

A duplikálódott *CYP21A2* gént tartalmazó humán LASB RCCX struktúrákat különleges genetikai átrendeződés alakítja ki

XXXI. OTDK dolgozat

készítette:

Szabó Julianna

Eötvös Loránd Tudományegyetem, II. évf. biológia Msc hallgató

témavezető:

Dr. Doleschall Márton, tudományos munkatárs
Semmelweis Egyetem, III. Sz. Belgyógyászati Klinika

Budapest, 2012

Tartalomjegyzék

Tartalomjegyzék.....	2
Rövidítés jegyzék.....	3
Absztrakt	4
1. Bevezető.....	5
1.1 A humán kópiaszám variációk általános jellemzése	5
1.2 A humán RCCX CNV szerkezete és génjei	10
1.3 A humán RCCX CNV struktúráját formáló genetikai átrendeződések jellemzése ..	13
2. Célkitűzés.....	15
3. Anyag és módszer	16
3.1 Vizsgálati minták	16
3.2 Az RCCX struktúra molekuláris haplotipizálása és meghatározása.....	16
3.3 A CYP21A2 gén amplifikálása és szekvenálása	18
3.4 A PCR termékek elválasztása DNS gélelektroforézis segítségével	20
3.5 Bioinformatikai elemzések	20
4. Eredmények és értékelésük.....	21
4.1 A haploid RCCX CNV struktúra és a <i>CYP21A2</i> haplotípus meghatározása.....	21
4.2 A két CYP21A2-t tartalmazó haploid LASB RCCX struktúrák	22
4.3 A két <i>CYP21A2</i> -t tartalmazó RCCX struktúrák <i>CYP21A2</i> haplotípusainak elhelyezkedése a genealógiai hálózaton	24
5. Összefoglalás	26
Irodalom	30
Köszönetnyilvánítás	33

Rövidítés jegyzék

ASLR-PCR	allél-specifikus long-range PCR
<i>C4</i>	komplement komponens 4
<i>C4A</i>	<i>C4</i> gén A altípus
<i>C4B</i>	<i>C4</i> gén B altípus
CAH	kongenitális adrenális hiperplázia (congenital adrenal hyperplasia, CAH)
CNM	kópiaszám mutáció
CNP	kópiaszám polimorfizmus
CNV	kópiaszám variációk (copy number variation)
<i>CYP21</i>	szteroid 21-hidroxiláz
<i>CYP21A1P</i>	szteroid 21-hidroxiláz pszeudogén
<i>CYP21A2</i>	szteroid 21-hidroxiláz aktív gén
HERV-K	endogén retrovírus inszerció (human endogenous retrovirus)
LCR	low copy repeat (nincs meggyökeresedett magyar fordítása)
MAF	ritkábbik allél gyakorisága (minor allele frequency)
MHC	fő hisztokompatibilitási komplex (major histocompatibility complex)
NAHR	nem-allélikus homológ rekombinációra (non-allelic homologous recombination)
PCR	polimeráz-láncreakció (polymerase chain reaction)
qPCR	kvantitatív PCR (quantitative PCR)
RCCX CNV	<u>RP-C4-CYP21-TNX</u> géneket tartalmazó kópiaszám variáció
SD	szegmentális duplikáció (segmental duplication)
SNP	egy nukleotidot érintő polimorfizmus (single nucleotide polymorphism)
<i>STK19</i>	szerin/treonin kináz 19 (régi neve RP1)
<i>STK19P</i>	szerin/treonin kináz 19 pszeudogén
SV	strukturális variációk (structural variation)
<i>TNX</i>	tenascin-X
<i>TNXA</i>	tenascin-X pszeudogén
<i>TNXB</i>	tenascin-X aktív gén
VNTR	változó számú tandem ismétlődés (variable number of tandem repeat)

Absztrakt

A humán RCCX egy multiallélikus, tandem kópiaszám variáció (copy number variation, CNV), amelynek az egyik génje a szteroid 21-hidroxiláz (*CYP21A2*). Az RCCX, mint a CNV-k általában, hajlamos az átkereszteződéstől eltérő genetikai átrendeződésekre, mint például az egyenlőtlen átkereszteződés és a génkonverzió.

Az egész RCCX CNV-t lefedő allél-specifikus long-range PCR-eket, az ebből induló *CYP21A2* allél-specifikus nested PCR-t, a *CYP21A2* teljes szekvenálást és qPCR alapú kópiaszám meghatározást használtunk az RCCX struktúrák és a *CYP21A2* haplotípusok kísérletes meghatározásához. A 332 genetikailag független személyből válogatott 6 duplikálódott *CYP21A2* gént tartalmazó LASB RCCX struktúrát hordozó személynek a vizsgálatát végeztük el.

A 6 kromoszóma LASB RCCX struktúrával és két *CYP21A2* génnel 3 típusba tartozott a *CYP21A2* haplotípusaik alapján. A 3 típus azonos RCCX struktúrát és két, csaknem azonos, de típusonként egymástól eltérő *CYP21A2* haplotípust hordozott. A genealógiai hálózat felhasználásával bizonyítottuk, hogy a 3 típus egymástól függetlenül alakult ki, de feltehetőleg azonos genetikai átrendeződési mechanizmussal. Az RCCX struktúrájából következően az átrendeződés mechanizmusa biztosan nem az általánosan elfogadott egyszeres egyenlőtlen átkereszteződés vagy kis kiterjedésű génkonverzió, hanem vagy kettős egyenlőtlen átkereszteződése, vagy nagy kiterjedésű génkonverzió.

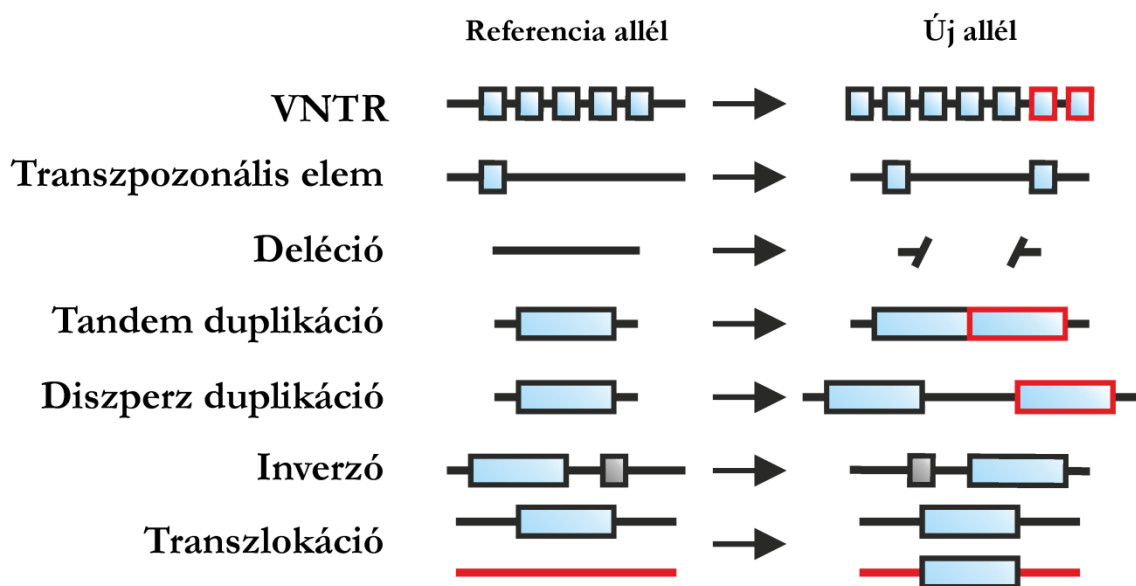
A fenti vizsgálatok nem csak az RCCX CNV átrendeződési mechanizmusaiba engednek mélyebb bepillantást, hanem általában véve a humán multiallélikus, tandem CNV-k átrendeződéseiért felelős mechanizmusokba is. Az CNV-k speciális átrendeződési mechanizmusainak megismerése a CNV-kel kapcsolt humán betegségek jobb megértéséhez és, végső soron, új terápiás lehetőségekhez vezethetnek.

1. Bevezető

1.1 A humán kópiaszám variációk általános jellemzése

A humán genom diverzitását különböző típusú genetikai polimorfizmusok alakítják ki, a legközismertebbek az egy nukleotidot érintő polimorfizmusok (single nucleotide polymorphisms, SNP) mellett ilyenek például a strukturális variációk (structural variation, SV) is (Hurles et al. 2008). Az SV-k fogalma viszonylag fiatal, még kiforratlan, a tudományos szakirodalomban némiképp ellentmondásos, átfedésben van több olyan rokon vagy szinonim fogalommal, mint a kópiaszám variációk (copy number variation, CNVs), kópiaszám polimorfizmusok (copy number polymorphisms, CNPs), kópiaszám mutációk (copy number mutations), szegmentális duplikációk (segmental duplications, SDs), vagy a meggyökeresedett magyar fordítással nem rendelkező low copy repeats (LCRs) (Freeman et al. 2006). Összességében az SV-k tág fogalma azokat a genetikai variációkat jelölik, amelyek megváltoztatják a kromoszóma struktúráját, de hagyományos fénymikroszkópos technikával nem észlelhetők. Tehát az SV-k mérettartománya a pár bázispártól a megabázisig terjed, így a fogalomba a tankönyvekben definiált „kromoszómaszerkezet változásai” (Weaver et al. 2000), nagyobb méretük okán, elvileg nem tartoznak bele, bár a mérethatár nem jól definiált, meglehetősen elmosódott. Ezek szerint a SV-k lehetnek változó számú tandem ismétlődések (variable number of tandem repeats, VNTRs), transzpozábilis elemek megléte vagy hiánya, deléciók, diszperz vagy tandem duplikációk, inverziók, transzlokációk (1. ábra). Az SV-k egyaránt magába foglalják a DNS kópiaszám változással járó (például CNV) és nem járó (pl.: inverzió) átrendeződéseket.

Más genetikai variációkkal, például az SNP-kkel és mutációkkal, szemben az SV-k vizsgálatát megnehezíti, hogy az első generációs DNS szekvenálással nem tanulmányozhatók könnyen (kivételt képeznek ez alól a rövidebb VNTR-ek és transzpozális elemek), így vizsgálatuk egy jó darabig háttérbe szorult (Hurles et al. 2008). Nem meglepő tehát, hogy egészen a közelmúltig az SNP-ket tekintették a genetikai variabilitás uralkodó formájának, és az egészséges fenotípus változatok fő alakítóinak. A DNS strukturális variációit sokáig csak specifikus kromoszóma átrendeződésekkel és genetikai rendellenességekkel hozták kapcsolatban, és az emlős genomokban való elterjedtségét csak jó pár évvel később mérték fel teljes mértékben (Freeman et al. 2006). Azonban a microarray technológia és a következő generációs szekvenálás térhódítása lehetővé tette az SV-k megfelelő és széles körű vizsgálatát, egyben feltárta a cseppet sem elhanyagolható szerepét a humán genomban.

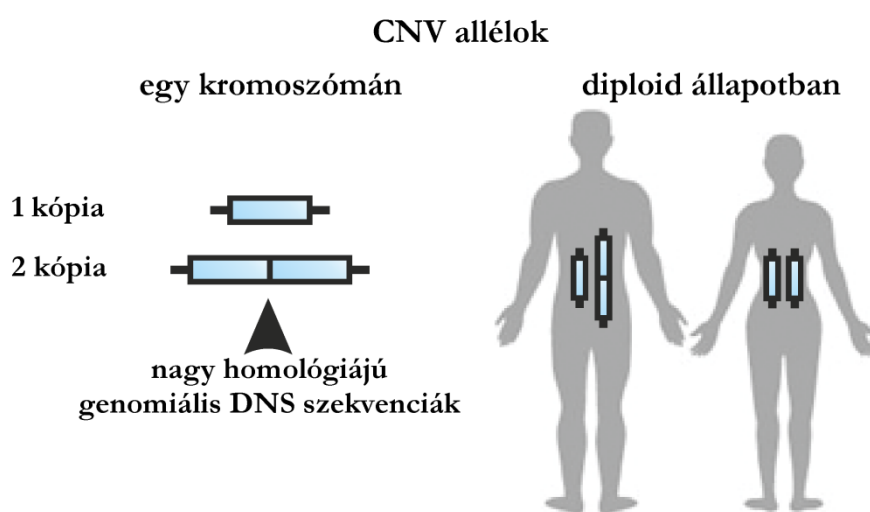


1. ábra: A strukturális változatok (SV-k) különböző típusai. Az SV-k lehetnek változó számú tandem ismétlődések (VNTRs), transzpozázilis elemek megléte vagy hiánya, deléción, diszperz vagy tandem duplikációk, inverziók, transzlokációk.

