

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

LIINA NAGIRNAJA

Inimese luteiniseeriva hormooni ja
kooriongonadotropiini beeta-subühiku geeniperekond:
struktuur ja potentsiaalne seotus korduvate
spontaanabortidega

Magistritöö

Juhendaja: Erakorraline professor Maris Laan, Ph.D.

TARTU
2005

Sisukord

Sisukord	2
Lühendid ja mõisted	4
Sissejuhatus	6
1. Kirjanduse ülevaade	7
1.1. Rekombinatsioon	7
1.2. Geenikonversioon	9
1.3. Rekombinatsioon ja alleelne aheldatus	13
1.4. Rekombinatsiooni <i>hotspot</i> 'd	15
1.4.1. Rekombinatsiooni hotspot'i detekteerimine	15
1.4.2. Rekombinatsiooni hotspot'i iseloomustus	16
1.4.3. Rekombinatsiooni hotspot ja palindroomid	19
1.5. Segmentaalsed duplikatsioonid ja genoomsed ümberkorraldused	21
1.5.1. Segmentaalsed duplikatsioonid	21
1.5.2. Segmentaalsete duplikatsioonide vahendatud ümberkorraldused	22
1.5.3. Segmentaalsete duplikatsioonide evolutsioon	24
1.6. Gonadotropiinid	26
1.5.1. Gonadotropiinide funktsioonid	27
1.5.2. Gonadotropiinide struktuur	28
1.5.3. Gonadotropiinide subühikuid kodeerivate geenide evolutsioon	29
1.6. Korduvad spontaanabordid	33
1.6.1. Üldiseloomustus	33
1.6.2. LHB/CGB geeniperekond ja spontaanabordid	34
1.6.2.1. Mutatsioonid	34
1.6.2.2. Struktuursed ümberkorraldused	35
2. Töö eesmärgid	36
3. Materjal ja meetodika	37
3.1. <i>In silico</i> analüüs	37
3.2. Analüüsi materjal	37
3.2.1. Rekombinatsiooni hotspot'i analüüsi materjal	37
3.2.2. Patsiendigrupp	37
3.2.3. Referentsvalim	38
3.3. DNA eraldamine	39
3.3.1. Vereproovid	39
3.3.2. Platsenta materjal	39
3.4. Praimerite disain	40
3.4.1. LHB/CGB geenid	40
3.4.2. Potentsiaalse rekombinatsiooni hotspot'i piirkond ning CGB8	40
3.5. Amplifikatsiooni protsess	42
3.5.1. LHB/CGB geenide amplifikatsioon	42
3.5.2. Potentsiaalse rekombinatsiooni hotspot'i piirkonna amplifikatsioon	43
3.6. Resekveneerimine	45
3.6.1. PCR produktide töötlus	45
3.6.2. Sekveneerimine	45
3.7. DNA järjestuste analüüs ja polümorfismide identifitseerimine	46
3.8. Statistiline analüüs	47
3.8.1. Markerite ning uuritava piirkonna varieeruvuse üldiseloomustus	47
3.8.2. Alleelse aheldatuse defineerimine	47

3.8.3. Rekombinatsiooni määra defineerimine ja haplotüüpide ennustamine	48
4. Tulemused.....	49
4.1. <i>LHB/CGB</i> geeniklastri detailne struktuur	49
4.1.1. <i>LHB/CGB</i> geeniklastri üldiseloomustus.....	49
4.1.2. Palindroomide ning sekundaarsete struktuuride ennustamine	52
4.2. Potentsiaalse rekombinatsiooni <i>hotspot</i> 'i ning <i>CGB8</i> iseloomustamine	53
4.2.1. Hotspot'i piirkonna amplifitseerimine ning sekveneerimine.....	53
4.2.2. Hotspot'i piirkonna polümorfismide jaotuvus	57
4.2.3. Haplotüüpide jaotuvus	59
4.2.4. Rekombinatsioonilise aktiivsuse analüüs	60
4.3. <i>LHB/CGB</i> geenide varieeruvus korduvate spontaanabortidega patsientidel <i>versus</i> referentspopulatsioonil	63
4.3.1. <i>LHB/CGB</i> geenide amplifitseerimine ning sekveneerimine	63
4.3.2. <i>LHB/CGB</i> geenide üldine varieeruvus patsientidel ja referentsvalimil	64
4.3.3. Haplotüüpide analüüs.....	67
5. Arutelu	71
5.1. <i>LHB/CGB</i> geeniklastri struktuur ja evolutsioon	71
5.2. Rekombinatsiooni <i>hotspot</i>	72
5.3. <i>LHB/CGB</i> geenide assotsiatsioon korduvate spontaanabortide esinemisega ...	74
Kokkuvõte.....	77
Summary	78
Tänuavaldused	79
Kasutatud kirjandus	80
Lisad.....	90

Lühendid ja mõisted

A	adeniin
Ah	aminohape
Bp	aluspaari
C	tsütosiin
CGB, CG β	kooriongonadotropiini β -subühiku geen
dATP	desoksüadenosiintrifosfaat
dCTP	desoksütsütidintrifosfaat
dGTP	desoksüguanosiintrifosfaat
DMSO	dimetüülsulfoksiid
DSB	kaheaheelaline lõige (<i>Double-Strand Break</i>)
DSBR	kaheaheelalise lõike parandus (<i>Double-Strand Break Repair</i>)
dTTP	desoksütümidintrifosfaat
EDTA	Etüleendiaminotetraädikhape
EL	nn loote hukk (<i>embryo loss</i>)
Exo I	eksonukleas I
FISH	<i>Fluorescence In Situ Hybridization</i>
FL	loote hukk (<i>fetal loss</i>)
FSH	folliikuleid stimuleeriv hormoon
G	guaniin
hCG	inimese kooriongonadotropiin
hCG β	inimese kooriongonadotropiini β -subühik
HW	Hardy-Weinbergi tasakaal (<i>Hardy Weinberg Equilibrium</i>)
Indel	insertsioon, deletsioon
Kb	1000 aluspaari
LCR	segmentaalne duplikatsioon (<i>Low-Copy Repeat</i>)
LD	ahelduse tasakaalustamatus (<i>Linkage Disequilibrium</i>)
LH	luteiniseeriv hormoon
LHB, LH β	luteiniseeriva hormooni β -subühiku geen
LH β	luteiniseeriva hormooni β -subühik
M	molaarsus
Mat	miljonit aastat tagasi
Mb	1000 kb
MIR	<i>Mammalian wide interspread repeat</i>
MHC	Suur koesobivuskompleks (<i>Major Histocompatibility Complex</i>)
Min	minut
μ l	mikroliiter
NAHR	Mitte-alleelne homoloogiline rekombinatsioon (<i>non-allelic homologous recombination</i>)
ng	nanogramm
NHEJ	mitte-homoloogiline otste seondumine (<i>non-homologous end joining</i>)
nt	nukleotiid
NTF	neurotrofiin
OD	optiline tihedus (<i>Optical Density</i>)
p	kromosoomi lühike õlg
PCR	polümeraasi ahelreaktsioon (<i>Polymerase Chain Reaction</i>)
ps	pseudogeen
q	kromosoomi pikk õlg

rpm	pööret minuti kohta
SA	spontaanabort
sAP	kreveti aluselin fosfataas (<i>Shrimp Alkaline Phosphatase</i>)
sek	sekund
SDSA	süntees-sõltuva ahela seondumine (<i>synthesis-dependent strand-annealing</i> , SDSA)
SNP	ühenukleotiidne polümorfism (<i>Single Nucleotide Polymorphism</i>)
<i>Taq</i>	<i>Thermus aquaticus</i> 'e DNA polümeraas
TSH	türoid stimuleeriv hormoon
<i>TSHB</i>	türoid stimuleeriva hormooni β -subühikut kodeeriv geen
<i>TSHβ</i>	türoid stimuleeriva hormooni β -subühik
T	tümiin
UTR	mittetransleeritav regioon (<i>untranslated region</i>)
U	ühik

Sissejuhatus

Korduvad spontaanabordid esinevad populatsioonis 1-2% fertiilsetel naistel. Vaatmata paljudele kliinilistele ja laboratoorsele uuringutele, ei osata siiani 50% juhtudest määratleda konkreetset raseduse katkemise põhjust. Üheks potentsiaalseks riskifaktoriks võib olla kooriongonadotropiini hormooni (hCG) düsfunktsionaalsus.

Kooriongonadotropiin on heterodimeerne glükoproteiin, mille esinemise põhjal ema seerumis diagnoositakse rasedust ning selle potentsiaalset kulgu. Arvukate uuringute tulemusena on leitud hCG-l olevat kriitiline roll raseduse I trimestril. Seega, vähene hCG tase ema seerumis viitab häiretele hormooni struktuuri ja/või metabolismis, mis võivad olla põhjustatud mutatsioonidest hCG subühikuid kodeerivates geenides.

HCG hormooni β -subühikuid kodeerivad geenid asuvad ühtse geeniklastrina 19. kromosoomis. *LHB/CGB* geeniperekond sisaldab duplikatsioonisündmuste teel tekkinud kõrge sarnasusega segmente nii geenide kui ka intergeensete alade piirkonnas ning lisaks veel rekombinatsiooni *hotspot*'i. Seega on tõenäoliselt tegemist kromosomaalsete ümberkorralduste suhtes väga ebastabiilse regiooniga, mille tagajärjeks võivad olla mutatsioonid ning geenikoopia arvu muutused. Siiani pole *LHB/CGB* geeniklastri aga tuvastatud ühtegi kromosoomi struktuuri anomaaliat ega mutatsiooni, mis seostuks korduvate spontaanabortide kujunemisega. Põhjuseks võib olla kasutatavate meetodite limiteeritus nii keerulise piirkonna struktuuri analüüsiks.

Käesoleva töö raames püstitati eesmärk kirjeldada detailselt *LHB/CGB* geeniklastri struktuuri ning analüüsida rekombinatsiooni toimumise aktiivsust antud piirkonnas. Seejärel viidi läbi pilootprojekt, kus resekveneeritud *LHB/CGB* geenide varieeruvust vaadeldi võrdlevalt korduvate spontaanabortidega patsientidel ning referentspopulatsioonil. *LHB/CGB* geeniklastri kirjeldusel saadud info põhjal disainitakse edaspidi FISH uuring, mis võimaldab identifitseerida erinevaid potentsiaalseid ümberkorraldusi korduvate spontaanabortidega patsientidel *versus* kontrollpopulatsioonil.

1. Kirjanduse ülevaade

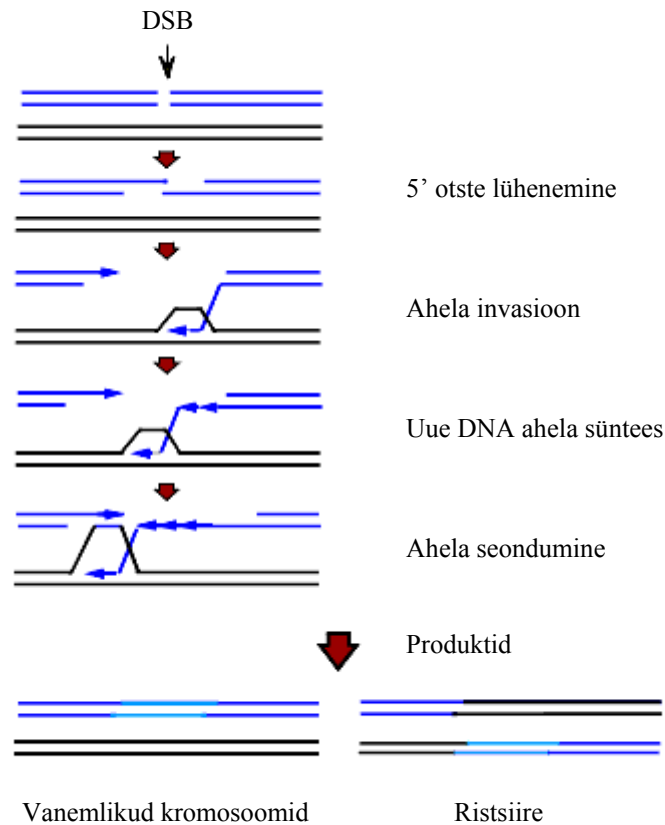
1.1. Rekombinatsioon

Meiootiline homoloogiliste kromosoomide rekombinatsiooni käigus leiab aset ristsiire, mis on peamiseks varieeruvuse genereerimise mehhanismiks populatsioonis indiviidide vahel. Rekombinatsiooni uuringud on pakkunud laialdast huvi nii reproduktsiooni aspektis kui ka evolutsioonilisest aspektist tänapäeval esinevate genoomide struktuuri kujundamisel.

Rekombinatsiooni tulemusena meioosis toimub homoloogiliste kromosoomide vahel ristsiire, mille tulemusena tekivad kas vanemlike kromosoomidega sarnased või uued rekombinantsed DNA molekulid (Joonis 1). Rekombinatsiooni toimumise eeltingimuseks on kaheaahelalise lõike (*double-strand break*, DSB) teke, mis initsieerib DSB paranduse raja (DSB *repair*, DSB_R). DSB_R käigus toimub homoloogiliste DNA ahelate vahel ristsiire ning moodustuvad rekombinantsed või vanemlikku tüüpi tütar-kromosoomid (Paques ja Haber, 1999). DSB-d on seega üheks olulisemaks determinandiks, mille alusel ennustada rekombinatsiooni poolest aktiivseid genoomi piirkondi. Seetõttu on praeguseks intensiivselt uuritud DSB tekkemehhanisme ning assotsiatsiooni rekombinatsiooni esinemissagedusega erinevate organismide genoomi põhjal, kuid ühtseid tulemusi on siiski vähe (de Massy, 2003).

Mudelorganismina DSB-de uurimisel on peamiselt kasutatud pärmi, mille puhul on leitud rekombinatsiooni aparatuuri ning mehhanismide sarnasust imetajate omaga. Tulemuste põhjal võib järeldada järgnevat (de Massy, 2003):

1. DSB-d ei esine genoomis juhuslikult, vaid on klasterdunud teatud piirkondadesse
2. Ei ole universaalset DNA järjestuse põhist motiivi, mille põhjal ennustada rekombinatsiooni esinemist
3. DSB-d esinevad peamiselt kromatiini kergesti ligipääsetavates regioonides.



Joonis 1. Rekombinatsiooni toimumise mehhanism. DSB – *double strand break*. (<http://www.bio.brandeis.edu/haberlab/jehsite/sdsa.html>).

DSB-de esinemist kergesti ligipääsetava kromatiini piirkondades seostatakse promootoritega, mis on avatuma struktuuriga ning tundlikumad DNA lõigete tekkimise suhtes. Alternatiivse võimalusena pakutakse spetsiifiliste transkriptsiooni faktorite osalust rekombinatsioonilise mehhanismi aktiveerimisel. Pärmi genoomi põhjal on leitud ka DSB-de esinemine kõrge GC-sisaldusega piirkondades, kuid antud tulemused on imetajate puhul leidnud vaid üksikuid tõendeid (Crawford et al., 2004, Jeffreys et al., 2000, de Massy, 2003). Ristsiirete analüüsil meiootilistes rakkudes on täheldatud, et DSB-d leiavad aset tunduvalt sagedamini kui detekteeritavad ristsiirded, mis viitab suurema osa DSB-de lahendamisele vanemlike kromosoomide tüüpi (suhtes 10:1) (Kauppi et al., 2004). Seega ei saa DSB-de esinemise põhjal otse ennustada vastava piirkonna ristsiirete toimumise intensiivsust.

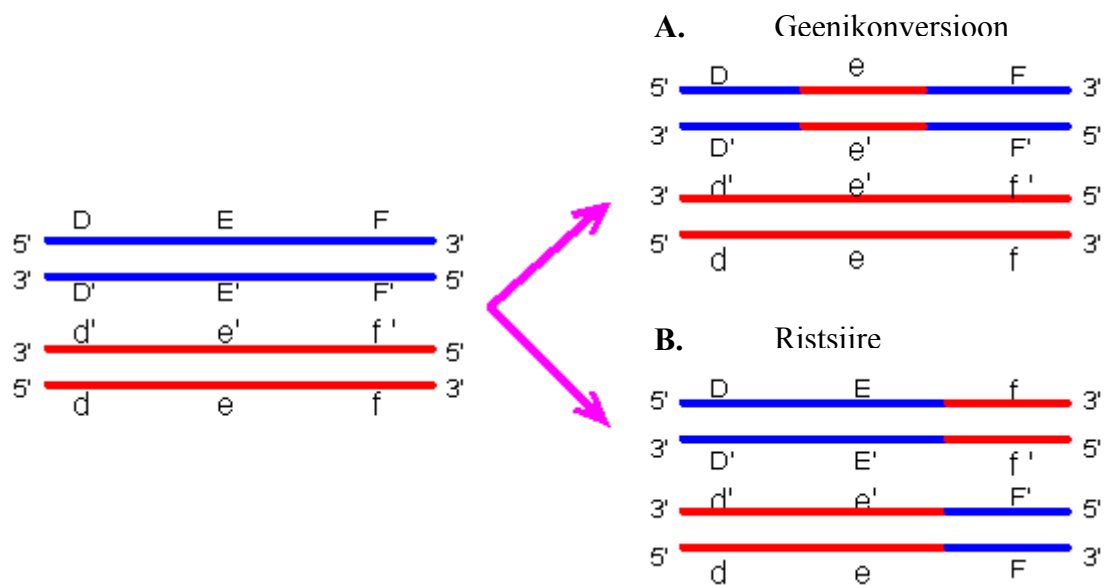
Inimese populatsiooni põhiste uuringute põhjal on defineeritud rekombinatsiooniliseks aktiivsuseks keskmiselt 1,1 cM/Mb (Kauppi et al., 2004). Leitud on varieeruvust nii erinevate indiviide tasemel kui ka sugude vahel, kus

keskmiselt kõrgem rekombinatsiooni tase on detekteeritud naistel (1,65 korda) (Tease et al., 2002, Yu et al., 1996, de Massay, 2003, Cullen et al., 2002). Samuti on uuritud vanuse seost rekombinatsiooni toimumise intensiivsusega, kuid tulemused on olnud negatiivsed nii meeste kui naiste gameetide puhul (Shi et al., 2002, Kong et al., 2004). Kong et al. töö põhjal ei korreleerunud rekombinatsiooniline aktiivsus vanusega, kuid samas leiti, et kõrge rekombinatsioonilise aktiivsusega naistel on suurem arv lapsi. Seega püstitati hüpotees, et kõrgem rekombinatsiooni tase naiste gameetides on reproduktiivseks eeliseks ning on positiivse selektsiooni all. Analoogilise tulemuseni on jõutud ka hiljutise Stefansson et al. (2005) poolt läbiviidud uuringu põhjal, kus leiti 17. kromosoomil 900 kb inversioon, millega assotsieerub kõrgem rekombinatsiooni tase ning suurem laste arv peres.

Üks huvitavamaid ning enim uuritud nähtus, mis on leitud rekombinatsiooni uuringutel, on ristsiirete toimumiskohtade klasterdumine nn rekombinatsiooni *hotspot*'dena. Nimelt kolokaliseeruvad DSB-d ning rekombinatsioonilised sündmused tihedalt samasse ~1-2 kb regiooni, mis piirneb vähe-rekombinatiivsete aladega (Kauppi et al., 2004). Kuna *hotspot*'del arvatakse olevat oluline roll genoomi struktuuri kujundamisel, siis on nende iseloomustamisele ning ennustamisele omistatud suurt tähelepanu.

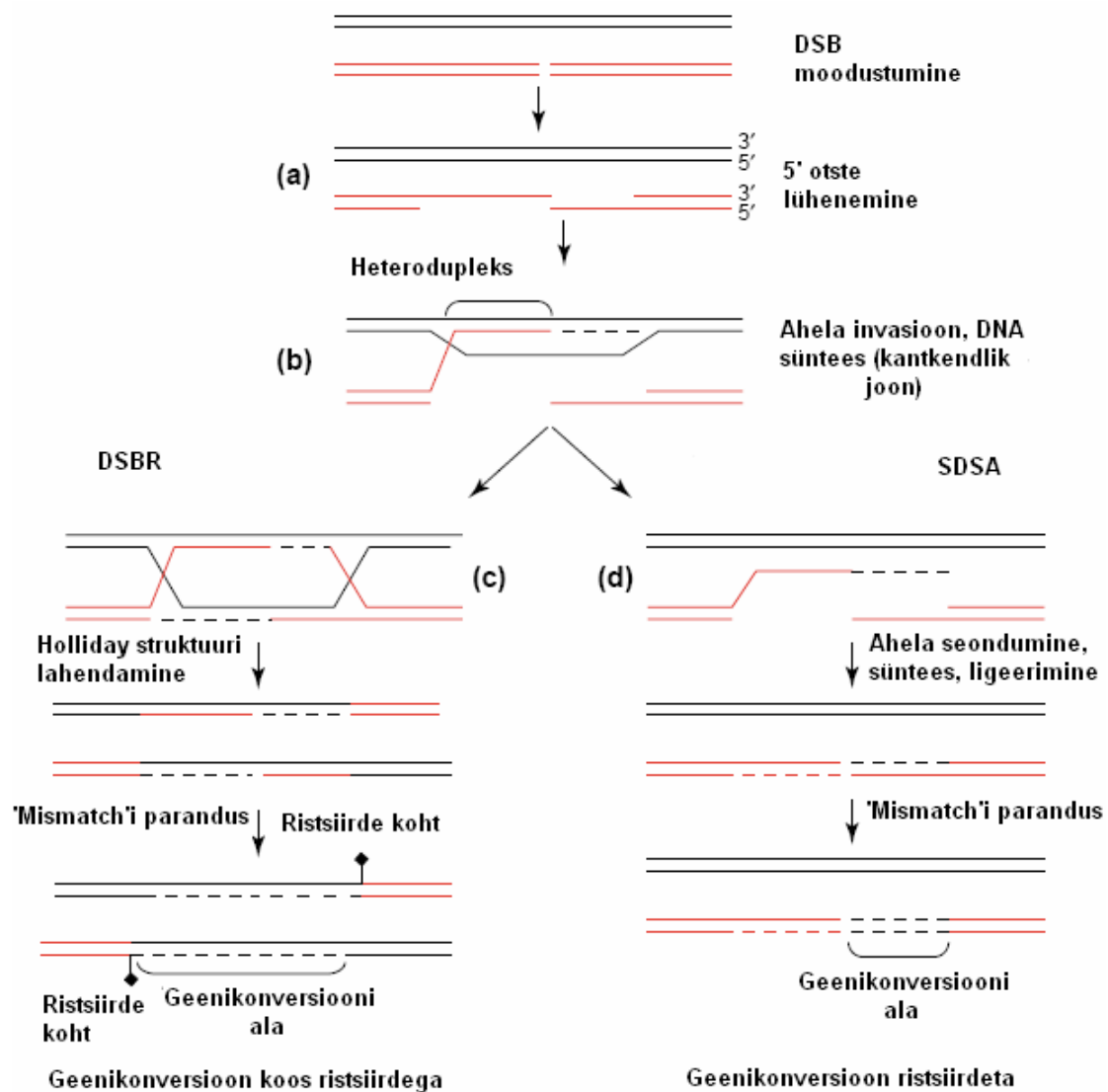
1.2. Geenikonversioon

Geenikonversiooniks nimetatakse nähtust, kus kahe kõrgelt homoloogse (>95%) järjestuse vahel toimub geneetilise info vahetamine ühesuunaliselt ilma ristsiirdeta (Joonis 2)(Papadakis ja Patrinos, 1999). Geenikonversiooni näol on tegemist olulise faktoriga geeniperekondade ning teiste kõrgelt sarnaste alade evolutsioneerimisel ning homogeniseerimisel.



Joonis 2. Geenikonversiooni võrdlus ristsiirdega. DEF, D'E'F', def, d'e'f' – erinevad markerid. (<http://www.web-books.com/MoBio/Free/Ch8D4.htm>)

Rekombinatsiooni *hotspot*'de uurimisel on leitud, et ristsiirded ning geenikonversiooni sündmused klasterduvad samasse piirkonda. Seetõttu toimub tõenäoliselt DNA segmendi konverteerimine *hotspot*'i alas DSBR mudeli järgi, kus tulemusena võivad tekkida nii ristsiire kui konversioon (Joonis 3). Alternatiivseks võimaluseks on süntees-sõltuva ahela seondumine (*synthesis-dependent strand-annealing*, SDSA), mille tulemuseks on ainult geenikonversioon (de Massy, 2003).



Joonis 3. Geenikonversiooni tekkemehhanismid (Allers ja Lichten, 2001). DSB – kahe ahela lõikamine (*double-strand break*), , DSBR – kaheaahelalise katkemise paranduse rada (*double-strand break repair*), SDSA - süntees-sõltuva ahela seondumine (*synthesis-dependent strand-annealing*).

Geenikonversiooni on enim kirjeldatud erinevate geeniperekondade põhjal, mis hõlmavad mitmeid duplikatsiooni teel tekkinud funktsionaalseid geene ning pseudogeene. Identifitseeritud konverteeritava ala pikkuseks on keskmiselt ~300 bp (Papadakis ja Patrinos, 1999, Zangenberg et al., 1995). Detekteeritava ala pikkus sõltub markerite tihedusest homoloogilistes piirkondades, mille alusel konverteeritava fragmendi piire määratleda. Geenikonversiooniga kaasneb paraloogete järjestuste homogeniseerimine materjali ülekandumise tõttu. Samas ühe järjestuse siseselt tõuseb

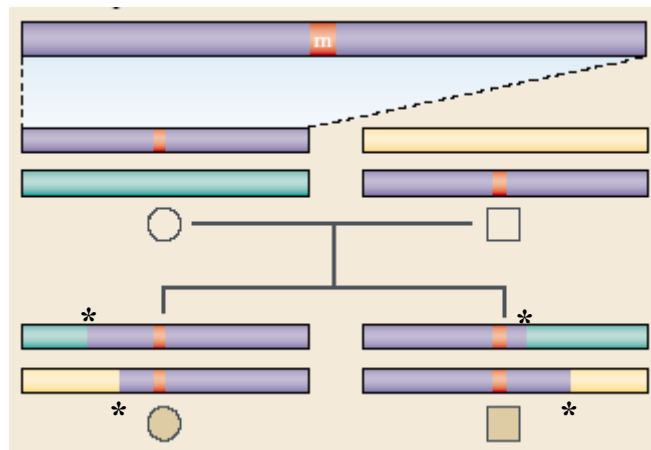
polümorfne varieeruvus kuna lisanduvad polümorfismid, mis on iseloomulikud paraloogsele segmendile (Bettencourt ja Feder, 2002, Papadakis ja Patrinos, 1999). Geenikonversioonile on iseloomulik ka suund GC sisalduse tõusule vastavas piirkonnas, kuna on näidatud eelistatult AT – polümorfismi sisaldavate kromosoomide konverteerimine DNA põhjal, mis pärineb GC – polümorfismiga kromosoomidelt (Galtier, 2003). Lisaks on leitud χ -järjestusi ning kromatiini sekundaarset struktuuri mõjutavaid elemente nagu (GT)_n, mis potentsiaalselt stimuleerivad rekombinatsioonilist ja/või geenikonversioonilist aktiivsust (Papadakis ja Patrinos, 1999, Wahls et al., 1990).

Globiini-sarnaste geenide perekonna põhjal on põhjalikult uuritud geenikonversiooni olemust ning rolli geenide evolutsioneerumisel (Papadakis ja Patrinos, 1999). Antud töö põhjal on järeldatud, et tänu kõrgelt sarnaste geenialade konverteerimisele toimub eriti 5' reguleerivate ning kodeerivate alade homogeniseerimine, mis tagab uute duplikaatide säilumise ning korrektse funktsioneerumise. Samas võidakse geenikonversiooni teel aktiveerida piirkonnas esinevaid pseudogeene tänu info ülekandumisele promootorite vahel ning genereerida uusi funktsionaalseid geene. Materjal võib ka vastupidises suunas liikuda ning seega olulisi geene inaktiveerida, mis sõltuvalt geeni iseloomust on letaalne või haigusliku väljundiga (Papadakis ja Patrinos, 1999). Sarnaseid näiteid DNA fragmentide ja assotsieeruvate polümorfismide/mutatsioonide ülekandumisest on teistegi geeniduplikaatide/pseudogeenide puhul ning teiste liikide siseselt (MHC klass I hiirel - Högstrand ja Böhme, 1994, CYP21/Cyp21P - Tusié-Luna ja White, 1995, HLA-DPB1 - 1995).

Praeguseks saadud tulemused viitavad geenikonversiooni olulisele rollile geeniperekondade evolutsioneerumisel ning funktsionaalsete geenide säilitamisel või tekitamisel. Seega on geeniperekondadega assotsieeruvate haiguste etioloogia uurimisel oluline arvestada lisaks erinevate mutatsioonide/ümberkorralduste võimaliku esinemisega ka geenikonversiooni poolt genereeritava varieeruvusega.

1.3. Rekombinatsioon ja alleelne aheldatus

Alleelne aheldatus ehk LD (*linkage disequilibrium*) (Joonis 4) on lähedalasuvate lookuste/markeri alleelide assotsiatsioon ehk koos pärandumine sagedamini kui on eeldatav nende iseseisva segregatsiooni puhul (Ardlie et al., 2002). Antud nähtuse uurimiseks on eriti suurt rõhku pööratud seoses komplekshaiguste geneetilise faktori determineerimisega LD-s olevate mutatsioonide/markeri kaudu.



Joonis 4. Aheldatuse iseloomustus. Lilla kromosoom kujutab haplotüüpi, millel on tekkinud mutatsioon (m, punane joon). Rekombinatsiooni toimumisel homoloogiliste kromosoomidega (roheline ja kollane) jääb tekkinud mutatsioon suure tõenäosusega endiselt oma haplotüübi (lilla) konteksti. (Cardon ja Bell, 2001) * -rekombinatsiooni toimumise koht.

