

Adatelemző módszerek alkalmazása a HIV-fertőzés kutatásában

Bartha István

Az Eötvös Loránd Tudományegyetem Természettudományi Karának Biológia Doktori Iskolája

Az iskola vezetője: Dr. Erdei Anna

Elméleti és evolúcióbiológia program

Programvezető: Dr. Szathmáry Eörs

Témavezető: Dr. Müller Viktor, tudományos főmunkatárs

Növényrendszertani, Ökológiai és Elméleti Biológiai Tanszék

2013

Tartalomjegyzék

1. Bevezetés	3
1.1. A HIV-fertőzés epidemiológiája	4
1.2. A HIV molekuláris biológiája és életrajza	5
1.3. A HIV patogenezise	10
1.4. A HIV-vírus evolúciója	14
1.5. Genetikai asszociációvizsgálat	15
1.6. A HIV-vírussal kapcsolatos humán genetikai faktorok irodalmának áttekintése .	20
2. A humán genom és a HIV-1 genom variációinak együttes asszociációvizsgálata	24
2.1. Eredmények	25
2.1.1. Rendelkezésre álló adatok	25
2.1.2. Az adatok összefésülése	26
2.1.3. Humán polimorfizmusok vs. HIV-mutációk	29
2.1.4. Virális mutációk vs. vírusszint	31
2.2. Diskusszió	34
2.3. Összefoglalás	38
2.4. Genetikai asszociációvizsgálat során alkalmazott módszerek	38
3. A HIV-felülfertőzés gyakoriságának becslése rutin genotipizálás során gyűjtött vírusszekvenciákból	47
3.1. Eredmények	49
3.1.1. Felhasznált adatok és leíró statisztikáik	49
3.1.2. Felülfertőzés kimutatása filogenetikai módszerekkel	49
3.2. Diskusszió	59
3.3. Összefoglalás	61
3.4. Felülfertőzés kimutatásában használt módszerek	61
3.5. Függelék	65
3.5.1. Szekvenciátávolságon alapuló módszerekkel nem lehet megbízhatóan kimutatni a kettős fertőzést	65
3.5.2. Előzetes szűrés maximum likelihood módszerrel	66

4. Részfeladatok egyéb projektekben	70
4.1. T-sejt populáció életidejének becslése	70
4.2. APOBEC3G indukálta hipermutáció nyomainak keresése HIV-szekvenciákban	71
4.3. Vizualizációs platformok készítése génexpressziós vizsgálatokhoz	71
5. Köszönetnyilvánítás	72

1. Bevezetés

A szerzett immunhiányos tünetegyüttest (Acquired Immune Deficiency Syndrome - AIDS) 1981-ben fedezték fel az Egyesült Államok nagyvárosainak homoszexuális közösségeiben (Masur, 1981; Mildvan, 1982). Az alig több mint három évtizede felismert betegség a legsúlyosabb közegészségügyi kihívások egyike lett, eddig megközelítőleg 25 millió ember halálát okozva. A Humán Immundeficiencia Vírus névre keresztelt kórokozót két évvel később, 1983-ban izolálták (Barre-Sinoussi, 1983; Gallo, 1983).

A World Health Organization (WHO) 34 millióra becsli a jelenleg a vírussal élők számát (UNAIDS, 2012). A patogén vírus ellen nem áll rendelkezésre végleges terápia, illetve hatásos oltóanyag. A fertőzés mortalitása magas, kezelés nélkül csaknem minden esetben halálhoz vezet. Továbbá eddig csupán négy esetben sikerült a vírust maradéktalanul eltávolítani a beteg szervezetéből (Check Hayden, 2013; Hütter, 2009). Ezek ellenére óriási előrelépések történtek a betegség megszelídítésében, folyamatos gyógyszeres kezelés mellett a tünetek megelőzhetőek, illetve jelentősen késleltethetőek.

1984-ben azonosították a vírus elsődleges célsejtjeit, a CD4+ T-limfocitákat (Klatzmann, 1984). A vírus molekuláris diverzitására 1985-ben derült fény, ekkor publikálták először a vírus genomját. Az első antiretrovirális hatóanyagot, a zidovudine-t (AZT), 1987-ben engedélyezték. Ez volt az első gyógymód, amely a több ezer HIV-fertőzött beteg rendelkezésére állt (Rowland-Jones, 2004). Önmagában ez az egy hatóanyag nem bizonyult megfelelőnek a vírus visszaszorításában, ezért 1996-ben több hatóanyag kombinálásával vezették be a nagy hatékonyságú antiretrovirális terápiát (highly active antiretroviral therapy - HAART, máshol cART). A kombinációs terápia képes a vérben mért vírusszintet a detekciós küszöb alá csökkenteni, és a páciensek várható élettartamát és életminőségét jelentős mértékben növelni. Az ART megjelenése óriási áttörést jelentett a vírussal élők életszínvonalában és a várható élettartamukban. A gyógyszerek ma már nemcsak a fejlett világban, hanem Fekete-Afrikában is elérhetőek, ahol hatmillió HIV-fertőzött áll terápia alatt (UNAIDS, 2012). A kezeléssel azonban nem tüntethető el a vírus teljesen a szervezetből. Ennek bizonyítéka, hogy a kezelés felfüggesztése esetén a vérbeli vírusszint előbb-utóbb ismét megemelkedik (Davey, 1999). Az első nagy léptékű vakcinakísérletekre 2003-ban került sor (Barouch & Korber, 2010; Sekaly, 2008), míg az antiretrovirális anyagok megelőző alkalmazásáról végzett kísérletek néhány éven belül követték azokat (Liu

AY Buchbinder SP, 2006; Abdool Karim, 2010; Grant, 2010). Az elmúlt évtizedek kutatásai során rengeteget megtudtunk az emberi immunrendszer működéséről és a vírus patogenezisééről, de széleskörű gyakorlati alkalmazásba vehető végleges, a fertőzést teljesen megszüntető terápia vagy oltóanyag kifejlesztése eddig sikertelennek bizonyult.

Az AIDS-re nyilvánvalóan nem tekintünk genetikai betegségként, azonban az epidemiológiájában jelentkező változatosságot legalább részben humán genetikai tényezők határozzák meg. Ugyanis nem mindenki fertőződik meg a vírussal, illetve a fertőzöttek között is nagy eltéréseket mutat a betegség tüneteinek a megjelenése (Fellay & McLaren, 2013). A fertőzött egyének különböző erősségű szerzett, illetve veleszületett immunitással rendelkeznek, illetve az antiretrovirális terápiára is eltérő hatékonysággal válaszol a szervezetük.

A dolgozatomban először röviden összefoglalom a HIV-vírussal és a genetikai asszociációvizsgálatokkal kapcsolatos általános szakirodalmat. A további fejezetekben tárgyalom az általam végzett két kutatási projektet azok specifikus irodalmi összefoglalójával, módszereivel és eredményeivel. Az első projekt a virális és a humán genom együttes genetikai asszociációvizsgálatával tárja fel azon lokuszokat a HIV- és az emberi genomban, amelyek kölcsönhatásban állhatnak egymással. A második projektben pedig a HIV felülfertőzés gyakoriságát vizsgálom klinikai adatbázisokból vett szekvenciák analízisével. Végül röviden ismertetem azokat a munkákat, amelyekben a doktori tanulmányaim alatt társszerzőként részfeladatokat végeztem.

1.1. A HIV-fertőzés epidemiológiája

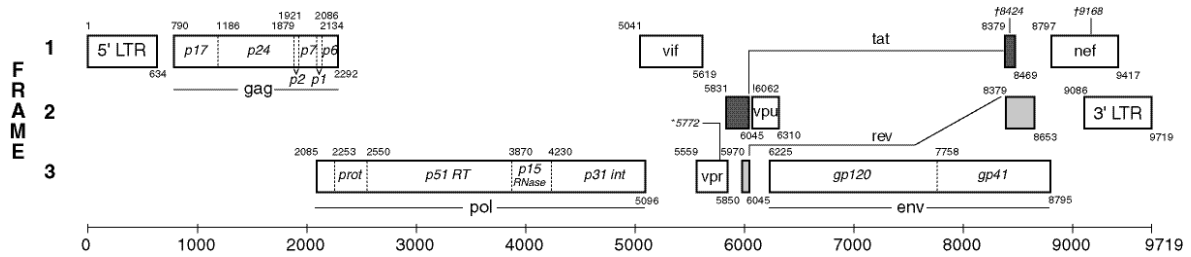
A HIV-vírus a nem ember főemlősöket fertőző Majom Immundeficiencia Vírusokból (Simian Immunodeficiency Virus - SIV) származtatható. Földrajzi eredete Közép- és Nyugat-Afrika. A vírusnak két főbb változatát ismerjük: a HIV-1 vírusok közvetlen ősei a csimpánzokat fertőző SIVcpz, illetve a gorillákat fertőző SIVgor, míg a HIV-2 vírusok őse a kormos mangábéket fertőző SIVsm vírus. A HIV-1 törzs egyik altípusa (M) felelős a jelenlegi világméretű járványért, míg a HIV-2 törzs elsősorban Nyugat-Afrikában fordul elő. A HIV-1 vírusokat molekuláris bélyegek alapján három csoportra osztjuk: valószínűleg mindhárom csoport (M, N, O) egy-egy egyedi és egyszer bekövetkezett fajok közötti átugrásból származik (Rambaut, 2004). Az utóbbi években egy negyedik csoportra (P) is fény derült (Vallari, 2011). A HIV-1 járvány kezdetét a becslések a 20. század első harmadára teszik. Ezt molekuláris filogenetikai vizsgálatok, és szociológiai vizsgálatok támasztják alá (de Sousa, 2010).

Az M csoport tovább bontható 9 altípusra (A-D,F-H,J,K altípusok) és számos rekombináns altípusra (circulating recombinant form - CRF). Ezek az altípusok külön járványokban a múlt század hetvenes éveiben törtek ki Közép-Afrikából. A nyugati világban a HIV-1 B altípusa a legelterjedtebb, ez a járvány 1966-ig, Haitiig vezethető vissza (Gilbert, 2007). Az A altípus nagy prevalenciával bír a korábbi Szovjetunió tagállamaiban, míg a C altípus főként Kelet-Ázsiában, Indiában és Brazíliában terjedt el, továbbá ez az uralkodó típus Dél-Afrikában is . Több cirkuláló rekombináns forma is kiterjedt járványokat okoz, ezek közül a legfontosabb a Thaiföldön igen gyakori CRF-01AE. Kelet-Afrikában jellemzően az A és D, illetve a CRF-02AG altípusok találhatóak meg. Globálisan a C altípus a leggyakoribb változat (UNAIDS, Global Report, 2012). Közép-Afrikában Kamerunban és a Kongói Demokratikus Köztársaságban a HIV-1 M csoport diverzitása olyan nagy, hogy nincs értelme elkülöníteni az alcsoportokat, azok beágyazódnak a helyi vírusok filogenetikai fájába . Valójában a Föld többi részén található alcsoportok egy-egy alapító törzs elterjedése révén jöttek létre (Rambaut, 2001).

A világon jelenleg 34 millió HIV-fertőzött ember él, ebből 23 millió Fekete-Afrikában, ahol a fertőzés prevalenciája a felnőtt lakosság körében elérheti az 20%-ot is. Nyugat-Európában és Észak-Amerikában 2,1 millió ember él a vírussal, míg a volt Szovjetunió tagállamaiban 1,5 millió. Ez utóbbi az egyetlen régió, ahol az elmúlt tíz évben jelentősen nőtt és jelenleg is növekszik a vírus előfordulási gyakorisága (0,3%-ról 0,9%-ra). Kelet-Európában a járvány robbanásszerű növekedést mutat a kábítószerhasználók között. Az új HIV-fertőzöttek száma a világon általában lassú csökkenést mutat köszönhetően a sokrétű megelőző programoknak és az antiretrovirális gyógyszerek hatékony és széleskörű alkalmazásának (UNAIDS, Global Report, 2012).

1.2. A HIV molekuláris biológiája és életciklusa

A vírusrészcseke (virion) egy membránnal körülvett vezikula amelyben a két egyszálú RNS molekulából álló genomot egy fehérjekapszid védi. A körülbelül 9700 nukleotid hosszú genom 9 gént kódol. A vírus főbb strukturális fehérjéi a *gag* gén által kódolt p17 mátrix, p24 kapszid és p7 nukleokapszid, a *pol* gén által kódolt proteáz, reverz transzkriptáz és integráz enzimek, valamint az *env* gén által kódolt gp120 és gp41 burokképletek. A Nef, Vif, Vpr, Vpu (HIV-2-ben Vpx), Tat és Rev fehérjék kiegészítő és szabályzó szerepet töltenek be, valamint ezek felelősek a gazdasejt védelmi mechanizmusainak leküzdéséért (Frankel & Young, 1998) (lásd



1. ábra. A vírus genomi szerveződésének ábrázolása. A kép a los alamosi adatbázis honlapjáról származik (<http://hiv.lanl.gov>).

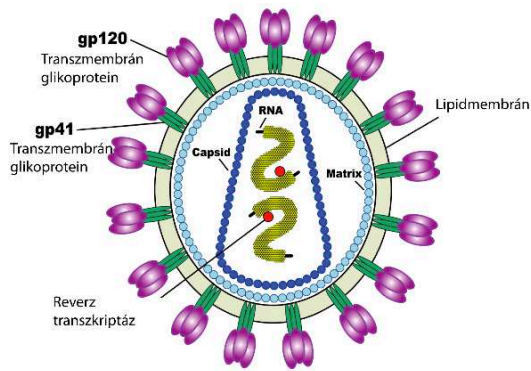
2. és 1. ábra).

A vírus gp120 fehérjéje a virion membránjának külső oldalán található, ez létesít kapcsolatot a CD4 receptorral (Poignard, 2001). A gp120 és a CD4 molekulák közötti interakció során megfigyelhető egy konzervatív konformációs epitóp, ami potenciális célpontja lehet HIV-ellenes vakcináknak, illetve egyéb hatóanyagoknak (Chen, 2009; Zhou, 2007). A vírusnak a citoplazmába való bejutáshoz a fő CD4 receptoron kívül egyéb koreceptorokhoz is kötődnie kell, ilyenek például a CCR5 és CXCR4 kemokinreceptorok.

A legtöbb természetes úton átadott vírus a CCR5 koreceptorhoz kapcsolódik, majd a betegség előrehaladtával a gazdán belüli evolúció során fejlődik ki a CXCR4 receptort preferáló variáns. Ez a koreceptor-váltás körülbelül a páciensek felében történik meg, és a betegség súlyosbodásához vezethet (Levy, 2009). A koreceptorok fontosságát jelzi, hogy az európai humán populációban 10%-os allélgyakorisággal jelen levő CCR5 Δ 32 mutáció homozigóta formája szinte teljes rezisztenciát biztosít a HIV ellen (Mulherin, 2003). A CD4 receptorhoz és a koreceptorokhoz való kötődés konformációváltozást indít el a gp41 fehérjében, amely ennek hatására a membránhoz horgonyozza a vírust és megindítja a membránfúziót (Buzon, 2010). Az utóbbi években bevezetett fúziógátló hatóanyagok ezt a konformációváltozást gátolják a vírus életciklusában (Henrich & Kuritzkes, 2013; Tilton & Doms, 2010). A CCR5 receptor allosztérikus gátlása szintén hatásos a vírusfertőzés ellen (Garcia-Perez, 2011). Utóbbi stratégia azért előnyös, mert nem a vírust, hanem a stabil genetikai állománnyal rendelkező gazdát célozza.

Ugyan a vírus elsődleges célsejtjei a CD4+ T-limfociták, egy sor más sejt is képes megfertőződni endocitózis útján, például makrofágok, mikroglia és dendritikus sejtek (Levy, 2009; He, 1997). A transzmisszió¹ utáni korai szakaszban valószínűleg fontos szerepet töltenek be a

¹Az az esemény, amikor a vírus egy emberről egy másikra jut.



2. ábra. A vírusrészecske vázlatos felépítése. A kép a Wikipediáról származik.

vírus elterjedésében a dendritikus sejtek, amelyek DC-SIGN receptorok segítségével megkötik és a nyirokcsomókba szállítják a virionokat. Ott immunológiai szinapszisekat létesítve a CD4+ T-limfocitákkal elősegítik a vírus szétterjedését a testben (Lore & Larsson, 2003).

A membránfúzió után a kapszidburok felbomlik, és a virionban lévő reverz transzkripció komplex bejut a citoplazmába. A burok felbomlásában és a sejt felszínhez közeli aktinfilamentumokon való átjutásban fontos szerepet játszik a Nef protein (Gallant, 2008). A reverz transzkripció során a vírus egyszálú RNS genomjából kétszálú cDNS szintetizálódik (Gallant, 2008). A reverz transzkriptáz blokkolása kiváló lehetőség a vírusreplikáció megakadályozására. Ez történhet nukleozid analógokkal (pl.: zidovudine), vagy nem nukleozid analógokkal (pl.: efavirenz), amelyek az enzim aktív helyén kívül fejtik ki hatásukat (Lapkouski, 2013; Knoll, 2007). Az utóbbi években felfedezett humán *SAMHDI* gén terméke a citoplazma nukleotidkoncentrációjának csökkentésével akadályozza meg a fertőzést nyugalmi állapotban levő dendritikus sejtekben, makrofágokban, illetve CD4+ T-limfocitákban (Schaller, 2012; Laguette, 2011; Descours, 2012). A reverz transzkripció befejeztével az új DNS molekula nem válik le a reverz transzkripció komplexéről, hanem úgynevezett preintegrációs komplexet alkot mind virális, mind gazda eredetű fehérjékkel (Gallant, 2008). A reverz transzkripció komplex és a preintegrációs komplex is aktívan kapcsolódik a sejt vázhoz, előbbi aktinfilamentumokhoz, utóbbi mikrotubulusok segítségével halad a sejt mag felé. A virális reverz transzkriptáz enzim nem rendelkezik hibajavító mechanizmusokkal, ezért hibaráta különösen magas (3×10^{-5} hiba/nukleotid/generáció) (Preston, 1988). Hatékony replikációval és rövid generációs idővel kombinálva ez teszi lehetővé a vírus igen gyors evolúcióját (Rambaut, 2004).

A preintegrációs komplex mérete jóval nagyobb, mint a nukleopórus átmérője, a nukleáris

transzporthoz ezért szükség van a nukleáris importfehérjék segítségével. A vírus p17 matrix és Vpr fehérjéin találunk nukleáris import szignálokat, amely fehérjék mind részei a preintegrációs komplexnek. (Gallant, 2008)

A sejtban az integráz segítségével a vírus kétszálú DNS-e beépül a gazda kromoszomális genomjába. Az integrálódott vírust provírusnak hívjuk. A vírus általában aktív kromoszómaterületekbe épül be, ezzel is növelve a saját replikációs hatékonyságát (Craigie & Bushman, 2012; Ocwieja, 2011). Az antiretrovirális hatóanyagok egy új osztályát képezik az integrázinhibitorok, amelyek a vírus örökítőanyagának beépülését gátolják a gazda genomjába (Messiaen, 2013).

Az integrált provírus lehet lappangó, látens állapotban, mikor a vírus fehérjéi nem expresszálódnak és a sejt nem termel fertőző virionokat. Ilyenkor a fertőzött sejt nem prezentál virális epitópeket a felszínén, így az immunrendszer elől a fertőzés rejtve marad. Látens állapot bekövetkezhet többek között akkor, ha az adott genomi régió heterokromatin állapotot vesz fel, vagy ha a sejt nyugalmi állapotban van és egyes alapvető transzkripció faktorok (NF- κ B, NFAT) a citoplazmában lokalizálódnak (Trono & Marzetta, 2011; Trono, 2010). Másrészt a provírusról folyhat aktív transzkripció és génexpresszió. Ehhez szükségesek az előbbi transzkripció faktorok és a virális Tat fehérje jelenléte a sejtban. A provírus 5' végén eukarióta promóter szekvenciák találhatóak, de a Tat nélkül ez a promóter gátlás alatt áll (Chou, 2013). A látens módon integrált provírus az egyik legfőbb oka annak, hogy a kombinációs terápia képtelen megszüntetni a fertőzést (Trono, 2010). Amíg a szervezetben van olyan hosszú életidejű nyugvó sejt (például memória CD4+ T-sejt (De Boer & Perelson, 2013)), amelynek a genomjában a provírus megtalálható, addig ebből a provírusból a virémia kiújulhat. Ugyanakkor elképzelhető olyan anatómiai vagy celluláris kompartment a szervezetben, ahol a hatóanyagok penetrációja rossz, így a vírus szuppresszív terápia alatt is szaporodhat. Ennek ellentmond, hogy hatásos terápia alatt evolúciós változások nem figyelhetők meg a vírus genomjában (Ruff, 2002; Tobin, 2005), és az intenzívebb terápia se akadályozza meg a virémia későbbi kiújulását (Dinosa, 2009).

Az igen tömör, körülbelül 9700 bázispár hosszú HIV genomban több gén kódja egymással átfed. Ez csökkenti az örökítőanyag méretét, de szükségessé teszi, hogy transzkripció során ugyanarról a szakaszból többféle splice-variáns készüljön. Főként a genom 3' vége felé eső felében kódolt auxiliáris gének transzkriptumai esnek át ezen a többszörös szerkesztésen (Pur-

cell & Martin, 1993). Ilyenek a *nef*, *tat* és *rev* gének, amelyek termékei szükségesek a genom további expressziójához és a hatékony vírusreplikációhoz. A genom nem szerkesztett (unspliced) transzkriptumai a Rev fehérjén lévő nukleáris export szignál segítségével szállítódnak ki a sejtmagból, ahol megkezdődik a transzláció és a vírus érése (Gallant, 2008). A Nef protein oly módon befolyásolja a gazdasejt állapotát, hogy az kedvezzen a vírusreplikációnak. A fehérje fontosságát jelzi, hogy diszfunkcionális *nef* gént tartalmazó vírussal fertőzött betegekben a betegségprogresszió jóval lassabb (Kirchhoff, 1995). A Nef internalizálja a sejt felszíni CD4 receptorokat és MHC I molekulákat, ezzel akadályozva a sejtet funkciójában, megnehezíti a sejt további fertőzését, és elfedi azt a citotoxikus immunválasz elől (Michel, 2005). A fehérje a gazdasejt apoptózisát gátolja, viszont serkenti a Fas-ligandum kihelyezését a membránra, ezzel a szomszédos sejtek apoptózisát előidézve (Baur, 2011). A gp120 és Vpu fehérjék az endoplazmatikus retikulumban kötődnek a készülő CD4 receptorokhoz, ezzel akadályozva azok exportját a sejt felszínre (Gallant, 2008).

Az expresszált Gag és Gag-Pol poliproteinek, illetve a vírusgenom RNS transzkriptumai a sejtmembrán belső oldalához szállítódnak. Ott összeáll a virion, ami a gazda vezikuláris membránrendszerét felhasználva lefűződik a sejtmembránról. (Gallant, 2008)

Az utóbbi években négy olyan úgynevezett restriktív faktort azonosítottak a főemlősök genomjában amelyek beavatkoznak a HIV (illetve SIV) életciklusába. A gazdaspecifikus fertőzőképes vírusok (ember esetén a HIV, csimpánz esetén a SIVcpz) mind alkalmazkodtak a specifikus gazda hasonló funkciójú génjeihez. Más gazdafajba oltva őket azonban nem minden esetben képesek replikációra, többek között azért nem, mert a gazda restriktív faktorait nem képesek ellensúlyozni (Laguette & Benkirane, 2012; Kirchhoff, 2010). Az első ilyen gén az *APOBEC3G*, ennek terméke becsomagolódik a virionba, majd a következő sejt fertőzése során deaminálja a vírus RNS genomját, ezzel hipermutációt okozva. Az gazda *APOBEC3G* génjeivel való kompatibilitásért felelős a *vif* gén, amely ubikvitináció útján lebontásra tereli az *APOBEC3G* fehérjét. (Kirchhoff, 2010)

Nem ember főemlősökben azonosították a *TRIM5 α* gént, amely terméke a kapszidburok megfelelő időben történő felbomlását akadályozza. A humán genomban lévő *TRIM5 α* -val a HIV-vírus *capsid* génje kompatibilis, de a rézusmajmokban található homológgal nem. A rézusmajmokat a HIV ezért nem képes megfertőzni. Érdekes, hogy egyetlen egy aminosav megváltoztatása a humán *TRIM5 α* fehérjében már képes megakadályozni a produktív fertőzést

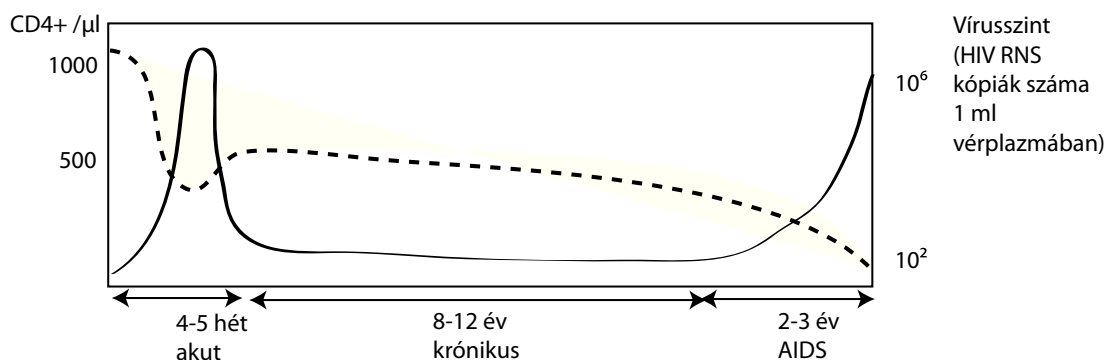
(Yap, 2005).

További restriktációs faktor a *tetherin* (más néven *BST-2*), amely a lefűződött és membránnal körülvevett virionokat horgonyozza a sejtfelszínhez. Ellensúlyozására a lentivírusok körében több megoldás is született, egyes esetekben a *nef*, máshol (A HIV-1-et is ideértve) pedig a *vpu* termékei felelősek a tetherin-kompetibilitásért (Neil, 2008; Kirchhoff, 2010).

A *SAMHD1* gén terméke nagymértékben hozzájárul ahhoz, hogy a nyugalmi állapotban lévő, differenciált mieloid sejtek, illetve T-limfociták rezisztensek a vírusfertőzésre (Descours, 2012). A gén terméke a sejtek citoplazmájában lokalizálódva lecsökkenti a szabad dNTP koncentrációt, ezzel megakadályozza a reverz transzkripciót. A HIV-2 genomjában található *vpx* terméke a SAMHD1 fehérjét proteaszómális degradációra tereli. (Laguet & Benkirane, 2012)

1.3. A HIV patogeneze

A fertőzés lefolyása akut, krónikus és szimptomatikus (AIDS) szakaszokra bontható (3. ábra). Az akut szakasz a transzmisszió után néhány hétig tart, rendkívül magas (akár 10 millió RNS kópia/ml) vérbeli vírusszint és a CD4+ T-sejtek jelentős pusztulása jellemzi. A krónikus fázis kezdetére a plazma vírusszintje csökken és beáll egy stabil értékre. Ez az úgynevezett setpoint vírusszint függ mind a vírus, mind a gazda genotípusától (Prince, 2012; Touloumi, 2013; Fellay, 2009). Az akut fázis után a CD4+ sejtszám szintén emelkedik, de soha nem éri el a kezdeti értéket. A vérben mérhető vírusszint fontos mutatója a betegségprogresszió várható sebességének, míg a vérbeli CD4+ T-limfociták számából a betegség stádiumára lehet következtetni. Ezek az értékek kezelés nélkül a krónikus fázis 8-12 éve alatt nagyon lassan romlanak. A krónikus fázis lassan növekvő vírusszintje egy folyamatos állóháborút rejt az gazda immunrendszere és a vírus között. Csupán a vírus replikációjának gátlásával a vírusszint két hét alatt a detekciós küszöb alá esik, felfedve a szervezetben belüli replikáció óriási mértékét (Perelson, 1996). Mikor a vérbeli CD4+ sejtszám 200 sejt/ μ l alá esik a betegség eléri az AIDS fázist. A CD4+ limfociták száma egyre csökken, a vírusszint pedig ismét megugrik. Az immunrendszer ilyenkor már nem képes ellátni feladatát, megjelennek különböző opportunist fertőzések tünetei. Kezelés nélkül a fertőzés majdnem minden esetben halálos. A vírus patogeneziséről összefoglaló található a következő cikkekben: (Moir, 2011; Levy, 2009; Grossman, 2006; Rowland-Jones, 2004; Coffin & Swanstrom, 2013; Stevenson, 2003; Douek, 2003).



3. ábra. A HIV-fertőzés lefolyása alatt a CD4+ T-limfociták és a vérbeli vírusszint változásai. Szaggatott vonal: CD4+ T limfociták száma. Folytonos vonal: vérbeli vírusszint.

A vírus gazdáról gazdára nyálkahártyán keresztül, leggyakrabban szexuális kontaktus útján terjed. Egyéb jelentős átviteli mód a vertikális anya-gyermek transzmisszió, fertőzött fecskendők és fertőzött vérkészítmények használata. A heteroszexuális kontaktus esetén a fertőzés valószínűségét 1:10 és 1:2000 közé becsülik (Shaw & Hunter, 2012), az átlagos érték 1:1000 a krónikus szakaszban. A férfiak körülmetélése jelentősen csökkenti a transzmisszió esélyét, míg a genitáliákat érintő gyulladással fertőzések növelik azt (Varghese, 2002). Az akut szakasz virémiájának következtében a szexuális transzmisszió esélye ebben a szakaszban jóval nagyobb, mint a krónikus fázisban (Shaw & Hunter, 2012).

