charcot-Marie-Tooth 1A, Prader-Willi/Angelman, DiGeorge, Williams sündroomid) või insertioone (Bailey *et al.*, 2002). Duplikatsioonidel võib olla suurem tähtsus inimeste fenotüübi määramisel ja haiguste tekkel, kui seni kirjeldatud (Eichler, 2001; Stankiewicz *et al.*, 2002).

Lähedase järjestusega geenid kuuluvad geeniperekondadesse, mis on organiseerunud ühte või mitmesse klastrisse. Näiteks DNA pakkimisvalkude, histoonide, 61 geeni asuvad seitsmes eri kromosoomis, kuid enamus (49 tk.) asuvad kahes klastris 6p21.3 ja 6p22

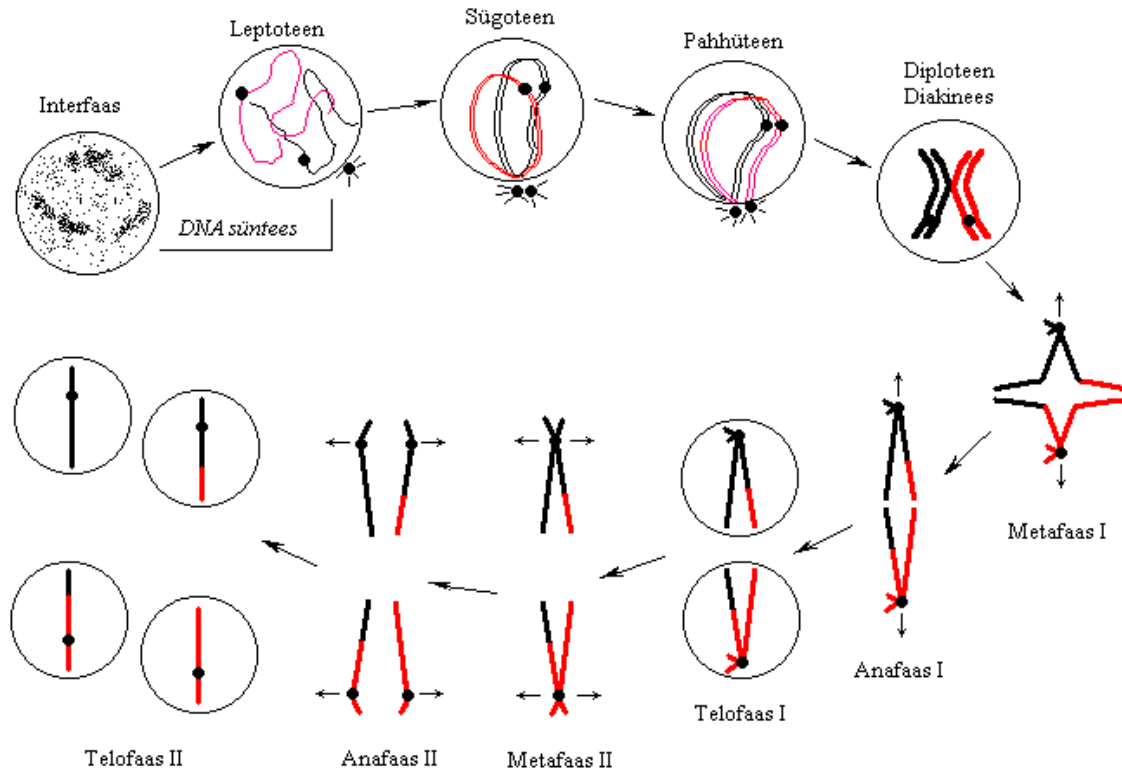
regioonis. Sarnaste geenide gruppeerumine on seotud evolutsioonis toimunud mittehomooloogilise intrakromosomaalse rekombinatsiooniga. Kui rekombinatsiooni sagedus on sõltuv kodeerivast alast, siis võib rekombinatsiooni *hotspot*-id olla seotud nn “avatud kromatiidi” struktuuriga ehk transkriptsiooniaktiivsete segmentidega.

1.2 Kromosomaalse materjali vahetus – meiootiline rekombinatsioon

1.2.1 Meioosi käik

Meioosi käigus jaguneb diploidne sugurakk, et toota haploidseid gameete suguliseks paljunemiseks. Meioos tagab bioloogilise populatsiooni varieeruvuse ja on lähtematerjal evolutsioonile. Meioosi ja mitoosi olulisimaks erinevuseks on meioosis toimuv geneetilise materjali vahetus. Kuigi meioos on kõrgelt konserveerunud protsess eukarüootidel, on siiski meioosi regulatsioonis tõestatud erinevused pärimi ja hiire puhul (Cohen *et al.*, 2001). Meioosi käik on kujutatud Joonisel 3. Kaks põhilist sündmust meioosi esimese jagunemise juures on homoloogsete kromosoomide paardumine ja nende paiknemine teineteisele poole meioosi käävist. Tütarkromosoomid eralduvad teineteisest meioosi teises jagunemises, MII-s (Joonis 3).

Premeiootiline S faas kestab kaks korda kauem kui mitootiline S faas. MI profaas kestab 16 päeva mehel ja aastaid naisel, mitoosi profaas kestab vaid mõned tunnid. Profaasi alguses (leptoteenis) hakkavad kromosoomid paistma peene niidina. FISH analüüs mehe kromosoomi 1 spetsiifiliste proovidega on näidanud, et enne meioosi profaasi hõlmavad mõlemad kromosoomid eraldi oma asukohta tuumas. Tsentromeersete ja telomeersete korduste FISH meetodil detekteeritud proovid on kinnitanud homoloogsete kromosoomide paardumise puudumist enne meioosi. Kui meioos algab, reorganiseeruvad pakitud kromosoomid mikroskoobis nähtavateks pikkadeks peenikesteks leptoteeni niitideks. Leptoteeni arenedes tulevad piki niidisarnase kromosoomi ahelat nähtavale kondenseerunud regioonid ja tulevad veelgi rohkem nähtavale, kui kromosoomid järk-järgult kontakteeruvad (Scherthan *et al.*, 1998).

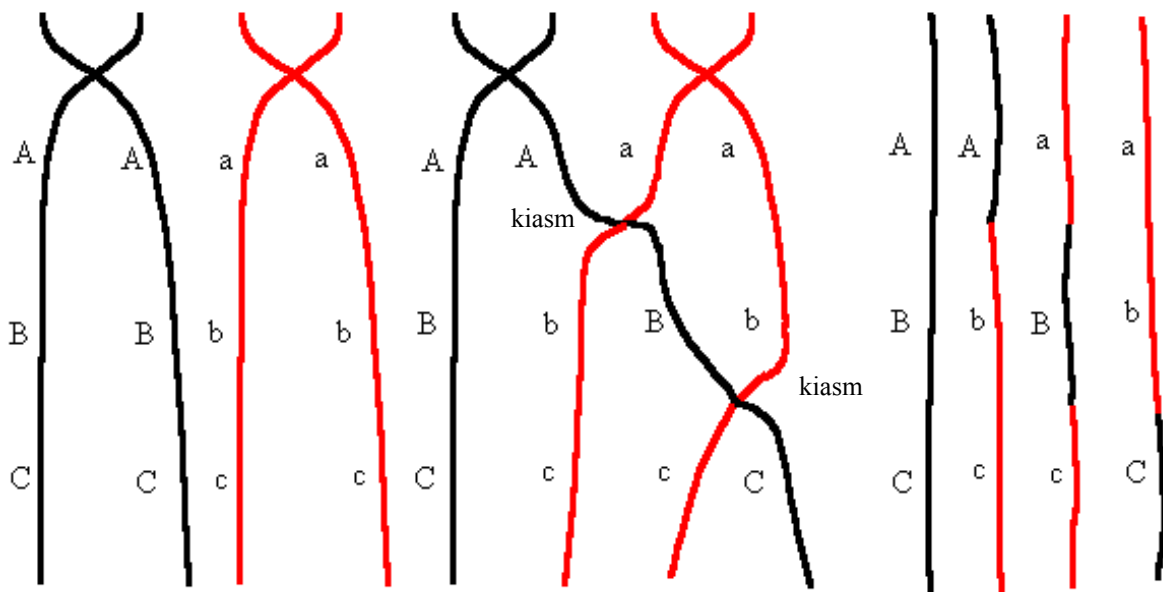


Joonis 3. Meioosi skeem. Ring ümber kromosoomide tähendab raku tuumamembraani olemasolu. Homoloogiline rekombinatsioon algab pahhüteenis. (Modifitseeritud joonis 9-1, Miller *et al.*, 2001).

Homoloogne paardumine algab sügoteenis. Varaseim nähtav sündmus on telomeeride koondumine tuumamembraanile ja tsentromeeride liikumine koondumiskohta tuuma membraanil, mida nimetatakse *bouquet* moodustumiseks. Homoloogsete kromosoomide paardumine algab mitmest kohast: igal paaril sügoteeni arenedes. Sel ajal on osad kromosoomi piirkonnad nähtavad peenikeste leptoteeni niitidena, ülejäänud osad on aga paardunud (moodustavad sünapsi) moodustamaks tihedaid pahhüteeni-faasi kromosoome veelgi enam kondenseerunud regioonidega (kromomeeridega). Tavaliselt toimub sünap homoloogsete kromosoomide täiesti sarnaste segmentide vahel. See on tõestatud homoloogsete kromosoomide puhul, kui ühes kromosoomis on inverseeritud regioon. Samuti on see näha mittehomoogsete kromosoomide paardumisel, kui tekivad translokatsioonid.

Homoloogide paardumise alguses on 46 pikka kromosoomi küllaltki läbipõimunud üksteisest. Elektonmikroskoobis on näha, et mittehomoogsete kromosoomide ristumised on sagedased sügoteenis nii mehe kui naise meioosis, kuid

väga harvad diploteenis ja MI-s. Lõpuks ei ole ühtegi ristumist enama kui kahe bivalendi vahel näha. “Selginemine” on korraldatud topoisomeraas II poolt, mis lubab erinevatel kromosoomidel teineteisest “läbi minna”. Pähhteenis on sünapsis valmis ja iga paardunud kromosoom koosneb pikkadest kromomeeride järjestustest, mis paistavad kui erineva suurusega helmed piki kromosoomi niiti. Iga homoloogne paar moodustab bivalendi. Bivalendi pikkus pähhteenis on 1/5 leptoteeni kromosoomi pikkusest. Kromomeeride muster sarnaneb G-vöödistusele mitootilise profaasi kromosoomis. Homoloogne rekombinatsioon, mille käigus vahetavad homoloogseid kromosoomi segmente kaks kromatiidi neljast (Joonis 4), toimub pähhteeni staadiumis, kuigi tulemusi ei näe enne, kui meioosi järgmises staadiumis – diploteenis.



Joonis 4. Homoloogsete segmentide vahetamise skeem. Kiasmid asuvad punaselt ja mustalt kujutatud kromosoomide ristumiskohas. (Modifitseeritud joonis 1.2, Ott, 1999).

Kromosoomide kondenseerumine jätkub läbi diploteeni. Kaks homoloogset kromosoomi, mis moodustavad bivalendi, hakkavad teadmata mehhanismiga üksteisest eemalduma kuni neid hoiavad koos vaid kiasmid. Kiasme on võimalik üle lugeda. Johansson eeldas juba 1909. aastal, et iga kiasm on ristsirde tulemus, hõlmates ühte kromatiidi mõlemast homoloogsest kromosoomist.

Tuuma membraan laguneb lühikese prometafaasi staadiumi jooksul ja bivalendid kogunevad metafaasi plaadile. Anafaasis liiguvad kromatiidid ühe funktsioneeruva

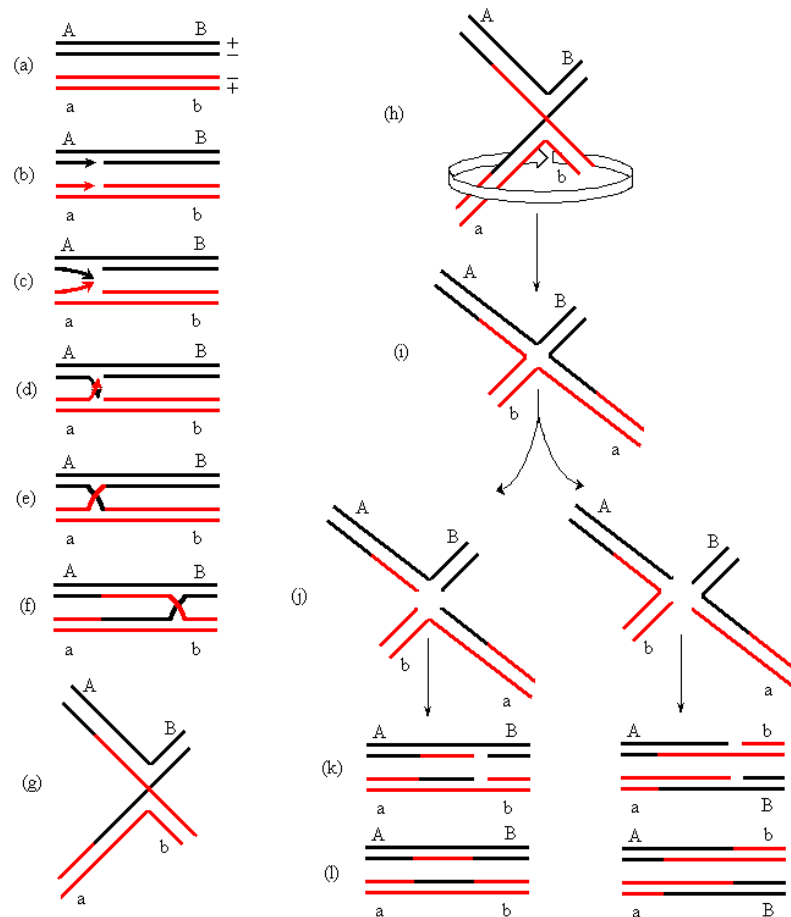
tsentromeeri poole. Telofaasi eel, interkineesi jooksul, kromosoomid dekondenseeruvad. MII sarnaneb mitosisile, erinev on vaid haploidne kromosoomide arv.

1.2.2 Rekombinatsiooni protsess ja jaotus

Seni on kiasmide arvu ja kromosoomide lahknemise arvu peetud võrdseks. Mehel on keskmiseks kiasmide arvuks 53 ja naisel lisaks 25% rohkem. Kiasmide visuaalne uurimine diakineesis on näidanud, et igas õlas on vähemalt üks kiasm ja pikemas õlas kaks kiasmi (Hulten, 1974). G grupi kromosoomides, mis on kõige väiksemad kromosoomid (21., 22. ja Y kromosoom), on pea-aegu alati üks kiasm. Kahe kiasmi vahel esineb interferents, mis välistab nende füüsilise ligistikku paiknemise (Hulten, 1974). Kiasmide liikumine ja arvu muutus meioosi käigus on inimesel väheuuritud. Kromosoomide ahelate vahetamise molekulaarseid mehhanisme on välja käidud mitmeid, kuid neist tuntuim on Holliday mudel (vt. Joonis 5).

Senistest rekombinatsiooni puudutavatest töödest on topeltrekombinandid (nn. liiga lähestikku asuvad rekombinatsioonipunktid) välja jäetud kui genotüpiseerimisvead. Housworth on oma töös näidanud, et peale interferentsi all olevat rekombinatsiooni, eksisteerib veel teine rekombinatsiooni rada – interferentsist mittesõltuv (Housworth *et al.*, 2003). Topeltrekombinantsed ristsiirded on uurimistöödest välja jäetud kui ka statistiliselt mittevajalik informatsioon geeni-kaardistamise tööde jaoks.

Praeguseks ei ole suudetud inimese genoomis veenvalt tõestada rekombinatsioonipunktide seotust ühegi konsensusjärjestusega. On leitud kõrge GC% ja rekombinatsiooni *hotspot*-ide seos pärmil *Saccharomyces cerevisiae* (Gerton *et al.*, 2000) ja Alu korduste seos inimesel (Dawson *et al.*, 2002).



Joonis 5. Holliday mudel ahelate vahetuseks rekombinatsioonis. Iga sirge kujutab ühte kromosoomi ahelat. Rekombinatsioonipunktis toimub kromosoomide läbilõikamine kahest kohast kas siis ühel ahelal või erinevatel ahelatel.

1.2.3 Rekombinatsioonis osalevad valgud

Väga olulist rolli DNA-le ligipääsemisele määravaid valke – histoonid (täpsemalt H3 ja H4), on põhiliselt uuritud atsetüleerimise vallas kui transkriptsiooni aktivatsiooni mõjutajad (Struhl, 1998). Transkriptsiooni aktivatsiooniga on seotud veel H3 metüleerimine (Rea *et al.*, 2000), H2A fosforüleerimine (Mahadevan *et al.*, 1991) ja H2B ubikuitineerimine (Robzyk *et al.*, 2000).

Inimese meioosis osalevate praeguseks teadaolevate valkude arv on 16: TRF1-telomeeri klasterdamine; SPO11 – endonukleas, mis produtseerib kaheaahelalisi katkestusi (DSB); ATM, ATR, p53 – DNA vigastuse kontroll; RAD50 – kaheaahelalise katkestuse parandus; RAD51 – homoloogne paardumine, ahela vahetus; RAD52 –

seostub üheaahelalise DNA-ga ja RAD51-ga; BRCA1, BRCA2, DMC1 – seondub RAD51-ga, DSB parandus; MRE11, NBS1 – seostub RAD50-ga, DSB parandus; MLH1 – mismatchi parandus; MSH4 – meioosispetsiifiline DNA parandus; DNA polümeraas β – täidab lühikesed “augud” (Miller et al., 2001). Loetletud valkudest on kõige olulisem RAD51, mis on pärmi RecA analoog. Kuigi komplementaarsete ahelate teineteisele lähenemine sünapsi jooksul on ebaselge, on teda, et RAD51 seostub üheaahelalise DNA-ga ja eraldub, kui komplementaarsed ahelad on teineteise leidnud (Ashley *et al.*, 1995).

Lisaks meiootilisele rekombinatsioonile on olemas mitootiline rekombinatsioon. Tavalistes rakkudes esineb mitootilist rekombinatsiooni 1-10 rakul 100000 –st (Gupta *et al.* 1997), kuid autosoomse retsessiivse haiguse, Bloomi sündroomi, puhul on mitootilise rekombinatsiooni sagedus lümfotsüütides 10-100 korda suurem. Bloomi sündroomi puhul toimub mitootiliste kiasmide tekkimine ka mittehomoloogsete kromosoomide vahel, põhjustades tihti kasvaja.

1.3 Ristsiirde jaotuvus genoomis ja geneetilised kaardid

1.3.1 Polümorfised markerid inimese genoomis

Inimese genoomis on hulgaliselt asuvaid varieeruvaid järjestusi, mida kasutatakse polümorfsete markeritena. Bialleelseid markereid nimetatakse +/- või 0/1 markeriteks. Rohkemate alleelivariantidega markerid on informatiivsemad kui bialleelsed markerid, kuna homosügootsuse tõenäosus langeb markeri alleelide arvu tõustes (Varilo *et al.*, 2003). Markeriks võib olla restiktaasi saidi esinemine ja mitteesinemine, mini- ja mikrosatelliitsete korduste erinev arv, SNP, Alu element või mõni muu DNA järjestus. Restiktsioonisaite (tegelikult SNP restiktsioonisaidis) kasutatakse restiktsioonifragmentide pikkuse polümorfismide (RFLP) juures. Minisatelliitjärjestused e. VNTR-d (variable number of tandem repeats) koosnevad 6-64 bp pikkustest tandeemsetest kordustest, on kõrge alleelide arvuga, kuid paiknevad genoomis valdavalt telomeerides.. Väga laialt kasutatud markerid on mikrosatelliitjärjestused e STR (short tandem repeat), mis koosnevad 2-6 bp motiividest. Kõige levinumad dinukleotiidsed

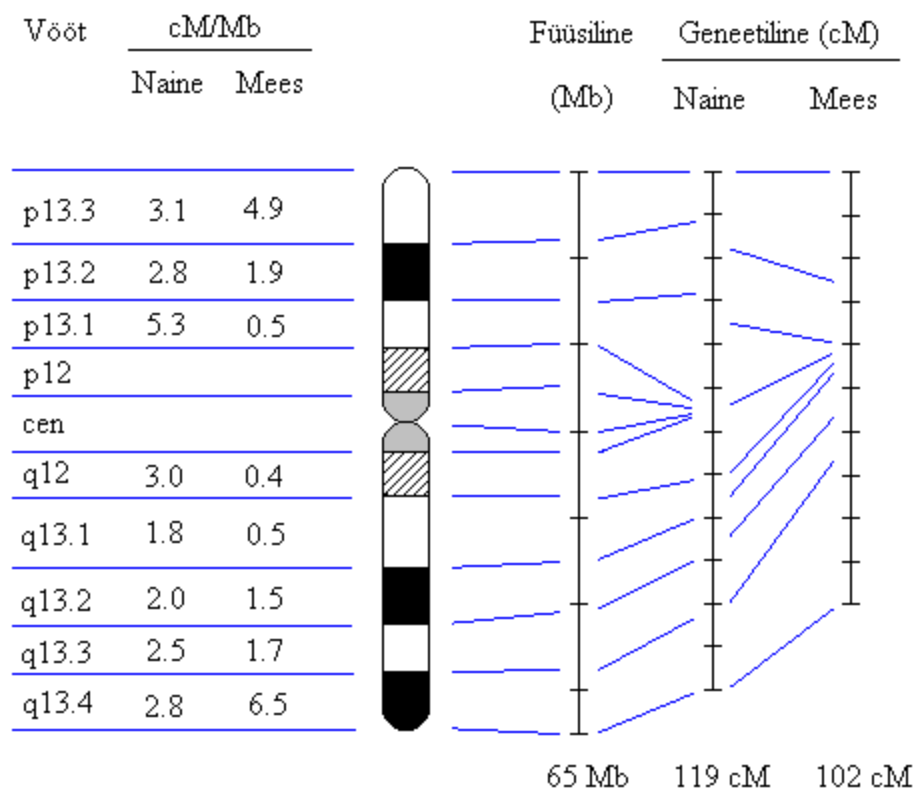
tandeemsed kordused on AC ja AT, vastavalt 50 ja 35%. Kõige levinumad trinukleotiidsed kordused on AAT ja AAC, vastavalt 33 ja 21% (Lander *et al.*, 2001).