Arvukate uuringute tulemusena on selgunud LD esinemine genoomis blokkidena, mille piirid sageli kattuvad meiootilise rekombinatsiooni *hotspot*'dega (Tishkoff ja Verrelli, 2003). Populatsioonis võib LD-bloki esineda mitu alleelsete variantide kombinatsiooni ehk haplotüüpi, mistõttu nimetatakse neid ka haplotüüpide blokkideks või haploblokkideks. Blokkide pikkuseks arvestatakse erinevatel hinnangutel keskmiselt kuni 170 kb, mis sõltub suuresti uuritavast populatsioonist, genoomi piirkonnast ning analüüsitud markerite sagedusest (Tishkoff ja Verrelli, 2003).

Populatsiooni demograafilisest ajaloost tulenevad faktorid (geenitriiv, migratsioon, pudelikaela efekt) mõjutavad oluliselt markerite alleelisagedusi ning LD-

struktuuri. Lisaks on oluline roll ka molekulaarbioloogilistel faktoritel nagu rekombinatsioon ja geenikonversioon (Ardlie et al., 2002). Populatsioonides väljaspool Aafrikat on LD-blokkide pikkus defineeritav pikemana, mida seostatakse väljarändega Aafrikast ning sellega kaasneva pudelikaela efektiga. Antud nähtus kirjeldab populatsiooni arvukuse järsku langust ning seeläbi alleelisageduste olulist muutumist ja LD-ulatuse suurenemist (Tishkoff ja Verrelli, 2003). Gabriel et al. (2002) andmetel sisaldab iga haploblokk 3-5 enamlevinud (>5%) haplotüüpi, mis on esindatud ~90% kromosoomidest igas populatsioonis.

LD ja haplotüüpide struktuuri andmeid on edukalt kasutatud komplekshaiguste (mitme geeni ning keskkonnast lähtuvate tegurite koosmõjul kujunev haigus, näiteks kasvaja) geneetilise faktori determineerimisel assotsiatsiooniuringutes. Antud tööde eesmärgiks on genotüpiseerida erinevatesse haploblokkidesse kuuluvaid ühenukleotiidseid polümorfisme (SNP) mitte-suguluses olevatel vastava haigusega patsientidel ja kontrollindiviididel ning eeldatavalt leida assotsiatsioon SNP-ga, mis on LD-s samas blokis esineva haigust põhjustava markeriga. Assotsiatsiooniuringute tulemusena on praeguseks identifitseeritud ~50 geeni, mis põhjustavat erinevaid haigusi või on eelsoodumust genereerivaks faktoriks (Wang et al., 2005).

Üheks oluliseks determinandiks LD-bloki piiride määratlemisel on rekombinatsiooni *hotspot*'de esinemine, nende sagedus genoomis ning intensiivsus. Kuna haploblokid on oluliseks vahendiks erinevate haiguste detekteerimisel, siis on igasugune info nende kujunemise kohta väärtuslik ning suurt tähelepanu on omistatud *hotspot*'de detailsele kirjeldamisele (Kauppi et al., 2004, Crawford et al., 2004).

1.4. Rekombinatsiooni *hotspot*'d

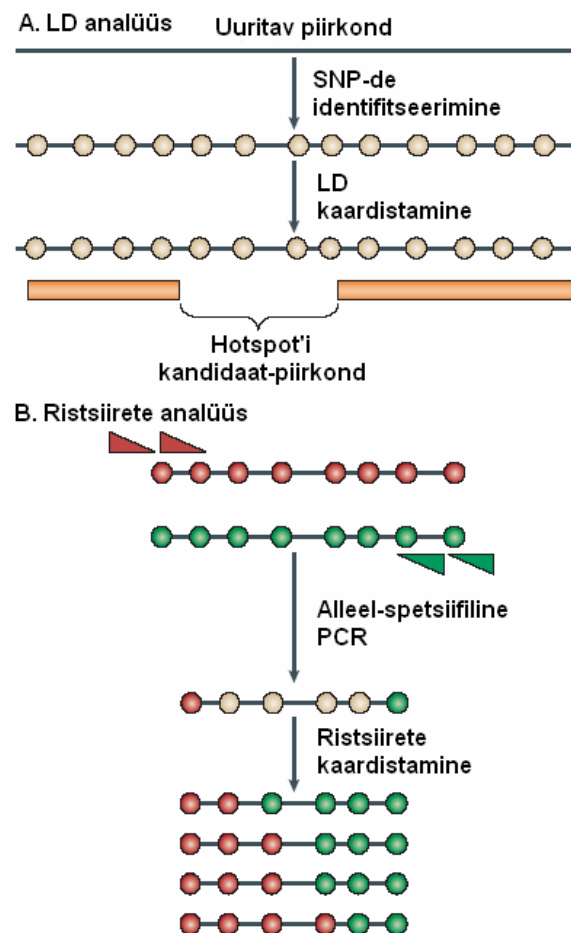
1.4.1. Rekombinatsiooni *hotspot*'i detekteerimine

Genoomis esinevate rekombinatsiooni *hotspot*'de lokaliseerimiseks on kasutusel kahte erinevat tüüpi meetodeid. Üheks võimaluseks on identifitseerida *hotspot*'i kaudselt LD-blokkide või statistiliste algoritmide abil polümorfsete positsioonide põhjal. Alternatiivseks meetodiks on kaardistada *hotspot*'i asukohta otseselt, kasutades haploidse sperma DNA-d.

Rekombinatsiooni *hotspot*'i esinemist piirkonnas on võimalik hinnata kaudselt LD-analüüsi kaudu. Kuna LD-blokkide piirid kattuvad suuresti ajalooliste rekombinatsiooni sündmustega, siis võib alleelse aheldatuse mustrit kasutada vihjena *hotspot*'de esinemise kohta genoomis (Tishkoff ja Verrelli, 2003, Phillips et al., 2003). Siiski ei ole LD-mustri interpreteerimine 1:1 suhtes *hotspot*'de esinemisega ning mitte alati ennustava efektiga (Kauppi et al., 2005).

Kaudselt on võimalik hinnata rekombinatsiooni *hotspot*'de esinemist ka statistiliste algoritmide abil. Li ja Stephens'i (2003) meetod arvutab välja rekombinatsioonilist aktiivsust haplotüüpide varieeruvuse alusel. Hudsoni algoritm (2001) võrdleb paarikaupa genotüpiseeritud lookuste infot, kalkuleerides *hotspot*'i esinemise tõenäosust analüüsitavate lookuste vahel. Piirkonna rekombinatsioonilise aktiivsuse parameetri saamiseks summeeritakse saadud tõenäosused.

Otseselt laboratoorsete meetoditega on võimalik detailsemalt rekombinatsiooni *hotspot*'e kirjeldada, kasutades analüüsitava materjalina sperma DNA-d. Jeffreys et al. on teostanud arvukalt uuringuid rakendades LD-mustrit viitava infona *hotspot*'de paiknemise kohta genoomis. Seejärel amplifitseeritakse LD-analüüsil leitud *hotspot*'i piirkond alleel-spetsiifiliste praimeritega, kasutades substraadina rekombinantset sperma DNA-d. Järgneb rekombinatsiooni saitide detekteerimine, võrdlemine referentsandmetega vere DNA-lt ning *hotspot*'i iseloomustamine (Joonis 5) (Jeffreys et al., 1990, Jeffreys et al., 2001, May et al., 2002, Jeffreys et al., 2003). Alternatiivse meetodina viiakse läbi kogu genoomi amplifikatsioon igal spermil eraldi ning seejärel analüüsitakse markereid huvipakkuvas piirkonnas, et identifitseerida rekombinatsiooni poolest aktiivseid alasid (Cullen et al., 2002).



Joonis 5. Rekombinatsiooni *hotspot*'de identifitseerimine otseselt rekombinantse sperma DNA põhjal (Kauppi et al., 2004).

1.4.2. Rekombinatsiooni *hotspot*'i iseloomustus

Rekombinatsioonisündmuste kaardistamisel on täheldatud, et ristsiirde ning sellega kaasneva geenikonversiooni punktid ei asu genoomis juhusliku jaotusega, vaid on pigem klasterdunud spetsiifilistesse piirkondadesse, mida nimetatakse rekombinatsiooni nõ kuumadeks punktideks ehk *hotspot*'deks.

Põletikureaktsioonides oluliste geenide analüüsil leiti, et rekombinatsioonisündmused esinesid *hotspot*'dena 12% geenide alas (de Silva et al., 2004). MHC lookuse analüüsil aga lokaliseeriti 38% rekombinatsiooni saitidest 6 *hotspot*'i piirkonda (Cullen et al., 2002). Laiendades antud andmeid kogu genoomile, järeldati, et rekombinatsioon toimub erinevatel hinnangutel umbes iga 65 kuni 850 kb tagant ning inimgenoom peaks sisaldama ca 3000-50000 rekombinatsiooni *hotspot*'i

(Kauppi et al., 2004, Cullen et al., 2002, Crawford et al., 2004). Arvestades potentsiaalset rekombinatsiooni *hotspot*'de arvu inimgenoomis, jääb aga praeguseks avastatute arv väga väikeseks (Tabel 1).

Tabel 1. Identifitseeritud rekombinatsiooni *hotspot*'d, nende intensiivsus, intensiivsuse võrdlus genoomi keskmisega ning pikkus. (de Massy, 2003).

Hotspot Inimene	Intensiivsus (cM)	Kordades üle genoomi keskmise	Pikkus (kb)
MS32	0.04	30	1.5
CEB1	0.16	70	2
b-Globin	0.9	90	11
PGM1	1.7	30	58
CMT1a	0.0015	1.20	1.4
DNA1	0.0005	0.3	1.9
DNA2	0.0037	3	1.3
DNA3	0.13	120	1.2
DMB1	0.0031	2	1.8
DMB2	0.028	30	1.2
TAP2	0.0058	10	1
PAR1	0.3	90	3.9
Hiir			
Ea	0.9	3750	0.4
Eb	0.1	40	4
Psm9	1	930	1.8
Gc7	-	-	2.4
Pb	1.5	470	5.3

Siiani iseloomustatud rekombinatsiooni *hotspot*'de piirkonnad hõlmavad enamasti 1-2 kb pikkuse piirkonna, mis viitab tõenäoliselt ühisele toimemehhanismile. Samas ei ole identifitseeritud DNA järjestuse põhiselt ühiseid motiive, mis seletaks kirjeldatud *hotspot*'de lokaliseerumist antud piirkonda. Potentsiaalsed motiivid/elemendid, mis võiksid olla mõnede *hotspot*'de funktsioneerimisel soodustavaks faktoriks on

1. *E. coli* χ -järjestus, mis stimuleerib rekombinatsioonlist aktiivsust üle spetsiifiliste valkude (Eggleston ja West, 1997, Badge et al., 2000). Arvatakse, et ka inimese genoomis esinevad χ -järjestused võiksid soodustada rekombinatsioonilist aktiivsust,

2. Translin valgu äratundmis-motiivid (Aoki et al., 1995, Badge et al., 2000). Translin tunneb ära spetsiifilised DNA järjestuse motiivid ning osaleb DSB kaudu tekkinud üheaheelaliste DNA otste seondumisel,
3. *Alu*-elemendid (Jeffreys ja Neumann, 2002): *Alu*-elemendid on primaatidele iseloomulikud 300 bp pikkused retrotransposonid, mida sageli seostatakse genoomsete ümberkorralduste toimumisega ning seega kromatiini struktuuri destabiliseerimisega (Houck et al., 1979, Bailey et al., 2003),
4. Ühenukleotiidne polümorfism. Jeffreys ja Neumann (2002) poolt identifitseeriti ühe MHC klass II piirkonna *hotspot*'i tsentris G/A polümorfism, mille puhul G-alleel assotsieerus vähenenud rekombinatsioonilise aktiivsusega. Leiti, et A-alleeli sisaldavalt kromosoomilt toimub eelistatult DSB initsieerimine ning vastavalt DSBR mudelile kopeeritakse info homoloogiliselt G-alleeli sisaldavalt kromosoomilt. Seega toimub rekombinatiivset aktiivsust supresseeriva G-alleeli ülekaalus transmissioon järglastele (suhtes 87G:13A *versus* 50:50 juhusliku ülekandumise puhul). Antud nähtust nimetatakse '*hotspot*'i konversiooni paradoksiks' ning bioinformaatiliselt on ennustatud analoogiliste *hotspot*'de elimineerumine populatsioonist 50% juures 70 generatsiooni jooksul (Pineda-Krch ja Redfield, 2005). Kas ja kuidas kirjeldatud *hotspot*'d populatsioonis püsivad ning kuidas antud paradoksaalne nähtus *in vivo* lahendub, ei osata veel hinnata.
5. (GT)_n piirkonnad: Pikad GT-dinukleotiidi kordused võivad põhjustada iseloomulikku kromatiini sekundaarset struktuuri, mis soodustab rekombinatsioonilise aktiivsuse tõusu kuni 20 korda (Wahls et al., 1990, Cullen et al., 2002),
6. Palindroomsed järjestused, milleks võivad olla ka pöördkordusjärjestustena esinevad *Alu*-d (Jeffreys ja Neumann, 2002), soodustavad sekundaarstruktuuride moodustumist ning stimuleerivad DSB teket (Gordenin et al., 1993).

Rekombinatsiooni *hotspot*'de detailsel analüüsil on leitud ristsiirete ja geenikonversiooni lokaliseerumine samasse piirkonda. DSB tekkele järgneva DSBR mudeli põhjal toimub suurema sagedusega initsieeriva kromatiidi konverteerimine kui ristsiire. Geenikonversioon esineb võrreldes ristsiirdega suhtes 4-15:1 ning

konverteeritavate alade pikkuseks on keskmiselt <300 bp (Kauppi et al., 2004, de Massy, 2003).

Rekombinatsioonilist aktiivsust on võrdlevalt vaadatud ka hiire ja šimpansi puhul. Hiire genoomis on identifitseeritud inimesega sarnaseid rekombinatsiooni *hotspot*'e, kuid vaatamata ortoloogsete fragmentide esinemisele ning geenide sarnasele järjestusele vastavates piirkondades, ei lokaliseeru need samadesse regioonidesse kui inimesel. (Guillon ja de Massy, 2002, Yauk et al., 2003, Kauppi et al., 2004,). Ptak et al. (2005) andmetel esinevad suured varieeruvused ka šimpansi ja inimese rekombinatsiooni *hotspot*'de võrdlusel. Saadud tulemused näitavad, et tõenäoliselt ei ole tegemist isegi suhteliselt lähedaste liikide vahel konserveerunud nähtusega ning võimalik, et *hotspot*'dele on iseloomulik dünaamiline esinemine ja lühiaegne funktsioneerimine genoomis.

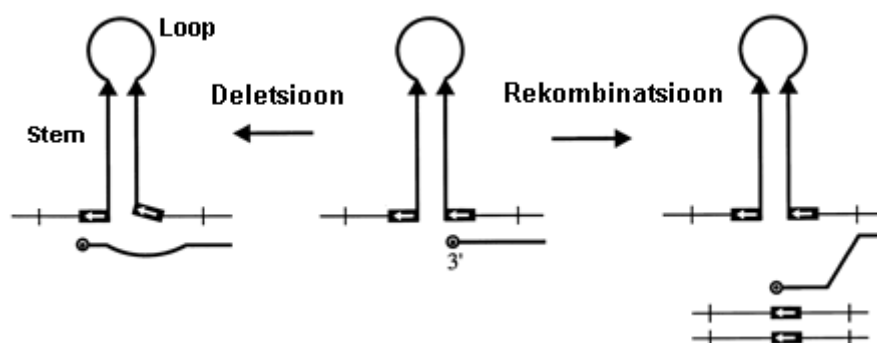
1.4.3. Rekombinatsiooni hotspot ja palindroomid

Pööratud kordusjärjestuste ehk palindroomide uurimiselt on näidatud nende assotsiatsioon rekombinatsiooni poolest aktiivsete piirkondadega. Arvatakse, et palindroomsed järjestused muudavad kromatiini struktuuri ebastabiilseks ning põhjustavad nii rekombinatsiooni *hotspot*'de kujunemist kui ka suuremate ümberkorralduste (deletsioonide) teket antud regioonis (Lobachev et al., 1998, Gordenin et al., 1993).

Juba aastal 1986 identifitseeriti rekombinatsiooni lõikekohtade vahetus läheduses kromosoomil pöördkordusjärjestused, mis potentsiaalselt on substraadiks rekombinatsiooni läbiviivatele valkudele (Krawinkel et al., 1986). Tandemsete ehk otsekordusjärjestustega analoogilist assotsiatsiooni ei identifitseeritud. Palindroomid, mis on kõrge sarnasusega, moodustavad sekundaarseid struktuure ning võivad korduste vahelises alas sisaldada ka unikaalset nn speisser-regiooni. Antud järjestuse iseloomu tõttu tekib nn *stem-loop* struktuur, mille *loop* osa on tõenäoliselt tundlik endonukleaasidele ning topoisomeraas I-le (Joonis 6) (Krawinkel et al., 1986, Cunningham et al., 2003). Seega toimub *loop* segmendis kromatiini lõikamine mõlemas kromosoomis, mille tulemuseks on DSB ning rekombinatsiooniline ja/või geenikonversiooniline aktiivsus (Cunningham et al., 2003, Nasar et al., 2000).

Palindroomseid järjestusi peetakse kromosoomi lõike indutseerimise tõttu ka kromosomaalsete ümberkorralduste tekitajaks. Suure tõenäosusega võib toimuda

stem-loop struktuuri deleteerumine või DSB tekke tagajärjel translokatsioon (Edelmann et al., 2001, Nasar et al., 2000). DSB tekkele alternatiivseks mehhanismiks on sekundaarsete struktuuride moodustumine replikatsiooni ajal. *Stem-loop* võib olla pidurdava efektiga ahela pikendamisel ning põhjustada kas (i) nn replikatsiooni libisemist, mille väljundiks on deletsioon, või (ii) sünteesitava ahela invasiooni homoloogilisse piirkonda ning seega rekombinatsiooni (Joonis 6) (Lobachev et al., 1998).



Joonis 6. Deletsiooni ning rekombinatsiooni tekkemehhanism *stem-loop* sekundaarse struktuuri vahendusel (Lobachev et al., 1998).

S. Cerevisiae kui mudelorganismi põhjal tehtud töö tulemusena leiti, et <30 bp pikkused palindroomsed kordused on genoomis stabiilsed ning ei põhjusta rekombinatsiooni ja ümberkorralduste aktiivsuse tõusu. Ebastabiilseteks osutusid >50 bp pikkused palindroomid (Nasar et al., 2000). Lisaks on leitud korrelatsioon rekombinatsiooni taseme ning *stem-loop*'i moodustvaate järjestuste pikkuse vahel. Samuti *S. cerevisiae* põhjal iseloomustatud palindroomide puhul ilmnes, et sekundaarstruktuuri stabiilsus ning rekombinatsiooni indutseerimise võime sõltub *stem* ning *loop* pikkuste suhtest (Lobachev et al., 1998). Juhul kui 323 bp pikkustest palindroomidest moodustuva *stem* osa vahele lisati 1112 bp asemel 58 bp pikkune *loop* fragment, suurenes 15 korda interkromosomaalsete ning 24 korda intrakromosomaalsete rekombinatsiooni sündmuste sagedus. Analoogseid palindroomseid järjestusi on leitud ka *Alu*-rikastes alades ning seostatud intensiivsete rekombinatsiooni *hotspot*'de ning ümberkorralduste esinemisega (Lobachev et al., 1998).

Iseloomulikke sekundaarseid struktuuri moodustavad pöördkordusjärjestused võivad modifitseerida kromatiini struktuuri avatumaks ning seetõttu olla ligipääsetavam rekombinatsiooni-spetsiifilistele valkudele (Nasa et al., 2000). Seega on palindroomsetel järjestustel, mida esineb sageli segmentaalsete duplikatsioonide ning geeniperekondade piirkonnas, rekombinatsiooni ning ümberkorraldusi tugevalt soodustav efekt.

1.5. Segmentaalsed duplikatsioonid ja genoomsed ümberkorraldused

1.5.1. Segmentaalsed duplikatsioonid

Viimaste aastate jooksul on järjest rohkem tähelepanu pööratud segmentaalsete duplikatsioonide uurimisele, kuna ilmnenud on nende kõrge sagedus inimese genoomis, oluline roll evolutsioonis tänapäeval nähtava genoomi struktuuri kujundamisel ning paljude genoomsete haiguste genereerimisel.

Segmentaalsed duplikatsioonid ehk LCR (*low-copy repeats*) on DNA kordusjärjestuste klass, mis on tekkinud pikkade genoomsete segmentide duplikatsiooni tagajärjel. LCR'de pikkus on >1kb (ca 1-400 kb) ning iseloomulikuks tunnuseks on paraloogete järjestuste sarnasus >95% (Shaw ja Lupski, 2004, Emanuel ja Shaikh, 2001). Üle inimese genoomi on leitud nende sagedus olevat ~5-10% ning tihti assotsieeruvad need geenirikaste piirkondadega. Bioinformaatiliste meetodite abil on identifitseeritud umbes 230 lookust, mille struktuur viitab LCR-de esinemisele (Stankiewicz ja Lupski, 2002, Mehan et al., 2004). Segmentaalsed duplikatsioonid võivad hõlmata nii üksikuid geene, geenifragmente, pseudogeene ning teisi paraloogete järjestusi, kui ka geeniklastreid, mis on üles ehitatud duplikatsiooni teel tekkinud sarnastes geenidest. Korduvad fragmendid võivad esineda tandeemsete või pöördkordusjärjestustena või keerulise, erineva orientatsiooniga fragmentidest koosneva struktuurina. Iseloomulikuks tunnuseks on veel suur *Alu* elementide sisaldus, mida seostatakse LCR-de evolutsioneerumise ning genoomi struktuuri ümberkorraldavate mehhanismidega (Stankiewicz ja Lupski, 2002). Kõrge sarnasuse tõttu arvatakse, et inimese genoomi segmentaalsed duplikatsioonid on kujunenud viimase ~34 miljoni aasta jooksul primaatide evolutsiooni ajal (Shaw ja Lupski, 2004, Bailey et al., 2003).

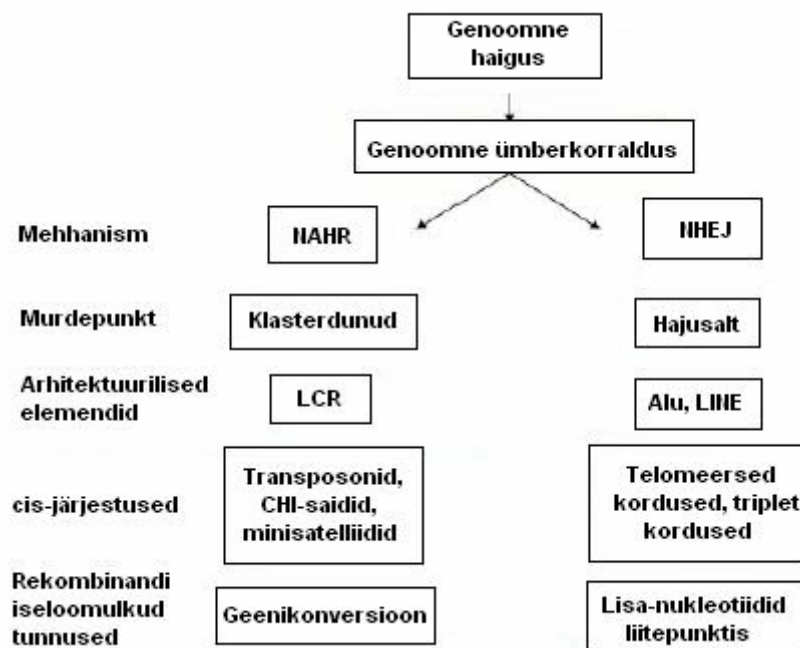
LCR-d jaotatakse genoomis lokaliseerumise alusel kahte klassi: (i) interkromosomaalsed ning (ii) intrakromosomaalsed. Interkromosomaalsed

kordusjärjestused esinevad mitte-homoloogilistel kromosoomidel, intrakromosomaalsed aga ühel kromosoomil või isegi samas kromosoomi regioonis (Emanuel ja Shaikh, 2001). Tänu kõrgele sarnasusele duplitseerunud piirkondade vahel on LCR-d substraadiks geenikonversioonile, mis omakorda aitab kaasa nende sarnasuse säilitamisele ning märkimisväärselt sagedasele esinemisele genoomis (Shaw ja Lupski, 2004).

1.5.2. Segmentaalsete duplikatsioonide vahendatud ümberkorraldused

Üks põhjusi, miks üha suurenevat tähelepanu pööratakse LCR-de detekteerimisele ning iseloomustamisele genoomis, on nende sage assotsieerumine kromosomaalsete ümberkorralduste ning genoomsete haigustega.

Segmentaalsed duplikatsioonid on >95% ulatuses sarnased, mistõttu on nad substraadiks alleelsele ning mitte-alleelsele homoloogilisele rekombinatsioonile, mille tulemuseks võivad olla erinevad kromosomaalsed ümberkorraldused. Konkreetse ümberkorralduse teke sõltub LCR-i pikkusest, sarnasuse astmest, duplitseerunud fragmentide vahelisest distantist ning orientatsioonist (Stankiewicz ja Lupski, 2002). Homoloogilise rekombinatsiooni toimumiseks on LCR-de poolt tagatud mitmed soodustavad faktorid: (i) minimaalne sarnase ala pikkus kahe duplitseerunud fragmendi vahel (~300-500bp), (ii) rekombinatsiooni soodustavate *cis*-elementide ning (iii) *stem-loop* sekundaarseid struktuure moodustavate palindroomsete korduste esinemine, mis põhjustavad DSB tekkimist (Emanuel ja Shaikh, 2001, Stankiewicz ja Lupski, 2002). Mitte-alleelse homoloogilise rekombinatsiooni (*non-allelic homologous recombination*, NAHR) toimumisel on täheldatud ümberkorralduste käigus tekkivate kromosoomi murdepunktide klasterdumist ehk rekombinatsiooni *hotspot*'de esinemist, millega võib kaasneda ka geenikonversiooniline aktiivsus (Joonis 7) (Shaw ja Lupski, 2004).

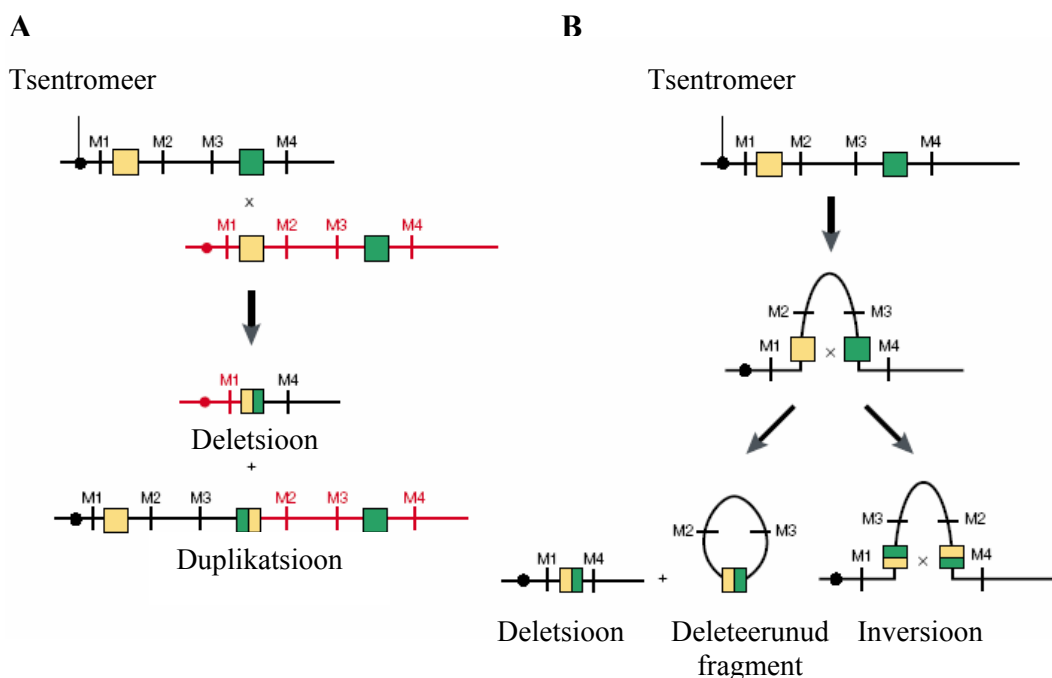


Joonis 7. Ümberkorralduste tekke-mehhanismid ning iseloomulikud tunnused (Shaw ja Lupski, 2004). NAHR - *non-allelic homologous recombination*, NHEJ - *non-homologous end joining*, LCR - *low-copy repeats*.

Alternatiivse rekombinatsiooni toimumise mehhanismiks on mitte-homoloogiline otste seondumine (*non-homologous end joining*, NHEJ), mille puhul murdepunktid esinevad hajusalt ning mehhanismi iseärasuse tõttu liidetakse otste seondumisel vahele lisa-nukleotiidid (Joonis 7) (Shaw ja Lupski, 2004).