A vírus az ép nyálkahártyán valószínűleg dendritikus sejtek segítségével jut át, amelyek eljuttatják azt a fertőzésre fogékony aktivált vagy részlegesen aktivált CD4+ T-sejtekhez (Miller & Shattock, 2003; Wu & KewalRamani, 2006; de Witte, 2008, 2007). Eleinte a nyálkahártyával asszociált nyirokszövetben sejtről sejtre terjed a vírus, majd eljut a nyirokcsomókba, ahonnan szétterjed az egész szervezetbe. Az akut vírusprodukciónak fő forrása a bélfallal asszociált nyirokszövet (gut associated lymphoid tissue - GALT), ahol a produktív fertőzésre különösen fogékony CD4+ CCR5+ T-sejtek nagy számban találhatóak (Kelleher & Zaunders, 2006). A hatékony vírusreplikáció következtében a GALT CD4+ populációja igen jelentős és irreverzibilis veszteséget szenved el. Ebben szerepet játszik a fertőzött sejtekkel szomszédos sejtekben indukált ún. bystander apoptózis (Chase, 2006; Finkel, 1995). Nyugalomban lévő CD4+ T-sejteket a vírus nem képes produktív módon megfertőzni, azonban működésképtelen virális DNS felhalmozódhat a sejtben, amely apoptózishoz vezet (Doitsh, 2010). A fertőzött sejtekre kifejtett

közvetlen vagy közvetett citotoxikus hatás jelentőségéről megoszlanak a vélemények, de *in vitro* kultúrákban a vírus elpusztítja a sejteket (Coffin & Swanstrom, 2013).

Az esetek többségében egyetlen egy virion létesít fertőzést a szervezetben (Keele, 2008). Ez egy igen erős alapítóhatás, az alapító vírusok szelektálódnak azokra a tulajdonságokra, amelyek szükségesek a gazdák közötti transzmisszióhoz (Moir, 2011; Keele, 2008). Ezek a korai változatok szinte minden esetben a CCR5 koreceptorhoz kötődve fertőzik meg a célsejtjeiket. Egyes vizsgálatok szerint a korai alapító vírustörzsek igen érzékenyek a neutralizáló antitestekre (ellentétben a későbbi változatokkal), de ezt nem minden tanulmány tudta reprodukálni (Derdeyn, 2004; Shaw & Hunter, 2012). Körülbelül 3 héttel a fertőzés után megjelenik a specifikus humorális és sejtes immunválasz is. Ezek jelentős szelekciós nyomást fejtenek ki a vírusra, de ez az immunválasz nem elég hatékony ahhoz hogy felvegye a harcot a nagyon gyors evolúcióra és masszív replikációra képes vírussal (McMichael, 2010).

A hatékony antiretrovirális terápia se képes teljesen meggyógyítani a betegeket, a kezelés felfüggesztését követően néhány héttel a vírusszint a kezelés előtti mértékre emelkedik. Ez arra utal, hogy a kezelés alatt is marad a szervezetben aktív, vagy aktiválható vírus (Bukrinsky, 1991; Grossman, 1998). A vírus kis mértékben szaporodhat olyan szövetekben, amelyeket a kezelés nem, vagy csak részlegesen ér el. Ezenkívül a vírus provírus formájában lappanghat hosszú életű nyugvó memória CD4+ sejtekben, amelyek időről időre aszimmetrikus osztódáson esnek át (vagy egyéb, nem azonosított kompartmentekben, amelyeket összességében látens rezervoárnak hív a szakirodalom). Ezek a rezervoárok a fertőzés kezdeti akut fázisában létrejönnek, általában a diagnózis felállítása előtt (Cohen, 1995; Trono, 2010; Trono & Marzetta, 2011).

A krónikus szakaszban már nemcsak a CD4+ T-sejtpopuláció sérül, hanem a CD8+ T-sejtek, az NK-sejtek és a B-sejtek működésében is zavarok jelentkeznek (Moir, 2011). Az immunrendszer különböző funkcionális és anatómiai kompartmentjeibe tartozó sejtpopulációk gyakoribb osztódást (megnövekedett turnover), fokozott aktivációt és túlzott differenciációt mutatnak. Hosszú távon ezek a hiperaktivációra utaló tünetek vezetnek az immunrendszer kimerüléséhez, valamint a sejtek megújulási képességeinek az elvesztéséhez (Biancotto, 2007; Nilsson, 2007). A vírus közvetett hatására utal egyrészt krónikus immunaktiváció globális jellege, másrészt a betegségprogresszió nagyon lassú időskálája (Yates, 2007). A hiperaktiváció oka nem teljesen tisztázott, egy megalapozott hipotézis szerint a bélfalán áthatoló mikrobiális anyagcseretermékek vagy alkotóelemek (pl.: lipopoliszacharid) okozzák (Pandrea, 2008; Brenchley & Price,

2006). A GALT-ban történt CD4+ sejtvesztés következtében ugyanis a bél immunrendszere jelentősen veszít védelmi funkciójából, nem képes megakadályozni a baktériumok (illetve azok anyagcseretermékeinek) áthatolását bélfalon (Brenchley & Douek, 2008). A krónikus immunaktiváció következményei az immunszövet és a nyirokcsomók szerkezetében fellépő hegesedések, amelyek megakadályozzák a limfociták megfelelő érését (Cohen, 1995), ezáltal korlátozva a limfocitapopulációk regenerációját. A kimerült limfociták hajlamosak az apoptózisra, illetve gyorsabban haladnak az immunszénescencia felé (Appay, 2007). Ennek következtében a periféria és a nyirokcsomók közti közlekedésük nem megfelelő, ami hozzájárulhat ahhoz, hogy nem alakul ki megfelelően széles specificitású immunválasz a vírus ellen (Douek, 2003; Douek, 2003). Az immunaktiváció jelentőségét jelzi, hogy mértéke nagyon pontosan korrelál a betegségprogresszióval (Deeks, 2004; Hazenberg, 2003). Ugyanakkor nem egyértelmű, hogy a vírus számára evolúciósan adaptív értékű-e az immunaktiváció kiváltása (Bartha, 2008).

A limfociták számán és aktivációjukon kívül néhány kvalitatív jellemzőjük is fontos szerepet tölt be. A fertőzést kezelés nélkül kontrollálni képes emberek CD4+ T-sejtjei nagyobb mennyiségű és többféle interleukint képesek kiválasztani (IFN- γ -t mellett IL-2-t is) (Owen, 2010), míg effektoros CD8+ T-sejtjei gyorsabb expanzióra képesek (Hersperger, 2011).

Az ART elkezdése visszасzorítja az immunaktivációt és csökkenti a limfocitapopulációk turnover-t (a sejtek élettideje nő). A kezelés megindítása után az esetek többségében a CD4+ populáció növekedésnek indul, de sosem éri el az eredeti szintet (Arts & Hazuda, 2012). Ebben szerepe lehet a regeneratív potenciál elvesztésének, az immunszövetek károsodásának és a tímusz csökkent funkciójának (Douek, 2003).

A fertőzésre adott reakcióban igen széles változatosságot találunk az egyes emberek között. A HIV-fertőzöttek körülbelül 10%-a kezelés nélkül is képes együttélni a vírussal és több mint 10 évig nem fejlődik ki bennük AIDS (ún. „long term non-progressorok”) (Moir, 2011). Ez betudható egy gyenge (például nem funkcionális Nef génnel rendelkező) vírussal, de az esetek többségében a beteg genetikai hátterében kell az okokat keresni. A páciensek 1%-ban a vírus-szint a detekciós küszöb alatt marad, ezt a csoportot „elite controllereknek” hívják. Bennük a CD4+ sejtszám normális szinten marad és erős (elsősorban sejtes) HIV-specifikus immunválasz figyelhető meg (Moir, 2011). Az elite controllerek és a long term non-progressorok között jelentős, de nem teljes átfedés van. Egyesek a magas kockázat ellenére sem fertőződnek meg a vírussal („exposed uninfected”), míg a betegek egy csoportja az átlagosnál érzékenyebb a ví-

rusra („fast progressors”). Ezen variációk humán genetikai polimorfizmusok meglétét sugallják, amelyek bizonyos allélikus konfigurációkban ellenállóbbá teszik a szervezetet a vírus ellen (van Manen, 2012). Ilyen polimorfizmusok azonosítása döntő lehet a patogenezis jobb megértése, és a járvány leküzdése céljából.

1.4. A HIV-vírus evolúciója

Retrovírus lévén a HIV örökítőanyagának másolása korántsem pontos, a mutációs ráta igen nagy (Mansky, 1998). Ezt összevetve az igen rövid generációs idővel, nagy szervezeten belüli kópiaszámmal és szaporodási rátával megérthető, hogy a vírus evolúciós hajlama kiemelkedően magas (Ribeiro, 2010; Perelson, 1996). Valóban, a HIV-vírus az egyik leggyorsabb evolúciót mutató ismert organizmus (Rambaut, 2004; Coffin & Swanstrom, 2013). Ez a gyors evolúció megfigyelhető mind egy adott páciensen belül, a fertőzés lefolyása során (Leslie, 2004), mind pedig epidemiológiai léptékekben gazdák között is (Rambaut, 2004).

Ezt kihasználva a vírus könnyen alkalmazkodik a különböző környezeti hatásokhoz, így a gazdák különböző immunológiai környezetéhez is. Jól megfigyelhető mutációkat váltanak ki a különböző humán HLA allélok a vírus szekvenciájában (Kawashima, 2009). A humorális immunválasz folyamatosan változó célpontok elleni szelekciója markáns nyomot hagy a szervezeten belüli víruspopuláció filogenetikai fáján a hosszú, irányított szelekcióra utaló ághosszakban (Rambaut, 2004). A transzmittált vírustörzs felszíni gp120 fehérjéje érzékeny az antitestekre, azonban az immunválaszt rövid időn belül követik az ellenálló vírusváltozatok, amelyek külső felszínén a kiterjedt cukor-oldalláncok akadályozzák meg az ellenanyagok kötődését a fehérjéhez (Wei, 2003). Továbbá, a vírus könnyen alkalmazkodik a gyógyszerhatóanyagokhoz, monoterápia alkalmazása esetén rövid idő alatt megjelennek a rezisztens törzsek (3TC esetén akár hetek alatt (Gallant, 2008)). Érdekes és fontos megfigyelés, hogy teljesen szuppresszív terápia alatt a vírusszekvenciákban evolúciós változások nem figyelhetők meg, a szervezeten belüli evolúció leáll, ami a vírusreplikáció teljes blokádjára utal (Ruff, 2002; Tobin, 2005). Drogrezisztencia mutációk ennek ellenére megjelenhetnek részlegesen szuppresszív terápia alatt (Tobin, 2005), vagy a rezisztens törzs kialakulhat már a kezelés megkezdése előtt (Hamers, 2010). A vírus szervezeten belüli diverzitása igen nagy (Yin, 2012). A szervezeten belüli evolúció során az immunrendszer, illetve a kezelés hatására felhalmozott mutációk új gazdába oltva revertálhatnak, adaptálódva az újonnan megváltozott környezethez (Leslie, 2004).

Fontos megjegyezni, hogy a HIV képes rekombinációra. Maga a HIV-1 őse, a csimpánzokat fertőző SIVcpz is két különböző nem ember főemlősöket fertőző SIV vírus rekombinánsa. A HIV-1 M csoportban megtalálunk mind altípusokon belüli, mind azok közötti rekombináns formákat (pl.: a jelentős elterjedéssel bíró cirkuláló rekombináns formákat (CRF)). Különböző in vitro és populációgenetikai vizsgálatok a rekombináció gyakoriságát a teljes genomra vonatkoztatva replikációs ciklusonként 1%-ra becslik (Neher & Leitner, 2010; Batorsky, 2011; Levy, 2004). A törzsek közötti rekombináció alapvető feltétele a sejt (és ezzel az egyén) kettős fertőzése, illetve a reverz transzkriptáz enzim képessége a szálak közötti váltásra. A lefűződő virionba az RNS örökítőanyag két kópiája csomagolódik be, megkönnyítve a rekombináns formák létrejöttét. A fertőzött sejtek genomjába integrált provírusok számát eleinte igen nagyra becsülték (medián 3,2 kópia / sejt), az újabb becslések szerint azonban a sejtek többsége csupán 1 kópiát hordoz (Jung, 2002; Josefsson, 2011). A rekombináció jelentős előnyöket biztosít a vírus számára, csakúgy, mint az összes szexuális úton szaporodó élőlény esetében. Két független törzs rekombinációja a többszörös escape mutációt hordozó haplotípusok (mind az immunrendszer, mint a gyógyszerek esetében) gyorsabb megjelenéséhez vezethet, illetve a fitnesszt csökkentő mutációk kivágása révén megakadályozza a nagy mutációs rátával és aszexuális módon szaporodó populációk „genetikai elszorodását” (Simon-Loriere & Holmes, 2011).

A HIV-vírus biológiájának általános összefoglalója után ismertetem a genetikai asszociációvizsgálatok módszerét.

1.5. Genetikai asszociációvizsgálat

A genetikai asszociációvizsgálatok célja olyan genetikai elemek azonosítása, amelyek variációja összefügghet egy betegség, vagy egyéb, akár kvantitatív fenotípus meglétével. Például, ha a vizsgálatba bevont betegek körében egy adott allél szignifikánsan nagyobb gyakoriságot mutat, mint a megfelelően kiválasztott kontrollcsoportban, akkor azt mondhatjuk, hogy az adott allél asszociál a betegséggel. Ilyen vizsgálat történhet egy család, vagy egy adott populáció egymással kevésbé rokon egyedeinek a bevonásával (Cichon, 2009). Vizsgálhatunk egyedi géneket, amelyekről más forrásból sejtjük, hogy szerepe van az adott fenotípus kialakításában (ún. candidate gene studies) vagy különböző mértékben haladhatunk a teljes genom lefedése felé (Abecasis, 2012).

Az angol szakirodalomban genome wide association study (GWAS) névvel illetik azt a genetikai asszociációvizsgálatot, amely egy populáció egymással nem rokon egyedein vizsgálja a populációban gyakran (>5%) előforduló, egy nukleotid hosszú genetikai változatokat (single nucleotide polymorphism, SNP). (Balding, 2006)

Az európai populációban körülbelül 50 kilobázisos blokkokban találunk SNP-eket amelyek egymással erős kapcsoltásban állnak (Gabriel, 2002). Ennek megfelelően elegendő a vizsgálatba bevont emberekben kevesebb (maximum egy millió) gyakori SNP-t genotipizálni ahhoz, hogy lefedjük a teljes genomban található változatosságot (Wall & Pritchard, 2003). Ha a betegséget kiváltó funkcionális, okozati (az angol szakirodalomban többnyire „causal”) változatot nem is genotipizáltuk, egy közelben lévő SNP, a genetikai kapcsolttság erejéig, ahhoz hasonlóan fog szegregálódni a fenotípusos csoportjaink között. A genotipizálandó SNP-k kiválasztását segítő jött létre az International HapMap Project, amely 2002 és 2007 között négy emberi populáció (Nigéria, Japán, Kína, USA - nyugat-európai leszármazással) 270 tagjában 3,1 millió gyakori SNP-t genotipizált (Frazer, 2007; Altshuler, 2010). Ennek segítségével kiválasztható az a néhány százezer, egymillió SNP amely egy adott (általában nyugat-európai) populáció haplotípus szerkezetét elégségesen lefedi, és olcsó genotipizáló csipek készíthetők belőlük.

A betegségek genetikai hátteréről több modell is napvilágot látott (Gibson, 2011). A „gyakori betegség - gyakori mutáció” (common disease common variant, CDCV) hipotézis feltételezi, hogy a betegséget néhány, a populációban gyakori (ősi) polimorfizmus okozza. Egy mutáció felhalmozódhat a populációban azért mert neutrális vagy jótékony, esetleg a múltban jótékony hatással bírt, de megváltozott környezetünkben káros. A CDCV hipotézist az utóbbi években elvégzett GWA vizsgálatok nem támasztják alá teljes mértékben (Gibson, 2011), pontosabban egyes fenotípusok megfelelnek neki, mások pedig nem. Gyakori polimorfizmusok kétségkívül befolyásolják a vizsgált gyakori betegségek kockázatát. Több mint 3600 SNP-ről tudunk, amelyek bizonyíthatóan asszociálnak valamely fenotípussal (Hindorff, 2012). Egyes esetekben a GWA vizsgálatban azonosított SNP jól magyarázza a betegség mechanizmusát és heritabilitását (pl.: Komplement Faktor H génben azonosított SNP az időskori sárgafolt sorvadásban (age related macular degeneration) (Haines, 2005)). Ennek ellenére a legtöbb vizsgálatban az azonosított polimorfizmusok a fenotípus variációjának csupán töredékét (körülbelül 15%) magyarázzák meg, még akkor is, ha az adott fenotípus szűken értelmezett heritabilitása² magas (Manolio,

²Az additív genetikai variancia által megmagyarázott fenotípusos variancia hányada a teljes fenotípusos vari-

2009). Ezt az inkonzisztenciát az irodalom „missing heritability” (hiányzó heritabilitás) néven említi, feloldására több egymást nem kizáró elmélet született. Egyrészt, a GWAS során talált polimorfizmusok az esetek többségében nem közvetlenül felelősek a fenotípusért, hanem kapcsolatban állnak az okozati SNP-vel (causal variant), ezért a kapcsoltsági viszony korlátozza az általuk megmagyarázható fenotípusos varianciát. Másrészt, a jelenleg széles körben alkalmazott GWAS technológia strukturális változatokat (elsősorban indeleket³) nem érzékel, ezek a polimorfizmusok nincsenek reprezentálva a vizsgálatokban (Manolio, 2009). Felmerült, hogy a szűken vett heritabilitás becslése nem pontos, túlbecsli a valódi értéket, így a GWAS során megmagyarázott genetikai variancia azt nem is érheti el (Zuk, 2012). Továbbá, megfogalmazhatunk új modelleket a betegség allélikus struktúrájáról, mint a CDCV hipotézis alternatíváit. A betegséget okozhatja sok, egyenként kis hatású (genotypic relative risk⁴ <1,1) mutáció amelyek egyenletesen helyezkednek el a frekvenciaspektrumon (tehát egyaránt lehetnek gyakori, és ritka polimorfizmusok is). Ezek közül a GWAS csak a gyakori, és nagyobb hatású mutációkat képes azonosítani. Ez a modell lényegében megegyezik Fisher poligénes öröklődést leíró modelljével, amely infinitezimálisan kicsi hatások összegeként írja le a kvantitatív tulajdonságok öröklését (Visscher, 2008). Utóbbi elképzelést alátámasztja, hogy a GWAS során megmért SNP-k együttesen képesek megmagyarázni a magasság és néhány betegség (reumatoid arthritisz, miokardiális infarktus, kettes típusú cukorbetegség, skizofrénia) fenotípusos varianciájának döntő hányadát (Stahl, 2012; Purcell, 2009; Yang, 2010). Egy másik hipotézis a ritka polimorfizmusokra helyezi a hangsúlyt, ez azt feltételezi, hogy a fenotípus háttérben nagyszámú ritka variáns (evolúciósan fiatal, esetleg teljesen egyedi, privát mutáció) áll, amelyek hatása azonban jelentős (GRR >2) (Cirulli & Goldstein, 2010). A GWAS során talált polimorfizmusok ezekkel állhatnak kapcsolatban, így természetesen képtelenek a teljes varianciát megmagyarázni. Továbbá elképzelhető, hogy egyes bonyolult fenotípusok (elsősorban pszichiátriai megbetegedések) sok, egymástól genetikailag független, de hasonló tüneteket kiváltó állapotok együttese. Ez szintén lehetetlenné teszi a megmagyarázott genetikai variancia pontos becslését. Gibson felsorakoztat 20 érvet mind a ritka allélok modellje, mind az infinitezimális modell mellett és ellen, demonstrálva, hogy jelenlegi tudásunkkal nehéz dönteni köztük, és minden bizonnyal mindkét elképzeléshez képest.

³Egy vagy több nukleotid inszerciója vagy delécioja.

⁴genotypic relative risk (GRR). A mutáns allél egy példányát heterozigóta formában hordozó egyén esélye a betegségre osztva az allélt nem hordozó egyén esélyével. Recesszív öröklésment esetén a két allélt hordozó esélye osztva az egy vagy 0 allélt hordozó esélyével.

lés részlegesen helyes (Gibson, 2011). Egy harmadik út a „missing heritability” problémájának megoldására a környezeti hatások, epigenetikus hatások, és génkölsönhatások (episztázis) figyelembevétele. Ezt a szakirodalom „broad sense heritability model” néven említi, de mivel az említett hatások nincsenek beleszámolva a szűken vett heritabilitásba, ezért szigorúan véve nem is részei a „missing heritability” problémájának. Ettől függetlenül, minden bizonnyal igen jelentős szerepet játszanak a betegségek genetikai hátterében (Gibson, 2011).

A mutációkat gyakoriságuk és hatásuk erőssége szerint sorolhatjuk csoportokba. A ritka (minor allele frequency ⁵ <1%), de nagy penetranciával⁶ bíró genetikai variációk felelősek a mendeli öröklődésű betegségekért. A nagy hatásméret miatt ezeket könnyű felismerni. Hasonlóan ritka, de kis vagy közepes hatással bíró mutációk felderítése ezzel ellentétben szinte lehetetlen, mivel túl nagy mintaméret szükséges egy megfelelő erejű statisztikai teszt kivitelezéséhez. A GWAS a frekvenciaspektrum másik végére koncentrál: az 5%-nál gyakoribb, de kis vagy közepes hatással bíró (GRR 1.2-2) SNP-k hatását megfelelő erővel ki lehet mutatni gyakorlatban kivitelezhető mintaméretek mellett is (ezer főtől néhány tízezer főig). (McCarthy, 2008; Bush & Moore, 2012)

Ahhoz, hogy egy új genetikai faktor felfedezését biztosan kijelenthessük, biztosítanunk kell, hogy a hatás, amelyet látunk valódi, és nem hamis. Ez egy igen súlyos probléma a GWA vizsgálatok során, hiszen több millió tesztet is elvégezhetünk. A szakma a következő konvencionális kritériumokat állította fel egy GWAS találat megbízhatóságának ellenőrzésére. A teljes kísérletnek meg kell felelnie a hagyományos frekventista hipotézisvizsgálat követelményeinek 5%-os szignifikanciaszinten. Ez Bonferroni-korrekciónál alkalmazva rendszerint $p = 5 \times 10^{-8}$ értékben határozza meg az egyes tesztek szignifikanciaszintjét (Balding, 2006). Ezenkívül az eredeti felfedezéstől független mintában meg kell ismételni a vizsgálatot a kérdéses markerre, és a hatásirányoknak egyeznie kell. A replikációs kísérletben természetesen szintén korrigálni kell, amennyiben több teszt került elvégzésre. Ha egy felfedezés ezeknek a követelményeknek nem felel meg, akkor érdemes azt különös körültekintéssel kezelni.

Egy helyesen elvégzett genetikai vizsgálat során minden kritériumnak eleget tevő felfedezés sem bizonyítja az érintett gének oki hatását. Egyrészt elképzelhető, hogy az adott SNP csupán kapcsolatban áll a valódi hatást kiváltó polimorfizmussal, másrészt a statisztikai eredményt

⁵Biallélikus polimorfizmust feltételezve a ritkább allél gyakorisága, angolul minor allele frequency, a továbbiakban MAF.

⁶A mutáns allélt hordozók között a mutáns fenotípus aránya.

minden esetben funkcionális vizsgálatoknak kell megerősítenie.

A vizsgált fenotípus lehet kettő (beteg - nem beteg, ún. case-control) vagy több értéket felvevő kategorikus változó, illetve kvantitatív érték. A fenotípus megfelelően körülhatárolt módon való definiálása, a mérés pontossága és torzítatlansága alapvetően meghatározza a vizsgálat eredményét (Balding, 2006; Bush & Moore, 2012).

A vizsgálatba bevont egyéneket úgy kell kiválasztani, hogy egy homogén populációt alkotósanak, és ne legyenek rejtett különbségek a minta részhalmazai között (McCarthy, 2008; Balding, 2006). Például case-control vizsgálat esetén a kontrollcsoportot a vizsgálati csoporthoz illesztve választjuk ki. Fontos figyelni a mintában szereplő emberek leszármazására, ugyanis a minta genetikai homogenitására törekszünk. Ha ugyanis a mintán belül egymással közelebbi rokonságban levő csoport van, és ezt nem vesszük figyelembe a statisztikai elemzésnél, akkor hibásan pozitív SNP-ket kaphatunk. Ha a beteg fenotípusú emberek felülreprezentálnak egy genetikailag közel rokon csoportot, akkor minden olyan SNP asszociálni fog a betegséggel, amelynek alléljai különbözőképpen szegregálnak ezen csoport és a teljes minta között. Ezt a hatást elkerülhetjük, ha megbizonyosodunk a minta homogenitásáról, vagy figyelembe vesszük a minta struktúráját az elemzés során. Mindkettőre alkalmas az úgynevezett eigenstrat módszer, amely főkomponens analízis segítségével deríti fel az adatokban rejlő struktúrákat (Patterson, 2006).

Mivel az SNP-k döntő hányada nem játszik szerepet a vizsgált fenotípus genetikai hátterében, így teljes genomi információ felhasználásával készült főkomponensek nem oltják ki az esetlegesen asszociáló polimorfizmusok hatását a regresszió során.

Magát az asszociációvizsgálatot, azaz az SNP és a fenotípus közötti kapcsolat felderítését végezhetjük lineáris vagy logisztikus regresszióval, illetve kontingenciatáblán alapuló tesztekkel (Pearson χ^2 -teszt vagy Fisher-féle egzakt teszt) (Balding, 2006; Bush & Moore, 2012). A regresszióval végzett tesztek esetén feltételezhetjük az additív genetikai modellt, tehát az adott lókuszon található alléldózis (0, 1 vagy 2 kópia található-e az egyik önkényesen kiválasztott allélból) és a fenotípus közötti kapcsolatot vizsgáljuk. ANOVA-val a genotípusosztályok hatásait külön-külön vagy összevonva vizsgálhatjuk. A tesztet általában minden SNP-re külön-külön végezzük el, azaz minden egyes SNP hatását önmagában vizsgáljuk. Episztázis fennállása esetén a próba erősebb, ha több SNP-t együtt tesztelünk, például interakciós tagokat téve a regresszióba (Gyenesei, 2012).

1.6. A HIV-vírussal kapcsolatos humán genetikai faktorok irodalmának áttekintése

Az irodalomban fellelhető tanulmányok ezidáig összesen 1350 GWA vizsgálatról számolnak be (Hindorff, 2012). Ebben a fejezetben megkísérlem összefoglalni a HIV-fertőzésre való hajlamot érintő vizsgálatokat, illetve az AIDS-betegségre hatást gyakorló humán genetikai polimorfizmusokról szóló eredményeket.

A HIV-fertőzés esetén a betegséggel legerősebben a hatodik kromoszóma 28. és 33. megabázisa között elhelyezkedő régió mutat asszociációkat (Stephens, 2012; Goulder & Watkins, 2008). Itt találhatóak az adaptív immunrendszerben kulcsszerepet betöltő *MHC-I* gének⁷. Az epitópok prezentálásának hatékonysága jelentős mértékben befolyásolja a betegség későbbi kimenetelét. Különböző allélok különböző epitópokat prezentálnak a CD8+ citotoxikus T-sejtek felé, amelyek így felismerhetik a fertőzött sejteket és elpusztíthatják azokat. Az emberi (és azt megelőző) evolúció során a sokféle kórokozó által kifejtett szelekciós nyomás következtében az MHC régió a humán genom legváltozatosabb szakaszává vált (Jeffery & Bangham, 2000). Az evolúciós nyomás hatására az epitópok bemutatásának egyre hatékonyabb és diverzebb formái alakultak ki az immunrendszerben. Ennek következtében minél sokrétűbb epitópokat képes prezentálni, illetve felismerni az adaptív immunrendszer, a populáció annál inkább védett a fertőző betegségek ellen.

Az előzőek alapján nem meglepő, hogy *HLA* gének homozigóta genotípusa gyorsabb betegségprogresszióval jár, mint a heterozigóta genotípusok (Carrington & O'Brien, 2003). A *HLA* gének alléljai közül egyesek jobban, mások rosszabbul képesek kötni a HIV epitópjait, ezzel befolyásolva a betegség kimenetelét. Például a *HLA-B*35* allél gyorsabb betegségprogresszióval párosul, míg a *B*57* és *B*27* allélek erősebb védelmet nyújtanak a vírus ellen (Goulder & Watkins, 2008; Carrington & O'Brien, 2003; Goulder, 2001). Utóbbi két allél gyakran megtalálható az elite controllerek körében. Lehetséges hatásmechanizmusként felmerült, hogy a *B*57* allél olyan epitópokat ismer fel, amik elől csak költséges mutációval tud a vírus kitérni (Leslie, 2004). Ezzel szemben a *B*27* allél a széles epitópspecificitásával képes leküzdeni a fertőzést (Kelleher, 2001).

⁷Major Histocompatibility Komplex - Az emberi MHC géneket Human Leukocyte Antigennek is nevezi az irodalom. Az irodalomban talált konvenciót követve a dolgozatban MHC régió néven a humán hatodik kromoszóma 28. és 33. megabázisa közötti részt értem, míg az egyes génekre HLA génként (pl.:*HLA-A*, *-B*). *MHC-I* géneken a *HLA-A*, *-B*, *-C*, *-E*, *-F* és *-G* géneket értem. Az *MHC-I* gének termékei felelősek a virális epitópok sejtfelszíni bemutatásáért a vírusfertőzött sejteken. A dolgozatban a kettes típusú MHC génekről nem esik szó.