Legfrissebb adatokból kiindulva valószínűsíthető hogy egy humán diploid genom átlagosan több mint 1000 SV-t tartalmaz (Hurles et al. 2008) összevetve a humán referencia szekvenciával. A humán genom komplexitásában résztvevő összes tényező hatásának megismeréséhez és értelmezéséhez a leghasznosabb minél több teljes humán genom szekvenciájának ismerete lenne, az ez irányú, nemzetközi 1000 Genomes projekt jelenleg a pilot stádiumnál tart (Durbin et al. 2010). Míg ez lehetővé válik, a különböző forrásokból származó, a genetikai variációkra vonatkozó információk összegyűjtése és elemzése kívánatos, hogy minél több ismeretet szerezzünk a humán szervezet genetikai diverzitásáról, a diverzitást kialakító mechanizmusokról, a humán evolúcióról és betegségekről (Stranger et al. 2007).

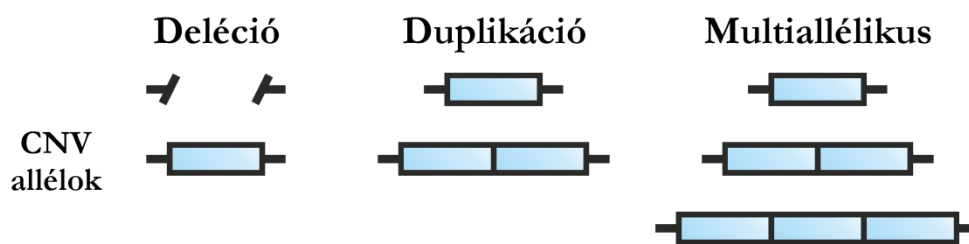
A CNV-k a kromoszómák kilobázistól megabázis mérettartományig terjedő strukturális variációi, amelyek területén egy nagy homológiával rendelkező genomiális régió különböző ismétlődésszámai fordulnak elő egy populációban fellelhető különböző kromoszómákon. A fenti meghatározás szerint a CNV-k az SV-k egy alhalmaza, amelyek definíciószerűen nem tartalmazzák a kisebb méretű SV-ket, mint a VNTR és a transzpozázilis elemek által okozott inszerció, illetve a kópiaszám változásával nem járó SV-ket, mint például az inverzió (Freeman et al. 2006). Ezen kívül a CNV-k fogalmát, épp úgy mint általában véve a variációkét, nem lehet értelmezni egyetlen egyedben, csak populációban. Tehát az adott

populációban különböző kópiaszámban fog előfordulni a kromoszómákon egy nagy homológiájú genomiális régió, például létezhet egy vagy két kópiát hordozó kromoszóma, amelyek kombinációi diploid állapotban kettő, három vagy négy kópiát eredményeznek a populáción belül (2. ábra). Ugyanez érvényes egy funkcionális gén kópiaszámára is, ha az ismétlődő genomiális régió hordozza a teljes gént. A különböző kópiaszámú kromoszómák allél gyakorisága alapján, megkülönböztethetünk kópiaszám polimorfizmusokat (CNP) és mutációkat (CNM), a báziscserékhez hasonlóan a 1% feletti gyakoriságú ritkábbik alléllal (minor allele frequency, MAF) rendelkeznek a CNP-k, és 1% alattival a CNM-ek.



2. ábra: A CNV allélok elhelyezkedése egy kromoszómán és diploid állapotban.

A CNV-k lehetnek egyszerű felépítésűek, ilyenek például a tandem duplikációk, és bonyolultabb felépítésűek, amelyek homológ szekvenciák nyerésével vagy vesztésével alakulnak ki és a genomban többfelé, elszórtan helyezkednek el (Redon et al. 2006). Az egyszerűbb felépítésűek három osztályba sorolhatóak, úgymint deléciók (0 vagy 1 kópia egy kromoszómán), duplikációk (1 vagy 2 kópia egy kromoszómán) és multiallélikus CNV-k (több mint 2 lehetséges kópiaszám) (Conrad et al. 2010) (3. ábra). A három osztály megoszlása az európai referencia populációban jelentős eltéréseket mutat, míg a deléciók körülbelül 70-80%-át alkotják a CNV-knek, addig a duplikációk csak 15%-át, a multiallélikus CNV-k pedig kevesebb mint 10%-át teszik ki az összesnek (Conrad et al. 2010).



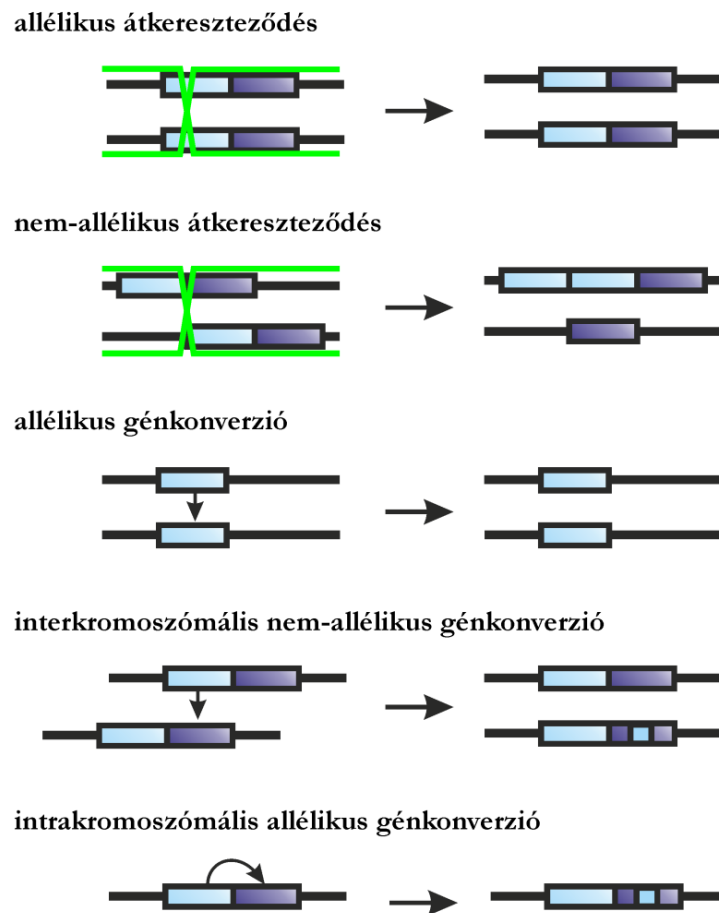
3. ábra: A egyszerű, szabályos CNV-k osztályai: deléción, duplikáción és multiallélikus CNV-k

A CNV-k 1-2%-át alkotják az emberi genomnak, de a genetikai diverzitáshoz való hozzájárulásuknál gyakran nagyobb funkcionális jelentőséget tulajdonítanak nekik (Zhang, 2009). A CNV-k felelősek az egészséges személyekben előforduló fenotípus változatok egy részéért (Freeman et al. 2006), habár nincsenek még megfelelő ismereteink arra nézve, hogy a CNV-k és az SNP-k mekkora arányban járulnak hozzá a fenotípus változatokhoz (Stranger et al. 2007). A különböző kópiaszámok, illetve az általuk hordozott különböző gén kópiaszámok befolyásolják a sejtek génexpressziós fenotípusát. Számos bizonyíték áll rendelkezésünkre arra nézve, hogy a megnövekedett kópiaszám pozitívan és negatívan (Például egy transzkripció represszor megváltozott kópiaszáma csökkentheti az általa szabályozott gének expressziós szintjét.) is korrelálhat a génexpressziós szintekkel (Stranger et al. 2007). Az egészséges fenotípusokon túl a betegségek és a betegségre hajlamosító tényezők széles spektrumáért szintén felelősek a CNV-k (Zhang et al. 2009).

Az egyszerűbb felépítésű multiallélikus CNV-k általában tandem elrendeződésűek, kiterjedésük gyakran nagy, és sok gént tartalmaznak. Jelentős szerepet játszanak a gének vagy genomok evolúciójában (Hurles et al. 2008). A multiallélikus CNV-k fixálódási fázisban lévő, dinamikusan változó fiatal duplikációknak is tekinthetőek, amelyek sokszor még pozitív evolúciós folyamatok hatása alatt állnak. Az evolúciós folyamatokban betöltött szerepük alapja az, hogy a multiallélikus CNV-k hajlamosak a nem-allélikus homológ rekombinációra (non-allelic homologous recombination, NAHR) (Hastings et al. 2009), mint az egyenlőtlen átkereszteződés vagy a nem-allélikus génkonverzió. (Cooper et al. 2011).

A NAHR a rekombináció azon formája, amely nem allélikus, nagy homológiával rendelkező, általában duplikálódott DNS szekvenciák között alakul ki (Beckmann et al. 2007). Az allélikus (interallélikus) homológ rekombinációnál nincsen változás a kromoszómastruktúrában, mivel a rekombináció a testvérekromatida vagy a homológ kromoszóma azonos szakaszáról történik. A NAHR esetében két kromoszóma más vagy egy kromoszóma nem azonos helyzetű (nem-allélikus) homológ szakaszával történik meg a rekombináció, így a kromoszóma strukturálisan megváltozhat. A NAHR két alapvető fajtája

közül az egyenlőtlen átkereszteződés reciprok, interkromoszómális és meiotikus, a nem-allélikus génkonverzó nem-reciprok, és lehet meiotikus és mitotikus, illetve inter- és intrakromoszómális egyaránt (4. ábra). Amíg az egyenlőtlen átkereszteződés nagyobb kiterjedésű strukturális átrendeződéseket és kópiaszám változást okozhat (Stankiewicz et al. 2002), addig az nem-allélikus génkonverziót viszonylag rövid szekvenciák áthelyeződésnek tekintik, amely a CNV-k kópiaszám változásaiban nem vesz részt (Chen et al. 2007). Az egyenlőtlen átkereszteződéssel keletkezett, duplikációt vagy deléciót hordozó kromoszómák az átkereszteződést követően mindkét utódsejtben megtalálhatók lesznek, eltérő kópiaszámot eredményezve. A NAHR lehetővé teszi az exon shuffling-ot, vagyis az exonok átrendeződését a különböző gének között. Ennek egyik nagy előnye hogy, így a fehérjék új doménekre tehetnek szert, és ez által új vagy módosított aktivitásra is (Hastings et al. 2009).



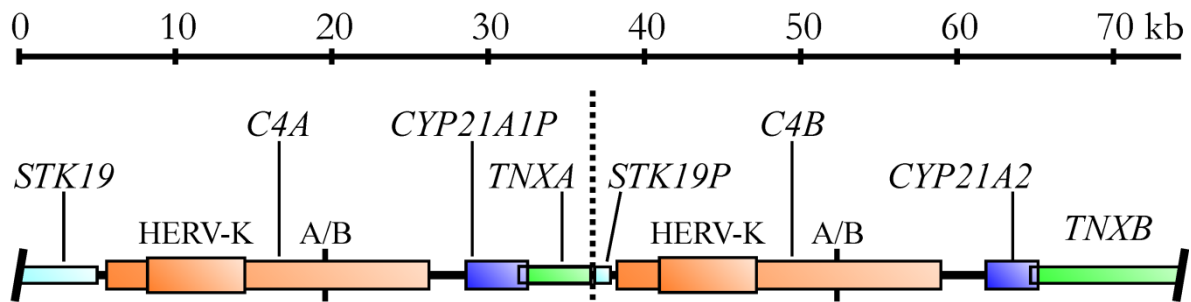
4. ábra: Genetikai átrendeződések a tandem duplikálódott régiókban belül. Az allélikus (interallélikus) homológ rekombinációnál nincs változás a kromoszómastruktúrában, mivel a rekombináció a testvérkromatida vagy a homológ kromoszóma azonos szakaszáról történik. A NAHR esetében két kromoszóma más vagy egy kromoszóma nem azonos helyzetű (nem-allélikus) homológ szakaszával történik meg a rekombináció, így a kromoszóma strukturálisan megváltozhat. A NAHR két alapvető fajtája közül az egyenlőtlen átkereszteződés reciprok, interkromoszómális, a nem-allélikus génkonverzó nem-reciprok, illetve inter- és intrakromoszómális is lehet.