Kõrge sarnasusega segmentaalsed duplikatsioonid muudavad kromosomaalse struktuuri vastuvõtlikuks erinevatele ümberkorraldustele, mis toimuvad peamiselt NAHR mehhanismi abil ning jaotuvad järgmiselt: (i) deletsioon, (ii) duplikatsioon, (iii) inversioon ja (iv) translokatsioon (Joonis 8) (Stankiewicz ja Lupski, 2002, Emanuel ja Shaikh, 2001). Tandeemsete duplitseerunud korduste (ühesuunalised fragmendid) osalusel tekivad peamiselt duplikatsioonid, pöördkordusjärjestuste ehk palindroomid puhul deletsioonid ning inversioonid. Kuna LCR-d paigutuvad sageli geenirikastesse piirkondadesse, siis kaasnevad kirjeldatud ümberkorraldustega sageli erinevad haiguslikud seisundid. 15. kromosoomis (15q11-q14) on geenirikas alas kirjeldatud kuus murdepunkti, mis lokaliseeruvad 15 LCR-i piirkonda ning on seotud mitme erineva genoomse haiguse etioloogiaga (deletsioonid ~70% Prader-Willi või Angelmani sündroomiga patsientidel, duplikatsioonid autismi puhul) (Pujana et al.,

2002). Näidetena võib tuua veel Charcot-Marie-Tooth haigus (duplikatsioon, Lupski, 1998), Williams-Beuren sündroom (deletsioon, Peoples et al., 2000), hemofiilia A puhul faktor VIII geeni düsfunktsionaalsus (inversioon, Lakich et al., 1993). Paljude genoomsete haiguste tekkemehhanismide uurimisel on leitud ümberkorralduste rikkas alas *hotspot*'e ning nendega assotsieeruvate elementide ja/või struktuuride (*chi*-järjestus, *stem-loop*'i moodustavad palindroomid jne) esinemine (Shaw ja Lupski, 2004).



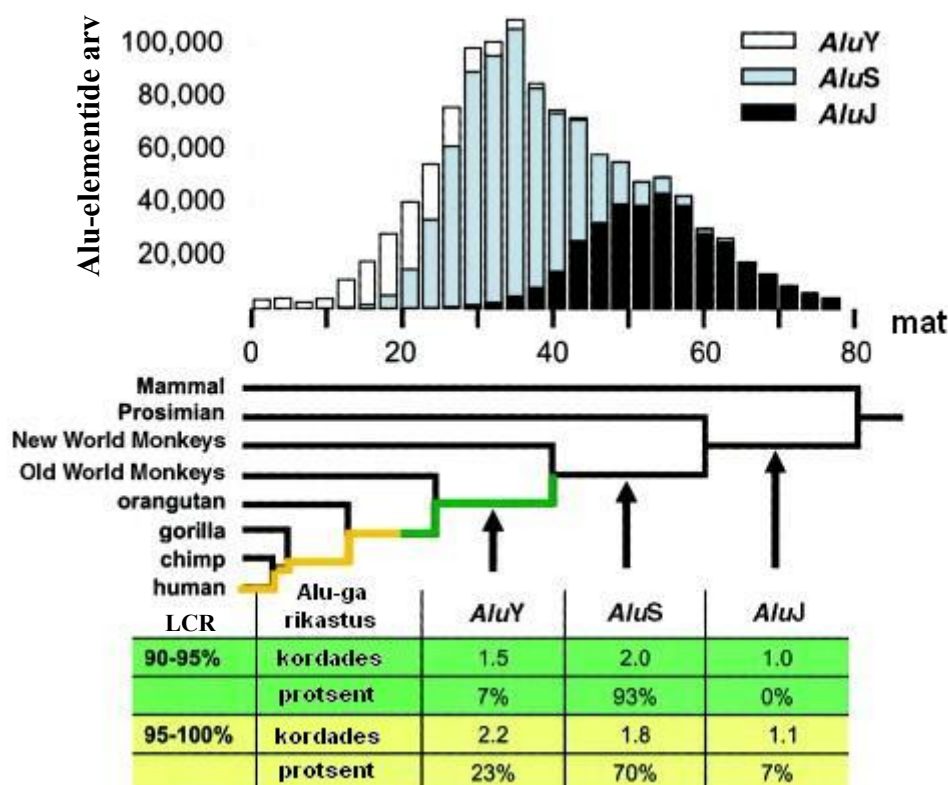
Joonis 8. LCR-vahendatud kromosomaalsete ümberkorralduste mehhanismid (Emanuel ja Shaikh, 2001). A – tandeemsed kordusjärjestused, B – pöördkordusjärjestused ehk palindroomid. M1-4 – erinevad markerid. LCR - *low-copy repeats*

1.5.3. Segmentaalsete duplikatsioonide evolutsioon

Lisaks kromosomaalsete ümberkorralduste ning erinevate sündroomide genereerimisele omastatakse LCR-dele olulist rolli ka primaatidele iseloomuliku genoomse struktuuri kujundamisel.

Seoses segmentaalsete duplikatsioonide kõrge esinemissagedusega inimeses, on uuritud detailsemalt nende pärinemist ning evolutsioonilist kujunemist. LCR-de ekspansiooni seostatakse *Alu*-elementidega, mida esineb duplitseerunud piirkondade liitekohtades oluliselt rohkem võrreldes genoomi keskmisega (23,9% versus 10,3%)

(Bailey et al., 2003). Enamik uuritud regioonidest identifitseeritud *Alu*-sid on suhteliselt noored (*AluY* 25-45 mat, *AluS* 25-45 mat) ning kolmandik neist on kiiresti levinud genoomis ~35-40 mat. Tuginedes *Alu*-järjestuste struktuurile LCR-de piirkonnas on püstitatud hüpotees, kus segmentaalsed duplikatsioonid on levinud üle kogu genoomi tänu *Alu-Alu* vahendatud rekombinatsioonile (Joonis 9) (Bailey et al., 2003).



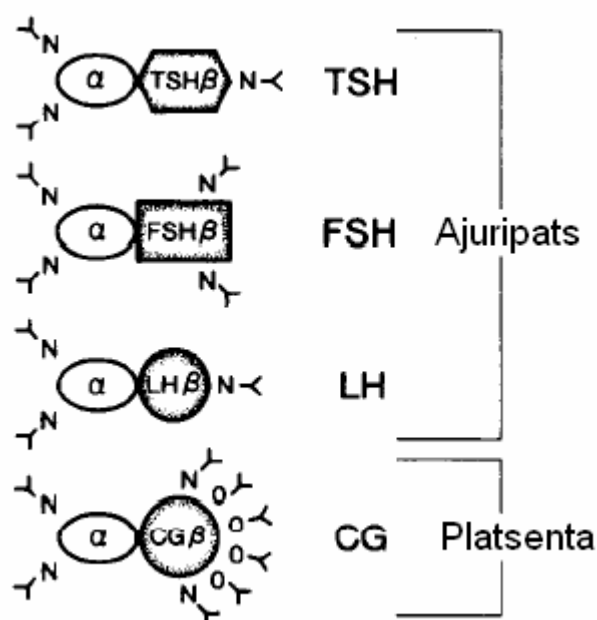
Joonis 9. *Alu*-elementide transponeerumise aktiivsus genoomis (Bailey et al., 2003). Histogrammil on kujutatud erinevate *Alu*-elementide ekspansiooni aega ja intensiivsust genoomis. Evolutsioonipuul on nooltega märgitud vastavat tüüpi *Alu*-elementide levimise aeg. Värvidega on märgitud vastavalt tabelis märgitud sarnasusega segmentaalsete duplikatsioonide leviku aeg. Tabelis on esitatud erineva sarnasuse astmega LCR-de rikastus *Alu*-dega kordades võrreldes kontrollpiirkondadega ning protsentuaalselt kõigist uuritud *Alu*-alaperekondadest. Mat – miljonit aastat tagasi, LCR – *low-copy repeat*.

Arvestades LCR-de esinemist peamiselt primaatidel ning kõrget sagedust (>95%), langeb nende levimine genoomis ajaliselt kokku *AluS* elementide leviku ja primaatide evolutsiooniga (Bailey et al., 2003, Gu et al., 2002, Cotton ja Page 2005). Tõenäoliselt muutis *Alu* elementide esialgne ekspansioon genoomi struktuuri ebastabiilseks, võimaldades *Alu-Alu* vahendatud rekombinatsiooni, teatud segmentide

duplikatsiooni ning kiiret levikut genoomis (Bailey et al., 2003, Shaw ja Lupski, 2004). LCR-de teke soodustas edasiste genoomsete ümberkorralduste toimumist ning suuremate duplikatsiooni-klastrite kujunemist. Seega võib geenirikastes piirkondades esinevaid segmentaalseid duplikatsioone vaadelda nende positsioonilise efekti (kromosomaalsete murdepunktide esinemine mitte-juhuslikult) (Shaw ja Lupski, 2004) taustal kui eelsoodumust genoomsete haiguste kujunemiseks.

1.6. Gonadotropiinid

Segmentaalsete duplikatsioonide toimumise tagajärjel on kujunenud geeniperekondi, mis on olulised reproduktsiooni seisukohast. Ühe näitena võib tuua teatud gonadotropiinide beeta-subühikuid kodeerivad geenid. Gonadotropiinid on glükoproteiinhormoonide perekond, kuhu kuuluvad luteiniseeriv hormoon ehk lutropiin (LH), folliikuleid stimuleeriv hormoon (FSH), kilpnääret stimuleeriv hormoon ehk türeotroopne hormoon (TSH) ning kooriongonadotropiinid (hCG) (Joonis 10).



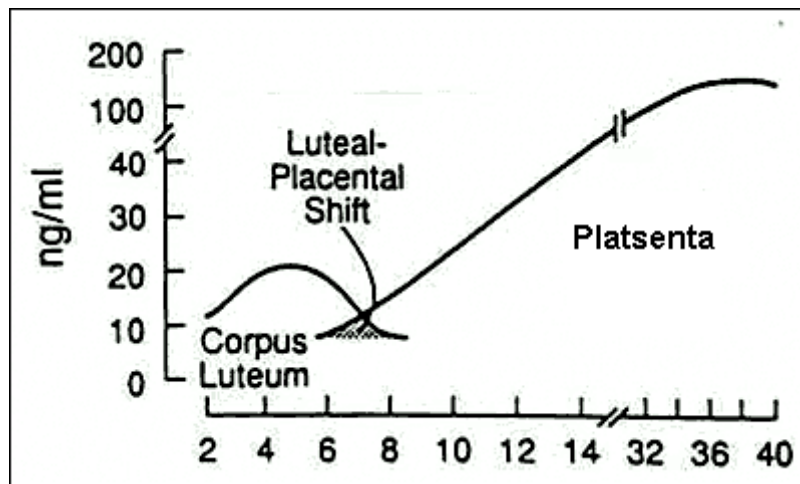
Joonis 10. Heterodimeersed glükoproteiinhormoonid. N ja O tähistavad vastavalt N- ning O-seoselist glükosüleerimist. (James ja Hollenberg, 1993)

1.5.1. Gonadotropiinide funktsioonid

Reproduktsooniga seotud hormoonid, gonadotropiinid, on glükosüleeritud valgud, millest LH, FSH ning TSH sünteesitakse ajuripatsis ning hCG peamiselt platsentas (Pierce ja Parsons, 1981, Dirnhofer et al., 1996).

Ajuripatsi LH ning FSH osalevad gametogeneesi regulatsioonis ning steroidhormoonide sünteesis gonaadides. FSH stimuleerib folliikulite küpsemist munasarjas ning osaleb spermatogeneesi regulatsioonil Sertoli rakkudes testises. LH indutseerib ovulatsiooni ning progesterooni sünteesi kollaskehas naistel ning androgeenide sünteesi Leydigi rakkudes meeste testises (Themmen ja Huhtaniemi, 2000). Samuti ajuripatsis sünteesitav TSH osaleb kilpnäärme hormonaalses metabolismis ning türoidhormoonid produktsioonis (Pierce ja Parsons, 1981).

Plantsentas sünteesitav hCG on üks esimesi ning olulisemaid determinante implantatsioonil ning kasutatakse seetõttu biomarkerina raseduse diagnoosimiseks I trimestril (Brown et al., 2000). HCG β -subühiku mRNA ekspresseeritakse *de novo* juba kahe-rakulises staadiumis ning hormooni sekreteeritakse 7 päeva vanuse blastotsüsti poolt (Jurisicova et al., 1999, Bonduelle et al., 1988). Pärast blastotsüsti pesastumist emaka limaskestast satub hCG otse ema verre, kus selle kontsentratsioon saavutab maksimumi 9-10 rasedusnädalal. HCG tase korreleerub paralleelselt moodustuva trofoblasti koe hulga (Jameson ja Hollenberg, 1993). Koorion gonadotropiin stimuleerib kollaskehas (*corpus luteum*) mitmete rasedumise normaalseks kuluks vajalike hormoonide, s.h. progesterooni, produktsiooni. Progesterooni ja hCG funktsioonideks on emaka limaskestast morfoloogiline ning immunoloogiline ettevalmistamine embrüo implantatsiooniks ja loote normaalseks arenguks ning ekstraembrüonaalsete kudede diferentseerumiseks (<http://www.endotext.com/female/female13/female13.htm>, Siiteri et al., 1977). Kollaskeha funktsioneerib ligikaudu 10. rasedusnädalani, kuid 6.-8. nädalal hakkab formeeruma platsenta ning järkjärgult siirdub nn rasedushormoonide produktsioon kollaskehast platsentale (*luteal-placental shift*) (Jameson ja Hollenberg, 1993) (Joonis 11). Seega on I trimestril hCG piisav produktsioon triviaalse olulisusega raseduse normaalseks kulgemiseks. Normist madalamaid hCG väärtusi seostatakse näiteks raseduse katkemise ja emakavälise raseduse esinemisega (Buyalos et al., 1992).



Joonis 11. Progesterooni süntees raseduse kulgemisel (nädalates). (<http://www.endotext.com/female/female13/female13.htm>)

1.5.2. Gonadotropiinide struktuur

Heterodimeersed glükoproteinhormoonid koosnevad kahest mitte-kovalentselt seotud alaühikust, α ja β . α -subühik on ühine kõigile gonadotropiinidele ning unikaalne β -subühik määrab iga hormooni spetsiifilisuse (Themmen ja Huhtaniemi, 2000, Pierce ja Parsons, 1981).

Gonadotropiinide α -subühik (92 ah) sisaldab kahte N-seoselist glükosüleerimise saiti ning sisaldab 10 tsüsteiini (5 disulfiid-sidet), mis on olulised subühiku korrektsel pakkumisel ning heterodimeerse hormooni assambleerimisel. β -subühik (111-145 ah) sisaldab kahte N-seoselist glükosüleerimise saiti (LH β -1 üks) ning 12 tsüsteiini 6 disulfiid-sideme moodustamiseks (Fan ja Hendrickson, 2005, Themmen ja Huhtaniemi, 2000, Miller-Lindholm et al., 1999). Võrreldes LH β -ga on HCG β valk 29 ah võrra pikem, mis tagab hCG pikema eluea võrreldes LH-ga (Themmen ja Huhtaniemi, 2000). Kahe gonadotropiinide alamühiku ekspressiooni ning translatsiooni reguleeritakse iseseisvalt, kuid paralleelselt, et tagada vajalik hulk aktiivset heterodimeerset valku. Limiteerivaks faktoriks aktiivse hormooni assambleerimisel on β -alaühiku kontsentratsioon. (Jameson ja Hollenberg, 1993, Pierce ja Parsons, 1981).

Glükoproteinhormoonide retseptorid on G-valgu seoselised ning sisaldavad spetsiifilist ekstratsellulaarset pikka N-terminaalset domääni, mis on vajalik hormooniga seondumiseks. FSH ja TSH seonduvad oma spetsiifilisele FSH ja TSH

retseptorile, mida kodeerivad geenid lokaliseeruvad vastavalt kromosoomil 2 (2p21-16) ning 14 (14q31). HCG ja LH seonduvad aga samale G-valgu seoselisele retseptorile, mis on kodeeritud 2. kromosoomil paikneva geeni poolt (2p21) (Themmen ja Huhtaniemi, 2000).

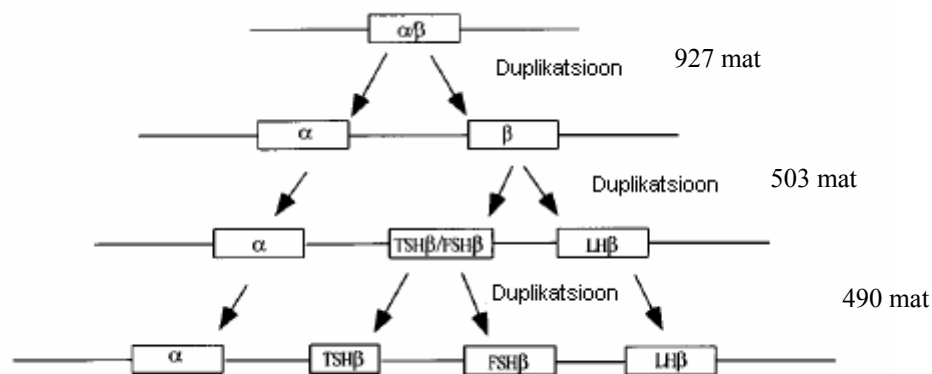
1.5.3. Gonadotropiinide subühikuid kodeerivate geenide evolutsioon

Gonadotropiine kodeerivate geenide struktuuri uuringute põhjal on ilmnunud glükoproteiinhormoonide keerukas evolutsiooniline areng, mis hõlmab mitmeid duplikatsiooni sündmusi. Senini ei ole aga ühtselt mõistetavat seletust, millist eelist omab tänapäeval esineva gonadotropiinide geeniklastri keeruline ning tõenäoliselt ümberkorralduste suhtes ebastabiilne struktuur.

E erinevate heterodimeersete glükoproteiinhormoonide subühikuid kodeerivad geenid paiknevad genoomis:

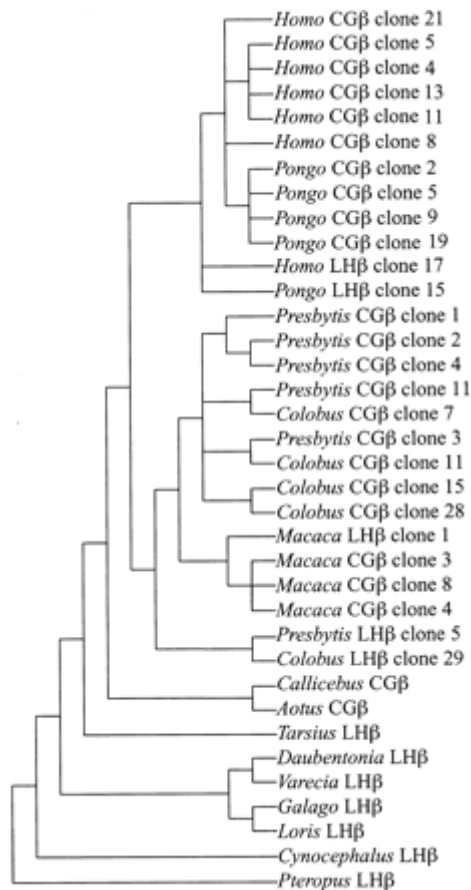
- (i) ühine α -subühik 6. kromosoomi pikas õlas (6q12.21),
- (ii) FSH β 11. kromosoomis (11p13),
- (iii) TSH β 1p13 ning
- (iv) LH β ja 6 hCG β geeni ühise geeniklastrina 19q13.32 piirkonnas (Themmen ja Huhtaniemi, 2000, Policastro et al., 1986)

Inimesel esineva gonadotropiinide perekonna α - ja β -subühikuid kodeerivad geenid arvatakse olevat tekkinud ühisest α/β eellasgeenist korduvate duplikatsioonisündmuste teel. Li ja Ford (1998) andmetel esines tõenäoliselt esimene duplikatsiooni sündmus 927 miljonit aastat tagasi (mat), mille tulemusena tekkis α -subühikut kodeeriv geen ning tänapäevastele gonadotropiinide β -subühikutele aluse pannud eellasgeen (Joonis 12)



Joonis 12. Gonadotropiinide subühikuid kodeerivate geenide evolutsioon (Li ja Ford, 1998).

503 mat tekkinud LHβ geeni evolutsioon on edasi kulgenud liigiti erinevaid teid pidi. Teatud liikidel on esindatud ainult üks LHβ geen (lammas, siga), hobustel modifitseeritud LHβ vorm, mida ekspresseeritakse nii ajuripatsis kui platsentas. Primaatidel on toimunud LHβ duplikatsioon, mille tulemusena on tekkinud platsentas ekspresseeritav CGβ geen (94 mat) (Li ja Ford, 1998). Erinevatel primaatidel on identifitseeritud erinev arv CGβ geenikoopiad ning kõige rohkem ($n=6 + 1 LHB$) on leitud inimesel. Maston ja Ruvolo (2002) uuringu raames määratleti CGβ ja LHβ lahknemise aeg veelgi hilisemaks (50-34 mat) ning lisaks tuvastati geenide analüüsil viiteid geenikonversiooni rollile klasteri evolutsioneerumisel. Erinevate liikide CGβ ja LHβ geenide klasterdamisel DNA kloonide järjestuse sarnasuse põhjal ilmnes, et liigi siseselt on CGβ ja LHβ geenid omavahel sarnasemad kui CGβ ja LHβ eraldi liikide vahel (Joonis 13). Tulemus viitab intensiivsele materjali vahetusele samas piirkonnas lokaliseeruvate LHβ ja CGβ geenide vahel liigi siseselt. Kuna geenikonversioon toimub piirkondades, kus esineb kõrge sarnasusega järjestusi, siis võib väita, et antud geeniklaster on ümberkorralduste aldis.



Joonis 13. LH β ja CG β geenide klasterdumine järjsetuse sarnasuse alusel liigiti (Maston ja Ruvolo, 1998).

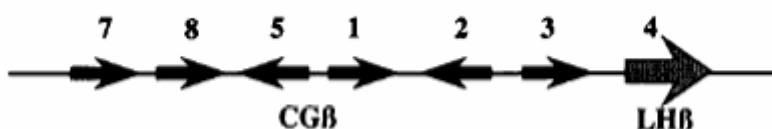
CG β on tekkinud LH β geenist duplikatsiooni teel ning lisaks omandanud ka raaminihke mutatsiooni vahetult enne stop-koodonit, mistõttu on tema valgu järjestus 30 ah võrra pikem (Fiddes ja Goodman, 1980, Pierce ja Parsons, 1981). Pikem N-terminaalne fragment (CTP) tagab CG β sekreteerimise rakust apikaalselt ema vereringesse (LH β sekreteeritakse basolateraalset) ning tunduvalt pikema eluea võrreldes LH β -ga (12h *versus* 30min). Valgu parem stabiilsus tagatakse nelja O-glükosüleerimise saidi esinemisega CTP fragmendis ning seostatakse positiivse selektsiooniga CG β evolutsioonil (Nakav et al., 2005, Jablonka-Shariff et al., 2002, Maston ja Ruvolo, 1998).

Lisaks raaminihet põhjustavale deletsioonile on uus CG β geen omandanud ka alternatiivse transkriptsiooni initsiatsiooni saidi. Kuigi CG β hõlmab ka LH β -le iseloomulikku TATA-box'i, initsieeritakse transkriptsioon sellest sõltumatult 366 bp võrra geenist eespool (Jameson ja Habener, 1996, Jameson ja Hollenberg, 1993).

Tõenäoliselt tagab uue promootori kasutusele võtmine hCG ekspressiooni platsentas ning samas paralleelselt LH β -spetsiifilise promootori võimalik säilumine seletaks LH β madalat ekspressiooni ka ajuripatsis (Dirnhofer et al., 1996).

Arvestades LH ja hCG erinevaid funktsioone ning produktsiooni kohta, on CG β geen pidanud lisaks duplitseerumisele ka muteeruma, omandama koespetsiifilisuse platsenta suhtes ning samuti nõ kaasama α -subühiku ekspressiooni platsentas (Maston ja Ruvolo, 1998). Alternatiivse hüpoteesi kohaselt ekspresseerus LH β nii ajuripatsis kui platsentas enne CG β lahknemist. Antud väidet toetab asjaolu, et hobusel esinev LH β /CG β geeni transkribeeritakse mõlemas organis ning inimese puhul on leitud vähesel määral hCG produktsiooni lisaks platsentale ka ajuripatsis (Dirnhofer et al., 1996, Maston ja Ruvolo, 1998).

Mitmete järjestikuste ümberkorralduste ning mutatsioonide tulemusena on inimesel välja kujunenud praeguseks keerulise struktuuriga *LHB/CGB* geeniklaster, mis koosneb ühest LH β ning kuuest CG β kodeerivast geenist (*CGB*, *CGB1*, *CGB2*, *CGB5*, *CGB7* ja *CGB8*) (Policastro et al., 1986) (Joonis 14). Lisaks on teadaolevalt piirkonnas ka kolm omavahel sarnast pseudogeeni *NTF6A*, *B* ja *G* (<http://www.ncbi.nlm.nih.gov/>). Kõik geenid ja samuti intergeensed alad on kuni ~96% sarnased (Jameson ja Hollenberg, 1993).



Joonis 14. *LHB/CGB* geeniklaster (Jameson ja Hollenberg, 1993). (*CGB3* sünonüüm on *CGB*).

1.6. Korduvad spontaanabordid

1.6.1. Üldiseloostus

Korduvat spontaanaborti defineeritakse kui kaks või enam järjestikust raseduse iseeneslikku katkemist <22. raseduse nädalal või kui loode kaalub sündides <500g. Spontaanabordi esinemissageduseks on hinnanguliselt 1-2% fertiilsetel naistel (Bricker ja Farquharson, 2002). Vanuses >36 aasta on täheldatud suuremat spontaanabordi (SA) kujunemise riski, mida seostatakse peamiselt aneuploidia (eriti trisoomiate) kõrgeenenud esinemissagedusega vanemas eas (Robinson et al., 2001). I trimestri SA esineb koekultuuri katsete põhjal aneuploidiat ~70% juhtudest, millest ~65% moodustavad trisoomiad (Fritz et al., 2001). Samas ei ole >36 a naistel leitud otsest seost struktuursete kromosoomi anomaaliatega (translokatsioonid, inversioonid) ning korduvate SA-de vahel. Goddijn et al. (2004) andmetel ei kaasne ema kõrgema eaga eelnevate SA-de suurema arvuga I trimestril kõrgeenenud riski struktuursete kromosomaalsete aberratsioonide esinemiseks.

Spontaanaborte võib laias laastus jagada kahte gruppi vastavalt loote südametegevuse esinemisele raseduse katkemisel: (i) varane raseduse katkemine, kus loote südametegevus ei ole visualiseeritav (*embryo loss*, EL) ja (ii) nn loote hukk (*fetal loss*, FL), mille korral on loote südametegevus ei ole detekteeritav. Korduvate iseeneslike raseduste katkemiste puhul on identifitseeritud sagedamini EL-i (suhtes 6:1 FL-ga) ning võrdselt erinevates etioloogilistes SA alamgruppides (Bricker ja Farquharson, 2002). Samuti on täheldatud, et korduvate SA puhul esineb normaalse kariotüübiga EL-i oluliselt sagedamini võrreldes kontrollgrupiga (vastavalt 66,7% versus 45,7%) (Morikawa et al., 2004). Seega on tõenäoliselt I trimestri raseduste katkemistel lisaks aneuploidiale oluline roll teistel faktoritel peale kromosoomi struktuuri anomaaliatega.

Korduvate spontaanabortide tekkepõhjused võivad olla endokriinsed, immunoloogilised, anatoomilised või geneetilised. Kriitilisteks staadiumiteks raseduse säilumisel peetakse implantatsiooni, trofoblasti invasiooni ja platsentatsiooni perioode (Quenby ja Farquharson, 1993). Vaatamata paljudele kliinilistele ja laboratoorsetele uuringutele antud vallas ei osata ~50% SA juhtudest konkreetset põhjust identifitseerida. Ebaselge põhjusega iseeneslike abortide etioloogilisi põhjusi selgitavad uuringud on seega olulised diagnostika, ravi ning patsientide nõustamise seisukohast.

1.6.2. *LHB/CGB geeniperekond ja spontaanabordid*

Glükoproteiini hCG, mida produtseeritakse platsentas blastotsüsti poolt, peetakse kõige olulisemaks biomarkeriks raseduse varajasel diagnoosimisel. HCG kontsentratsiooni põhjal hinnatakse raseduse suurust ning selle kulgu (Brown et al., 2000). Kõrvalekalded normaalsest hCG tasemest viitavad võimalikele häiretele glükoproteiini valgu ehituses, funktsioonis või metabolismis, mille üheks tagajärgedeks võib olla raseduse iseeneslik katkemine. Seetõttu on antud valku kodeeriv *LHB/CGB* geeniklaster atraktiivne kandidaatpiirkond mutatsioonide, ümberkorralduste või deletsioonide/duplikatsioonide detekteerimiseks iseeneslike raseduse katkemistega naistel.

1.6.2.1. Mutatsioonid

Seoses gonadotropiinide kriitilise rolliga reproduktiivfunktsioonis on *LHB/CGB* geeniklastrit mutatsioonide suhtes ka detailsemalt uuritud, kuid usaldusväärseid tulemusi on märkimisväärselt vähe. Varieeruvusi on kirjeldatud pigem geenides, mis kodeerivad glükoproteiinide retseptoreid ning nende metabolismi mõjutavaid hormone (Themmen ja Huhtaniemi, 2000, Layman, 2002, Layman, 1999). Samuti on leitud mutatsioone teistes glükoproteiinide geenides: (i) *FSHB* (Matthews et al., 1993) ning (ii) *TSHB* (Dacou-Voutetakis et al., 1990). Gonadotropiinide α -subühikut kodeerivast lookusest pole leitud ühtegi mutatsiooni. Kuna tegemist on nii FSH, TSH kui ka gonadotropiinide ühise subühikuga, siis tõenäoliselt on α -subühiku inaktiveerumisel inimesel letaalne väljund (Layman, 2002).