Mind a HIV, mind más vírusok vagy daganatok csökkenthetik a fertőzött sejt felszínén expresszált MHC-I molekulák számát (Kerkau, 1989). A természetes ölősejtek (NK-sejtek) feladata kiszűrni azokat a sejteket, amelyeken nincs elég HLA kifejezve. Ez a funkció az NK-sejtek KIR receptorainak és a célsejt HLA-I molekuláinak a kölcsönhatásán alapszik. A *KIR* gén *3DS1* allélja a *HLA-Bw4* allélcsoport tagjaival együtt episztatikus hatásban növeli a betegségprogressziót. Ugyanez a *3DS1* allél más *HLA-B* allélokkal együtt pedig véd a fertőzés ellen. (Goulder & Watkins, 2008)

A vírusfertőzés feltétele a virális burokkészlet és bizonyos sejtfelszíni kemokinreceptorok egymáshoz való kapcsolódása. A nyálkahártyán átjutó vírusok általában a CCR5 kemokinreceptorhoz kötődnek, míg később a gazdán belüli evolúció során kialakulhat egy CXCR4 tropikus változat is. A CCR5 receptor és ligandumaiban bekövetkező mutációknak fontos szerepük van a HIV-fertőzésre adott válaszban. A nyugat-európai és észak-európai populációban 1%-os gyakorisággal előforduló homozigóta CCR5 Δ 32 genotípus majdnem teljes védelmet biztosít az R5 tropikus vírus ellen (Dean, 1996; Novembre, 2005), míg a mutáció heterozigóta formában nem nyújt védelmet a fertőzés ellen, de a setpoint vírusszintet csökkenti (Walli, 1998). A *CCR2* gén V64I mutációja az előbbitől függetlenül lassítja a betegségprogressziót (Smith, 1997).

Az első GWAS kísérletet, amely a HIV-vel szembeni tűrőképességet vizsgálta Fellay és munkatársai végezték 2007-ben (Fellay, 2007). Ebben fenotípusként 486 európai származású beteg setpoint vírusszintjét, illetve CD4+ T-sejtszámát vizsgálták. A vizsgálat azonosított 2 SNP-t, amely a vírusszinttel, valamint egy SNP-t, amely a CD4+ sejtek számával asszociál, mindhárom az MHC régióban található. Ezeket az eredményeket később más csoportok független mintán is megerősítették, illetve a szerzők később 2300 páciensen megismételték a vizsgálatot (Fellay, 2009). A vírusszinttel asszociált SNP-k közül az erősebb (rs2395029) a *HCP5* génben található, de az okozati hatást nem ez a gén okozza. Az érintett SNP nagyon erős kapcsoltságban áll a HLA B*5701 haplotípussal. A 2009-es vizsgálat kilenc páciensben megfigyelt egy törést a *HCP5* gén és *HLA-B* gén között. Továbbá a *HLA-B*57:01* allélt hordozó betegek vírusszintjei alacsonynak bizonyultak az SNP minor allélja nélkül is. Ez azt bizonyítja, hogy az okozati mutáció a HLA-B génben található. Ezt később megerősítették (lásd lent). A másik vírusszinttel asszociált SNP a *HLA-C* gén promótere előtt 35kb-sal található. Funkciójának felderítése bonyolult, egyrészt az SNP asszociál a gén expressziós szintjével (egy közeli miRNS kötőhelyen keresztül), másrészt kapcsoltságban áll mind *HLA-B*, mind *HLA-C* haplotípusok-

kal, így az SNP jelölheti ezen haplotípusok hatását is. A polimorfizmus kapcsoltágban áll egy miRNS kötőhellyel, amely a *HLA-C* gén expressziós szintjét szabályozza, és a *HLA-C* fehérje szintje korrelál a betegségprogresszióval (Apps, 2013). A CD4+-sejtszámmal asszociált SNP a *ZNRD1* génben található, a hatás legalább két független kísérletben is mutatkozott, de bővebbet nem tudunk róla.

Egy további nagy mintán elvégzett GWAS a fenti eredményeket megerősítette, és továbblépve részleteire bontotta a *HLA-B* génen belül tapasztalt hatásokat (Pereyra, 2010). Az International HIV Controllers Study GWA vizsgálata 2648 progresszor páciens és 974 kontroller páciens genomjait hasonlította össze egy case-control vizsgálatban (Pereyra, 2010). Az SNP-ken elvégzett teszteken felül, vizsgálták a klasszikus *HLA*-allélek hatásait, és a *HLA* fehérjék aminosavjainak a hatásait. Miután a genotipizált SNP-k variációiból levezették egy független minta segítségével a vizsgálatba bevont egyének *HLA* haplotípusait, és a *HLA* géneken belüli polimorf aminosavak állapotát, lehetővé vált ezek részletesebb vizsgálata. A *HLA-B* fehérje 67., 70., és 97. pozíciójában lévő aminosavak variációja erősebb hatásnak bizonyult mind az SNP-nél, mind a klasszikus haplotípusoknál. Utóbbi haplotípusokat hozzárendelték a különböző pozícióban levő aminosav oldalláncok kombinációhoz. Mind a három pozíció a fehérje peptidkötő mélyedésnek az alján található, erősen valószínűsítve, hogy ezek a variációk ténylegesen befolyásolják az epitóp-*HLA* kötődés erősségét, illetve minőségét.

Az itt említett genetikai faktorokon kívül az említett két tanulmány további, ezektől független polimorfizmusokat (a *MICA*, *PSORS1C3*, *ZNRD1*, *TRIM10*, *NOTCH4* génekben) azonosított, de ezeket nem erősítették meg független vizsgálatok (Pereyra, 2010; Fellay, 2009, 2007).

Ezidáig a legnagyobb HIV-víruson végzett genetikai asszociációvizsgálatot McLaren és munkatársai publikálták (McLaren, 2013). Ebben a 6300 HIV-fertőzött páciens és 7200 HIV-negatív kontroll gyakori genetikai variációit hasonlították össze, olyan faktorok után kutatva, amelyek befolyásolhatják a fertőzésre való hajlamot. Nem találtak a fenotípussal szignifikánsan asszociáló humán genetikai markert.

Más publikált vizsgálatok is történtek az előzőeknél kisebb mintán, és más fenotípust vizsgálva. Egyes szerzők az extrém fenotípust mutató pácienseket (pl.: gyors progresszorok, LTNP páciensek, vagy elit kontrollerek) hasonlítottak össze case-control vizsgálatban. Ezek a vizsgálatok megerősítették az MHC régió hatását, és egyes esetekben az MHC régió kívül eső szignifikáns asszociációkat is publikáltak (Limou, 2009; Le Clerc, 2009; Dalmasso, 2008; Troyer,

2011; Ballana, 2012; Limou, 2012). Későbbi tanulmányoknak azonban az MHC régió kívüli hatásokat, a *CCR* gének kivételével, nem sikerült reprodukálni.

Fontos megemlíteni a candidate gene vizsgálatok eredményeit, és azok utóéletét. Ezek a vizsgálatok a GWAS előtti időszakban több mint 100 gént (pl.: adaptív és örökölt immunrendszer génjei, antivirális restrikciós faktorok, vagy HIV replikációhoz szükséges gének) hoztak összefüggésbe a vírussal (köztük a már említett *CCR5* polimorfizmusokat). Utóbbi kivételével egyiket sem sikerült reprodukálni (Fellay & McLaren, 2013). Mind a 2009-es Fellay, mind a 2010-es International HIV Controllers Study által végzett nagy elemszámú vizsgálatok megpróbálták validálni ezen variánsokat, de a tesztek nem érték el a szignifikanciaszintet (Pereyra, 2010; Fellay, 2009). A korai candidate gene vizsgálatok félrevezető eredményének az oka lehet a populációstruktúra figyelmen kívül hagyása, a többszörös tesztre végzett korrekció elhagyása vagy enyhítése, és a kicsi mintaméret (Fellay & McLaren, 2013).

2. A humán genom és a HIV-1 genom variációinak együttes asszociációvizsgálata

A gazdaszervezet és a patogénje közti kapcsolat evolúciós változást indít el mindkét fél genomjában. Az emberi és nem emberi főemlősök genetikai állományában a retrovírusok által kifejtett szelekciós nyomás jól felismerhető lenyomatokat hagyott: a *TRIM5 α* , az *APOBEC3G* és a *tetherin* (Bst-2) antiretrovirális faktorok mind a legerősebb szelekciós szignált tartalmazó gének közé tartoznak (Ortiz, 2009). A HIV-vírus különleges abból a szempontból, hogy a genomjában zajló evolúciós változások nemcsak gazdaváltások során, hanem egy gazdán belül is megfigyelhetők, és a divergencia sebessége egy gazdában gyorsabb, mint a gazdák közötti evolúciós során (Lythgoe & Fraser, 2012). Így különös figyelmet érdemelnek a szervezeten belüli evolúció során bekövetkező drogrezisztencia-mutációk, és az *MHC-I* gének által bemutatott epitópok escape mutációi (Goulder, 2001; Kawashima, 2009). Ezeken kívül, az utóbbi években a természetes ölüsejteken kifejeződő killer immunoglobulin-like receptor (KIR) molekulák által célbavett virális szekvenciaszakaszokon is hasonló, szervezeten belüli adaptív változásokat írtak le (Alter, 2011). Főemlős modellekben pedig a *TRIM5 α* szelekciós hatása figyelhető meg a SIV-vírus *capsid* génjére (Kirmaier, 2010), azonban a humán *TRIM5 α* gén nem vezet escape mutációkhoz, ami valószínűleg a patogén adaptációjának a következménye (Rahm, 2013). Gazda és patogén genetikai adatokat együttesen felhasználva máig csupán a *HLA*- és a *KIR*-gének asszociációvizsgálatát végezték el (Kawashima, 2009; Moore, 2002; Bhattacharya, 2007; Alter, 2011), a teljes emberi genomot feltáró asszociációvizsgálatra mindezidáig nem került sor.

A rendelkezésünkre álló a teljes genomot lefedő humán genotípusadatokat és HIV-szekvenciákat felhasználva a doktori munkám során munkatársaimmal arra kerestük a választ, hogy mely humán genetikai faktorok képesek evolúciós változásokat kiváltani a vírus genomjában, annak szervezeten belüli evolúciója során. Továbbá a párosított, ugyanabból az egyénből származó genotípusokat és majdnem teljes HIV-genomokat felhasználva lehetőségünk volt meghatározni a két szervezet genetikai kölcsönhatásának a lókuszeit.

páciensek száma	genotipizáló platform	kohorsz
140	Illumina 1M	ACTG
1	Illumina 1M duo	SHCS
1	Illumina 1M v1	SHCS
15	Illumina 550K v1	WAHCS
55	Illumina 550K v1	SHCS
24	Illumina 550K v3	SHCS
47	Illumina 650k	IHCS
61	Illumina OmniExpress12v1H	HOMER
6	Illumina OmniExpress12v1H	MADRID
122	Illumina OmniExpress12v1H	SHCS
518	Affymetrix 6.0	HOMER
990		

1. táblázat. A genetikai asszociációvizsgálathoz rendelkezésünkre álló páciensek számai kohorsz és genotipizáló csipek szerint lebontva. ACTG = AIDS Clinical Trials Group Network; HOMER = HAART Observational Medical Evaluation and Research Study; IHCS = International HIV Controllers Study; SHCS = Swiss HIV Cohort Study; WAHCS = Western Australian HIV Cohort Study

2.1. Eredmények

2.1.1. Rendelkezésre álló adatok

Humán genotípus adatok és HIV-szekvenciák álltak rendelkezésünkre 990 (56 nő, 340 férfi, 594 ismeretlen nemű) krónikusan HIV-fertőzött, kezelés alatt nem álló, nyugat-európai leszármazással bíró páciensből. A vizsgált személyek hat különböző földrajzi helyről származtak, és a genotipizálásuk is különböző laborokban, különböző genotipizáló platformokon történt. A páciensek kohorszönkénti és genotipizáló csipenkénti eloszlását az 1. táblázat mutatja. A genotipizáló platformok diverzitása miatt az első minőségellenőrző lépéseket platformonként külön-külön végeztük el, összesen négy csoportban (Affymetrix 6.0, Illumina 1M, Illumina OmniExpress, illetve összevontuk egy csoportba az Illumina 550K és 650K csipeket - utóbbi csoportra Illumina 500K néven fogok hivatkozni).

Ezeket a csoportokat külön-külön elvégeztük az SNP-k alapvető ellenőrzését. Elvetettük azon SNP-eket, amelyek nem voltak Hardy-Weinberg egyensúlyban, amelyek ritka allél frekvenciája kisebb volt, mint 5%, és amelyek genotipizálása nem sikerült a páciensek több mint 1%-ban. Továbbá minden platformon χ^2 próbával összehasonlítottuk az SNP-k allélfrekvenciáját a HapMap CEU populációban mérhető allélfrekvenciával, illetve a többi platformon mért

allélfrekvenciával. Azokat az SNP-eket figyelmen kívül hagytuk, amelyek allélfrekvenciái szignifikánsan eltértek valamely populációk között (részletesebben lásd alább).

A vizsgálatba bevont összes páciens nyugat-európai leszármazású volt, ezt platformonként főkomponens analízissel ellenőriztük. A 4. ábra mutatja a vizsgálati populáció és a HapMap populációk egymáshoz viszonyított helyzetét a genotípusmátrixból készült első két főkomponensen. A vizsgálatba bevont páciensek jól elkülönülten együtt csoportosulnak a nyugat-európai leszármazással bíró CEU populációval. Szintén főkomponens analízissel ellenőriztük a vizsgált populációk genetikai homogenitását. A 5. ábra mutatja a vizsgálati populációkon végzett főkomponens analízis eredményeit. A populációk genetikai varianciája kicsi, az első két főkomponens mindegyik platform esetén kevesebb mint 2%-ot magyaráz az összvarianciából.

A svájci kohorsz által követett páciensek esetében a vérbeli (logaritmikus skálán megadott) vírusszintmérések átlaga is rendelkezésre állt, amire a továbbiakban setpoint vírusszint néven fogok hivatkozni.

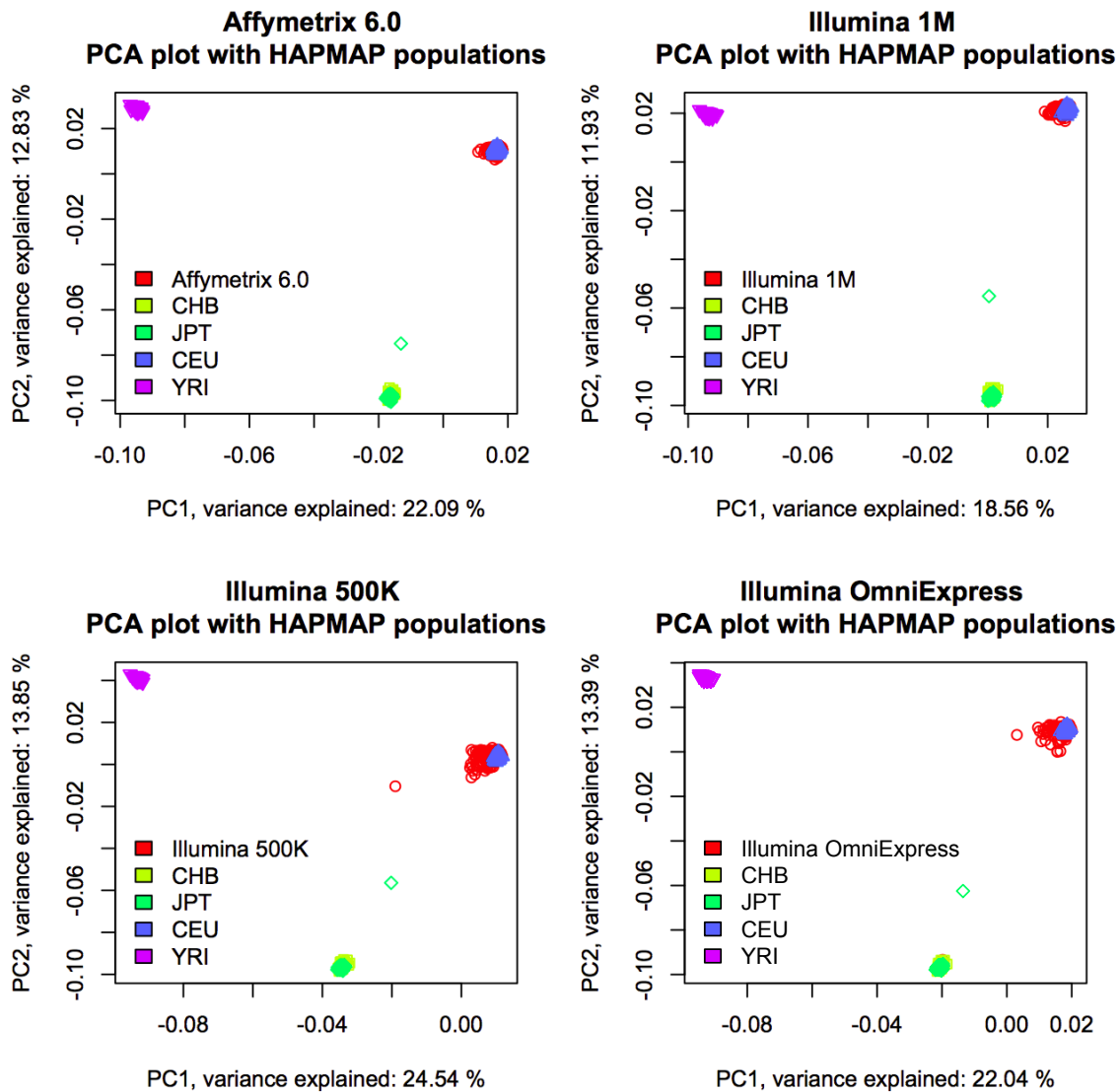
A vizsgálatba bevont páciensek mind B altípusú vírustörzsszel fertőződtek meg, ezt a Rega v2 szoftverrel ellenőriztük (Alcantara, 2009; de Oliveira, 2005). Populációs szekvenálással kinyert konszenzusszekvenciák álltak rendelkezésünkre egy kivételével a vírus összes fehérjéjéről. A HIV-vírus gp120 fehérjéje nagyon variábilis, szekvenálása technikailag nehézkes, ezért erről a régióról nem volt elegendő szekvenciaadat. Továbbá a Pol poliprotein Proteázt megelőző 55 kodon hosszúságú részéről⁸ se állt rendelkezésünkre nukleotidszekvencia. A maradék kilenc fehérjében (Gag, Pol, Gp41, Nef, Rev, Tat, Vpu, Vpr és Vif) a vizsgálat során 913 aminosavpozíciót találtunk variábilisnak legalább 20 páciensben (részletezve a 2. táblázatban). Ebben a 913 pozícióban összesen 1998 különböző aminosavat figyeltünk meg (lásd később).

2.1.2. Az adatok összefésülése

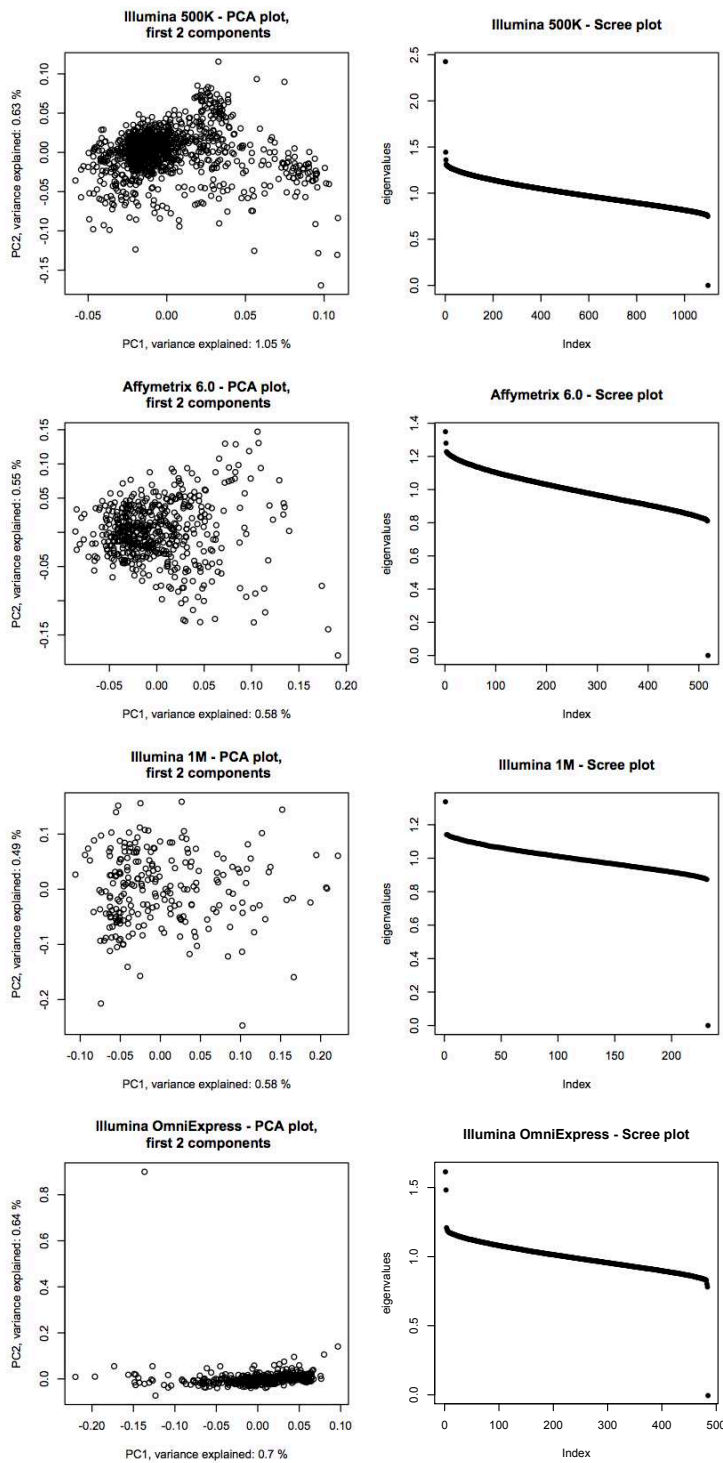
A különböző genotipizáló platformok más-más markereket tartalmaznak, és a közös markereket is esetleg torzítottan mérik. Ezért szükséges volt, hogy az eltérő forrásból származó adatokat egységesítsük, és feltárjuk a szisztematikus eltéréseket (ún. batch hatást⁹) a platformok között.

⁸A hxb2 törzsön ez a következő szekvencia: FFREDLAFLQ GKAREFSSEQ TRANSPTRRE LQVWGRDNNS PSEAGADRQG TVSFNF

⁹Batch hatás néven illetünk minden olyan mesterséges, az adatgenerálás vagy feldolgozás során létrejött szisztematikus hibát, amely specifikus a körülményekre (például labor, kohorsz, mintavételi módszer, feldolgozási módszer)



4. ábra. A vizsgálati személyek és a HapMap populációk közös genotípusmátrixán végzett főkomponens analízisek eredményei. A vizsgálati személyek pirossal, a háttérpopulációk egyéb színnel vannak jelölve. A grafikonokon az első két főkomponenst ábrázoltam egymás ellenében. Megfigyelhető, hogy a vizsgálati személyek közös csoportot alkotnak a HapMap CEU populációval. Populációk rövidítései: CHB - Peking, Kína; YRI - Ibadán, Nigéria; CEU - nyugat- és észak-európai leszármazással bíró lakosok az Egyesült Államokból; JPT - Tokió, Japán.



5. ábra. A genotípusmátrixokon elvégzett főkomponens analízisek eredményei. Balra az első két főkomponens egymás ellenében ábrázolva, jobbra a sajátértékek csökkenő sorrendben.

Fehérje	Variábilis pozíciók száma
gag	161
gp41	148
nef	103
pol	213
rev	65
tat	53
vif	88
vpr	33
vpu	49

2. táblázat. A legalább 20 mintában variábilisnak talált aminosav-pozíciók száma a HIV-vírus különböző fehérjeiben.

A mintákat platform szerint négy csoportba soroltuk: Affymetrix 6.0, Illumina OmniExpress, Illumina 1M és a maradék legalább 500 ezer, maximum 650 ezer markert tartalmazó Illumina platformok (lásd 1. táblázat). A több platformon is genotipizált SNP-k genotípuseloszlását páronként összehasonlítottuk a csoportok között χ^2 próbán alapuló függetlenségvizsgálattal. Elvetettük azon SNP-eket, amelyek esetében a genotípusok eloszlása nem volt független a platformtól. A tesztek szignifikanciaszintjének a beállításához Benjamini-Hochberg procedúrát alkalmaztunk 20%-s False Discovery Rate-el (Benjamini & Hochberg, 1995). A maradék markereken a mach szoftver (Li, 2010) segítségével négy csoportban genotípus imputációt végeztünk HapMap Phase 3 referencia-haplotípusokat használva. Az imputáció eredményét a valós genotípusok és a hozzárendelt (imputed) alléldózisok szoftver által becsült korrelációjával ellenőriztük. Elvetettük azokat a markereket amelyek esetében a korrelációs együttható 0,3-nél kisebb volt. Utóbbi érték a mach szoftver dokumentációjában ajánlott. Az így kapott markereken újból elvégeztük azok szűrését a genotípus-frekvenciák összehasonlítása alapján annak érdekében, hogy az esetleges csoportspecifikus batch hatást kizárjuk, illetve csökkentjük. A maradék 1 107 448 markert egységes adathalmazként kezeltük a továbbiakban, amelyek állapota minden páciensre rendelkezésünkre álltak. A fenti lépéseket a 3. táblázat foglalja össze.

2.1.3. Humán polimorfizmusok vs. HIV-mutációk

A vírus szekvenciákat génenként illesztettük, majd nukleotidról aminosavkódra fordítottuk azokat. A vírus kilenc génjében összesen 913 aminosav-pozíciót találtunk, amelyek legalább 20 egyénben variábilisak voltak. Ebben a 913 pozícióban összesen 1998 különböző aminosavat

	OmniExpress	1M	550K-650K	Affymetrix 6.0
Minőségellenőrző lépések előtt	730 525	878 351	549 455	934 968
Minőségellenőrző lépések után (lásd Módszerek)	580 088	775 125	411 026	659 995
Imputáció után	1 180 010	1 180 790	1 174 832	1 168 986
Közös SNP-k összefésülése után		1 107 448		

3. táblázat. A markerek száma a genotípusadatok előkészítésének különböző szakaszaiban.

figyeltünk meg, ezeket a mutációkat bináris változókkal reprezentáltuk.

A 1998 bináris változó és 1 107 448 SNP között összesen 1998×1107448 logisztikus regressziót végeztünk el. Az asszociációvizsgálatok során additív genetikai modellt feltételeztünk, azaz az SNP-k alléldózisait teszteltük a függő változó ellen. Ebben a humán polimorfizmusok és HIV-mutációk között végzett asszociációvizsgálatban szignifikáns hatásokat találtunk az MHC régió és több, az egész HIV-genomban egyenletesen elhelyezkedő aminosav-pozíció között (7. ábra). A Gag és Nef fehérjékben szignifikánsan több asszociált aminosav-pozíció található, mint a genom többi részében (Gag: 6,8% vs. 2,6% $p=0.001$; Nef: 11% vs. 2,6% $p=1,2 \times 10^{-5}$, binomiális teszt). Bonferroni-korrekción (a tesztenkénti szignifikanciaszint 2.2×10^{-11}) után 40 virális aminosav mutatott szignifikáns asszociációt összesen 1260 humán polimorfizmussal az MHC régióban (7. ábra és 4-6. táblázatok). A legerősebb asszociációt a Nef 135. kodonja mutatja ($p = 1 \times 10^{-49}$). Szignifikáns hatást nem találtunk az MHC régió kivül eső genomi régiókban. Az MHC régióban a lókuszok erős kapcsoltságban állnak egymással, ennek tudjuk be az asszociált SNP-k magas számát, ezért egy adott aminosav-pozícióra csak a legerősebb asszociációt mutató SNP-t vettük figyelembe a továbbiakban.

Annak érdekében, hogy ezt az MHC régióban talált hatást független komponenseire bontsuk egy iterációs eljárásban kondicionális regressziók sorozatát végeztük. Először a legerősebben asszociált SNP-re kondicionálva (azaz magyarázó változóként szerepeltetve a modellben) ismételtük meg a regressziókat, majd az így kapott legerősebben asszociált SNP-t ismét hozzáadtuk a modellhez. Ezt az eljárást minden szignifikáns asszociációt adó virális aminosav-pozícióra külön végeztük el.