A multiallélikus CNV-k terén megszerzett tudásunk ellenére csak a közelmúltban kezdték értékelni az átrendeződési mechanizmusok relatív hozzájárulását a CNV-k és azok gén tartalmának kialakításához és fenntartásához (Mills et al. 2011). A multiallélikus tandem CNV-ket, és azok duplikálódott géntartalmát nagyon nehéz genotipizálni a genom egészét vizsgáló (genome-wide) technikák segítségével, illetve a kópiaszámok csak gyengén jellemezhetők SNP-kkel (Alkan et al. 2011). Így, sajnos, a legdinamikusabban változó géneket tartalmazó multiallélikus CNV-k fölött gyakran átsiklik a genom egészét vizsgáló tanulmányok figyelme (Sudmant et al. 2010).

1.2 A humán RCCX CNV szerkezete és génjei

Az RCCX egy közepes méretű, multiallélikus, tandem CNV, amely a fő hisztokompatibilitási komplex (major histocompatibility complex, MHC) III régiójában helyezkedik el (Horton et al. 2004). A monomoduláris, bimoduláris és trimoduláris RCCX CNV allélok (1, 2 és 3 kópia egy kromoszómán) a közönségesek, előfordulásuknak gyakorisága európai populációkon belül rendre körülbelül 15%, 75% és 10%. Négy gén foglal helyet egymás mellett az RCCX CNV egy modulján (kópiáján) belül, a szerin/treonin kináz 19 (*STK19*), a komplement komponens 4 (*C4*), a szteroid 21-hidroxiláz (*CYP21*) és a tenascin-X (*TNX*). Az összes modult figyelembe véve az *STK19*, *CYP21* és *TNX* génnek egy aktív (*STK19*, *CYP21A2*, *TNXB*) génje található egy kromoszómán, és nulla, egy vagy két pszeudogénje (*STK19P*, *CYP21A1P*, *TNXA*) az egy kromoszómán található modulszám függvényében. A *C4* génnek csak aktív változatai léteznek (*C4A* és *C4B*) (Blanchong et al. 2000) (5. ábra).

Az aktív *STK19* gén közvetlenül a *C4A* gén mellett helyezkedik el az ábrán, és egy szerin/threonin kináz típusú nukleáris proteint kódol. Csonkolódott formája a *TNXB* és a *C4B* között lokalizálódik. A *TNXB* gén a RCCX CNV 3'-végén található, egy a szívben és vázizmokban kifejeződő extracelluláris mátrix proteint kódol, a tenascin-X fehérjét. A gén csonkolt formája a *CYP21A1P* és az *STK19P* gének között foglal helyet (Yang et al. 1999). Az *STK19P* és *TNXA* pszeudogének, a *CYP21A1P* pszeudogénnel ellentétben, a modulok duplikálódása során csonkolódtak.



5. ábra: A bimoduláris RCCX CNV allél általános felépítése. Szaggatott függőleges vonal jelzi a modulok határát. Az RCCX struktúra ábrázolása felett a régióra jellemző gének rövidítései olvashatók, valamint a HERV-K inszerció megléte a C4 géneken belül.

A *CYP21A2* gén a *C4B* és *TNXB* gén között helyezkedik el az ábrán. Az 5'-irányba eső szélső és középső modulok *CYP21* génjei a legtöbbször pszeudogének (Yang et al. 1999). A *CYP21A1P* és *CYP21A2* gének 3'-vége átfedésben van a *TNXA* és *TNXB* gének 3'-végével, mivel ez utóbbiak leolvasási iránya ellentétes a modul többi génjével (Milner et al. 2001). A *CYP21A2* létfontosságú a glukokortikoid (kortizol) és mineralokortikoid (aldoszteron) hormonok bioszintézisében (White et al. 2000). *CYP21A2* és *CYP21A1P* génekben egyenként 10-10 exon helyezkedik el, DNS szekvenciájuk az exonokban 98%-ban, míg az intronokban 96%-ban megegyezik, tehát homológiájuk igen magas szintű (Yang et al. 1999). A nagymértékű hasonlóság ellenére, a pszeudogének számos, inaktív fehérjeterméket okozó mutációt hordoznak az aktív génhez képest, mint például egy 8 bp deléció a 3. exonban, négy kapcsolt szubsztitúció a 6. exonban. Ezen kívül számos más nukleotid eltérés is előfordul bennük a kódoló illetve nem-kódoló szakaszokban egyaránt (Yang et al. 1999). A *CYP21A2* gén mutációi által okozott szteroid 21-hidroxiláz deficiencia a humán szervezet metabolizmusának egyik leggyakoribb örökölt, veleszületett rendellenessége, és messze leggyakoribb oka a kongenitális adrenális hiperpláziának (congenital adrenal hyperplasia, CAH), amely csökkent vagy teljesen hiányzó kortizol és aldoszteron szinttel jár együtt (Sudmant et al. 2010). A CAH-ot okozó mutációk több mint 95%-áért NAHR esemény felelős. A mutációk 75%-a nem-allélikus génkonverzióból származó pont-, vagy kisebb kiterjedésű mutáció, amely a *CYP21A1P*-ről kerül át a *CYP21A2*-re, illetve nagyjából 20%-a egyenlőtlen átkereszteződéssel létrejövő nagy méretű (körülbelül 30 kb-os) deléció, amely a *CYP21A2* teljes vagy részleges deléciójával jár együtt (Concolino, 2010).

A C4 gének az *STK19* és *CYP21* gének között helyezkednek el, A C4 gén két, immunológiai funkcióban részben különböző altípusra osztható, a *C4A*-ra és a *C4B*-re. A két variáns szekvenciája több mint 99%-ban megegyezik, az immunológia szubfunkcionalizációt öt közeli nukleotid szubsztitúció (négy aminosav cseréje) okozza a 26. exonban (Yu et al.

2003). A *C4* gének 9. intronjában egy HERV-K endogén retrovírus inszerciójából származó deléciós CNV található. A deléción két CNV allélja kétféle méretű *C4* gént eredményez egy hosszabbat (long, L) és egy rövidebbet (short, S).

A különböző kópiaszám variációk, illetve az RCCX génjeinek változatai komplex struktúrájú CNV-t eredményeznek, amelyet hagyományosan az RCCX modulok számával, valamint modulonként a HERV-K inszerciójának meglétével vagy hiányával és a *C4* gén altípusával írnak le (Yu et al. 2003), annak ellenére, hogy ezek a polimorfizmusok eltérő jellegűek és méretűek. Modulonként az 5'-irányból először a HERV-K jelenlétét (L vagy S), majd a *C4* gén típusát jelöljük (A vagy B), annyiszor ismételve a két betűt, ahány modul található az RCCX CNV-ben a kromoszómán (6. ábra). Például a LA/LASB egy diploid állapotban három kópiával rendelkező egyén genotípusát jelöli, egy monomoduláris kromoszómán HERV-K inszercióval és *C4A* génnel, valamint egy bimoduláris kromoszómán 5'-irányban ugyancsak HERV-K inszercióval és *C4A* génnel, illetve 3'-irányban HERV-k inszerció nélkül és *C4B* génnel.

Az RCCX CNV felépítésének tanulmányozása nem könnyen kivitelezhető a gyakorlatban, például az egy modulban található HERV-K deléciós CNV allélja és a *C4* gén altípusa egészen a közelmúltig nem volt kísérletesen meghatározva. Kutatócsoportunk egy sorozat allél-specifikus long-range és kvantitatív PCR-t (ASLR- és quantitative PCR, qPCR) fejlesztett ki (6. ábra), amely az RCCX CNV egész variábilis régióját lefedi, és ezzel az eddigi legrészletesebb betekintést nyerte az RCCX CNV felépítésébe (Szilagyi et al. 2006; Bánlaki et al. 2012b). A technikai előrelépés ellenére az RCCX CNV moduljainak nagyfokú hasonlósága miatt, a diploid RCCX struktúra kombinációknak csak egy része határozható meg kísérletesen. A teljes egészében kísérletes diploid RCCX struktúra kombinációk meghatározásához a fenti technikákat és a családvizsgálat módszerét együttesen használta fel kutatócsoportunk (Bánlaki et al. 2012a), először szolgáltatva teljes egészében kísérletesen meghatározott eredményeket egy multiallélikus humán CNV szerkezetéről. Az egészséges magyar populációban (92 genetikailag független személy) a bimoduláris RCCX struktúrák fordulnak elő leggyakrabban (78,8%), a monomoduláris struktúrák (14,1%) elterjedtebbek a trimodulárisokhoz (7,1%) képest, illetve az előfordulások statisztikai hibahatáron belül megegyeznek egy másik európai populációival (Yang et al. 2003). A bimoduláris struktúrák közül a leggyakoribb az LALB (52,6%), majd az LASB (34,5%) és LALA (11,1%) struktúra néhány ritkább mellett. A monomoduláris struktúrák közül az LA (57,2%) némiképp gyakoribb, mint az SB (42,8%). Trimoduláris struktúrákból van a legtöbb változat, ezek közül az LASBSB (46,2%) a leggyakoribb, majd a LALASB és LBSASB struktúrák (15,4%)

következnek, végül az LALALB, LASALB, LALBLB struktúrák (7,6%) zárják a sort. Külön érdekesség, hogy az LASB struktúrák 4,2% (a teljes populációhoz viszonyítva 1,1%) mindkét moduljában a szokásostól eltérően *CYP21A2*-t hordoz (Bánlaci et al. 2012a).

1.3 A humán RCCX CNV struktúráját formáló genetikai átrendeződések jellemzése

A NAHR mechanizmusok, éppúgy mint általában a tandem duplikációk és multiallélikus CNV-k, hozzájárulnak az RCCX CNV komplexitásának létrejöttéhez, mivel a tandem duplikálódott *C4* és a *CYP21* gének hajlamosak a nem megfelelően illeszkedő interkromoszómális párosodásra. Egyfelől, az egyenlőtlen átkereszteződés alakítja ki a mono- és trimoduláris (Blanchong et al. 2000; Redon et al. 2006) és a ritka kvadrimoduláris RCCX CNV allélokat (Chung et al. 2002; Koppens et al. 2002b), a több mint egy *CYP21A2*-vel rendelkező kromoszómákat (Koppens et al. 2002b), a csak *CYP21A1P*-vel rendelkező kromoszómákat (Koppens et al. 2003), valamint a *CYP21A1P/CYP21A2* kiméra géneket (Koppens et al. 2002a; Lee 2005). Másrészt a *CYP21A1P* mutációi kis kiterjedésű nem-allélikus génkonverzióval helyeződnek át *CYP21A2*-be (Tusie-Luna et al. 1995; Hurles et al. 2008).