LHB/CGB regioonis on leitud üks aminohappelist muutust põhjustav *missense* mutatsioon *LHB*-s (Weiss et al., 1992) ning üks *missense* G>A transitsioon *CGB5* geenis (Miller-Lindholm et al., 1999). *CGB5*-s leitud aminohappeline muutus takistab disulfiid-sidemete moodustumist ning seega korrektset valgu pakkumist. Alternatiivne *CGB5* produkt ei ole võimeline seonduma α -subühikuga ning 85% sekreteeritakse monomeersena (*versus* 50% *wild-type* valguga). Disulfiid-sildade olulisust gonadotropiini normaalsel funktsioneerimisel on kinnitanud ka teised uuringud (Suganuma et al., 1989, Feng et al., 1995) ning seega võib potentsiaalselt olla spontaanabordi põhjustajaks madala hCG tasemega rasedatel naistel. Kato et al.

(2002) uuringu tulemusena aga ei leitud korduvate SA-ga patsientide ning kontrollindiviidide võrdlusel erinevusi hCG molekuli struktuuris ega bioloogilises aktiivsuses.

1.6.2.2. Struktuursed ümberkorraldused

Lisaks erinevatele mutatsioonidele on tandeemseid ja pöördkordusjärjestusi sisaldav *LHB/CGB* geeniklaaster potentsiaalselt aldis erinevatele struktuursetele ümberkorraldustele nagu deletsioonid, duplikatsioonid ning inversioonid. Analogiliste piirkondade puhul on eelnevalt identifitseeritud mitmeid näiteid, kus haiguslik fenotüüp on ilmnunud segmentide või tervete geenide deleteerumisel/duplitseerumisel (Christian et al., 1999, Robinson et al., 1996, Pentao et al., 1992). Kirjeldatud ümberkorraldusi on uuritud *LHB* ning *CGB5* geenide puhul Layman et al. (1997) poolt. Uuringu eesmärgiks oli identifitseerida potentsiaalseid deletsioone korduvate SA-ga patsientidel ja viljatutel naistel ning duplikatsioone patsientidel, kellel esineb trofoblasti koe kasvaja. Kogu geeni hõlmavaid deletsioone ega duplikatsioone vastavates etioloogilistes gruppides antud uuringu raames ei detekteeritud. Siiski ei välista kirjeldatud tulemus ümberkorralduste esinemist *LHB/CGB* geeniklastri, kuna kasutatud meetod on CG β geenide spetsiifiliseks eristamiseks limiteeritud.

2. Töö eesmärgid

Käesolev töö on üks osa laiahaardelisest projektist, mille eesmärgiks on esmalt detailselt iseloomustada *LHB/CGB* geeniklastri struktuuri ja varieeruvust ning seejärel analüüsida selle seotust korduvate spontaanabortide tekkega. Töö koosneb järgnevatest etappidest:

- I. *LHB/CGB* geeniklastri detailse ülesehituse kirjeldamine,
- II. *LHB/CGB* geenide üldise varieeruvuse analüüs üldpopulatsioonis (samuti rekombinatsiooni *hotspot*'i iseloomustamine) (Hallast, Nagirnaja et al., käsikiri),
- III. Rekombinatsiooni *hotspot*'i detailne analüüs otseselt rekombinantse sperma DNA põhjal üldpopulatsioonis (DNA materjal kogutud),
- IV. Pilootprojekti läbiviimine *LHB/CGB* geenide varieeruvuse võrdluseks korduvate spontaanabortidega patsientidel ning referentsina üldpopulatsioonis,
- V. *LHB/CGB* geenide varieeruvuse võrdlus korduvate spontaanabortidega patsientidel ning kontrollpopulatsioonina komplikatsioonideta sünnitajatel (analüüsi materjal kogutud),
- VI. Etapis I saadud info põhjal uurida potentsiaalseid kromosomaalseid ümberkorraldusi võrdlevalt patsientidel ning kontrollindiviididel hübriidisatsioonil põhineva meetodiga FISH (meetod on optimeerimisel).

Käesoleva magistritöö eesmärkideks oli läbi viia etapid I, IV ning etapist II rekombinatsiooni *hotspot*'i iseloomustamise osa, mis on omakorda ettevalmistuseks ülejäänud projekti etappidele.

3. Materjal ja metoodika

3.1. *In silico* analüüs

LHB/CGB geeniklastri detailse struktuuri kirjelduse aluseks on piirkonna erinevate segmentide paarikaupa võrdlus veebipõhiste programmide abil (globaalne joondus, <http://www2.igh.cnrs.fr/bin/align-guess.cgi>, ja BLAST, <http://www.ncbi.nlm.nih.gov/>). Alu-elementide sisaldust uuritavas piirkonnas määrati samuti veebipõhise programmiga (<http://searchlauncher.bcm.tmc.edu/>). *In silico* analüüsil kasutatud referentsjärjestus on saadud NCBI avalikust andmebaasist (<http://www.ncbi.nlm.nih.gov/>) (lookus nr NG_000019).

Palindroomsete järjestuste ennustamisel on rakendatud EMBOSS'i einverted programmi (Rice et al. 2000) ja sekundaarsete struktuuride ennustamisel RNAstructure programmi (versioon 4.11) (Mathews et al., 2004) ning veebipõhist programmi mfold (<http://www.bioinfo.rpi.edu/applications/mfold/>). EMBOSS einverted programm otsib järjestuse sarnasuse põhiselt pööratud kordusjärjestusi ehk palindroome. RNAstructure ning mfold võimaldavad lisaks RNA-le ka DNA järjestuse alusel kalkuleerida minimaalset seondumise vaba energiat (ΔG) ning seega identifitseerida uuritava piirkonna enim soodustatud sekundaarset struktuuri.

3.2. Analüüsi materjal

3.2.1. Rekombinatsiooni hotspot'i analüüsi materjal

Potentsiaalse rekombinatsiooni *hotspot*'i kirjeldamiseks on kasutatud 11 Eesti populatsiooni indiviidi DNA-d, mis on eraldatud 2 ml vereproovist (kogutud koostöös Margus Punabiga, M.D., TÜ Kliinikumi androloogia kabinet).

3.2.2. Patsiendigrupp

Käesoleva projekti läbiviimiseks on TÜ Kliinikumi Inimueuringute Eetikakomitee luba (nr. 117/9, 16.06.2003). Antud töös kasutatud patsiendigruppi kuuluva 24 indiviidi vere- ja platsentaproovid on kogutud koostöös Kristiina Rulliga (M.D., TÜ Kliinikumi Naistekliinik). Valim koosneb:

1. 9 naispatsiendi platsenta materjal

2. 5 naispatsiendi vereproovid

3. 5 paari vereproovid (naispatsient ja tema meespartner).

Meeste analüüsi materjal on uuringus vajalik, kuna *LHB/CGB* geenide puhul puudub informatsioon imprintingu kohta ning seega võimalikud raseduse katkemist soodustavad geenivariandid võivad olla kodeeritud isapoolselt. Võimalusel on analüüsiks kogutud platsenta materjal vanemate vereproovide asemel, kuna platsenta sisaldab nii ema- kui isapoolset geenivarianti, mis potentsiaalselt võisid põhjustada raseduse iseeneslikku katkemist.

Valikukriteeriumid patsientide grupi koostamiseks on järgnevad:

1. Kaks või enam raseduse iseenesliku katkemise või raseduse peetumise juhtu
2. Raseduse katkemine I trimestril. Võimalikult homogeense grupi koostamiseks on uuringust välja jäetud patsiendid hilisema perioodi spontaanabortidega, kuna peale I trimestrit võivad ilmneda lisafaktorid (näiteks emakakaela puudulikkus), mis ei ole antud uuringu seisukohast olulised.
3. Raseduse katkemine on ebaselge põhjusega ning patsient on negatiivse vastusega järgnevate analüüside suhtes:
 - a. Kromosoomide analüüs perifeersest verest (kromosomaalsete aberratsioonide välistamiseks)
 - b. Ultraheli (anatomiliste tegurite välistamiseks)
 - c. Hormoonanalüüsid (teatud hormonaalsete häirete välistamiseks)
 - d. Verehüübimise testid (verehüübimise süsteemi funktsiooni häirete välistamiseks)
 - e. Autoantikehade esinemise analüüs
4. Meespartnerite kromosoomide analüüsi vastus on negatiivne.

3.2.3. Referentsvalim

LHB/CGB geenide (v.a. *CGB8*) võrdluseks patsientidega kasutati 47 eestlase (TÜ MRI biotehnoloogia õppetooli DNA-pank, Lõuna- Eesti valim) genotüüpide infot, mis on eelnevalt identifitseeritud *LHB/CGB* geenide resekveneerimisel (P. Hallast, 2004). *CGB8* geeni võrdluseks on kasutatud 11 indiviidi materjali, mille põhjal on analüüsitud potentsiaalse rekombinatsiooni *hotspot*'i piirkonda, kus lokaliseerub ka *CGB8*.

3.3. DNA eraldamine

3.3.1. Vereproovid

DNA eraldamiseks EDTA-verest (2-9 ml) lisati 1ml kohta 4 ml puhvit A [0.32 M sahharoos, 1 mM Tris-HCl (pH 7.5), 5 mM MgCl₂, 1% Triton X-100, vesi] ning tsentrifuugiti 10 min 3000 rpm 4°C juures. Seejärel eemaldati supernatant ning korrati sama etappi 2 ml puhver A-ga. Peale supernatandi uut eemaldamist lisati sademele 0,9 ml puhvri B [400 mM NaCl, 2 mM EDTA (pH 8), 10 mM Tris-HCl (pH 7.5), vesi], toksiti sade lahusesse, lisati 100 µl puhvrit C (10% SDS, 2 mg/ml proteinase K, vesi) ning inkubeeriti üleöö 37°C juures. Järgmisel päeval lisati inkubeeritud lahusele 270 µl küllastatud NaCl-i, raputati segu tugevalt ning tsentrifuugiti 15 min 4000 rpm toatemperatuuril. Supernatant tõsteti seejärel uude tuubi, lisati kahekordne kogus 96 või 100 % etanooli, segati õrnalt ning DNA eemaldati Pasteuri pipetiga. DNA lahustati sõltuvalt kogusest 250-1500 µl TE-s (pH 7,5, 10 mM Tris-HCl pH 7,5, 1mM EDTA).

3.3.2. Platsenta materjal

Uuringus kasutatud platsenta on peale kogumist koheselt külmutatud vedelas lämmastikus –70°C juures. DNA eraldati ca 30 mg külmutatud koeproovist kasutades MBI Fermentase kiti (Genomic DNA Purification Kit, #K0512) ning vastavalt tootja poolt varustatud protokollile. Lisaks viidi läbi ka ribonukleaas A töötlus, kuna platsenta on transkriptsiooniliselt väga aktiivne. Ribonukleaas A-d (MBI Fermentase) lisati vastavalt kiti protokollile DNA eraldamise vaheetapis 0,2 mg/ml ning inkubeeriti 37°C 10 min. Eraldatud DNA lahustati 100 µl TE-s (pH 7,5).

Verest ning platsentast eraldatud DNA kontsentratsiooni ja puhtust mõõdeti spektrofotomeetriliselt (HeliosTMBeta, Thermo Spectronic) vähemalt peale kahepäevast lahustumist 4°C juures ning kvaliteeti kontrolliti geelelektroforeesil. Puhtuse määramisel kasutati mõõtmisel saadud tulemusi lainepikkustel 260 nm ning 280 nm ning arvutati suhe OD₂₆₀/OD₂₈₀. Väärtust ~1,7-2 defineeriti kui puhas ning edasiseks analüüsiks sobiv genoomne DNA.

3.4. Praimerite disain

3.4.1. *LHB/CGB* geenid

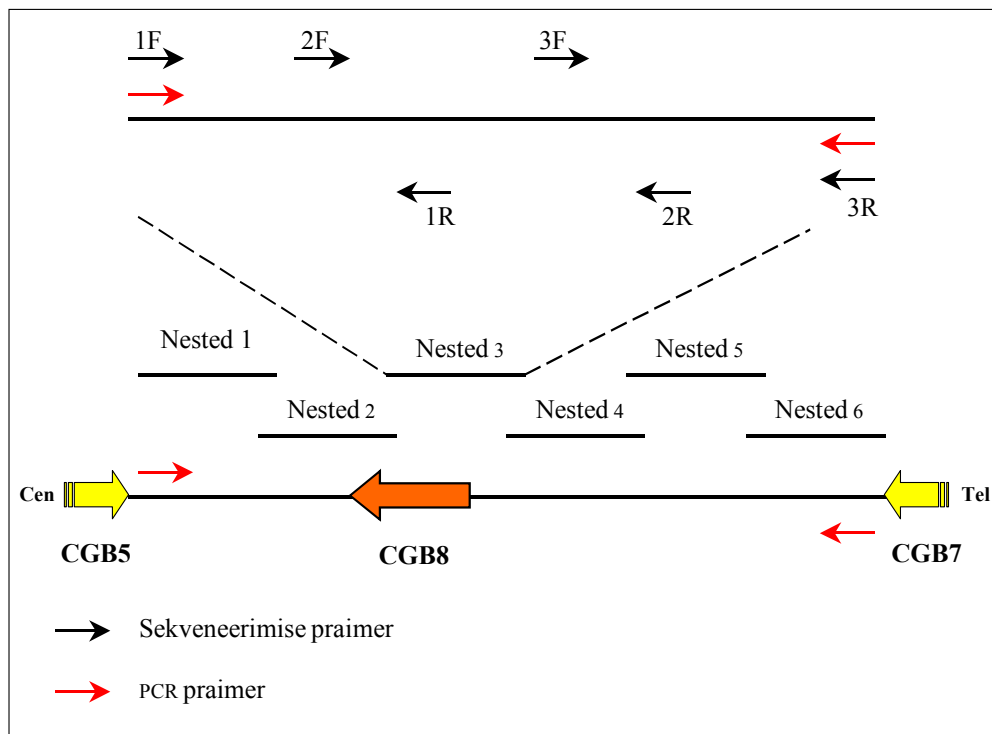
Kõikide praimerite disainimisel kasutati Whitehead Institute for Biomedical Research ja Howard Medical Institute's välja töötatud veebipõhist programmi Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) tuginedes NCBI andmebaasi referentsjärjestusele (<http://www.ncbi.nlm.nih.gov/>) (lookus nr NG_000019). *LHB*, *CGB*, *CGB1*, *CGB2*, *CGB5* ja *CGB7* geenidele oli võimalik disainida geeni-spetsiifilisi PCR praimereid, mis amplifitseerivad unikaalse soovitud fragmendi (P. Hallast, 2004) (Lisa 1). Seoses *LHB/CGB* geeniklastri komplekse struktuuriga ei olnud *CGB8*-le võimalik disainida geeni-spetsiifilisi PCR praimereid ning amplifitseeriti seetõttu paralleelselt potentsiaalse rekombinatsiooni *hotspot*'i piirkonnaga (vt allpool).

LHB/CGB geenide sekveneerimise praimerite disain on detailsemalt kirjeldatud P. Hallast magistritöös (2004) (Lisa 1).

3.4.2. Potentsiaalse rekombinatsiooni *hotspot*'i piirkond ning *CGB8*

Eelnevalt identifitseeritud kõrge rekombinatsioonilise aktiivsusega piirkonna (Hallast, Nagirnaja et al., käsikiri) amplifitseerimiseks disainiti praimerid *CGB5* ja *CGB7* vahelisse piirkonda sarnaselt punktile 3.4.1. PCR praimerid asetsevad üksteisest 8,3 kb kaugusel ning amplifitseeritav piirkond hõlmab ka *CGB8* geeni (Joonis 15). Spetsiifilisuse tõstmiseks amplifitseeriti 8,3 kb piirkond edasi nested-PCR meetodil, kus nn nested ehk sisemised PCR praimerid asuvad susbstraadina kasutatava PCR-produkti seesmistel aladel. Nested praimerid valiti nii, et saadavad produktid (6 tk) kattuksid ca 80-100 bp. Nested1 ja nested6 fragmentide amplifitseerimisel kasutati ühe praimerina ka vastavalt *forward* või *reverse* 8,3 kb PCR praimerit. Resekveneerimiseks valiti välja kokku 6 praimerit (3 kummagi ahela jaoks) spetsiifiliselt iga nested produkti jaoks (Lisa 2). Enamus nested-PCR praimeritest olid kasutatavad ka vastava produkti sekveneerimiseks. Uued sekveneerimispraimerid disainiti ainult nested fragmentide seesmistele osadele ning nested1 ja 6 puhul ka vastavalt 1F ning 3R otsmise PCR-praimerite asemele. Kõik sekveneeritavad segmendid kattuvad ottest ca 80 bp.

Uuringus kasutatud PCR praimerite unikaalsust on kontrollitud inimese genoomi vastu (BLAST) tuginedes NCBI referentsjärjestusele. *LHB/CGB* geenide puhul olid aktsepteeritavad praimerite paarid, millest vähemalt üks on genoomis unikaalne. Kahte unikaalset praimerit ühe geeni kohta ei olnud võimalik geeniklastri keerulise struktuuri tõttu disainida. Nested praimerid olid unikaalsed 8,3 kb PCR produkti järjestuse vastu kontrollides.



Joonis 15. Potentsiaalse rekombinatsiooni *hotspot*'i ning *CGB8* piirkonna PCR-i ning sekvenerimise praimerite paigutus genoomis.

3.5. Amplifikatsiooni protsess

3.5.1. *LHB/CGB* geenide amplifikatsioon

CGB, *CGB1*, *CGB2*, *CGB5*, *CGB7* ja *LHB* geenid amplifitseeriti genoomselt DNA-lt (100 ng). PCR reaktsioonisegu (25 µl mahus) sisaldas veel 2,5 µl 10x PCR puhvrit koos 15 mM MgCl₂'ga (MBI Fermentas, Vilnius, Leedu), 0,2 mM dATP, dCTP, dGTP ja dTTP (Amersham Biosciences Inc., Milwaukee, WI, USA), 10 pmol iga praimerit (Metabion GmbH, Martinsried, SLV) ja 1U DNA polümeraasi (Long PCR Enzyme Mix, MBI Fermentas). Iga reaktsiooniseguga viidi alati läbi ka negatiivse kontrolli PCR (ilma DNA-ta). Amplifikatsioon viidi läbi PTC-200 (MJ Research, Inc. Watertown, MA, USA) PCR masinaga järgmistel tingimustel:

Algne denaturatsioon	95°C	5 min		
Denaturatsioon	95°C	20 sek	}	10 tsüklit
Praimerite seondumine	68°C	30 sek		
Temperatuuri langus tsükli kohta 1°C				
Ekstensioon	68°C	2 min		
Denaturatsioon	95°C	20 sek	}	10 tsüklit
Praimerite seondumine	56°C	30 sek		
Ekstensioon	68°C	2 min		
Denaturatsioon	95°C	20 sek	}	10 tsüklit
Praimerite seondumine	54°C	30 sek		
Ekstensioon	68°C	2 min		
Denaturatsioon	95°C	20 sek	}	10 tsüklit
Praimerite seondumine	51°C	30 sek		
Ekstensioon	68°C	2 min		
Inkubatsioon	68°C	10 min		

3.5.2. Potentsiaalse rekombinatsiooni hotspot'i piirkonna amplifikatsioon

8,3 kb pikkune produkt amplifitseeriti genoomselt DNA-lt (100 ng). Erinevalt *LHB/CGB* geenide amplifikatsioonil kasutatavast reaktsioonisegust sisaldas käesolev PCR-i segu ka DMSO-t (3%, MBI Fermentas). Paralleelselt viidi läbi sama seguga ka ilma DNA-ta PCR negatiivse kontrollina. Amplifikatsioon viidi läbi GeneAmp® PCR System 2700 (Applied Biosystems) PCR masinaga järgmistel tingimustel:

Algne denaturatsioon	94°C	5 min		
Denaturatsioon	94°C	20 sek	}	4 tsüklit
Praimerite seondumine	68°C	30 sek		
Temperatuuri langus tsükli kohta 1°C				
Ekstensioon	68°C	8 min		
Denaturatsioon	94°C	20 sek	}	11 tsüklit
Praimerite seondumine	64°C	30 sek		
Ekstensioon	68°C	8 min		
Denaturatsioon	94°C	20 sek	}	25 tsüklit
Praimerite seondumine	64°C	30 sek		
Ekstensioon	68°C	8 min		
Aja pikenemine iga tsükli kohta +5 sek				
Inkubatsioon	68°C	10 min		

8,3 kb PCR produkti põhjal viidi edasi läbi nested-PCR. Reaktsioonisegu (25 µl mahus) sisaldas 0,5-1 µl 8,3 kb PCR produkti, 2,5 µl 10x PCR puhvrit (Solis BioDyne, Estonia), 2,5 µl 25 mM MgCl₂ (Solis BioDyne, Estonia), 0,2 mM dATP, dCTP, dGTP ja dTTP (Amersham Biosciences Inc., Milwaukee, WI, USA), 10 pmol iga praimerit (Metabion GmbH, Martinsried, SLV), 3% DMSO-t (MBI Fermentas) ja 1-2U DNA polümeraasi (Smart-Taq Hot DNA Polymerase, Naxo, Estonia). Nested

fragmentide amplifikatsioon viidi läbi GeneAmp® PCR System 2700 (Applied Biosystems) PCR masinaga järgmistel tingimustel:

Algne denaturatsioon	95°C	15 min		
Denaturatsioon	95°C	20 sek	}	10 tsükli
Praimerite seondumine	68°C	30 sek		
Temperatuuri langus tsükli kohta	1°C			
Ekstensioon	68°C	2 min		
Denaturatsioon	95°C	20 sek	}	10 tsükli
Praimerite seondumine	56°C	30 sek		
Ekstensioon	68°C	2 min		
Denaturatsioon	95°C	20 sek	}	20 tsükli
Praimerite seondumine	54°C	30 sek		
Ekstensioon	68°C	2 min		
Inkubatsioon	68°C	10 min		

CGB8 geeni amplifikatsiooniks patsientidel viidi läbi esmalt 8,3 kb fragmendi PCR genoomselt DNA-lt ning seejärel 8,3 kb produktilt kahe nested piirkonna amplifikatsioon (nested2 ja nested3), mis hõlmasid uuritava geeni piirkonda (Joonis 15). Vastavad kaks nested produkti sekveneeriti *CGB8* geeni ulatuses kokku 8 sekveneerimispraimeriga (nested 2 – praimerid 2F, 2R, 3F ja 3R ning nested 3 – praimerid 1F, 1R, 2F ja 2R).

3.6. Resekveneerimine

3.6.1. PCR produktide töötlus

Kõik PCR produktid kontrolliti geelelektroforeesil kõrvuti koos negatiivse kontrolliga. Seejärel töödeldi ExoI eksonukleasiga (MBI Fermentas) ja kreveti aluselise fosfataasiga (SAP, USB Corporation, Cleveland, Ohio, USA): ~20 µl produktile lisati 1U ExoI ja 1,5U sAP ensüümi. Reaktsioonid viidi läbi GeneAmp® PCR System 2700 (Applied Biosystems) PCR masina abil järgmistel tingimustel:

Inkubeerimine	37°C	20 min
Inaktiveerimine	80°C	15 min

3.6.2. Sekveneerimine

LHB/CGB geenide (v.a. *CGB8*) resekveneerimisel kasutati iga geeni kohta kokku 6 praimerit: geenispetsiifilised 2 otsmist ning 4 praimeriga 2 erinevat komplekti *sense* (*CGB2*, *CGB5*) ja *antisense* (*CGB*, *CGB1*, *CGB7*, *LHB*) geenide jaoks. Sekveneeritava ala pikkus jäi enamike geenidel vahemikku 1,5-1,6 kb, *CGB7* puhul 2,1 kb.

CGB8 sekveneerimisel kasutati 8 sekveneerimispraimerit ning substraadina kahte nested produkti (nested2 ja nested3), mis katsid geeni *CGB8* piirkonda. Rekombinatsiooni *htospot*'i piirkonnast amplifitseeritud 6 nested PCR produkti sekveneeriti kokku 6 erineva sekveneerimispraimeriga.

Sekveneerimisreaktsioonides kasutati DYEnamic ET Terminator Cycle sekveneerimiskiti (Amersham Biosciences). Reaktsioonisegu (10 µl mahus) sisaldas 1 µl sekveneerimise pre-mix'i, 3 µl lahjenduspuhvrit [100 µl 10x PCR puhver (Solis BioDyne, Estonia), 55 µl 25 mM MgCl₂ (Solis BioDyne, Estonia), 95 µl ddH₂O], 3 µl ddH₂O, 1 µl 2pM praimerit ja 2 µl puhastatud PCR produkti. Reaktsioonid viidi läbi GeneAmp® PCR System 2700 (Applied Biosystems) PCR masina abil järgmistel tingimustel:

Denaturatsioon	95°C	30 sek	} 50 tsüklit
Praimerite seondumine	50°C	15 sek	
Ekstensioon	60°C	1 min	

Sekveneerimisreaktsioonide sadestamiseks lisati igale proovile 1,5 µl 10 M ammooniumatsetaadi ja dekstraani segu (1 µl ammooniumatsetaati ja 0,5 µl dekstraani). Proovid segati, lisati 35 µl 96% -20°C etanooli, pandi 15 minutiks -20°C juurde sadenema ning tsentrifuugiti 15 min 13 000 rpm. Seejärel supernatant eemaldati, lisati 200 µl 70% +20C etanooli ning tsentrifuugiti 10 min 13 000 rpm. Peale supernatandi eemaldamist kuivatati proove 10-15 min 37°C juures ning lisati 3 µl *ABI loading solution*'it (Formamide loading dye, Amersham Biosciences). Proovid sekveneeriti ABI Prism 377 masina abil.

3.7. DNA järjestuste analüüs ja polümorfismide identifitseerimine

Sekveneerimisetapil saadud DNA järjestused visualiseeriti BioEdit Sequence Alignment Editor programmi abil (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>), kus toimus järjestuse kvaliteedi hindamine. Halva kvaliteediga sekventsipuhul vastav proov sekveneeriti uuesti.

Hea kvaliteediga järjestused assambleeriti konsensusjärjestuseks programmiga Phrap (<http://www.phrap.org>), skaneeriti programmiga PolyPhred ning analüüsiti programmiga Consed (Gordon et al. 2001). PolyPhred programmi poolt identifitseeritud potentsiaalsed heterosügootsed positsioonid kontrolliti ning kinnitati käsitsi. Iga kinnitatud polümorfne positsioon esines mõlemal ahelal.

Järjestuste assambleerimisel saadud konsensusjärjestusi võrreldi NCBI andmebaasi referentsjärjestusega (<http://www.ncbi.nlm.nih.gov/>) erinevuste leidmiseks ning sekveneerimise kvaliteedi kontrolliks. Antud analüüsiks kasutati veebipõhist paarikaupa globaalset joondust, (<http://www2.igh.cnrs.fr/bin/align-guess.cgi>). GC sisalduse arvutamiseks kasutati programmi Oligo Calculator (http://old.mclab.com/toolbox/oligo_calculator.jsp).

3.8. Statistiline analüüs

Sekveneerimise tulemuste analüüsil (i) iseloomustati identifitseeritud markereid, (ii) hinnati alleelse aheldatuse (LD) tugevust, (iii) piirkonna rekombinatsioonilist ning (iv) ennustati haplotüüpide jaotuvust populatsioonis.

3.8.1. Markerite ning uuritava piirkonna varieeruvuse üldiseloomustus

Markerite iseloomustamiseks määrati esmalt nende alleelisagedused ning kontrolliti vastavust Hardy-Weinbergi tasakaalule ($p^2 + 2pq + q^2 = 1$) programmiga HaploView (Version 3). Statistiliselt oluliseks piirväärtuseks HW tasakaalu hindamisel loeti $P < 0,05$. (P-väärtus on tõenäosus, et antud tulemused on saadud juhuslikult).

Analüüsitava piirkonna varieeruvuse kirjeldamiseks kasutati järgnevaid parameetreid (programmiga DnaSP, Version4.0): (i) nukleotiidne mitmekesisus (π), mis on keskmine nukleotiidide erinevus kahe järjestuse vahel, (ii) theta (θ), mis sõltub olemasolevate polümorfsete positsioonide arvust, järjestuse pikkusest ja valimi suurusest ning (iii) Tajima's D, mis testib hüpoteesi, et kõik mutatsioonid on selektiivselt neutraalsed, tuginedes π ja θ erinevustele. Juhul kui π ja θ on võrdsed ning Tajima's D läheneb nullile, on tegemist standartse neutraalse mudeliga. Positiivse Tajima's D viitab positiivsele selektsioonile antud piirkonnas ning negatiivne väärtus puhastavale valikule, hitch-hiking'u efektile või populatsiooni kasvule.

3.8.2. Alleelse aheldatuse defineerimine

Kahe lookuse/markeri vahelise alleelse aheldatuse määratlemiseks kasutati parameetrit D' , mis lähtub D väärtusest. D hindab saadud haplotüüpide sageduste erinevust oodatavast (markerid segregueeruvad juhuslikult) (Lewontin, 1960). D' on aga suhteline D väärtus võrreldes selle maksimumiga antud alleelisageduste puhul ($D' = D/D_{\max}$, $D' \geq 0$) (Lewontin, 1964). Juhul kui $D' = 1$, siis ei ole uuritavaid markereid eraldatud meiotilise rekombinatsiooni käigus ning esineb alleelne aheldatus. Rekombinatsiooniliste sündmuste esinemisel antud kahe markeri vahel läheneb D' väärtus nullile. D' väärtuse determineerimiseks kasutati programmi HaploView (Version 3), mis kalkuleerib lisaks D' -le ka *LOD score* väärtust (kümnnendlogaritm

tõenäosusest, et LD markerite vahel esineb *versus* LD markerite vahel ei esine). Markereid peetakse tugevalt aheldunuks D' väärtustel $>0,7$ (Gabriel et al. 2002) ning *LOD score* väärtusel >3 . Programm loeb mitte-informatiivseteks LD väärtusi, mille puhul D' on kõrge, kuid *LOD score* madal (näiteks madala alleelisagedusega markerid).