Viimase aja tööd on siiski rohkem spetsialiseerunud bialleelsete SNP-de kasutamisele kui kõige tihedamalt inimese genoomi katvatele markeritele. SNP-d võivad olla ka tri- ja tetra-alleelsed, kuid neid esineb väga harva. SNP-ks loetakse ühealuselist DNA järjestuse muutust (v.a. insertioon/deletsioon), mis esineb populatsioonis minoorse alleeliga vähemalt 1%. Kõigist polümorfsetest markeritest moodustavad SNP-d 90% (Collins *et al.*, 1997). SNP-d tekivad sugurakkude küpsemise käigus juhuslikult iga 10^9 - 10^{11} nukleotiidi lisamisel DNA polümeraasi poolt (Cooper *et al.*, 1995). Seega tekib üks SNP 1-15 inimese kohta. SNP-ks kinnistumine sõltub juba konkreetse indiviidi järglaste (järglaste jne.) saamise võimalustest. SNP-de arvuks on ennustatud 3-11 miljonit, mis teeb keskmiseks sageduseks üks SNP 300-1000 bp kohta.

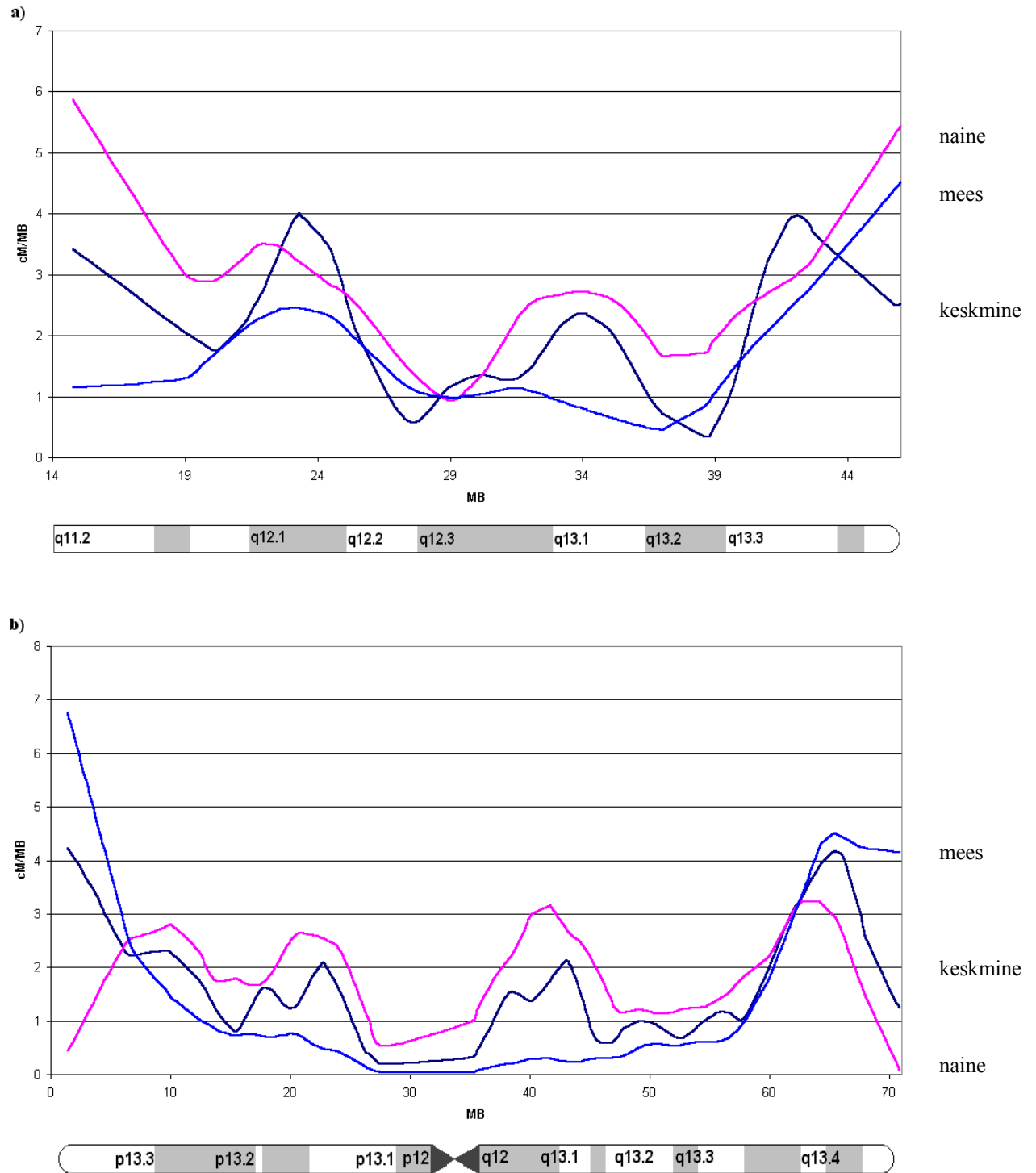
1.3.2 Inimese genoomi kaardid

Viimase dekaadi jooksul on lokaliseeritud väga palju inimese geene. Inimese Genoomi Projekti edukaks kulgemiseks aitas oluliselt kaasa erinevate füüsiliste ja geneetiliste kaartide kasutamine.

Geneetilise kaardi aluseks on kahe markeri vaheline geneetiline kaugus, mis on defineeritud kui keskmiselt oodatud ristsiirete arv ühel kromosoomil ja seda vahemaad mõõdetakse Morganites ($1\text{cM} = 1\%$ rekombinatsioone põlvkonnas, Joonis 6). Rekombinatsiooni aste näitab ristsiirete arvu 1 miljoni nukleotiidi kohta (cM/Mb). Mida kõrgem on rekombinatsiooniaste, seda intensiivsemalt antud piirkonnas rekombinatsioon toimub (Joonis 7). Otsese geneetilise kaardi aluseks on visuaalselt üleloetud kiasmid ja nende lokaliseerimine. Kaudselt saadud geneetiliste kaartide hulka kuuluvad molekulaarbioloogiliste meetoditega saadud tulemused. Rekombinatsioonipunkti on genotüpiseerides võimalik leida kolme põlvkonna või paljulapseliste perekondade andmete põhjal.



Joonis 6. Füüsilise ja geneetilise kaardi omavaheline võrdlus kromosoomis 19. (modifitseeritud joonis 11-3, Strachan *et al.*, 1999). Sinine joon tähendab ühe markeri asukohta. Geneetilise kaardi puhul määrab markeri asukoha ristsirde aktiivsus. Naise geneetiline kaart on mehe omast pikem 25% aktiivsema rekombinatsioonilise aktiivsuse tõttu. Mehe suurem rekombinatsiooniline aktiivsus on koondunud kromosoomide telomeersetesse osadesse, naisel tsentromeeri ümber.



Joonis 7. *Island Map* mehe, naise ja keskmiste rekombinatsiooniastmete graafikud inimese 22. kromosoomi (a) ja 19. kromosoomi (b) kohta. Rekombinatsiooni aste näitab ristsiirete arvu 1 miljoni nukleotiidi kohta, cM/Mb.

Olulisimad geneetilised kaardid on *Genethon Map* (Dib *et al.*, 1996), *Marshfield Map* (Broman *et al.*, 1998) ja *Island Map* (Kong *et al.*, 2002). Kõigi kaartide puhul on kasutatud mikrosatelliitseid markereid. Kõigi kaartide pikkused on küllalt lähedased, 3500-3700 cM. *Genethon*-i ja *Marshfield*-i kaartide loomiseks on kasutatud CEPH Euroopa päritolu USA perekondade DNA-d. *Centre d'Etude du Polymorphisme Humain* on teaduslik labor, mis esmaselt kogus 61 perekonna proovid läbi kolme põlvkonna, <http://www.cephb.fr/>.

Tabel 3. Olulisemate geneetiliste kaartide andmed.

	<i>Genethon Map</i>	<i>Marshfield Map</i>	<i>Island Map</i>
Teostus aasta	1996	1998	2002
Markeri tüüp	(CA) _n	(CA) _n	(CA) _n
Markerite arv	5264	8031	5136
Indiviidide arv	134	8 perekonda	869
Meiooside arv	186	188	1257
Pikkus (cM)	3699	3488	3614
DNA päritolu	CEPH Utah	CEPH Utah	Island

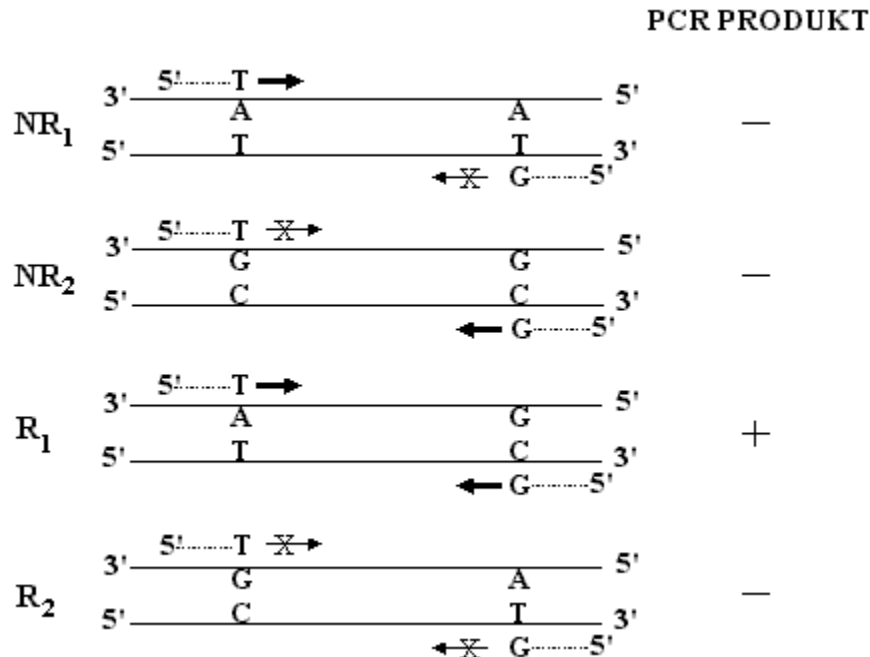
Geneetiline kaart on soospetsiifiline, indiviidispetsiifiline ja isegi meioosispetsiifiline, kuna rekombinatsiooniastmete erinevus on tõestatud nii sugude vahel (Mohrenweiser *et al.*, 1998), samast soost indiviidide vahel (Yu *et al.*, 1996) kui ka ühe vanema laste vahel (Kong *et al.*, 2002). Üldjuhul on naistel kõigis kromosoomides rekombinatsioon aktiivsem tsentromeeri ümber, mehel telomeeride lähedal (Kong *et al.*, 2002). Indiviidide ja meiooside arvu koha pealt on *Island Map* kõige informatiivsem (Tabel 3). Kõigi mainitud kaartide suureks puuduseks on markerite arvu vähesus, kuna kasutati mikrosatelliitseid markereid, mis asuvad genoomis keskmiselt iga 0,5 Mb tagant. Samas on väga tähtis markerite polümorfsus. Kuigi mikrosatelliitseid markereid loetakse väga polümorfseteks, võivad monomorfset markerid kaardi täpsust e. informatiivsust veelgi alla viia. Pärast inimese genoomi sekveneerimist omab enam tähtsust markerite tihedus.

1.3.3 Otsesed meioosi uuringud

Otseseid meioosi uuringuid on võimalik läbi viia spermatoosoidide peal kas siis ühe spermatoosoidiga (*single-sperm typing*) või totaalsperma peal (*sperm typing*).

Ühe spermatoosoidiga meetodi juures eraldatakse kõigepealt iga üksik gameet käsitsi (Li *et al.*, 1988, Lien *et al.*, 1993) või automaatselt rakusortoriga (Cui *et al.*, 1989). Ühe gameedi genotüüpiseerimise juures saadakse väga täpset informatsiooni mehe meiootilise rekombinatsiooni kohta, mida perekonna andmetega ei ole võimalik saada. Nendeks on geenitriiv, kromosomaalsete anomaaliade efektid rekombinatsioonis (Shi *et al.*, 2001), ristsiirete sageduste erinevused eri indiviididel (Yu *et al.*, 1996; Lien *et al.*, 2000; Cullen *et al.*, 2002) ja vanuse efektid rekombinatsioonile (Shi *et al.* 2002). Ühe spermatoosoidi genotüüpiseerimise juures lokaliseeritakse rekombinatsioonipunkt lõigu poolitamise meetodil, kus võrreldakse alleelspetsiifiliste praimeritega rekombinandi olemasolu või puudumist. Seni on mahukaim ühe spermatoosoidi genotüüpiseerimine tehtud 3,3 Mb alas MHC regioonis 20000 spermatoosoidile 12-l indiviidil (Cullen *et al.*, 2002). Antud töö juures kasutati 48 polümorfset mikrosatelliitset markerit fikseerimaks kolmekümnet vahemikku, kus ristsiire toimus. Nendes vahemikes oli kokku 325 rekombinanti. Kolmekümnest vahemikust 6-s võib olla rekombinatsiooni *hotspot*.

Totaalse spermi analüüsi juures kasutatakse alleelspetsiifilisi PCR praimereid, millega toimub selektiivne rekombinantsete molekulide amplifikatsioon. Rekombinantset PCR produktid on detekteeritavad, kui neid amplifitseeritakse (joonis 8). Selle meetodi juures on suurim raam rekombinatsioonipunkti leidmiseks 5 kb, mistõttu on uuritud juba tuntud aktiivseid rekombinatsioonipunkte MHC regioonis (Kauppi *et al.* 2003, Jeffreys *et al.* 2000, Jeffreys *et al.* 2001, Jeffreys *et al.* 2002), DMD geenis (Oudet *et al.* 1992), fosfoglükomutaasi (PGM1) geenis (Yip *et al.* 1999), Inimese lipoproteiin lipaasi geenis (Templeton *et al.* 2000), SHOX geenis (May *et al.* 2002). Hiire MHC $E\beta$ geenis on lokaliseeritud üks rekombinatsiooni *hotspot*. 6,5 miljoni spermi analüüsi käigus hiirel summeeriti 516 rekombinatsioonijuhtu 1,0-1,6 kb regioonis (Yauk *et al.*, 2003). Inimese MHC klass II regioonis on DNA ja TAP2 geenide vahel kuus rekombinatsiooni *hotspot*-i, milledes detekteeriti kokku 472 ristsiiret 7,6 miljoni spermatoosoidi kohta (Jeffreys *et al.*, 2000; Jeffreys *et al.*, 2001).



Joonis 8. Selektiivne rekombinantsete molekulide amplifikatsioon. Kahe alleelspetsiifilise praimeriga lokaliseeritakse piirkond, kus ristsiire toimunud on, antud joonisel R₁ (TC). Enne on veel kahe alleelspetsiifilise praimeriga määratud vanema markerite haplotüüp, antud joonisel NR₁ ja NR₂ (5'-3' suunas TT ja CC).

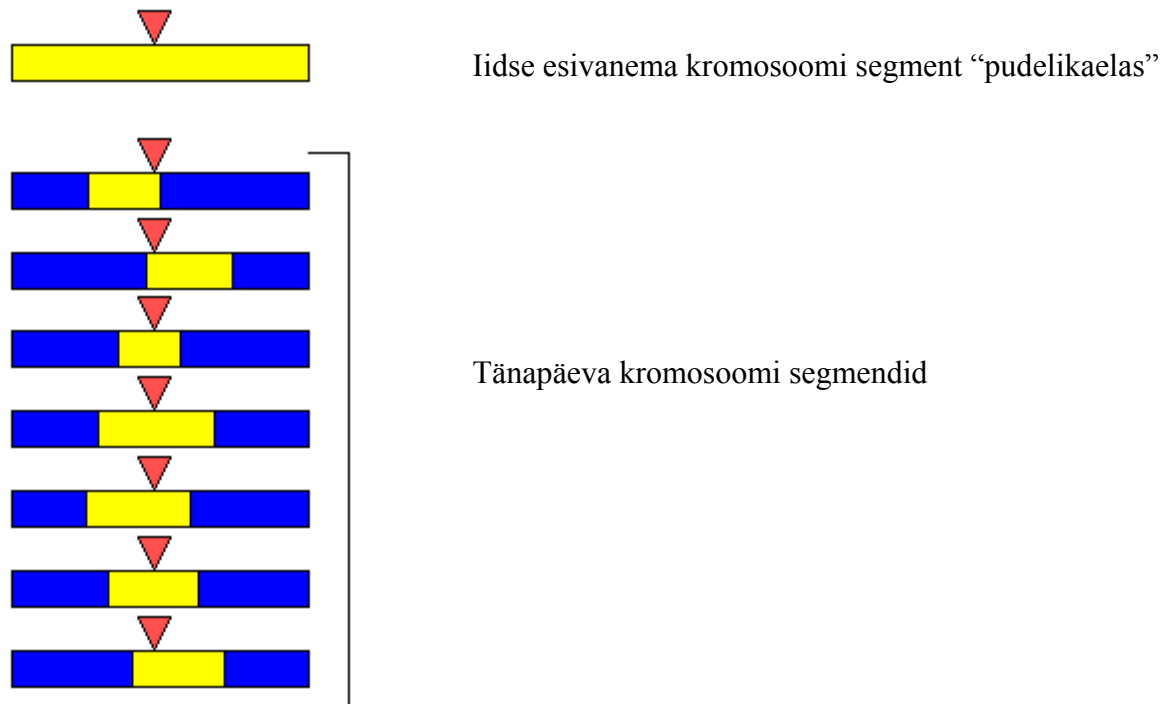
1.4 Haplotüübi blokid inimese genoomis

1.4.1 *Linkage disequilibrium* ehk lookuste alleelne aheldatus

Geneetiliste markerite omavahelist seotust mõõdetakse tasakaalustamata aheldatusega e. LD-ga. Kaks olulisimat LD mõõdikut on:

1. $|D'|$ (Lewontin, 1964),
2. r^2 (mõnikord tähistatakse ka Δ^2).

Mõlemate mõõdikute väärtused on 0 (ei ole LD-d) ja 1 (täielik LD) vahel, kuid nende sisu on väga erinev. $|D'| = 1$, kui kahe bialleelse markeri haplotüübi variantidest on esindatud kaks või kolm. $|D'| < 1$, kui on esindatud kõik neli haplotüüpi. Kui $|D'| < 1$, siis on kahe markeri vahel toimunud ristsiirde juht. r^2 näitab statistilist korrelatsiooni kahe markeri vahel ja on võrdne 1-ga vaid kahe haplotüübi variandi esinemisel. r^2 on oluliselt tundlikum ja seetõttu on kahe markerivahelise assotsiatsiooni leidmiseks vaja proovide hulka suurendada $1/r^2$ korda.



Joonis 9. LD iidse esivanema mutatsiooni ümber. Sinise alaga on tähistatud rekombinatsioonis vahetunud kromosoomi lõike. Kollase ala markerid on omavahel tugevas LD-s (Ardlie *et al.*, 2002).

