

3.	Multifaktoriális betegségek genomikai vizsgálati módszerei	1
3.1.	<i>Statisztikai alapok</i>	1
3.1.1.	<i>Hardy Weinberg eloszlás</i>	1
3.1.2.	<i>Kapcsoltság és haplotípus</i>	3
3.1.3.	<i>LOD score</i>	4
3.1.4.	<i>Pozícionális klónozás</i>	5
3.1.5.	<i>Founder populációk</i>	5
3.1.6.	<i>Asszociációs vizsgálatok</i>	6
3.1.7.	<i>Mintagyűjtések típusai</i>	7
3.1.8.	<i>Kockázatszámítás</i>	8
3.1.9.	<i>Genetikai markerek</i>	9
3.1.10.	<i>Power analízis</i>	10
3.2.	<i>Betegségek genomikai hátterének vizsgálati módszerei</i>	10
3.2.1.	<i>Genetikai variációk kimutatása</i>	10
3.2.2.	<i>GWAS</i>	12
3.2.3.	<i>GWAS eredmények értékelése</i>	13
3.2.4.	<i>Parciális genomszűrések</i>	14
3.2.5.	<i>Személyre szabott genomika</i>	14
3.2.6.	<i>Újgenerációs szekvenálás (NGS)</i>	15
3.2.7.	<i>Génexpresszió mérés</i>	15
3.2.8.	<i>Egyéb mikroarray alapú módszerek</i>	16
3.3.	<i>Állatmodellek</i>	16
3.3.1.	<i>Állatmodellek előnyei</i>	16
3.3.2.	<i>Állatmodellek hátrányai</i>	18
3.3.3.	<i>Kísérleti betegségmodellek</i>	18
3.3.4.	<i>QTL analízis egérben</i>	18
3.4.	<i>Irodalom</i>	28
3.5.	<i>Fejezethez tartozó kérdések</i>	29

3. Multifaktoriális betegségek genomikai vizsgálati módszerei

Ebben a fejezetben a multifaktoriális betegségek genomikai vizsgálatok fontosabb módszereit, és a hozzájuk tartozó elméleti alapokat foglaljuk össze. A csak genetikai vizsgálatokra alkalmas módszereket, illetve a módszerek technikai hátteréről más forrásokból tájékozódhatunk. A genetikai, statisztikai alapokról itt csak annyiban lesz szó, amennyi szorosan kapcsolódik a tárgyalt módszerek megértéséhez. Statisztikát, bioinformatikát részletesebben a következő fejezetben tárgyalunk.

3.1. Statisztikai alapok

3.1.1. Hardy Weinberg eloszlás

Hardy Weinberg egyensúly (HWE) a genotípusok várható eloszlását írja egy random populációban.

Az allélok relatív gyakoriságát két egyenlettel írhatjuk le.

$$p + q = 1 \text{ és } p^2 + 2pq + q^2 = 1$$

Ahol:

- p a gyakoribb (major) allél frekvenciája
- q a ritka (minor) allél frekvenciája
- p^2 és q^2 a homozigóta egyedek gyakorisága
- $2pq$ a heterozigóta egyedek gyakorisága

A genotípus várható eloszlásának számszerű értékeit úgy számolhatjuk ki, hogy a kapott értékeket megszorozzuk a vizsgált populációban levő egyedek számával.

A HWE-től való eltérés ellenőrzését minden populációgenetikai vizsgálatnál el kell végezni, amelyben a vizsgált csoportok genotípus eloszlását vizsgáljuk valamilyen szempontból. Az összehasonlítást, amelyben az elméletileg várható eloszlást hasonlítjuk össze a kapott eloszlással, χ^2 statisztikával lehet elvégezni, és manapság remek *on line* internetes programok állnak rendelkezésre, amelyekkel könnyedén el tudjuk végezni a tesztet. Ilyen web oldal pl. a <http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl>, amellyel egyszerre több allélnál is el lehet végezni a vizsgálatot, ráadásul, ha két populációt hasonlítunk össze, az oldal asszociációs tesztek is végez.