Ezután megvizsgáltuk, hogy a legerősebb asszociációt mutató SNP-k mely *MHC-I* allélok-kal fordulnak elő egy haplotípuson. A *HLA* allélok SNP-kből történő imputációja hasonló, mint a 2.4 fejezetben ismertetett genotípus imputáció, de szekvenciaalapú MHC genotipizálás szük-

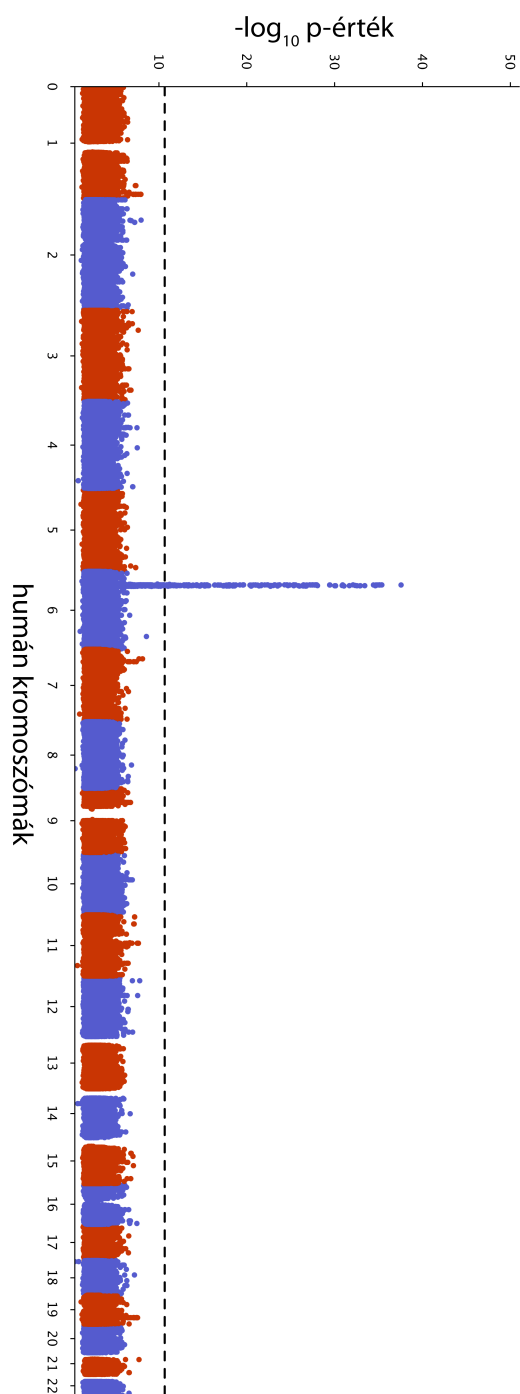
séges hozzá a referencia panelen. Az imputációt korábban publikált módszerekkel végezték a munkatársaim ((Pereyra, 2010)). A *HLA* allélok imputációját követően megállapítottuk a 49 SNP és a legalább 5%-os gyakorisággal jelenlevő *HLA* allélok között a kapcsoltság (LD) mértékét (Pearson-féle korreláció, illetve a D' formájában, lásd 2.4. fejezet). Mint a 4-6. táblázatok mutatják, a legtöbb esetben igen magas kapcsoltságot találtunk az SNP és azon *MHC-I* allél között, amely allél az adott víruspozícióban található epitópot célozza. Például a legerősebb asszociációt a Nef 135. kodon és az rs16896119 SNP között találtuk: ez a kodon egy ismert *HLA-A*24:02* epitópban található (RYPLTFGW 134-141), míg az rs16896119 SNP majdnem tökéletes kapcsoltságban áll a *A*24:02* alléllal ($D'=1.00$ $r^2=0.87$). A 40-ből összesen 15 aminosavpozíció található olyan ismert és publikált epitópokban, amelyeket az adott SNP-vel kapcsoltságban álló *MHC-I* allél prezentál. (Az epitópok listája Carlson és munkatársai 2012-es cikkéből származik, amely magába foglalja a Los Alamos-i adatbázist is (Yusim, 2009; Carlson, 2012).) A maradék 25 kodon vagy eddig fel nem tárt epitóp, vagy kompenzációs mutáció lehet.

Ezek az eredmények bizonyítják, hogy a gazda-parazita genomok együttes asszociációvizsgálatával sikeresen tárható fel a *HLA*-epitópok, illetve másodlagos escape mutációk térképe a parazita genomján.

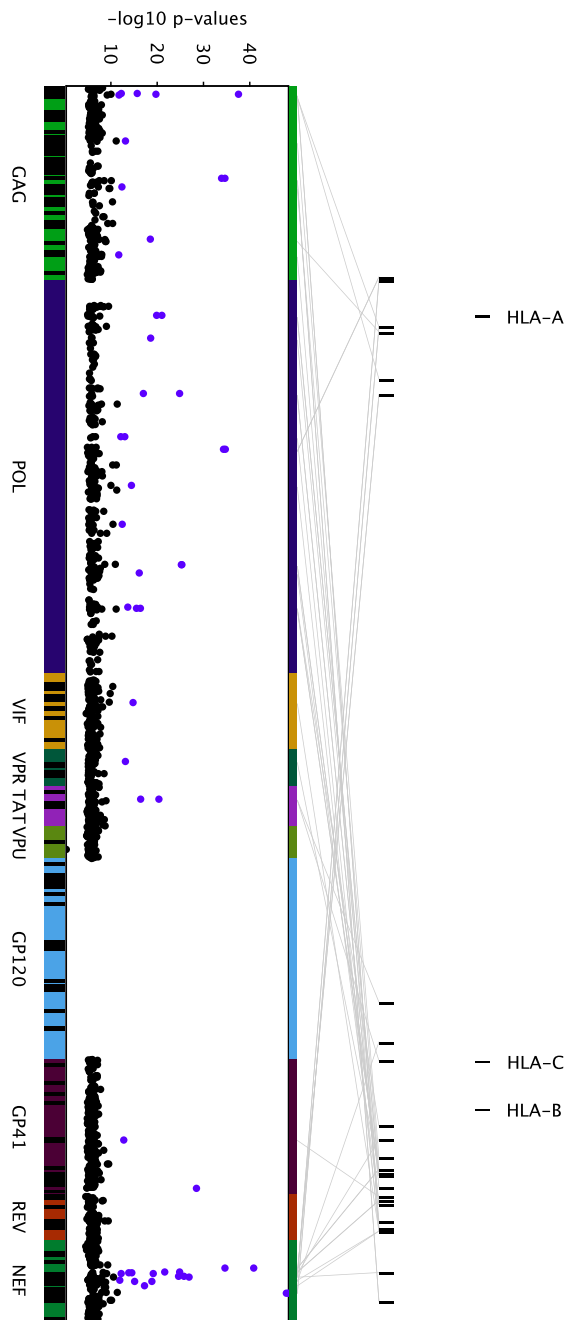
Az így feltárt polimorfizmusokat a 4-6. táblázatok tartalmazzák.

2.1.4. Virális mutációk vs. vírusszint

1998 lineáris regressziót végeztünk a vírusszint és a HIV-mutációkat reprezentáló bináris változók között. Ezt a vizsgálatot csupán 186, a svájci kohorszban követett páciens adatain tudtam elvégezni. Egyik mutáció se mutatott szignifikáns hatást a vírusszintre. Az előző fejezetben ismertetett humán genetikai faktorokkal asszociáló mutációk eredményei a 8. ábrán láthatóak. Ezen - a gazda polimorfizmusokkal szignifikáns módon asszociáló - kodonok hatása -0.2 -től $+0.15$ \log_{10} kópia/ml-ig terjed (8. ábra). Azonban meg kell jegyeznünk, hogy a nagy számú statisztikai teszt miatt a szignifikancia szintet 2.5×10^{-5} -re kellett állítanunk, ilyen szigorú szignifikanciaszint és ilyen gyenge hatások mellett pedig a 186 elemszámból álló adathalmaz



6. ábra. A humán polimorfizmusok és a vírusmutációk közti asszociációvizsgálat eredménye a humán genomra vetítve (ún. Manhattan-plot). Az x tengelyen a kromoszómák láthatók egymás után 1-től 22-ig (váltakozó színnel). Az y tengelyen az asszociáció erőssége látható. Az ábra a vírus Gag fehérjének összes mutációjával készült asszociációkat ábrázolja, minden pont egy SNP-vírusmutáció párt reprezentál. A Vpu és a Rev fehérjék kivételével a többi génről hasonló ábra készíthető (azokban nem találtunk szignifikáns asszociációt). A 6. kromoszómán található csúcs az MHC-régióból származik. A vízszintes szaggatott vonal a szignifikanciaszintet jelöli.



7. ábra. A humán polimorfizmusok és a vírusmutációk közti asszociációvizsgálat eredménye a vírus proteomjára vetítve. Az ábra a vírus génjeit és az azokban található asszociációk erősségét szemlélteti. A fekete sávozás az ismert epitópokat jelöli. Minden pont egy SNP-vírusmutáció párt jelöl. Egy adott virális aminosav-pozícióban, csak a legerősebb asszociáció van ábrázolva. A szignifikanciaszintet (2.2×10^{-11}) elérő összes asszociáció az MHC-régióból származik. A grafikon felett a humán genom MHC régiójának itt releváns génjei és SNP-i láthatóak. A szignifikáns asszociációkat szürke vonalak reprezentálják.

kicsi.

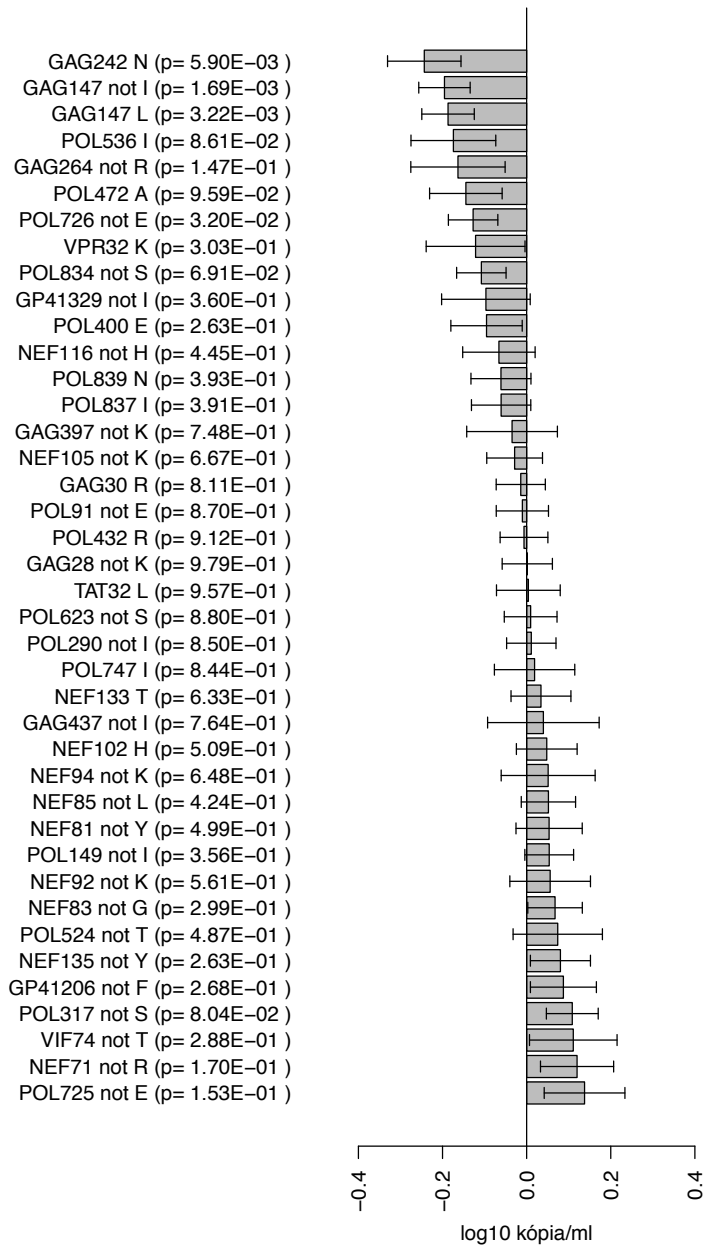
2.2. Diszkusszió

A HIV-vel kapcsolatban eddig elvégzett humán genomikai vizsgálatok mind klinikai állapotra (rezisztencia, betegségprogresszió) vagy laboreredményekre (CD4 T-sejt szám, setpoint vírus-szint) fókuszáltak. Azonban ezek a fenotípusok nem tökéletesek, igen nagy varianciát mutathatnak (inherens módon, vagy pedig a nem egységes protokollok miatt). Ez a variancia pedig súlyosan érintheti a nagy, több kohorszon átívelő genomikai asszociációvizsgálatok erejét (Evangelou, 2011). Továbbá, ezek a klinikai- és laboreredmények hosszú követési időt igényelnek az inherens variabilitásuk miatt, ami nehezzé teszi az adatgyűjtést.

Hogy áthidaljuk ezt a problémát, új módszert használtunk a fertőző betegségek gazda oldali genomikai vizsgálatára. A humán polimorfizmusok és a vírusszekvenciák egyidejű felhasználásával lehetőségünk nyílt egy lépésben megkeresni a vírusgenom azon pontjait, ahol az konfliktusban áll a gazdával. Módszerünk során a vírus genetikai variációját mint köztes fenotípust használtuk. Hasonló köztes változókat sikerrel alkalmaztak egyéb GWA tanulmányokban is, olyan asszociációkat felfedve, amelyek a komplex klinikai változókkal rejtve maradtak volna (pl.: szérumszint IgE koncentráció asztma esetén (Moffatt, 2010) vagy képzett eljárások alapján képzett változók pszichiátriai betegségek esetén (Rasetti & Weinberger, 2011)). A patogén szekvenciája sok esetben egyszerűbb, könnyebben értelmezhető, és egységesebb fenotípusváltozó lehet, mint a hagyományos klinikai változók.

Ez az új megközelítés lehetővé tette, hogy egy kísérletben feltérképezzük a HIV-vírus és gazdájának genomikai kölcsönhatásait. A legerősebb asszociációkat a *MHC-I* gének alléljaival kapcsolatban álló markerek és azon kodonok között találtuk, amelyek ezen allél epitópjai-ban találhatóak. Ez az eredmény bizonyítja, hogy a módszerrel valós, biológiailag értelmezhető eredmények nyerhetők, anélkül, hogy bármiféle előzetes tudással rendelkezünk a két organizmus felépítéséről, biológiájáról, illetve együttműködésükről. A vírust adaptációra kényszerítő gazda faktorokat találhatunk, ha ezen faktorok lenyomatait keressük a patogén genomjában. Továbbá asszociációkat találtunk olyan virális kodonokban, melyek nem részei ismert epitópoknak: ezek lehetnek új, eddig még le nem írt epitópok, vagy másodlagos kompenzációs mutációk.

A vírus Gag és a Nef fehérjében tapasztaltuk a legnagyobb szelekciós nyomást (a legtöbb



8. ábra. A 40 MHC-vel asszociált HIV aminosav hatása a setpoint vírusszintre. A regressziókat a bináris változókkal végeztük. Egy adott pozícióban minden legalább 20 példányban megfigyelhető aminosavról készült egy bináris indikátorváltozó. Az ábra a legerősebb hatást mutató változókat ábrázolja egy adott pozícióról.

asszociált pozíciót), ez a különbség szignifikáns volt, mikor a vírusgenom többi részével hasonlítottuk össze. Ez konzisztens azon korábbi megfigyelésekkel, melyek a Gag p24 fehérje fontosságát hangsúlyozzák a megfelelő sejtes immunválasz kialakításában: lassabb betegség-progresszió figyelhető meg azon páciensekben, akik ezen fehérje ellen alakítanak ki citotoxikus immunválaszt (Borghans, 2007; Brennan, 2012), illetve az elit controller státusz tovább fennmaradhat ezekben a páciensekben (Dyer, 2008). Ez szintén megerősíti, hogy a módszer biológiailag releváns hatásokat képes kimutatni.

Új genetikai faktort, amely hatással lenne a HIV-vírusra nem azonosítottunk. Ekkora mintaméret mellett és ennyi tesztet elvégezve csak nagyon erős hatásokat lehet kimutatni. A vizsgálatnak 80%-os ereje¹⁰ volt arra, hogy egy 10%-os gyakoriságú 3,9 odd ratio-val rendelkező genetikai marker hatását kimutassa, feltételezve, hogy a virális mutáció 100 szekvenciában jelen van (Purcell, 2003). Nem szabad elfelejteni, hogy a módszerünk csak polimorf genetikai faktorok esetében alkalmazható, a populációban már rögzült mutációk hatását felderíteni természetesen képtelen. Továbbá csak olyan humán genetikai faktorok hatását mutatja ki, amelyek a vírus szekvenciáját a szervezeten belüli evolúcióban befolyásolhatják, így a restriktív faktorok hatását nem. Az egyik legfontosabb hiányossága a tanulmányunknak az, hogy a vírus gp120 fehérjéjét nem tudtuk megvizsgálni. Ez a fehérje központi szerepet tölt be a vírus külső burka és a gazdasejtek közti interakcióban, így joggal várhatjuk, hogy szekvenciája szelekciós nyomás alatt áll.

Megemlítendő, hogy az asszociációvizsgálatokat additív genetikai modellt feltételezve végeztük el. Ez megfelelő erővel képes az additív, a domináns és a gyakori recesszív változatok hatását kimutatni, azonban ritka recesszív mutációk felfedéséhez az érzékenysége csekély (Lettre, 2007).

A legfontosabb eredményünket a 6. ábra foglalja össze: látható, hogy a vírus szervezeten belüli evolúciójára az MHC régióban található gyakori mutációk igen erős hatással vannak, míg ezen a régió kívül nincs más gyakori mutáció észrevehető hatással a vírusra. Felmerül a kérdés, miért nincs? Figyelembe kell vennünk, hogy a vizsgált fenotípusból következően, csak azon humán polimorfizmusok hatásait látjuk, amelyek a krónikus szakaszban lévő fertőzés lefolyása alatt szelekciós nyomást fejtenek ki a vírusra. Tehát nem figyelhetünk meg olyan

¹⁰ Annak a valószínűsége, hogy a teszt elveti a null hipotézist abban az esetben, ha az alternatív hipotézis valóban igaz. Ha β jelöli a másodfajú hiba valószínűségét, akkor $1-\beta$.

hatásokat, amelyek rezisztenciát biztosítanak a HIV-fertőzés ellen (például a CCR5 Δ 32 mutációjának homozigóta formáját, vagy egyéb hipotetikus faktorok hatásait). Állatkísérletek azonban azt mutatják, hogy a veleszületett immunrendszer antivirális faktorai képesek befolyásolni a vírus szervezeten belüli evolúcióját (Kirmaier, 2010). Ezen faktorok hatásait tehát látnánk, ha azok hatásai elegendően erősek, valamint mutációik elegendően gyakoriak lennének a populációban. Ritka mutációk hatásait se a kis mintaméret, se az alkalmazott genotipizáló technika következtében nem állt módunkban felderíteni. Az MHC-régióban talált jel pedig bizonyítja, hogy a vizsgálatunk megfelelően erős volt ahhoz, hogy a gyakori mutációk hatásait kimutassa.

Elgondolkodhatunk tehát azon, miért nincs elegendően erős és gyakori variáció az MHC-n kívüli régiókban. Azt a lehetőséget, hogy ilyen variációk nem képesek létrejönni, azaz az adaptív immunrendszer kivételével más nem képes felvenni a harcot a vírussal, ellenpélda ismeretében elvetem (restrikciós faktorok, CCR5 Δ 32 heterozigóta formában lásd Bevezető). Elképzelhető, hogy volt kedvező variáció más génekben, de az az emberi generációk során rögzült, és további „javulás” már nem érhető el. A monomorf, invariáns állapot természetesen genetikai asszociációvizsgálat előtt láthatatlan marad. Ebben az esetben az érintett génekben a kedvezőtlen mutációk nem terjednének el a populációban, a vizsgálatunk pedig nem alkalmas ritka variációk hatásainak a felderítésére. Ha létezik ilyen hipotetikus mutáció a jelenlegi emberi populációban, de a hatása kicsi, vagy frekvenciája alacsony, akkor a vizsgálatunk nem képes azt kimutatni. Tehát annak ellenére, hogy gyakori variációkat csekély mintaméreten felhasználva igen szép eredményeket értünk el az MHC-régióban, az előbbieik alapján a fenotípusra ható humán variációk jobb feltáráshoz nagyobb mintaméretre és fejlettebb szekvenálási technikákra van szükség.

A CCR5 Δ 32 mutáció heterozigóta formában nem biztosít rezisztenciát a vírus ellen, de lassítja a betegségprogressziót. Ezért jogosan várjuk, hogy a vírus szervezeten belüli evolúcióját ez a mutáció befolyásolja. Sajnos a genotipizáló csipek nem tartalmazzak olyan SNP-t amely elegendően közel, illetve elegendően erős kapcsoltságban áll a CCR5 génnel, így a mutáció hatásainak felfedéséhez sokkal nagyobb mintaméret lenne szükséges. Ezen a helyzeten a genotípus imputáció sem segít, mert a HapMap populációiba nem válogattak be olyan egyéneket (McLaren, 2013), akik hordozzák ezt a mutációt, így referenciahaplotípus hiányában nem lehet megállapítani a környező SNP-k állapotáról, hogy az adott egyén hordozza-e a mutációt.

2.3. Összefoglalás

Egy új, mind a gazda, mind a parazita genomját felhasználó, előzetes hipotéziseket nélkülöző módszerrel asszociációkat fedtünk fel a HIV-vírus genetikai variációja és humán polimorfizmusok között. Ez a stratégia lehetővé teszi a gazda-patogén kapcsolat széleskörű genomikai leírását és feltárja azokat a lókusztokat, ahol a két genom egymással „konfliktusban” van. A módszer kiterjeszhető, alkalmazható más gazda-parazita rendszerekre is, feltéve, hogy a gazda variábilis genetikai faktorai elegendő szelekciós hatást fejt ki a patogénre ahhoz, hogy az escape mutációkra kényszerüljön.

2.4. Genetikai asszociációvizsgálat során alkalmazott módszerek

Genotípus hozzárendelés („imputáció”) és a haplotípusok levezetése GWAS adatokból

Joggal feltételezhető, hogy a humán populációban megtalálható gyakori polimorfizmusok nem új mutációval jelennek meg minden hordozó egyénben, hanem a polimorfizmus egyszer (vagy néhányszor) jelent meg a populációban, majd szelekció vagy genetikai sodródás révén terjedt el, és más genomrégiókkal rekombinálódhatott is. A polimorfizmusok tehát haplotípusokba szerveződnek és haplotípus-blokkokban öröklődnek, amely blokkok crossing over során kereszteződhetnek. GWAS-csippel történő genotipizálás során a haplotípusokra nem derül fény, azaz nincs közvetlen információnk arról, hogy két polimorfizmus mely alléljai találhatóak ugyanazon a kromoszómán (úgynevezett fázis nélküli genotípusok). Ha azonban egy adott genomi régióból elegendő számú polimorfizmus állapotát ismerjük, és rendelkezésünkre áll egy referencia haplotípus-készlet, akkor következtethetünk az adott egyén által hordozott haplotípusokra az adott geográfiai régióban. Minden egyén genotípusaihoz megpróbálhatunk hozzárendelni két, a referencia haplotípusokból képzett mozaikot oly módon, hogy a mintában mért fázis nélküli, és a mozaikban levő fázissal rendelkező SNP-k állapotai megegyezzenek. A haplotípus ismeretében pedig megjósolhatjuk az adott populáció egyéb, közvetlenül nem genotipizált polimorfizmusainak az állapotát. Ezt a módszert nevezi az angol szakirodalom genotype imputation-nek. Egy egyén haplotípusainak meghatározása (phasing) ugyanezekben az alapelveken nyugszik.

A genotípus-imputáció kivitelezéséhez szükség van egy, az adott populációra vonatkozó referencia haplotípuskészletre. Mi a HapMap projekt nyugat-európai leszármazással bíró populációjáról készült haplotípusokat használtuk erre a célra (Frazer, 2007; Altshuler, 2010). Ez a

panel szülő-utód triók genotipizálása után készült, mivel ilyen triókban a haplotípusok meghatározása pontosabb (hiszen az utód genotípusából következtetni lehet a szülők haplotípusaira). A referencia haplotípusok segítségével lehetőség nyílik a referencián meglévő, de a mi mintáinkban közvetlenül nem genotipizált SNP-k állapotainak meghatározására. Természetesen ez a meghatározás nem tökéletes, ezért az eredmény minden lókuszra a genotípusok egy lehetséges eloszlása csupán.

A genotype imputation kivitelezésére a mach szoftvert használtuk (Li, 2010). A szoftver honlapján található utasítások alapján alkalmaztuk a szoftvert az alábbi parancsokkal. A kapott imputált genotípusokat a ritkábbik allél frekvenciája és a szoftver által számolt korrelációmérték (r^2) alapján szűrtük. Csak azokat a SNP-eket tartottuk meg, ahol a ritkább allél frekvenciája $\geq 0,05$ és az r^2 érték $\geq 0,3$.

```
$mach1 --autoflip -d mintaDat -p mintaPed -s referenciaSNP \\  
-h referenciaHaplotipusok --greedy --rounds 20 --prefix aa  
$mach1 --autoflip -d mintaDat -p mintaPed -s referenciaSNP \\  
-h referenciaHaplotipusok --greedy --crossover aa.recfile \\  
--errormap aa.eratefile --mle --mldetails --prefix kimenet
```

A páciensek feljegyzett nemének és genotípusának összevetése

Ha egy páciens feljegyzett neme férfi volt, de az X kromoszómán elhelyezkedő polimorfizmusokra heterozigóta, akkor a pácienset kihagytuk a további vizsgálatokból. Ezt a vizsgálatot a plink szoftverrel végeztük.

A vizsgálatba bevont páciensek leszármazásának ellenőrzése

A vizsgálatba bevont egyének genotípusait és a nyugat-európai HapMap CEU populáció (mint referenciapopuláció) tagjainak a genotípusait összesítve főkomponens analízist végeztünk a smartpca szoftverrel (Price, 2006). A smartpca szoftvert alapbeállításokon használtuk kivéve az alábbi három kapcsolót. Ezek segítségével biztosítottuk, hogy a főkomponens analízis egymástól független, kapcsoltságban nem álló SNP-ken történjen.

```
killr2: YES -- LD alapú szűrés bekapcsolása  
r2thresh: 0.2 -- LD alapú szűrés küszöbparamétere  
numoutlieriter: 0 -- ne keressen kieső pácienseket
```


A humán genotípus adatok minőségellenőrzése és összefésülése

A különböző gyártmányú GWAS csipekről, és különböző kohorszokból származó adatok összefésülése előtt több minőségellenőrző lépést végeztünk, minden kohorsz mintáira külön-külön. Elvetettünk egy SNP-t, ha kevesebb, mint a kohorsz pácienseinek 99%-ban sikerült helyesen megmérni (genotipizálni) az állapotát; ha a ritkábbik allél gyakorisága az adott populációban kevesebb mint 5% volt; illetve ha genotípus eloszlása az adott populációban szignifikánsan eltért a Hardy-Weinberg-egyensúlytól (0,00005-es szignifikanciaszintet használva). Ezeket a szűrőket a plink szoftver segítségével alkalmaztuk. Azon páciensek adatait nem elemeztük tovább, akiknek a genotípusai nem feleltek meg a páciens feljegyzett nemének (lásd fentebb), illetve akik DNS mintáiból a genotipizálás nem sikerült (azaz a sikeresen mért SNP-k száma kevesebb, mint 99%-a a vártnak).

A különböző gyártmányú GWAS csipek különböző SNP-k állapotait mérik, így a vizsgálatba bevont páciensekről különböző SNP-k álltak rendelkezésünkre. Ahhoz, hogy ezeket az adatokat egységesen elemezhesük szükséges volt a rendelkezésre álló SNP-k közös halmazát kialakítani. Sajnos a csipek diverzitása miatt a közös metszet csupán néhány száz-ezer SNP-t tartalmazott, ezért a genotípus imputáció módszerét alkalmazva hoztuk a különböző forrásból származó adatokat „közös nevezőre” (lásd fentebb). Az imputációt a HapMap CEU populáció referencia-haplotípusait ((Frazer, 2007; Altshuler, 2010), <http://www.sph.umich.edu/csg/abecasis/MACH/download/HapMap3.r2.b36.html>) felhasználva végeztük, külön az Illumina és külön az Affymetrix által gyártott platformokra (csipekre), összesen négy külön csoportban (Illumina 1M, Illumina OmniExpress12v1H, Illumina 550K és 650K, Affymetrix 6.0). Mind a genotype imputation előtt, mind azután összehasonlítottuk az egyes polimorfizmusok genotípus-eloszlásait mind a csoportok között, mind a HapMap referenciapopulációval. Ehhez az összehasonlításhoz χ^2 -próbán alapuló függetlenségvizsgálatot használtunk. Amennyiben egy SNP genotípus-eloszlása nem volt független a platformtól, úgy azt az SNP-t figyelmen kívül hagytuk. A próbák során Benjamini-Hochberg procedúra segítségével korrigáltuk a p-értékeket, 20%-os False Discovery Rate-et használva (Benjamini & Hochberg, 1995). Utóbbi kritériumot azért választottuk a Bonferroni-korrekciónak helyett, mert a Bonferroni-korrekciónak túl konzervatív (alacsony p-érték küszöb miatt kevesebb markert vetne el), és ebben a minőségellenőrzési lépésben inkább elvetettük a markert, mintsem az később hamis pozitív eredményt adjon.

Ennek eredményeképpen minden páciensről rendelkezésünkre álltak ugyanazoknak a polimorfizmusoknak az állapotai, így azokat egységesen elemezhettük a továbbiakban. (Az imputáció részleteit másik fejezetben tárgyalom.)

Asszociációvizsgálatok a humán- és a vírusgenom között

Célunk a vírus genomjában található mutációk és humán polimorfizmusok között esetleg fennálló asszociációk felderítése volt. Ehhez logisztikus regressziók sorozatát végeztük el a vírus mutációit kifejező bináris változók (egy adott pozícióban az adott minta hordozza-e a mutációt) és a humán polimorfizmusok alléldózisa között. A vírusszekvenciákat génenként illesztettük Muscle szoftver segítségével (Edgar, 2004), lefordítottuk őket, majd az illesztés segítségével létrehoztuk azokat a bináris változókat, amelyeket felhasználtunk a regresszióban. Minden aminosav-pozícióban, minden A aminosavra létrehoztunk egy bináris változót, amely 1 értéket vesz fel, ha az adott minta (páciens) az adott pozícióban A értékű, míg 0-t ha nem A (9. ábra). Elvetettük azon bináris változókat, amelyeknél a ritkábbik állapot kevesebb mint 20 esetben fordult elő. A maradék változókat, egymástól függetlenül minden SNP-vel külön-külön teszteltük, additív genetikai modell feltételezve. A regressziókban a függő változó szerepét a vírusszekvenciákból származtatott bináris változók töltötték be, az SNP-k alléldózisai pedig magyarázó változók voltak. Továbbá a kohorszra, és a maradék humán populációstruktúrára további magyarázóváltozók segítségével korrigáltuk a regressziót. A humán populációstruktúrát főkomponens analízis segítségével tártuk fel, és a regresszióban a páciensek első két főkomponensen felvett koordinátaival reprezentáltuk. A genotípusmátrix sajátvektorainak és sajátértékeinek kiszámolásához a smartpca szoftvert használtuk (Price, 2006).