3.8.3. Rekombinatsiooni määra defineerimine ja haplotüüpide ennustamine

Rekombinatsioonilise aktiivsuse (ρ) ning haplotüüpide mustri ennustamiseks rakendati programmi PHASE (Version 2.1.), mis kasutab analüüsil sekveneerimisel identifitseeritud populatsiooni varieeruvuse andmeid vastavalt Stephens et al., (2001) algoritmile. Rho (ρ) on keskmine rekombinatsiooni sagedus populatsioonis. $\rho=4N_e r$, kus N_e on efektiivne populatsiooni suurus ja r rekombinatsiooni määr generatsiooni kohta. Rekombinatsioonilise aktiivsuse hindamiseks kahe järjestikuse markeri vahel leitakse parameetri λ väärtus, mis iseloomustab, mitu korda on rekombinatsioon antud kahe markeri vahel kõrgem kui ρ . Juhul kui $\lambda = 1$, ei erine rekombinatsiooniline aktiivsus kahe markeri vahel rho-st. Lambda väärtusel >1 viitab aga kõrgenenud rekombinatsioonilisele aktiivsusele ning potentsiaalse rekombinatsiooni *hotspot*'i esinemisele. PHASE programmi parameetritena kasutati: tsüklite arv – 1000; *thinning interval* - 1; *burn-in* – 100.

Valimi haplotüüpide jaotuvuse hindamiseks geeniti kasutati programmi PHASE (Version 2.1.), mis ennustab populatsiooni varieeruvuse andmete põhjal erinevate haplotüüpide mustrit ning sagedusi.

4. Tulemused

4.1. *LHB/CGB* geeniklastri detailne struktuur

4.1.1. *LHB/CGB* geeniklastri üldiseloostus

LHB/CGB geeniklastri *in silico* analüüsetapi tulemusena ilmnes antud genoomse regiooni kompleksne struktuur. Piirkond on ~40 kb pikk ning asetseb 19. kromosoomi pikas õlas (19q13.32) *LHB/CGB* klaster hõlmab seitse LH ja hCG beeta-subühikuid kodeerivat duplitseerunud geeni, mis on omavahel kõrge sarnasusega (*CGB*, *CGB1*, *CGB2*, *CGB5*, *CGB7*, *CGB8*, *LHB*). Geenide pikkused on vahemikus 1365-1466 bp ning on valdavalt kodeeritud *antisense* ahelalt (v.a. *CGB2* ja *CGB5*) (Tabel 2) (Joonis 16). Klastrist tsentromeeri poole, vahetult enne *LHB* geeni, asetseb *RUVBL2*, mis on bakterite *RUVB* geeni homoloog ning ei oma mingit sarnasust *LHB/CGB* geenidega. Uuritud piirkonna telomeerses osas paikneb neutrofiin 5 (*NTF5*), mis on olnud tõenäoliselt duplikatsioonisündmuste üheks substraadiks klastri evolutsioneerumisel.

Praeguse *LHB/CGB* geeniklastri struktuuri (Joonis 16) põhjal püstitasin hüpoteesi, et see on kujunenud järgmiste evolutsiooniliste sündmuste kaudu:

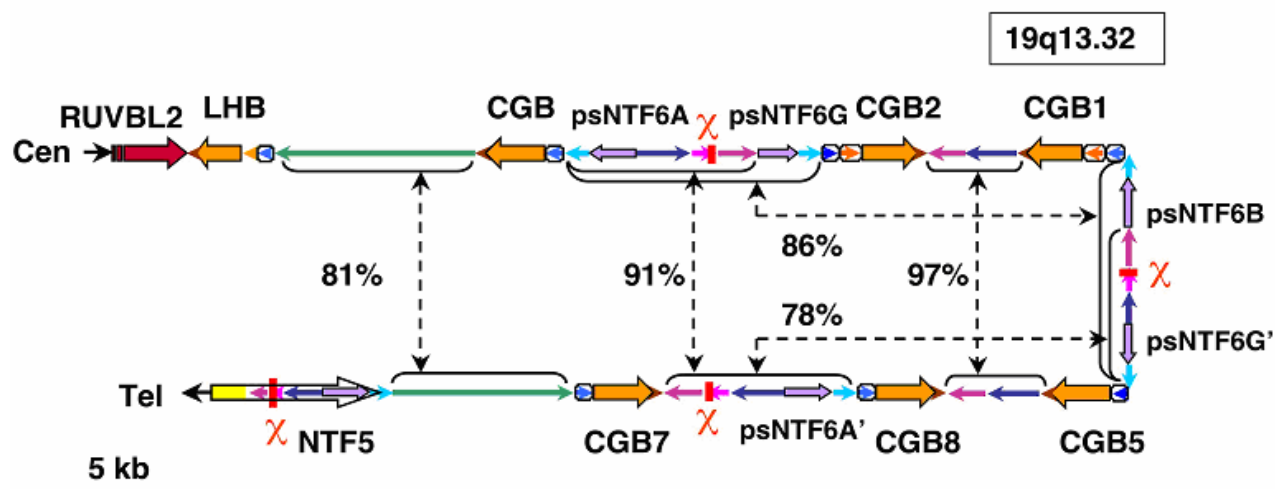
- (1) *LHB* ja *NTF5* fragmendi (intron 2, ekson 3 ja 3'UTR) ning nende vahelise intergeense ala duplikatsioon,
- (2) ainult *LHB* ja *NTF5* fragmenti hõlmavad järjestikulised duplikatsioonid,
- (3) interkromosomaalne (kromosoom 19 *vs.* kromosoom 2 või 3) *Alu*-vahendatud mitte-homoloogiline rekombinatsioon, mille tagajärjel inserterus *CGB1* ja *CGB2*-spetsiifiline fragment, mis paikneb geenide 3' osas. Antud mehhanismi abil on tekkinud *CGB1* ja *CGB2* –l alternatiivne promootor ja ekson 1.

Kirjeldatud evolutsiooniliste sündmuste tulemusena pani *LHB* duplikatsioon aluse *CGB* geenidele ja *NTF5* viimase eksoni duplikatsioon aluse viiele neutrofiin 6 pseudogeeni (*psNTF6*), mis paiknevad sama orientatsiooniga iga *CGB* geeni 5' otsas (v.a. *LHB* ja *CGB7*). Eelnevalt olid antud piirkonnas teada pseudogeene *psNTF6A/B* ja *psNTF6G*. Käesoleva analüüsi käigus aga identifitseeriti veel *psNTF6A'* ja *psNTF6G'*, mis on nimetatud vastavalt järjestuse sarnasusele teadaolevate pseudogeenidega.

LHB/CGB geenide intergeensed piirkonnad jäävad vahemikku 2330-7536 bp, on *Alu*-rikkad (10-56%) ning 81-97% sarnased (Joonis 16). Kogu geeniklaster on ka kõrge GC sisaldusega (>56%), mis on tunduvalt üle genoomi keskmise (45%). Geenide piirkondades on GC sisaldus isegi kuni 65%.

Tabel 2. *LHB/CGB* geenide pikkused, orientatsioon ning sarnasus

	Geen	Pikkus (bp)	Orientatsioon	Omavaheline DNA järjestuse sarnasus	
1	<i>LHB</i>	1110	<i>Antisense</i>	~85% <i>CGB1,2</i>	
3	<i>CGB2</i>	1365	<i>Sense</i>	~99 %	~85%
4	<i>CGB1</i>	1365	<i>Antisense</i>		
2	<i>CGB</i>	1466	<i>Antisense</i>	~97–99%	
5	<i>CGB5</i>	1466	<i>Sense</i>		
6	<i>CGB8</i>	1466	<i>Antisense</i>		
7	<i>CGB7</i>	1466	<i>Antisense</i>		



Joonis 16. *LHB/CGB* geeniklastri struktuur. Ühte värvi nooled tähistavad sarnase järjestusega piirkondi. Laiemad nooled esindavad geenide piirkondi ning pseudogeene, peenemad nooled aga sarnase järjestusega intergeensete segmentide alasid. Kastides sinised nooled märgivad kõikidele geenidele ühiseid promootorite alasid, oranžid kastides nooled aga *CGB1* ja *CGB2* spetsiifilisi potentsiaalseid reguleerivaid alasid. Kollane piirkond *NTF5* geenis tähistab antud klastris unikaalset järjestust. χ – *E.coli* chi-järjestus.

LHB/CGB geeniklastri detailsemal analüüsil järjestuste omavahelisel võrdlusel ilmnes *CGB* geenide väga kõrge homoloogia. Ühelt poolt *CGB*, *CGB1*, *CGB5* ja *CGB7* ning teisalt *CGB1* ja *CGB2* on omavahel sarnased 97-99% (Tabel 2). *LHB* omab homoloogiat ülejäänud geenidega ca 92-93%, kuna erineb lühema mRNA poolest nii 5' kui 3' osas. Samuti sisaldab ta diferentseerunud eksonit 3, mis kannab hormooni-spetsiifilisust. *CGB1* ja *CGB2* on küll omavahel äärmiselt sarnased (~99%), kuid erinevad järjestuse põhjal rohkem teistest klastri geenidest (sarnasus ~85%). Detailsemal analüüsil identifitseeriti *CGB1* ja *CGB2* 5' otsas ka unikaalne järjestus, mis võib olla antud geenidele spetsiifiline regulatoorne ala. (Joonis 16) Kuna *CGB1* ja *CGB2* kasutavad alternatiivset esimest eksonit ja ekson 2 ja 3 puhul on toimunud 1 aluspaariline raaminihe, siis pole nende geenide kodeeritud hüpoteetilisel valgul mingit sarnasust hCG beeta-subühikut kodeerivate geenide (*CGB*, *CGB5*, 7,8) valguga. Kogu genoomi võrdlemisel antud järjestusega leidsid homoloogilised piirkonnad kromosoom 2 ning 3 mitte-kodeerivates alades. Lisaks omavad *CGB1* ja *CGB2* ka promootori regiooni, mis on ühine kõigile *LHB/CGB* klastri geenidele.

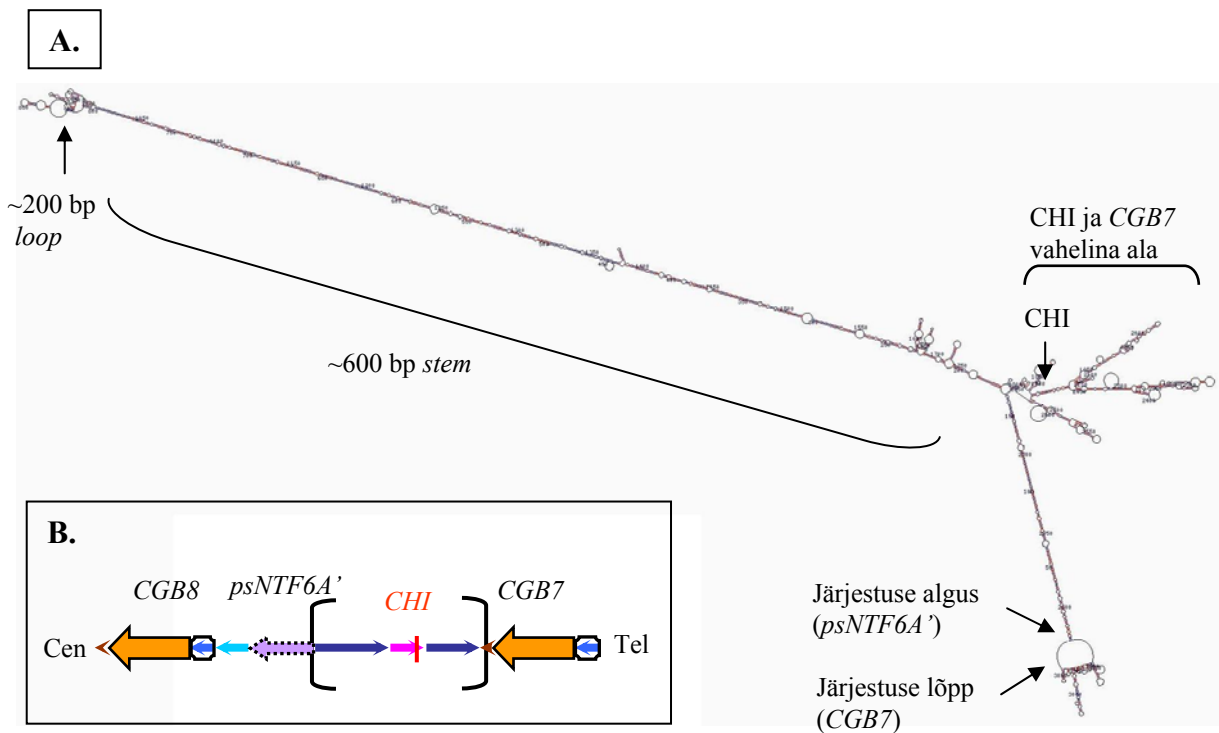
Intergeensete piirkondade järjestuste omavahelisel võrdlusel ilmnes äärmiselt keeruline korduvate elementide struktuur. Antud klastri geenidevahelised alad on sarnased osale *NTF5* geenist ning tõenäoliselt tekkinud sealt *Alu*-vahendatud duplikatsiooni teel. *LHB* ja *CGB* ning *NTF5* ja *CGB7* vahelised piirkonnad ei oma sarnasust teiste intergeensete aladega klastris ning on ka *Alu*-vaesed (Joonis 16). *CGB-CGB2*, *CGB1-CGB5* ning *CGB8-CGB7* on kõige sarnasemad *NTF5*-le ning hõlmavad *Alu*-rikast segmenti, mis sisaldab rekombinatsiooni ning geenikonversiooni stimuleerivad χ -saiti. Antud piirkondades on kas hilisemate ümberkorralduste tulemusena või klastri duplikatsioonisündmuste käigus tekkinud järjestustes varieeruvused: *CGB7-CGB8* alas on puudu fragment koos pseudogeeninga, *CGB1-CGB5* alas on toimunud pseudogeenide vahelise ala inversioon ning kaotsi läinud osa järjestusest, mis eelneb *psNTF6G*'-le (Joonis 16). Seetõttu on häiritud ka sekundaarse struktuuri moodustumine antud piirkonnas (punkt 4.1.2.). *CGB2-CGB1* ning *CGB5-CGB8* intergeensetes regioonides puuduvad nii pseudogeenid kui *Alu*-rikas fragment χ -järjestusega.

4.1.2. Palindroomide ning sekundaarsete struktuuride ennustamine

LHB/CGB geeniklastri detailsel analüüsil identifitseeriti keeruline sarnastest fragmentidest koosnev struktuur, mis võib eeldatavalt olla aluseks erinevate palindroomsete *stem-loop* sekundaarsete struktuuride moodustumisele. Viimaste ennustamiseks kasutati programme EMBOSS einverted, RNAstructure ja mfold ning saadud tulemused suuresti kattusid.

EMBOSS einverted ennustatud tulemuste põhjal sisaldab *LHB/CGB* regioon kaks paari palindroomseid järjestusi, mis võiksid potentsiaalselt moodustada stabiilseid sekundaarseid struktuure. *CGB-CGB1* vahelises alas (Joon 15) esinevad palindroomsetena intergeense ala otsad (2094 bp), mis hõlmavad ka pseudogeene ning jäävad üksteisest 2788 bp kaugusele. *CGB8* ja *CGB7* vahelises piirkonnas leidsid palindroomsed järjestused ~100 bp χ -saidist tsentromeeri suunas äärmiselt *Alu*-rikkas DNA segmendis. Palindroomid olid pikkusega 625 bp ning nende vaheline ala 222 bp. Vaatamata χ -järjestust sisaldava fragmendi esinemisele ka teistes klastri piirkondades, ei ole seal palindroomide ning stabiilsete sekundaarsete struktuuride moodustumine soodustatud DNA järjestuse varieeruvuse tõttu.

RNAstructure ning mfold programmidega leitud sarnased DNA *stem-loop* sekundaarsed struktuurid hõlmasid samasid palindroomide, mis olid defineeritud EMBOSS einverted poolt. *CGB* ja *CGB1* vahele ennustati struktuur, kus pöördkordusjärjestused moodustasid pika (~2000 bp) nn *stem* osa ning nende vaheline ala ~2780 bp pikkuse nn *loop* osa. *CGB8-CGB7* regioonis joonistus ~600 bp pikkuse *stem*'i ning ~200 bp pikkuse *loop*'ga sekundaarne struktuur (Joonis 17), mis võib potentsiaalselt olla aluseks kõrgele rekombinatsioonilisele aktiivsusele ning ümberkorraldustele piirkonnas.

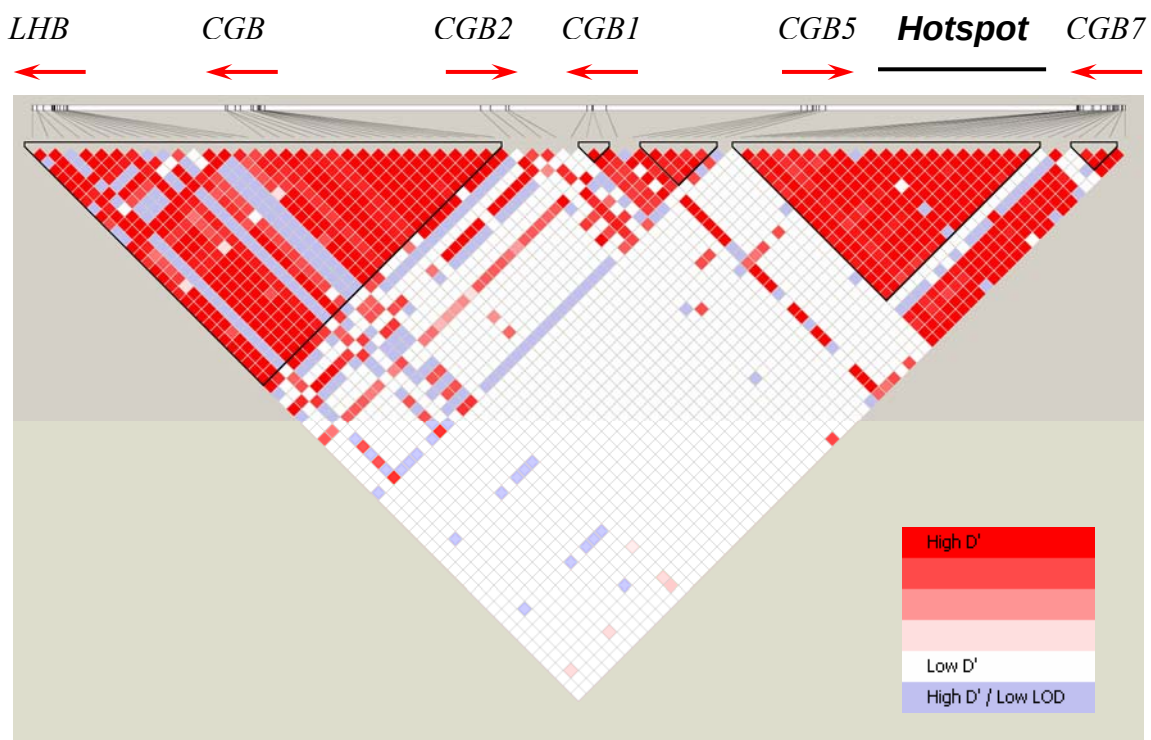


Joonis 17. Mfold programmiga ennustatud *psNTF6A'* ja *CGB7* vahelise DNA fragmendi sekundaarne struktuur. **A.**- Mfold programmi sekundaarne struktuur. Identifitseeritav on ~600 bp pikkune *stem* osa ning ~200 bp pikkune *loop* fragment. Märkitud on uuritava järjestuse algus- (*psNTF6A'* telomeeri poolsest otsast) ja lõpp-punkt (*CGB7* tsentromeeri poolse otsani). Samuti on lisatud *E.coli* χ -järjestuse paiknemine ning ära näidatud χ ja *CGB7* vahelise ala ebaspetsiifiline struktuur. **B.** Sekundaarse struktuuri analüüsis kasutatud järjestuse paiknemine *LHB/CGB* geeniklastris.

4.2. Potentsiaalse rekombinatsiooni *hotspot*'i ning *CGB8* iseloomustamine

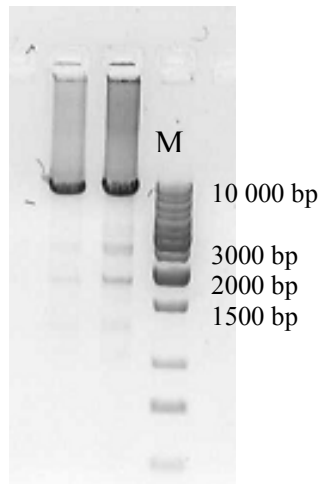
4.2.1. *Hotspot*'i piirkonna amplifitseerimine ning sekveneerimine

Eelnevalt *LHB/CGB* geeniperekonna resekveneerimisel ning varieeruvuse analüüsil (P. Hallast, 2004) identifitseeriti *CGB7* ja *CGB5* vahelises piirkonnas oluliselt kõrgem rekombinatsiooni määr võrreldes ülejäänud piirkonnaga. Samuti oli detekteeritav järsk alleelse aheldatuse taseme langus antud regioonis (Joonis 18), mis viitab potentsiaalse rekombinatsiooni *hotspot*'i esinemisele piirkonnas. Geeniklastri arvukate korduvate järjestuste tõttu ei olnud antud töö raames võimalik *CGB8* eraldi amplifitseerida ning iseloomustada.



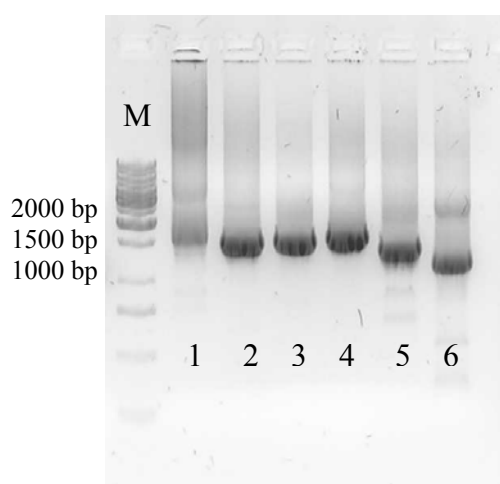
Joonis 18. Sekveneeritud *LHB/CGB* geenide (v.a. *CGB8*) alleelse aheldatuse muster ning ennustatud rekombinatsiooni *hotspot* Eesti populatsioonis (P.Hallast, 2004). Tugevas LD-s olevad alad on tähistatud punase/roosaga, madala LD-ga piirkonnad valgega ning mitte-informatiivsed alad sinisega. Musta joonega on ümbritsetud programmi poolt ennustatud LD-blokid.

Identifitseeritud potentsiaalse rekombinatsiooni *hotspot*'i detailsemaks analüüsiks amplifitseeriti üles kogu piirkond geenide *CGB5* ja *CGB7* vahel, hõlmates ka seni resekveneerimata *CGB8* (Joonis 15). Praimerid disainiti *CGB5* ja *CGB7* geenide 3' otsa, amplifitseerides seega 8338 bp pikkuse ala (Joonis 19). Seoses kirjeldatud genoomse regiooni komplekse ülesehitusega ei olnud võimalik kasutada PCR-l 100% unikaalseid praimereid. Sobivaimad disainitud praimerid seonduvad lisaks ka amplifitseeritava piirkonna seesmistele aladele. Kuna aga väljaspool amplifitseeritavat 8,3 kb pikkust ala disainitud praimeritel seondumiskohti ei esine, siis ei sega PCR-l tekkivad alternatiivsed produktid järgnevaid analüüsi etappe.



Joonis 19 Amplifitseeritud produkt (8,3 kb) rekombinatsiooni *hotspot*'i piirkonnast. 0,8 % agarosgeel. M- GeneRuler™ 1kb DNA Ladder (MBI Fermentas)

Järgmise etapina viidi läbi nested-PCR, kus substraadiks oli eelnevalt amplifitseeritud 8,3 kb PCR produkt. Kokku kasutati 6 paari nested praimereid (kaks neist kattusid 8,3 kb PCR-l kasutatavate praimeritega), mis olid 8,3 kb piirkonna suhtes unikaalsed, ülekattuvate otstega ning tagasid oodatava pikkusega produkti (Joonis 20.).



Nr geelil	Piirkond	Pikkus (bp)
1	Nested 1	1557
2	Nested 2	1577
3	Nested 3	1604
4	Nested 4	1675
5	Nested 5	1454
6	Nested 6	1195

Joonis 20. Nested-PCR produktid 1% agarosgeelil. M- GeneRuler™ 1kb DNA Ladder (MBI Fermentas). Tabelis on välja toodud vastavate nested produktide pikkused

Kõik korrektselt amplifitseeritud nested-PCR produktid sekveneeriti kokku 6 erineva praimeriga (3 praimerit *sense* ja 3 *antisense* ahela jaoks) (Joonis 15). Iga indiviidi kohta laekus seega 36 DNA järjestust potentsiaalse rekombinatsiooni *hotspot*'i piirkonnast. Järjestuse kvaliteedi kontrollimiseks viidi läbi võrdlus NCBI andmebaasi referentsjärjestusega, kus ainsateks erinevusteks olid polümorfised positsioonid. Kuna sekveneeritud piirkond on andmebaasiga võrreldes korrektne ning alternatiivseid järjestuste variante sekveneerimisel ei detekteeritud, siis võib kinnitada, et 8,3 kb PCR-l amplifitseeritud alternatiivsed produktid pärinevad sama piirkonna sisemistelt aladelt ning ei tekita tulemuste interpreteerimisel komplikatsioone.

4.2.2. Hotspot'i piirkonna polümorfismide jaotuvus

CGB5 ja *CGB7* vahel paikneva potentsiaalse rekombinatsiooni *hotspot*'i kirjeldamiseks sekveneeriti antud piirkond kokku 11 indiviidil Eesti populatsioonist. Töö tulemusena identifitseeriti 64 markerit, millest 20 asuvad *CGB5* ja *CGB8* vahelises alas, 36 *CGB8* ja *CGB7* vahelises regioonis ning 8 SNP-d *CGB8* geenis (Tabel 3) (Lisa 3). Insertioone/deletioone (indel) leidis kokku 6, mis asetsesid ainult intergeensetes piirkondades.

Tabel 3. Potentsiaalse rekombinatsiooni *hotspot*'i piirkonna varieeruvus

Piirkond	Pikkus (bp)	Polüm. arv	Indel	Min. al. sag. >10%	Singleton'de arv	Polüm./ kb
5' <i>CGB8</i> -st	2147	20	4	16	4	~9
<i>CGB8</i>	1466	8	0	5	3	~5
3' <i>CGB8</i> -st	4725	36	2	27	7	~7
Kokku	8338	64	8	48	14	~8

Keskmiseks markeri sageduseks kogu sekveneeritud ala kohta tuleb 8/kb. Intergeensete piirkondade polümorfismide tihedus on oodatavalt kõrgem (~7-9/kb) võrreldes *CGB8* järjestusega (~5/kb). Ülejäänud *LHB/CGB* perekonna geenide varieeruvus on keskmiselt võrdväärne (~5,3/kb) (P. Hallast, 2004).

Analüüsitud markerite alleelisagedused sekveneeritud piirkonnas olid valdavalt >10% - kogu potentsiaalse *hotspot*'i regioonis kokku 72%, *CGB8* ala piires 62,5% (ülejäänud klasteri geenides keskmiselt 72%). Singleton positsioone (varieeruvus detekteeritud ainult ühel kromosoomil valimis) leiti kokku 14 8,3 kb pikkuse järjestuse peale (Tabel 3).

CGB8 geenist identifitseeritud SNP-d lokaliseerusid kõik intronite või UTR-de piirkonda ning seega puuduvad potentsiaalselt aminohappelist muutust või raaminihet genereerivad markerid (Tabel 4.). Markerite suurem tihedus 5' UTR-s ning intron 1-s on korrelatsioonis teiste *LHB/CGB* perekonna geenide põhjal leitud andmetega (P. Hallast, 2004). *CGB8* polümorfismidest on NCBI andmebaasis esindatud ainult kaks (Tabel 4).

Tabel 4. SNP-de jaotuvus *CGB8* geenis ning esinemine NCBI andmebaasis

Piirkond	SNP-de arv	Arv NCBI's
5' UTR	4	1 (rs13345685)
Intron 1	3	-
Intron 2	1	1 (rs13345575)

Kõikide sekveneerimisel leitud markerite allelisageduste vastavust Hardy-Weinbergi tasakaalule kontrolliti HaploView programmi abil (P -väärtus <0.05) ning probleemseid positsioone ei identifitseeritud.

Kogu sekveneeritud *hotspot*'i piirkonna varieeruvuse analüüsil hinnati π , θ ja Tajima's D parameetreid. *CGB8* ala nukleotiidne varieeruvus (Tabel 5) oli sarnane klasteri keskel olevate *CGB5*, 1, 2 geenidega ($\pi \approx 0.00155$) (Hallast, Nagirnaja et al., käsikiri). θ ja Tajima's D väärtused olid samuti kõige sarnasemad kõrvalasuva *CGB5*-ga. Statistiliselt olulist erinevust null-hüpooteesist (kõik mutatsioonid on selektiivselt neutraalsed, $P < 0.05$) Tajima's D puhul ei esinenud. *CGB8*-ga piirnevate intergeensete alade nukleotiidne varieeruvus oli suurem, mis on oodatav tulemus mitte-kodeerivate alade puhul (Tabel 5).

Tabel 5. Rekombinatsiooni *hotspot*'i piirkonna nukleotiidne varieeruvus ja neutraalsustest võrreldes *CGB5*-ga.