Általánosan elfogadott, hogy a mono- és trimoduláris RCCX struktúrák a bimoduláris RCCX struktúrákból keletkeznek egyenlőtlen átkereszteződés hatására (Yang et al. 1999; Blanchong et al. 2000). Bár a magyar és más európai populációk nagyjából 60%-a bimoduláris RCCX struktúrákat hordoz homozigóta formában, azonban körülbelül a személyek 10%-a a bimoduláris mellett trimoduláris RCCX struktúrát hordoz. A trimoduláris RCCX struktúrák tovább növelik az egyenlőtlen átkereszteződések bekövetkezésének esélyét, illetve az egyenlőtlen átkereszteződéssel létrejövő lehetséges, egymástól különböző struktúrák számát. A teljes hosszúságú *CYP21A2* haplotípusokat, mint komplex genetikai markereket (Hoehe 2003) felhasználó elemzés alapján egyes RCCX struktúrák többször is létre jöhetnek a humán intraspecifikus evolúció során, ilyen például az LA monomoduláris RCCX struktúra, amit tehát több, független egyenlőtlen átkereszteződés alakított ki. Azonban a RCCX struktúrák egy része, még a meglehetősen gyakoriak is, mint például az SB RCCX struktúra, egyetlen egyenlőtlen átkereszteződésre vezethetőek vissza (Bánlaci et al. 2012b). Az eddig legrészletesebben leírt egyenlőtlen átkereszteződés az RCCX CNV-ben egy kvadrimoduláris struktúra keletkezésének nyomon követése, amelyben a kiindulási RCCX struktúrák (egy bi- és egy trimoduláris struktúra) jellemzése mellett az RCCX-et körülvevő MHC markereket is megfigyelték, és egyértelműen azonosították a NAHR mechanizmust (Chung et al. 2002).

Különböző tanulmányokban többfajta olyan bi-; tri-; és kvadrimoduláris RCCX struktúrát találtak már, amelyek több mint egy *CYP21A2*-t tartalmaznak (Koppens et al. 2002b). Ezek közül az egyik legtöbbször leírt az egy hosszú és két rövid *C4* génnel, illetve egy *C4A* és két *C4B* génnel asszociált trimoduláris struktúra (Blanchong et al. 2000; Chung et al. 2002; Koppens et al. 2002b; Baumgartner-Parzer et al. 2007; Kleinle et al. 2009), amely minden valószínűség szerint azonos a LBSASB RCCX struktúrával (Bánlaki et al. 2012a). Ezen kívül még leírtak a két hosszú *C4* gént, és egy-egy *C4A*-t és *C4B*-t hordozó struktúrát (Chung et al. 2002; Koppens et al. 2002b), illetve a LASB struktúrát két *CYP21A2*-vel (Bánlaki et al. 2012a).

A csak *CYP21A1P*-vel rendelkező kromoszómák esetében, melyek a CAH egyik leggyakoribb okozói, a *CYP21A2* pszeudogén deléciójának legvalószínűbb koncepciója az, hogy egy aktív *CYP21A2* gént tartalmazó monomoduláris struktúra és egy hagyományos bimoduláris struktúra között történik rekombináció. Az ennek eredményeként létrejövő egyik kromoszóma két aktív *CYP21A2*-t fog tartalmazni, amíg a másik csak egy *CYP21A1P*-t (Koppens et al. 2003). Ugyanígy a *CYP21A1P/CYP21A2* kiméra gének létrejötte is az RCCX struktúrában lejátszódó egyenlőtlen átkereszteződések eredménye. A *CYP21A1P* és a *CYP21A2* gének közötti egyenlőtlen átkereszteződés egyes esetekben képezhet olyan kiméra gént, ami az 5'-régiójában *CYP21A1P* eredetű és számos mutációt tartalmaz, míg a 3'-régiójában *CYP21A2* génből származik (Lee 2005; Friaes et al. 2006). A *TNXB/TNXA* kiméra gének szintén egy bimoduláris és egy monomoduláris kromoszóma egyenlőtlen átkereszteződése által jönnek létre (Koppens et al. 2002a).

Az RCCX CNV-t már több, mint 20 éve kitűnő vizsgálati rendszernek tartják a NAHR tanulmányozásához (Tusie-Luna et al. 1995). Ennek ismeretében, és a CAH elterjedtségének ismeretében nem meglepő a tény, hogy a legtöbb génkonverziót (elsősorban nem allélikust) pont a *CYP21* géneket vizsgálva írták le emberben (Lawson et al. 2009). Az átrendeződési mechanizmusok emberben történő vizsgálata azonban nem egyszerű, mivel általában csak a rekombináció végső termékei elemezhetőek (Koppens et al. 2002b; Chen et al. 2007).

2. Célkitűzés

Célkitűzésünk a két *CYP21A2*-t hordozó LASB RCCX struktúrák és a kialakulásukért felelős genetikai átrendeződések tanulmányozása volt. Mivel az RCCX, mint a multiallélikus, tandem CNV-k általában, hajlamos a meiotikus átkereszteződéstől eltérő genetikai átrendeződésekre, mint például az egyenlőtlen átkereszteződés és az nem-allélikus génkonverzió, így elsősorban ezekről a mechanizmusokról szerettünk volna a lehetőségeinkhez mérten minél többet megtudni. A fenti vizsgálatok nem csak az RCCX CNV átrendeződési mechanizmusába, hanem általában véve a humán multiallélikus, tandem CNV-k átrendeződéseiért felelős mechanizmusokba is mélyebb bepillantást kívántunk tenni. A CNV-k speciális átrendeződési mechanizmusainak megismerése a CNV-kkel kapcsolt humán betegségek jobb megértéséhez és, végső soron, új terápiás lehetőségekhez vezethetnek.

3. Anyag és módszer

3.1 Vizsgálati minták

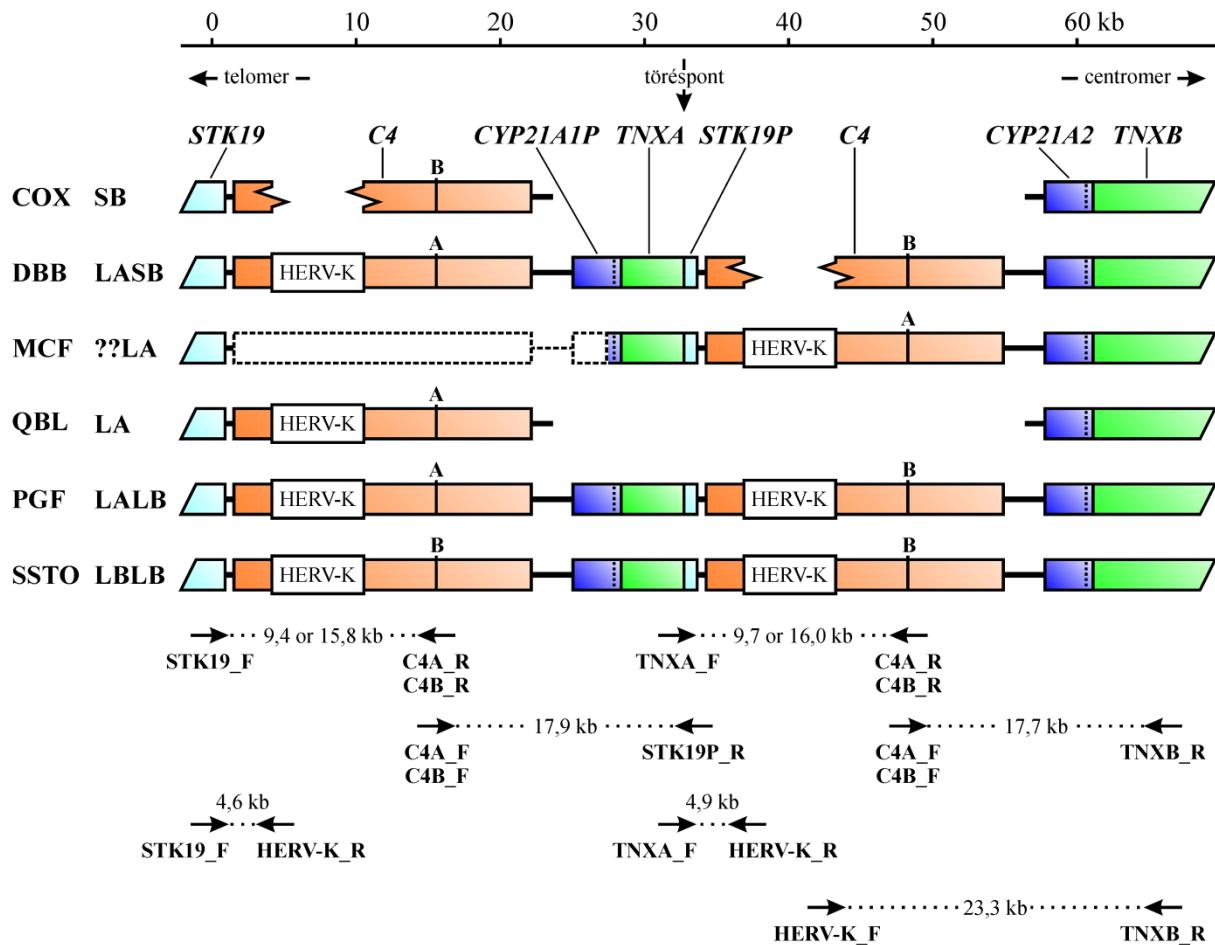
A jelen dolgozathoz a kutatócsoportunk két előző tanulmányból származó (Bánlaci et al. 2012a; Bánlaci et al. 2012b) duplikálódott *CYP21A2* gént hordozó LASB RCCX struktúrát hordozó személyeket használtuk fel a tanulmányok eredeti 332 genetikailag független személyéből válogatva. Végeredményben hat genetikailag független személy mintáját vizsgáltunk meg, vizsgálati személyeinkből háromban már meghatároztuk az RCCX struktúrát az előző tanulmányokban, háromban viszont nem, illetve a különböző modulokban elhelyezkedő *CYP21A2* haplotípusok szintén nem voltak meghatározva ezelőtt. Minden beteg beleegyezését adta a vizsgálatok elvégzéséhez, a tanulmányt a Magyar Nemzeti Etikai Bizottság és a Helsinki Nyilatkozat elveinek megfelelően végeztük el.

3.2 Az RCCX struktúra molekuláris haplotipizálása és meghatározása

Az RCCX struktúrákat ASLR-PCR-rel és qPCR-es *C4*, *CYP21*, és *HERV-K* kópiaszám vizsgálattal határoztuk meg (6. ábra).

A *C4A* és a *C4B* allél-specifikus long-range PCR forward és reverse primereket (*C4A_F*, *C4B_F*, *C4A_R* és *C4B_R*), együtt használtuk a különböző régiókra specifikus primerekkel. Az *STK19* és *TNXB* génekre specifikus primerek (*STK19P_F* és *TNXA_R*) csak is az aktív génhez illeszkedtek, míg a *STK19P* és *TNXA* génekre specifikus primerek (*STK19P_R* és *TNXA_F*) az aktív és a pszeudogénre is képesek voltak kötődni, azonban ezek csak akkor eredményezhettek PCR terméket a *C4* allél-specifikus primerekkel, ha a pszeudogénre tapadtak le. Végül a *HERV-K* specifikus primerek (*HERV-K_F* és *HERV-K_R*) csak arra a *C4* génre illeszkedtek, amely *HERV-K* inszerciót tartalmazott. Az említett primerekből 11 primer párt állítottunk össze:

- 6 primer pár tartalmazott *STK19_F* vagy *TNXA_F* forward primereket *C4A_R*, *C4B_R* vagy *HERV-K_R* reverse primerekkel kombinálva.
- 4 primer pár tartalmazott *C4A_F* vagy *C4B_F* forward primereket *STK19P_R* vagy *TNXB_R* reverse primerekkel kombinálva.
- 1 primer pár tartalmazott *HERV-K_F* forward primert *TNXB_R* reverse primerrel kombinálva.



6. ábra: Az RCCX CNV felépítése illetve az ASLR-PCR primerek és termékek elhelyezkedése. Hat publikált, HLA homozigóta sejtvonal RCCX struktúráinak ábrázolása szemlélteti a régióra jellemző változatosságot. Az ábrázolt régió az *STK19* gén 4. exonjának telomer végétől a *TNXB* 28. exonjának centromer végéig tart. A hiányzó szekvenciákat szaggatott vonal jelöli. A bimoduláris RCCX struktúrák variábilis régiója két teljes gént, a komplement komponens 4 (*C4A*; *C4B*), steroid 21-hidroxiáz (*CYP21A2*; *CYP21A1P*), és két részleges gént, szerin/reonin kináz 19 (*STK19*; *STK19P*), tenascin-X (*TNXA*; *TNXB*) tartalmaz. A *HERV-K* delécia a *C4* génben lokalizálódik. A 6. kromoszóma végének iránya és a bimoduláris struktúrák modul töréspontja a kilobázis mérettartományú skála alatt található. Az RCCX struktúrák sematikus rövidítései és a sejtvonalak nevei baloldalt látható. A PCR termékek hosszát és az ASLR-PCR primerek neveit és pozícióit a struktúrák ábrázolása alatt olvashatóak.