Peale põhiliste LD määra mõjutajate, rekombinatsiooni ja mutatsioonide (Joonis 9), on veel populatsiooni demograafilisest ajaloost põhjustatud (1-4) ja lookusega toimuvate protsessidega seotud (5-8) faktoreid:

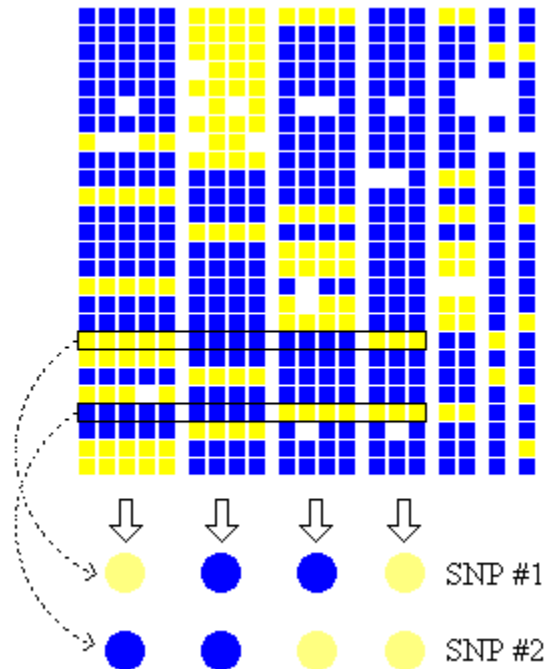
1. Geenitriiv – selle sündmuse käigus muutub lookuse sagedus populatsioonis iga põlvkonna vältel. Sagedamini juhtub väikeses populatsioonis.
2. Populatsiooni kasv – kiire populatsiooni kasv vähendab LD kahandades geenitriivi.
3. Populatsioonide segunemine või migratsioon – alguses on LD proportsionaalne alleeli sageduste erinevusega populatsioonide vahel ja see on sõltumatu markerite vahelisest kaugusest. Järgmiste põlvkondade jooksul kaob kiirest “tehislik” LD aheldamata markerite vahel, samal ajal kaob LD naabermarkerite vahel palju aeglasemalt.
4. Populatsiooni struktuur – inimese LD mustri kujunemisel peaksid olulised olema alampopulatsioonid. Ekstreemne inbriiding toodab kõrget LD määra, kuid ei muuda variatsiooni taset taimede mudelorganismi, *Arabidopsis thaliana*-ga tehtud tööde põhjal (Nordborg *et al.*, 2002).

5. Looduslik valik – eelistatuim haplotüüp võib kiiresti tõusta domineerivaks või koguni fikseeruda. Teise variandina võib ühe kromosoomi mitme lookuse kombinatsioon olla loodusliku valiku all ja mõjutada LD ulatust veelgi laiemalt.
6. Erinev rekombinatsiooni kiirus – toimub nagu *hotspot*-i teoorias (vt. allpool).
7. Erinev mutatsiooni kiirus – mõned SNP-d CpG saarte lähedal on kõrgema mutatsiooni kiirusega, mistõttu näitavad väikest LD-d ka olematu ajaloolise rekombinatsiooni olemasolul.
8. Geeni konversioon - selline juht on võrdne kahe väga lähestikku oleva mittehomoloogse ristsiirde juhuga.

1.4.2 Haplotüübid ja -blokid

Haplotüübi moodustavad vähemalt kaks ühel kromosoomi ahelal asuvat markeri varianti. Nende kahe markeri teiste alleelide kombinatsioonid moodustavad uued haplotüübid. Markerite arvu ja variantide kasvades võimalike haplotüüpide arv suureneb. Kromosoomil füüsiliselt järjestikku asuvad markerid moodustavad haplobloki, kui markerite vahel on tugev LD. Haploblokke on püütud defineerida kahte moodi:

1. “Finantsblokk” e. iga blokk moodustub minimaalse arvu markerite abil, mis katab 80% olemasolevatest haploblokkide koguarvust (Patil *et al.* 2001, joonis 10).
2. “Rekombinatsiooniline haploblokk” e. haplotüübi blokid on seotud rekombinatsiooni *hotspot*-idega (Gabriel *et al.*, 2002; Phillips *et al.* 2003). *Hotspot* on 1-3 kb pikkune kromosoomi lõik, kus toimub eriti palju ristsiirdeid, mistõttu markerite vahel puudub LD ning selles piirkonnas on bloki pikkuseks vaid 1 marker. Kahe *hotspot*-i vaheline ala on tugevas LD-s ($D' > 0,8$) ja see moodustabki ühe bloki.



Joonis 10. Haplotüübiblokk Patil'i järgi (Patil *et al.* 2001). Iga ruut tähistab markeri väärtust (sinine- sagedasem alleel, kollane- harvem alleel), iga tulp eraldi kromosoomi, iga rida markerit. Markerite väärtuste klasterdamisel saadakse küllalt korrapärane struktuur. Kahe markeri kombinatsioon kirjeldab neli haploblokki ja samas 80% haploblokkide koguarvust.

Esimene definitsioon lähtub finantsilisest vajadusest minimiseerida SNP-de arvu laiaulatuslike genoomiuuringute tegemiseks, selle meetodiga detekteeriti 21. kromosoomis 4135 blokki (Patil *et al.* 2001, joonis 10).

Jeffreys'e järgi on *hotspot*-i pikkuseks 1-3 kb, kusjuures ristsiirded ei ole jaotunud normaaljaotusega ümber *hotspot*-i keskpunkti. Normaaljaotust meenutav kõver on ühest otsast välja venitatud vastavalt markerite (SNP) muustrile (Jeffreys *et al.* 2002). Tõenäoliselt mängib siin rolli veel kiasmi liikumissuund. Ühel ja samal indiviidil on erinevad *hotspot*-id erineva "võimsusega", mistõttu mõni neist on ristsiirde juures sagedamini kasutamisel (Jeffreys *et al.* 2001).

Hiire rekombinatsioonis klasterduvad ristsiirded samuti *hotspot*-idesse. Hiljuti kirjeldati 8 *hotspot*-i hiire MHC E_β geeni alas. Hiire-uuringud on näidanud, et *hotspot*-ide kuju on samuti sõltuv MHC haplotüübist (Yauk *et al.* 2003) nagu ka inimesel (Jeffreys *et al.*, 2002).

1.4.3 Teooriaid haploblokkide tekkest

Asutaja efekt ja pudelikael inimese väljumisel Aafrikast

Out of Africa teooria kohaselt lahkus 50000-100000 aasta eest 10-50 tuhandeline inimeste grupp Aafrikast (Rogers *et al.* 1992). Teooriat tõendavad faktid, et aafriklaste mtDNA geneetiline variaablus on inimeste populatsioonidest suurim (Cann *et al.* 1987). Aafrika kasuks räägib ka tõsiasi, et tänapäevase *Homo sapiensi* lähimad sugulased šimpansid elavad samuti Aafrikas. Kahe šimpansi liigi geneetiline varieeruvus on suurem kui Euroopa, Aasia ja Aafrika inimpopulatsioonidel. Uurides geneetilist distantssi Aafrika ja teiste inimpopulatsioonide vahel, eeldatakse *Homo sapiensi* kujunemist 100000 kuni 400000 aastat tagasi. Kuigi *Out of Africa* teooria on püstitatud lähtudes mtDNA töödest, võib sarnast lähenemist rakendada genoomse DNA lõikude (iga haploblokk eraldi) kohta. Siin moodustavad pudelikaela Aafrikast väljarännanud inimesed, kelle asutaja efekti tõttu tugevas alleelses aheldatuses olevad haplotüübi mustrid ei ole jõutud lõplikult ära lõhkuda. Selle teooria poolt räägib fakt, et aafriklastel on LD ulatus poole lühem, kui eurooplastel (Gabriel *et al.*, 2002).

Rekombinatsioonil on põhiline roll LD mustri kujundamisel

Inimese genoomist on kõige uuritum osa MHC klass II regioon. Nimetatud regioon sisaldab kolme tuntud rekombinatsiooni *hotspot*-i ja ühte 60 kb pikkust LD blokki. Kuigi markerite sagedused olid inglastel, saamidel ja zimbabwelastel erinevad, oli neil kõigil olemas nimetatud 60 kb LD blokk. Selline haplotüüpide uurimine on tõestanud, et LD mustri määravad ära rekombinatsiooni *hotspot*-id, mitte populatsiooni ajalugu (Kauppi *et al.* 2003).

Hotspot-i teooria kohaselt on enamus kromosoomipiirkondi juba ära lõhutud ja LD muster vaid kujutab lõhkumise vahemaid. Mõnes genoomipiirkonnas on haploblokid (ehk siis ka LD blokid) pikemad, teistes piirkondades lühemad. Edaspidised rekombinatsioonid LD mustrit ei muuda, kuna rekombinatsioonipunktid satuvad jällegi vaid haploblokkide vahel asuvasse *hotspot*-ide piirkondadesse, kus iga blokk on tavaliselt 1 marker. Kuna MHC regioon on väga väike lõik inimese kromosoomidest, siis ei ole teada, kus teistes regioonides sama mehhanismi kasutada võib (Zhang *et al.*, 2003).

Juhuslikult toimuvad ristsiirded võivad jaotada kromosoomilõigud blokkideks

Nimetatud väide baseerub arvutisimulatsioonidel, kus olulisteks arvutuste parameetriteks on SNP-de arv, SNP-de alleelisagedus, “kromosoomilõigu” pikkus, populatsiooni suurus ja kasv, generatsioonide arv.

Zhang'i arvutisimulatsioonis jälgiti LD blokkide kujunemist läbi põlvkondade juhuslikult jaotuva ja MHC regiooni rekombinatsiooni spetsiifiliselt jaotuva ristsiirete toimumise järgi. Mudelis, kus kasutati 100 võrdsetel vahemaadel asuvat SNP-d (alleelisagedusega 0,5), jälgiti LD blokkide struktuuri 10 000 indiviidiga populatsioonis 2000 põlvkonna jooksul iga 125 põlvkonna järel. Uusi SNP-sid juurde ei tekkinud. Kuigi iga simulatsiooniga saadi juhuslikult jaotuva ristsiiretega erinev tulemus, on blokiline struktuur nähtav kuigi lühemate blokkidena, kui MHC spetsiifiliselt jaotaval ristsiiretel (Zhang *et al.*, 2003).

Teises mudelis lubati uusi SNP-sid sagedusega 10^{-8} ja oli üks pudelikael kümnetuhandendas põlvkonnas 400 põlvkonna jooksul, kuni populatsioon kasvas jälle 10000 indiviidini. Samuti valiti pudelikaelast väljujateks 1000 indiviidi, kellel kõigil olid 10 SNP-d täielikult aheldunud. LD blokke vaadeldi iga 1000 põlvkonna tagant enne pudelikaela ja iga 125 põlvkonna järel pärast pudelikaela. Teine mudel andis sama tulemuse, kuid blokkide vahel olevad alad on pikemad (Zhang *et al.*, 2003).

Andersoni ja Slatkini (Anderson & Slatkin, 2004) simulatsioonid näitasid, et 5q31 regioonis Daly *et al.* poolt identifitseeritud SNP mustrid ei ole tekkinud mitte ühtlase rekombinatsiooni jaotuvusega jääva suurusega populatsioonis, vaid on tekkinud kas *hotspot*-ide olemasolul jääva suurusega populatsioonis või *hotspot*-ide puudumisel kiirekasvulises populatsioonis (Anderson & Slatkin, 2004). Nende teooria järgi erineb lokaalne rekombinatsiooni aste populatsiooni kasvu ja *hotspot*-ide tõttu. Oma töös simuleerisid nad 516 kromosoomi “makesamples” programmiga (Hudson 2002), lubades rekombinatsioone 500 erinevas kohas neljal erineval juhul: a) ühtlane rekombinatsioon jääva suurusega populatsioonis, b) *hotspot*-ide lubamine jääva suurusega populatsioonis, c) pudelikaelata ja *hotspot*-ideta kasvava suurusega populatsioonis ning d) pudelikaelaga ja *hotspot*-ideta kasvava suurusega populatsioonis (Anderson *et al.*, 2004).

2 Töö eesmärgid

Töö hüpoteesid:

- a. Rekombinatsioonil on põhiline roll LD mustri kujundamisel.
- b. Rekombinatsioonipunktid ei ole inimese genoomis ühtlaselt jaotunud.
- c. Rekombinatsioonipunktid assotsieeruvad teatud genoomsete struktuurielementidega.

Käesolevas töös:

1. Kasutades tervet kromosoom 19 ja 22 hõlmavat genotüpiseerimise andmestikku võrdlen LD mustreid Eesti populatsiooni, Saksa triode (isa, ema, laps) ja Euroopa päritolu USA perekondade (CEPH perekonnad) vahel ning kõrvutan meiootilise ristsiirde jaotusega.
2. Võrdlen rekombinatsioonipunktide jaotuvust kromosoomis 19 ja 22 CEPH perekondade materjalil.
3. Uurin rekombinatsiooni *hotspot*-e lähtuvalt hüpoteesist, et *hotspot*-id assotseeruvad tuntud genoomsete struktuurielementidega nagu T-vöödid, aktiivse transkribeerimise ala, kõrge GC% ja Alu korduste rikkad alad.

3 Materjal ja metoodika

3.1 Kromosoom 22 LD kaardi koostamine populatsiooni genotüpiseerimise andmetega*

3.1.1 Kromosoom 22 chip *in silico*

Selleks, et teha valik markeritest terves kromosoomis 22, saadi koostööpartneritelt, *The Wellcome Trust Sanger Institute*-lt (<http://www.sanger.ac.uk>) tol hetkel veel publitseerimata, 1638 SNP ja INS/DEL andmed (<http://www.biodata.ee/~tarmo/magistritoo/>). Andmed sisaldasid SNP-dega ja INS/DEL-dega piirnevaid järjestusi, markeri tüüpi ja variante ning markeri asukoha q ðla algusest alates. Markeri ees ja järel olevaid 25 aluspaari kontrolliti 3' otsas sekundaarstruktuuride tekkimise võimaluse suhtes. APEX metoodika spetsiifilisuse tõttu langes välja 266 markerit.

In silico sobivatele markeritele disainiti kaks APEX praimerit, üks sense ja teine antisense ahelale vastavalt. PCR-praimerite disain tehti 300 markeri osas käsitsi programmiga GCG ja ülejäänute osas kasutati automaatset PCR-praimerite disainimise programmi (Andreson, 2002). PCR ja APEX praimerid (<http://www.biodata.ee/~tarmo/magistritoo/>) disainiti 2001.a. annotatsiooni põhjal.

Edasisest töötlustest langesid välja veel 93 markerit PCR-praimerite ebaõnnestunud disaini ja APEX praimerite genoomis unikaalsusele mittevastavuse tõttu.

3.1.2 Genotüpiseerimine

Genotüpiseerimise platform - APEX

Meie laboris kasutati genotüpiseerimise platvormina APEX meetodit, mis baseerub komplementaarsete DNA fragmentide hübridisatsioonil ja DNA polümeraasi poolt läbiviidud primerechtsensioonil (Shumaker *et al.*, 1996).

* Osa käesolevast tööst on läbiviidud grupitööna, mille kõigis etappides töö autor osales (Dawson *et al.*, 2002).

Kõik 25-meersed oligonukleotiidid (MWG-Biotech AG) olid immobiliseeritud 5' otsast C12 aminolinkeri kaudu aminosilaaniga töödeldud klaasi pinnale (Guo *et al.*, 1994). Oligonukleotiidide kandmiseks klaasile kasutati Affymetrix 417 robotit (Affymetrix Inc.). Oligonukleotiidide seostumist kontrolliti värvides klaasi SYBRTM Gold'iga (Molecular Probes Europe BV).

APEX reaktsioonid viidi läbi vastavalt Tõnissoni jt poolt välja töötatud protokollile (Tõnisson *et al.*, 2000). Florestseeruvate signaalide detekteerimiseks kasutati GenoramaTM QuattroImageri ja analüüsimiseks GenoramaTM arvutiprogrammi (Asper Biotech AS). Analüüsimisel hinnati signaalide intensiivsust erinevates kanalites.

DNA proovid

Kokku genotüpiseeriti 265 indiviidi, 101 vabatahtlikku indiviidi Eesti populatsioonist, 132 triiot Saksa populatsioonist (isa-ema-laps) ja 32 indiviidi CEPH perekondadest (6 perekonda), kokku 265 indiviidi.

PCR

Kõik PCR reaktsioonid viidi läbi omaette reaktsioonikambris 384 plaadiformaadis Primus Multiblock PCR-System masinaga (MWG Biotech AG). PCR reaktsioonisegu maht oli 15 µl: 1,5 µl 10x Taq DNA polümeraasi puhvrit (750 mM Tris-HCl (pH 8.8), 200 mM (NH₄)₂SO₄, 0,1 % Tween 20, MBI Fermentas), 25 mM MgCl₂, 1,2 µl 2.5 mM dNTP (25% dATP, 25% dCTP, 25% dGTP, 20% dTTP, 5% dUTP) (MBI Fermentas), 1 U Taq DNA polümeraas (OÜ Naxo), 2 µl 5 µM praimeripaari stocklahust, 9,5 µl milliQ H₂O.

1045 G+C rikastele praimeripaarile kasutati PCR programmi protokoll MODHINX1 ja ülejäänud 234 madalama sulamistemperatuuriga praimeripaarile HINX2 (Tabel 4).

Tabel 4. PCR programmid MODHINX1 ja HINX2

	MODHINX1		HINX2	
Esmane denaturatsioon	95 °C	5.00 min	95 °C	5.00 min
I tsükli denaturatsioon	95 °C	0.20 sek	95 °C	0.20 sek
Praimerite seondumine	68 °C	0.25 sek	58 °C	0.25 sek
	-1 °C	tsükli kohta	-0.5 °C	tsükli kohta
Ekstensioon	72 °C	0.30 sek	72 °C	0.30 sek
I tsükkel	8 korda		15 korda	
II tsükli denaturatsioon	95 °C	0.20 sek	95 °C	0.20 sek
Praimerite seondumine	60 °C	0.25 sek	50 °C	0.25 sek
Ekstensioon	72 °C	0.30 sek	72 °C	0.30 sek
II tsükkel	12 korda	15 korda		
III tsükli denaturatsioon 95 °C	0.20 sek			
Praimerite seondumine	60 °C	0.25 sek		
	-1 °C	tsükli kohta		
Ekstensioon	72 °C	0.30 sek		
III tsükkel	4 korda			
IV tsükli denaturatsioon	95 °C	0.20 sek		
Praimerite seondumine	56 °C	0.25 sek		
Ekstensioon	72 °C	0.30 sek		
IV tsükkel	12 korda			
Inkubatsioon	72 °C	5.00 min	72 °C	5.00 min
Maha jahutamine	4 °C	Püsivalt	4 °C	Püsivalt

Fragmenteerimine

Amplifitseeritud produktid segati kokku, puhastati etanooli (EtOH) ja ammooniumatsetaadi (NH₄OAc) sadestamisega ja fragmenteeriti hübriidsatsiooni efektiivsuse tõstmiseks. Fragmenteerimine toimus uratsiil-N-glükosülaasiga (UNG, Epicentre Technologies) ning töödeldi kreveti aluselise fosfataasiga (sAP, Amersham Biosciences Inc.). UNG ensüümi toimel hüdrolyüsib uratsüülnukleotiidi ja deoksüriboosi vaheline N-glükosiid side. Fragmentatsiooni reaktsioonisegu maht oli 55,5 µl: 45 µl puhastatud PCR produkte, 2,5 µl UNG (1U/µl) ja sAP (1U/µl) ning 5,5 µl 10x Epicentre UNG puhvrit (50 mM Tris-HCl (pH 9.0), 20 mM (NH₄)₂SO₄). Fragmentatsiooniks inkubeeriti reaktsioonisegu 1 h 37 °C juures. Fragmentatsiooni edukuse kontrollimiseks kanti 1% agarosgeelil kõrvuti radadele 3 µl produkti, mida oli hoitud 10 min 95 °C juures ja 3 µl kuumutamata produkti.

3.1.3 LD mustrite analüüs SNP Assistant programmiga

SNP assistant (<http://www.biodata.ee>) on uus geneetiliste andmete analüüsi programm, mida käesoleva töö juures kasutati LD *pairwise* arvutusteks ja visuaalseks

presenteerimiseks. LD arvutusi tehti d^2 baasil (mõõdab sageduste erinevusi, (Collins et al. 2001).

$$d^2 = D/(p_{11}+p_{21})(p_{12}+p_{22})$$

D määrab kahe lookuse vahelise assotsiatsiooni kõrvalekallet. $D = p_{ij} - p_i p_j$, kus p_{ij} on kahe alleeli i ja j esinemissagedused ning p_i ja p_j erinevate alleelide esinemissagedused (Lewontin 1988).

Markerite juures fikseeriti signaali esinemissagedust $> 65\%$, Hardy-Weinbergi tasakaalu (HWE) p -väärtust üle 0,05 ja minoorse alleeli sagedust üle 20%.