Piirkond	π	θ	D
<i>CGB5-CGB8</i>	0.00257	0.00219	0.61
<i>CGB8</i>	0.00152	0.00178	-0.54
<i>CGB8-CGB7</i>	0.00227	0.00208	0.35
<i>CGB5</i>	0.00155	0.00153	0.03

4.2.3. Haplotüüpide jaotuvus

CGB5 ja *CGB7* vahelise ala resekveneerimisel leitud markerite genotüüpide põhjal ennustati ka piirkonna haplotüüpide mustrit kasutades Li ja Stephens'i (2003) algoritmi programmis PHASE. Geen *CGB8* hõlmab 10 erinevast haplotüüpi, millest kolm on ülejäänutest kõrgema sagedusega (>10%) (Tabel 6).

Tabel 6. *CGB8* haplotüüpide muster 11 indiviidi (22 kromosoomi) genotüüpide põhjal. S- singeleton (haplotüübi esinemine ainult ühel kromosoomil).

Haplotüüp	Esinemise arv (22 krom)	Esinemise %
<i>CGB8</i> 1	7	31.8
<i>CGB8</i> 2	3	13.6
<i>CGB8</i> 3	3	13.6
<i>CGB8</i> 4	2	9
<i>CGB8</i> 5	2	9
<i>CGB8</i> 6	1	S
<i>CGB8</i> 7	1	S
<i>CGB8</i> 8	1	S
<i>CGB8</i> 9	1	S
<i>CGB8</i> 10	1	S

Intergeensed piirkonnad osutusid võrreldes *CGB8*-ga tunduvalt varieeruvamaks. Lisaks suuremale markerite arvule on ka haplotüüpide jaotuvus mitmekesisem. *CGB5-CGB8* piirkond hõlmas 23 polümorfismi, mis moodustasid kokku 14 erinevat haplotüüpi, millest kolme sagedused on >10% (Tabel 7) *CGB8-CGB7* intergeenne ala on eelmisega võrreldes pikem, sisaldab rohkem markereid (36) ning moodustab ka suurema arvu haplotüüpe. Erinevalt *CGB5-CGB8* piirkonnast ei ole antud fragmendis kujunenud enamlevinutaid haplotüüpe, mille sagedus ületaks 10% läve (Tabel 7). Suur arv madala sagedusega haplotüüpe võib viidata kõrgeenenud mutatsioonide tekkimise tasemele, kuna tegemist on siiski intergeense alaga, või kõrgeenenud rekombinatsioonilisele aktiivsusele.

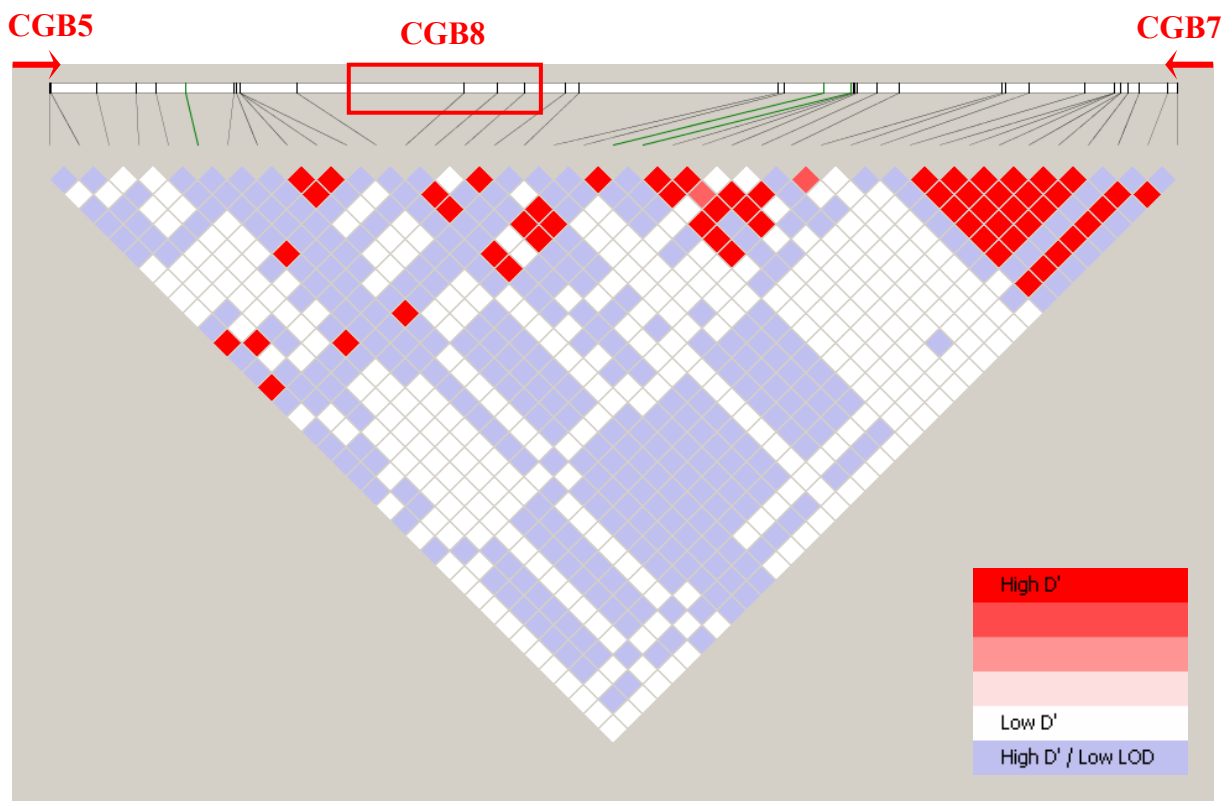
Tabel 7. Intergeensete alade haplotüüpide muster 11 indiviidi (22 kromosoomi) genotüüpide põhjal. S- singeleton (haplotüübi esinemine ainult ühel kromosoomil).

Haplotüüp	Esinemise arv (22 krom)	Esinemise %	Haplotüüp	Esinemise arv (22 krom)	Esinemise %
CGB5-CGB8_1	4	18.2	CGB8-CGB7_1	2	9
CGB5-CGB8_2	3	13.6	CGB8-CGB7_2	2	9
CGB5-CGB8_3	3	13.6	CGB8-CGB7_3	2	9
CGB5-CGB8_4	2	9	CGB8-CGB7_4	2	9
CGB5-CGB8_5	1	S	CGB8-CGB7_5	1	S
CGB5-CGB8_6	1	S	CGB8-CGB7_6	1	S
CGB5-CGB8_7	1	S	CGB8-CGB7_7	1	S
CGB5-CGB8_8	1	S	CGB8-CGB7_8	1	S
CGB5-CGB8_9	1	S	CGB8-CGB7_9	1	S
CGB5-CGB8_10	1	S	CGB8-CGB7_10	1	S
CGB5-CGB8_11	1	S	CGB8-CGB7_11	1	S
CGB5-CGB8_12	1	S	CGB8-CGB7_12	1	S
CGB5-CGB8_13	1	S	CGB8-CGB7_13	1	S
CGB5-CGB8_14	1	S	CGB8-CGB7_14	1	S
			CGB8-CGB7_15	1	S
			CGB8-CGB7_16	1	S
			CGB8-CGB7_17	1	S
			CGB8-CGB7_18	1	S

4.2.4. Rekombinatsioonilise aktiivsuse analüüs

Meiootiliste ristsiirete toimumisel kõrge sagedusega ehk rekombinatsiooni *hotspot*'i esinemisel teatud piirkonnas langeb alleelse aheldatuse tase antud fragmendis kahel pool ristsiirde saiti olevate markerite vahel. Seega kontrollimaks CGB5 ja CGB7 vahelisel alal leiduva rekombinatsiooni *hotspot*'i hüpoteesi, viid läbi sekveneerimisel kogutud andmete põhjal alleelse aheldatuse analüüs ning rekombinatsiooni määra (ρ) hindamine.

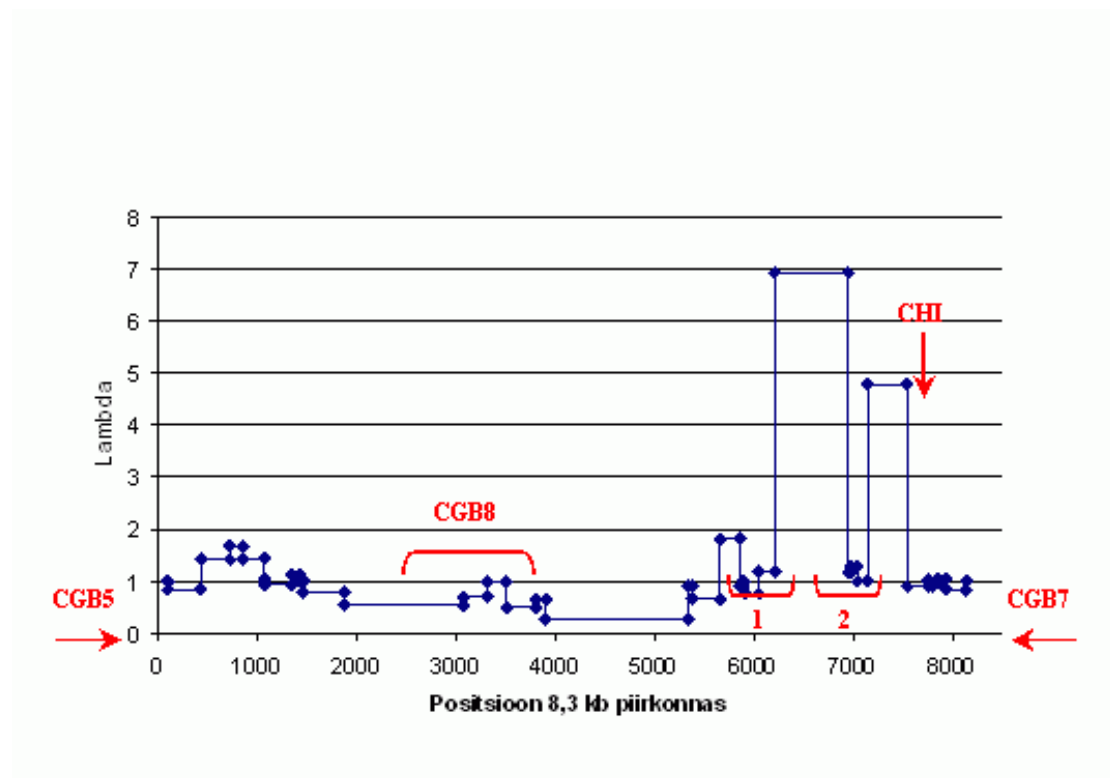
LD analüüsil ennustatud alleelse aheldatuse mustris (Lewontin'i D' koos LOD score'i väärtusega) ei moodustunud ühtegi LD-blokki uuritavate markerite vahel (Joonis 21). Kõrgem LD tase on nähtav ainult CGB7 läheduses, mis on varemalt detekteeritud kui LD-s olev piirkond (Joonis 18., P.Hallast, 2004). LD-blokkide puudumine on ka loogiline, eeldades et tegemist on rekombinatsiooniliselt aktiivse piirkonnaga ning seega pole võimalust alleelse aheldatuse kujunemiseks. Selguse mõttes on LD analüüsist välja jäetud madala sagedusega (<10%) markerid, mida käsitletakse programmi poolt mitte-informatiivsetena.



Joonis 21 Alleelse aheldatuse muster *CGB5* ja *CGB7* vahelises intergeenses piirkonnas. Markerite minoorse alleeli sagedus >10%. Tugevas LD-s olevad alad on tähistatud punase/roosaga, madala LD-ga piirkonnad valgega ning mitte-informatiivsed alad sinisega.

Lisaks alleelse aheldatuse analüüsile määratleti ka antud piirkonna rekombinatsiooni taset iseloomustav parameeter ρ kasutades statistilist algoritmi programmis PHASE (Li ja Stephens, 2003) ($\rho=4N_e r$, kus N_e on efektiivne populatsiooni suurus ja r rekombinatsiooni määr generatsiooni kohta). Kogu 8,3 kb regiooni tausta ρ väärtuseks saadi $\rho = 3,6/\text{kb}$, mis on ligi 10x kõrgem genoomi keskmisest ($0,4/\text{kb}$). Samuti on välja arvatud lambda väärtus (λ), mis iseloomustab iga järjestikuse markeri vahel esinevat rekombinatsioonilist aktiivsust võrreldes tausta ρ -ga. Tulemuste põhjal ilmnas, et *CGB8* ning *CGB7* vahelises alas on rekombinatsiooni toimumine 7x sagedasem võrreldes ülejäänud analüüsitud piirkonnaga (Joonis 22). Kuna tausta rekombinatsiooni määr on ligi 10x kõrgem

genoomi keskmisest, siis võib leitud *hotspot*'i aktiivsust lugeda 70x kõrgemaks genoomi keskmisest.



Joonis 22 .Rekombinatsiooniline aktiivsus *CGB5* ja *CGB7* vahelises piirkonnas. (markerite alleelisagedused >10%). Graafikule on märgitud *E. coli* χ -järjestuse ning *CGB8* suhteline asukoht. Numbritega 1 ja 2 on tähistatud kahe pöördkordusjärjestuse paiknemine rekombinatsiooni *hotspot*'s, mis potentsiaalselt moodustavad ühe *stem-loop* struktuuri.

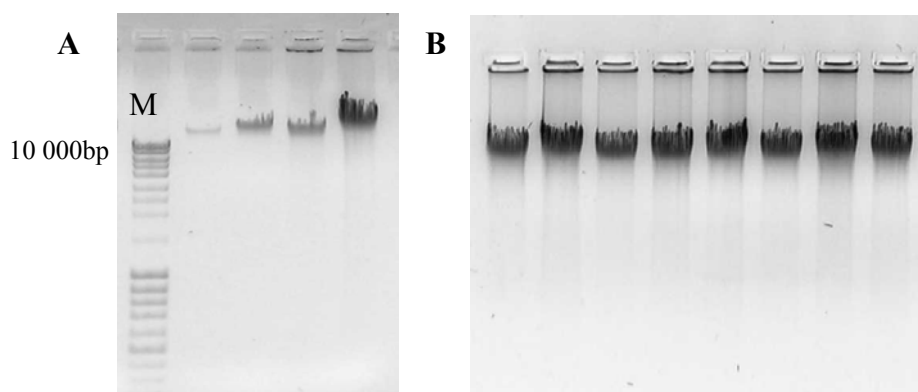
Identifitseeritud *hotspot*'i vahetus läheduses (<100bp) 3' suunas lokaliseerub rekombinatsiooni stimuleeriv χ -järjestus ning täpselt *hotspot*'i alas stabiilset sekundaarset struktuuri moodustavad palindroomid (Joonis 22).

4.3. *LHB/CGB* geenide varieeruvus korduvate spontaanabortidega patsientidel versus referentspopulatsioonil

4.3.1. *LHB/CGB* geenide amplifitseerimine ning sekveneerimine

LHB/CGB geenide analüüsiks ning võrdluseks kontroll-populatsiooniga koguti 24 korduvate spontaanabortidega patsientide ja nende meespartnerite proovist koosnev DNA-pank. Võrdlusena kasutati eesti populatsiooni valimi (47 indiviidi) genotüüpide materjali, mis pärinevad eelnevalt sekveneeritud *LHB/CGB* geenide analüüsi andmetest (P. Hallast, 2004). *CGB8* võrdluseks patsientide andmetega kasutati potentsiaalse rekombinatsiooni *hotspot*'i analüüsil saadud 11 eestlase genotüüpide andmeid.

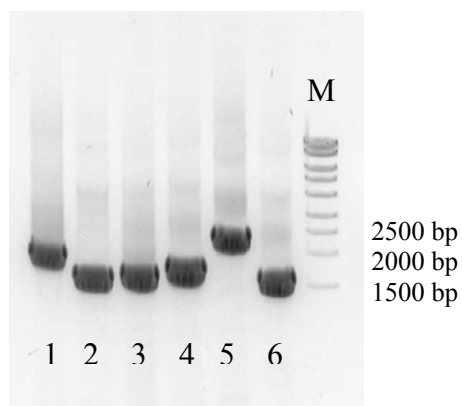
LHB/CGB geenide amplifitseerimiseks eraldati esmalt vere- ja platsentaproovidest genoomne DNA, mille kvaliteeti kontrolliti agarosgeelil ning spektrofotomeetriselt. Kõikide eraldatud DNA-de kvaliteet agarosgeelil vastas nõuetele (Joonis 23) ning spektrofotomeetri andmete põhjal jäi enamike DNA-de puhtuse aste vahemikku 1,72-2. Kontrollitud DNA-de kontsentratsioon mõõdeti spektrofotomeetri abil ning keskmiseks saagikuseks oli verede puhul 38 µg (5-9 ml-st) ning platsentade puhul 24 µg (~30 mg koetükist).



Joonis 23. Vereproovidest ning platsentast eraldatud genoomne DNA 0,8% agarosgeelil. A – Platsenta genoomne DNA, B – Vere genoomne DNA. M- MassRuler™ DNA Ladder Mix (MBI Fermentas).

LHB/CGB geenide amplifikatsiooniks patsientide grupi genoomselt DNA-lt kasutati praimereid, mis on eelnevalt välja töötatud ning kvaliteedi suhtes kontrollitud (Hallast, Nagirnaja et al, käsikiri). *CGB8* analüüsimiseks kasutati nested-PCR

strateegiat nagu on kirjeldatud osas 3.4.2. Ülejäänud *LHB/CGB* geenide amplifikatsioonil kasutati geeni-spetsiifilisi praimereid, millest vähemalt üks on unikaalse järjestusega ning tagavad teadaolevalt korrektse produkti (Joonis 24). Kõikide produktide õigsust kontrolliti geelelektroforeesil, kus kõrvalekaldeid oodatavast pikkusest ei esinenud



Nr geelil	Piirkond	Pikkus (bp)
1	<i>CGB</i>	1887
2	<i>CGB1</i>	1600
3	<i>CGB2</i>	1652
4	<i>CGB5</i>	1757
5	<i>CGB7</i>	2364
6	<i>LHB</i>	1599

Joonis 24. Amplifitseeritud *LHB/CGB* geenide (v.a. *CGB8*) produktid 1% agarosegelil. (geenide nimetused koos oodatava pikkusega) M - MassRuler™ DNA Ladder, High Range (MBI Fermentas)

4.3.2. *LHB/CGB* geenide üldine varieeruvus patsientidel ja referentsvalimil

LHB/CGB geenide resekveneerimisel identifitseeriti kokku 120 markerit, mis hõlmasid nii insertioone kui ka deletsioone (n=5) (Lisa 4). Patsientide ning referentsvalimi polümorfsete positsioonide võrdlusel esinesid väikesed erinevused: (i) patsientidel ei tuvastatud kokku 25 polümorfismi, mis olid esindatud referentsindiviididel, kuid samas (ii) identifitseeriti patsientide grupi resekveneerimisel 15 unikaalset polümorfismi. Mõlemal juhul on erinevusi põhjustavate positsioonide minoorse alleeli sagedus <10%. Paarikaupa markerite alleelisageduste võrdlusel ei ilmnenud olulisi kõikumisi, mis viitaks mõne alleeli rikastumisele või alaesindatusele patsientide valimis.

LHB/CGB geenides on identifitseeritud mitmeid aminohappe muutust põhjustavaid polümorfisme, mis võivad potentsiaalselt häirida valkude normaalset funktsioneerimist. Seetõttu kontrolliti antud markerite esinemist ka patsientidel. Ainsad positsioonid, mis oluliselt (>10%) erinesid patsientidel võrreldes referentspopulatsiooniga, olid *CGB7* geenis lokaliseeruvad kolm markerit: c7E417, c7E422, c7E423 (Tabel 8). Esindatud olid ka kõik ülejäänud teadaolevad eksonites olevad varieeruvused. Lisaks identifitseeriti ühel indiviidil üks uus mutatsioon *CGB7* geeni ekson 2-s (Lisa 4, c7E473), kus C deletsiooni tagajärjel toimub raaminihe ning seega tõenäoliselt on hCG valk mitte-funktsionaalne.

Erinevuste puudumine kahe valimi vahel enamiku positsioonide puhul, mis põhjustavad aminohappelist muutust, võib seletada kahel viisil. Esiteks ei pruugi aset leidvad muutused kaasa tuua valgu funktsionaalseid häireid. Alternatiivne hüpotees on, et referentsina kasutatavas üldpopulatsioonis esineb indiviide, kellel on samuti probleeme raseduse normaalse kulgemisega. Seetõttu on edasi planeeritud läbi viia projekti järgmine etapp, kus korduvate spontaanabortidega patsiente võrreldakse sobiva kontrollpopulatsiooniga, kellel ei ole esinenud problemaatilisi raseduse juhtumeid.

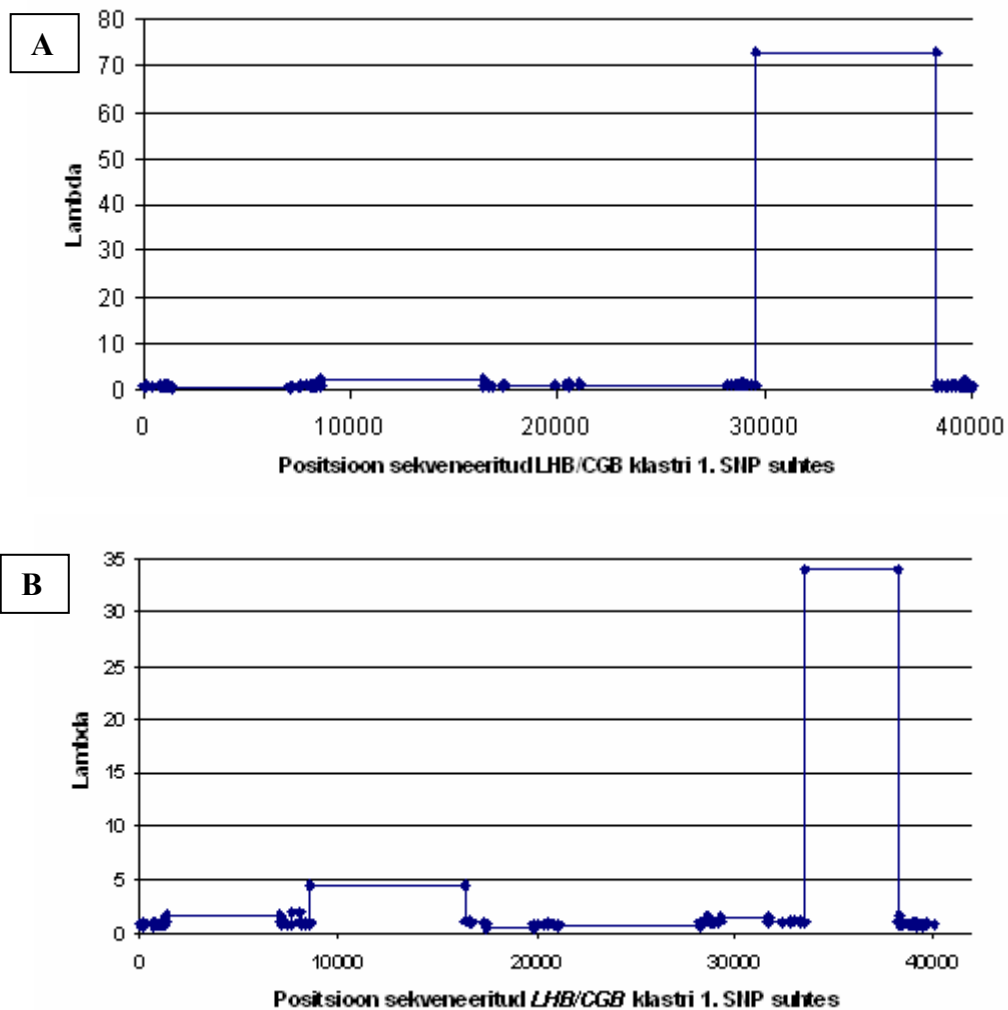
Tabel 8. *LHB/CGB* geenide eksonites leiduvad polümorfismid, vastavalt polümorfismile koodoni ja aminohappeline muutus ning sagedus patsientidel *vs* referentsvalimil. * - aminohappe muutust põhjustav positsioon.

Geen	Nr	SNP positsioon	Ekson	Muutus	Amino-hape	Koodon	Sagedus patsientidel	Sagedus referentspop-s
LHB	1	397*	2	ATG-ATA	Met-Ile	15	-	0,02
	2	434*	2	TGG-CGG	Trp-Arg	28	0.094	0.12
	3	466	2	GTC-GTG	Val-Val	38	0.312	0.3
	4	484	2	CCA-CCC	Pro-Pro	44	0.364	0.41
	5	873	3	GGT-GGC	Gly-Gly	95	Mitte-informatiivne	0.29
CGB	6	1026	3	CCC-CCG	Pro-Pro	146	0,333	0.33
	7	1038	3	AGC-AGT	Ser-Ser	150	0,354	0.33
CGB2	8	531	2	CCG-CCA	Pro-Pro	44	0.125	0.13
CGB1	9	458*	2	GAG-GGG	Glu-Gly	20	0.271	0.20
CGB5	10	424	2	CCG-CCA	Pro-Pro	24	0.333	0.20
	11	893	3	TAC-TAT	Tyr-Tyr	102	0.042	0,020
CGB7	12	417*	2	AGG-AAG	Arg-Lys	22	0.333	0.41
	13	422*	2	ATG-CTG	Met-Leu	24	0.333	0.41
	14	423*	2	ATG-ACG	Met-Thr	24	0.333	0.41
	15	867*	3	CGC-TGC	Arg-Cys	94	0.021	0,05
	16	1061	3	TCA-TCG	Ser-Ser	158	Mitte-informatiivne	0.47

LHB/CGB geeniklastri alleelse aheldatuse muster korduvate spontaanabortidega patsientidel (n=24) ning üldpopulatsioonil (n=47) oli vaatamata indiviidide erinevale arvule sarnane. Rekombinatiooni parameetrite ρ ja λ (Li ja Stephens'i algoritm, 2003) (Tabel 9) võrdlusel oli küll λ väärtus rekombinatsiooni *hotspot*'i piirkonnas üldpopulatsioonis kõrgem, mis võib tuleneda suuremast indiviidide arvust, kuid üldine muster kahe valimi vahel ei erine (Joonis 25). Kuna nii alleelse aheldatuse kui ka rekombinatsiooni muster on kahel võrdluse all oleval populatsioonil sarnane, siis on ka järgnevalt ennustatud *LHB/CGB* geenide haplotüüpid võrreldavad antud valimite vahel.

Tabel 9. Rekombinatsiooni parameetrite ρ ja λ väärtused võrdlevalt korduvate spontaanabortidega patsientidel ning üldpopulatsioonil. Polümorfismid minoorse alleli sagedusega >10%. Λ – rekombinatsiooni suhteline aktiivsus kordades kahe markeri vahel võrreldes ρ -ga. (Genoomi keskmine $\rho=0,4/\text{kb}$).

Parameeter	Patsiendid	Üldpopulatsioon
Klastri ρ /kb	0.56	0.44
<i>CGB-CGB2</i> λ	4.4	2.2
<i>CGB5-CGB7</i> λ	34	72.6



Joonis 25. Rekombinatsiooni parameetri λ (lambda) väärtused kogu analüüsitud LHB/CGB klatri ulatuses võrdlevalt korduvate spontaanabortidega patsientidel ning üldpopulatsioonis. A – üldpopulatsioon, B – patsiendid. Λ – rekombinatsiooni suhteline aktiivsus kordades kahe markeri vahel võrreldes ρ -ga.

4.3.3. Haplotüüpide analüüs

Haplotüüpide sageduste ning mustri määratlemiseks viidi läbi resekveneerimisel identifitseeritud genotüüpide info analüüs Li ja Stephens'i (2003) algoritmil põhineva programmiga PHASE. Ennustatud haplotüüpide võrdlusel patsientide grupi ning üldpopulatsioonis põhineva valimi vahel ilmnemiseid mitmed olulised erinevused.

Jaotuvuse analüüsiks järjestati ennustatud haplotüübid geeni kaupa ning võrreldi sageduse alusel kahe populatsiooni vahel. Oluliseks erinevuseks loeti

sageduse kõikumist ~10 % või rohkem. Antud kriteeriumile vastas mitmed erinevate geenide piirkonnas leiduvad haplotüübid:

1. *CGB1* geeni enimlevinud haplotüüp oli patsientidel laesindatud 9,4% võrra (referentsvalim n=47, kastiga on ümbritsetud aminohappe muutust põhjustav alleel),

Nr	<i>CGB1</i> haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1*	GGTTGGGCT	25,00	34,04
2	GGTTGGGAT	25,00	27,66
3	GGT□AGGCT	20,83	20,21
4	GGCTGGGCT	12,50	7,45
5	GGCTGGGAT	4,17	5,32
6	AGTTGGGAT	4,17	3,19

2. rekombinatsiooni *hotspot*'i alaga piirnev *CGB8* geenis on haplotüüp 2 ülesindatud >11% võrra (referentsvalim n=11),

Nr	<i>CGB8</i> haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1	CTTGCGCGACGCC	45,83	50,00
2*	CTTGCGGGACACA	29,17	18,18
3	CTTGCGCGTCGCC	8,33	9,09
4	CTTGCAATGGC	8,33	4,55
5	CTTGCGCGACGGC	4,17	4,55

3. kõige suurema varieeruvusega tulemusi andis *CGB5*, kus enimlevinud haplotüüp 1 on ~24% võrra patsientidel laesindatud, kuid samas esineb haplotüüp 2 sagedusega ~21% võrreldes ~4%-ga referentspopulatsioonis. Samuti on *CGB5* piirkonda ennustatud haplotüüp, mille sageduseks on üldpopulatsioonis ~10%, kuid patsientidel puudub (referentsvalim n=47).