Az ASLR-PCR-t 10 µl-es végtérfogatban végeztük, amelyben a reakcióelegy a következő összetevőket tartalmazta:

- 1 U LongAmp *Taq* DNS polimeráz;
- 1x LongAmp *Taq* reakció puffer;
- 300 µM dNTP;
- 0,4 µM a megfelelő primerekből;
- 10-100 ng genomialis DNS (a DNS minőségétől függően).

Az alábbi táblázatban tüntetjük fel a különböző primerek esetében használt PCR profilokat.

Név	SA	SB	TA	TB	AS	BS	AT	BT	SH	TH	HT
forward primer	STK19F		TNXA_F		C4A_F	C4B_F	C4A_F	C4B_F	STK19_F	TNXA_F	HERV-K_F
reverse primer	C4A_R	C4B_R	C4A_R	C4B_R	STK19P_R		TNXB_R		HERV-K_R		TNXB_R
ciklus szám	30	30	30	30	30	30	32	30	30	30	32
annealing T (°C)	64	64	66	66	64	64	64	64	66	66	64
extenziós idő	13 min 30 sec				15 min				4 min		21 min

A *C4A*, *C4B*, *CYP21A1P*, *CYP21A2* gének és a HERV-K inszerció diploid kópiaszámát qPCR segítségével határoztuk meg, ahogy azt korábbi tanulmányokban leírták (Szilagyi et al. 2006; Wu et al. 2007). *C4* specifikus Taqman próbákat (AppliedBiosystem) 6-FAM fluoreszcens festékkel, az RNázP belső referencia gén próbáját VIC festékkel jelöltük meg.

3.3 A *CYP21A2* gén amplifikálása és szekvenálása

A *CYP21A2* gének 5'- és 3'-résznek megfelelő allél-specifikus nested PCR alkalmazásával vizsgáltuk (7. ábra).

Templátnak az ASLR-PCR eljárásból származó geno- vagy haplotipikus termékeket használtuk fel:

- a C4A_F vagy C4B_F forward primerek és STK19P_R vagy TNXB_R reverse primerek termékét
- HERV-K_F és TNXB_R primerek termékét

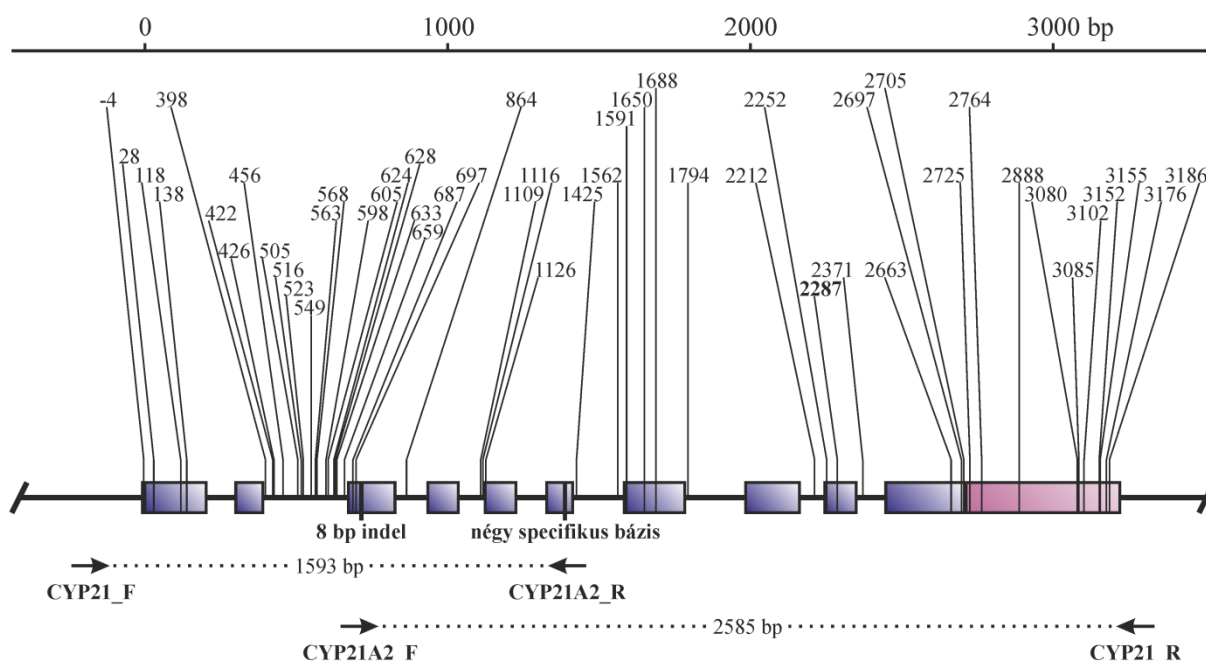
Az 5'-rész allél-specifikus primere (*CYP21A2_R*) a 6. exon allél-specifikus nukleotid szubsztitúciójaira illeszkedtek, a hozzá tartozó forward primer (*CYP21_F*) az 5'-határoló régióra tapadt. A 3'-rész amplifikálásához az allél-specifikus primer (*CYP21A2_F*) a két gén között a 3. exonban lévő 8 bp-os különbséget használták ki, és a 3'-határoló régióban tapadó primerrel (*CYP21_R*) képeztek párt.

A nested PCR-t 15 µl-es végtérfogatban végeztük, amelyben a reakcióelegy a következő összetevőket tartalmazta:

- 1 U GoTaq DNS polimeráz (Promega);
- 1 x GoTaq colorless Flexi puffer;
- 1,5 mM MgCl₂;
- 200 µM dNTP;

- 133 nM a megfelelő primerekből;
- ~4 ng ASLR-PCR termék.

A ciklus első lépése a 95°C-on 5 percig tartó elődenaturáció, majd ezt 15 ciklus követte, ami ciklusonként három lépésből állt: 10 másodperc denaturálás 95°C-on, 5 másodperces annealing 64 °C-on, végül 90 másodperc extenzió 72°C-on (CYP21_F és CYP21A2_R primerpárokkal), vagy 150 másodperc extenzió szintén 72°C-on (CYP21A2_F; CYP21_R primerpárokkal). Végül a nested PCR reakció befejezéseként 72°C-on 5 perces utóextenzió következett.



7. ábra. A *CYP21A2* szegregációs helyeinek ábrázolása. Az exonokat kék és rózsaszín színek szimbolizálják, kék szín jelzi a kódoló régiót, rózsaszín a nem transzlálódó régiót. A szegregációs helyeket a pozícióik száma jelöli a *CYP21A2* kódoló régió kezdeti szakaszától induló számozással. A vastagított számú szegregáló hely csak a duplikálódott *CYP21A2* gént hordozó RCCX struktúrákban található meg. A primerek pozíciója, neve és a PCR termék hossza az ábra alján látható.

A nested PCR eljárást követően a teljes *CYP21A2* gént megszekvenáltuk kapilláris gélelektroforézissel. A nested PCR termékeket a tisztítás során exonukleáz I. enzimmal és rAPid alkalikus-foszfáttal kezeltük, majd a DNS mindkét szálát megszekvenáltuk 7-7 szekvenáló primerrel. A szekvenálás során Big Dye Terminator Sequencing Kit v3.1 (Applied Biosystem) kitet használtunk, és ABI 3100 Genetic Analyzer (Applied Biosystems) készüléken futtattuk.

3.4 A PCR termékek elválasztása DNS gélelektroforézis segítségével

A DNS fragmensek elválasztására agaróz gélelektroforézist használtunk. Az elektroforézishez az ASLR-PCR termékeihez 0,7 %-os, és az ebből induló allél-specifikus nested PCR termékeihez 1% agaróz gél készítettünk. A gélelektroforézist 1% TBE oldatban 60 V feszültségen, 45 percig futott. A gél festéséhez etídium-bromidot alkalmaztunk, mely lehetővé teszi, hogy UV fénnel megvilágítva a szeparálódott DNS sávok láthatóvá váljanak. Az ASLR-PCR termékei esetében λ fág Hind III létrát, amíg az allél -specifikus nested PCR termékeihez 1 kb-os létrát használtam a elválasztott fragmensek méreteinek megállapításához.

3.5 Bioinformatikai elemzések

A *CYP21A2* gén kromatogramjainak leolvasásához a CLC DNA Workbench v5.7.1 (CLC bio) programot használtuk, a program eredményeit két különböző ember vizuálisan ellenőrizte. Azokat a *CYP21A2* haplotípusokat, amelyeket kísérletesen nem tudtunk meghatározni, PHASE v2.1.1 program segítségével rekonstruáltuk a genotípus adatokból. A kísérletesen meghatározott, illetve rekonstruált haplotípusokból Network v4.6 programmal median-joining hálózatot építettünk fel. Az eljárás során két parciális orángután *CYP21A2* szekvenciát egyesítettünk és használtunk kulcsoportként a hálózat felállításához, mivel a publikus szekvencia adatbázisokban nem állt rendelkezésre teljes hosszúságú emberszabású majom *CYP21A2* ortológ szekvencia. Ezért az orángutánokban lévő gap (1764-től 2341-ig PGF szekvencián) mutációt az ENSEMBL-ből egészítettük ki egy előző tanulmány részleges ortológ szekvenciájából.

4. Eredmények és értékelésük

4.1 A haploid RCCX CNV struktúra és a *CYP21A2* haplotípus meghatározása

Az RCCX CNV struktúrák szerveződésének meghatározásához az RCCX modulok számát, a *C4* gén változatait, a *CYP21* gén változatait és a HERV-K inszerciót vizsgáltuk ASLR-PCR és qPCR eljárásokkal (6. ábra). Habár az ASLR-PCR eljárás során a *C4* gén altípusát és a HERV-K inszerció meglétét vagy hiányát egyszerre tudtuk meghatározni egy modulon belül, a modul számokat és a modulok közötti kapcsolatokat az ASLR-PCR-ek és a qPCR által kapott eredményekből következtettük ki. A teljes hosszúságú *CYP21A2* haplotípusokat az ASLR-PCR termékekből határoztuk meg allél-specifikus nested PCR eljárással, és az azt követő *CYP21A2* szekvenálással. Azokban esetekben, amikor egy személyben a haplotipikus PCR termékek hiányoztak a *CYP21A2* gént csak genotipizálni tudtuk, a hat személy között egyetlen egy volt ilyen. A RCCX CNV-t tanulmányozó eljárásokkal ellentétben, az ASLR-PCR eljárással nem csak a kromoszómákat genetikai információ tartalmát tudtuk szétválasztani egymástól (haplotipizálni), hanem a különböző elhelyezkedésű (paralóg) modulok információ tartalmát is (Bánlaci et al. 2012b). Ez tette lehetővé, hogy külön meghatározzuk mindkét, egy kromoszómán elhelyezkedő *CYP21A2* szekvenciáját.