Hardy Weinberg Equilibrium valem on:

$$q^2 + 2qp + p^2 = 1,$$

kus q ja p on alleelide sagedused populatsioonis. HWE kasutatakse markeri usaldusväarsuse hindamisel. HWE seaduse kehtib kõige paremini, kui marker vastab järgmistele tingimustele: a) populatsioon on suur, b) populatsioonis toimub juhuslik paardumine, c) ei toimu uusi mutatsioone, d) populatsiooni ei ole sisenenud migrante ja e) puudub looduslik valik.

Hardy Weinbergi tasakaalu usaldust hii-ruut (p -väärtus) arvutati valemiga:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}},$$

kus oodatud $A_i B_j$ haplotüüpide arv (e_{ij}) on avaldatav valemiga:

$$e_{ij} = n \tilde{p}_{i.} \tilde{p}_{.j},$$

ühe vabadusastmega kahaalleelsete andmete juures (Weir 1979).

3.2 CEPH perekondade kromosoom 22 ja 19 rekombinatsioonipunktide kirjeldus

3.2.1 CEPH perekondade genotüüpiseerimisandmed

Kromosoom 22

Rekombinatsioonipunktide detailseks kirjeldamiseks kasutati koostööpartnerite (Dawson *et al.*, 2002) CEPH perekondade genotüüpiseerimisandmeid, kuna neil olid genotüüpiseeritud suurema arvu markeritega rohkem arv perekondi. Koostööpartnerid kasutasid genotüüpiseerimise platvormina Invader tehnoloogiat ([Third Wave Technologies](#)

Inc.). Kromosoom 22 andmed on pärit Sanger Keskuse kodulehelt <http://www.sanger.ac.uk/HGP/Chr22/LD> (Dawson *et al.*, 2002) failidest ceph.dat, ceph.ped ja ceph.map.

Antud töös vajaminevad andmed paiknevad failides ceph.map (markerite nimed ja asukohad) ning ceph.ped (indiviidide koodid, sugulussidemed ja genotüübid).

Kromosoom 19

Selleks, et võrrelda kromosoom 22 perekonna ristsiirde punktide jaotust muude genoomsete piirkondadega, kasutati kromosoom 19, samuti CEPH perekondade SNP-del baseeruvat genotüpiseeritud andmestikku. Kromosoom 19 andmed olid saadud kasutades [Affymetrix](http://www.affymetrix.com) tehnoloogiat. Kromosoom 19 genotüpiseerimise andmed on pärit Lon Cardon'i kodulehel <http://bioinformatics.well.ox.ac.uk/~lon/chr19/data/> (Phillips *et al.*, 2003) asuvast failist ceph19.tar.gz. Lahtipakkimisel tekivad failid chr19.built29.dat, chr19.built29.ped ja chr19.built29.map. Antud töös kasutatud andmed asuvad sama laiendiga (*.map ja *.ped) ja seega analoogse sisuga failides.

Mõlema kromosoomi võrdlevad andmed on Tabelis 5. Kromosoom 22-st genotüpiseeriti vaid q õlg, sest p õla moodustavad valdavalt kordusjärjestused. Markerite tihedus jääb samasse suurusjärku e. 20 000 bp juurde. Neli perekonda olid genotüpiseeritud mõlema kromosoomi puhul. Kahjuks ei õnnestunud 19. kromosoomi andmete juures teada saada laste CEPH koodi, mistõttu lapsed võivad sama perekonna juures olla erinevad.

Tabel 5. Kromosoomide ja markerite võrdlus. Kattuvad perekonnad on trükitud rasvases kirjas.

	Kromosoom 19	Kromosoom 22
Suurus (p õlg ; q õlg)	63,7 Mb (24 Mb; 30,8Mb)	49,3 Mb (14,4 Mb; 34,9 Mb)
Markereid	3297	1504
Kasutatud markereid	3269	1473
Marker bp kohta	16 800	22 600 (q õlg)
Genotüpiseeritud CEPH perekonnad	1331 , 1333, 1334, 1340, 1341, 1347 , 1362 , 1408, 1416 , 1423	13291, 1331 , 1332, 1347 , 1362 , 1413, 1416
Genotüpiseeritud indiviidid	Indiviidi kood teadmata	Indiviidi kood teada

3.2.2 Ensembl andmebaasid

Kuna kasutatud markerid kromosoomist 19 on järjestatud kromosoomide assambleerimise Ensembl v.29 käigus ja kromosoom 22 markerid on järjestatud aastast 2001, siis kontrolliti markereid viimase avalikult kasutatava genoomi annotatsiooni Ensembl v.33 kromosoomide järjestuste vastu. Markerite positsiooni andmed võeti Ensembl v.33 andmebaasist Homo_sapiens_lite_15_33 tabelist snp (ftp://ftp.ensembl.edu/pub/human-15.33/data/mysql/homo_sapiens_lite_15_33/snp.txt.table.gz).

Andmete kõrvutamise tulemusel jäeti edasisest analüüsist välja 24 chr19 markerit ja 18 chr22 markerit, kuna v.33 ei kirjeldanud neid enam markeritena 19. või 22. kromosoomis. Välja tuli jätta ka kaks 19. ja üksteist 22. kromosoomi markerit, kuna andsid kahte või rohkemasse kohta vaste. Samuti oli vajalik markerite järjekorra muutus mõne markeri osas 22. kromosoomis ja paljude markerite blokkide osas 19. kromosoomis. Liigmahukaid andmeid käesoleva töö trükitud lisades ei ole, need on saadaval käesoleva töö kodulehel <http://www.biodata.ee/~tarmo/magistritoo/> ja töö juurde kaasaskäival CD-l.

3.2.3 Rekombinatsioonipunktide lokaliseerimine CRI-MAP programmiga

CRI-MAP (<http://linkage.rockefeller.edu/soft/crimap>) programmi põhieesmärgiks on kiire, enamalt automatiseeritud multilookus ahelduvuse kaardi tegemine. Käesolevas töös kasutati vaid ühte mitmest CRI-MAP'i funktsioonidest, CHROMPIC, mis lokaliseerib kaks markerit, mille vahel on toimunud ristsiire.

Rekombinatsioonipunktide detailsem lokaliseerimine

Edasi valmistati CRI-MAP programmi jaoks sisendfailid mõlema kromosoomi andmete jaoks. Selleks kasutati MS Wordpad programmi ja kopeeriti andmeid *.map ja *.ped failidest uude faili laiendiga *.gen. Sisendfaili ülesehitus on järgmine:

Perekondade arv

Markerite arv

Markerid

...

Perekonna kood

Perekonna suurus

Indiviidi kood Indiviidi isa kood Indiviidi ema kood Indiviidi sugu

Genotüübid ...

...

Sisendfailid on saadaval käesoleva töö kodulehel

<http://www.biodata.ee/~tarmo/magistritoo/> .

Käesoleva töö jaoks oli vaja CRI-MAP programmi lubatud markerite arvu suurendada ja ümber kompileerida. Antud töö juures on kõik CRI-MAP programmi arvutused tehtud Intel Celeron 850 MHz 1GB operatsioonisüsteemiga arvutil.

CRI-MAP programmi esimese käsu täitmise jooksul koostati parameetrite fail. Markerite järjestuste muutus kanti parameetrite faili.

CHROMPIC funktsiooniga saadud tulemused olid muudetud kasutatud markerite järjekorra põhiseks ja edasi töödeldi MS Excelis.

Rekombinatsioonipunktide jaotuvuskõver ja liikuv keskmine

Rekombinatsioonipunktide jaotuvuskõvera koostamiseks kirjutati väike andmete töötlemise programm programmeerimiskeeles Perl, mis arvutas rekombinatsioonipunktide arvu 250 kb kaupa p õlg q õlg suunas. Tulemus kuvati MS Excel programmis X-Y graafikus kuhu lisati liikuva keskmise graafik viie järjestikuse punkti baasil.

Alu korduste, GC% ja kodeeriva ala statistika

Alu korduste, GC% ja kodeeriva ala numbriliste väärtuste väljakirjutamiseks kasutati samuti Perl'i programm. Alu korduste andmed saadi Ensembl v.33 andmebaasist Homo_sapiens_lite_15_33 tabelist *repeat*

(ftp://ftp.ensembl.edu/pub/human-5.33/data/mysql/homo_sapiens_lite_15_33/repeat.txt.table.gz).

Ensembli korduste tabel ei ole kasutajasõbralik, sisaldades identseid ja ülekattuvaid järjestusi algselt topelt sekveneeritud kontiigides s.t., et on piirkondi, kus kordused on topelt. Seetõttu jäeti korduste tabelist välja Alu järjestused, millede algusest ja lõpust 50 bp ulatuses oli samasisuline, kuid erineva ID koodiga Alu järjestus.

GC% arvutused tehti Ensembli v.33 annotatsiooni põhjal 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 250, 500 ja 1000 kb raamis. Kromosoomi andmed saadi täispikkadest kromosoomijärjestustest:

ftp://ftp.ensembl.edu/pub/human-15.33/data/golden_path/22.fa.gz

ftp://ftp.ensembl.edu/pub/human-15.33/data/golden_path/19.fa.gz

Transkribeeritava ala arvutustes võeti arvesse ainult Ensembli poolt aksepteeritud geenid ja arvutustes lähtuti geeni ala pikkuse suhe arvutusakna pikkusesse (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 250, 500 ja 1000 kb). Geeni ala andmed saadi Ensembl v.33 andmebaasist Homo_sapiens_lite_15_33 tabelist *gene*

(ftp://ftp.ensembl.edu/pub/human-15.33/data/mysql/homo_sapiens_lite_15_33/gene.txt.table.gz).

Alu ala, GC% ja geeniala korrelatsiooni leidmiseks rekombinatsioonilise aktiivsusega kasutati parameetrilist, lineaarset korrelatsiooni eeldavat Pearsoni testi ja mitteparameetrilist Spearmani testi statistikaprogrammis R (<http://www.r-project.org>).

Pearsoni testi on parameetriline test ja tema väärtus r on 1 ja -1 vahel ning ta näitab kahe andmemassiivi omavahelist positiivset või negatiivset lineaarset seost.

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

, kus X ja Y on andmemassiivide väärtused ning n on andmete arv kummaski massiivis.

Spearmani test on mitteparameetriline test ja teda arvutatakse järgneva valemi järgi:

$$r = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

, kus N on andmete arv kummaski massiivis (XY) ning D on andmepaaride ($X-Y$) järjekorra numbrite vahe. Alla 0,33 puhul on andmemassiivide vahel nõrk korrelatsioon, 0,34-0,66 näitab keskmist korrelatsiooni ning üle 0,66 näitab kõrget korrelatsiooni.

4 Tulemused

4.1 Inimese kromosoom 22 LD kaart*

Selleks, et erinevate andmekogumite LD mustrid oleksid võrreldavad, võtsin võrdluse alla Eesti populatsiooni 101 indiviidi, Saksa triode vanemad 82 indiviidiga ja CEPH perekondade esimese põlvkonna 26 indiviidiga. Selleks, et määrata markeritevahelisi d^2 väärtusi, kasutasin SNP Assistant programmi. Kõigi andmekogumite d^2 väärtused asuvad käesoleva töö kodulehel. LD arvutuse visualiseerimisel kasutatud parameetrid on Tabelis 6. Eesti ja Saksa andmete puhul oli genotüüpide töökindlus e. *call-rate* madal, mistõttu pidin kasutama ka madalama usaldusväärsusega andmeid. Minoorse alleeli sageduse piiriks fikseerisin $> 10\%$ ja HWE testi kriteeriumiks $p < 0,05$. Kasutades mainitud filtreid, jäi LD mustri joonistamiseks sõelale 351 markerit Eesti, 463 markerit Saksa ja 1116 markerit CEPH-i andmekogumite jaoks. Kromosoomi piirkonnad, mida nimetatud markerid katsid, on kirjas Tabelis 7. LD mustrid on Joonisel 11-13.

Tabel 6. Erinevate andmekogumite võrdlev tabel LD graafiliseks presenteerimiseks.

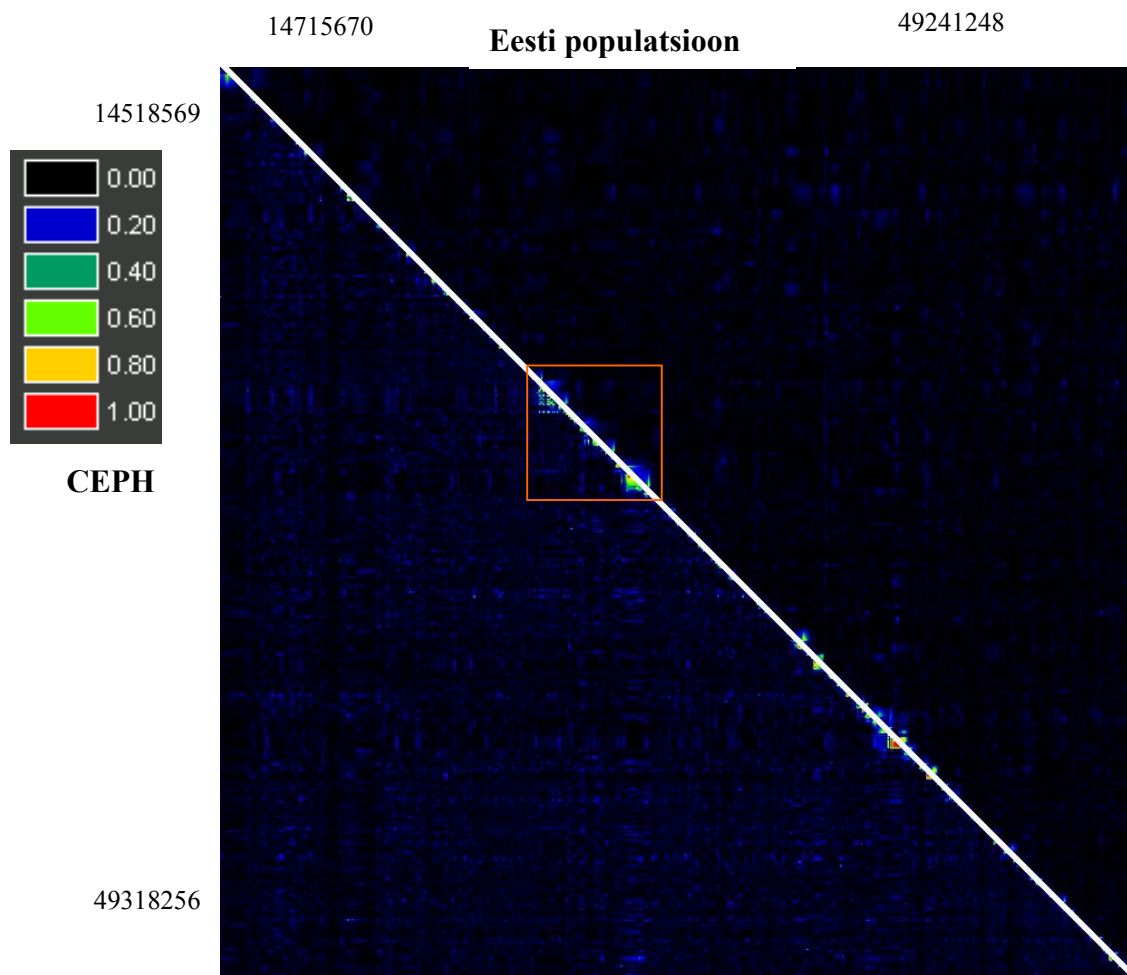
Andmekogum	Genotüpi-seerimis koht	Meetod	Indiviide	<i>call-rate</i> (%)	Minimaalne alleeli-sagedus (%)	HWE p-väärtus
Eesti populatsioon	Tartu	APEX	101	80	10	0,05
Saksa triiod (vanemad)	Tartu	APEX	82 ¹	80	10	0,05
CEPH perekonnad	Sanger Keskus	Invader	26 ¹	80	10	0,05

¹ LD mustri analüüs teostati valimil, mis sisaldas ainult I põlvkonna genotüüpe

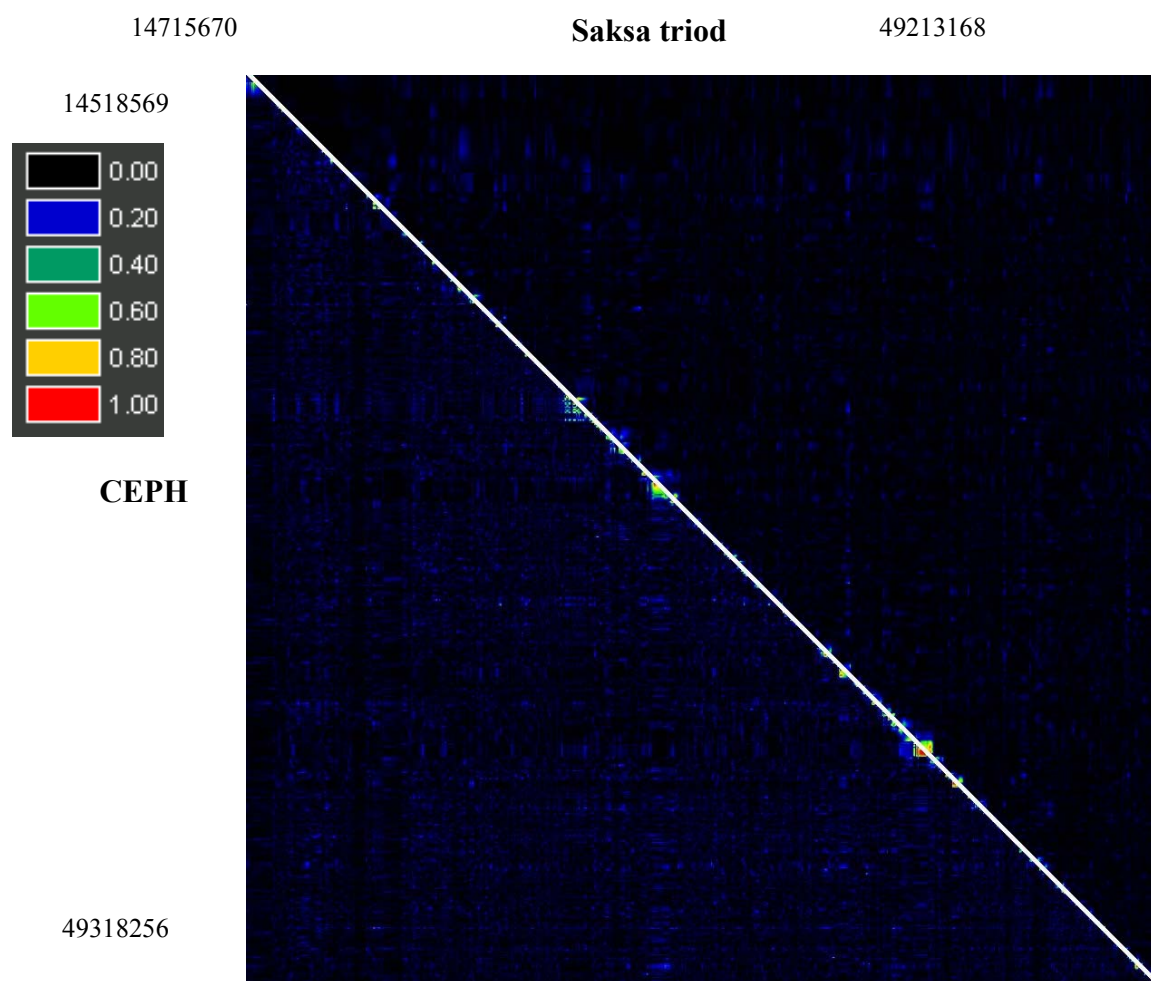
* Osa käesolevast tööst on läbiviidud grupitööna, mille kõigis etappides töö autor osales (Dawson et al., 2002).

Tabel 7. Erinevates andmekogumites kasutatud markerite arv ja kromosoomi 22 piirkond, mida vaadeldi.

Andmekogum	Markereid LD kaardil	Kaardi algus (bp)	Kaardi lõpp (bp)
Eesti populatsioon	351	14715670	49241248
Saksa triiod (vanemad)	463	14715670	49213168
CEPH perekonnad	1116	14518569	49318256

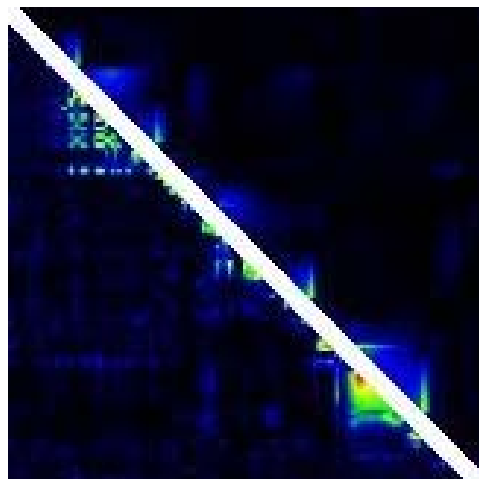


Joonis 11. Kromosoom 22 LD muster. Diagonaalist allpool on CEPH perekondade LD muster ja üleval pool eestlaste populatsiooni muster. Nii horisontaal- kui vertikaaltelge pidi jookseb positsioon kromosoomil. d^2 väärtustele vastav värv on kõrvalolevas selgituses. Punase ruudu kohta vaata joonis 13.