Nr	<i>CGB5</i> haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1*	ACGTCGCGGCCACCG	37,50	61,70
2*	ACCGCATAGCCACCG	20,83	4,26
3	ACGGCGTGGCTACCG	10,42	11,70
4*	GCCGCATAGCCACCG	-	9,57
5	GCGTCGCGGCCACCG	4,17	-
6	ACCGCATAGCCACTG	4,17	-
7	GCGTCGCGGCTACCG	4,17	-
5	GCCGCATAGCCACTG	-	2,13

4. *CGB7* geeni puhul on samuti nähtavad erinevused, kus kõige sagedasem haplotüüp 1 esineb patsientidel 16% võrra sagedamini. (referentsvalim n=11, kastiga on ümbritsetud aminohappe muutust põhjustav alleel).

Nr	<i>CGB7</i> haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1*	CAGCCCTWGAGTGWATCCATGGTTTCCTA	27,08	11,70
2	CAGCCCCWGAGTGWATCCGTGGTTTGGTG	25,00	18,09
3	CCAGGATWGGGTCWGGTACATCGGCGTA	18,75	25,53
4	CCAGGATWGAGCGWGGTACATTGGCGCA	12,50	9,57
5	CAGCCCCWGAGTGWATCCGTGGTTTGGTA	-	6,38
6	CAGCCCCWGAGTGWATCCGTGGTTTGCTG	2,08	3,19
7	CAGCCCCWGAATGWATCCGTGGTTTGGTA	-	3,19
8	CAGCCCCWAGTGWATCCGTGGTTTGGTG	-	2,13
9	CAGCCCTWGAGTGWATCCGTGGTTTGGTG	-	2,13
10	CAGCCATWGAGTGWATCCATGGTTTCCTA	-	2,13

Geenides *LHB*, *CGB* ning *CGB2* olulisi erinevusi ei leitud (Tabel 10).

Tabel 10. Haplotüüpide jaotuvus geenides *LHB*, *CGB* ning *CGB2* võrdlevatl korduvate spontaanabortidega patsientidel (n=24) ning üldpopulatsioonil (n=47). D – T-deletsioon. Kastiga on ümbritsetud aminohappe muutust põhjustav alleel.

Nr	<i>LHB</i> haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1	CGGGDGGCAGCTGACTCCT	56,25	51,06
2	TTGATATGACACTGACTCTA	22,92	14,89
3	CTAGDGTGCGCTGACCTTT	6,25	8,51
4	TTGATATGACACCGACTCTA	4,17	9,57
5	TTGATATGACACTTACTCTA	2,08	4,26
6	CGGGDGGCAGCTGACTCCT	2,08	2,13
7	CGGGDGGCATGCTGACTCCT	2,08	2,13

Nr	<i>CGB</i> haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1	GACATGGCCAGGGGTTTCGTA	50,00	53,19
2	TGGGCGGGTCTGGACGTCCG	20,83	20,21
3	GACATGGCCAGGTGTTTCGTA	10,42	4,26
4	GGGGCGGGTCTGGACGTCCG	4,17	1,06
5	GACACGGCCAGGGGTTTCGTA	2,08	6,38
6	TGGGCGGCCAGAGGCTCGTA	2,08	2,13
7	TGGGCGGGTCTGGGCGTCCG	2,08	2,13
8	TGGGCGGGTCTGTGCGTCCG	2,08	2,13

Nr	CGB2 haplotüübid	Sagedus patsientidel	Sagedus referents valimis
1	GCCGGCAT	29,17	32,98
2	GTCGGCGT	25,00	27,66
3	GTCGGAGT	25,00	21,28
4	GTCGAAGT	12,50	10,64
5	CCTGGCAT	4,17	2,13

Lisaks haplotüüpide üldisele jaotusele kontrolliti, kas aminohappe muutust põhjustavate alleelide esinemine korreleerub haplotüüpidega, mille puhul identifitseeriti oluline sageduse varieeruvuse kahe valimi vahel. Märkimisväärne on *CGB7* geeni polümorfismide c7E417, c7E422 ja c7E423 aminohapet muutvate alleelide koosesinemine kahel haplotüübil, millest üks, haplotüüp 3 (tabel käesoleva peatüki punkt 4-s), on üldpopulatsioonis enimlevinud. Samas on haplotüüp 3 alaesindatud patsientidel, kellel domineerivad alternatiivsed markerite kombinatsioonid, mis ei hõlma aminohappelist muutust põhjustavaid allelele.

Seega on identifitseeritud resekvenerimise tulemusena nii haplotüüpe, mis võivad potentsiaalselt olla riskifaktoriks (eelsoodumuseks) korduva spontaanabordi kujunemiseks, kui ka markerite kombinatioone, mille esinemine tundub pigem olevat normaalset raseduse kulgu tagava efektiga (c7E417, c7E422 ja c7E423).

5. Arutelu

5.1. *LHB/CGB* geeniklastri struktuur ja evolutsioon

Eelnevate uuringute raames on inimese genoomist leitud kokku 6 kooriongonadotropiini β -subühikuid kodeerivat geeni ja 3 *NTF6* pseudogeeni ning täpselt määratletud nende asukoht ja orientatsioon üksteise suhtes. Detailsemal piirkonna analüüsil käesoleva töö raames aga ilmnes äärmiselt keeruline segmentaalsete duplikatsioonide muster ning seni teadaoleva info puudlikkus. Saadud tulemused viitavad mitmetele evolutsioonilistele duplikatsioonisündmustele, mutatsioonidele ning potentsiaalsele ebastabiilsusele ümberkorralduste suhtes.

LHB/CGB geeniklastri kirjeldamisel selgus nii geenide kui ka intergeensete alade suur sarnasus. Lisaks identifitseeriti *CGB1-1* ja *CGB2-1* oma unikaalne, võimaliku promootori, järjestus ning intergeensete alade sarnasus klastri telomeerses otsas paiknevale *NTF5*-le. Antud piirkond on *Alu*-rikas ning sisaldab rekombinatsiooni ja geenikonversiooni stimuleerivat *E. coli* χ -motiivi. *LHB/CGB* geenide piirkonna struktuur on seletatav korduvate mitte-alleelse homoloogilise rekombinatsiooni sündmuste esinemisega primaatide evolutsioonis, millega on kaasnenud arvukad duplikatsioonid, inversioonid ning deletsioonid. *LHB/CGB* geeniklastri on tõenäoliselt aluse pannud esialgne *LHB* geeni duplikatsioon 94 mat, mis hõlmas ka osa *NTF5* geenist. Kuna *NTF5* sisaldab *Alu*-rikast ala ning segmentaalsete duplikatsioonide levik genoomis kattub kirjanduse põhjal ajaliselt *Alu*-elementide ekspansiooniga primaatidel, siis võis *LHB/CGB* geeniperekond kujuneda üle *Alu-Alu* vahendatud rekombinatsiooni. Segmentaalsete duplikatsioonidele ning *Alu*-dele on iseloomulik ka erinevate ümberkorralduste soodustamine, mis seletaks *LHB/CGB* geeniklastri keerulist struktuurset ülesehitust tandemsete ning pöördkordusjärjestustena. Kromosomaalsete ümberkorralduste esinemisele viitab ka palindroomsetena paiknevate *CGB1* ja *CGB5* vahelise ala inversioon (Joonis 16).

Lisaks mitmetele rekombinatsioonisündmustele on *LHB/CGB* geeniklastris viiteid ka geenikonversiooni toimumisest. Esiteks on piirkonnas kõrge GC sisaldus (>56%, geenides kuni 65%), mis on üle genoomi keskmise (~45%). Antud nähtust seostatakse sageli geenikonversiooniga, kuna on leitud eelistatult AT konverteerimist GC-ks. Samuti seletaks geenikonversioon lisaks positiivsele selektsioonile mitme sarnase funktsionaalse *CGB* geeni säilimist paralleelselt, kuna eeldatavalt peaksid

duplitseerunud geenid muteeruma ning omandama uue funktsiooni või inaktiveeruma. Geenikonversiooni teel toimub aga info ülekandumine sarnaste geenide vahel (eriti 5' regulaatorsetes alades) ning seega paralleelselt mitme duplikaadi funktsioneerimist.

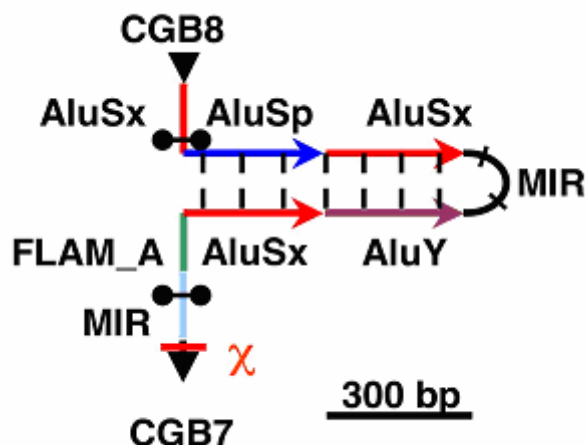
LHB/CGB geeniklastris on geenikonversioon tugevalt soodustatud LCR-de esinemise tõttu ning Hallast, Nagirnaja et al. (2005, käsikiri) poolt on ka kinnitatud konversiooni intensiivne toimumine antud piirkonnas, kus akseptorina käituvad klasteri äärealadel paiknevad geenid. Antud nähtus on seletatav akseptor-geenide (*CGB* ja *CGB7*) vahetus läheduses paiknevate nn rekombinatsiooni *warm*- ja *hotspot*'ga, mis omakorda veel destabiliseerivad *LHB/CGB* geeniklastri struktuuri stimuleerides rekombinatsioonilist aktiivsust ning uusi ümberkorraldusi.

5.2. Rekombinatsiooni *hotspot*

LHB/CGB geenide resekveneerimisel ning järgneval LD ja rekombinatsiooni parameetrite hindamisel leiti piirkonnas kõrgenenud rekombinatsioonilise aktiivsusega ala ehk *hotspot*, mis ületas märkimisväärselt regiooni keskmist (~73 korda) (Hallast, 2004). Identifitseeritud piirkonda analüüsiti seejärel detailsemalt resekveneerides kogu *hotspot*'i hõlmav fragment, mille tulemusena identifitseeriti rekombinatsioonisündmuste klasterdumine geenide *CGB8* ja *CGB7* vahelisse alasse.

Rekombinatsiooni intensiivsus leitud *hotspot*'s oli võrreldes genoomi keskmisega ca 70 korda kõrgem. Järjestuse detailsemal analüüsil selgusid mitmed faktorid, mis kirjanduse põhjal võiksid seletada niivõrd intensiivset rekombinatsiooni esinemist just antud piirkonnas:

1. vahetult *hotspot*'i läheduses (<100 nt) paiknev rekombinatsiooni soodustav *E. coli* χ -järjestus,
2. *NTF5* geenist pärit piirkonna rikastatus *Alu*-elementidega (76%), millega kaasnevad *Alu-Alu* vahendatud ümberkorraldused/rekombinatsioon (Bailey et al., 2003),
3. *stem-loop* sekundaarset struktuuri tekitavad pöördkordusjärjestused, mis hõlmavad suhteliselt noori palindroomseid *AluS* elemente (Joonis 26). Kirjanduse andmetel võib palindroomide esinemine piirkonnas tõsta rekombinatsioonilist aktiivsust mitmekümne kordselt (Lobachev et al., 1998).



Joonis 26. *Alu*-palindroomide poolt moodustatud *stem-loop* sekundaarne struktuur *hotspot*'i piirkonnas. Mustad jooned tähistavad suhtelist rekombinatsiooni *hotspot*'i ala piiride paiknemist. χ – *E. coli* rekombinatsiooni stimuleeriv järjestus. MIR – (*Mammalian wide interspread repeat*)

Analoogiline *hotspot*'i ehitus on leitav ka teistes genoomi piirkondades (Jeffreys ja Neumann, 2002). *LHB/CGB* klastris on kirjeldatud *hotspot*'i sarnane piirkonna ehitus ka *CGB* ja *CGB2* vahelises regioonis, mis LD ning rho ja lambda väärtuste hinnangul hõlmab ka potentsiaalset rekombinatsiooni *warmspot*'i. Sekundaarsete struktuuride ennustamisel aga ei moodustunud antud alasse nii stabiilset *stem-loop*'i kui *hotspot*'i alas. *CGB1* ja *CGB5* vahelises regioonis on aga toimunud klatri evolutsiooni käigus inversioon ning praktiliselt üleni on deleteerunud üks *stem-loop*'i moodustavatest palindroomidest.

Seoses erinevate järjestuste iseärasustega *hotspot*'i piirkonnas, võib rekombinatsiooniline aktiivsus antud regioonis olla seletatav mitme mehhanismiga. Ühelt poole soodustab rekombinatsiooni *E. coli* χ – järjestus, teisalt aga initsieeritakse rekombinatsioon, tekitades DSB-i spetsiifiliste sekundaarsete struktuuride tõttu. Alternatiivse hüpoteesi kohaselt (või paralleelselt) võib oluline roll olla ka kromatiini avatusel ning ligipääsetavusel DSB tekamiseks vajalikule Dnaasile ning rekombinatsiooni läbiviivatele valkudele. Kuna kirjeldatud *LHB/CGB* geeniklatri sisemised alad hõlmavad järjest sarnaseid duplitseerunud fragmente, siis võivad antud piirkonnad olla ka rohkem pakitud. Seega oleks seletatav *warm*- ja *hotspot*'i paiknemine just klatri äärealadele, kus tõenäoliselt on kromatiini struktuur avatum.

Kolme erineva rekombinatsiooni stimuleeriva faktori (*Alu*-d, χ - järjestus ja *stem-loop*) koosinemine toetab resekveneerimise andmete analüüsil saadud tulemusi ning seega saab väita, et antud piirkonnas esineb suure tõenäosusega rekombinatsiooni *hotspot*. Detailsem *hotspot*'i kirjeldamine otseselt haploidse sperma DNA põhjal (käesolevaks hetkeks DNA materjal kogutud, koostöö M. Punabiga, M.D. TÜ Kliinikumi Androloogia kabinet) võimaldaks teha põhjendatud järeldusi konkreetse rekombinatsiooni mehhanismi toimumise kohta ning määratleda eksperimentaalselt rekombinatsiooni ning geenikonversiooni intensiivsust.

5.3. *LHB/CGB* geenide assotsiatsioon korduvate spontaanabortide esinemisega

Kooriongonadotropiin on platsentas ekspresseeritav glükoproteiinhormoon, mille kontsentratsiooni alusel naise seerumist diagnoositakse raseduse olemasolu ning potentsiaalset kulgu. Vähene hCG tase on märk komplikatsioonidest raseduse normaalsel kulgemisel ning võib viia spontaanaborti kujunemisele.

HCG funktsioonide uuringutel on leitud selle kriitiline olulisus raseduse I trimestril, kus peamiseks determinandiks blastotsüsti pesastumisel on hCG piisav ekspressiooni tase varajase embrüo poolt. Seega võib eeldada, et mutatsioonid, koopia-arvu muutused ning ümberkorraldused hCG beeta-subühikuid kodeerivate geenide alas võivad viia spontaanabortide tekkimiseni. Praeguseks on kirjeldatud *LHB/CGB* geeniklastris ainult ühte mutatsiooni, mis põhjustab β -subühiku valesti pakkumist ning seega hormooni inaktiveerumist. Uuritud on ka geenide deletsioone/duplikatsioone, kuid siiani mingeid ümberkorraldusi identifitseeritud ei ole. Siiski ei tähenda see kromosomaalsete struktuuri ümberkorralduste või mutatsioonide puudumist *LHB/CGB* geeniklastris, kuna seni kasutatud meetodid võivad olla vähe-spetsiifilised ning mitte-informatiivsed piirkonna keerulise struktuuri tõttu.

Viiteid *LHB/CGB* geenide rollile spontaanabortide tekkimisel on hiljuti leitud ekspressiooniuringute raames. Viimaste andmete kohaselt leiti korrelatsioon hCG hormooni β -subühikut kodeerivate geenide *CGB*, *CGB5*, *CGB7* ja *CGB8* mRNA madala seerumi kontsentratsiooni ning korduvate spontaanabortide esinemise vahel (Rull ja Laan, 2005, ajakirja esitatud). Lisaks oli *CGB1* ja *CGB2* mRNA detekteeritav kontrollindiviididel I trimestri jooksul, kuid korduvate spontaanaborti patsientidel ei leitud mingeid märke transkriptsioonilisest aktiivsusest antud geenidelt. Kuigi

erinevate *CGB* geenide jaoks on identifitseeritud individuaalne ekspressiooniprofiil, on vähe informatsiooni selle funktsionaalse väljundi kohta. Arvatakse, et mitme erineva *CGB* geeni transkribeerimine (i) kindlustab piisava hulga hCG-d, et tagada raseduse säilumine varajases staadiumis või (ii) on mitme transkriptsiooniliselt aktiivse geenikoopia olemasolu nõ kaitseks kiiresti evolutsioneeruva klatri puhul (Jameson ja Hollenberg, 1993). Samas võib Rull ja Laan poolt esitatud *CGB1* ja *CGB2* andmete põhjal eeldada, et esineda võivad veel seni identifitseerimata *CGB* geenide valgulised produktid, mille funktsiooni ning olulisust ei osata hetkel hinnata. Seega võib järeldada, et spontaanabordi kujunemisel võivad olla määrava rolliga nii mutatsioonid, kromosomaalsed ümberkorraldused kui ka *CGB* geenide alternatiivne transkriptsiooni regulatsioon. Lisaks on leitud viiteid võimalikule imprintingu mustriks *CGB* geenides, kus hCGβ on ekspresseeritud isapoolselt (Marshall et al., 1998, McFadden et al., 1993). Seetõttu on hCG geenide analüüsil korduvate spontaanabortidega patsientidel informatiivne uuringusse kaasata ka meespartnerid.

Käesoleva töö raames disainiti pilootprojekt *LHB/CGB* geeniperekonna varieeruvuste võrdlevaks analüüsiks I trimestri korduvate spontaanabortidega patsientidel ja nende meespartneritel ning üldpopulatsioonil põhineval valimil. Projekti tulemusena identifitseeriti patsientidel alternatiivne haplotüüpide jaotuvus, mis kinnitab püstitatud hüpoteesi, et *LHB/CGB* geenide võivad osaleda spontaanabordi kujunemisel. Mitmed ennustatud *CGB* geenide haplotüübid olid rikastatud patsientide valimis (geenid *CGB5*, 8 ja 7) ning samas esines ka alleelide kombinatsioone, mille sagedus oli patsientidel tunduvalt madalam või puudusid hoopis. Seega võivad *LHB/CGB* geenid hõlmata nii spontaanaborti põhjustavaid riski-haplotüüpe kui ka normaalset rasedust tagavaid nn protektiivseid haplotüüpe.

Arvestades *LHB/CGB* geeniklatri ebastabiilsust segmentaalsete duplikatsioonide, *Alu*-elementide ning *E. coli* χ -järjestuste tõttu, on ümberkorralduste ning varieeruvuste esinemine *LHB/CGB* geeniklatriks eeldatav tulemus. Täpsemate andmete saamiseks viiakse järgnevalt läbi uuring, kus pilootprojekti raames sekveneeritud patsientidele on valitud kontrollindiviidid, kellel on rasedus kulgenud komplikatsioonideta (analüüsimaterjal kogutud, koostöö Kristiina Rulliga, M.D.). Lisaks on konkreetsete ümberkorralduste identifitseerimiseks plaanis kasutada hübriidisatsioonil põhinevat FISH metoodikat, mis võimaldab detailsemalt kirjeldada *LHB/CGB* geeniklatri varieeruvust korduvate spontaanabortidega patsientidel ning referentspopulatsioonil.

Käesoleva töö tulemused on olulised (i) segmentaalseid duplikatsioone sisaldava geeniklastri olemuse selgitamise seisukohast, mis viitab võimalikele varieeruvusi genereerivatele sündmustele ka teistes duplikatsiooni teel tekkinud klastrites. (ii) Kuna 50% spontaanabortidest klassifitseeritakse kui identifitseerimata põhjusega juhtumid, siis on käesolev uuring oluline ka spontaanabortide päriliku soodumuse määratlemise, diagnostika, ravi ning patsientide nõustamise seisukohast.

Kokkuvõte

Käesoleva magistritöö kirjanduse ülevaates on iseloomustatud genoomis laialdaselt levinud segmentaalseid duplikatsioone ning kirjeldatud erinevaid ümberkorralduste tekkemehhanisme. Detailsemalt on analüüsitud geenikonversiooni ja rekombinatsiooni nähtusi ning rekombinatsiooni *hotspot*'de omadusi. Lisaks on iseloomustatud hormooni kooriongonadotropiini, *LHB/CGB* geeniklastrit ning esitatud põgus ülevaade spontaanabortide tekkepõhjuste kohta.

Eksperimentaalne osa koosnes kolmest etapist. Esmalt määratleti *in silico* *LHB/CGB* geeniklastri detailne struktuur. Seejärel iseloomustati eelnevalt *LHB/CGB* geeniklastris ennustatud rekombinatsiooni *hotspot*'i regiooni, sidudes resekveneerimisel ning *in silico* piirkonna analüüsil saadud andmeid. Viimase etapina analüüsiti *LHB/CGB* geenide potentsiaalset seotust korduvate spontaanabortidega. Selleks viidi läbi nn pilootprojekt, kus võrreldi sekveneerimisel saadud genotüüpide andmeid I trimestri korduvate spontaanabortidega patsientidel ning üldpopulatsioonil põhineval valimil.

Saadud tulemuste põhjal on praeguseks teada keerulise ülesehitusega *LHB/CGB* geeniklastri detailne struktuur ning rekombinatsiooni *hotspot*'i järjestusepõhine iseloomustus, mis viitavad antud piirkonna ebastabiilsusele ümberkorralduste suhtes. Samuti leiti pilootprojekti raames korduvate spontaanabortidega patsientidel iseloomulik *LHB/CGB* geenide haplotüüpide sageduste jaotuvus, mis kinnitab hCG kriitilist funktsiooni raseduse säilitamisel.

Edasiste plaanide hulka kuuluvad: (i) *LHB/CGB* geenide ning intergeensete alade koopiaarvu määramine korduvate spontaanabortidega patsientidel FISH meetodil; (ii) *LHB/CGB* geenide iseloomustamine patsientidele vastaval kontrollpopulatsioonil nii sekveneerimise kui ka FISH analüüsi kaudu; (iii) võimalusel identifitseeritud rekombinatsiooni *hotspot*'i piirkonna detailsem analüüs, kasutades otsesest eksperimentaalset meetodit rekombinantse sperma DNA põhjal.

Summary

Chorionic gonadotropin (hCG) is a glycoprotein hormone with a crucial function for normal outcome of pregnancy. Low levels of hCG mRNA in mother's serum indicates a possible malfunction of hCG protein that can lead to miscarriage. Therefore an analysis of duplicated *LHB/CGB* genes that encode luteinizing hormone and hCG β -subunits would provide useful information on genetic susceptibility to recurrent miscarriages.

In this study a general description of segmental duplications and recombination is presented with the focus on their involvement in generating genetic variability. Also a comprehensive review of research papers on gonadotropins, their function in pregnancy and mutations in *LHB/CGB* genes that have been found so far, is given.

A detailed analysis of the structure of *LHB/CGB* gene cluster was undertaken in order to describe a complicated pattern of segmental duplications and provide thorough background information for further genetic analysis. As a result of the following resequencing study of *LHB/CGB* gene cluster, an intense recombination hotspot has been detected. A hypothesis has been postulated that the hotspot destabilizes the region and promotes different chromosomal rearrangements because of a palindromic secondary structure, *Alu*-sequences and recombination simulating χ -motif.

An extensive pilot-study has been designed in order to compare variations that occur in *LHB/CGB* gene cluster between patients of recurrent miscarriages and a reference population. A haplotype estimation analysis revealed several haplotypes of *LHB/CGB* genes that might have a causative role in recurrent spontaneous abortions. In order to provide well-founded explanation of different variations that might alternate the normal function of hCG, an association study will be conducted with patients of recurrent miscarriages and matching control individuals. Further plans also involve a FISH (fluorescence *in situ* hybridisation) detection to reveal possible chromosomal rearrangements in patients and reference individuals, based on the detailed information of the structure of *CGB/LHB* gene cluster

Tänuavaldused

Sooviksin tänada oma juhendajat, Maris Laant, suurepärase juhendamise, põneva teema ning samuti osutatud usalduse ja toetuse eest. Tänan ka Kristiina Rulli hea koostöö, soovitude ning töö korrigeerimise eest. Samuti tänan Pille Hallastit ning Tõnu Margust, kes on osutanud märkimisväärset abi käesoleva töö valmimisel. Suured tänud veel Viljo Soole ning kõikidele mu suurepärasele kolleegidele, sõpradele ja perele, kes on mind igati aidanud ning toetanud.

Kasutatud kirjandus

- Allers, T., and M. Lichten. 2001.** Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell* 106: 47-57.
- Aoki, K., K. Suzuki, T. Sugano, T. Tasaka, K. Nakahara, O. Kuge, A. Omori, and M. Kasai. 1995.** A novel gene, Translin, encodes a recombination hotspot binding protein associated with chromosomal translocations. *Nat Genet* 10: 167-74.
- Ardlie, K. G., L. Kruglyak, and M. Seielstad. 2002.** Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3: 299-309.
- Badge, R. M., J. Yardley, A. J. Jeffreys, and J. A. Armour. 2000.** Crossover breakpoint mapping identifies a subtelomeric hotspot for male meiotic recombination. *Hum Mol Genet* 9: 1239-44.
- Bailey, J. A., G. Liu, and E. E. Eichler. 2003.** An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 73: 823-34.
- Bettencourt, B. R., and M. E. Feder. 2002.** Rapid concerted evolution via gene conversion at the *Drosophila* hsp70 genes. *J Mol Evol* 54: 569-86.
- Bo, M., and I. Boime. 1992.** Identification of the transcriptionally active genes of the chorionic gonadotropin beta gene cluster in vivo. *J Biol Chem* 267: 3179-84.
- Bo, M., and I. Boime. 1992.** Identification of the transcriptionally active genes of the chorionic gonadotropin beta gene cluster in vivo. *J Biol Chem* 267: 3179-84.
- Bonduelle, M. L., R. Dodd, I. Liebaers, A. Van Steirteghem, R. Williamson, and R. Akhurst. 1988.** Chorionic gonadotrophin-beta mRNA, a trophoblast marker, is expressed in human 8-cell embryos derived from tripronucleate zygotes. *Hum Reprod* 3: 909-14.
- Bricker, L., and R. G. Farquharson. 2002.** Types of pregnancy loss in recurrent miscarriage: implications for research and clinical practice. *Hum Reprod* 17: 1345-50.
- Buyalos, R. P., L. M. Glassman, S. M. Rifka, R. J. Falk, P. O. Macarthy, V. J. Tyson, and M. DiMattina. 1992.** Serum beta-human chorionic gonadotropin, estradiol and progesterone as early predictors of pathologic pregnancy. *J Reprod Med* 37: 261-6.
- Cardon, L. R. and J. I. Bell. 2001.** Association study designs for complex diseases.

Nat Rev Genet 2: 91-99.

- Christian, S. L., J. A. Fantes, S. K. Mewborn, B. Huang, and D. H. Ledbetter. 1999.** Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). *Hum Mol Genet* 8: 1025-37.
- Cotton, J. A., and R. D. Page. 2005.** Rates and patterns of gene duplication and loss in the human genome. *Proc Biol Sci* 272: 277-83.
- Crawford, D. C., T. Bhangale, N. Li, G. Hellenthal, M. J. Rieder, D. A. Nickerson, and M. Stephens. 2004.** Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 36: 700-6.
- Cross, J. C., Z. Werb, and S. J. Fisher. 1994.** Implantation and the placenta: key pieces of the development puzzle. *Science* 266: 1508-18.
- Cullen, M., S. P. Perfetto, W. Klitz, G. Nelson, and M. Carrington. 2002.** High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am J Hum Genet* 71: 759-76.
- Cunningham, L. A., A. G. Cote, C. Cam-Ozdemir, and S. M. Lewis. 2003.** Rapid, stabilizing palindrome rearrangements in somatic cells by the center-break mechanism. *Mol Cell Biol* 23: 8740-50.
- Dacou-Voutetakis, C., D. M. Feltquate, M. Drakopoulou, I. A. Kourides, and N. C. Dracopoli. 1990.** Familial hypothyroidism caused by a nonsense mutation in the thyroid-stimulating hormone beta-subunit gene. *Am J Hum Genet* 46: 988-93.
- de Massy, B. 2003.** Distribution of meiotic recombination sites. *Trends Genet* 19: 514-22.
- de Silva, E., L. A. Kelley, and M. P. Stumpf. 2004.** The extent and importance of intragenic recombination. *Hum Genomics* 1: 410-20.
- Dirnhofer, S., M. Hermann, A. Hittmair, R. Hoermann, K. Kapelari, and P. Berger. 1996.** Expression of the human chorionic gonadotropin-beta gene cluster in human pituitaries and alternate use of exon 1. *J Clin Endocrinol Metab* 81: 4212-7.
- Edelmann, L., E. Spiteri, K. Koren, V. Pulijaal, M. G. Bialer, A. Shanske, R. Goldberg, and B. E. Morrow. 2001.** AT-rich palindromes mediate the constitutional t(11;22) translocation. *Am J Hum Genet* 68: 1-13.