A haploid RCCX struktúrák kikövetkeztetésének menetét a hat publikált, HLA homozigóta sejtvonal RCCX szekvencia illesztésén mutatjuk be (6. ábra). Diploid fázisban a COX és QBL sejtvonal két monomoduláris RCCX struktúrája nem képez PCR terméket a *TNXA* és *STK19P* génekből, azonban az *STK19_F* és a *C4A_R* primerek 15,8 kb fragmenst eredményeznek a QBL-nek megfelelően, továbbá az *STK19_F* és a *C4B_R* primerek 9,4 kb terméket állítanak elő a COX-szal megegyezően. A *TNXB_R* primer a *C4A_F* és *C4B_F* primerekkel (két külön reakcióban) két 17,7 kb méretű haplotípus terméket hoznak létre a COX és QBL diploid kombinációjából. A két ASLR-PCR termékből mindkét *CYP21A2* haplotípus amplifikálható és megszekvenálható, így a megfelelő RCCX struktúrákhoz párosíthatóak a *CYP21A2* haplotípusok. Továbbá a HERV-K_F és *TNXB_R* primerek által adott 23,3 kb termékből szintén meghatározható a QBL-hez kapcsolódó *CYP21A2* haplotípus, mivel ilyen ASLR-PCR termék nem alakul ki COX-ból. A két azonos monomoduláris RCCX struktúra szerveződése a fenti módon leírható, azonban a *TNXB_R* primer ASLR-PCR termékei nem különültek el egymástól, így a két kromoszóma *CYP21A2* haplotípusait csak genotipizálni tudjuk. Tehát a *CYP21A2* haplotípusokat csak olyan személyekben tudjuk

meghatározni, amely két különböző *C4* génnel és két eltérő HERV-K CNV alléllal rendelkeznek a *TNXB*-vel szomszédos régióban.

A fentihez hasonló következtetések más RCCX struktúrák diploid kombinációira is alkalmazhatóak. Általánosságban nézve az RCCX kombinációk közül számosan pusztán az ASLR-PCR eredményei alapján egyértelműen meghatározhatóak voltak a haploid struktúrák, még akkor is, ha az összes elképzelhető RCCX struktúrát figyelembe vettük. Ezekben az esetekben az ASLR-PCR és a qPCR kópiaszámra vonatkozó eredményei azonosak voltak, ami mindkét eljárás megbízhatóságát igazolta. Bizonyos kombináció ugyanazt az ASLR-PCR mintázatot mutatták, de a kópiaszámuk alapján meg tudtuk különböztetni őket egymástól és a haploid struktúrák kikövetkeztethetőek voltak. Mellesleg megjegyezve, bizonyos kombinációkat nem lehet a kísérleti eredményekből kikövetkeztetni, ekkor csak a lehetséges kombinációk száma szűkíthető, de a hat vizsgált személy közül egy sem volt ilyen.

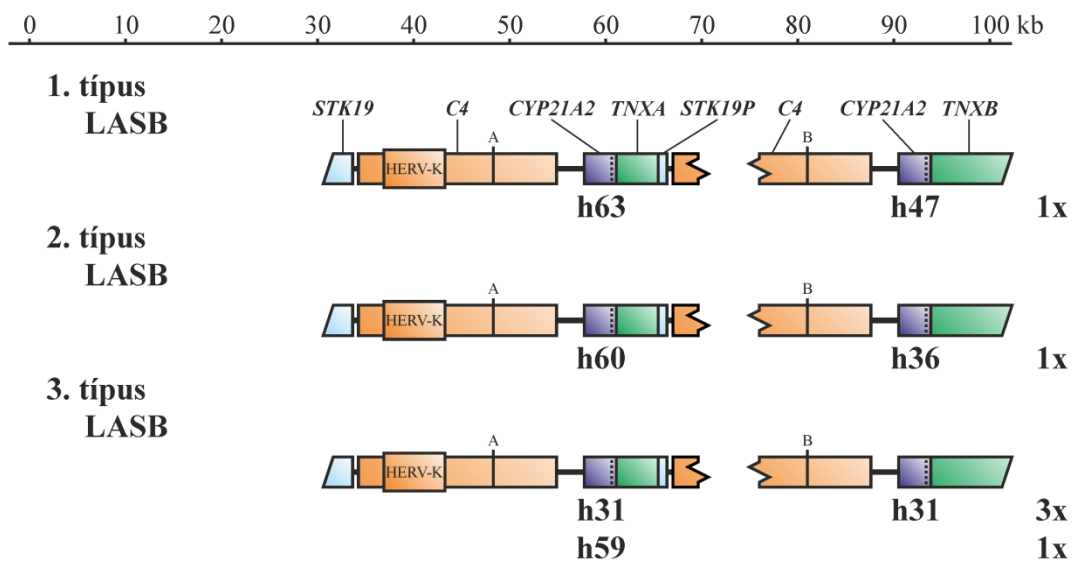
A *CYP21A2* szekvenálás során 1 új szegregáló helyet (mutációt és polimorfizust együttesen) tártunk fel a 48 szegregáló helyhez képest (7. ábra), amelyek az egy *CYP21A2* hordozó RCCX struktúrák haplotípusaiban találhatók (Bánlaci et al. 2012b). A *CYP21A2* haplotípusok meghatározása az ASLR-PCR haploid termékeinek a hiányában nem történhetett meg egy személyben, így ebben az esetekben a *CYP21A2* gént csak genotipizáltuk. Ennél a mintánál, PHASE program segítségével készítettük el a haplotípus rekonstrukciót, amely megfelelt a 0,99-es konfidencia határnak (confidence probability threshold).

4.2 A két *CYP21A2*-t tartalmazó haploid LASB RCCX struktúrák

A vizsgálataink során három olyan egyedi LASB RCCX struktúrát találtunk, amely duplikálódott *CYP21A2* géneket tartalmazott (8. ábra). A három LASB struktúra közül kettőt csak egy mintában azonosítottunk, miközben egyet négyszer is leírtunk. A haploid RCCX struktúrákat és a hozzájuk tartozó *CYP21A2* haplotípusokat csak egy esetben nem tudtuk kísérletesen meghatározni, de ez pont annál a típusban történt, ahol két teljesen azonosat meghatároztunk.

Az első LASB struktúra egyik esetében a duplikálódott *CYP21A2* gének közül az RCCX struktúra 3'-felőli (centroméra felőli) h47 haplotípust, míg a 5'-felőli (teloméra felőli) h63 haplotípust hordozott. A h47 *CYP21A2* haplotípus szintén előfordul a normális, egy 5'-modulban elhelyezkedő *CYP21A1P* pszeudogént, és egy 3'-modulban elhelyezkedő *CYP21A2* gént hordozó LASB struktúrában is, amelyből feltehetőleg a duplikálódott *CYP21A2*-t tartalmazó RCCX struktúra származik. A h63 haplotípus az 52 általunk talált szegregáló hely

közül az 5'-vég felé eső 45 hely alléljában megegyezik a h47 haplotípussal, így az 5'-vég felőli része a h47 haplotípus ezen részének megfelelő szekvenciájával volt azonos. A maradék 7 szegregáló hely alléljai a *CYP21A1P* pszeudogénnek megfelelőek, tehát a haplotípus 3'-vége a *CYP21A1P* pszeudogénből ered. Ráadásul 2765-ös hely A allélja ritka, de a *CYP21A1P* pszeudogént és h47 haplotípust hordozó LASB struktúrához tartozó pszeudogénben pont megtalálható, ami az azonos kromoszóma vagy kromoszómák közötti mechanizmusra utal.



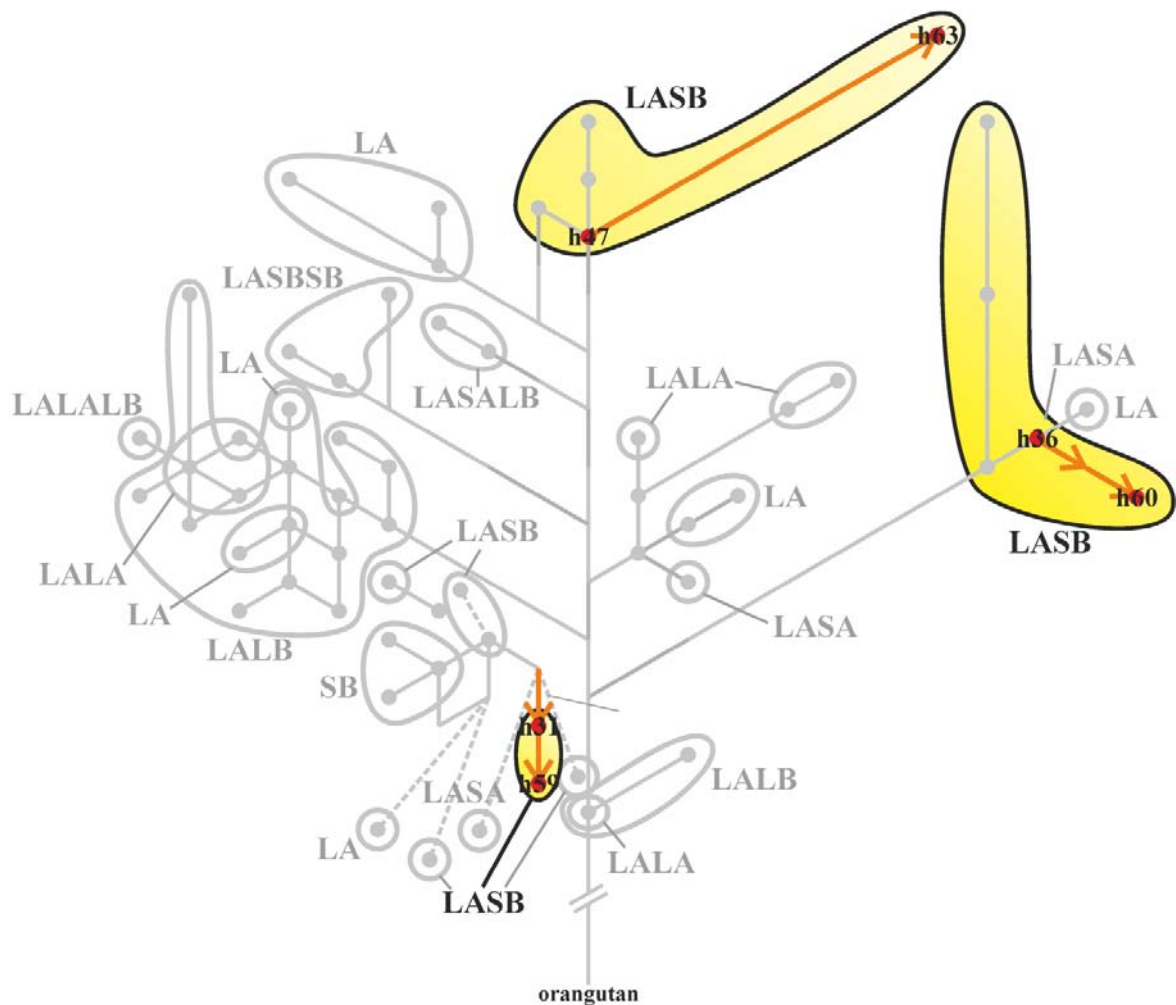
8. ábra. Három egyedi két aktív *CYP21A2* gént tartalmazó RCCX struktúra felépítése. A kék színnel jelzett *CYP21A2* gének alatt látható azok haplotípusa. Bal oldalt az RCCX struktúrák rövidítéseit, jobb oldalt a leírt esetek számát tüntettük fel.

A második LASB struktúra 5'-vég felé eső *CYP21A2* génje h60 haplotípust, míg az 3'-vég felé eső h36 haplotípust tartalmazott. A 3'-modulban elhelyezkedő h36 haplotípus normál, 5'-pszeudogént hordozó LASB struktúrában is előfordult. A két haplotípus csaknem azonos utalva az eredetüknek kapcsoltágára, a h60 haplotípus egy ritka mutációban (1126) és gyakoribb polimorfizmusban (2705) tért el a h36 haplotípustól.

A harmadik LASB struktúrát négy különálló esetben írtuk le, ezek közül háromszor mindkét *CYP21A2* gén esetén h31 haplotípust határoztunk meg. Egy mintánál azonban a 3'-vég felőli *CYP21A2* h31 haplotípusa mellett egy h59 haplotípust figyeltünk meg, ami csak egy teljesen egyedi mutációban (2287) tért el az előbbtől. A h31 haplotípus szintén megtalálható normális, az 5'-modulban *CYP21A1P* pszeudogént hordozó LASB struktúrában is. Ha a három LASB struktúra *CYP21A2* haplotípusait összevetjük, 6-12 szegregáló hely allél különbséget találunk, amely a struktúrák független kialakulását bizonyítja.