A HWE elméletével itt csak annyiban foglalkozunk, hogy mit jelenthet általában ezekben a vizsgálatokban, ha a kapott eloszlás szignifikánsan különbözik a várttól. Az eltérés több okra vezethető vissza:

- A genotipizálás hibás volt.
- A mintavétel nem véletlenszerű. pl.: sok rokon van a mintában.
- Beltenyészet (*inbred*) populáció. Mindkét utóbbi esetben a homozigóták aránya megnő.
- A vizsgált allél egy ismétlődő (*repeat*), pl. CNV régióban helyezkedik el. Ebben az esetben általában a heterozigóták arány nő meg.
- A vizsgált allél valamilyen szerepet játszik a vizsgált populáció fenotípusában. Pl.: a cisztikus fibrózis (CF) monogénes betegség nagy részéért felelős CFTR gén $\Delta F508$ -es allélt vizsgálva, CF-esekben a homozigóták túlsúlya mutatható ki.

Ha egy genotípusnál eltérést tapasztalunk a HWE-ben a kontroll populációban, akkor az általában kizárja azt az allélt a további vizsgálatokból, hiszen hibás eredményekhez vezethet. A beteg populáció esetében is (case-control vizsgálatoknál) az első négy eset mindegyikében hasonló a helyzet, de itt már ennek megállapítása eseti elemzéseket igényel, pl. másik módszerrel is el kell végezni a genotipizálást, vagy a populációban tapasztalható rokoni kapcsolatokat más módszerekkel is vizsgálni kell. De a genomikai típusú elemzéseknél, amikor egyszerre nem egy, hanem esetleg több 100 ezer, vagy millió genotípust vizsgálunk, ez a többi genotípus elemzésével tisztázható. Tulajdonképpen számunkra a legutolsó eset a legérdekesebb. Ha egy genotípus hajlamosít a vizsgált betegségre, akkor az a vártnál gyakrabban fordulhat elő a beteg populációban, mint az egészségesben. Ha véd a betegség kialakulásával szemben, akkor pedig ritkábban. Mindkét esetben értékes információhoz jutottunk az illető genotípusról, amit persze más módszerekkel még igazolni kell, hiszen, mivel itt egy statisztikai próbát végzünk, nem zárható ki a véletlenszerű tévedés lehetősége. Ritkán ezek ellentéte is előfordulhat, azaz pl. egy másik lókuszon található allél okozza a betegséget, és a gyakrabban, vagy ritkábban előforduló genotípus befolyásolja az egyedek túlélésének esélyét. Természetesen, mivel itt is asszociációról van szó, erre is érvényesek azok a megállapítások, amelyet az asszociáció címszó alatt tárgyalunk.

3.1.2. Kapcsoltság és haplotípus

Ismert, hogy az ivarsejtek fejlődésekor a **meiózis** során a homológ kromoszómák között **crossing over**, vagy **genetikai rekombináció** zajlik le, azaz a két homológ kromoszóma genetikai anyaga részben kicserélődik egymással (1. ábra). Például, az emberi hímvarsejtek meiózisa során átlag 49 crossing over történik. Ennek, szempontunkból az a jelentősége, hogy a korábban egy kromoszómán, két egymás melletti marker elkerülhet egymás mellől. Mivel az asszociációs és kapcsoltsági vizsgálatok esetén genetikai markereket használunk, ennek, mint látni fogjuk, fontos következményei vannak. Annak jellemzésére, hogy egy populációban két allél milyen eséllyel öröklődik egyszerre, vezették be a **linkage disequilibrium (LD)**, kapcsoltsági kiegyensúlyozatlanság) fogalmát. Ez a fogalom azt takarja, hogyha két allél egymástól függetlenül öröklődik, akkor a populációs eloszlásuk egymáshoz képest random, véletlenszerű (azaz köztük egyensúly van). Ha azonban valami oknál fogva, (általában azért mert egymás közelében vannak így nem egymástól függetlenül öröklődnek) ez a véletlenszerű eloszlás megszűnik, azt mondjuk, hogy kapcsoltsan öröklődnek, és valamilyen fokú LD van közöttük. Egy másik **definíció**: egy kromoszómán lévő két markerpozíció között linkage disequilibrium áll fenn, ha a két pozíción található allélok tekintve bizonyos allélkombináció gyakorisága eltér az egyes allélok gyakoriságának szorzatától.