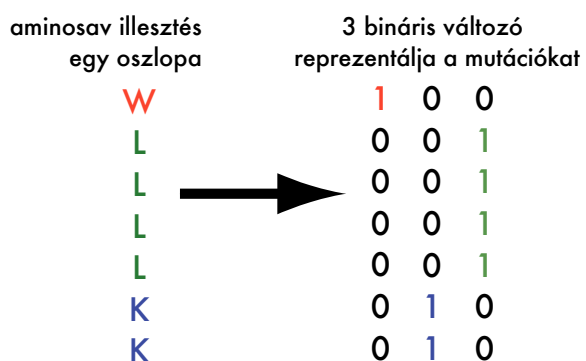
A regresszióban szereplő modell: $V = SNP + Kohorsz + PCA_1 + PCA_2$, ahol az SNP és a PCA_x változók arányskálán, a $Kohorsz$ változó pedig nominális skálán volt értelmezve.

Student-féle t-tesztet használtunk annak eldöntésére, hogy a regresszióban becsült hatás méret (odds ratio) szignifikánsan eltér-e nullától. A statisztikai tesztet és a regressziót plink (Purcell, 2007) szoftverrel végeztük.

Abban az esetben, ha egy virális mutáció szignifikáns asszociációt mutatott valamely SNP-vel, akkor a legerősebben asszociáló humán polimorfizmusra kondicionálva (magyarázóváltozóként felvéve) megismételtük a többi SNP tesztjét. Ezt addig ismételtük, amíg valamely SNP szignifikáns maradt. Az egymást követő iterációkban az összes megelőző lépésben szereplő

HIV	POZ	SNP	P	HLA	EPITÓP
GAG	26	rs2523684	2.04×10^{-16}	<i>B*15:01</i> ($D'=0.96$ $r^2=0.91$)	-
GAG	28	rs2571382	2.74×10^{-38}	<i>A*03:01</i> ($D'=1.00$ $r^2=0.80$)	RLRPGGKKK (20-28)
		rs1611430	1.86×10^{-20}	-	-
GAG	30	rs16896119	1.75×10^{-12}	<i>A*24:02</i> ($D'=1.00$ $r^2=0.87$)	KYKLVKHW (28-36)
GAG	147	rs12212594	7.08×10^{-14}	<i>C*06:02</i> ($D'=0.99$ $r^2=0.85$)	-
GAG	242	rs28732144	2.31×10^{-35}	<i>B*57:01</i> ($D'=0.95$ $r^2=0.83$)	TSTLQEQIGW (240-249)
		rs4346874	2.80×10^{-22}	-	-
		rs6929464	1.80×10^{-18}	<i>C*06:02</i> ($D'=0.88$ $r^2=0.38$)	-
GAG	264	rs3819299	4.14×10^{-13}	<i>B*08:01</i> ($D'=1.00$ $r^2=0.01$)	EIYKRWII (260-267)
GAG	397	rs7760172	3.02×10^{-19}	-	-
GAG	437	rs12206131	1.99×10^{-12}	<i>B*13:02</i> ($D'=1.00$ $r^2=0.96$)	RQANFLGKI (429-437)
GP41	206	rs3763288	1.66×10^{-13}	<i>B*15:01</i> ($D'=0.98$ $r^2=0.96$)	-
GP41	329	rs1875565	3.27×10^{-29}	<i>B*15:01</i> ($D'=0.99$ $r^2=0.01$)	-

4. táblázat. A HIV-vírus fehérjéinek egyes pozíciói és humán SNP-k között talált asszociációk. Egy adott virális kodonnal több független SNP asszociálhat, ezeket a szövegben részletezett iterációs eljárással határoztuk meg. A markerekhez feltüntettem a vele kapcsoltságban álló HLA allélokot és az allél által esetleg felismert epitópszekvenciát.



9. ábra. Vírusmutációk reprezentálása a logisztikus regressziókban. A többszörös szekven-
ciaillesztés minden oszlopából 1 vagy több nullákat és egyeseket tartalmazó bináris változót
készítettünk. Minden megfigyelt aminosavra készült egy ilyen változó.

HIV	POZ	SNP	P	HLA	EPITÓP
NEF	71	rs2523612	1.41×10^{-41}	<i>B*07:02</i> ($D'=0.99$ $r^2=0.95$) <i>C*07:02</i> ($D'=0.99$ $r^2=0.84$)	FPVTPQVPLR (68-77) -
		rs2523612	4.28×10^{-12}	<i>B*07:02</i> ($D'=0.99$ $r^2=0.95$) <i>C*07:02</i> ($D'=0.99$ $r^2=0.84$)	FPVTPQVPLR (68-77) -
NEF	81	rs6910087	1.41×10^{-25}	<i>B*07:02</i> ($D'=1.00$ $r^2=0.02$)	RPMTYKAAL (77-85)
NEF	83	rs2232236	3.00×10^{-15}	-	-
		rs12111032	5.80×10^{-12}	<i>B*07:02</i> ($D'=0.92$ $r^2=0.04$) <i>B*15:01</i> ($D'=0.93$ $r^2=0.17$) <i>C*03:04</i> ($D'=1.00$ $r^2=0.25$) <i>C*06:02</i> ($D'=1.00$ $r^2=0.30$)	RPMTYKAAL (77-85) - - -
NEF	85	rs34131062	6.40×10^{-12}	-	-
		rs9259004	1.04×10^{-14}	-	-
		rs28367646	7.09×10^{-20}	-	-
		rs12111032	5.80×10^{-12}	<i>B*07:02</i> ($D'=0.92$ $r^2=0.04$) <i>B*15:01</i> ($D'=0.93$ $r^2=0.17$) <i>C*03:04</i> ($D'=1.00$ $r^2=0.25$) <i>C*06:02</i> ($D'=1.00$ $r^2=0.30$)	RPMTYKAAL (77-85) - - -
		rs9461680	2.95×10^{-14}	<i>B*07:02</i> ($D'=0.84$ $r^2=0.02$) <i>C*03:04</i> ($D'=1.00$ $r^2=0.36$)	RPMTYKAAL (77-85) -
NEF	92	rs3888722	1.63×10^{-26}	<i>A*11:01</i> ($D'=1.00$ $r^2=0.99$)	AVDLSHFLK (84-92)
NEF	94	rs2734583	1.29×10^{-27}	<i>B*08:01</i> ($D'=0.94$ $r^2=0.88$)	FLKEKGGL (90-97)
		rs2535340	1.48×10^{-26}	<i>B*08:01</i> ($D'=0.88$ $r^2=0.74$) <i>C*07:01</i> ($D'=0.81$ $r^2=0.45$)	FLKEKGGL (90-97) -
NEF	102	rs2255221	1.20×10^{-12}	-	-
NEF	105	rs3130557	1.40×10^{-19}	<i>B*08:01</i> ($D'=0.95$ $r^2=0.85$)	-
		rs3130696	1.62×10^{-13}	<i>B*08:01</i> ($D'=0.98$ $r^2=0.31$) <i>C*07:01</i> ($D'=1.00$ $r^2=0.50$) <i>C*07:02</i> ($D'=1.00$ $r^2=0.05$)	- - KRQEILDLWVY (105-115)
NEF	116	rs2395029	5.45×10^{-18}	<i>B*57:01</i> ($D'=1.00$ $r^2=0.98$)	HTQGYFPDW (116-124)
NEF	133	rs16896119	4.13×10^{-12}	<i>A*24:02</i> ($D'=1.00$ $r^2=0.87$)	-
		rs5010528	1.80×10^{-11}	<i>C*04:01</i> ($D'=1.00$ $r^2=1.00$)	-
NEF	135	rs16896119	4.36×10^{-49}	<i>A*24:02</i> ($D'=1.00$ $r^2=0.87$)	RYPLTFGW (134-141)

5. táblázat. Az ábraalírás a 4. táblázat alatt található.

HIV	POZ	SNP	P	HLA	EPITÓP
POL	91	rs2596551	9.72×10^{-22}	<i>B*44:02</i> (D'=1.00 $r^2 = 0.49$)	-
		rs9266462	1.12×10^{-14}	-	-
POL	149	rs9266775	2.57×10^{-19}	<i>B*15:01</i> (D'=1.00 $r^2 = 0.53$)	-
				<i>B*51:01</i> (D'=1.00 $r^2 = 0.01$)	EKEGKISKI (141-149)
		rs2523684	7.71×10^{-19}	<i>B*15:01</i> (D'=0.96 $r^2 = 0.91$)	-
POL	290	rs2848713	1.44×10^{-25}	<i>B*51:01</i> (D'=0.95 $r^2 = 0.59$)	-
		rs2395030	6.20×10^{-16}	-	-
		rs4959053	3.67×10^{-11}	-	-
POL	317	rs2523612	4.28×10^{-12}	<i>B*07:02</i> (D'=0.99 $r^2 = 0.95$)	-
				<i>C*07:02</i> (D'=0.99 $r^2 = 0.84$)	-
POL	400	rs2395029	9.85×10^{-14}	<i>B*57:01</i> (D'=1.00 $r^2 = 0.98$)	-
		rs13210132	7.33×10^{-13}	<i>C*06:02</i> (D'=0.89 $r^2 = 0.44$)	-
POL	432	rs1625636	1.90×10^{-35}	<i>A*03:01</i> (D'=0.99 $r^2 = 0.76$)	-
		rs11757235	1.40×10^{-16}	-	-
		rs9260706	2.99×10^{-24}	-	-
POL	472	rs2523684	6.66×10^{-12}	<i>B*15:01</i> (D'=0.96 $r^2 = 0.91$)	-
POL	524	rs6930572	3.65×10^{-15}	<i>B*13:02</i> (D'=0.95 $r^2 = 0.90$)	-
POL	536	rs6929434	5.43×10^{-12}	<i>C*12:03</i> (D'=0.99 $r^2 = 0.94$)	-
POL	623	rs519417	3.57×10^{-13}	<i>A*11:01</i> (D'=0.83 $r^2 = 0.01$)	QIEQLIKK (619-627)
				<i>B*08:01</i> (D'=0.89 $r^2 = 0.76$)	-
				<i>C*07:01</i> (D'=0.88 $r^2 = 0.48$)	-
		rs886424	6.30×10^{-12}	<i>A*11:01</i> (D'=1.00 $r^2 = 0.01$)	QIEQLIKK (619-627)
				<i>B*08:01</i> (D'=0.82 $r^2 = 0.62$)	-
POL	725	rs2596551	1.10×10^{-11}	<i>B*44:02</i> (D'=1.00 $r^2 = 0.49$)	-
		rs9266462	1.12×10^{-14}	-	-
POL	726	rs2596551	4.60×10^{-26}	<i>B*44:02</i> (D'=1.00 $r^2 = 0.49$)	-
		rs9266433	2.69×10^{-16}	-	-
		rs2523590	8.20×10^{-14}	<i>B*18:01</i> (D'=1.00 $r^2 = 0.13$)	-
				<i>B*44:02</i> (D'=1.00 $r^2 = 0.22$)	-
				<i>C*05:01</i> (D'=0.97 $r^2 = 0.21$)	-
POL	747	rs2848713	7.47×10^{-17}	<i>B*51:01</i> (D'=0.95 $r^2 = 0.59$)	-
		rs17192463	2.41×10^{-11}	-	-
POL	834	rs2524073	2.23×10^{-14}	<i>C*05:01</i> (D'=1.00 $r^2 = 0.62$)	-
		rs9266462	1.12×10^{-14}	-	-
POL	837	rs2523575	4.22×10^{-17}	<i>B*18:01</i> (D'=0.97 $r^2 = 0.15$)	-
				<i>B*44:02</i> (D'=1.00 $r^2 = 0.31$)	-
				<i>C*05:01</i> (D'=0.96 $r^2 = 0.27$)	-
POL	839	rs2395029	7.33×10^{-12}	<i>B*57:01</i> (D'=1.00 $r^2 = 0.98$)	KTAVQMAVF (832-840)
TAT	32	rs16898614	4.26×10^{-21}	<i>C*12:03</i> (D'=0.88 $r^2 = 0.68$)	CCFHCQVC (30-37)
VIF	74	rs2395029	1.63×10^{-15}	<i>B*57:01</i> (D'=1.00 $r^2 = 0.98$)	-
VPR	32	rs3819299	7.31×10^{-14}	-	-

6. táblázat. Az ábraalírás a 4. táblázat alatt található.

már kondicionált SNP is szerepelt. Ezzel a módszerrel felderíthettük a független asszociációs szignálokat egy-egy kromoszómális régióból.

Asszociációvizsgálat a vírus mutációi és a vírusszint között

Az asszociációvizsgálatot lineáris regressziók sorozatával végeztük el a plink szoftvert használva. További korrekciós magyarázóváltozóként a páciensek nemét vontuk be a regresszióba.

Több statisztikai teszt elvégzése után végzett korrekció

Minden esetben Bonferroni-korrekciót alkalmaztunk az asszociációvizsgálatok szignifikanciájának megállapítására abban az esetben, ha több statisztikai teszt került elvégzésre.

Lókuszok közti kapcsoltság kifejezése

A *HLA*-allélok és az SNP-k közti kapcsoltságot Pearson-féle korrelációval és D' -mértékkel fejeztük ki, mindkettőt a plink szoftverrel számoltuk. A D' az adott haplotípus megfigyelt, illetve függetlenség esetén várt frekvenciájának a különbsége, az adott allélfrekvenciák esetén elérhető maximumra normalizálva.

A vizsgálat során felhasznált szoftverek

Az egyes részfeladatokhoz felhasznált szoftverek összesítését a 7. táblázat tartalmazza.

A szerző hozzájárulása a tanulmányhoz

A szerző a projekt munkatársaitól megkapta a vizsgálatban résztvevő páciensek genotípusadatait és HIV-szekvenciáit. Ő végezte el az összes fent ismertetett adatelemző és adatfeldolgozó lépést (kivéve a *HLA*-allélok imputációját). A felhasznált szoftvereket a 7. táblázat tartalmazza, az ezeken kívül szükséges szoftvereket a szerző saját maga készítette.

Részfeladat	Szoftver	Megjegyzés és hivatkozás
Diszkrét genotípusmátrix manipulálása	plink v.1.07	(Purcell, 2007)
Alléldózisokat tartalmazó genotípusmátrix manipulálása	saját	
Főkomponens-analízis a genotípusmátrixon	smartpca (eigenstrat)	(Price, 2006)
genotype imputation	mach	(Li, 2010)
bináris változók generálása	saját	
szekvenciaillesztés	Muscle v.3.8.31	(Edgar, 2004)
nukleotidszekvenciák fordítása	emboss transeq és saját	(Rice, 2000)
genotípuseloszlások összehasonlítása	saját	χ^2 -eloszlás innen (Commons Math 2012)
lineáris és logisztikus regressziók	plink v.1.07	(Purcell, 2007)
részfeladatok koordinálása számítógépklaszteren	saját	
grafikonok rajzolása	R, ggal, saját	(R Core Team, 2013; Seifert, 2009)

7. táblázat. Az egyes részfeladatokhoz felhasznált szoftverek.

3. A HIV-felülfertőzés gyakoriságának becslése rutin genotipizálás során gyűjtött vírusszekvenciákból

Második kutatási projektben a HIV-felülfertőződés gyakoriságát becsültem egy klinikai adatbázisban található nukleotidszekvenciákból. Tudomásunk szerint ezideáig a jelen munka a felülfertőződés gyakoriságának a legtöbb pácienset felhasználó becslése.

HIV-felülfertőződésről (szuperinfekció) akkor beszélünk, ha egy HIV-fertőzött páciens egy újabb vírustörzssel fertőződik meg (Smith, 2005a). Ennek lehetnek mind klinikai, mind epidemiológiai következményei a HIV-járvány szempontjából. Elsősorban ez teszi lehetővé a vírus evolúcióját elősegítő rekombinációt (Bretscher, 2004; Fraser, 2005; Mostowy, 2011; Althaus & Bonhoeffer, 2005) két különálló törzs között (Templeton, 2009). Klinikai szempontból pedig a vírusszint megemelkedéséhez, és a progresszió felgyorsulásához vezethet abban az esetben, ha a felülfertőző törzs könnyebben kerüli ki a CD8 immunválaszt (Yang, 2005; Cornelissen, 2012). Továbbá, egy terápiával alacsony szinten tartott gyógyszerérzékeny vírustörzset egy gyógyszerrezisztens vírus felülfertőzhet (Smith, 2005b).

Két egymástól különböző, azaz nem egymásból leszármazó vírustörzs megléte egy páciens mintái között felülfertőződésre utal. Hasonló szituáció adódhat két külön törzs szimultán transzmissziója során (Smith, 2005a), azonban az átvitt vírusrészecskék alacsony száma miatt (Keele, 2008) úgy gondoljuk, hogy az ilyen transzmisszió nagyon ritka, és a továbbiakban minden duális fertőzést felülfertőzésnek tekintünk.

A felülfertőződés gyakoriságának a becslését bonyolítja, hogy ez a jelenség lehet tranzienst, mikor az eredeti rezidens törzs nem szorul ki teljesen, hanem változó koncentrációban együtt él a felülfertőző törzs mellett (Templeton, 2009). Ilyenkor a két vírustörzs aránya jelentősen változhat az idő folyamán, hol az egyik kerül túlsúlyba, hol a másik. A felülfertőző törzs jelen lehet a szervezetben alacsony koncentrációban, esetleg egy idő után teljesen eltűnhet: ilyen esetekben a felülfertőződés észrevétlen maradhat a konszenzusszekvenciát eredményező populációszekvenálás vagy ritka mintavételek mellett. Ezért a HIV-felülfertőződés gyakoriságára vonatkozó irodalmi adatok jelentősen különböznek egymástól, amit a különböző vizsgálatok eltérő mintavételi stratégiájának tulajdonítunk. Egyes szerzők alacsony (<5%) prevalenciát találtak akár populációs szekvenálást alkalmaztak (pl.: (Gonzales, 2003)), akár következő generációs mély szekvenálást (pl.: (Tsui, 2004; Redd, 2012)). Ezzel ellentétben más tanulmányok jóval maga-

sabb prevalenciát mértek, például egy amerikai kohorszban klonális szekvenálást és gyakori mintavételt alkalmazva 23 felülfertőződést találtak az 58 vizsgált páciens között (Templeton, 2009). Jurriaans és munkatársai két esetet talált a 14 páciensből álló vizsgálati csoportban (Jurriaans, 2008), míg egy másik csoport 5 esetről tett jelentést 14 magas rizikójú kenyai asszony között (Piantadosi, 2008).

Nem tisztázott, hogy egy már kialakult vírusszint gátolja-e, hogy egy újabb vírus megtelepedjen a szervezetben (Nethe, 2005; Piantadosi, 2007; Ronen, 2013). Azonban a hatékony antiretrovirális terápia a vérben mérhető vírusszintet nagyon alacsony szintre csökkenti, amit a célsejt-populációk fokozatos regenerációja és a HIV-ellenes immunválasz gyengülése követ (Ogg, 1999). Így elképzelhető, hogy ez megteremti a környezetet egy új drogrezisztens vírustörzs inváziójához. Felvetettük ezért a lehetőséget, hogy a felülfertőződés a drogrezisztencia-mutációk (DRM) egy fontos terjedési útja is lehet.

A jelenlegi tanulmányban a felülfertőződés gyakoriságát becsüljük egy nagy európai populációban (a nemzetközi együttműködésben létrehozott Virolab és EuResist adatbázisok alapján), amelynek tagjai az esetek többségében terápia alatt állnak, és alacsony, de detektálható vírusszinttel rendelkeznek. Ez a kezelt populáció megfelelő annak eldöntésére, hogy a felülfertőződés hozzájárul-e a drogrezisztencia-mutációk terjedéséhez. Az adatbázisban a terápia beállításához szükséges rezisztenciatesztelés során keletkezett, rutin módon gyűjtött HIV-szekvenciák találhatóak. Egy olyan felülfertőződés, amelynek következtében az addig terápiával kontrollált vírusszint hirtelen megemelkedik, nagy valószínűséggel vezethet újbóli gyógyszerrezisztenciateszthez, ami pedig lehetővé teszi ennek az eseménynek a kimutatását az adatbázisból. A rutin módon gyűjtött genotípus adatok populációs szekvenálással készültek, így a tranziens és kisebb mértékben lévő felülfertőzéseket ezzel a módszerrel nem lehet kimutatni. Utóbbi eseteknek viszont kicsi a jelentőségük a drogrezisztencia-mutációk terjedésében. A rutin módon gyűjtött populációs szekvenálással kinyert HIV-szekvenciákról korábban már megmutatták, hogy alkalmasok a felülfertőződés kimutatására (Hué, 2004).

Ezek alapján, a tanulmány elsődleges célja a sikeres, a rezidens vírustörzs lecserélésével járó felülfertőzések kimutatása, és azok gyakoriságának becslése volt a Virolab és EuResist adatbázisokban. Utóbbi adatbázisok klinikai adatokat és rutin HIV-genotipizálás során keletkezett nukleotidszekvenciákat tartalmaznak német, belga, olasz, svéd és spanyol kórházakból.

3.1. Eredmények

3.1.1. Felhasznált adatok és leíró statisztikáik

4425 olyan páciens volt a Virolab és Euresist konzorciumok adatbázisában, akikről legalább két időpontból származtak vírusszekvenciák. Tőlük összesen 13816 szekvencia állt a rendelkezésünkre. A legtöbb szekvencia a *pol* régiót fedte (proteáz és/vagy részleges reverz transzkriptáz géneket); átlagos hosszuk 1026 nukleotid volt. Egy pácienshez tartozó szekvenciák számának mediánja 2 volt (minimum 2, maximum 29).

3653 páciens állt kezelés alatt a mintavétel idején vagy azt megelőzően. 772 páciensről kezelési előzmény nem állt rendelkezésre. A páciensek rizikócsoporthoz szerinti megoszlása: 1123 (25%) heteroszexuális, 910 (20%) homoszexuális, 1000 (22%) intravénás droghasználó páciens, 1236 (27%) ismeretlen, 156 (3%) egyéb. 1308 (29%) nőről állt rendelkezésünkre adat. 3878 páciensről volt legalább egy vírusszint adatunk, míg közülük 3552-nek volt legalább egyszer 10000 kópia/ml felett a vérbeli vírusszintje. A vérbeli vírusszintek mediánja az összes páciens összes mérésére vonatkoztatva 3,3 log₁₀ kópia/ml (IQR:1,9-4,4 log₁₀ kópia/ml). A CD4+ T-sejtek számának mediánja az összes páciens összes mérésére vonatkoztatva 349 sejt/μl (IQR:208-521 sejt/μl). A páciensek korának mediánja a legkorábbi rendelkezésre álló adat időpontjában 35,8 év volt (IQR: 29,3 - 40,9 év). Egy páciens legkorábbi és a legkésőbbi szekvenciái között eltelt idő mediánja 756 nap (IQR:367-1497 nap) volt. A HIV-szekvenciák 86% a B altípushoz tartozott (lásd 8. táblázat).

3.1.2. Felülfertőzés kimutatása filogenetikai módszerekkel

Módszerünk egy adott páciensről származó szekvenciák és a páciensről független háttérszekvenciák közös filogenetikai elemzésén alapszik. Egy ilyen filogenetikai fában, ha a páciens összes szekvenciáját magába foglaló legkisebb részfa nem monofiletikus az adott páciensre nézve (tehát tartalmaz háttérszekvenciákat is), akkor ennek a részfának a közös őse nem a páciensben volt (10.A ábra), vagy pedig a részfába tartozó idegen szekvenciák néhány transzmissziós lépésen keresztül mégis a páciensről származnak (10.B ábra).

Ezek alapján a vizsgálatunkban a következő két kritérium segítségével döntöttük el a filogenetikai fákról, hogy azok felülfertőződést tükröznek-e: először, a vizsgált páciens legalább két szekvenciája erősen támogatott részfat alkot idegen szekvenciákkal. Más szóval, legalább

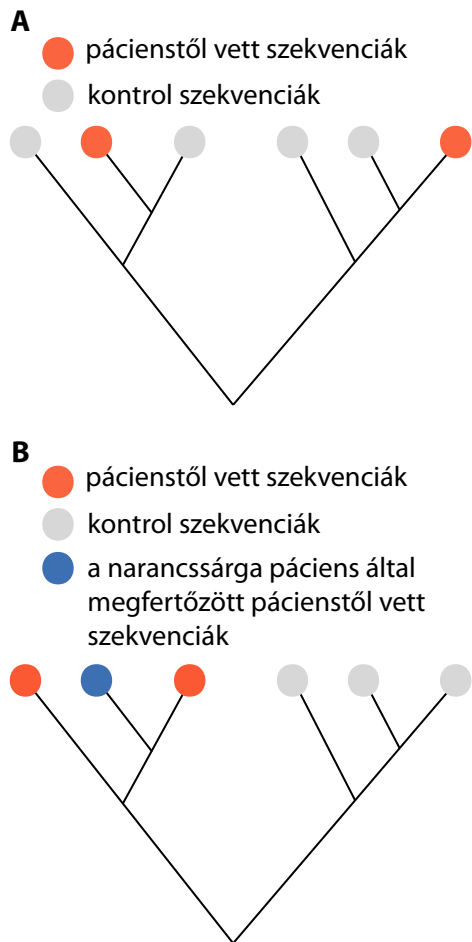
Altípus	Szekvenciák száma
HIV-1 Subtype B	12238
HIV-1 Subtype C	289
HIV-1 CRF 02_AG	275
HIV-1 Subtype A(A1)	241
HIV-1 Subtype F(F1)	159
HIV-1 Subtype G	97
HIV-1 Subtype A(01_AE)	86
HIV-1 CRF 06_CPX	28
HIV-1 Subtype G(02_AG)	27
HIV-1 Subtype D	16
HIV-1 Subtype J	12
HIV-1 Subtype A(A2)	11
HIV-1 F(F)	7
HIV-1 CRF 12_BF	6
HIV-1 CRF 01_AE	5
HIV-1 CRF 11_CPX	4
HIV-1 CRF 13_CPX	2
HIV-1 CRF 10_CD	2
Nem definiált	717

8. táblázat. Az elemzésünkben felhasznált HIV-szekvenciák altípus szerinti eloszlása. Az altípusokba való besoroláshoz a Rega szolgáltatást használtuk. (Alcantara, 2009; de Oliveira, 2005)

két megbízható szekvenciaklaszternek jelen kell lennie a fában, melyek legalább egy szekvenciát tartalmaznak mind a páciensről, mind a háttérszekvenciák csoportjából (lásd 3.4. fejezet és 10. ábra). Ezt a kritériumot azért választottuk, hogy elkerüljük annak a lehetőségét, hogy a fában kis megbízhatósággal elhelyezett szekvenciák miatt hibásan felülfertőzöttnek ítéljük egy pácienset.

A második feltételünk pedig, hogy a legkisebb részfa mérete legalább 50 legyen, amely magában foglalja a páciens összes szekvenciáját. Utóbbi kritériummal azt szeretnénk kizárni, hogy egy személyből eredeztethető transzmissziós láncot képező vírustörzs hamis szuperinfekciót jelezen. A konkrét küszöbértéket a beékelten idegen szekvenciák kétcsúcsú eloszlása alapján állítottuk fel (11. ábra és 3.4. fejezet). Továbbá ez a második kritérium szükséges azon esetek kikerüléséhez, amikor a páciensről származó minták esetleg megjelennek a háttérszekvenciák halmazában (pl. korábban anonim feltölthették a Genbankba).

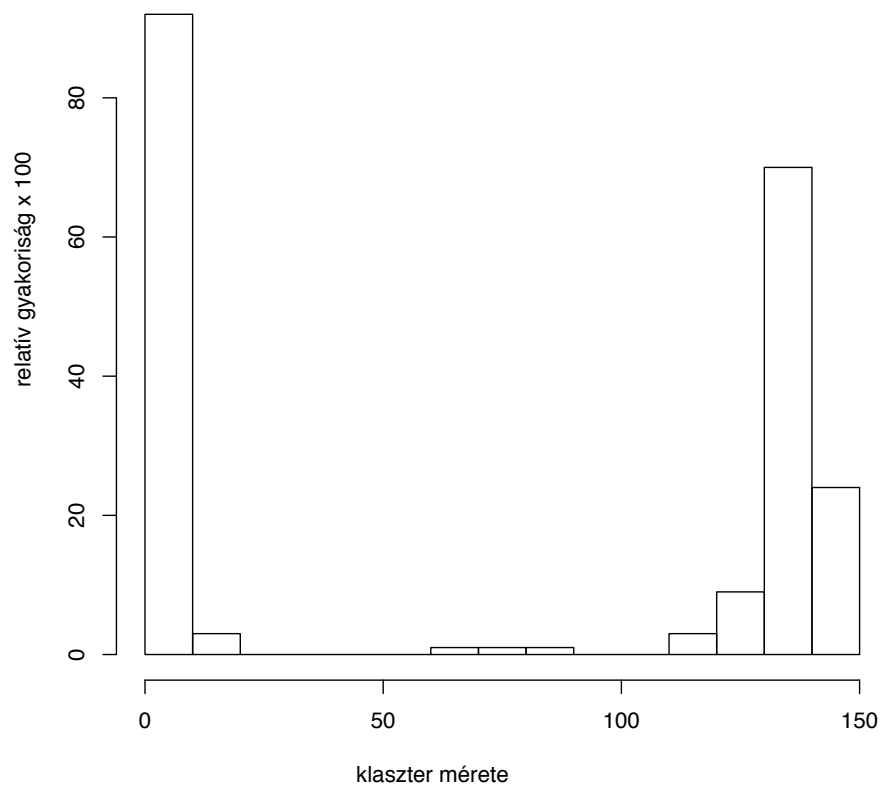
A filogenetikai fa építése számításigényes feladat, ezért valamiféle előszűrésre volt szükségünk, amivel elkülöníthetjük azokat a pácienseket, akiket alaposabban tovább vizsgáltunk.