Joonis 12. Kromosoom 22 LD muster. Diagonaalist allpool on CEPH perekondade LD muster ja üleval pool Saksa triode muster. Nii horisontaal- kui vertikaaltelge pidi jookseb positsioon kromosoomil. d^2 väärtustele vastav värv on kõrvalolevas selgituses.

Joonis 13. Suurendus joonisel 11 asuvast regioonist 22. kromosoomil. Selgelt on näha blokkide sarnast struktuuri.



Kõige rohkem on markereid kasutusel CEPH-I perekonna andmetes, mistõttu vastava andmekogumi kaart tuli ka tihedam, kuigi valimi suurus on väikseim. Kõigi andmekogumite baasil joonistub välja lühemaid ja pikemaid LD blokke. Kõrvutades Eesti populatsiooni, Saksa triode ja CEPHi perekondade mustreid, näeme selget korrapära blokkide paiknemisel (Joonis 11-13). Sarnane LD muster viitab kas sarnasele demograafilisele ajaloole ja/või rekombinatsiooni põhirollile inimese genoomi LD mustrite kujundamises.

22. kromosoomi detailne LD kaart CEPH perekondade põhjal on publitseeritud koostööpartneritega ajakirjas Nature 2002.aastal (Dawson *et al.*, 2002).

4.2 Kromosoomide 22 ja 19 rekombinatsiooni aktiivsuse iseloomustus ja rekombinatsioonipunktide jaotuvus

Selleks, et iseloomustada rekombinatsioonipunkte, arvutati rekombinatsioonipunktide asukohad CRIMAP programmi CHROMPIC funktsiooniga. Mõlemas kromosoomis lokaliseeritud rekombinatsioonipunktid on Lisas 1 ja 2. Kromosoomis 22 oli ristsiirde punkte 109 ja 19. kromosoomis 132 punkti.

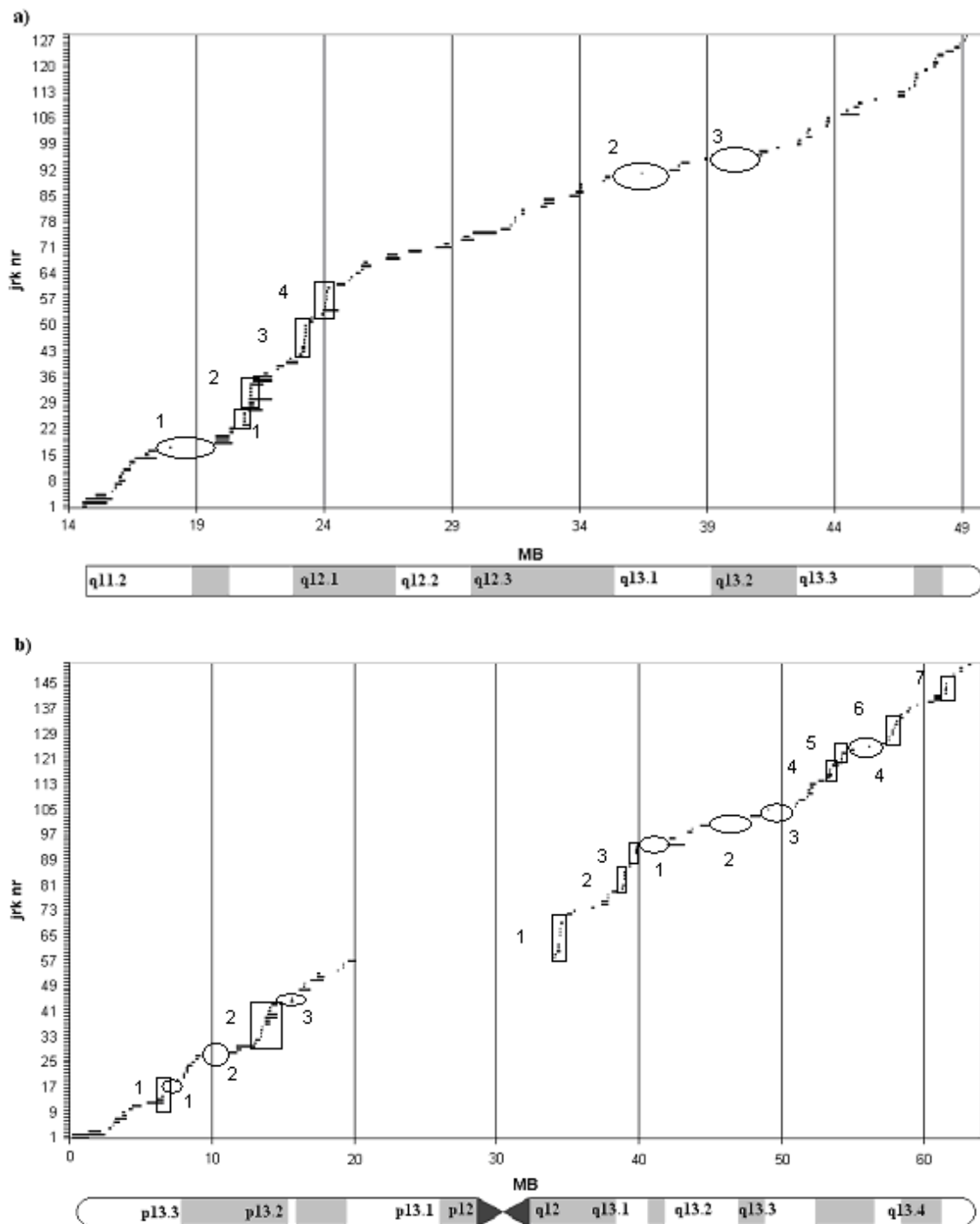
Rekombinatsioonipunktide jaotus indiiditi on näidatud Tabelis 8. Rekombinatsioonipunktide arv vastab üldjuhul reeglile, mille järgi on lubatud maksimaalselt kaks igas õlas. Huvitav on perekonna 1347 laste ekstreemselt kõrge ristsiirete arv mõlemas kromosoomis ning perekonna 1341 kõrge näitaja kromosoom 19 juures (Tabel 8).

Selleks, et algselt visuaalselt lokaliseerida kõrge ja madala rekombinatsioonilise aktiivsusega piirkonnad, koostati Joonis 14. Iga lokaliseeritud piirkond on kujutatud kas riskülikuna (aktiivne ala) või ellipsina (madala aktiivsusega ala). X-teljel on asukoht kromosoomis. Y-teljel on ristsiirde järjekorranumber kromosoomi p-telomeerist alates. Mida pikem on ristsiirde ala joonistatud, seda vähem informatiivseid markereid vastavas proovis antud kohas oli. Mida lühem, seda täpsemini õnnestus ristsiirde koht lokaliseerida. Graafikult väljajoonistuvad kiiremini tõusvad alad kirjeldavad rekombinatsiooniliselt aktiivsemaid alasid (edaspidi *hotspot* piirkond). Tühjad alad graafikul kirjeldavad rekombinatsiooni vaeseid alasid e. kõrbeid (edaspidi *coldspot*).

Ristkülikud ja ellipsid tähistavad detailsema analüüsi jaoks väljavalitud kromosoomi piirkondi (Joonis 14).

Tabel 8. Ristsiirete arv perekonniti ja indiviiditi kromosoomis 19 ja 22. Ristsiirde kohad tuvastati valdavalt kolmanda põlvkonna esindajatel. Perekonna ja indiviidi koodid (va. 19. kromosoomi töös kasutatud DNA-d) on pärit [Corielli](#) andmebaasist.

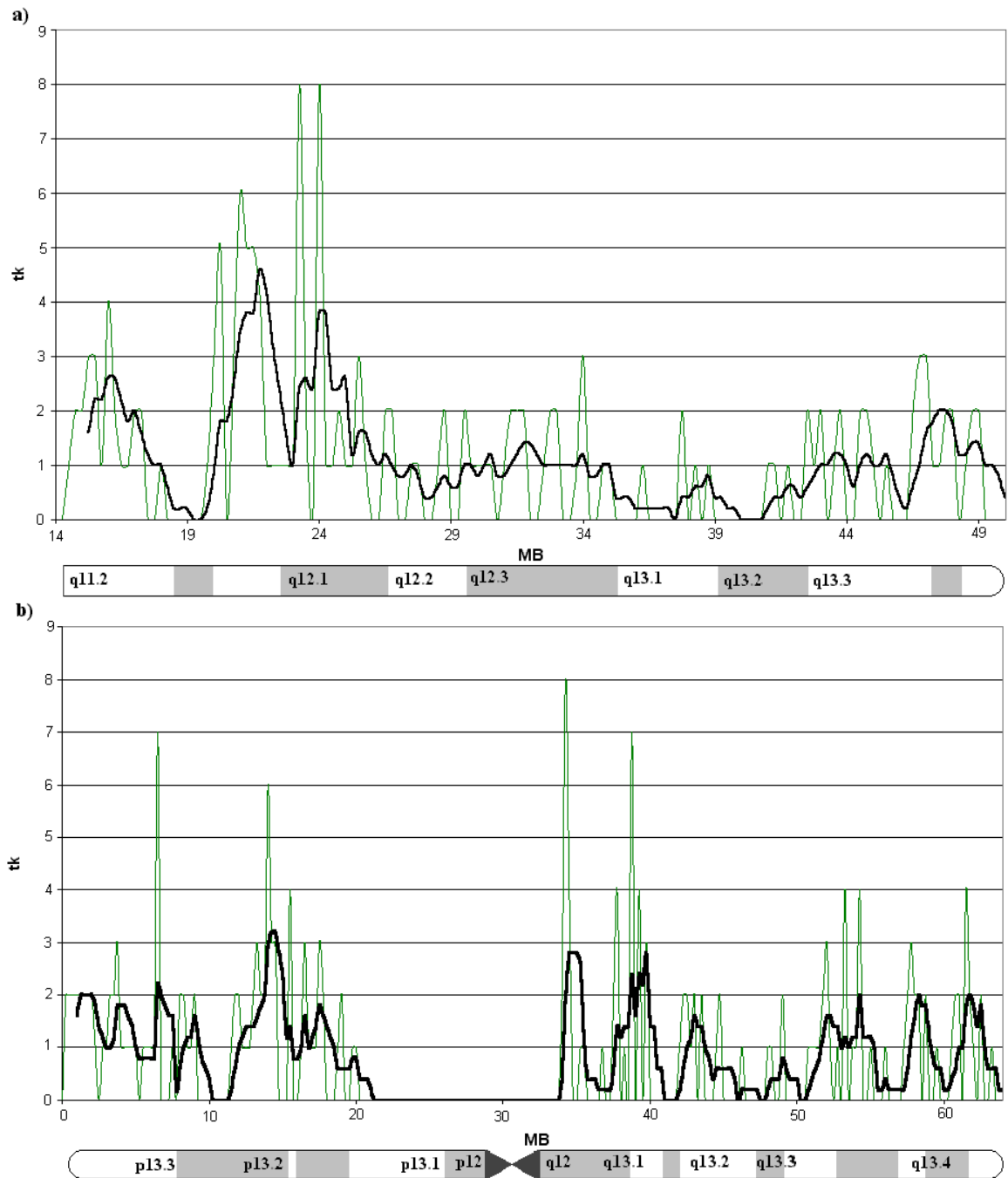
perekond	indiviid	Ristsiirete arv vanemate meioosis					
		19p ♂	19p ♀	19q ♂	19q ♀	22q ♂	22q ♀
13291	6980					1	3
	6996					1	1
	7437					0	2
	7694					4	2
	6995					3	0
	7013					2	1
	7042					5	0
	7046					5	2
	7014					2	0
	7434					4	0
	6998					2	3
	7015					0	0
	7047					3	1
	7060					2	1
1413	12110					0	3
	12113					0	3
	12115					0	4
1332	12090					0	0
1331	7005					1	2
	11818					0	1
	7033					0	2
	7	5	0	2	0		
	8	3	1	1	0		
1347	11875					31	2
	7	8	3	9	9		
	8	5	3	3	8		
1362	11984					1	1
	7	1	3	5	0		
	8	3	1	2	1		
1416	12244					10	4
	12245					11	2
	7	1	0	1	6		
	8	0	0	1	7		
1423	7	1	0	0	0		
	8	1	1	0	1		
1333	7	0	0	1	1		
	8	0	0	1	1		
1334	7	1	0	1	0		
	8	1	0	1	1		
1340	7	1	1	3	0		
	8	0	1	1	1		
1341	7	1	0	3	0		
	8	4	3	8	7		
1408	7	1	1	2	1		
	8	1	1	1	4		



Joonis 14. Rekombinatsioonipunktide jaotuvus 22. (a) ja 19. (b) kromosoomil.

X-teljel on kromosoomi positsiooni number füüsiliselt kaardilt ja Y-teljel ristsiirde järjekorranumber. Jooned tähistavad ala, mille sees üks ristsiire on toimunud. Mida lühem joon, seda informatiivsemad, üksteisele lähemal paiknevad markerid on konkreetsel indiviidil olnud. Ristkülik näitab *hotspot*-i ala, ellips *coldspot*-i ala. Nende juures olev number näitab valitud regiooni kromosoomi õlas (Tabel 13).

Rekombinatsioonipunktide jaotus 250 kb raamis on Joonisel 15. Mõlemas kromosoomis oli 250 kb sees maksimaalselt 8 ristsiiret. Keskmine rekombinatsiooniline aktiivsus on mõlemas kromosoomis sama, 1,5 ristsiiret miljoni aluspaari kohta. Selleks, et tasandada ristsiirete arvu järske muutusi, kasutati liikuva keskmise funktsiooni. Liikuv keskmine arvutati viie eelneva punkti keskmise põhjal. Väljajoonistuv kõver annab sama sisuga graafiku, nagu seda on rekombinatsiooniastmete kõver. Viie punkti keskmine annab 1,25 Mb täpsuselise graafiku ja joonistuv graafik peab olema sarnane sama suurusastme rekombinatsiooniastmete graafikuga (deCODE). Kõrvutades deCODE graafikuid (Joonis 7) ja käesoleva töö Joonist 15, võime näha korrapära, kuigi ei täheldatud rekombinatsiooni aktiivsuse suurenemist telomeerides, mis on näha *Island Map*-s. Käesolevas töös ei leitud seost rekombinatsioonilisel aktiivsusel ja kromosoomi vöödistuse muustril.



Joonis 15. Rekombinatsioonipunktide arv 250 kb aknas ja liikuv keskmine viie viimase väärtuse järgi 22. (a) ja 19. (b) kromosoomis. Vöödistusel ja rekombinatsioonilisel aktiivsusel seost ei ole. Liikuva keskmise graafik sarnaned Island Map rekombinatsiooniastmete graafikuga Joonisel 7.

4.3 Rekombinatsioonipunktide analüüs genoomsete näitajate vastu

Käesolevas töös olid võrdluse all genoomsed struktuurielemendid, mida on varasemates töödes seostatud rekombinatsioonilise aktiivsusega (vöödid, kodeeriva ala %, Alu korduste arv, GC%) ja rekombinatsioonipunktid erineva pikkusega raamides (5 kb - 1 Mb). Genoomsete struktuurielementide assotsiatsiooni rekombinatsioonilise aktiivsusega testiti parameetrilise (lineaarset assotsiatsiooni eeldavat) Pearsoni ja mitteparameetrilise Spearmani testi abil (Tabel 9-11). Lisas 3-5 on graafiline võrdlus üksikute GC%, Alu korduste arvu ja kodeeriva ala % väärtuste osas, mis on arvutatud 250 kb aknas.

Testide tulemusel võime järeldada, et:

1. Spearmani test on vähem usaldusväärne (nagu ka pärmil *Saccharomyces cerevisiae*, Petes *et al.*, 2002);
2. 19. kromosoomis puudub ja 22. kromosoomis on nõrk korrelatsioon kodeeriva ala ja rekombinatsioonilise aktiivsuse vahel (Tabel 9);
3. 19. kromosoomis on positiivne korrelatsioon Alu korduste sisalduse ja rekombinatsioonilise aktiivsuse vahel. Kromosoomis 22 see aga puudus (Tabel 10). Kromosoomis 22 on Alu elemente kolmandiku võrra vähem, kui kromosoomis 19 (chr19 on 984 tk./Mb, chr 22 661 tk./Mb);
4. 19. kromosoomis on positiivne korrelatsioon GC% ja rekombinatsioonilise aktiivsuse vahel. Kromosoomis 22 see aga puudus (Tabel 11);
5. Parimaks raamiks korrelatsiooni uuringutes on 5-100 kB, mis tõendab, et seni tehtud geneetilised kaardid on liiga hõredad korrelatsiooni leidmiseks rekombinatsioonilise aktiivsuse ja genoomi struktuurielementide vahel.

Tabel 9. Pearsoni ja Spearmani testi väärtused kodeeriva ala ja rekombinatsioonilise aktiivsuse vahel 19. ja 22. kromosoomis. Kromosoomis 22 on nõrk korrelatsioon.

Geeniala Aken (kb)	19chr				22chr			
	Pearson p		Spearman p		Pearson p		Spearman p	
5	-0,01729	0,06771	0,05706	1,62E-09	0,13845	2,20E-16	-0,02059	0,08768
10	-0,01479	0,26920	0,06443	1,46E-06	0,14059	2,22E-16	-0,02164	0,20420
20	-0,00955	0,61400	0,07973	2,48E-05	0,13271	3,07E-08	-0,02450	0,30860
30	-0,01570	0,49860	0,08534	0,00023	0,13428	4,75E-06	-0,03129	0,28840
40	-0,00333	0,90100	0,10144	0,00015	0,13018	0,00012	-0,03145	0,35430
50	-0,00379	0,89930	0,08997	0,00262	0,12841	0,00068	-0,01906	0,61570
60	0,00660	0,84060	0,11081	0,00071	0,11344	0,00615	-0,04157	0,31670
70	0,00506	0,88650	0,11518	0,00112	0,14430	0,00123	-0,01621	0,71780
80	0,01417	0,70840	0,10748	0,00446	0,12891	0,00697	-0,02126	0,65750
90	0,03375	0,40120	0,11667	0,00362	0,12157	0,01658	-0,04714	0,35420
100	0,01916	0,65120	0,11661	0,00580	0,11285	0,03508	-0,02013	0,70770
250	0,04612	0,49130	0,15411	0,02083	0,01138	0,89350	-0,10738	0,20480
500	0,08203	0,38770	0,17657	0,06145	-0,01278	0,91580	-0,11728	0,32930
1000	0,04453	0,74220	0,14480	0,28160	0,16629	0,33240	-0,11188	0,51470

Tabel 10. Pearsoni ja Spearmani testi väärtused Alu korduste ala ja rekombinatsioonilise aktiivsuse vahel 19. ja 22. kromosoomis. Kromosoomis 19 on nõrk korrelatsioon.

Aluala Aken (kb)	19chr				22chr			
	Pearson p		Spearman p		Pearson p		Spearman p	
5	0,12106	2,20E-16	0,12415	2,20E-16	-0,04336	0,00032	-0,02442	0,04281
10	0,13495	2,20E-16	0,14098	2,20E-16	-0,04892	0,00409	-0,02661	0,11850
20	0,14612	8,66E-15	0,15395	2,88E-16	-0,05351	0,02612	-0,02906	0,22720
30	0,13891	1,77E-09	0,14998	7,89E-11	-0,06167	0,03627	-0,03613	0,22020
40	0,14286	8,31E-08	0,16220	1,09E-09	-0,05630	0,09720	-0,02866	0,39860
50	0,12929	1,46E-05	0,15274	3,01E-07	-0,05457	0,15040	-0,03242	0,39300
60	0,13327	4,52E-05	0,16166	7,36E-07	-0,06464	0,11930	-0,03885	0,34940
70	0,12482	0,00041	0,15413	1,24E-05	-0,06417	0,15230	-0,06482	0,14810
80	0,12709	0,00076	0,17539	3,20E-06	-0,05416	0,25860	-0,02964	0,53640
90	0,09750	0,01508	0,14873	0,00020	-0,04671	0,35890	-0,04062	0,42480
100	0,11109	0,00857	0,17859	2,23E-05	-0,09047	0,09149	-0,09183	0,08670
250	0,07205	0,28190	0,16109	0,01566	-0,10481	0,21610	-0,18537	0,02789
500	0,06127	0,51910	0,15136	0,10940	-0,11983	0,31960	-0,27139	0,02234
1000	0,13156	0,32930	0,28157	0,03418	-0,20822	0,22300	-0,29290	0,08322

Tabel 11. Pearsoni ja Spearmani testi väärtused GC protsendi ja rekombinatsioonilise aktiivsuse vahel 19. ja 22. kromosoomis. Kromosoomis 19 on nõrk korrelatsioon.