4.3 A két *CYP21A2*-t tartalmazó RCCX struktúrák *CYP21A2* haplotípusainak elhelyezkedése a genealógiai hálózaton

Az genealógiai haplotípus hálózatok alkalmasak az intraspecifikus szint különleges tulajdonságainak megjelenítésére, mint például a létező ősi haplotípusok, a többszörös elágazások és a hálózaton belüli hurkok (Posada et al. 2001). A két *CYP21A2*-t tartalmazó RCCX struktúrák *CYP21A2* haplotípusainak származását median-joining módszerrel épített genealógiai hálózaton követtük nyomon, az genealógiai hálózat felépítéséhez a korábban már publikált egy *CYP21A2*-t tartalmazó kromoszómákban fellelhető *CYP21A2* haplotípusokat használtuk fel (Bánlaki et al. 2012a).



9. ábra. Két *CPY21A2* gént tartalmazó struktúrák ábrázolása a normál *CYP21A2* haplotípusokból levezett filogenetikai hálózatában. Az általunk vizsgált egyedi, duplikálódott *CYP21A2* géneket tartalmazó RCCX struktúrák haplotípusainak neveit sárga színnel jelöltük a hálózatban. Piros nyilak szimbolizálják az allélhely változásokat.

Az alkalmazott módszer feltételeinek a *CYP21A2* haplotípus variánsok teljes mértékben megfeleltek, mivel a gén rövidsége miatt átkereszteződések nincsenek a génen belül, és a meiotikus rekombináció aránya is alacsony ebben a régiókban. A hálózat felállításában csak a *CYP21A2* haplotípus variánsok vettek részt, az RCCX struktúrákat nem vettük figyelembe. Ezért ebben a tanulmányban a felhasznált *CYP21A2* haplotípusok komplex genetikai markereknek tekinthetők, az így alkalmazott haplotípusok a teljes RCCX CNV genealógiáját jellemzik. A kapott hálózat evolúciós irányultságát két szekvencia ötvöztetésével felállított orángután ortológ adta meg. Ez a szekvenciát nevezzük a hálózat kulcsoportjának vagy gyökerének.

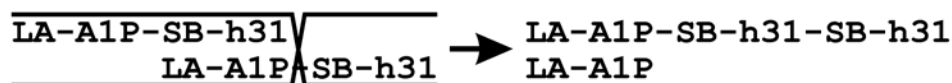
A *CYP21A2* haplotípus hálózat egy faszzerű struktúrát alkotott, hurkokkal és sok többszörös elágazással (9. ábra). A *CYP21A2* haplotípusokra rávetítve a hozzájuk tartozó RCCX struktúrákat, az RCCX struktúrák jól elkülöníthető klaszterekbe csoportosulását figyeltük meg. Az általunk vizsgált egyedi, duplikálódott *CYP21A2* géneket tartalmazó LASB RCCX struktúrák szintén LASB struktúrához tartozó haplotípusokból származnak, így elmondható, hogy a keletkezésük során nem módosult az RCCX struktúrájuk. Tehát a HERV-k inszerció és *C4* gén altípus állapota nem változott meg, amiből egyértelmű következtetések vonhatóak le a *CYP21A2* duplikáció lehetséges mechanizmusaira.

5. Összefoglalás

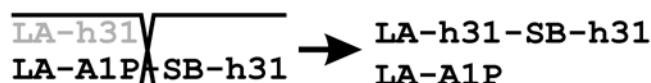
A meghatározott haploid RCCX struktúrákra, a hordozott *CYP21A2* haplotípusokra illetve a haplotípusok szegregáló helyeinek alléljeire alapozva három azonos vagy nagyon hasonló *CYP21A2* haplotípust tartalmazó LASB struktúrákat határoztunk meg. A nagyszámú szegregáló helyet tartalmazó haplotípusok ilyen szintű konvergens kialakulása gyakorlatilag kizárt, még akkor is a duplikálódott régiókra jellemző nem-allélikus gén konverziók homogenizáló hatását (Teshima et al. 2004) is figyelembe vesszük, így a két azonos vagy csaknem azonos haplotípus eredete kapcsolt. A három LASB struktúrához tartozó haplotípusok egymástól nagymértékben eltérőek, amely természetesen a genealógiai hálózaton is eltérő elhelyezkedést eredményez, így a három LASB struktúra független alakult ki. Mivel mindhárom két *CYP21A2*-t tartalmazó LASB struktúrának még létezik az 5'-modulban *CYP21A1P* pszeudogént hordozó normál változata (létező, ősi haplotípus), az evolúciós származásuk minden valószínűség szerint intraspecifikus, és a pszeudogént tartalmazó struktúrákból levezethető. Ráadásul, a normál, egy *CYP21A2*-t tartalmazó RCCX struktúrák haplotípusainak és a két *CYP21A2*-t tartalmazó LASB struktúrák haplotípusainak polimorfizmus különbségei is evolúciós következtetéseket tesznek lehetővé, mivel a haplotípusok szétválása között eltelt időre a haplotípusok közötti eltérések számából következtethetünk. A h47-et és h63-at hordozó LASB struktúra esetében a h47-ből származó *CYP21A2* szekvenciák teljesen azonosak, így a struktúra kialakulása fiatalnak tekinthető, mivel a két függetlenné vált szekvenciának nem volt ideje újabb változásokat beszerezni. Majdnem ugyanez mondható el a h31 haplotípust hordozó struktúráról, mivel a h31 haplotípus duplikációját tartalmazta három minta is és csak egy minta esetében figyelhettük meg a h31 haplotípus helyett h59 haplotípust az 5'-modulban, ami csak egyetlen szegregáló hely allélben (2287) tér el az előbbitől. Tehát a második LASB struktúra esetében szintén a közel múltban megtörtént eseményt feltételezhetünk, bár a haplotípus variációk szétválása már elkezdődött.

A három LASB típus azonos tulajdonságai (azonos vagy csaknem azonos *CYP21A2* haplotípus az 5'- és 3'-modulban, létező normál, pszeudogén tartalmú LASB struktúra azonos 3'-*CYP21A2* haplotípussal) alapján feltehető, hogy a struktúrákat azonos genetikai átrendeződési mechanizmus alakította ki. Habár az RCCX CNV átrendeződési mechanizmusairól általánosan ismert tény, hogy a mono- és trimoduláris RCCX struktúrák bimoduláris struktúrák egyenlőtlen átkereszteződéséből jöttek létre, (Koppens et al. 2003; Concolino et al. 2010) de az egy *CYP21A2*-t tartalmazó LASB struktúrákból egyszeres

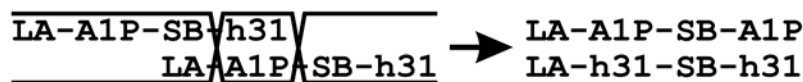
egyenlőtlen átkeresztezéssel nem alakulhat ki két azonos *CYP21A2*-t tartalmazó LASB struktúra, csak LASBSB struktúra (10. ábra). Különböző modulszámú struktúrák közötti egyenlőtlen átkeresztezések is elképzelhetők, egy azonos *CYP21A2* haplotípust hordozó LA és LASB kromoszóma egyszeres egyenlőtlen átkereszteződésével létrejöhet LASB struktúra két *CYP21A2*-vel, de egyetlen egy LA struktúrát se találtunk a három két *CYP21A2*-t tartalmazó LASB típusának megfelelő 3'-*CYP21A2* haplotípussal a teljes vizsgálati anyagot képző 332 személyben. Az RCCX régióra ugyancsak jellemző kis kiterjedésű génkonverziók (Bandelt et al. 1999; Blanchong et al. 2000) se jöhetnek szóba, mint potenciális mechanizmus, mivel az bizonyítottan átkerült kromoszóma szakaszok (teljes vagy majdnem teljes *CYP21A2* gén) túl hosszúak ehhez.



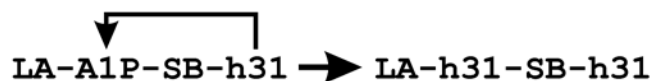
egyszeres egyenlőtlen átkereszteződés
a két *CYP21A2*-t tartalmazó LASB RCCX struktúrák között



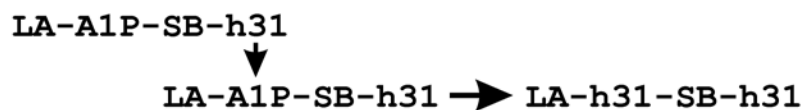
egyszeres egyenlőtlen átkereszteződés
hipotetikus LA RCCX struktúra és
azonos *CYP21A2*-t tartalmazó LASB RCCX struktúra között



kettős egyenlőtlen átkereszteződés a testvérkromatidák között



intrakromoszómális nagy kiterjedésű génkonverzió



interkromoszómális nagy kiterjedésű génkonverzió a testvérkromatidák között

10. ábra. A duplikálódott *CYP21A2*-t hordozó LASB RCCX struktúrák kizárható és lehetséges mechanizmusai. Az ábra csak h31 haplotípusra nézve szemlélteti a lehetséges mechanizmusokat, azonban ezek az összes két *CYP21A2* gént tartalmazó LASB haplotípusra vonatkoznak (h36; h47).

A három LASB típushoz vezető lehetséges átrendeződési mechanizmusok közül a kettős egyenlőtlen átkereszteződés a testvérkromatidák között, a intrakromoszómális nagy kiterjedésű konverzió, és az interkromoszómális konverzió a testvérkromatidák között jöhet szóba. Mind a három lehetséges mechanizmusban közös, hogy az azonos RCCX struktúrán belül, vagy struktúrák között játszódik le, ez a tulajdonságuk a három LASB típus genealógiai származásából fakad. Ezt támasztja alá a h63 haplotípus speciális mutációja (2765) a két *CYP21A2*-t LASB struktúra 5'-modul *CYP21* lókuszán, ez a mutáció ugyanis egy másik ritka haplotípustól eltekintve ugyanennek a LASB struktúrának a normál, *CYP21A1P* pszeudogént tartalmazó változatán fordul csak elő.

A fentihez hasonló nagy-méretű átrendeződéseket megfigyelték már kisebb gyakorisággal a 3'-modul *CYP21* lókuszán is, ezekben az esetekben a *CYP21A1P* pszeudogénnek nagyobb darabjai inzertálódnak bele a *CYP21A2*-be illetve az azt körülvevő határoló régióba. Konkrétan négy ilyen típusú átrendeződést írtak le egy európai populáció esetében. Az egyik parciális génkonverzió az 1. exontól a 7. exonig terjedt ki, ez volt a leggyakoribb az összes többi közül. Egy másik esetben az 1. exontól a 3. exonig inzertálódott a *CYP21A1P* pszeudogén egy része, míg a maradék két esetben ez a 3. exontól a 7. illetve a 8. exonig nyúlt. Az utóbbi három parciális génkonverzió már jóval kisebb esetben fordult elő. (Friaes et al. 2006)

Bár a pontos mechanizmus nem levezethető az általunk rendelkezésre álló adatokból, csak a két széles körben elfogadott NAHR mechanizmus zárható ki, más duplikálódott humán gének esetében intrakromoszómális átrendeződéseket írtak le a közelmúltban (Ganster et al. 2010; Roehl et al. 2010), amelyek a mi esetünkben is az egyik lehetséges magyarázata a duplikálódott *CYP21A2*-t tartalmazó LASB RCCX struktúra keletkezésének. Az egyik ilyen esetet a 7. kromoszómába beágyazott duplikációban írták le, melyben egy DNS hibajavító aktív gén és a pszeudogénje közötti szekvencia kicserélődések során hasonló kiméra gének keletkeznek, mint az RCCX CNV régió esetében. A kiméra génekhez ugyan különböző haplotípusok tartoznak, de ezek visszavezethetők egyetlen ősi intrakromoszómális átkereszteződésre. Három különböző variáns esetében is leírtak egy olyan 5 kb méretet meghaladó inszerciót ami a normál génben volt jelen, azonban bizonyosan a pszeudogénből származott. Ezek a több mint 5 kb kiterjedésű szegmensek jócskán meghaladják azt a szekvencia méretet, amit a szakirodalomban a génkonverzióknak tulajdonítanak (Ganster et al. 2010). Ugyancsak egy humán duplikálódott régióban, a 17. kromoszómán a neurofibromin 1 régióban, írtak le kiméra génekre alapozva intrakromoszómális átrendeződéseket, amelyek valószínűleg mitotikus eredetűek (Roehl et al. 2010).