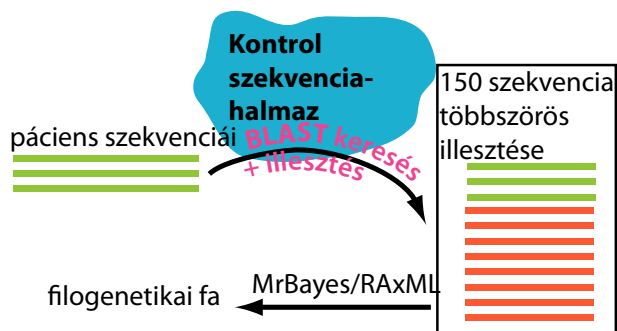
10. ábra. Sematikus filogenetikai fák, amelyek felülfertőződést, illetve transzmissziós klasztert ábrázolnak. A narancssárga levelek a vizsgált személy szekvenciáit jelzik, a szürke levelek idegen háttérszekvenciákat jeleznek. Az A ábra felülfertőződést illusztrál, a B ábra transzmissziós klasztert. A vizsgált egyén szekvenciái mindkét esetben polifiletikus leszármazásúak, azonban felülfertőződés esetében a páciens szekvenciáit befoglaló részfa nagyobb.

Ezért először előzetes elemzéseket végeztünk az adataink egy kis részén, amivel beállítottuk a filogenetikai rekonstrukció és a fák kiértékelésének optimális paramétereit. Ezeket az előzetes vizsgálatokat az eredmények után a Függelék fejezetben tárgyalom.

A 4425 páciens szekvenciáit külön-külön, egymástól függetlenül vizsgáltuk. Mindegyik pácienshez készítettünk egy filogenetikai fát előre rögzített számú háttérszekvenciát felhasználva. A háttérszekvenciákat 70374 HIV-szekvencia lokális (helyben létrehozott) BLAST adatbázisából választottuk ki úgy, hogy a páciens minden szekvenciájához azonos számú hasonló háttérszekvenciát választottunk a BLAST adatbázisból (lásd 12. ábra és 3.4. fejezet). A filogenetikai rekonstrukcióhoz a RAxML (egy gyors és megbízható implementációja a maximum likelihood



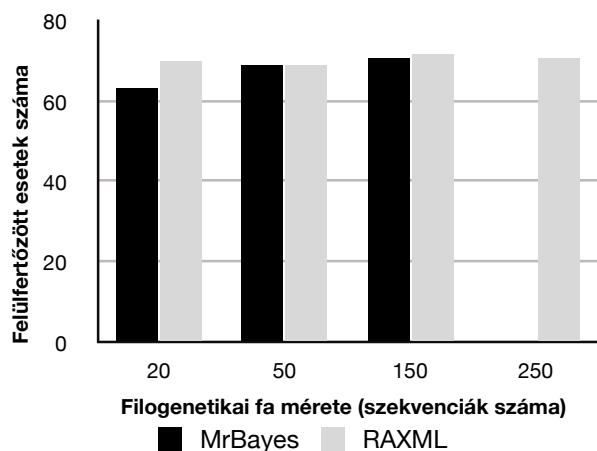
11. ábra. A páciens mintáiból származó szekvenciák közé ékelődő háttérszekvenciák száma azokban a páciensekben, akiknek a szekvenciái polifiletikus részfát alkottak. A beékelte háttérszekvenciák száma jól elkülönülő módon vagy kevés (kisebb mint 20), vagy sok (nagyobb mint 50). Ezért feltételeztük, hogy azokban az esetekben ahol kevés idegen szekvencia ékelődik be, ott egymást fertőző páciensektől származó közel rokon szekvenciák által alkotott klaszterről lehet szó. A filogenetikai fák 150 szekvenciával készültek.



12. ábra. A kettős fertőzés filogenetikai fákon alapuló kimutatása. Minden pácienshez kiválasztottuk egy nagy BLAST adatbázisból a saját szekvenciáihoz legközelebb eső háttérszekvenciákat, majd az így nyert háttérszekvenciákat és a páciens szekvenciáit felhasználva filogenetikai fát készítettünk.

filogenetikai módszernek (Stamatakis, 2006)) és a MrBayes (amely a jóval lassabb, de alaposabb bayes-i megközelítést alkalmazza (Ronquist, 2012)) szoftvereket használtuk. Először 170 előre kiválasztott páciensen összehasonlítottuk a két szoftver viselkedését, továbbá beállítottuk az optimális megbízhatósági küszöbértéket a fák kiértékeléséhez. Ezt a 170 pácienset a Függlekben tárgyalt előzetes elemzések során választottuk ki úgy, hogy ez a halmaz felülfertőzést gyanús esetekben fel volt dúsulva.

Ezen vizsgálatok alapján megállapítottuk, hogy a nem monofiletikus leszármazással bíró esetek száma (felülfertőződés, vagy transzmissziós klaszter) csökken, ahogy növeljük a kiértékelés során használt megbízhatósági küszöböt (9. táblázat). Azonban ez a küszöbérték beállítható úgy, hogy a két szoftvert használva azonos páciensek halmazát kapjuk eredményül. Azaz az irodalomban gyakran alkalmazott 0,95-ös bayesian küszöbvel kapott polifiletikus leszármazással bíró egyének halmaza reprodukálható RAxML segítségével 0,6-os küszöböt alkalmazva (9. táblázat). Emiatt úgy ítéltük, hogy biztonsággal használhatjuk a jóval gyorsabb RAxML szoftvert az összes páciens vizsgálata során. Ahhoz, hogy megtaláljuk a háttérszekvenciák halmazának optimális méretét, mind MrBayes, mind RAxML segítségével megismételtük az elemzéseket 20, 50, 150 és 250 szekvenciát tartalmazó halmazokon (ugyanazon 170 páciens adatait felhasználva). Nem volt látható összefüggés a fa mérete és a felülfertőződéssel gyanúsítható páciensek száma között (lásd 13. ábra), és a felülfertőződéssel gyanúsított páciensek is azonosak voltak a különböző esetekben (10. táblázat). Ez azt mutatja, hogy a BLAST kereséssel



13. ábra. Előzetes vizsgálat 170 páciensben: a grafikon a felülfertőződéssel gyanúsítható páciensek számát ábrázolja a filogenetikai fa méretének (szekvenciák számának) a függvényében. Nem látható számottevő növekedés a talált felülfertőződések esetében a háttérszekvenciák számának növelése során.

megbízhatósági küszöb		Metszet	csak RAXML	csak MB	RAXML összes	MB összes
MrBayes	RAXML					
0,7	0,6	70	2	4	72	74
0,7	0,7	63	2	11	65	74
0,7	0,95	40	2	34	42	74
0,8	0,6	70	2	3	72	73
0,8	0,7	63	2	10	65	73
0,8	0,95	40	2	33	42	73
● 0,95	0,6	69	3	2	72	71
0,95	0,7	62	3	9	65	71
0,95	0,95	40	2	31	42	71

9. táblázat. A bayesian és maximum likelihood módszerek eredményeinek összevetése, és a megfelelő megbízhatósági küszöb kiválasztása. A táblázatban a felülfertőzéssel gyanúsítható esetek száma szerepel különböző szoftverrel és különböző megbízhatósági küszöbvel számolva. Minden esetben 150 szekvenciáról készült a filogenetikai fa. Az eredmények 170 páciens adatait felhasználva készültek. A teli kör jelzi azt az értéket, amelyet végül a további vizsgálatokban használtunk.

faméret, küszöb	RAxML 150 0,6	MrBayes 150 0,95	RAxML 20 0,6	MrBayes 20 0,95	RAxML 250 0,6	RAxML 50 0,6	MrBayes 50 0,95
RAxML 150 0,6	72						
MrBayes 150 0,95	69	71					
RAxML 20 0,6	67	66	70				
MrBayes 20 0,95	61	62	62	63			
RAxML 250 0,6	69	67	66	59	71		
RAxML 50 0,6	69	66	65	59	67	69	
MrBayes 50 0,95	67	68	64	62	65	66	69

10. táblázat. A bayesian és maximum likelihood módszerek eredményeinek összevetése és a faméret változtatásának hatása az eredményekre. A táblázat celláiban a különböző famérettel, illetve szoftverrel készült vizsgálatok eredményeinek a metszete szerepel. A fejlécek a filogenetikai szoftvert, a faméretet, és a megbízhatósági küszöböt mutatják.

kiválasztott háttérszekvenciák számát alacsonyan tarthatjuk.

Ezen megelőző, „kalibrációs” vizsgálatok alapján a 4425 páciens végső elemzését a gyorsabb futásidő miatt RAxML szoftverrel végeztük el 150 szekvenciát tartalmazó szekvencia-halmazokon, míg a kiértékelés során 0,6-os megbízhatósági küszöbértéket alkalmaztunk (ez konzisztens a bayesian fák 0,95-ös küszöbével). A 4425 páciens közül 201 szekvenciái alkottak polifiletikus részfat. Felülfertőződéssel gyanúsítható esetből 107-et találtunk, míg 94 esetben közel rokon vírustörzsek által alkotott transzmissziós klasztert gyanítottunk (azaz a páciens szekvenciáit befoglaló minimális részfa mérete 50-nél kisebb volt - lásd 3.4. fejezet). A 107 eset közül az egyik hamisan pozitívnak bizonyult, ennek a páciensnek az egyik mintája a PCR reakció során rekombináldott egy labortörzssel - így ezt eltávolítottuk az eredmények összesítéséből és a további validációs vizsgálatokból.

Validáció Eredményeink megerősítése céljából az eltárolt minták egy részének újbóli szekvenálását kezdeményeztük a közreműködő klinikai/virológiai partnereknél. Ideális esetben mind a 107 felülfertőződéssel gyanúsított páciens mintáit ellenőrizni kellett volna, ám ennek anyagi és technikai akadályai voltak (sok esetben az amplifikáció nem sikerült a régi mintákból). Végül 14 felülfertőződéssel gyanúsított, és 4 kontroll esetben került sor validációs szekvenálásra, amelynek eredményeit ezután analizáltam. A felülfertőződéssel gyanúsított esetekben a felülfertőzést közvetlenül megelőző és közvetlenül követő mintákat szekvenálták újra. Két esetben az eredeti minták nem álltak rendelkezésre, itt egy időben közeli alternatív mintát használtak.

A *pol* gén részleges amplifikációja és szekvenálása történt 6 esetben, 4 esetben az *env* C2-C4 régiója lett szekvenálva, míg 8 esetben mindkettő.

Az új szekvenciák elemzését az előző fejezetben leírtakhoz hasonlóan végeztük. Ezek alapján két esetben tudtuk megerősíteni a felülfertőződést a 14 gyanús páciens közül (11. táblázat). Kilenc esetben az új *pol* szekvenciák monofiletikus csoportot alkottak, megcáfolva a korábbi eredményeket. Mind a kilenc esetben az új szekvenciák szoros monofiletikus csoportot alkottak a páciens eredeti szekvenciáival, egy eredeti szekvenciát leszámítva. Ez arra utal, hogy az az eredeti szekvencia, amely a felülfertőződésre utaló jelet okozta nem a páciensről származik, hanem felcserélt minták következménye. A kilencből három esetben az *env* gén C2-C4 variábilis régióit is sikerült szekvenálni, ezek megerősítették a fenti eredményeket.

Két esetben csak az *env* génből sikerült az amplifikáció, ezek is monofiletikus csoportot alkottak, ami arra utal, hogy valójában nem történt felülfertőzés, azonban ebben a két esetben a *pol* szekvenciák validációjának hiányában nincs bizonyítékunk a minták elcserélésére.

Végül egy esetben az új szekvenciák elemzése polifiletikus leszármazást talált, de a legkisebb részfa mérete kicsi volt, így elképzelhető, hogy ez egy közeli szekvenciákból álló lokális transzmissziós klaszter (a szekvenciák földrajzi eredete megegyezett). Ennek a páciensnek az eredeti szekvenciáit MrBayes programmal megvizsgálva megerősítettük, hogy a RAxML fában néhány háttérszekvencia rossz helyen volt, így hibásan megnövelve a legkisebb részfa méretét. Az újonnan amplifikált *env* szekvenciák monofiletikus csoportot alkottak ebben az esetben.

A két megerősített felülfertőződött páciens esetében mind a *pol*, mind az *env* szekvenciák szuperinfekcióra utaltak, igaz, az *env* szekvenciák esetében a megbízhatósági értékek éppen alatta maradtak a kritériumnak. Az egyik páciensről 7 szekvencia is rendelkezésünkre állt, amelyek közül hat közös klasztert alkotott, és egy szekvencia reprezentálta a különálló vírustörzset. Ezt a különálló szekvenciát időben követték is és meg is előzték a nagyobb klaszterhez tartozó minták: ebből a mintázatból tranziens felülfertőzésre következtethetünk. A páciens CD4 sejt-száma 430 sejt/ μ l-ről 125 sejt/ μ l-re esett a felülfertőződés lehetséges időintervallumában, és a CD4 sejt-szám később sem állt vissza. A szekvenciák drogrezisztenciára való hajlamát megvizsgálva nem tapasztaltunk megemelkedett rezisztenciát a felülfertőző törzsben azon hatóanyagok ellen, amit a páciens akkor éppen kapott (12. táblázat). A másik felülfertőzött páciensről csak két minta állt rendelkezésünkre, megemelkedett drogrezisztenciát ott sem tapasztaltunk. Ennek

Páciens	Gén	Eredmény	Felcserélt minta dátuma	Megjegyzés
Case1	POL	monofiletikus	10/26/2000	
Case2	POL ENV	monofiletikus	02/19/2004	
Case3	POL ENV	monofiletikus	04/19/2004	a felülfer- tőződést közvetlenül megelőző minta nem állt ren- delkezésre, ehelyett egy időben közelit használtak
Case4	POL ENV	monofiletikus	05/05/2004	
Case5	ENV	monofiletikus		az eredeti minták he- lyett alterna- tív mintákat amplifikáltak
Case6	ENV	monofiletikus		
Case7	POL	monofiletikus	01/25/2006	
Case8	POL ENV	közel rokon szekvenciák által alkotott polifiletikus klaszter		
Case9	POL ENV	monofiletikus	19/07/1999	
Case10	POL	monofiletikus	04/20/1998	
Case11	POL	monofiletikus	09/10/1997	
Case12	POL	monofiletikus	11/24/1998	
• Case13	POL ENV	nem monofiletikus		
• Case14	POL ENV	nem monofiletikus		
Control1	POL ENV	monofiletikus		
Control2	POL ENV	monofiletikus		
Control3	ENV	monofiletikus		
Control4	ENV	monofiletikus		

11. táblázat. 14 felülfertőződést mutató páciens és 4 kontroll páciens validációja. A mintákból a 3.4. fejezetben ismertetett módon amplifikálták a *pol* és *env* gének megfelelő részeit. A teli körök a megerősített felülfertőzéseket mutatják.

Regime	Case13		Case14		DBC+	
	Residens törzs	Felül-fertőző törzs	Residens törzs	Felül-fertőző törzs	Residens törzs	Felül-fertőző törzs
3TC	HR	HR	PL	S	●S	HR
ABC	HR	LL	R	S	S	●HR
ATV/r	S	LL	S	S	S	S
AZT	HR	R	●HR	●LL	●PL	HR
D4T	●HR	●R	HR	PL	LL	●HR
DDI	●HR	PL	●R	●S	PL	HR
DLV	S	S	S	S	S	S
DRV/r	S	S	S	S	S	S
EFV	S	●S	S	S	S	S
ETR	S	S	S	S	S	S
EVG	S	S	S	S	S	S
FPV/r	S	LL	S	S	S	S
FTC	HR	HR	PL	S	S	HR
IDV/r	S	R	●S	●S	S	S
LPV/r	S	PL	S	S	S	S
NFV	●S	●HR	S	S	S	S
NVP	S	S	S	S	S	S
RAL	S	S	S	S	S	S
SQV/r	S	R	S	S	S	S
TDF	R	PL	R	S	PL	●LL
TPV/r	S	S	S	S	S	S

12. táblázat. A két validált felülfertőződést hordozó páciens és a magas DBC számmal rendelkező páciens (lásd 3.1. fejezet) drogrezisztencia-pontjai. A pontokat a stanfordi HIVdb programmal számítottuk (Rhee, 2003), majd kategóriákba soroltuk: HR - highly resistant (nagyon rezisztens), R - resistant (rezisztens), LL - low level resistance (alacsony szintű rezisztencia), PL - potential low level resistance (potenciális alacsony szintű rezisztencia), S - susceptible (fogékony). A teli körök azokat a hatóanyagokat jelölik, amelyeket a beteg az adott időszakban szedett.

a páciensnek a CD4 sejtszáma stabil volt.

A négy validált kontroll páciens közül egyik se bizonyult felülfertőzöttnek.

Degenerált bázispárok számán alapuló módszer Végül Cornelissen módszerét alkalmazva (Cornelissen, 2007), megszámoztuk a szekvenciákban található degenerált bázisokat jelentő betűk számát (egyes helyeken erre mixture-ként utalnak, itt a továbbiakban degenerate base count, DBC néven utalok rá). Populációs szekvenálás során akkor illesztünk degenerált kódot a szekvenciába, ha a kapilláris gélelektroforézis spektrumán egynél több bázispár ad csúcsot

egy adott pozícióban. Ez arra utal, hogy két vírustörzs is amplifikálódott, és a kevert minta került szekvenálásra. Az, hogy a második amplifikált vírustörzs valódi-e, vagy szennyeződés, a végső szekvencia alapján nem lehet eldönteni. Cornelissen és munkatársai azt találták, hogy azokban a mintákban, ahol a DBC szám 44-nél magasabb volt, ott 73% eséllyel találtak később felülfertőződést klonális szekvenálást alkalmazva.

Az általunk vizsgált adatbázisban 79 páciensnek volt legalább egy olyan szekvenciája, ahol a DBC szám 44-nél magasabb volt (5 páciensnek két ilyen szekvenciája volt, és egynek 3). A 79 páciens közül hatot gyanúsítottunk felülfertőződéssel a filogenetikai vizsgálatunk alapján. Egy esetben (Case2 a 11. táblázatban) a páciens mintáit újraszekvenálták, ami felfedte az eredeti minták elcserélését - érdekes módon, éppen az a minta származott egy idegen páciensből, amelyikben a DBC szám magas volt. Ebben az esetben mind a *pol*, mind az *env* szekvenciák elemzése erős monofiletikus leszármazásra utalt, így a felülfertőződést itt kizártuk.

A 79 páciens közül 23 esetben a magas DBC számmal bíró szekvenciát időben megelőzték, és követték is alacsony DBC számú minták, ami arra utal, hogy a páciensben található víruspopuláció egy tranzien időszakban heterogén volt. Az egyik ilyen esetben a magas DBC mintát határoló szekvenciák két külön szekvenciaklasztert alkottak a filogenetikai fán, ami nagyon erősen arra utal, hogy ez a páciens valóban felülfertőzött volt. Ezt az esetet ugyan nem tudtuk újbóli szekvenálással megerősíteni, de mind a mi filogenetikai elemzésünk, mind a degenerált bázispárok számolása sikeres, a vírustörzsek lecserélődésével végződő szuperinfekcióra utalt. Ebben az esetben a magas DBC számmal rendelkező minta az átmeneti állapotot jelentheti. Továbbá, ebben a páciensben a drogrezisztencia jelentős mértékben emelkedett a felülfertőzés után azokkal a drogokkal szemben, amelyet a páciens éppen szedett (lásd 12. táblázat).

A többi 22 esetben a filogenetikai elemzés nem támasztotta alá a degenerált bázispárok magas száma által felvetett gyanúját a felülfertőződésnek. Tranziens, vagy kisebbségben maradó felülfertőző törzs jelenlétét viszont nem tudjuk kizárni.

3.2. Diskusszió

A tanulmány során létrehoztunk egy „protokollt”, amelynek segítségével gyorsan és megbízhatóan kimutatható a felülfertőződés nagy adatbázisokban. Módszerünk publikus és privát adatbázisokban található vírusszekvenciákon alapszik, hiszen az egymástól független leszármazással bíró vírustörzsek kimutatásához elegendően nagy mintára van szükségünk a szekvenciák

mögött meghúzódó járványból. A vizsgálati módszerünket alaposan ellenőriztük, a felmerült döntési pontokban a lehetséges opciókat mind megvizsgálva jutottunk ahhoz a „protokollhoz”, amivel végül az eredményeket elkészítettük (lásd 3.5 és 3.1 .fejezetek).

Filogenetikai elemzésünk 107 páciens felülfertőződésére mutatott bizonyítékot a 4425 páciens közül. Azonban a lefagyasztott minták újraszekvenálása ezt 14 esetből csupán két esetben erősítette meg. Ennek következtében nem állt módunkban pontos becslést készíteni a felülfertőződés prevalenciájáról a populációban, azonban azt kijelenthetjük, hogy az alacsony (107/4425 is csak 2%-ot jelentene, és a validáció alapján a legtöbb eset valószínűleg elcserélt mintából származik).

Ez az eredmény nem alkalmazható a szuperinfekció gyakoriságának becslésére kezelés alatt nem álló páciensek esetében, de azt mutatja, hogy ART alatt a felülfertőződés viszonylag ritka jelenség. Az egész tanulmány során egy esetet találtunk, ahol a megerősített felülfertőzés megnövekedett drogrezisztenciával járt. Ez azt mutatja, hogy a felülfertőzés nem játszik fontos szerepet a drogrezisztencia terjedésében.

Vizsgálatunkat a rutin klinikai ellátásban gyűjtött szekvenciákra alapoztuk, ez némileg torzított mintavételezéshez vezethetett. Csak azokat a pácienseket tudtuk ugyanis vizsgálni, akiktől legalább két minta állt rendelkezésre, és elképzelhető, hogy a második vagy harmadik mintavétel oka a vírusszint megemelkedése, míg a stabilan alacsony vírusszinttel bíró páciensek mintáit kisebb valószínűséggel vetik alá ismételt rezisztenciatesztnek. Egy ilyen torzítás azonban csak növelheti a felülfertőződés becsült gyakoriságát, amely így is alacsonynak bizonyult.

Továbbá a vizsgálati populációban igen magas (82%) volt a terápia alatt állók száma. Ezért az eredményeinket nem lehet ART alatt nem álló populációra kivetíteni. Azonban, ha a drogrezisztencia-mutációk terjedését a felülfertőzés segítené, akkor azt a terápia alatt állók csoportjában könnyebben észrevettük volna, hiszen a kezelés szelekciós előnyt biztosít a drogrezisztens törzsnek.

A szekvenciában megjelenő degenerált bázispárok magas száma két különálló vírustörzs jelenlétére utal a mintában (Cornelissen, 2007). Ezt a heterogén állapotot követheti a rezidens vírustörzs lecserélődése. Vizsgálatunkban 23 esetet találtunk, ahol egy magas DBC számmal bíró mintát időben megelőz, és követ is egy vagy több homogén minta. Ezek közül csupán egy esetben tudtuk megerősíteni az eredeti vírustörzs lecserélődését, a többi 22 eset legfeljebb tranziens felülfertőződés lehetett. A tranziens felülfertőződés azonban nem valószínű, hogy hozzájárul a

drogrezisztencia terjedéséhez.

A vizsgálatunk során nem tettünk különbséget felülfertőződés és az általánosabb kettős fertőzés között. Kettős fertőzés lehet a felülfertőzés mellett az az eset is, amikor két különböző leszármazással bíró vírustörzs egy transzmissziós esemény során (esetleg röviddel egymás után) fertőzi meg az addig HIV-negatív személyt (Smith, 2005b). Azonban az alapító vírusok nagyon alacsony száma (Keele, 2008) miatt ez a jelenség minden bizonnyal igen ritka. Továbbá ehhez az is szükséges, hogy a fertőző páciensben mindkét vírustörzs hasonló, és elegendően nagy koncentrációban legyen jelen a sikeres kettős fertőzéshez. Ilyen ritka események összejátszását elhanyagolhatónak tekintjük, és úgy gondoljuk, hogy az azonosított kettős fertőzések valószínűleg felülfertőződést tükröznek.

A részleges *pol* szekvenciák alapján a sikeres, a rezidens törzs lecserélődésével járó felülfertőződést voltunk képesek kimutatni (Hué, 2004; Lemey & Vandamme, 2005). Azonban egyes esetekben a felülfertőzést követően az új törzs rekombinálódhat a rezidens törzssel, és az új rekombináns törzs kiszoríthatja a többit. A mi vizsgálatunk ezeket az eseteket csak akkor képes kimutatni, ha az új rekombináns törzs *pol* génje a felülfertőző vírustól származik. A rekombináns formák megbízható felderítéséhez teljes genomi szekvenciákra lenne szükség.

3.3. Összefoglalás

Eredményeinkből levonjuk a következtetést, hogy a rutin módon gyűjtött HIV-szekvenciák alkalmasak a felülfertőződés kimutatására, azonban körültekintő validáció szükséges, hogy az elcserélt mintákból származó hamis eseteket ki lehessen zárni. Vizsgálati csoportunk (krónikusan fertőzött, antivirális kezelés alatt álló páciensek) ideális volt ahhoz, hogy a felülfertőződés szerepét tisztázhassuk a drogrezisztencia terjedésében. A felülfertőződés ritkasága ebben a populációban tehát arra utal, hogy az nem játszik meghatározó szerepet a drogrezisztencia terjedésében.

3.4. Felülfertőzés kimutatásában használt módszerek

Páciens és szekvenciaadatok

Az EU 6. Keretprogram támogatásával létrehozott Eurezist és Virolab adatbázisokat elemeztük (témavezetőm vezetésével az ELTE a Virolab konzorcium tagja volt), amelyekben klinikai adatok, az antivirális terápia előzményei és HIV-szekvenciák találhatók német, spanyol, olasz,

belga és svéd kórházakból. Azon betegek adatait vizsgáltuk, akikből legalább két különböző időpontból rendelkezésre állt vírusszekvencia.

Kettős fertőzés kimutatása

Annak eldöntésére, hogy egy páciens két különböző eredetű vírussal fertőzött-e, a következő módszert használtuk. Filogenetikai fát készítettünk a páciensről vett vírusszekvenciák (továbbiakban a páciens saját szekvenciái) és bizonyos számú (lásd később) a páciensről független vírusszekvenciák (továbbiakban háttérszekvenciák) együtteséből. Ha a saját szekvenciák nem alkottak monofiletikus részfat, és a páciens saját szekvenciái közé elegendő számú (lásd később) háttérszekvencia ékelődött be, akkor az adott esetet kettős fertőzésként kezeltük. A háttérszekvenciákat BLAST kereséssel (Lemey, 2005; Altschul, 1990) választottuk ki a rendelkezésünkre álló összes HIV-szekvenciából, azaz a Virolab és az Euresist adatbázisokból és a Los Alamos-i publikus adatbázisból (kivéve az adott egyén szekvenciáit). A lehetséges háttérszekvenciák halmazában 13816 szekvencia származott az ezen tanulmány alapját képező szekvenciákból, 12314 szekvencia származott a Virolab és Euresist adatbázisok egyéb adataiból, és 44244 szekvenciát töltöttünk le a Los Alamos-i adatbázisból. Minden pácienshez készítettünk egy előre definiált méretű szekvenciahalmazt, azaz kiválasztottuk minden szekvenciájához a legmagasabb BLAST pontot adó háttérszekvenciákat oly módon, hogy a páciens összes szekvenciája azonos számú háttérszekvenciával legyen reprezentálva. Ezeken a szekvenciahalmazokon végeztük el a filogenetikai elemzést minden páciensre külön-külön.

Az alábbiakban részletezem az előbb ismertetett módszer egyes lépéseit.

BLAST adatbázis készítése Letöltöttünk 44244, a HXB2¹¹ számozás szerinti 2253-3290. pozíciók közötti HIV-1 B szekvenciát a Los Alamos-i adatbázisból (<http://www.hiv.lanl.gov/>). Ehhez hozzávettük a Virolab és Euresist adatbázisokban található ugyanezt a régiót lefedő 26130 szekvenciát. A nukleotidszekvenciák illesztését Muscle v3.8.31 szoftverrel végeztük (Edgar, 2004). A *pol* gén konzerváltságát kihasználva, ezeket a szekvenciákat 1000 darabos kötegekben illesztettük, majd ezeket a részleteket összefésültük a HXB2 referenciaszekvencia segítségével. Az ismert rezisztenciapozíciókat (Rhee, 2003) eltávolítottuk (13. táblázat), és minden legalább 50%-ban üres (gap-et tartalmazó) oszlopot kitöröltünk. Az így kapott illesztést

¹¹A HXB2 egy referenciaként használt HIV-szekvencia <http://www.hiv.lanl.gov/>

használtuk a további BLAST keresésekben, amelyhez egy helyi BLAST adatbázist készítettünk. Az eredmények validációjára szolgáló *env* szekvenciák elemzéséhez hasonló módon készült a BLAST adatbázis, ehhez 13725 szekvenciát töltöttünk le Los Alamosból. Az *env* illesztést aminosav szekvenciákon végeztük Muscle-al, majd visszafordítottuk azt nukleinsav szekvenciákra. A fordításhoz az Emboss szoftvercsomag transeq szoftverét (Rice, 2000), illetve saját szoftvereket használtunk.