GC%	19chr				22chr			
Aken (kb)	Pearson p		Spearman p		Pearson p		Spearman p	
5	0,21075	2,20E-16	0,20064	2,20E-16	-0,07829	7,90E-11	-0,08472	1,95E-12
10	0,22191	2,20E-16	0,21034	2,20E-16	-0,08542	5,17E-07	-0,09417	3,08E-08
20	0,23087	2,20E-16	0,21929	2,20E-16	-0,09147	0,00014	-0,10107	2,56E-05
30	0,23813	2,20E-16	0,22908	2,20E-16	-0,08408	0,00428	-0,09916	0,00075
40	0,23857	2,20E-16	0,23287	2,20E-16	-0,08583	0,01136	-0,09346	0,00584
50	0,23525	1,55E-15	0,23830	8,52E-16	-0,08708	0,02158	-0,10081	0,00780
60	0,22861	1,67E-12	0,23206	8,85E-13	-0,09526	0,02154	-0,10990	0,00799
70	0,21930	3,72E-10	0,23008	5,30E-11	-0,07937	0,07651	-0,09442	0,03502
80	0,21020	2,02E-08	0,22574	1,76E-09	-0,08920	0,06244	-0,09270	0,05283
90	0,20053	4,69E-07	0,21409	7,72E-08	-0,07406	0,15400	-0,09216	0,06979
100	0,20659	8,36E-07	0,23536	1,97E-08	-0,10026	0,06135	-0,11165	0,03712
250	0,11164	0,09483	0,16008	0,01633	-0,07872	0,35350	-0,08502	0,31570
500	0,03447	0,71700	0,07350	0,43850	-0,05602	0,64260	-0,02482	0,83700
1000	0,05944	0,66050	0,11266	0,40290	-0,08204	0,63430	-0,05126	0,76610

4.4 *Hotspot*-ide ja *coldspot*-ide genoomse järjestuse ja struktuuri analüüs

Mõlemal kromosoomil lokaliseeriti kõrge ja madala rekombinatsioonilise aktiivsuse piirkonnad (Joonis 14 ning Lisad 1 ja 2). Kõrge rekombinatsiooniga alad võeti võimalikult kitsalt, seevastu madala rekombinatsiooniga alad võimalikult laiad (Tabel 13). Vastavad alad määrati esmalt visuaalselt graafikult (Joonis 14) ning seejärel piiritleti vastavate markerite positsiooninumbriaga täpsemalt. Tihedas alas pidi olema ligistikku vähemalt viis rekombinatsioonijuhtu 500 kb kohta ja hõredas alas kuni üks rekombinatsioonijuht 1,5 Mb kohta. Kokku piiritleti 23 ala (Tabel 12 ja 13).

Tabel 12. Kõrge ja madala rekombinatsioonipunktide arv kromosoomiõlgade lõikes.

Kromosoomi õlg	Kõrge rekomb. ala	Madal rekomb. ala
19p	2	3
19q	7	4
22q	4	3

Kõigil 23-l regioonil arvutati GC% ning Alu ja geeniala protsent kogu regiooni ulatuses (Tabel 13). Üleüldist erinevust kõrge ja madala intensiivsusega piirkondade vahel pole. Kõrge rekombinatsioonilise aktiivsusega piirkondade seast viis sisaldasid vähe geeniala ja kaheksa palju geeniala. Silma hakkab kõrge rekombinatsioonilise piirkondade q õla tsentromeerse osa suhteliselt madal Alu % ja kõrge geeniala %. Varasemalt on vastavaid elemente peetud kooskäivateks (Lander *et al.*, 2001).

Mõlema kromosoomi q õla esimene kõrge ristsiirete piirkond on nii Alu kui ka geenidevaene. Selliselt kirjeldatud kõrge rekombinatsioonilise aktiivsusega piirkonda ei ole varem kirjeldatud. *Islan Map*-i rekombinatsiooniastmete graafikult ei tule vastav koht tänu harvadele markeritele esile. Kromosoom 19 puhul sulanduvad *deCODE* kaardi kaks teravat *hotspot*-i kokku üheks 19q tsentromeeri kõrvale. Kuna *Island Map*-is on telomeerides suurem rekombinatsiooniline aktiivsus ja et imetajate evolutsioonis on kromosoomide samad segmendid ühel liigil telomeeris, teisel tsentromeeri lähedal, siis kontrollisime q õlgade tsentromeerseid alasi telomeerispetsiifiliste minisatelliitsete

järjestustega (GGGCAGGARG)_n. Leidsime, et antud piirkonnad ei sisalda nimetatud kordusi rohkem, kui kogu kromosoom keskmiselt.

Tabel 13. Kõrge ja madala rekombinatsioonisagedustega alade detailsem iseloomustus. Tabelis on nimetus, regiooni algus, lõpp ja pikkus. GC% on arvutatud kogu regiooni peale. Alu ja geeniala protsent on arvutatud vastavate elementide alla jäävate aluspaaride arvu ja kogu regiooni pikkuse järgi.

Nimetus	Algus (bp)	Lõpp (bp)	Pikkus (Kb)	GC%	Alu ala %	Geeni ala %
22q_kõrge1	20841759	20881216	39	45	3	1
22q_kõrge2	21028512	21246682	218	44	13	93
22q_kõrge3	23043639	23320222	277	47	18	70
22q_kõrge4	23915832	24211215	295	47	18	10
19p_kõrge1	6526268	6637406	111	47	32	28
19p_kõrge2	13016633	14589111	1572	51	35	45
19q_kõrge1	34039448	34565647	526	45	8	1
19q_kõrge2	38808471	38999862	191	51	14	78
19q_kõrge3	39340737	39833954	493	44	23	78
19q_kõrge4	53239555	53459264	220	47	27	70
19q_kõrge5	53986638	54443164	457	52	35	66
19q_kõrge6	57535849	58096379	561	44	31	69
19q_kõrge7	61444236	61763972	320	42	10	84
22q_madal1	18007006	19731638	1725	52	20	49
22q_madal2	35188680	37559771	2371	51	19	52
22q_madal3	38972653	41060127	2087	48	30	58
19p_madal1	6637406	7586455	949	48	34	52
19p_madal2	9190884	11289340	2098	49	36	62
19p_madal3	14589111	15546311	957	47	27	54
19q_madal1	39833954	42004778	2171	48	25	62
19q_madal2	44807869	47805493	2998	48	22	54
19q_madal3	49083536	50934073	1851	48	27	62
19q_madal4	55039196	57079537	2040	49	20	43

Kõigi uuritud genoomsete elementide, GC protsendi, Alu ala ja transkribeeritava ala protsendi, osas arvutati keskmised väärtused (Tabel 14). Keskmise GC% ja Alu ala % olid peaaegu identsed. Geeniala protsent erines ka väga vähe. Samas, vaadates eelmist tabelit võib täheldada, et geeniala protsent on suurem just kõrge rekombinatsiooniga piirkondade hulgas (Tabel 13).

Tabel 14. Kõrge ja madala rekombinatsiooniga piirkondade genoomsete elementide keskmiste väärtuste risttabel.

	Keskmine GC%	Keskmine Alu ala %	Keskmine geeni ala %
Kõrge	48 (42-52)	25 (3-35)	53 (1-93)
Madal	49 (47-52)	25 (19-36)	55 (43-62)

5 Arutelu

5.1 Eesti populatsiooni, Saksa triode ja CEPH perekondade LD mustri võrdlus inimese 22. kromosoomi põhjal

Erinevate populatsioonide LD mustri võrdluse juures on määravaks markerite õige valik. Siinses töös kasutatud markerid on SNP massilise avastamise algetapist (Mullikin *et al.*, 2000), mil meil oli valikuvõimalus vaid 1623 SNP seast. Ensembl v.17.33 pakub teadlastele kasutamiseks 22. kromosoomis juba 120765 SNP-d. Käesoleva töö juures puuduvad Eesti populatsiooni valimis täielikult informatiivsed markerid CEPH perekondade juures eksisteeriva kõige pikema LD bloki sees. Samas on CEPH perekondade ja Saksa triode LD mustrid peaaegu identsed. Hetkel jääbki selgusetuks, mis põhjustas informatiivsete markerite puuduse Eesti populatsiooni valimis antud piirkonnas. Selleks võivad olla tehnoloogilise platvormi erinevused (koostööpartnerid kasutasid Invaderit [Third Wave Technologies](#), Dawson *et al.*, 2002), markerite ebaõige valik (praegu puudub korralik SNP-sageduste analüüs Eesti populatsioonis) või populatsiooni iseärasus. Populatsiooni iseärasuseks võib olla uuritava piirkonna markerite homosügootsus, mida on täheldatud ka Zimbabwe populatsiooni LD blokkide võrdlemisel aasialaste ja eurooplastega (Kauppi *et al.*, 2003).

Ristsiirete jaotuse juures on küsimus, kas ristsiirded on jaotunud juhuslikult või mittejuhuslikult. Juhusliku ristsiirete puhul peaks jaotuvus olema ühtlaselt üle kromosoomi. Käesoleva töö juures kasutatud Euroopa päritolu individid moodustavad 1/6 maailma elanikkonnast. Kui eeldada, et maailmas on üleüldse elanud 20 miljardit *Homo sapiensi* läbi aegade ja elanikkonna proportsioon kontingenditi on olnud sama ning, et iga meioosi käigus on olnud keskmiselt 50 kiasmi, siis saame summaarseks ristsiirete arvuks 166 miljardit. Kui vastav number jagada läbi inimese genoomi pikkusega saame, et ühtlase jaotuvuse juures peaks ristsiire toimunud olema iga nukleotiidipaari vahelt 55 korda. Seega, igasugune LD blokkide olemasolu tõestab ristsiirete ebaühtlast jaotuvust, kui ei oleks minoorse alleeli minimaalse esinemissageduse piirangut. Kuna aga 20 % piirang on nõutav, siis nähtav LD muster näitab ajaloolist mustrit, mille daatum on raskesti määratav. Võib-olla, et käesolevas töös nähtavad LD

bloki on populatsiooni pudelikaela haploblokid. Või on osad (nt. pikad) blokid tingitud transverveerunud kromosoomi segmentidest? Transversioonide sagedust inimese ajaloos tõendab pikkade kromosoomide lõikude duplitseerumine sama kromosoomi eri telomeeris (Bailey *et al.*, 2002a), mis takistab ristsiirde toimumist kuni teine kromosoom ei ole sama transversiooniga.

Võttes aluseks CEPH'i perekondade, Saksa triode vanemate ja Eesti populatsiooni LD mustrid võime väita, et ei ole võimalik eristada kahte tsenaariumit: a) LD muster sõltub rekombinatsioonist või b) populatsiooni hilisest ajaloost.

5.2 Rekombinatsioonipunktide jaotus 22. ja 19. kromosoomis

Käesolev töö annab olulist uut informatsiooni rekombinatsioonipunktide jaotuse kohta meioosis, kuna seni on kiasmide ja rekombinatsioonipunktide arvu peetud võrdseks. Selle teooria järgi on 19. kromosoomis kuni kolm ristsiirde kohta ja 22. kromosoomis kuni kaks ristsiirde kohta (Hulten, 1974). Seniste geneetiliste kaartide tegemisel on alati topeltrekombinandid välja praagitud. Käesoleva töö järgi ei ole selline tegevus rekombinatsiooni intensiivsuse seisukohalt õige. Kahel täiesti eri meetodit kasutaval teadlaste grupil on genotüpiseeritud samu perekondi erinevatel kromosoomidel ja see on andnud tulemused, kus ühel perekonna lastel on ekstreemselt kõrge rekombinatsiooniline aktiivsus – 10 korda kõrgem. Rekombinatsiooni interferents peaks sellise võimaluse välistama. Üks teadlaste grupp on samuti kahtluse alla seadnud, et kiasmide arv = ristsiirete arv, kuna eksisteerib veel teine, interferentsist mittesõltuv rekombinatsioon (Housworth *et al.*, 2003). Geneetiliste kaartide konstrueerimisel on kasutatud mikrosatelliitseid markereid, mis on genoomis vähema sagedusega. Markerite hõredamast asustusest ja markerite spetsiifikast tingituna võibki tulla erinev tulemus kui seda on käesoleval töö. Samuti ei ole meil teada, kas geneetiliste kaartide konstrueerimise käigus ei ole osa andmeid jäänud publitseerimata. CEPH perekond 1347 puhul said erinevad teadlaste grupid sama tulemuse: traditsioonilisest 15 korda kõrgem ristsiirete arv 22. kromosoomis (31 versus 2) ja 19. kromosoomis 6 korda suurem (17 versus 3). Väiksemate markerite arvu puhul võivad paar markerit reeglist eralduda ja see juhtum topeltrekombinandiks nimetada, kuid käesolevas töös on nn. topeltrekombinante liiga palju ja ei riskiks ühtegi neist kõrvaldada. CEPH perekond nr 1347 on lisauurimist

väärt kui abivahend rekombinatsiooni aktiivsuse regulatsioonimehhanismide tundmaõppimisel.

Käesolev töö kinnitab rekombinatsioonipunktide ebaühtlast jaotuvust piki kromosoomi. Käesolevas töös koostatud liikuva keskmise ja *Island Map*-i rekombinatsiooniastmete graafikud kromosoomide 22 ja 19 kohta on sarnase kujuga. Erinevus on telomeerides, kus sinne töö kõrget rekombinatsioonilist aktiivsust ei tuvastanud. Käesoleva töö markerite tihedus lubab ristsiirde kohti täpsemalt lokaliseerida. *Island Map* markerite lokaliseeris rekombinatsioonijuhud 19. kromosoomis keskmiselt 2,2 Mb täpsusega (autori arvutused CRI-MAP programmi CHROMPIC funktsiooniga) olgugi, et markerite lokalisatsioon on 0,5 Mb vahedega. Jällegi on oluline markerite informatiivsus.

Käesolevas töös leidsime kogu kromosoomi hõlmava analüüsi juures nõrga korrelatsiooni rekombinatsioonilise aktiivsuse ja Alu ala % ning GC% vahel 19. kromosoomis. 22. kromosoomis vastavat korrelatsiooni ei olnud, küll aga oli korrelatsioon geeniala % ja rekombinatsioonilisel aktiivsusel 22. kromosoomis. 19. kromosoomiga tehtud testid olid sama akna juures usaldusväärsemad rohkema arvu andmepunktide tõttu, kui seda olid kromosoom 22 tehtud korrelatsiooni testid. *Island Map*-iga tehtud kogu genoomi hõlmavate materjalide analüüsi juures leiti tugev korrelatsioon GC% ja rekombinatsioonilise aktiivsuse vahel, kuid sealse töö aknaks oli 3Mb (Kong *et al.*, 2002). *Genethon Map* baasil tehtud analüüs ei andnud kõrget korrelatsiooni rekombinatsiooni astmete ja GC% ning Alu korduste vahel kasutades 250 kb akent (Yu *et al.*, 2001). Võrreldes inimese ja pärmis rekombinatsioonilise aktiivsuse ja GC% vahelist korrelatsiooni võib öelda, et pärmil on vastav korrelatsioon eriti kõrge nii 0,5 kb akna kui ka 100 kb akna juures (Pearsoni test $r = 0,71-0,96$). Samas on rekombinatsioonilise aktiivsuse parameetrikaks hoopis kaheahelaliste katkestuste arv (DSB), millega vastav korrelatsioon arvutatud on (Petes *et al.*, 2002)

5.3 Ristsiirete *hotspot*-ide kirjeldus

MHC piirkonnas lokaliseeritud rekombinatsiooni *hotspot*-ide pikkuseks on määratud 1,5-3 kb. MHC regioon on üks enimuuritud piirkondi inimese genoomis, mistõttu vastavate *hotspot*-ide leidmine oli loogiline teaduslik etapp. Ülejäänud inimese

genoomiga on olukord olnud raskem, kuna varasemate geneetiliste kaartide täpsus on olnud keskmiselt 0,5 Mb ja meiooside arv on jäänud alla 200 indiviidi. *Island Map* on teinud järgmise etapi väga suure hulga indiviidide genotüpiseerimisega, mis on andnud oluliselt informatiivsema geneetilise kaardi, kuid markerite arv on jäänud tagasihoidlikuks. Antud töös tuvastati ristsiirde kohad kõige kitsamalt, mis annab võimaluse defineerida *hotspot*-ide asukohad veelgi täpsemalt, kuna kasutatud markeritevaheline keskmine kaugus oli 23 kb.

Käesolevas töös lokaliseeritud rekombinatsiooni *hotspot*-ide juures ei leitud assotsiatsiooni kromosoomide struktuursete elementide nagu vöödid, Alu elemendid ja GC%, vahel seost. Alude esinemissagedus on genoomis sedavõrd suur, et ta assotsieerub pea iga genoomse motiiviga. Ainsana oli tugevamat assotsiatsiooni näha rekombinatsioonilisel aktiivsusel ja geenialadel, kuna suurem osa *hotspot*-e asus geenirikastes piirkondades. See kinnitab teooriat, et avatud kromatiinistruktuur laseb DNA-le ligi ka rekombinatsiooni aktiivsust määravaid valke. Siiski leiti siinses töös erandlik rekombinatsiooni hotspot nii 19. kui ka 22. kromosoomis. Mõlemad *hotspot*-id asuvad q õla tsentromeerses osas ja mõlemas on väga madal transkribeeriva järjestuse osatähtsus ja Alu korduste arv. Tavaliselt käivad mainitud struktuurielemendid koos. On väga võimalik, et inimesel on ka mitut tüüpi *hotspot*-e. Pärmis *Saccharomyces cerevisiae*-l on leitud kolme tüüpi *hotspot*-e: a) α *hotspot* eeldab transkriptsioonifaktori seondumist, b) β *hotspot* assotseerub nukleasitundliku kromatiiniga ja c) γ *hotspot* assotseerub kõrge GC protsendiga. Võib-olla ongi tsentromeerilähedane *hotspot* pärmis β *hotspot*-i analoog inimesel, mille nukleasitundlikkuse taga on näit. DNA kordusjärjestused, mis muudavad lokaalse DNA arhitektuuri nukleasidele kättesaadavaks.

5.4 Järeldused

Käesoleva töö raames tehtud katsete ja analüüsi põhjal võime järeldada, et:

- rekombinatsioonipunktid ei ole kromosoomides ühtlaselt jaotunud;
- LD mustrid on sarnased Euroopa päritoluga populatsioonides: eesti, saksa ja CEPH-perekonnad perekondadel sarnased;
- kuni siiani rekombinatsiooniliselt passiivseteks peetud tsentromeeri lähedased piirkonnad sisaldavad rekombinatsiooni *hotspot*-e nii kromosoomis 19 kui ka 22;
- mikrosatelliitseid markereid kasutanud analüüside põhjal ristsiireterikasteks peetud telomerides ei täheldatud kõrgeenenud rekombinatsioonilist aktiivsust;
- rekombinatsiooni *hotspot*-ide ja *coldspot*-ide vahel on järsk üleminek;
- rekombinatsioonil on pärilik taust, kuna varieerub perekonniti;
- rekombinatsioonilisel aktiivsusel on nõrk korrelatsioon GC% ja Alu ala %-ga 19. kromosoomis ning kodeeriva ala %-ga 22. kromosoomis.

Kokkuvõte

Geneetilise materjali rekombineerumine toimub meioosis. Vead meioosi käigus põhjustavad nii raskeid nn. Genoomseid haigusi kui ka iseeneslikke aborte. Samas on nad oluliseks materjaliks eluslooduse evolutsioonis. Üheks võimaluseks meioosi uurimiseks on suurte perekondade genotüpiseerimine kõrgtihedalt paiknevate markeritega.

Käesoleva töö eesmärk on uurida kasutades kõrgtihedaid SNP markereid ristsiirde punktide jaotust ja korrelatsiooni struktuursete genoomielementidega (Alu kordused, kodeeriv ala, vöödistus, GC%) inimese 19. ja 22. kromosoomis.

Töös võrdleme ka LD mustreid erinevates Euroopa päritoluga populatsioonides: a) CEPH perekondades, b) Saksa triodes ja c) Eesti populatsioonide vahel.

Saksa triode ja Eesti populatsiooni valim kromosoom 22 markerite genotüpiseerimisplatform põhines APEX meetodil, mis baseerub komplementaarsete DNA fragmentide hübridisatsioonil ja DNA polümeraasi poolt läbiviidud praimerektensioonil klaaskiibil. CEPH perekondade kromosoom 19 ja 22 genotüübid, mille abil lokaliseeriti rekombinatsioonide toimumiskohad, saadi publitseeritud materjalidest (Phillips *et al.*, 2003; Dawson *et al.*, 2002) .