Ezen kívül *in vitro* emlős sejtes rendszerben a nagy kiterjedésű génkonverziók is leírtak (Neuwirth et al. 2007), de ezeknek *in vivo* létezését mai napig megkérdőjelezzük emberben (Chen et al. 2007). Interkromoszómális génkonverzióval specifikusan asszociálódott átkereszteződéseket emlőssejtek esetén még nagyon ritkán írtak le, azonban két izogén humán B-limfoblasztóma sejtvonal esetében ez sikerült. Egy egyik sejtvonalban az összefüggő, komplex génkonverzió hossza meghaladta a 11.2 kb-t. (Neuwirth et al. 2007). Bár az nem világos, hogy humán, nagy kiterjedésű génkonverziós események léteznek-e *in vivo*, de egyértelműen a mai napig nem sikerült kizárni, tehát, mint lehetőséggel, ezzel is számolni kell a szóba jövő mechanizmusoknál az általunk vizsgált modellben.

Összességében a három felvázolt lehetséges NAHR mechanizmus közül bármelyik is legyen a tényleges, az elfogadott kis kiterjedésű génkonverzió és egyszeres egyenlőtlen átkereszteződésen kívül egy harmadik mechanizmus is jellemzője az RCCX CNV-nek, és általánosítva, a duplikálódott régióknak.

Irodalom

- Alkan, C., B. P. Coe and E. E. Eichler (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**(5): 363-376.
- Bandelt, H. J., P. Forster and A. Rohl (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**(1): 37-48.
- Bánlaki, Z., M. Doleschall, K. Rajczy, G. Fust and Á. Szilágyi (2012a). Fine-tuned characterization of RCCX copy number variants and their relation to extended MHC haplotypes. *Megjelenés alatt*.
- Bánlaki, Z., J. Szabó, Á. Szilágyi, A. Patócs, Z. Prohászka, G. Fust and M. Doleschall (2012b). Intraspecific evolution of human RCCX copy number variation traced by the haplotypes of CYP21A2 gene. *Megjelenés alatt*.
- Baumgartner-Parzer, S. M., G. Fischer and H. Vierhapper (2007). Predisposition for de novo gene aberrations in the offspring of mothers with a duplicated CYP21A2 gene. *J Clin Endocrinol Metab* **92**(3): 1164-1167.
- Beckmann, J. S., X. Estivill and S. E. Antonarakis (2007). Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* **8**(8): 639-646.
- Blanchong, C. A., B. Zhou, K. L. Rupert, E. K. Chung, K. N. Jones, J. F. Sotos, W. B. Zipf, R. M. Rennebohm and C. Yung Yu (2000). Deficiencies of human complement component C4A and C4B and heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in caucasians. The load of RCCX genetic diversity on major histocompatibility complex-associated disease. *J Exp Med* **191**(12): 2183-2196.
- Chen, J. M., D. N. Cooper, N. Chuzhanova, C. Ferec and G. P. Patrinos (2007). Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* **8**(10): 762-775.
- Chung, E. K., Y. Yang, R. M. Rennebohm, M. L. Lokki, G. C. Higgins, K. N. Jones, B. Zhou, C. A. Blanchong and C. Y. Yu (2002). Genetic sophistication of human complement components C4A and C4B and RP-C4-CYP21-TNX (RCCX) modules in the major histocompatibility complex. *Am J Hum Genet* **71**(4): 823-837.
- Concolino, P., E. Mello, C. Zuppi and E. Capoluongo (2010). Molecular diagnosis of congenital adrenal hyperplasia due to 21-hydroxylase deficiency: an update of new CYP21A2 mutations. *Clin Chem Lab Med* **48**(8): 1057-1062.
- Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* **464**(7289): 704-712.
- Cooper, D. N., A. Bacolla, C. Ferec, K. M. Vasquez, H. Kehrer-Sawatzki and J. M. Chen (2011). On the sequence-directed nature of human gene mutation: the role of genomic architecture and the local DNA sequence environment in mediating gene mutations underlying human inherited disease. *Hum Mutat* **32**(10): 1075-1099.
- Durbin, R. M., G. R. Abecasis, D. L. Altshuler, A. Auton, L. D. Brooks, R. A. Gibbs, M. E. Hurles and G. A. McVean (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**(7319): 1061-1073.
- Freeman, J. L., G. H. Perry, L. Feuk, R. Redon, S. A. McCarroll, D. M. Altshuler, H. Aburatani, K. W. Jones, C. Tyler-Smith, M. E. Hurles, et al. (2006). Copy number variation: new insights in genome diversity. *Genome Res* **16**(8): 949-961.
- Friaes, A., A. T. Rego, J. M. Aragues, L. F. Moura, A. Mirante, M. R. Mascarenhas, T. T. Kay, L. A. Lopes, J. C. Rodrigues, S. Guerra, et al. (2006). CYP21A2 mutations in

- Portuguese patients with congenital adrenal hyperplasia: identification of two novel mutations and characterization of four different partial gene conversions. *Mol Genet Metab* **88**(1): 58-65.
- Ganster, C., A. Wernstedt, H. Kehrer-Sawatzki, L. Messiaen, K. Schmidt, N. Rahner, K. Heinemann, C. Fonatsch, J. Zschocke and K. Wimmer (2010). Functional PMS2 hybrid alleles containing a pseudogene-specific missense variant trace back to a single ancient intrachromosomal recombination event. *Hum Mutat* **31**(5): 552-560.
- Hastings, P. J., J. R. Lupski, S. M. Rosenberg and G. Ira (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* **10**(8): 551-564.
- Hoehe, M. R. (2003). Haplotypes and the systematic analysis of genetic variation in genes and genomes. *Pharmacogenomics* **4**(5): 547-570.
- Horton, R., L. Wilming, V. Rand, R. C. Lovering, E. A. Bruford, V. K. Khodiyar, M. J. Lush, S. Povey, C. C. Talbot, Jr., M. W. Wright, et al. (2004). Gene map of the extended human MHC. *Nat Rev Genet* **5**(12): 889-899.
- Hurles, M. E., E. T. Dermitzakis and C. Tyler-Smith (2008). The functional impact of structural variation in humans. *Trends Genet* **24**(5): 238-245.
- Kleinle, S., R. Lang, G. F. Fischer, H. Vierhapper, F. Waldhauser, M. Fodinger and S. M. Baumgartner-Parzer (2009). Duplications of the functional CYP21A2 gene are primarily restricted to Q318X alleles: evidence for a founder effect. *J Clin Endocrinol Metab* **94**(10): 3954-3958.
- Koppens, P. F., T. Hoogenboezem and H. J. Degenhart (2002a). Carriership of a defective tenascin-X gene in steroid 21-hydroxylase deficiency patients: TNXB -TNXA hybrids in apparent large-scale gene conversions. *Hum Mol Genet* **11**(21): 2581-2590.
- Koppens, P. F., T. Hoogenboezem and H. J. Degenhart (2002b). Duplication of the CYP21A2 gene complicates mutation analysis of steroid 21-hydroxylase deficiency: characteristics of three unusual haplotypes. *Hum Genet* **111**(4-5): 405-410.
- Koppens, P. F., H. J. Smeets, I. J. de Wijs and H. J. Degenhart (2003). Mapping of a de novo unequal crossover causing a deletion of the steroid 21-hydroxylase (CYP21A2) gene and a non-functional hybrid tenascin-X (TNXB) gene. *J Med Genet* **40**(5): e53.
- Lawson, M. J., J. Jiao, W. Fan and L. Zhang (2009). A pattern analysis of gene conversion literature. *Comp Funct Genomics*: 761512.
- Lee, H. H. (2005). Chimeric CYP21P/CYP21 and TNXA/TNXB genes in the RCCX module. *Mol Genet Metab* **84**(1): 4-8.
- Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, et al. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature* **470**(7332): 59-65.
- Milner, C. M. and R. D. Campbell (2001). Genetic organization of the human MHC class III region. *Front Biosci* **6**: D914-926.
- Neuwirth, E. A., M. Honma and A. J. Grososky (2007). Interchromosomal crossover in human cells is associated with long gene conversion tracts. *Mol Cell Biol* **27**(15): 5261-5274.
- Posada, D. and K. A. Crandall (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* **16**(1): 37-45.
- Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, et al. (2006). Global variation in copy number in the human genome. *Nature* **444**(7118): 444-454.
- Roehl, A. C., J. Vogt, T. Mussotter, A. N. Zickler, H. Spoti, J. Hogel, N. A. Chuzhanova, K. Wimmer, L. Kluwe, V. F. Mautner, et al. (2010). Intrachromosomal mitotic nonallelic homologous recombination is the major molecular mechanism underlying type-2 NF1 deletions. *Hum Mutat* **31**(10): 1163-1173.

- Stankiewicz, P. and J. R. Lupski (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet* **18**(2): 74-82.
- Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. de Grassi, C. Lee, et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**(5813): 848-853.
- Sudmant, P. H., J. O. Kitzman, F. Antonacci, C. Alkan, M. Malig, A. Tsalenko, N. Sampas, L. Bruhn, J. Shendure and E. E. Eichler (2010). Diversity of human copy number variation and multicopy genes. *Science* **330**(6004): 641-646.
- Szilagyi, A., B. Blasko, D. Szilassy, G. Fust, M. Sasvari-Szekely and Z. Ronai (2006). Real-time PCR quantification of human complement C4A and C4B genes. *BMC Genet* **7**: 1.
- Teshima, K. M. and H. Innan (2004). The effect of gene conversion on the divergence between duplicated genes. *Genetics* **166**(3): 1553-1560.
- 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**(23): 10796-10800.
- Weaver, R. W. and P. W. Hedrick (2000). *Genetika*. Budapest, Panem.
- White, P. C. and P. W. Speiser (2000). Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev* **21**(3): 245-291.
- Wu, Y. L., S. L. Savelli, Y. Yang, B. Zhou, B. H. Rovin, D. J. Birmingham, H. N. Nagaraja, L. A. Hebert and C. Y. Yu (2007). Sensitive and specific real-time polymerase chain reaction assays to accurately determine copy number variations (CNVs) of human complement C4A, C4B, C4-long, C4-short, and RCCX modules: elucidation of C4 CNVs in 50 consanguineous subjects with defined HLA genotypes. *J Immunol* **179**(5): 3012-3025.
- Yang, Y., E. K. Chung, B. Zhou, C. A. Blanchong, C. Y. Yu, G. Fust, M. Kovacs, A. Vatay, C. Szalai, I. Karadi, et al. (2003). Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with C4 gene size and polygenic variations, hemolytic activities, and body mass index. *J Immunol* **171**(5): 2734-2745.
- Yang, Z., A. R. Mendoza, T. R. Welch, W. B. Zipf and C. Y. Yu (1999). Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module). A mechanism for gene deletions and disease associations. *J Biol Chem* **274**(17): 12147-12156.
- Yu, C. Y., E. K. Chung, Y. Yang, C. A. Blanchong, N. Jacobsen, K. Saxena, Z. Yang, W. Miller, L. Varga and G. Fust (2003). Dancing with complement C4 and the RP-C4-CYP21-TNX (RCCX) modules of the major histocompatibility complex. *Prog Nucleic Acid Res Mol Biol* **75**: 217-292.
- Zhang, F., W. Gu, M. E. Hurles and J. R. Lupski (2009). Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* **10**: 451-481.

Köszönetnyilvánítás

Szeretném megköszönni az OTDK dolgozat írása során nyújtott segítségét, tanácsait és értékes információit Dr. Doleschall Mártonnak, témavezetőmnek. Továbbá köszönöm a támogatást Prof. Dr. Füst Györgynek, Dr. Prohászka Zoltánnak és Dr. Szilágyi Ágnesnek, akik munkám során segítségemre voltak.