Filogenetikai fa készítése Filogenetikai fa készítéséhez RAxML v7.2 (Stamatakis, 2008; Stamatakis, 2006) és MrBayes v3.1.8 (Ronquist, 2012) szoftvereket használtunk. Mindkét esetben az általános, időben megfordítható szubsztitúciós modellt (General Time Reversible, GTR) feltételeztük, ennek (nuisance) paramétereit az adatokból becsültük. A RAxML esetében az elágazások megbízhatósági értékeit (branch support values) bootstrap replikátumok segítségével számoltuk. A bootstrap replikátumok számát a szoftver dinamikusan állította be (-autoMR kapcsoló a RAxML-ben). MrBayes esetében 10000000 lépést végeztünk 4 láncon 3 párhuzamos futással. A szekvenciák közti GTR szubsztitúciós modellen alapuló távolságot faépítés nélkül maximum likelihood módszerrel becsültük a PAUP szoftverben (Swofford, 2003).

Filogenetikai fák kiértékelése Egy általam erre a célra ruby és scala nyelven készített program végezte a filogenetikai fák kiértékelését, azaz annak eldöntését, hogy a fában a vizsgált páciens szekvenciái monofiletikus csoportot alkotnak-e vagy sem, és ha nem, akkor a páciens szekvenciái közé ékelődött háttérszekvenciák száma elér-e egy általunk meghatározott értéket (lásd később). A fából minden olyan belső csomópontot eltávolítottunk, amely nem érte el a kívánt megbízhatósági értéket. Ez az érték a maximum likelihood fák esetében 0,6, a bayesian fák esetében 0,95 volt. A következő két kritériumnak kellett egy fának megfelelnie ahhoz, hogy az adott esetet kettős fertőzésnek tekinthessük: I. A páciens legalább kettő szekvenciája közelebbi rokonságot mutat idegen szekvenciákkal, mint a páciens egyéb szekvenciáival (lásd 10. ábra). Ez a kritérium erősebb, mintha csupán a polifiletikus összetételt követelnénk meg: ahhoz elég lenne, ha a páciens egy szekvenciája közelebb klasztereződne egy idegen szekvenciával, mint a többivel. Ezt az erős kritériumot azért vezettük be, hogy a páciens megbízhatatlan módon csoportosuló szekvenciái ne okozzanak esetlegesen hibásan pozitív eredményt. Továbbá II.: a legkisebb olyan részfa mérete, amely magában foglalja a páciens összes szekvenciáját, legyen

Gén	Pozíció
RT	41
RT	65
RT	67
RT	69
RT	70
RT	74
RT	75
RT	77
RT	115
RT	116
RT	151
RT	184
RT	210
RT	215
RT	219
RT	100
RT	101
RT	103
RT	106
RT	179
RT	181
RT	188
RT	190
RT	225
RT	230
PR	23
PR	24
PR	30
PR	32
PR	46
PR	47
PR	48
PR	50
PR	53
PR	54
PR	73
PR	76
PR	82
PR	83
PR	84
PR	85
PR	88
PR	90

13. táblázat. Drogrezisztenciapozíciók listája, amelyeket eltávolítottunk a szekvenciaillesztésekből.

legalább 50. Utóbbi kritériumot azért állítottuk fel, mert úgy gondoljuk, hogy kis számú közbeékeltszekvencia inkább transzmissziós klaszterre (a vizsgált páciensből eredő fertőzési láncra) utal, mintsem igazi felülfertőzésre. A küszöbértéket (50) a beékeltszekvenciák eloszlása alapján állítottuk fel, amely kétszűcsű eloszlást mutatott (lásd Eredmények és 11. ábra).

Validációs szekvenálás Az eredmények megerősítése végett a minták újbóli szekvenálását kezdeményeztünk azokban az esetekben, ahol a vizsgálatunk kettős fertőzés gyanúját vetette fel. Validáció céljából ugyanezt a *pol* régiót és az *env* gén C2-C4 régióját (hxb2 7021-7560) szekvenáltattuk újra azokból a mintákból, amelyek az analízisünk alapján feltételezett felülfertőzést közvetlenül megelőzték, illetve követték. (Rachinger, 2010; Smith, 2005b; Campbell, 2009; Cornelissen, 2007) Az új szekvenciákon ugyanezt az elemzést végeztük el, mint az eredeti szekvenciákon. Hogy eldönthessük, hogyan viszonyul egymáshoz az új és a régi szekvencia, a filogenetikai fák készítésénél a háttérszekvenciák mellé mind az új, mind a régi szekvenciákat beválogattuk.

HIV-1 altípus meghatározása A szekvenciák altípusát a Rega v2 szoftverrel határoztuk meg (Alcantara, 2009; de Oliveira, 2005).

Drogrezisztencia kvantifikálása A szekvenciák drogrezisztenciájának megállapításához a Stanford HIVdb Sequence Analysis szolgáltatást használtuk (<http://hivdb.stanford.edu>, (Rhee, 2003)). Magasabb pontszám erősebb rezisztenciát és gyengébb virológiai választ jelent a kezelés során.

3.5. Függelék

3.5.1. Szekvenciatávolságon alapuló módszerekkel nem lehet megbízhatóan kimutatni a kettős fertőzést

Kettős fertőzés kimutatásához el kell dönteni, hogy a páciensből származó HIV-szekvenciák közös őse jelen volt-e az adott személyben. Ha nem, az felülfertőződést, vagy kettős fertőzést jelent. Ez a döntés nem lehet pontos, mivel nem ismerjük a világ összes HIV-szekvenciáját, csupán ritka mintavételek állnak rendelkezésünkre azok törzsfájából. Eleinte megpróbáltuk elkerülni a filogenetikai fa építését a nagy számításigény miatt, ezért kipróbáltunk néhány divergencián alapuló módszert.

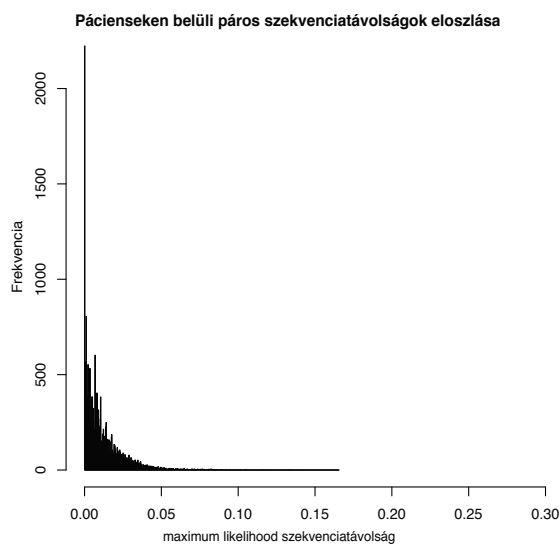
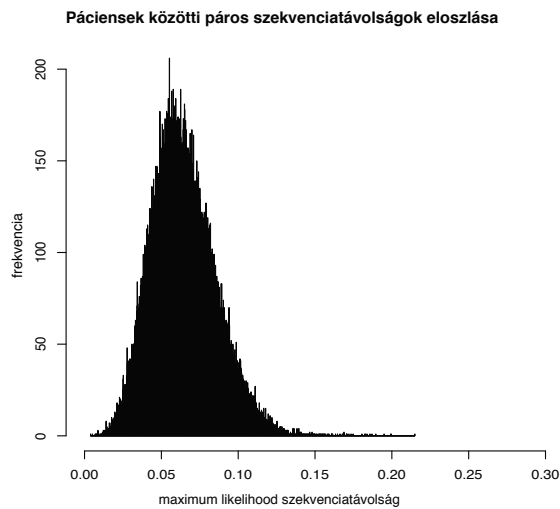
Állandó sebességgel haladó molekuláris órát feltételezve felállíthatnánk egy küszöbértéket a szekvenciák evolúciójának sebességére az átlagos gazdán belüli divergencia alapján. Azon szekvenciák utalhatnak felülfertőződésre, amelyek jobban eltérnek egymástól, mint az átlagos egy gazdán belüli divergencia. Azonban, az egy gazdán belüli és gazdák között páros szekvenciatávolságok eloszlásai átfednek, lehetetlenné téve két szekvencia egységnyi idő alatti divergenciájára egy olyan egyértelműen interpretálható küszöbérték felállítását, amely érték elkülönítené a kettősen fertőzött pácienseket a többitől (14. ábra). Ezek alapján megállapítottuk, hogy pusztán szekvenciatávolságok alapján nem tudjuk eldönteni egy páciensről, hogy felülfertőzött-e.

Továbbá, a mintavételi idők között eltelt idő nem jelzi megbízhatóan a két szekvencia közti generációk számát, mert a vírus archiválódhat a látens rezervoárban, majd újra aktiválódhat később (Yin, 2012). Ez a probléma megakadályozza annak az azonosítását is, hogy két vírustörzs közül melyik fertőzte meg előbb a beteget.

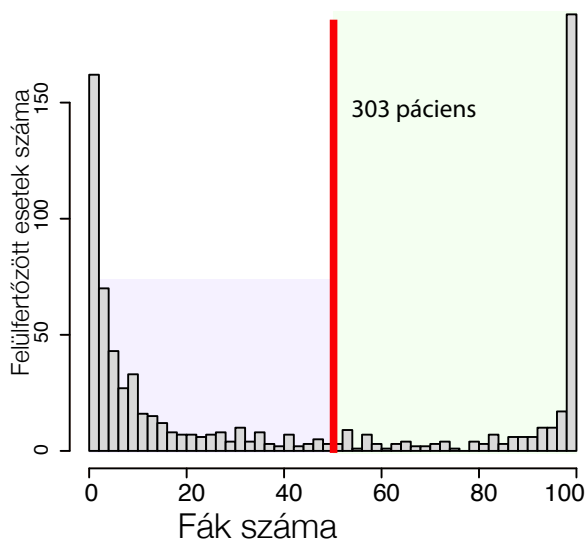
Végül, a divergencián alapuló módszerek általános problémája, hogy a HIV-vírus esetében a gazdán belüli és gazdák közötti evolúció léptéke, időskálája is átfed (Hill, 2012). Egy gazdából származó két vírusváltozat kialakulhatott a gazdán belüli evolúció során, vagy több gazdában, transzmissziós események sorozata során. Ha egy fertőzött páciens felülfertőződik egy vírussal, akkor ez a vírus ugyanannyi ideje divergál a közös őstől, mint a rezidens vírus, így pusztán a szekvenciák távolságából nem tudjuk eldönteni, hogy kettős fertőzéssel (felülfertőzéssel) van-e dolgunk, vagy sem. Ezen problémák miatt úgy döntöttünk, hogy filogenetikai fák alapján fogjuk eldönteni a páciensekről, hogy felülfertőzöttek-e (Lemey, 2005).

3.5.2. Előzetes szűrés maximum likelihood módszerrel

Első lépésként RAxML szoftver segítségével mind a 13816 szekvenciából készítettünk filogenetikai fát. Mivel a maximum likelihood keresés lokális optimumokba futhat, ezért a faépítést 100 véletlenszerűen inicializált replikátumon megismételtük. Mind a száz fában külön-külön megkerestük a felülfertőződéssel gyanúsítható pácienseket (a fenti kritériumokat használva), és minden pácienshez feljegyeztük, hogy hány replikátum fában mutatott felülfertőzést. Ez a kétcsúcsú eloszlás látható a 15. ábrán, amely alapján úgy döntöttünk, hogy azt a 303 a páciens vizsgáljuk meg részletesebben, akik legalább 50 replikátum fában felülfertőzöttnek bizonyultak.



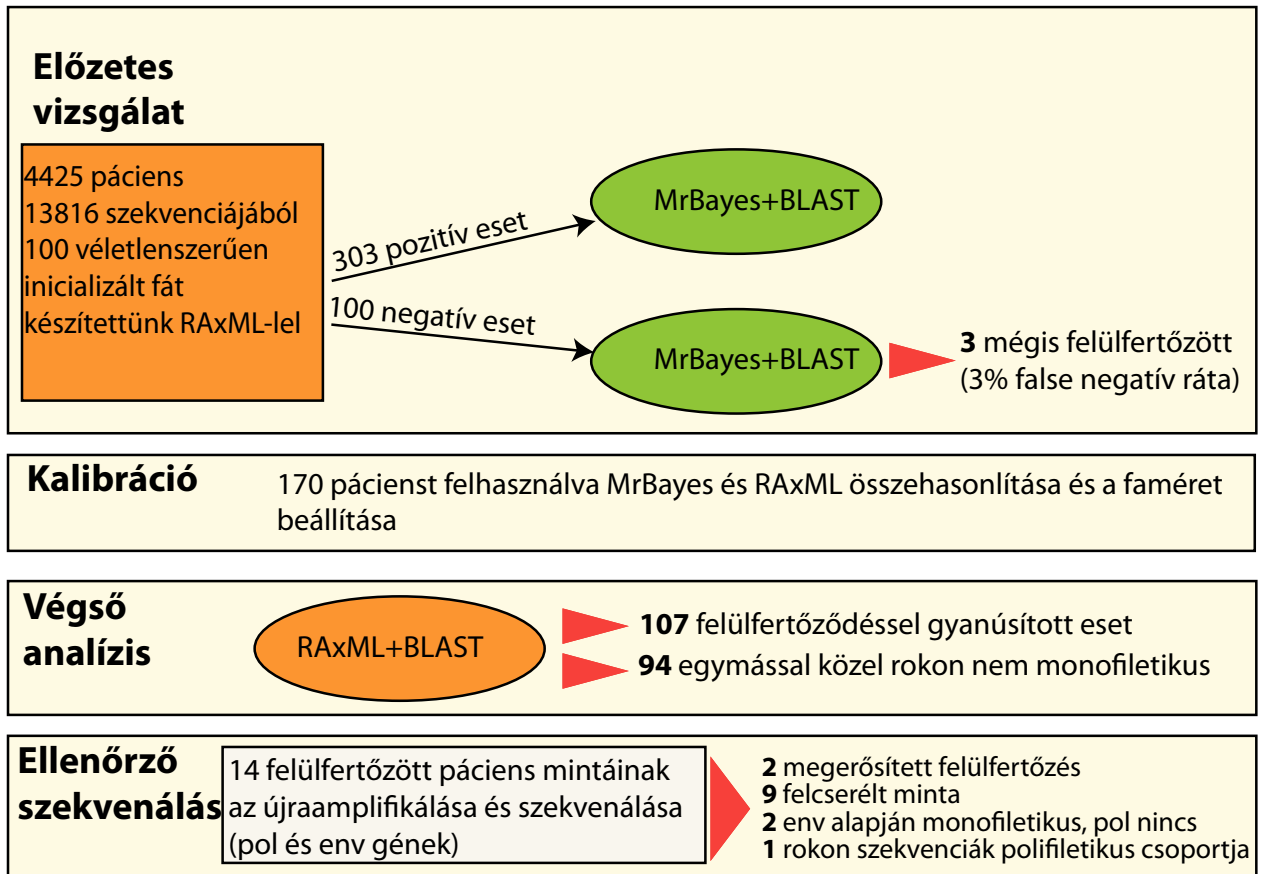
14. ábra. Az egy páciensből, illetve különböző páciensekből származó szekvenciák közötti páros szekvenciátávolságok eloszlásai. A szekvenciátávolságokat maximum likelihood módszerrel becsültük GTR szubsztitúciós modellt feltételezve, PAUP* szoftverrel. Felső: 80000 véletlenszerűen kiválasztott, különböző páciensekből származó szekvenciapár távolságainak az eloszlása. Alsó: az összes, páciensen belüli páros szekvenciátávolság eloszlása.



15. ábra. Előzetes vizsgálat RAxML segítségével: 100 véletlenszerűen inicializált maximum likelihood fa eredményei. A fákat 4425 páciens összes (13816) szekvenciája alapján építettük. A hisztogram azt mutatja, hogy adott számú fában hány páciens mutatott felülfertőződést. A kétszcúsú eloszlás alapján azt a 303 páciens vizsgáltuk meg részletesebben, akik legalább 50 fában felülfertőzöttnek bizonyultak. A hisztogram nem ábrázolja azt a kategóriát, amelyben azok a páciensek szerepelnek akik egyetlen egy fában se tűntek felülfertőzöttnek.

Ezt a 303 páciens a következő módon vizsgáltuk. Mindegyik páciens saját szekvenciáihoz kiválasztottunk 150 háttérszekvenciát egy nagy BLAST adatbázisból (lásd 3.4. fejezet) oly módon, hogy a saját szekvenciák egyenlő számú idegen szekvenciával legyenek reprezentálva. Ha egy páciens több, különálló vírustörzssel van megfertőzve, akkor a módszer biztosítja, hogy mindegyik vírustörzs szerepeljen a 150 háttérszekvencia között, míg a faépítéshez használt szekvenciák számát alacsonyan tartja. Az így kiválasztott szekvenciákon minden páciensre külön-külön filogenetikai fát építettünk a MrBayes szoftverrel. Ezekben a fában 0,95-ös megbízhatósági küszöbértéket használva 67 felülfertőzéssel gyanúsítható páciens találtunk, megerősítve, hogy a 303 páciens által alkotott halmaz valóban feldúsult a felülfertőzéses esetekben.

Az első RAxML szűrés szenzitivitását is meg szeretnénk volna becsülni, ezért 100 véletlenszerűen kiválasztott, de addig felülfertőződést nem mutató páciensen szintén elvégeztük a fenti vizsgálatot a BLAST keresés és a bayesian faépítés segítségével. Ez 3 felülfertőzéssel gyanúsítható páciens talált, amely eredmény rávilágított arra, hogy az első heurisztikus szűrés nem rendelkezik megfelelő érzékenységgel. A 3%-os „hamis negatív” arány a teljes adathalmazra vetítve 127 felülfertőzéssel gyanúsítható esetet vétene el. Ezért úgy döntöttünk, hogy a fenti-ekben leírt előzetes szűrést nem alkalmazzuk, és minden páciens egyenként megvizsgálunk a



16. ábra. A filogenetikai elemzés lépései vázlatosan, a Függelékben szereplő lépésekkel együtt.

BLAST keresésen alapuló módszerrel.

A fenti 67 felülfertőződéssel gyanúsított páciens mellé kiválasztottunk még véletlenszerűen 103-at, majd az így kapott 170 páciensből álló, felülfertőződéssel gyanúsított esetekben feldúsított halmazon kalibráltuk a módszer végző paramétereit (lásd 3.1. fejezet).

A fent leírt lépésekről egy diagram látható a 16. ábrán.

A szerző hozzájárulása a tanulmányhoz

Miután a Virolab és Euresist adatbázisok kért részeit megkaptuk, a szerző végezte el az összes adatfeldolgozási, adatelemzési és programozási feladatot. A szerző a témavezető tanácsai alapján tervezte a vizsgálatokat és algoritmusokat, illetve értelmezte azok eredményeit.

4. Részfeladatok egyéb projektekből

Doktori tanulmányaim alatt a fenti két nagy projekten kívül más munkákban kisebb részfeladatokat végeztem. Ezeket röviden összegzem az alábbiakban.

4.1. T-sejt populáció életidejének becslése

A T-sejtek életideje fontos immunológiai paraméter a HIV-fertőzéssel kapcsolatos kutatásokban. Ezt a paramétert *in vivo* rendszerben jelöléses vizsgálattal mérik. Ennek során a kísérleti személlyel vagy állattal a DNS-be beépülő jelölőanyagot (BrdU vagy deutérium) itatnak heteken keresztül. A jelölési szakasz során a sejtek az osztódás során felveszik és beépítik a „festéket” a saját DNS-ükbe. A jelölési ún. uplabeling szakaszt követő downlabeling fázisban pedig továbbadják az utódsejteknek a festéket, így egységnyi sejtben a festék mennyisége fokozatosan csökken. A sejtek életidejére (osztódási gyakoriságukra) az uplabeling és a downlabeling szakaszokban mért egységnyi sejtben levő festékből lehet következtetni.

Az adatok kiértékeléséhez megfelelő modellt kell illeszteni a mért pontokra, azonban heterogén (több kompartment) sejtpopuláció esetén apriori nehéz kiválasztani a kompartmentek pontos számát, ez pedig befolyásolhatja a becsült átlagos turnover-t.

Feladatom annak feltérképezése volt, hogy hogyan befolyásolja a becsült paraméter megbízhatóságát az, ha a valóságtól eltérő számú kompartmentet feltételezünk a modellben. Mesterségesen generált adatokra illesztettem különböző számú kompartmentből álló modelleket, majd nem parametrikus bootstrappal empirikus konfidenciaintervallumot készítettem a különböző modellek alatt becsült paraméterekről.

A jelölési szakaszra illesztett modellek:

$$L(t) = \sum_{i=1}^{n-1} \alpha_i (1 - e^{-d_i t}) + (1 - \sum_i \alpha_i) (1 - e^{-d_n t}), \quad \sum \alpha_i \leq 1. \quad (1)$$

A mesterséges adatokat $n = 2$ -vel generáltam, majd minden adatpontot egy egyenletes eloszlású ($p \in [1 \pm \sigma]$) véletlen számmal szoroztam be.

Nem parametrikus bootstrap Az illesztendő adatpontok száma N . Visszatevéssel kihúzunk ebből a halmazból N -et. 500 ilyen replikátumok készültek, mindegyikén elvégzem az illesztést és az átlagos turnover így kapott eloszlásából konfidenciaintervallumot készíték.

Eredmény A valódi átlagos turnover a konfidenciaintervallumon belül helyezkedett el akkor is, ha a valósnál nagyobb vagy kisebb kompartmentú modellt illeszttem.

Ebből a következő publikáció született: Westera L, Drylewicz J, den Braber I, Mugwagwa T, van der Maas I, Kwast L, Volman T, van de Weg-Schrijver EHR, **Bartha I**, Spierenburg G, Gaiser K, Ackermans MT, Asquith B, de Boer RJ, Tesselaar K, Borghans J a M: Closing the gap between T-cell life span estimates from stable isotope-labeling studies in mice and men. Blood 2013.

4.2. APOBEC3G indukálta hipermutáció nyomainak keresése HIV-szekvenciákban

Az antivirális APOBEC3G fehérje G->A hipermutáció révén képes hatástalanítani a vírus RNS kópiáit. A projekt egy kis részfeladatában a lehetséges hipermutációs eseményeket számoltam meg HIV-szekvenciákban. A mutációs eseményt egy feltételezett ősi referenciaszekvenciához viszonyítottuk. Feltételeztük, hogy az APOBEC3G fehérje a

G[AG] [^C]

(POSIX reguláris kifejezés) mintában az első guanint változtatja adeninné. Ezzel a módszerrel az RT és a Gp120 fehérjékben találtunk hipermutációra utaló lenyomatokat.

Ebből a következő publikáció született: Snoeck J, Fellay J, **Bartha I**, Douek DC, Telenti A: Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. Retrovirology 2011, 8:87.

4.3. Vizualizációs platformok készítése génexpressziós vizsgálatokhoz

HIV-fertőzés által indukált expressziós változások CD4-sejtekben Kollégáim CD4 sejtek expresszióját RNA-seq következő generációs szekvenálással mérték HIV-fertőzés előtt, és 24 órán keresztül két óránként a fertőzés után. Feladatom a nagy mennyiségű feldolgozott adat köré egy vizualizációs platform létrehozása volt, ahol tetszőleges gének expressziós mintázatának változásait böngészeti az olvasó.

Ebből a projektből a következő publikáció született: Mohammadi P, Desfarges S, **Bartha I**, Joos B, Zangger N, Muñoz M, Günthard HF, Beerenwinkel N, Telenti A, Ciuffi A: 24 hours in the life of HIV-1 in a T cell line. PLoS pathogens 2013, 9:e1003161.

5. Köszönetnyilvánítás

Köszönettel tartozom Müller Viktornak az elmúlt hat évben nyújtott útmutatásaiért, támogatás-áért és töretlen bizalmáért. Lausanne-i tanulmányutamat a svájci Sciex ösztöndíj tette lehetővé, ahol Amalio Telenti és Jacques Fellay segítségével végeztem kutatásaimat. Paul McLaren mindig segítőkész válaszáért hálás vagyok. Köszönöm Rovó Zitának és Rovó Petrának, hogy átnézték a dolgozatomat és kritikájukkal segítettek a dolgozat elkészítését. Köszönettel tartozom Gémesi Irén és Scheuer Zsuzsanna gimnáziumi tanárainknak.

Köszönöm a szüleimtől és a családomtól kapott támogatást és türelmet.

Hivatkozások

- Abdool Karim Q, Abdool Karim SS, Frohlich JA *et al.* Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science (New York, NY)* 2010. 329(5996):1168–74.
- Abecasis GR, Auton A, Brooks LD *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012. 491(7422):56–65.
- Alcantara LCJ, Cassol S, Libin P *et al.* A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic acids research* 2009. 37(Web Server issue):W634–42.
- Alter G, Heckerman D, Schneidewind A *et al.* HIV-1 adaptation to NK-cell-mediated immune pressure. *Nature* 2011. 476(7358):96–100.
- Althaus CL & Bonhoeffer S. Stochastic interplay between mutation and recombination during the acquisition of drug resistance mutations in human immunodeficiency virus type 1. *Journal of virology* 2005. 79(21):13572–8.
- Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *Journal of molecular biology* 1990. 215(3):403–410.
- Altshuler DM, Gibbs Ra, Peltonen L *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010. 467(7311):52–8.
- Appay V, Almeida JR, Sauce D *et al.* Accelerated immune senescence and HIV-1 infection. *Experimental gerontology* 2007. 42(5):432–7.
- Apps R, Qi Y, Carlson JM *et al.* Influence of HLA-C expression level on HIV control. *Science (New York, NY)* 2013. 340(6128):87–91.
- Arts EJ & Hazuda DJ. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor perspectives in medicine* 2012. 2(4):a007161.
- Balding DJ. A tutorial on statistical methods for population association studies. *Nature reviews Genetics* 2006. 7(10):781–91.
- Ballana E, Ruiz-de Andres A, Mothe B *et al.* Differential prevalence of the HLA-C – 35 CC genotype among viremic long term non-progressor and elite controller HIV+ individuals. *Immunobiology* 2012. 217(9):889–94.
- Barouch DH & Korber B. HIV-1 vaccine development after STEP. *Annual review of medicine*

2010. 61:153–67.
- Barre-Sinoussi F, Chermann JC, Rey F *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983. 220(4599):868–871.
- Bartha I, Simon P & Müller V. Has HIV evolved to induce immune pathogenesis? *Trends in immunology* 2008. 29(7):322–8.
- Batorsky R, Kearney MF, Palmer SE *et al.* Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proceedings of the National Academy of Sciences of the United States of America* 2011. 108(14):5661–6.
- Baur AS. HIV-Nef and AIDS pathogenesis: are we barking up the wrong tree? *Trends in microbiology* 2011. 19(9):435–40.
- Benjamini Y & Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995:289–300.
- Bhattacharya T, Daniels M, Heckerman D *et al.* Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science (New York, NY)* 2007. 315(5818):1583–6.
- Biancotto A, Grivel JC, Iglehart SJ *et al.* Abnormal activation and cytokine spectra in lymph nodes of people chronically infected with HIV-1. *Blood* 2007. 109(10):4272–4279.
- Borghans JaM, Mø lgaard A, de Boer RJ *et al.* HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PloS one* 2007. 2(9):e920.
- Brenchley JM & Douek DC. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *Mucosal immunology* 2008. 1(1):23–30.
- Brenchley JM & Price DA. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature Medicine* 2006. 12(12):1365.
- Brennan Ca, Ibarondo FJ, Sugar Ca *et al.* Early HLA-B*57-restricted CD8+ T lymphocyte responses predict HIV-1 disease progression. *Journal of virology* 2012. 86(19):10505–16.
- Bretscher MT, Althaus CL, Müller V *et al.* Recombination in HIV and the evolution of drug resistance: for better or for worse? *BioEssays* 2004. 26(2):180–8.
- Bukrinsky MI, Stanwick TL, Dempsey MP *et al.* Quiescent T lymphocytes as an inducible

- virus reservoir in HIV-1 infection. *Science* 1991. 254(5030):423–427.
- Bush WS & Moore JH. Chapter 11: Genome-wide association studies. *PLoS computational biology* 2012. 8(12):e1002822.
- Buzon V, Natrajan G, Schibli D *et al.* Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS pathogens* 2010. 6(5):e1000880.
- Campbell MS, Gottlieb GS, Hawes SE *et al.* HIV-1 superinfection in the antiretroviral therapy era: are seroconcordant sexual partners at risk? *PLoS One* 2009. 4(5):e5690.
- Carlson JM, Listgarten J, Pfeifer N *et al.* Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of HIV-1. *Journal of Virology* 2012. 86(9):5230–5243.
- Carrington M & O'Brien SJ. The influence of HLA genotype on AIDS. *Annual review of medicine* 2003. 54:535–51.
- Chase A, Zhou Y & Siliciano RF. HIV-1-induced depletion of CD4+ T cells in the gut: mechanism and therapeutic implications. *Trends in pharmacological sciences* 2006. 27(1):4–7.
- Check Hayden E. Infant's vanquished HIV leaves doctors puzzled. *Nature* 2013.
- Chen L, Kwon YD, Zhou T *et al.* Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science (New York, NY)* 2009. 326(5956):1123–7.
- Chou S, Upton H, Bao K *et al.* HIV-1 Tat recruits transcription elongation factors dispersed along a flexible AFF4 scaffold. *Proceedings of the National Academy of Sciences of the United States of America* 2013. 110(2):E123–31.
- Cichon S, Craddock N, Daly M *et al.* Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *The American journal of psychiatry* 2009. 166(5):540–56.
- Cirulli ET & Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews Genetics* 2010. 11(6):415–25.
- Coffin J & Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor perspectives in medicine* 2013. 3(1):a012526.
- Cohen OJ, Pantaleo G, Schwartzentruber DJ *et al.* Pathogenic insights from studies of lymphoid tissue from HIV-infected individuals. *J Acquir Immune Defic Syndr Hum Retrovirol* 1995. 10 Suppl 1:S6–14.
- Cornelissen M, Jurriaans S, Kozaczynska K *et al.* Routine HIV-1 genotyping as a tool to identify