Käesoleva töö tulemustena saime tunnistust, et a) rekombinatsioonipunktid ei ole kromosoomides ühtlaselt jaotunud, b) LD mustrid on sarnased Euroopa päritoluga populatsioonides: eesti, saksa ja CEPH; perekonnad perekondadel sarnased, c) kuni siiani rekombinatsiooniliselt passiivseteks peetud tsentromeeri lähedased piirkonnad sisaldavad rekombinatsiooni *hotspot*-e nii kromosoomis 19 kui ka 22, d) mikrosatelliitseid markereid kasutanud analüüside põhjal ristsiirete rikasteks peetud telomeerides ei täheldatud kõrgeenenud rekombinatsioonilist aktiivsust, e) rekombinatsiooni *hotspot*-ide ja *coldspot*-ide vahel on järsk üleminek, f) rekombinatsioonil on pärilik taust, kuna varieerub perekonniti ja g) rekombinatsioonilisel aktiivsusel on nõrk korrelatsioon GC% ja Alu ala protsendiga 19. kromosoomis ning kodeeriva ala protsendiga 22. kromosoomis.

Töö tulemused aitavad paremini mõista rekombinatsiooni aktiivsust erinevates kromosoomi piirkondades, eri populatsioonides ja indiviidides.

Inimese genoomi meiotilise ristsiirete *hotspot*-ide detailsem analüüs seisab veel ees.

Summary

Recombination of genetic material occurs during meiosis. However, mistakes during meiosis can cause serious diseases and miscarriages. Also, recombinational arrangements provide the major material for evolution of living organism.

The subject of the current paper is to study the distribution of crossover spots in human chromosomes 22-nd and 19-th using high-density SNP markers. In addition, I explored the potential correlation of the identified hotspots with structural elements of human genome(Alu repeats, gene regions, banding, GC%).

In this paper we compare LD pattern between European populations of different origins: a) CEPH families, b) German trios and c) Estonian population.

German trios and a set from Estonian population was genotyped in APEX method based on hybridisation of complementary DNA fragments and primer extension by DNA polymerase in glass slide. In order to identify recombination hotspots, genotypes of CEPH families were obtained from published datasets (Phillips et al., 2003; Dawson et al., 2002).

This study provided new information on the following aspects:

- a) recombination spots are not randomly distributed across chromosomes,
- b) LD patterns are similar in different populations of European origin,
- c) regions close to centromeres in chromosome 22 and 19 consist also recombination hotspots,
- d) no increased recombinational activity was obtained at telomeres,
- e) there is a sharp border between recombination hotspot and coldspot,
- f) recombinational activity has a genetic background proved by differences in hotspot concentration in different families,
- g) recombination activity has the correlation between GC% and Alu repeats in chromosome 19 and transcribed regions in chromosome 22.

The results of the current thesis will help to understand the variation of recombinational activity in different chromosomal regions, in different populations and individuals.

The detailed analysis of meiotic crossovers of hotspots in human genome is still ahead.

Tänuavaldused

Eelkõige tänan oma juhendajat Maris Laant, kelle abistava käe all oli antud töö võimalik läbi viia. Tänan Biodata AS ja Asper Biotech AS, kelle tehnilised ressursid lubasid käesoleva töö bioinformaatilist poolt teha. Tänan prof. Andres Metspalu, kes 22. kromosoomi projekti Tartusse tõi ja selle teemaga mind kokku viis. Samuti tänan väga paljusid inimesi, kes osalesid töös kasutatud genotüüpide genereerimisel TUMRI-s, Asper Biotech AS-s, Sanger Keskuses ja Orchid Biosciences Ltd.-s.

Kasutatud kirjandus

- Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD** (1994) Molecular biology of the cell. 3th edition. Garland Publishing, Inc.
- Ahmad T, Neville M, Marshall SE, Armuzzi A, Mulcahy-Hawes K, Crawshaw J, Sato H, *et al*** (2003) Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 12:647-56.
- Anderson EC, Slatkin M** (2004) Population-genetic basis of haplotype blocks in the 5q31 region. *Am J Hum Genet* 74:40-49.
- Andreson R** (2002) Erinevate *in silico* meetodite võrdlus PCR praimerite kvaliteedi parandamiseks. Magistritöö.
- Aoki K, Suzuki K, Sugano T, Tasaka T, Nakahara K, Kuge O, Omori A, *et al*** (1995) A novel gene, Translin, encodes a recombination hotspot binding protein associated with chromosomal translocations. *Nat Genet* 10:167-74.
- Ardlie KG, Kruglyak L, Seielstad M** (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309.
- Arnheim N, Calabrese P, Nordborg M** (2003) Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. *Am J Hum Genet* 73:5-16.
- Ashley T, Plug AW, Xu J, *et al.*** (1995) Dynamic changes in Rad51 distribution in chromatin during meiosis in male and female vertebrates. *Chromosoma* 104:19-28.
- Badge RM, Yardley J, Jeffreys AJ, Armour JA** (2000) Crossover breakpoint mapping identifies a subtelomeric hotspot for male meiotic recombination. *Hum Mol Genet* 9:1239-44.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE** (2002a) Recent segmental duplications in the human genome. *Science* 297:1003-7.
- Bailey JA, Yavor AM, Viggiano L, Misceo D, Horvath JE, Archidiacono N, Schwartz S, *et al*** (2002b) Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am J Hum Genet* 70:83-100.

- Bernardi G** (2000) Isochore and the evolutionary genomics of vertebrates. *Gene* 241:3-17.
- Batzer MA, Deininger PL** (2002) Alu repeats and human genomic diversity. *Nat Rev Genet* 3:370-9.
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL** (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861-9.
- Brzustowicz LM, Merette C, Xie X, Townsend L, Gilliam TC, Ott J** (1993) Molecular and statistical approaches to the detection and correction of errors in genotype databases. *Am J Hum Genet* 53:1137-45.
- Camp NJ** (1997) Genomewide transmission/disequilibrium testing--consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 61:1424-30.
- Cann RL, Stoneking M, Wilson AC** (1987) Mitochondrial DNA and human evolution. *Nature* 325:31-6.
- Cardon LR, Abecasis GR** (2003) Using haplotype blocks to map human complex trait loci. *Trends Genet* 19:135-40.
- Cohen PE, Pollard JW** (2001) Regulation of meiotic recombination and prophase I progression in mammals. *Bioessays* 23:996-1009.
- Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M** (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am J Hum Genet* 71:759-76.
- Collins A, Ennis S, Taillon-Miller P, Kwok PY, Morton NE** (2001) Allelic association with SNPs: metrics, populations, and the linkage disequilibrium map. *Hum Mutat* 17:255-62.
- Collins FS, Guyer MS, Charkravarti A** (1997) Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580-1.
- Cui XF, Li HH, Goradia TM, Lange K, Kazazian HH, Jr., Galas D, Arnheim N** (1989) Single-sperm typing: determination of genetic distance between the G gamma-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. *Proc Natl Acad Sci U S A* 86:9389-93.
- Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES** (2001) High-resolution haplotype structure in the human genome. *Nat Genet* 29:229-232.

Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Tibling T, Tinsley E, Kirbi S, Carter D, Papaspyridonos M, Livingstone S, Ganskel R, Lõhmussaar E, Zernant J, Tõnisson N, Remm M, Mägi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I (2002) A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 418:544-8.

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, *et al* (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-4.

Eichler EE (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* 17:661-9.

Filipski J, Thiery JP, Bernardi G (1973) An analysis of the bovine genome by Cs₂SO₄-Ag density gradient centrifugation. *J Mol Biol* 80:177-97.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, *et al* (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225-9.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD (2000) Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 97:11383-90.

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. *Nature Genet* 25:12-14.

Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res* 22:5456-65.

Gupta PK, Sahota A, Boyadjiev SA, Bye S, Shao C, O'Neill JP, Hunter TC, *et al* (1997) High frequency in vivo loss of heterozygosity is primarily a consequence of mitotic recombination. *Cancer Res* 57:1188-93.

Housworth EA, Stahl FW (2003) Crossover interference in humans. *Am J Hum Genet* 73:188-97.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337-338.

Hulten M (1974) Chiasma distribution at diakinesis in the normal human male. *Hereditas* 76:55-78.

- Jeffreys AJ, Ritchie A, Neumann R** (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum Mol Genet* 9:725-33.
- Jeffreys AJ, Kauppi L, Neumann R** (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217-22.
- Jeffreys AJ, Neumann R** (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat Genet* 31:267-71.
- Kauppi L, Sajantila A, Jeffreys AJ** (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum Mol Genet* 12:33-40.
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, *et al.*** (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241-7.
- Lander ES, Linton LM, Birren B, Nusbaum C, *et al.*** (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Laurie DA, Hulten MA** (1985) Further studies on bivalent chiasma frequency in human males with normal karyotypes. *Ann Hum Genet* 49:189-201.
- Lewontin RC** (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49:49-67.
- Lewontin RC** (1988) On measures of gametic disequilibrium. *Genetics* 120:849-52.
- Li HH, Gyllenstein UB, Cui XF, Saiki RK, Erlich HA, Arnheim N** (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 335:414-7.
- Lien S, Kaminski S, Alestrom P, Rogne S** (1993) A simple and powerful method for linkage analysis by amplification of DNA from single sperm cells. *Genomics* 16:41-4.
- Lien S, Szyda J, Schechinger B, Rappold G, Arnheim N** (2000) Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. *Am J Hum Genet* 66:557-66.
- Mahadevan LC, Willis AC, Barratt MJ** (1991) Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. *Cell* 65:775-83.

Matise TC, Porter CJ, Buyske S, Cuttichia AJ, Sulman EP, White PS (2002) Systematic evaluation of map quality: human chromosome 22. *Am J Hum Genet* 70:1398-410.

May CA, Shone AC, Kalaydjieva L, Sajantila A, Jeffreys AJ (2002) Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX. *Nat Genet* 31:272-5.

Miller OJ, Therman E (2001) *Human Chromosomes*. Springer-Verlag

Mohrenweiser HW, Tsujimoto S, Gordon L, Olsen AS (1998) Regions of sex-specific hypo- and hyper-recombination identified through integration of 180 genetic markers into the metric physical map of human chromosome 19. *Genomics* 47:153-62.

Mullikin JC, Hunt SE, Cole CG, Mortimore BJ, Rice CM, Burton J, Matthews LH, et al (2000) An SNP map of human chromosome 22. *Nature* 407:516-20.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, et al (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406:593-9.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, et al (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* 30: 190-193.

Ott J (1999) *Analysis of human genetic linkage*. The John Hopkins University Press.

Oudet C, Hanauer A, Clemens P, Caskey T, Mandel JL (1992) Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. *Hum Mol Genet* 1:599-603.

Pal C, Hurst LD (2003) Evidence for co-evolution of gene order and recombination rate. *Nat Genet* 33:392-5.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, et al (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719-23.

Petes TD (2001) Meiotic recombination hot spots and cold spots. *Nat Rev Genet* 2:360-9.

Petes TD, Merker JD (2002) Context Dependence of Meiotic Recombination hotspots in yeast: the relationship between recombination activity of a reporter construct and base composition. *Genetics* 162:2049-2052.

Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, *et al* (2003) Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 33:382-7.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, *et al* (2001) Linkage disequilibrium in the human genome. *Nature* 411:199-204.

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516-7.

Remm M, Metspalu A (2002) High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model. *Curr Opin Chem Biol* 6:24-30.

Robzyk K, Recht J, Osley MA (2000) Rad6-dependent ubiquitination of histone H2B in yeast. *Science* 287:501-4.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552-69.

Saccone S, Caccio S, Kasuda J, *et al.* (1996) Identification of the gene-richest bands in human chromosomes. *Gene* 174:85-94.

Scherthan H, Weich S, Schwegler H, Heyting C, Harle M, Cremer T (1996) Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. *J Cell Biol* 134:1109-25.

Shi Q, Spriggs E, Field LL, Ko E, Barclay L, Martin RH (2001) Single sperm typing demonstrates that reduced recombination is associated with the production of aneuploid 24,XY human sperm. *Am J Med Genet* 99:34-8.

Shi Q, Spriggs E, Field LL, Rademaker A, Ko E, Barclay L, Martin RH (2002) Absence of age effect on meiotic recombination between human X and Y chromosomes. *Am J Hum Genet* 71:254-61.

Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A (2003) Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 12:771-6.

Shumaker JM, Metspalu A, Caskey CT (1996) Mutation detection by solid phase primer extension. *Hum Mutat* 7:346-54.

Stankiewicz P, Lupski JR (2002) Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* 12:312-9.

Strachan T, Read AP (1999) *Human Molecular Genetics*. John Wiley & Sons

- Struhl K** (1998) Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* 12:599-606.
- Zhang K, Akey JM, Wang N, Xiong M, Chakraborty R, Jin L** (2003) Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. *Hum Genet* 113:51-59.
- Tease C, Hartshorne GM, Hulten MA** (2002) Patterns of meiotic recombination in human fetal oocytes. *Am J Hum Genet* 70:1469-79.
- Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF** (2000) Recombinational and mutational hotspots within the human lipoprotein lipase gene. *Am J Hum Genet* 66:69-83.
- Tõnisson N, Kurg A, Kaasik K, Lõhmussaar E, Metspalu A** (2000) Unravelling genetic data by arrayed primer extension. *Clin Chem Lab Med* 38:165-70.
- Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, Peltonen L** (2003) The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum Mol Genet* 12:51-9.
- Wall JD, Pritchard JK** (2003) Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 4:587-97.
- Wang N, Akey JM, Zhang K, Chakraborty R, Jin L** (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am J Hum Genet* 71:1227-34.
- Weir BS** (1979) Inferences about linkage disequilibrium. *Biometrics* 35:235-54
- Yauk CL, Bois PR, Jeffreys AJ** (2003) High-resolution sperm typing of meiotic recombination in the mouse MHC Ebeta gene. *Embo J* 22:1389-97.
- Yip SP, Lovegrove JU, Rana NA, Hopkinson DA, Whitehouse DB** (1999) Mapping recombination hotspots in human phosphoglucomutase (PGM1). *Hum Mol Genet* 8:1699-706.
- Yu A, Zhao C, Fan Y, Jeng W, Mungall AJ, Deloukas P, *et al.*** (2001) Comparison of human genetic and sequence-based physical maps. *Nature* 409:951-953.
- Yu J, Lazzeroni L, Qin J, Huang MM, Navidi W, Erlich H, Arnheim N** (1996) Individual variation in recombination among human males. *Am J Hum Genet* 59:1186-92.

Lisad

1. Rekombinatsioonipunktide asukohad inimese 22. kromosoomil.
2. Rekombinatsioonipunktide asukohad inimese 19. kromosoomil.
3. Geeniala protsendi ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.
4. Alu korduste ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.
5. GC% ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.

Lisa 1. Rekombinatsioonipunktide asukohad inimese 22. kromosoomil.

Ristsiirde punktid on kirjeldatud alguse ja lõpu asukohaga kromosoomis 22 ning markeri järjekorra-numbriga CRI-MAP programmi parameetrite failis. Pikkus näitab lokaliseerimise täpsust aluspaarides. Indiviidi koodis on kirjas perekonna kood, indiviidi kood ning ristsiirde päritolu ("M" puhul emapoolne, "P" puhul isapoolne).

Jrk	Algus		Lõpp		Pikkus	Indiviidi kood	Ristsiirete
Nr.	marker	bp	marker	bp	bp		arv
1	3	14562673	6	14695214	132541	1413-12115-M	1
2	4	14569744	16	15502643	932899	13291-7437-M	1
3	6	14695214	20	15683916	988702	1347-11875-M	1
4	13	15061627	14	15444674	383047	13291-6996-P	1
5	22	15699058	23	15713557	14499	1347-11875-M	1
6	26	15809478	28	15877010	67532	13291-6980-M	1
7	26	15809478	35	16064963	255485	1347-11875-P	1
8	32	15974398	40	16217285	242887	13291-7042-P	1
9	33	16008029	35	16064963	56934	1331-7033-M	1
10	37	16084489	39	16159854	75365	13291-7047-M	1
11	39	16159854	44	16406098	246244	13291-7046-M	1
12	45	16413340	46	16421717	8377	1416-12244-M	1
13	46	16421717	52	16579984	158267	13291-7437-M	1
14	55	16604328	68	17408794	804466	1416-12244-M	1
15	63	16975969	64	17110317	134348	13291-7046-P	1
16	64	17110317	68	17408794	298477	13291-7046-P	1
17	84	17959226	85	18007006	47780	1347-11875-P	1
18	145	19731638	156	20405419	673781	1347-11875-P	1
19	149	19771848	153	20268492	496644	1416-12245-P 1416-12244-P	2
20	154	20301488	157	20453769	152281	1416-12245-P 1416-12244-P	2
21	167	20841759	172	21050926	209167	1347-11875-P	1
22	168	20873046	169	20881216	8170	1413-12115-M 1413-12113-M 1413-12110-M	3
23	170	21028512	187	21588262	559750	13291-7046-P	1
24	173	21075769	179	21246682	170913	13291-6998-P 13291-7694-P	2
25	173	21075769	197	21951923	876154	13291-7047-P	1
26	174	21108637	175	21163978	55341	1413-12115-M 1413-12113-M 1413-12110-M	3
27	174	21108637	188	21625884	517247	1331-11818-M	1
28	179	21246682	197	21951923	705241	13291-6998-P 13291-7694-P	2
29	189	21680305	191	21793821	113516	1416-12244-M	1
30	205	22170101	208	22215228	45127	1416-12245-M	1
31	206	22178836	216	22441140	262304	13291-7694-P	1
32	223	22544912	237	22965757	420845	1347-11875-P	1
33	227	22658419	228	22730046	71627	13291-7694-M	1
34	239	23043639	242	23131358	87719	1347-11875-P	1
35	241	23111352	249	23268212	156860	13291-7434-P 13291-6995-P	2
36	246	23217575	249	23268212	50637	13291-7047-P	1
37	247	23229842	249	23268212	38370	1331-7005-M	1
38	249	23268212	252	23320222	52010	1331-7005-M 13291-7047-P 13291-7434-P 13291-6995-P	4
39	255	23470307	262	23568614	98307	1362-11984-M	1
40	258	23519118	260	23548474	29356	13291-6980-P	1
41	283	23915832	289	24037924	122092	13291-7042-P	1
42	288	23990793	309	24546324	555531	1347-11875-P	1
43	289	24037924	290	24070453	32529	13291-7042-P 13291-6995-P	2
44	291	24075427	293	24092647	17220	13291-7434-P	1
45	293	24092647	294	24112805	20158	13291-7434-P	1
46	294	24112805	295	24121721	8916	13291-7042-P	1
47	296	24138988	299	24211215	72227	13291-7042-P	1
48	308	24506321	322	24810717	304396	13291-7060-P	1
49	336	24967488	337	24976103	8615	1347-11875-P	1
50	341	25053559	343	25083171	29612	1347-11875-P	1
51	354	25287992	355	25375853	87861	1331-7005-P	1

Lisa 1 järg

Jrk	Algus		Löpp		Pikkus	Indiviidi kood	Ristsiirete
Nr.	marker	bp	marker	bp	bp		arv
52	359	25469276	362	25514328	45052	1347-11875-P	1
53	361	25497380	382	25836779	339399	13291-6998-M	1
54	367	25577907	373	25679902	101995	1362-11984-P	1
55	417	26462560	440	26955341	492781	13291-6980-M	1
56	419	26495244	436	26882972	387728	13291-6998-M	1
57	458	27317737	471	27786176	468439	13291-6998-M	1
58	500	28372286	527	28963664	591378	1347-11875-P	1
59	516	28699474	524	28902325	202851	13291-7060-P	1
60	542	29375269	559	29854326	479057	13291-7013-P	1
61	546	29475873	552	29671306	195433	1347-11875-P	1
62	558	29839031	578	30738683	899652	1347-11875-P	1
63	587	30931344	605	31270873	339529	1347-11875-P	1
64	607	31303634	608	31326012	22378	1347-11875-P	1
65	616	31483170	620	31502332	19162	1416-12245-P 1416-12244-P	2
66	642	31750883	646	31833550	82667	1416-12245-P 1416-12244-P	2
67	670	32477453	677	32710637	233184	1347-11875-P	1
68	676	32613531	693	33000457	386926	1416-12245-P 1416-12244-P	2
69	723	33631761	733	33971415	339654	1413-12110-M	1
70	730	33872177	740	34130930	258753	1347-11875-P	1
71	737	34058262	738	34073692	15430	1416-12245-P 1416-12244-P	2
72	765	34914183	768	34966146	51963	1413-12115-M	1
73	771	35005751	780	35188680	182929	1347-11875-P	1
74	841	36455107	842	36470860	15753	1347-11875-P	1
75	888	37559771	905	37908353	348582	13291-7013-M	1
76	904	37889240	906	37936158	46918	1347-11875-P	1
77	908	38001292	926	38308361	307069	1413-12113-M	1
78	950	38916724	954	38972653	55929	1347-11875-P	1
79	1023	41060127	1027	41167616	107489	1347-11875-P	1
80	1023	41060127	1035	41376362	316235	13291-7013-P	1
81	1058	41759181	1060	41810024	50843	13291-7060-M	1
82	1097	42547972	1103	42686267	138295	1347-11875-P	1
83	1098	42572149	1102	42676270	104121	13291-7694-M	1
84	1118	42923959	1126	43079835	155876	1347-11875-P	1
85	1119	42931066	1120	42940132	9066	13291-7046-P	1
86	1120	42940132	1123	43029917	89785	13291-7046-P	1
87	1160	43687327	1163	43728875	41548	13291-7046-M	1
88	1160	43687327	1168	43809232	121905	1347-11875-P	1
89	1163	43728875	1167	43794362	65487	1416-12244-M	1
90	1193	44255484	1217	44953848	698364	13291-6980-M	1
91	1203	44459358	1205	44506285	46927	1347-11875-P	1
92	1213	44765056	1216	44933162	168106	13291-7014-P	1
93	1216	44933162	1222	45066639	133477	13291-7014-P	1
94	1250	45553486	1258	45659590	106104	1347-11875-P	1
95	1296	46491638	1307	46762198	270560	1416-12245-M 1347-11875-P	2
96	1326	46924845	1327	46932241	7396	1347-11875-P	1
97	1346	47135364	1347	47157302	21938	1416-12245-P 1416-12244-P	2
98	1348	47171544	1350	47241223	69679	1416-12245-P	1
99	1348	47171544	1352	47258849	87305	1416-12244-P	1
100	1365	47551905	1366	47608698	56793	1416-12245-P	1
101	1381	47877046	1385	47948253	71207	1331-7033-M	1
102	1381	47877046	1391	48048454	171408	13291-6996-M	1
103	1386	47966795	1387	47977670	10875	13291-7694-P	1
104	1391	48048454	1406	48233705	185251	1416-12245-P	1
105	1419	48368029	1434	48652927	284898	1347-11875-P	1
106	1438	48721441	1444	48870263	148822	1416-12245-P	1
107	1445	48879147	1446	48887332	8185	1416-12244-P	1
108	1451	49080716	1453	49089926	9210	1416-12244-P	1
109	1466	49160661	1468	49178502	17841	1347-11875-P	1

Lisa 2. Rekombinatsioonipunktide asukohad inimese 19. kromosoomil.