Példa: két SNP, mindkettőn 50-50% gyakorisággal A, ill. G nukleotidok fordulhatnak elő. Ha nincs LD a két pozíció között, azaz egymástól függetlenül öröklődnek akkor az AG kombinációnak $50\% \times 50\% = 25\%$ gyakorisággal kell előfordulnia egy populációban. Ha a várt 25%-os együttes előfordulás helyett 40%-ban fordul elő az AG kombináció egy populációban, azt jelenti, hogy nem egymástól függetlenül öröklődnek.

Az LD-t koeficienssel szokták jellemezni. Leggyakrabban két ilyen koeficienszt használnak: a standardizált LD koeficienszt D' -vel az ún. korrelációs koeficienszt pedig r^2 -tel jelölik. A két koeficienszt eltérő módon számolják (lásd:

http://en.wikipedia.org/wiki/Linkage_disequilibrium), de az értékük két szélső értéke és azok jelentése megegyezik egymással. Mindkét koeficiens esetében 0 azt jelenti, hogy a két allél egymástól függetlenül öröklődik (azaz egymással egyensúlyban „equilibriumban” van), míg az 1-es érték teljes kapcsoltságot jelent, azaz a két allél abban a populációban mindig együtt fordul elő. Ezt úgy szokták interpretálni, hogy a két allél egymással teljes LD-ben van. Az 1 közeli értékek mindig erős kapcsoltságra utalnak.

Ha két vagy több allél egymás mellett van, és egyszerre fordulnak elő, akkor azt mondjuk, hogy egy **haplotípuson** vannak. Egy másik definíció szerint: Ha több, egymás melletti allél gyakran fordul elő különböző emberekben egyszerre, azaz együtt öröklődnek (köztük csak ritkán van crossing over) akkor azt mondjuk, hogy ezek az allélok egy haplotípuson vannak. Hogy egy adott populációban milyen gyakran fordulnak elő egyszerre az allélok, annak jellemzésére a **haplotípus frekvenciát** szoktuk használni. A különböző populációkban eltérő haplotípusokat és haplotípus frekvenciákat lehet találni. Ennek feltérképezésére indult el 2002-ben a **HapMap project**, a HGP folytatásaként (ld. <http://hapmap.ncbi.nlm.nih.gov/> és http://en.wikipedia.org/wiki/International_HapMap_Project), melynek azóta már 3 fázisa is lezajlott (1,2).

Genomikai vizsgálatoknál fel szokták rajzolni a vizsgált populációk, és SNP-k haplotípus térképét. A leggyakrabban használt ábrázolási módot a 2. ábrán mutatjuk be. A haplotípusok megállapítására és frekvenciájuk kiszámítására szintén *on line* szoftverek állnak rendelkezésre, pl. Haploview 4.1: <http://www.broad.mit.edu/mpg/haploview/>.

Még térjünk vissza a kapcsoltságra. Genomikai/genetikai vizsgálatoknál a kapcsoltságot két értelemben szoktuk használni. Az első jelentése, ahogy előbb is kifejtettük: kapcsoltság lehet két genetikai lókuszon elhelyezkedő két allél között (azaz együtt öröklődnek). A másik jelentése, hogy kapcsoltság lehet egy allél, vagy egy haplotípus és egy fenotípus között (pl. betegség, hajszín, szemszín, IQ, koleszterinszint stb.). Ilyenkor feltételezni lehet, hogy ez az allél (vagy a vele kapcsoltságban lévő másik allél, vagy egy egész haplotípus) befolyásolja annak a fenotípusnak a manifesztálódását.