- dual infections. *AIDS (London, England)* 2007. 21(7):807–11.
- Cornelissen M, Pasternak AO, Grijzen ML *et al.* HIV-1 dual infection is associated with faster CD4+ T-cell decline in a cohort of men with primary HIV infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012. 54(4):539–47.
- Craigie R & Bushman FD. HIV DNA Integration. *Cold Spring Harbor perspectives in medicine* 2012. 2(7):a006890.
- Dalmaso C, Carpentier W, Meyer L *et al.* Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. *PLoS one* 2008. 3(12):e3907.
- Davey RT, Bhat N, Yoder C *et al.* HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proceedings of the National Academy of Sciences of the United States of America* 1999. 96(26):15109–14.
- De Boer RJ & Perelson AS. Quantifying T lymphocyte turnover. *Journal of theoretical biology* 2013. 327:45–87.
- de Oliveira T, Deforche K, Cassol S *et al.* An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2005. 21(19):3797–3800.
- de Sousa JaD, Müller V, Lemey P *et al.* High GUD Incidence in the Early 20th Century Created a Particularly Permissive Time Window for the Origin and Initial Spread of Epidemic HIV Strains. *PLoS ONE* 2010. 5(4):e9936.
- de Witte L, Nabatov A & Geijtenbeek TBH. Distinct roles for DC-SIGN+-dendritic cells and Langerhans cells in HIV-1 transmission. *Trends in molecular medicine* 2008. 14(1):12–9.
- de Witte L, Nabatov A, Pion M *et al.* Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. *Nature medicine* 2007. 13(3):367–71.
- Dean M, Carrington M, Winkler C *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE . *Science (New York, NY)* 1996. 273(5283):1856–1862.
- Deeks SG, Kitchen CMR, Liu L *et al.* Immune activation set point during early HIV infection

- predicts subsequent CD4+ T-cell changes independent of viral load. *Blood* 2004. 104(4):942–947.
- Derdeyn Ca, Decker JM, Bibollet-Ruche F *et al.* Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science (New York, NY)* 2004. 303(5666):2019–22.
- Descours B, Cribier A, Chable-Bessia C *et al.* SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4(+) T-cells. *Retrovirology* 2012. 9(1):87.
- Dinoso JB, Kim SY, Wiegand aM *et al.* Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences of the United States of America* 2009. 106(23):9403–8.
- Doitsh G, Cavrois M, Lassen KG *et al.* Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell* 2010. 143(5):789–801.
- Douek DC. Disrupting T-cell homeostasis: how HIV-1 infection causes disease. *AIDS reviews* 2003. 5(3):172–7.
- Douek DC, Picker LJ & Koup RA. T Cell Dynamics in {HIV}-1 Infection. *Annu Rev Immunol* 2003. 21:265–304.
- Dyer WB, Zaunders JJ, Yuan FF *et al.* Mechanisms of HIV non-progression; robust and sustained CD4+ T-cell proliferative responses to p24 antigen correlate with control of viraemia and lack of disease progression after long-term transfusion-acquired HIV-1 infection. *Retrovirology* 2008. 5:112.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 2004. 32(5):1792–7.
- Evangelou E, Fellay J, Colombo S *et al.* Impact of phenotype definition on genome-wide association signals: empirical evaluation in human immunodeficiency virus type 1 infection. *American journal of epidemiology* 2011. 173(11):1336–42.
- Fellay J, Ge D, Shianna KV *et al.* Common genetic variation and the control of HIV-1 in humans. *PLoS genetics* 2009. 5(12):e1000791.
- Fellay J & McLaren P. Host Genetics and Genomics. In *Encyclopedia of AIDS*. Springer 2013.
- Fellay J, Shianna KV, Ge D *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007. 317(5840):944–947.
- Finkel TH, Tudor-Williams G, Banda NK *et al.* Apoptosis occurs predominantly in bystander

- cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. *Nature Medicine* 1995. 1(2):129–134.
- Frankel AD & Young JA. HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry* 1998. 67(67):1–25.
- Fraser C. HIV recombination: what is the impact on antiretroviral therapy? *Journal of the Royal Society, Interface / the Royal Society* 2005. 2(5):489–503.
- Frazer Ka, Ballinger DG, Cox DR *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007. 449(7164):851–61.
- Gabriel SB, Schaffner SF, Nguyen H *et al.* The structure of haplotype blocks in the human genome. *Science (New York, NY)* 2002. 296(5576):2225–9.
- Gallant J. *Global HIV/AIDS Medicine*. Elsevier Health Sciences 2008. ISBN 978-1-4160-2882-6.
- Gallo RC, Sarin PS, Gelmann EP *et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 1983. 220(4599):865–867.
- Garcia-Perez J, Rueda P, Alcami J *et al.* Allosteric model of maraviroc binding to CC chemokine receptor 5 (CCR5). *The Journal of biological chemistry* 2011. 286(38):33409–21.
- Gibson G. Rare and common variants: twenty arguments. *Nature reviews Genetics* 2011. 13(2):135–45.
- Gilbert MTP, Rambaut A, Wlasiuk G *et al.* The emergence of HIV/AIDS in the Americas and beyond. *Proceedings of the National Academy of Sciences of the United States of America* 2007. 104(47):18566–70.
- Gonzales MJ, Delwart E, Rhee SY *et al.* Lack of detectable human immunodeficiency virus type 1 superinfection during 1072 person-years of observation. *J Infect Dis* 2003. 188(3):397–405.
- Goulder PJ, Brander C, Tang Y *et al.* Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* 2001. 412(6844):334–8.
- Goulder PJR & Watkins DI. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature reviews Immunology* 2008. 8(8):619–30.
- Grant RM, Lama JR, Anderson PL *et al.* Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. *New England Journal of Medicine* 2010. 363(27):2587–2599.
- Grossman Z, Feinberg MB & Paul WE. Multiple modes of cellular activation and virus trans-

- mission in HIV infection: a role for chronically and latently infected cells in sustaining viral replication. *Proceedings of the National Academy of Sciences of the United States of America* 1998. 95(11):6314–9.
- Grossman Z, Meier-Schellersheim M, Paul WE *et al.* Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. *Nature medicine* 2006. 12(3):289–95.
- Gyenesi A, Moody J, Semple CaM *et al.* High-throughput analysis of epistasis in genome-wide association studies with BiForce. *Bioinformatics (Oxford, England)* 2012. 28(15):1957–64.
- Haines JL, Hauser MA, Schmidt S *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005. 308(5720):419–421.
- Hamers RL, Siwale M, Wallis CL *et al.* HIV-1 drug resistance mutations are present in six percent of persons initiating antiretroviral therapy in Lusaka, Zambia. *Journal of acquired immune deficiency syndromes (1999)* 2010. 55(1):95–101.
- Hazenber MD, Otto SA, van Benthem BHB *et al.* Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *AIDS* 2003. 17(13):1881–1888.
- He J, Chen Y, Farzan M *et al.* CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature* 1997:645–649.
- Henrich TJ & Kuritzkes DR. HIV-1 entry inhibitors: recent development and clinical use. *Current opinion in virology* 2013. 3(1):51–7.
- Hersperger AR, Martin JN, Shin LY *et al.* Increased HIV-specific CD8+ T-cell cytotoxic potential in HIV elite controllers is associated with T-bet expression. *Blood* 2011. 117(14):3799–808.
- Hill AL, Rosenbloom DIS & Nowak Ma. Evolutionary dynamics of HIV at multiple spatial and temporal scales. *Journal of molecular medicine (Berlin, Germany)* 2012. 90(5):543–61.
- Hindorff L, MacArthur J, Morales J *et al.* A Catalog of Published Genome-Wide Association Studies 2012.
- Hué S, Clewley JP, Cane PA *et al.* HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS (London, England)* 2004. 18(5):719–28.
- Hütter G, Nowak D, Mossner M *et al.* Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *The New England journal of medicine* 2009. 360(7):692–8.

- Jeffery KJ & Bangham CR. Do infectious diseases drive MHC diversity? *Microbes and infection / Institut Pasteur* 2000. 2(11):1335–41.
- Josefsson L, King MS, Makitalo B *et al.* Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proceedings of the National Academy of Sciences of the United States of America* 2011. 108(27):11199–204.
- Jung A, Maier R, Vartanian JP *et al.* Multiply infected spleen cells in HIV patients. *Nature* 2002. 418.
- Jurriaans S, Kozaczynska K, Zorgdrager F *et al.* A sudden rise in viral load is infrequently associated with HIV-1 superinfection. *Journal of Acquired Immune Deficiency Syndromes* 2008. 47(1):69–73.
- Kawashima Y, Pfafferoth K, Frater J *et al.* Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* 2009. 458(7238):641–5.
- Keele BF, Giorgi EE, Salazar-Gonzalez JF *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* 2008. 105(21):7552–7.
- Kelleher aD, Long C, Holmes EC *et al.* Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *The Journal of experimental medicine* 2001. 193(3):375–86.
- Kelleher AD & Zaunders JJ. Decimated or missing in action: CD4+ T cells as targets and effectors in the pathogenesis of primary HIV infection. *Current HIV/AIDS reports* 2006. 3(1):5–12.
- Kerkau T, Schmitt-Landgraf R, Schimpl A *et al.* Downregulation of HLA class I antigens in HIV-1-infected cells. *AIDS research and human retroviruses* 1989. 5(6):613–620.
- Kirchhoff F. Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell host & microbe* 2010. 8(1):55–67.
- Kirchhoff F, Greenough TC, Brettler DB *et al.* Absence of Intact nef Sequences in a Long-Term Survivor with Nonprogressive HIV-1 Infection. *New England Journal of Medicine* 1995. 332(4):228–232.
- Kirmaier A, Wu F, Newman RM *et al.* TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new

- species. *PLoS biology* 2010. 8(8).
- Klatzmann D, Champagne E, Chamaret S *et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 1984. 312(5996):767–768.
- Knoll B, Lassmann B & Temesgen Z. Current status of HIV infection: a review for non-HIV-treating physicians. *International journal of dermatology* 2007. 46(12):1219–28.
- Laguette N & Benkirane M. How SAMHD1 changes our view of viral restriction. *Trends in immunology* 2012. 33(1):26–33.
- Laguette N, Sobhian B, Casartelli N *et al.* SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* 2011. 474(7353):654–7.
- Lapkouski M, Tian L, Miller JT *et al.* Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nature structural & molecular biology* 2013. 20(2):230–6.
- Le Clerc S, Limou S, Coulonges C *et al.* Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). *The Journal of infectious diseases* 2009. 200(8):1194–201.
- Lemey P, Van Dooren S, Van Laethem K *et al.* Molecular testing of multiple HIV-1 transmissions in a criminal case. *AIDS* 2005. 19(15):1649–1658.
- Lemey P & Vandamme AM. Exploring full-genome sequences for phylogenetic support of HIV-1 transmission events. *AIDS (London, England)* 2005. 19(14):1551–2.
- Leslie aJ, Pfafferott KJ, Chetty P *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine* 2004. 10(3):282–9.
- Lettre G, Lange C & Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genetic epidemiology* 2007. 31(4):358–62.
- Levy DN, Aldrovandi GM, Kutsch O *et al.* Dynamics of HIV-1 recombination in its natural target cells. *Proceedings of the National Academy of Sciences of the United States of America* 2004. 101(12):4204–9.
- Levy Ja. HIV pathogenesis: 25 years of progress and persistent challenges. *AIDS (London, England)* 2009. 23(2):147–60.
- Li Y, Willer CJ, Ding J *et al.* MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology* 2010. 34(8):816–834.

- Limou S, Delaneau O, van Manen D *et al.* Multicohort genomewide association study reveals a new signal of protection against HIV-1 acquisition. *The Journal of infectious diseases* 2012. 205(7):1155–62.
- Limou S, Le Clerc S, Coulonges C *et al.* Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *The Journal of infectious diseases* 2009. 199(3):419–26.
- Liu AY, Buchbinder SP, GRM. Preexposure prophylaxis for hiv: Unproven promise and potential pitfalls. *JAMA* 2006. 296(7):863–865.
- Lore K & Larsson M. The role of dendritic cells in the pathogenesis of HIV-1 infection. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* 2003. 111(7-8):776–88.
- Lythgoe Ka & Fraser C. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proceedings Biological sciences / The Royal Society* 2012. 279(1741):3367–75.
- Manolio Ta, Collins FS, Cox NJ *et al.* Finding the missing heritability of complex diseases. *Nature* 2009. 461(7265):747–53.
- Mansky LM. Retrovirus mutation rates and their role in genetic variation. *The Journal of general virology* 1998. 79 (Pt 6):1337–45.
- Masur H, Michelis MA, Greene JB *et al.* An outbreak of community-acquired *Pneumocystis carinii* pneumonia. *New England Journal of Medicine* 1981. 305(24):1431–1438.
- McCarthy MI, Abecasis GR, Cardon LR *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews Genetics* 2008. 9(5):356–69.
- McLaren PJ, Coulonges C, Ripke S *et al.* Association Study of Common Genetic Variants and HIV-1 Acquisition in 6,300 Infected Cases and 7,200 Controls. *PLoS pathogens* 2013. 9(7):e1003515.
- McMichael AJ, Borrow P, Tomaras GD *et al.* The immune response during acute HIV-1 infection: clues for vaccine development. *Nature reviews Immunology* 2010. 10(1):11–23.
- Messiaen P, Wensing AMJ, Fun A *et al.* Clinical Use of HIV Integrase Inhibitors: A Systematic Review and Meta-Analysis. *PLoS ONE* 2013. 8(1):e52562.
- Michel N, Allespach I, Venzke S *et al.* The Nef protein of human immunodeficiency virus establishes superinfection immunity by a dual strategy to downregulate cell-surface CCR5

- and CD4. *Current biology : CB* 2005. 15(8):714–23.
- Mildvan D, Mathur U, Enlow R *et al.* Persistent, generalized lymphadenopathy among homosexual males. *MMWR Morbidity and mortality weekly report* 1982. 31(19):249–51.
- Miller CJ & Shattock RJ. Target cells in vaginal HIV transmission. *Microbes and infection / Institut Pasteur* 2003. 5(1):59–67.
- Moffatt MF, Gut IG, Demenais F *et al.* A large-scale, consortium-based genomewide association study of asthma. *The New England journal of medicine* 2010. 363(13):1211–21.
- Moir S, Chun TW & Fauci AS. Pathogenic mechanisms of HIV disease. *Annual review of pathology* 2011. 6:223–48.
- Moore CB, John M, James IR *et al.* Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science (New York, NY)* 2002. 296(5572):1439–43.
- Mostowy R, Kouyos RD, Fouchet D *et al.* The role of recombination for the coevolutionary dynamics of HIV and the immune response. *PloS one* 2011. 6(2):e16052.
- Mulherin Sa, O'Brien TR, Ioannidis JP *et al.* Effects of CCR5-Delta32 and CCR2-64I alleles on HIV-1 disease progression: the protection varies with duration of infection. *AIDS (London, England)* 2003. 17(3):377–87.
- Neher Ra & Leitner T. Recombination rate and selection strength in HIV intra-patient evolution. *PLoS computational biology* 2010. 6(1):e1000660.
- Neil SJD, Zang T & Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 2008. 451(7177):425–30.
- Nethe M, Berkhout B & van der Kuyl AC. Retroviral superinfection resistance. *Retrovirology* 2005. 2:52.
- Nilsson J, Kinloch-de Loes S, Granath A *et al.* Early immune activation in gut-associated and peripheral lymphoid tissue during acute {HIV} infection. *AIDS* 2007. 21(5):565–574.
- Novembre J, Galvani AP & Slatkin M. The geographic spread of the CCR5 Delta32 HIV-resistance allele. *PLoS biology* 2005. 3(11):e339.
- Ocwieja KE, Brady TL, Ronen K *et al.* HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS pathogens* 2011. 7(3):e1001313.
- Ogg GS, Jin X, Bonhoeffer S *et al.* Decay Kinetics of Human Immunodeficiency Virus-Specific Effector Cytotoxic T Lymphocytes after Combination Antiretroviral Therapy. *Journal of*

- Virology* 1999. 73(1):797–800.
- Ortiz M, Guex N, Patin E *et al.* Evolutionary trajectories of primate genes involved in HIV pathogenesis. *Molecular biology and evolution* 2009. 26(12):2865–75.
- Owen RE, Heitman JW, Hirschhorn DF *et al.* HIV+ elite controllers have low HIV-specific T-cell activation yet maintain strong, polyfunctional T-cell responses. *AIDS (London, England)* 2010. 24(8):1095–105.
- Pandrea I, Gaufin T, Brenchley JM *et al.* Cutting edge: Experimentally induced immune activation in natural hosts of simian immunodeficiency virus induces significant increases in viral replication and {CD}4+ {T} cell depletion. *J Immunol* 2008. 181(10):6687–6691.
- Patterson N, Price AL & Reich D. Population Structure and Eigenanalysis. *PLoS Genetics* 2006. 2(12):20.
- Perelson AS, Neumann AU, Markowitz M *et al.* HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996. 271(5255):1582–1586.
- Pereyra F, Jia X & McLaren P. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science (New York, ...)* 2010. 330(6010):1551–1557.
- Piantadosi A, Chohan B, Chohan V *et al.* Chronic HIV-1 infection frequently fails to protect against superinfection. *PLoS pathogens* 2007. 3(11):e177.
- Piantadosi A, Ngayo MO, Chohan B *et al.* Examination of a second region of the HIV type 1 genome reveals additional cases of superinfection. *AIDS research and human retroviruses* 2008. 24(9):1221.
- Poignard P, Saphire EO, Parren PW *et al.* gp120: Biologic aspects of structural features. *Annual review of immunology* 2001. 19:253–74.
- Preston B, Poiesz B & Loeb L. Fidelity of HIV-1 reverse transcriptase. *Science* 1988. 242(4882):1168–1171.
- Price AL, Patterson NJ, Plenge RM *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* 2006. 38(8):904–9.
- Prince J, Claiborne D, Heckerman D *et al.* Impact of transmitted CTL escape mutations on replicative capacity and HIV pathogenesis in early infection. *Retrovirology* 2012. 9(Suppl 2):O57.
- Purcell DF & Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA

- modulates viral protein expression, replication, and infectivity. *Journal of Virology* 1993. 67(11):6365–6378.
- Purcell S, Cherny SS & Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics (Oxford, England)* 2003. 19(1):149–50.
- Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 2007. 81(3):559–575.
- Purcell SM, Wray NR, Stone JL *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009. 460(7256):748–52.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2013.
- Rachinger A, van de Ven TD, Burger JA *et al.* Evaluation of pre-screening methods for the identification of HIV-1 superinfection. *Journal of virological methods* 2010. 165(2):311–7.
- Rahm N, Gfeller D, Snoeck J *et al.* Susceptibility and adaptation to human TRIM5 α alleles at positive selected sites in HIV-1 capsid. *Virology* 2013:1–9.
- Rambaut A, Posada D, Crandall KA *et al.* The causes and consequences of HIV evolution. *Nature reviews Genetics* 2004. 5(1):52–61.
- Rambaut A, Robertson DL, Pybus OG *et al.* Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature* 2001. 410(6832):1047–8.
- Rasetti R & Weinberger DR. Intermediate phenotypes in psychiatric disorders. *Current opinion in genetics & development* 2011. 21(3):340–8.
- Redd AD, Mullis CE, Serwadda D *et al.* The Rates of HIV Superinfection and Primary HIV Incidence in a General Population in Rakai, Uganda. *The Journal of infectious diseases* 2012. 206(2):267–74.
- Rhee SY, Gonzales MJ, Kantor R *et al.* Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic acids research* 2003. 31(1):298–303.
- Ribeiro RM, Qin L, Chavez LL *et al.* Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *Journal of virology* 2010. 84(12):6096–102.

- Rice P, Longden I, Bleasby A *et al.* EMBOSS: the European molecular biology open software suite. *Trends in genetics* 2000. 16(6):276–277.
- Ronen K, McCoy CO, Matsen Fa *et al.* HIV-1 Superinfection Occurs Less Frequently Than Initial Infection in a Cohort of High-Risk Kenyan Women. *PLoS Pathogens* 2013. 9(8):e1003593.
- Ronquist F, Teslenko M, van der Mark P *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 2012. 61(3):539–42.
- Rowland-Jones S. AIDS pathogenesis: what have two decades of research taught us? *Bulletin et mémoires de l'Académie royale de médecine de Belgique* 2004. 159(Pt 2):171–5.
- Ruff CT, Ray SC, Kwon P *et al.* Persistence of Wild-Type Virus and Lack of Temporal Structure in the Latent Reservoir for Human Immunodeficiency Virus Type 1 in Pediatric Patients with Extensive Antiretroviral Exposure. *Journal of Virology* 2002. 76(18):9481–9492.
- Schaller T, Goujon C & Malim MH. AIDS/HIV. HIV interplay with SAMHD1. *Science (New York, NY)* 2012. 335(6074):1313–4.
- Seifert E. GRAL graphing library 2009.
- Sekaly RP. The failed HIV Merck vaccine study: a step back or a launching point for future vaccine development? *The Journal of experimental medicine* 2008. 205(1):7–12.
- Shaw GM & Hunter E. HIV Transmission. *Cold Spring Harbor perspectives in medicine* 2012.
- Simon-Loriere E & Holmes EC. Why do RNA viruses recombine? *Nature reviews Microbiology* 2011. 9(8):617–26.
- Smith DM, Richman DD & Little SJ. HIV superinfection. *J Infect Dis* 2005a. 192(3):438–444.
- Smith DM, Wong JK, Hightower GK *et al.* HIV drug resistance acquired through superinfection. *AIDS* 2005b. 19(12):1251–1256.
- Smith MW, Dean M, Carrington M *et al.* Contrasting Genetic Influence of CCR2 and CCR5 Variants on HIV-1 Infection and Disease Progression. *Science* 1997. 277(5328):959–965.
- Stahl Ea, Wegmann D, Trynka G *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature genetics* 2012. 44(5):483–9.
- Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006. 22(21):2688–2690.

- Stamatakis A, Hoover P & Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic biology* 2008. 57(5):758–71.
- Stephens HAF. Immunogenetic surveillance of HIV/AIDS. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2012. 12(7):1481–91.
- Stevenson M. HIV-1 pathogenesis. *Nature Medicine* 2003. 9(7):853–860.
- Swofford DL. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. 2003.
- Templeton AR, Kramer MG, Jarvis J *et al*. Multiple-infection and recombination in HIV-1 within a longitudinal cohort of women. *Retrovirology* 2009. 6:54.
- Tilton JC & Doms RW. Entry inhibitors in the treatment of HIV-1 infection. *Antiviral research* 2010. 85(1):91–100.
- Tobin NH, Learn GH, Holte SE *et al*. Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. *Journal of virology* 2005. 79(15):9625–34.
- Touloumi G, Pantazis N, Pillay D *et al*. Impact of HIV-1 subtype on CD4 count at HIV seroconversion, rate of decline, and viral load set point in European seroconverter cohorts. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2013. 56(6):888–97.
- Trono D & Marzetta F. Profaning the ultimate sanctuary: HIV latency in hematopoietic stem cells. *Cell host & microbe* 2011. 9(3):170–2.
- Trono D, Van Lint C, Rouzioux C *et al*. HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals. *Science (New York, NY)* 2010. 329(5988):174–80.
- Troyer JL, Nelson GW, Lautenberger Ja *et al*. Genome-wide association study implicates PARD3B-based AIDS restriction. *The Journal of infectious diseases* 2011. 203(10):1491–502.
- Tsui R, Herring B, Barbour J *et al*. Human immunodeficiency virus type 1 superinfection was not detected following 215 years of injection drug user exposure. *Journal of Virology* 2004. 78(1):94–103.
- UNAIDS. *Global report: UNAIDS report on the global AIDS epidemic 2012*. Geneva. ISBN 978-92-9173-592-1 2012.

- Vallari A, Holzmayer V, Harris B *et al.* Confirmation of putative HIV-1 group P in Cameroon. *Journal of virology* 2011. 85(3):1403–7.
- van Manen D, van 't Wout AB & Schuitemaker H. Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology* 2012. 9:70.
- Varghese B, Maher JE, Peterman Ta *et al.* Reducing the risk of sexual HIV transmission: quantifying the per-act risk for HIV on the basis of choice of partner, sex act, and condom use. *Sexually transmitted diseases* 2002. 29(1):38–43.
- Visscher PM, Hill WG & Wray NR. Heritability in the genomics era—concepts and misconceptions. *Nature reviews Genetics* 2008. 9(4):255–66.
- Wall JD & Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nature reviews Genetics* 2003. 4(8):587–97.
- Walli R, Reinhart B, Luckow B *et al.* HIV-1-Infected Long-Term Slow Progressors Heterozygous for [DELTA]32-CCR5 Show Significantly Lower Plasma Viral Load Than Wild-Type Slow Progressors. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 1998. 18(3).
- Wei X, Decker JM, Wang S *et al.* Antibody neutralization and escape by HIV-1. *Nature* 2003. 422(6929):307–12.
- Wu L & KewalRamani VN. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nature reviews Immunology* 2006. 6(11):859–68.
- Yang J, Benyamin B, McEvoy BP *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 2010. 42(7):565–9.
- Yang O, Daar E & Jamieson B. Human immunodeficiency virus type 1 clade B superinfection: evidence for differential immune containment of distinct clade B strains. *Journal of Virology* 2005. 79(2):860–868.
- Yap MW, Nisole S & Stoye JP. A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Current Biology* 2005. 15:73–78.
- Yates A, Stark J, Klein N *et al.* Understanding the Slow Depletion of Memory CD4+ T Cells in HIV Infection. *PLoS Medicine* 2007. 4(5):e177.
- Yin L, Liu L, Sun Y *et al.* High-resolution deep sequencing reveals biodiversity, population structure, and persistence of HIV-1 quasispecies within host ecosystems. *Retrovirology* 2012. 9(1):108.

- Yusim K, Korber BTM, Brander C *et al.* HIV molecular immunology 2009. *Los Alamos, NM: Los Alamos National Laboratory, Theoretical Biology and Biophysics* 2009:3–24.
- Zhou T, Xu L, Dey B *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* 2007. 445(7129):732–7.
- Zuk O, Hechter E, Sunyaev SR *et al.* The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America* 2012. 109(4):1193–8.

Kivonat

Napjaink egyik jelentős egészségügyi kihívása a HIV-vírus által okozott világméretű járvány. Dolgozatomban két adatelemzési munkát mutatok be, amelyek a HIV-fertőzés epidemiológiájába és patogenezisébe engednek bepillantást.

Az első munkában a humán genom által a vírusra kifejtett szelekciós nyomást térképeztem fel. Ennek során genetikai asszociációvizsgálatot végeztünk teljes genomot lefedő gyakori humán mutációk és a vírus aminosavszekvenciájában megtalálható variáció között. Azonosítottunk 40 virális aminosavpozíciót, amelyek szignifikánsan asszociálnak a humán genomban található Fő Hisztokompatibilitási Komplex (MHC) polimorfizmusaival. Megerősítettük az MHC kiemelt szerepét a HIV-fertőzésre adott válaszban.

Második munkámban a HIV-fertőzés epidemiológiájában fontos szerepet betöltő felülfertőződés gyakoriságát becsültem nukleotidszekvenciákból. A HIV-felülfertőződés lehetővé teszi a különböző virális törzsek rekombinációját, illetve meggyorsíthatja a betegség előrehaladását. Továbbá egy esetleges drogrezisztens vírussal való felülfertőződés hatástalaníthatja az addig sikeres terápiát. Rutin genotipizálás során gyűjtött nukleotidszekvenciák álltak rendelkezésünkre egy terápia alatt álló páciensek adatait tartalmazó klinikai adatbázisból. A szekvenciák filogenetikai elemzésével azonosítottuk azokat a pácienseket, amelyek felülfertőződéssel gyanúsíthatóak. Megállapítottuk, hogy a terápia alatt álló páciensek körében a felülfertőződés ritka, azonban jelentős számú felcserélt mintát azonosítottunk.

A fentiekén kívül kisebb részfeladatokat végeztem egyéb, a HIV-fertőzés biológiájával kapcsolatos kutatásokban: nyugvó T-sejtek életidejének becslése, APOBEC3G potenciális hipermutációs helyeinek megkeresése, és génexpressziós adatok vizualizálása.

Abstract

HIV infection is a profound public health issue, especially in Sub-Saharan Africa. During my PhD training I performed two data analysis projects which provided insights into the epidemiology and pathogenesis of the virus.

HIV-1 sequence diversity is affected by selection pressures arising from host genomic factors. Using paired human and viral data we tested for associations between host DNA polymorphism and HIV-1 sequence variation. A total of 40 HIV-1 amino acid variants were associated with SNPs mapping to the Major Histocompatibility Complex region. The proposed genome-to-genome approach highlights sites of genomic conflict, and confirms the extensive effect of MHC on the within host evolution of the virus.

Superinfection with drug resistant HIV strains could potentially contribute to compromised therapy in patients initially infected with drug-sensitive virus and receiving antiretroviral therapy. To investigate the importance of this potential route to drug resistance, we developed a bioinformatics pipeline to detect superinfection from routinely collected genotyping data, and assessed whether superinfection contributed to increased drug resistance in a large European cohort of viremic, drug treated patients. Most cases of non-monophyletic clustering in patient phylogenies arose from sample or sequence mix-up rather than from superinfection. Superinfection was rare in our mainly treatment experienced cohort, therefore we conclude that in our large cohort, superinfection with drug resistant HIV did not compromise the efficiency of antiretroviral treatment.