Ristsiirde punktid on kirjeldatud alguse ja lõpu asukohaga kromosoomis 19 ning markeri järjekorra-numbriga CRI-MAP programmi parameetrite failis. Pikkus näitab lokaliseerimise täpsust aluspaarides. Indiviidi koodis on kirjas perekonna kood, indiviidi kood ning ristsiirde päritolu ("M" puhul emapoolne, "P" puhul isapoolne).

Jrk	Algus		Lõpp		Pikkus	Indiviidi kood	Ristsiirde
Nr.	marker	bp	marker	bp	bp		arv
1	2	232583	10	1382042	1149459	1423-8-M	1
2	3	232901	27	2497610	2264709	1408-8-M	1
3	10	1382042	22	2195336	813294	1341-8-M	1
4	48	2781360	61	2953227	171867	1408-7-M	1
5	67	3058119	74	3089990	31871	1347-8-M	1
6	87	3149661	94	3557549	407888	1362-8-M	1
7	89	3215315	115	3973718	758403	1423-7-P	1
8	105	3736849	116	4009426	272577	1340-8-M	1
9	106	3775177	109	3906520	131343	1347-7-M	1
10	125	4228332	136	4393325	164993	1347-8-P	1
11	136	4393325	150	5031933	638608	1347-7-P	1
12	178	5477004	216	6632796	1155792	1340-7-M	1
13	194	6251818	201	6585524	333706	1362-7-M	1
14	197	6526268	201	6585524	59256	1362-8-P	1
15	204	6589564	206	6589732	168	1362-8-P 1362-7-M	2
16	215	6625817	219	6633885	8068	1347-7-M	1
17	220	6634247	221	6637406	3159	1347-7-M	1
18	278	7586455	283	7660256	73801	1408-8-P	1
19	315	7988255	318	8068882	80627	1331-8-P 1331-7-P	2
20	327	8216959	329	8297593	80634	1331-8-P 1331-7-P	2
21	328	8284510	330	8634439	349929	1362-7-M	1
22	331	8648236	332	8761169	112933	1341-7-P	1
23	355	8915403	359	9062609	147206	1331-8-M	1
24	358	8922645	369	9190884	268239	1334-8-P	1
25	464	11289340	472	11724408	435068	1347-7-P	1
26	474	11725212	484	12020592	295380	1347-7-P	1
27	474	11725212	511	13016633	1291421	1347-8-M	1
28	511	13016633	516	13121554	104921	1347-8-M	1
29	516	13121554	524	13384811	263257	1362-7-P	1
30	530	13435621	544	13499228	63607	1347-8-P 1347-7-P	2
31	545	13499276	546	13512379	13103	1347-7-P	1
32	546	13512379	547	13545203	32824	1347-8-P	1
33	559	13730693	566	14078943	348250	1331-8-P 1331-7-P	2
34	561	13797039	574	14585453	788414	1416-7-P	1
35	564	13917943	574	14585453	667510	1423-8-P	1
36	565	14028786	568	14140293	111507	1347-7-P	1
37	566	14078943	568	14140293	61350	1362-8-P	1
38	570	14221781	575	14589111	367330	1347-7-P	1
39	632	15546311	654	15671651	125340	1341-8-M	1
40	652	15622264	653	15642741	20477	1341-8-P	1
41	655	15684377	657	15688663	4286	1341-8-M	1
42	659	15695208	661	15699331	4123	1341-8-P	1
43	713	16127726	744	16913733	786007	1334-7-P	1
44	730	16457220	732	16503130	45910	1341-8-P	1
45	734	16511066	737	16545022	33956	1341-8-P	1
46	746	16994677	768	17802518	807841	1340-7-P	1
47	757	17386226	769	17889471	503245	1347-7-P	1
48	758	17392064	763	17608601	216537	1408-7-P	1
49	823	18864075	824	18880626	16551	1331-7-P	1

Lisa 2 järg

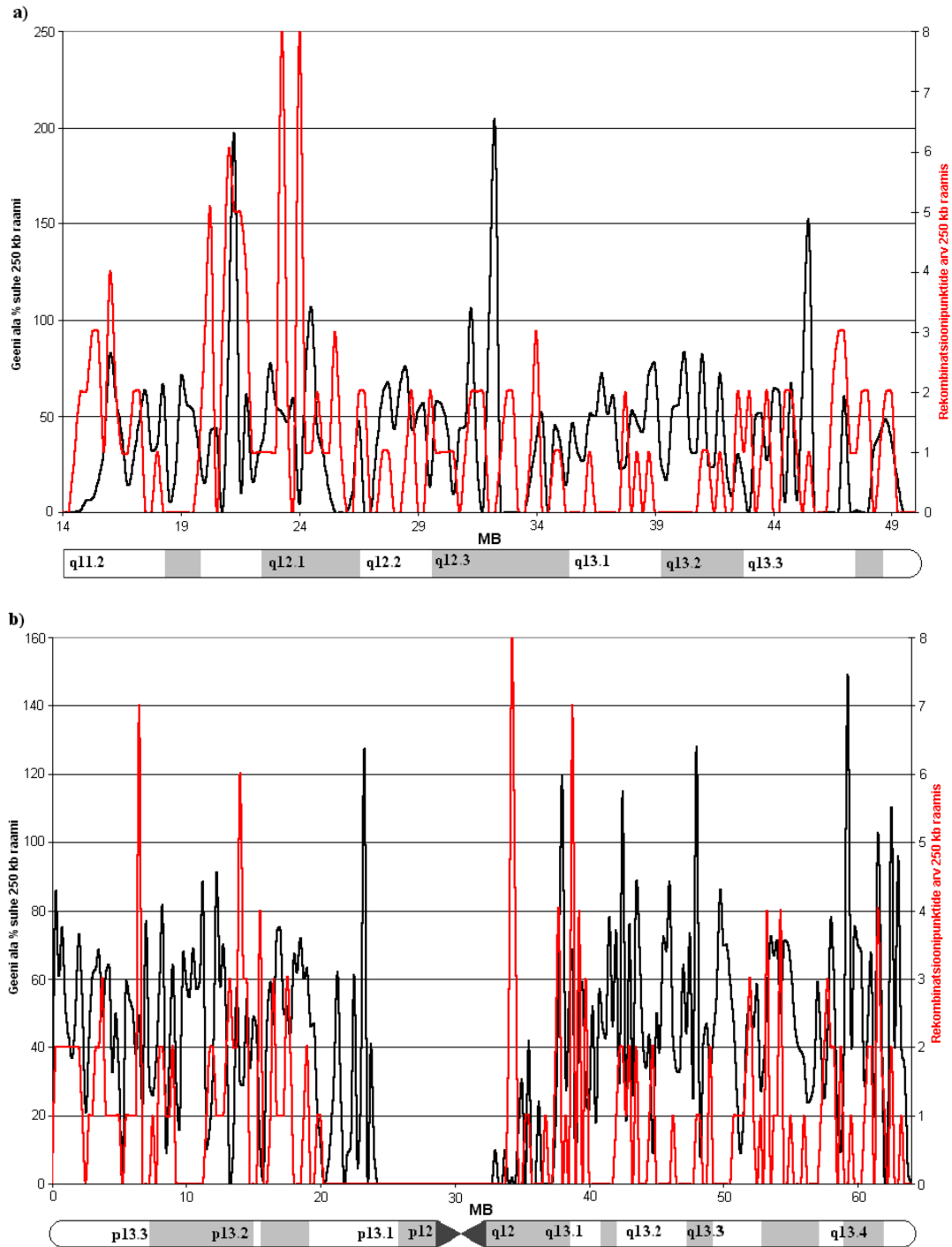
Jrk	Algus		Lõpp		Pikkus	Indiviidi kood	Ristsiirete
Nr.	marker	bp	marker	bp	bp		arv
50	832	19186369	833	19223880	37511	1331-7-P	1
51	832	19186369	835	19232565	46196	1347-8-P	1
52	854	19561003	873	20078400	517397	1347-8-P	1
53	1319	34039448	1331	34074770	35322	1347-8-M	1
54	1338	34084446	1354	34159046	74600	1347-8-M	1
55	1372	34246224	1404	34470671	224447	1347-8-M	1
56	1384	34329338	1390	34414690	85352	1362-8-P 1362-7-P	2
57	1391	34430380	1392	34436712	6332	1362-8-P 1362-7-P	2
58	1394	34436721	1397	34446728	10007	1347-8-P 1347-7-P	2
59	1400	34456787	1404	34470671	13884	1347-7-P	1
60	1430	34546185	1431	34554485	8300	1347-7-P	1
61	1430	34546185	1435	34565647	19462	1347-8-M	1
62	1434	34565357	1435	34565647	290	1347-8-P 1347-7-P	2
63	1479	35001305	1489	35271413	270108	1333-7-P	1
64	1509	35402083	1517	35503420	101337	1341-7-P	1
65	1647	36712891	1651	36814133	101242	1331-7-P	1
66	1728	37389691	1754	37762140	372449	1416-8-M 1416-7-M	2
67	1757	37782000	1765	37808444	26444	1416-8-M 1416-7-M	2
68	1786	38086437	1795	38432288	345851	1334-8-P	1
69	1829	38808471	1840	38883202	74731	1362-7-P	1
70	1841	38886498	1843	38907089	20591	1341-8-P	1
71	1851	38921990	1852	38924941	2951	1341-8-P	1
72	1851	38921990	1854	38937038	15048	1408-8-M	1
73	1856	38941927	1858	38957745	15818	1408-7-P	1
74	1860	38967044	1864	38999862	32818	1408-8-M	1
75	1863	38999687	1864	38999862	175	1408-7-P	1
76	1893	39298122	1898	39340525	42403	1347-8-M 1347-7-M	2
77	1899	39340737	1900	39343724	2987	1347-8-M 1347-7-M	2
78	1924	39745586	1928	39815627	70041	1340-8-P	1
79	1926	39761921	1927	39802811	40890	1341-7-P	1
80	1927	39802811	1929	39833954	31143	1341-7-P	1
81	2016	42004778	2074	43201847	1197069	1347-7-M	1
82	2037	42157202	2038	42164246	7044	1341-8-M	1
83	2039	42164548	2043	42541414	376866	1341-8-M	1
84	2077	43203426	2079	43216037	12611	1347-7-M	1
85	2094	43366055	2101	43669687	303632	1340-7-P	1
86	2102	43674855	2103	43692669	17814	1340-7-P	1
87	2141	44321711	2167	44799132	477421	1341-8-P	1
88	2168	44807050	2169	44807869	819	1341-8-P	1
89	2280	46242135	2288	46380353	138218	1408-8-P	1
90	2360	47805493	2374	48486076	680583	1416-8-P	1
91	2436	49082176	2437	49082340	164	1341-8-M	1
92	2438	49082370	2439	49083536	1166	1341-8-M	1
93	2564	50934073	2565	50938335	4262	1347-7-P	1
94	2566	50985252	2567	51092787	107535	1347-7-P	1
95	2587	51249430	2607	51657992	408562	1341-8-M	1
96	2612	51788697	2617	51838251	49554	1341-8-M	1
97	2619	51841035	2627	52182818	341783	1347-7-M	1
98	2620	51852929	2624	51963988	111059	1341-8-P	1
99	2625	51964694	2628	52225204	260510	1341-8-P	1
100	2625	51964694	2633	52337542	372848	1347-8-P	1
101	2641	52581708	2652	52939729	358021	1347-7-M	1
102	2659	53239555	2668	53315742	76187	1341-8-P	1
103	2659	53239555	2690	53459264	219709	1416-7-P	1
104	2669	53317595	2673	53334953	17358	1341-8-P	1
105	2685	53395435	2689	53440742	45307	1333-8-P	1
106	2694	53549032	2707	53924751	375719	1333-7-M	1
107	2712	53986638	2725	54250146	263508	1408-8-M	1
108	2723	54194505	2725	54250146	55641	1347-7-P	1
109	2725	54250146	2726	54276613	26467	1408-8-M	1
110	2725	54250146	2729	54443164	193018	1347-7-P	1
111	2744	54906818	2747	55039196	132378	1347-7-M	1
112	2801	56123975	2806	56158480	34505	1340-8-M	1

Lisa 2 järg

Jrk	Algus		Lõpp		Pikkus	Indiviidi kood	Ristsiirete
Nr.	marker	bp	marker	bp	bp		arv
113	2878	57079537	2893	57363709	284172	1340-7-P	1
114	2904	57535849	2908	57561030	25181	1416-8-M 1416-7-M	2
115	2944	57736173	2951	57792081	55908	1416-8-M 1416-7-M	2
116	2954	57865593	2957	57913463	47870	1362-7-P	1
117	2960	57986365	2965	58089846	103481	1362-7-P	1
118	2964	58057450	2967	58096379	38929	1362-8-M	1
119	2993	58312037	3001	58488180	176143	1341-8-M	1
120	2995	58321312	2996	58435292	113980	1347-7-P	1
121	3018	58782568	3021	58897162	114594	1331-7-P	1
122	3022	58899141	3023	58975035	75894	1408-7-M	1
123	3040	59517419	3041	59528631	11212	1331-8-P	1
124	3101	60264432	3104	60664370	399938	1423-8-M	1
125	3106	60696660	3128	61174814	478154	1334-8-M	1
126	3108	60747137	3121	61127070	379933	1334-7-P	1
127	3149	61444236	3156	61578478	134242	1347-8-M 1347-7-M	2
128	3155	61564122	3156	61578478	14356	1416-8-M 1416-7-M	2
129	3173	61757963	3174	61763972	6009	1416-8-M 1416-7-M	2
130	3206	62043226	3207	62112417	69191	1333-8-M	1
131	3222	62498390	3232	62663926	165536	1347-8-M 1347-7-M	2
132	3251	63112062	3256	63281018	168956	1416-8-M	1

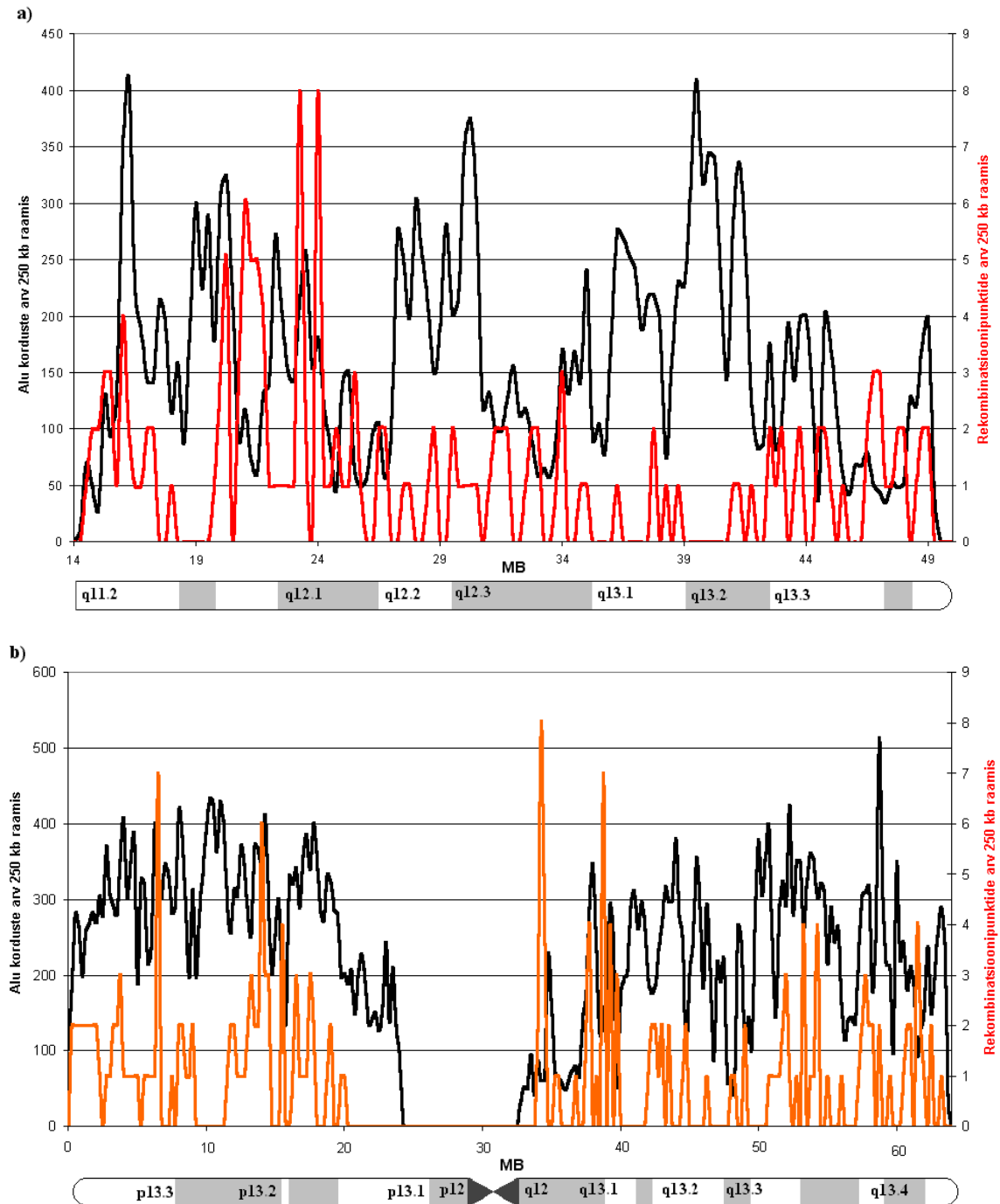
Lisa 3. Geeniala % ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.

Geeniala % (must joon) ja rekombinatsioonide arvu (punane joon) võrdlus 250 kb aknas 22. (a) ja 19. (b) kromosoomis



Lisa 4. Alu korduste ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.

Alu korduste (must joon) ja rekombinatsioonide (punane joon) arvu võrdlus 250 kb aknas 22. (a) ja 19. (b) kromosoomis.



Lisa 5. GC% ja rekombinatsioonide arvu võrdlus 250 kb-s aknas.

GC% (must joon) ja rekombinatsioonide (punane joon) arvu võrdlus 250 kb aknas 22. (a) ja 19. (b) kromosoomis