3.1.3. *LOD score*

Bár a *LOD score* jelentősége az utóbbi években csökkent, régebbi genomikai publikációkban szinte mindig ezt az értéket használták annak jellemzésére, hogy egy genetikai marker és egy fenotípus között milyen erős a kapcsoltság. Itt most a teljesség igénye nélkül a *LOD score* jelentését tisztázzuk. Részletesebben lehet tájékozódni ingyenesen letölthető *on line* könyvekből, pl. <http://www.ncbi.nlm.nih.gov/books/NBK7580/>, vagy más internetes forrásokból, pl. http://en.wikipedia.org/wiki/Genetic_linkage.

A *LOD* jelentése **logarithm of odds**, azaz az **esély 10-es alapú logaritmus**a. A *score* az angol érték szó. A *LOD score* egy statisztikai fogalom. A statisztikában sokszor használjuk a nullhipotézist, ami általában egy hipotézis elvetését jelöli, és azt vizsgáljuk, hogy mennyire állja meg a helyét a nullhipotézis. A $p = 0,05$ -ös érték azt jelenti, hogy ekkora az esélye a nullhipotézisnek, azaz 0,95 (95%), az esélye, hogy nem igaz. Más szavakkal, 0,05-ös p értéknél 20 esetben egyszer (100 esetben ötször) tévedhetünk. Kapcsoltsági vizsgálatokban a nullhipotézis azt jelenti, hogy a két lókuszt között nincs kapcsoltság. **LOD = logarithm of odds, annak az esélynek a 10-es alapú logaritmus, hogy két lókuszt genetikailag kapcsolt, összehasonlítva annak az esélyével, hogy nem kapcsolt.** Pl. $LOD = 3$ azt jelenti, hogy 1000-szeres a valószínűsége (10^3) annak, hogy két lókuszt együtt öröklődik, mint annak, hogy nem. A *LOD score* =3 felel meg a $p=0,05$ -ös valószínűségnek (magyarázatát ld. pl. az előbb említett könyvben).

A *LOD score*-os elemzést használták, amikor még a 90-es évek elejének környékén feltérképezték a monogénes betegségek genetikai hátterét. Ilyenkor ún. **parametrikus LOD score**-t használtak, azaz paraméterként meg lehetett adni a családfa vizsgálatok alapján az öröklődés típusát (pl. recesszív), és a betegség penetranciáját (**penetrancia** = %-ban kifejezett érték, amely azt mutatja, hogy a betegséget okozó mutációt hordozók hány százaléka beteg). A multifaktoriális betegségeknek ilyen paramétereket nem lehet megadni, ezért itt a **nem-parametrikus LOD score** számítást használják.

A *LOD score*-t pl. a teljes genomszűréskor használják (ld. később). Ezekben a vizsgálatokhoz olyan családokat keresnek, amelyben egy testvérpár mindkét (vagy több testvér esetén több), tagja beteg a tanulmányozott betegségben. Ezt **affected sib pair (ASP)** analízisnek hívják. Ezekben a vizsgálatokban a szülőknél (esetleg egyéb rokonokban) és nagyszülőknél is elvégzik a genomikai szűrést. A bonyolult statisztikai elemzést pénzért megvásárolható szoftverekkel (pl. GeneHunter) lehet elvégezni. Minél magasabb *LOD* értéket kapunk egy markerről, annál nagyobb a valószínűsége, hogy az illető markernek van valamilyen köze a betegséghez, azaz pl. kapcsolt egy mutációhoz, amely hozzájárul a betegségre való hajlammal. Ezért ezt a vizsgálatot **kapcsoltsági vizsgálatnak** is szokták nevezni. A 3. ábra egy ilyen kapcsoltsági vizsgálat eredményét, azaz a *LOD score* görbét mutatja. Általában *LOD score* >3 esetén beszélünk bizonyított kapcsoltságról, míg <-2 esetén biztosan kizárjuk a kapcsoltságot.