- Eggleston, A. K., and S. C. West. 1997.** Recombination initiation: easy as A, B, C, D... chi? *Curr Biol* 7: R745-9.
- Emanuel, B. S., and T. H. Shaikh. 2001.** Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat Rev Genet* 2: 791-800.
- Fan, Q. R., and W. A. Hendrickson. 2005.** Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* 433: 269-77.
- Feng, W., M. M. Matzuk, K. Mountjoy, E. Bedows, R. W. Ruddon, and I. Boime. 1995.** The asparagine-linked oligosaccharides of the human chorionic gonadotropin beta subunit facilitate correct disulfide bond pairing. *J Biol Chem* 270: 11851-9.
- Fiddes, J. C., and H. M. Goodman. 1980.** The cDNA for the beta-subunit of human chorionic gonadotropin suggests evolution of a gene by readthrough into the 3'-untranslated region. *Nature* 286: 684-7.
- Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack, and A. Di Rienzo. 2001.** Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am J Hum Genet* 69: 831-43.
- Fritz, B., C. Hallermann, J. Olert, B. Fuchs, M. Bruns, M. Aslan, S. Schmidt, W. Coerdts, H. Muntefering, and H. Rehder. 2001.** Cytogenetic analyses of culture failures by comparative genomic hybridisation (CGH)-Re-evaluation of chromosome aberration rates in early spontaneous abortions. *Eur J Hum Genet* 9: 539-47.
- Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, and D. Altshuler. 2002.** The structure of haplotype blocks in the human genome. *Science* 296: 2225-9.
- Galtier, N. 2003.** Gene conversion drives GC content evolution in mammalian histones. *Trends Genet* 19: 65-8.
- Goddijn, M., J. H. Joosten, A. C. Knegt, F. van derVeen, M. T. Franssen, G. J. Bonse, and N. J. Leschot. 2004.** Clinical relevance of diagnosing structural chromosome abnormalities in couples with repeated miscarriage. *Hum Reprod* 19: 1013-7.
- Gordenin, D. A., K. S. Lobachev, N. P. Degtyareva, A. L. Malkova, E. Perkins,**

- and M. A. Resnick. 1993.** Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol Cell Biol* 13: 5315-22.
- Gu, X., Y. Wang, and J. Gu. 2002.** Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet* 31: 205-9.
- Guillon, H., and B. de Massy. 2002.** An initiation site for meiotic crossing-over and gene conversion in the mouse. *Nat Genet* 32: 296-9.
- Hallast, P. 2004.** Geenikonversiooni roll lokaalsete alleelse aheldatuse (LD) mustrite kujundamises inimese genoomis kooriongonadotropiini beta subühiku (CGB) geeniperekonna näitel. Magistritöö.
- Hallast, P., Nagirnaja, L., Margus, T. and Laan, M. 2005.** "Gene conversion shapes regional SNP diversity and linkage disequilibrium patterns in segmental duplications: Human Luteinizing Hormone/ Chorionic Gonadotropin Hormone Beta gene cluster". Käsikiri.
- Hogstrand, K., and J. Bohme. 1994.** A determination of the frequency of gene conversion in unmanipulated mouse sperm. *Proc Natl Acad Sci U S A* 91: 9921-5.
- Homan, G., S. Brown, J. Moran, S. Homan, and J. Kerin. 2000.** Human chorionic gonadotropin as a predictor of outcome in assisted reproductive technology pregnancies. *Fertil Steril* 73: 270-4.
- Houck, C. M., F. P. Rinehart, and C. W. Schmid. 1979.** A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* 132: 289-306.
- Hsu, S. Y. 2004.** Bioinformatics in reproductive biology--functional annotation based on comparative sequence analysis. *J Reprod Immunol* 63: 75-83.
- Hudson, R. R. 2001.** Two-locus sampling distributions and their application. *Genetics* 159: 1805-17.
- Hudson, R. R. 2001.** Two-locus sampling distributions and their application. *Genetics* 159: 1805-17.
- Jablonka-Shariff, A., V. Garcia-Campayo, and I. Boime. 2002.** Evolution of lutropin to chorionic gonadotropin generates a specific routing signal for apical release in vivo. *J Biol Chem* 277: 879-82.
- Jablonka-Shariff, A., V. Garcia-Campayo, and I. Boime. 2002.** Evolution of lutropin to chorionic gonadotropin generates a specific routing signal for apical release in vivo. *J Biol Chem* 277: 879-82.

- Jameson, J. L., C. M. Lindell, and J. F. Habener. 1986.** Evolution of different transcriptional start sites in the human luteinizing hormone and chorionic gonadotropin beta-subunit genes. *DNA* 5: 227-34.
- Jameson, J. L., and A. N. Hollenberg. 1993.** Regulation of chorionic gonadotropin gene expression. *Endocr Rev* 14: 203-21.
- Jeffreys, A. J., R. Neumann, and V. Wilson. 1990.** Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* 60: 473-85.
- Jeffreys, A. J., A. Ritchie, and R. Neumann. 2000.** High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum Mol Genet* 9: 725-33.
- Jeffreys, A. J., L. Kauppi, and R. Neumann. 2001.** Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29: 217-22.
- Jeffreys, A. J., and R. Neumann. 2002.** Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat Genet* 31: 267-71.
- Jeffreys, A. J., J. K. Holloway, L. Kauppi, C. A. May, R. Neumann, M. T. Slingsby, and A. J. Webb. 2004.** Meiotic recombination hot spots and human DNA diversity. *Philos Trans R Soc Lond B Biol Sci* 359: 141-52.
- Juriscova, A., M. Antenos, K. Kapasi, J. Meriano, and R. F. Casper. 1999.** Variability in the expression of trophoctodermal markers beta-human chorionic gonadotrophin, human leukocyte antigen-G and pregnancy specific beta-1 glycoprotein by the human blastocyst. *Hum Reprod* 14: 1852-8.
- Kato, K., M. H. Mostafa, K. Mann, A. E. Schindler, and R. Hoermann. 2002.** Human chorionic gonadotropin exhibits normal biological activity in patients with recurrent pregnancy loss. *Gynecol Endocrinol* 16: 179-86.
- Kauppi, L., M. P. Stumpf, and A. J. Jeffreys. 2005.** Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region. *Genomics*.
- Knofler, M., L. Saleh, S. Bauer, B. Galos, H. Rotheneder, P. Husslein, and H. Helmer. 2004.** Transcriptional regulation of the human chorionic gonadotropin beta gene during villous trophoblast differentiation. *Endocrinology* 145: 1685-94.
- Kong, A., J. Barnard, D. F. Gudbjartsson, G. Thorleifsson, G. Jonsdottir, S.**

- Sigurdardottir, B. Richardsson, J. Jonsdottir, T. Thorgeirsson, M. L. Frigge, N. E. Lamb, S. Sherman, J. R. Gulcher, and K. Stefansson. 2004.** Recombination rate and reproductive success in humans. *Nat Genet* 36: 1203-6.
- Krawinkel, U., G. Zobelein, and A. L. Bothwell. 1986.** Palindromic sequences are associated with sites of DNA breakage during gene conversion. *Nucleic Acids Res* 14: 3871-82.
- Lakich, D., H. H. Kazazian, Jr., S. E. Antonarakis, and J. Gitschier. 1993.** Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet* 5: 236-41.
- Layman, L. C., J. L. Edwards, W. E. Osborne, D. B. Peak, D. G. Gallup, S. P. Tho, R. H. Reindollar, D. J. Roach, P. G. McDonough, and K. D. Lanclos. 1997.** Human chorionic gonadotrophin-beta gene sequences in women with disorders of HCG production. *Mol Hum Reprod* 3: 315-20.
- Layman, L. C. 1999.** Mutations in human gonadotropin genes and their physiologic significance in puberty and reproduction. *Fertil Steril* 71: 201-18.
- Layman, L. C. 2002.** Human gene mutations causing infertility. *J Med Genet* 39: 153-61.
- Lewontin, R. C. 1964.** The Interaction of Selection and Linkage. II. Optimum Models. *Genetics* 50: 757-82.
- Lewontin, R. C., K. Kojima. 1960.** The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 458-472.
- Li, M. D., and J. J. Ford. 1998.** A comprehensive evolutionary analysis based on nucleotide and amino acid sequences of the alpha- and beta-subunits of glycoprotein hormone gene family. *J Endocrinol* 156: 529-42.
- Li, N., and M. Stephens. 2003.** Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165: 2213-33.
- Lobachev, K. S., B. M. Shor, H. T. Tran, W. Taylor, J. D. Keen, M. A. Resnick, and D. A. Gordenin. 1998.** Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* 148: 1507-24.
- Lupski, J. R. 1998.** Charcot-Marie-Tooth disease: lessons in genetic mechanisms. *Mol Med* 4: 3-11.

- Marshall, V. S., L. J. Wilton, and H. D. Moore. 1998.** Parthenogenetic activation of marmoset (*Callithrix jacchus*) oocytes and the development of marmoset parthenogenones in vitro and in vivo. *Biol Reprod* 59: 1491-7.
- Maston, G. A., and M. Ruvolo. 2002.** Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol* 19: 320-35.
- Matthews, C. H., S. Borgato, P. Beck-Peccoz, M. Adams, Y. Tone, G. Gambino, S. Casagrande, G. Tedeschini, A. Benedetti, and V. K. Chatterjee. 1993.** Primary amenorrhoea and infertility due to a mutation in the beta-subunit of follicle-stimulating hormone. *Nat Genet* 5: 83-6.
- May, C. A., A. C. Shone, L. Kalaydjieva, A. Sajantila, and A. J. Jeffreys. 2002.** Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX. *Nat Genet* 31: 272-5.
- McFadden, D. E., L. C. Kwong, I. Y. Yam, and S. Langlois. 1993.** Parental origin of triploidy in human fetuses: evidence for genomic imprinting. *Hum Genet* 92: 465-9.
- Mehan, M. R., N. B. Freimer, and R. A. Ophoff. 2004.** A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. *Hum Genomics* 1: 335-44.
- Miller-Lindholm, A. K., E. Bedows, C. F. Bartels, J. Ramey, V. Maclin, and R. W. Ruddon. 1999.** A naturally occurring genetic variant in the human chorionic gonadotropin-beta gene 5 is assembly inefficient. *Endocrinology* 140: 3496-506.
- Morikawa, M., H. Yamada, E. H. Kato, S. Shimada, T. Yamada, and H. Minakami. 2004.** Embryo loss pattern is predominant in miscarriages with normal chromosome karyotype among women with repeated miscarriage. *Hum Reprod* 19: 2644-7.
- Nakabayashi, K., H. Matsumi, A. Bhalla, J. Bae, S. Mosselman, S. Y. Hsu, and A. J. Hsueh. 2002.** Thyrostimulin, a heterodimer of two new human glycoprotein hormone subunits, activates the thyroid-stimulating hormone receptor. *J Clin Invest* 109: 1445-52.
- Nakay, S., A. Jablonka-Shariff, S. Kaner, P. Chadna-Mohanty, H. E. Grotjan, and D. Ben-Menahem. 2005.** The LHbeta gene of several mammals embeds a carboxyl-terminal peptide-like sequence revealing a critical role for mucin

- oligosaccharides in the evolution of lutropin to chorionic gonadotropin in the animal phyla. *J Biol Chem* 280: 16676-84.
- Nasar, F., C. Jankowski, and D. K. Nag. 2000.** Long palindromic sequences induce double-strand breaks during meiosis in yeast. *Mol Cell Biol* 20: 3449-58.
- Navarro, A., and E. Gazave. 2005.** Inversions with classical style and trendy lines. *Nat Genet* 37: 115-6.
- Papadakis, M. N., and G. P. Patrinos. 1999.** Contribution of gene conversion in the evolution of the human beta-like globin gene family. *Hum Genet* 104: 117-25.
- Paques, F., and J. E. Haber. 1999.** Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63: 349-404.
- Pentao, L., C. A. Wise, A. C. Chinault, P. I. Patel, and J. R. Lupski. 1992.** Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat Genet* 2: 292-300.
- Peoples, R., Y. Franke, Y. K. Wang, L. Perez-Jurado, T. Paperna, M. Cisco, and U. Francke. 2000.** A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome--deletion region at 7q11.23. *Am J Hum Genet* 66: 47-68.
- Phillips, M. S., R. Lawrence, R. Sachidanandam, A. P. Morris, D. J. Balding, M. A. Donaldson, J. F. Studebaker, W. M. Ankener, S. V. Alfisi, F. S. Kuo, A. L. Camisa, V. Pazorov, K. E. Scott, B. J. Carey, J. Faith, G. Katari, H. A. Bhatti, J. M. Cyr, V. Derohannessian, C. Elosua, A. M. Forman, N. M. Grecco, C. R. Hock, J. M. Kuebler, J. A. Lathrop, M. A. Mockler, E. P. Nachtman, S. L. Restine, S. A. Varde, M. J. Hozza, C. A. Gelfand, J. Broxholme, G. R. Abecasis, M. T. Boyce-Jacino, and L. R. Cardon. 2003.** Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 33: 382-7.
- Pierce, J. G., and T. F. Parsons. 1981.** Glycoprotein hormones: structure and function. *Annu Rev Biochem* 50: 465-95.
- Pineda-Krch, M., and R. J. Redfield. 2005.** Persistence and loss of meiotic recombination hotspots. *Genetics* 169: 2319-33.
- Policastro, P. F., S. Daniels-McQueen, G. Carle, and I. Boime. 1986.** A map of the hCG beta-LH beta gene cluster. *J Biol Chem* 261: 5907-16.
- Policastro, P. F., S. Daniels-McQueen, G. Carle, and I. Boime. 1986.** A map of the

- hCG beta-LH beta gene cluster. *J Biol Chem* 261: 5907-16.
- Ptak, S. E., D. A. Hinds, K. Koehler, B. Nickel, N. Patil, D. G. Ballinger, M. Przeworski, K. A. Frazer, and S. Paabo. 2005.** Fine-scale recombination patterns differ between chimpanzees and humans. *Nat Genet* 37: 429-34.
- Pujana, M. A., M. Nadal, M. Guitart, L. Armengol, M. Gratacos, and X. Estivill. 2002.** Human chromosome 15q11-q14 regions of rearrangements contain clusters of LCR15 duplicons. *Eur J Hum Genet* 10: 26-35.
- Quenby, S. M., and R. G. Farquharson. 1993.** Predicting recurring miscarriage: what is important? *Obstet Gynecol* 82: 132-8.
- Robinson, W. P., J. Waslynska, F. Bernasconi, M. Wang, S. Clark, D. Kotzot, and A. Schinzel. 1996.** Delineation of 7q11.2 deletions associated with Williams-Beuren syndrome and mapping of a repetitive sequence to within and to either side of the common deletion. *Genomics* 34: 17-23.
- Robinson, W. P., D. E. McFadden, and M. D. Stephenson. 2001.** The origin of abnormalities in recurrent aneuploidy/polyploidy. *Am J Hum Genet* 69: 1245-54.
- Rull, K. and Laan, M. 2005.** "Expression of β -subunit of human chorionic gonadotropin genes during the normal and failed pregnancy". Submitted
- Shaw, C. J., and J. R. Lupski. 2004.** Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* 13 Spec No 1: R57-64.
- Shi, Q., E. Spriggs, L. L. Field, A. Rademaker, E. Ko, L. Barclay, and R. H. Martin. 2002.** Absence of age effect on meiotic recombination between human X and Y chromosomes. *Am J Hum Genet* 71: 254-61.
- Siiteri, P. K., F. Febres, L. E. Clemens, R. J. Chang, B. Gondos, and D. Stites. 1977.** Progesterone and maintenance of pregnancy: is progesterone nature's immunosuppressant? *Ann N Y Acad Sci* 286: 384-97.
- Stankiewicz, P., and J. R. Lupski. 2002.** Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18: 74-82.
- Suganuma, N., M. M. Matzuk, and I. Boime. 1989.** Elimination of disulfide bonds affects assembly and secretion of the human chorionic gonadotropin beta subunit. *J Biol Chem* 264: 19302-7.
- Tanabe, M., I. Ieiri, N. Nagata, K. Inoue, S. Ito, Y. Kanamori, M. Takahashi, Y. Kurata, J. Kigawa, S. Higuchi, N. Terakawa, and K. Otsubo. 2001.**

- Expression of P-glycoprotein in human placenta: relation to genetic polymorphism of the multidrug resistance (MDR)-1 gene. *J Pharmacol Exp Ther* 297: 1137-43.
- Tease, C., G. M. Hartshorne, and M. A. Hulten. 2002.** Patterns of meiotic recombination in human fetal oocytes. *Am J Hum Genet* 70: 1469-79.
- Themmen, A. P. N., and I. T. Huhtaniemi. 2000.** Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev* 21: 551-83.
- Tishkoff, S. A., and B. C. Verrelli. 2003.** Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. *Curr Opin Genet Dev* 13: 569-75.
- Tusie-Luna, M. T., and P. C. White. 1995.** Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. *Proc Natl Acad Sci U S A* 92: 10796-800.
- Wahls, W. P., L. J. Wallace, and P. D. Moore. 1990.** The Z-DNA motif d(TG)30 promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol Cell Biol* 10: 785-93.
- Wall, J. D. 2001.** Insights from linked single nucleotide polymorphisms: what we can learn from linkage disequilibrium. *Curr Opin Genet Dev* 11: 647-51.
- Wang, W. Y., B. J. Barratt, D. G. Clayton, and J. A. Todd. 2005.** Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 6: 109-18.
- Weiss, J., L. Axelrod, R. W. Whitcomb, P. E. Harris, W. F. Crowley, and J. L. Jameson. 1992.** Hypogonadism caused by a single amino acid substitution in the beta subunit of luteinizing hormone. *N Engl J Med* 326: 179-83.
- Yauk, C. L., P. R. Bois, and A. J. Jeffreys. 2003.** High-resolution sperm typing of meiotic recombination in the mouse MHC Ebeta gene. *Embo J* 22: 1389-97.
- Yu, J., L. Lazzeroni, J. Qin, M. M. Huang, W. Navidi, H. Erlich, and N. Arnheim. 1996.** Individual variation in recombination among human males. *Am J Hum Genet* 59: 1186-92.
- Zangenberg, G., M. M. Huang, N. Arnheim, and H. Erlich. 1995.** New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nat Genet* 10: 407-14.

Lisad

Lisa 1. *LHB/CGB* geenide PCR ja sekveneerimise praimerid ning PCR produktide pikkused (P. Hallast, 2004)

Lisa 2. *Hotspot*‘i piirkonna PCR ja sekveneermise praimerid ning PCR produktide pikkused.

Lisa 3. *CGB5* ja *CGB7* vahel esineva rekombinatsiooni *hotspot*'i piirkonna polümorfismid .

Geen/ markeri nr	Markeri pos.	Markeri kood	Mazoorne Alleel	Minoorne Alleel	Minoorse alleeli sagedus
1	<i>CGB5-8</i>	98	A	T	0.182
2	<i>CGB5-8</i>	110	A	T	0.182
3	<i>CGB5-8</i>	119	G	T	0.045
4	<i>CGB5-8</i>	439	G	T	0.2
5	<i>CGB5-8</i>	728	G	A	0.182
6	<i>CGB5-8</i>	864	G	A	0.5
7	<i>CGB5-8</i>	916	G	A	0.091
8	<i>CGB5-8</i>	1079	TTTG	-	0.318
9	<i>CGB5-8</i>	1345	C	A	0.25
10	<i>CGB5-8</i>	1430	T	C	0.15
11	<i>CGB5-8</i>	1447	G	C	0.227
12	<i>CGB5-8</i>	1468	T	G	0.25
13	<i>CGB5-8</i>	1469	G	T	0.25
14	<i>CGB5-8</i>	1472	A	G	0.25
15	<i>CGB5-8</i>	1473	G	A	0.045
16	<i>CGB5-8</i>	1476	-	G	0.05
17	<i>CGB5-8</i>	1622	G	-	0.091
18	<i>CGB5-8</i>	c8E1369	G	A	0.318
19	<i>CGB5-8</i>	c8E1151	T	C	0.045
20	<i>CGB5-8</i>	c8E1146	TG	-	0.045
21	Intron 2	c8E673	G	A	0.091
22	Intron 1	c8E199	G	A	0.045
23	Intron 1	c8E169	C	G	0.227
24	Intron 1	c8E124	G	A	0.045
25	5'UTR	c8E-72	A	T	0.136
26	5'UTR	c8E-172	C	T	0.045
27	5'UTR	c8E-265	G	A	0.273
28	5'UTR	c8E-268	C	G	0.091
29	<i>CGB8-7</i>	c8E-558	C	A	0.2
30	<i>CGB8-7</i>	c8E-659	A	G	0.45
31	<i>CGB8-7</i>	c8E-786	C	G	0.045
32	<i>CGB8-7</i>	c8E-787	A	G	0.045
33	<i>CGB8-7</i>	4100	T	C	0.045
34	<i>CGB8-7</i>	5334	A	G	0.227
35	<i>CGB8-7</i>	5384	C	T	0.273
36	<i>CGB8-7</i>	5664	C	-	0.273
37	<i>CGB8-7</i>	5862	-	A	0.45
38	<i>CGB8-7</i>	5884	C	T	0.318
39	<i>CGB8-7</i>	5890	T	C	0.4
40	<i>CGB8-7</i>	5910	C	G	0.318
41	<i>CGB8-7</i>	5937	T	C	0.125
42	<i>CGB8-7</i>	5952	G	A	0.083
43	<i>CGB8-7</i>	5953	A	G	0.083
44	<i>CGB8-7</i>	5981	A	G	0.167
45	<i>CGB8-7</i>	6051	C	T	0.3

46	<i>CGB8-7</i>	6211	C	T	0.5
47	<i>CGB8-7</i>	6494	A	G	0.045
48	<i>CGB8-7</i>	6951	A	C	0.5
49	<i>CGB8-7</i>	6977	T	C	0.364
50	<i>CGB8-7</i>	6979	A	G	0.045
51	<i>CGB8-7</i>	7043	G	T	0.45
				-	0.045
52	<i>CGB8-7</i>	7146	C	T	0.273
53	<i>CGB8-7</i>	7467	C	T	0.045
54	<i>CGB8-7</i>	7542	T	C	0.136
55	<i>CGB8-7</i>	7755	A	G	0.045
56	<i>CGB8-7</i>	7756	A	C	0.136
57	<i>CGB8-7</i>	7758	T	C	0.136
58	<i>CGB8-7</i>	7760	T	C	0.136
59	<i>CGB8-7</i>	7801	G	A	0.136
60	<i>CGB8-7</i>	7803	A	G	0.136
61	<i>CGB8-7</i>	7852	A	T	0.136
62	<i>CGB8-7</i>	7937	C	T	0.5
63	<i>CGB8-7</i>	8137	C	G	0.136
64	<i>CGB8-7</i>	8205	C	G	0.364

Lisa 4. *LHB/CGB* geenides esinevad polümorfismid võrdlevalt patsientidel ning referentspopulatsioonis. *- aminohappe muutust põhjustav polümorfism.

Geen/markeri nr	Markeri pos.	Markeri kood	Mazoorne Alleel	Minoorne Alleel	Minoorse alleeli sagedus	
<i>LHB</i>					Referents	Patsiendid
1	3' downstream	LE1281	C	T	0.298	0.326
2	3' downstream	LE1109	G	T	0.415	0.35
3	3' UTR	LE1099	G	A	0.117	0.087
4	Ekson 3	LE873	G	A	0.289	-
5	Intron 2	LE553	-	T	0.298	0.312
6	Intron 2	LE546	G	A	0.298	0.312
7	Ekson 2	LE484	G	T	0.415	0.364
8	Ekson 2	LE466	C	G	0.298	0.312
9	Ekson 2*	LE434	A	G	0.117	0.094
10	Ekson 2*	LE397	C	T	0.021	-
11	Intron 1	LE342	G	A	0.319	0.31
12	Intron 1	LE264	C	G	0.021	-
13	Intron 1	LE263	T	C	0.117	0.043
14	Intron 1	LE246	G	T	0.043	0.021
15	Intron 1	LE219	A	G	0.011	-
16	Intron 1	LE154	C	A	0.011	-
17	Intron 1	LE125	T	C	0.117	0.083
18	Intron 1	LE122	C	T	0.117	0.083
19	Intron 1	LE24	C	T	0.426	0.391
20	5' upstream	LE-34	T	A	0.309	0.348
<i>CGB</i>						
21	3' downstream	cE1110	G	T	0.319	0.217
22	Ekson 3	cE1038	A	G	0.330	0.354
23	Ekson 3	cE1026	C	G	0.330	0.333
24	Intron 2	cE751	A	G	0.309	0.333
25	Intron 2	cE554	T	C	0.394	0.386
26	Intron 1	cE321	G	A	0.011	-
27	Intron 1	cE302	G	A	0.011	-
28	Intron 1	cE168	C	G	0.298	0.292
29	Intron 1	cE163	C	T	0.298	0.292
30	Intron 1	cE101	A	C	0.298	0.292
31	Intron 1	cE88	G	T	0.298	0.292
32	Intron 1	cE86	G	A	0.032	0.042
33	5' UTR	cE-76	G	T	0.074	0.104
34	5' UTR	cE-77	G	A	0.245	0.261
35	5' UTR	cE-111	T	C	0.330	0.13
36	5' UTR	cE-144	T	G	-	0.292
37	5' UTR	cE-180	C	T	0.298	0.292
38	5' UTR	cE-185	G	C	0.293	0.271
39	5' UTR	cE-324	T	C	-	0.292
40	5' UTR	cE-328	A	G	0.287	0.292
<i>CGB2</i>						
41	Intron 1	c2E140	G	C	0.021	0.042

42	Intron 1	c2E188	T	C	0.372	0.333
43	Intron 1	c2E391	C	T	0.021	0.043
44	Intron 1	c2E421	G	A	0.011	-
45	Ekson 2	c2E531	G	A	0.128	0.125
46	3' UTR	c2E1087	C	A	0.340	0.396
47	3' downstream	c2E1211	G	A	0.372	0.375
48	3' downstream	c2E1273	T	C	0.011	-
CGB1						
49	3' UTR	c1E1140	G	A	0.033	0.021
50	Intron 2	c1E840	G	A	-	0.021
51	Intron 2	c1E643	T	C	0.138	0.13
52	Ekson 2*	c1E458	T	C	0.202	0.271
53	Intron 1	c1E452	G	A	0.189	0.292
54	Intron 1	c1E432	G	A	-	0.042
55	Intron 1	c1E150	G	T	-	0.022
56	5' UTR	c1E-76	C	A	0.367	0.271
57	5' UTR	c1E-116	-	TCCCC	-	0.042
CGB5						
58	5' UTR	c5E-228	A	G	0.138	0.104
59	5' UTR	c5E-209	C	T	0.011	0.022
60	5' UTR	c5E-21	G	C	0.202	0.333
61	Intron 1	c5E179	T	G	0.330	0.458
62	Intron 1	c5E186	C	T	0.021	0.021
63	Intron 1	c5E215	G	A	0.202	0.375
64	Intron 1	c5E301	C	T	0.330	0.438
65	Ekson 2	c5E424	G	A	0.202	0.333
66	Intron 2	c5E547	G	C	-	0.021
67	Intron 2	c5E553	C	G	0.011	-
68	Intron 2	c5E673	C	T	0.128	0.188
69	Intron 2	c5E746	A	T	-	0.021
70	Intron 2	c5E750	C	T	0.021	0.021
71	Ekson 3	c5E893	C	T	0.021	0.042
72	3' downstream	c5E1214	G	A	0.032	-
CGB8						
73	3' downstream	c8E1369	G	A	0.318	-
74	3' downstream	c8E1298	C	G	-	0.083
75	3' downstream	c8E1151	T	C	0.045	-
76	3' downstream	c8E1146	TG	-	0.045	-
77	Intron 2	c8E673	G	A	0.091	-
78	Intron 2	c8E553	C	G	-	0.022
79	Intron 1	c8E199	G	A	0.045	-
80	Intron 1	c8E169	C	G	0.227	0.292
81	Intron 1	c8E124	G	A	0.045	-
82	5' UTR	c8E-72	A	T	0.136	0.083
83	5' UTR	c8E-172	C	T	0.045	-
84	5' UTR	c8E-265	G	A	0.273	0.312
85	5' UTR	c8E-268	C	G	0.091	0.042
86	5' upstream	c8E-558	C	A	0.2	0.333
CGB7						
87	3' downstream	c7E1435	C	T	-	0.021

88	3' downstream	c7E1278	A	C	0.436	0.354
89	3' downstream	c7E1277	G	A	0.415	0.354
90	3' downstream	c7E1261	C	G	0.426	0.333
91	3' downstream	c7E1241	C	G	0.426	0.333
92	3' downstream	c7E1224	C	A	0.468	0.333
93	3' downstream	c7E1151	T	C	0.404	0.354
94	3' downstream	c7I1136	-	GT	-	0.042
95	Ekson 3	c7E1061	T	C	0.467	-
96	Ekson 3*	c7E867	G	A	-	0.021
97	Intron 2	c7E750	A	G	-	0.188
98	Intron 2	c7E724	G	A	0.032	-
99	Intron 2	c7E683	T	C	0.117	0.146
100	Intron 2	c7E553	G	C	0.287	0.125
				T	0.030	0.024
101	Ekson 2*	c7E473	-	C	-	0.042
102	Ekson 2*	c7E423	A	G	0.415	0.333
103	Ekson 2*	c7E422	T	G	0.415	0.333
104	Ekson 2*	c7E417	C	T	0.415	0.333
105	Intron 1	c7E339	C	T	0.011	-
106	Intron 1	c7E168	A	G	0.436	0.354
107	Intron 1	c7E135	T	C	0.404	0.333
108	Intron 1	c7E124	G	A	0.415	0.333
109	Intron 1	c7E60	G	T	0.426	0.354
110	5' UTR	c7E-35	T	C	0.298	0.196
111	5' UTR	c7E-46	T	G	0.415	0.348
112	5' UTR	c7E-58	T	G	0.415	0.326
113	5' UTR	c7E-97	C	G	0.426	0.333
114	5' UTR	c7E-147	G	C	0.191	0.324
115	5' UTR	c7E-201	G	T	0.138	-
116	5' UTR	c7E-252	C	T	0.422	-
117	5' UTR	c7E-357	G	A	0.476	-
118	5' upstream	c7E-379	C	T	0.422	-
119	5' upstream	c7E-468	T	C	0.117	0.146
120	5' upstream	c7E-697	A	G	0.237	0.348