3.1.4. Pozicionális klónozás

Pozicionális klónozásnak nevezzük azt a technikát, amikor pl. a kapcsoltsági vizsgálatoknál az asszociált, a LOD score csúcs alá eső genomrégiókat különböző újabb markerekkel tovább szűkítjük, majd pl. *in silico* és laboratóriumi módszerekkel (pl. génexpressziós mérés, DNS szekvenálás, funkcionális vizsgálatok) megkeressük az asszociációért felelős gént, és annak mutációját. Részletesebben: http://en.wikipedia.org/wiki/Genetic_screen#Positional_cloning.

3.1.5. Founder populációk

Mivel két lókuszt közötti, a meióziskor bekövetkező crossing over valószínűsége megközelítőleg arányos a két lókuszt egymástól való távolságával, ezt felhasználták, az ún. **genetikai távolság** becslésére, és bevezették (Thomas Hunt Morgan Nobel-díjas genetikus tiszteletére) a **centiMorgan (cM)** mértékegységet. Ennek alapján **két lókuszt között 1 cM a genetikai távolság, ha annak a valószínűsége, hogy köztük crossing over következik be 1%.** Régebben, a humán genom megszekvenálása előtt ezt a mértékegységet használták két lókuszt közötti távolság megadásakor. Mivel a rekombináció mértéke (a két homológ X kromoszóma miatt) enyhén nagyobb a nőkben, ezért a genetikai távolságok általában kisebbek férfiakban. Manapság alkalmazása kezd kiszorulni, és egyre inkább a bázisokban megadott fizikai távolságot használjuk, bár markerek távolsága esetén, főleg családvizsgálatokban, sokszor praktikus a genetikai távolságot (is) megadni. Körülbelül, 1 cM = 1 Mb (megabázis) fizikai távolságnak felel meg.

A meióziskor bekövetkező crossing over következtében, ha egy mutáció keletkezik egy családban, és az, nemzedékeken át továbböröklődik, a közelében található lókusztok egy idő után egy bizonyos eséllyel elkerülnek mellőle. Minél messzebb van egy lókuszt, annál nagyobb eséllyel. Ennek, kapcsoltsági vizsgálatoknál az a következménye, hogyha egy betegséget okozó mutációt egy kapcsolt marker segítségével szeretnénk detektálni, annál közelebbi markert kell használni, minél régebben keletkezett a mutáció. Ez teljes genomszűréseknél azt jelenti, hogy nagyon sűrű genetikai markereket kell alkalmaznunk, hogy jó eséllyel használjunk olyan markert, ami a betegséget okozó mutációval kapcsolt. Minél távolabbi rokonságban állnak egy populáció tagjai egymástól, annál sűrűbben elhelyezkedő markereket kell használnunk. Ehhez még azt is hozzá kell tennünk, hogyha történelmi távlatokban gondolkodunk, akkor minden ember rokonságban áll egymással. Például, becslések szerint az UK jelenlegi lakosságából két egymással nem rokon embernek átlagosan 22 generációval ezelőtt volt közös őse, azaz 44 meiózis választja el őket egymástól. Ennek az a következménye, hogy a közös ősből 3 cM-ra levő lókusztok esetén $(1-0,03)^{44}=0,26$ az esélye, hogy a két nem-rokon emberben is egymás mellett maradjanak. Ez úgy jön ki, hogy 3 cM azt jelöli, hogy 3% az esély a rekombinációra a két lókuszt között. Annak az esélye, hogy nincs rekombináció $1-0,03 = 0,97$, amit a 44 generáció miatt, ennyiszor kell összeszorozni. 20cM távolságban levő lókusztok esetén, $(1-0,2)^{44}=5 \times 10^{-5}$ az esély ugyanerre. Mivel a humán genom cM-ben kifejezett mérete 3.000 cM, ki lehet számolni, hogy az UK populációban, milyen sűrűn kell a markereket elhelyezni, hogy jó (mondjuk 95%-os) esélyünk legyen arra, hogy minden mutációt megtalálunk.

A humán kapcsoltsági vizsgálatokban itt lehetett felhasználni, az ún. „founder populációkat”.

Founder populáció: kisszámú ősről visszavezethető beltenyésztet populáció, azaz olyan populáció, melyet vissza lehet vezetni kisszámú családra, vagy egyénre. Ezek valamilyen oknál fogva izoláltan élnek (pl. földrajzi (pl. kis szigeten élnek), vagy társadalmi, vallási okokból csak egymás között köthetnek házasságot), emiatt a rokoni távolság sokkal kisebb közöttük, mint egy nyitottabb populációban. Ilyen populációt alkotnak, pl. a finnek, quebec-i francia-kanadai populáció, izlandiak, a hutterite, vagy az amish közösségek. Egy vizsgálatban bizonyították is ezt:

20 mikroszatellita markert vizsgáltak 664 brit és 430 finn egyénben a 18q21 régióban. 1 cM távolságban levő markerek között teljes volt az LD mindkét populációban. 1-3 cM esetén 75 markerből 20 volt kapcsolt egymással a brit populációban (20/75 (26,6%)), 61 a finn populációban (81,3%). >3cM távolságú markerek esetén, a briteknél már nem voltak LD-ben egymással a markerek: 0/62 (0%); a finnekben viszont még 20 egymással LD-ben lévő markert találtak: 20/66 (30,3% finn). Tehát csak a finn populációban volt esély arra, hogy egy markerrel kimutassanak egy >3 cM-ra elhelyezkedő mutációt. Emiatt genetikai vizsgálatokban előszeretettel használják a founder populációkat, és számos betegség genetikai hátterének tisztázásakor jelentős eredményeket értek el ezek felhasználásával. Például, egyes cégek founder populációkat ajánlanak fel ezekhez a vizsgálatokhoz, ld.: <http://www.genizon.com/english/discovery/founder.html>, ahol a quebec-i francia-kanadai populációt ajánlják genetikai vizsgálatokhoz.

3.1.6. Asszociációs vizsgálatok

Valamilyen jellemző (pl. betegség) genetikai hátterének tisztázására jelenleg a legnépszerűbb módszer az asszociációs vizsgálat. Ilyenkor marker-genotipizálással, majd statisztikai módszerekkel azt vizsgáljuk, hogy egy marker milyen eséllyel asszociál egy betegséggel.

Az **asszociáció** egy statisztikai kijelentés. Ha egy marker asszociál egy fenotípussal az azt jelenti, hogy az adott allél (marker) szignifikánsan gyakrabban fordul elő együtt az adott fenotípussal, mint az várható.

Pozitív asszociációnak számos oka lehet:

- Direkt hatás: a vizsgált allél okozza a betegséget
- Természetes szelekció. Az illető allél megnöveli a túlélés esélyét a tanulmányozott betegséggel szemben
- Populációs rétegződés (**population stratification**): egyes népcsoportokban bizonyos allélok gyakrabban fordulnak elő. Pl. evőpálcika gén (HLA-A1 gyakoribb a kínaiakban)
- Statisztikai hiba (ún. egyes típusú hiba, azaz hamis pozitivitás)
- A vizsgált allél LD-ben van a betegségben szerepet játszó alléllal (pl. mutációval).

A fenti okok közül, itt kettőt részletezünk (a statisztikai hibákról, ld. pl. az előző fejezet). Az egyik a populációs rétegződés, ami az egyik legnehezebben korrigálható problémát okozza. Ez azt jelenti, hogy ilyen típusú populációs vizsgálatoknál fontos szempont, hogy az összehasonlítható két populáció populációs összetétele megegyezzen egymással. Hogy milyen problémát okozhat, ha ez nem teljesül, a legismertebb elméleti példázat az evőpálcika-gén esete. Ez a példázat röviden azt mondja: tegyük fel, hogy azt akarjuk, kideríteni, hogy van-e annak a képességnek, hogy valaki tud-e evőpálcikával enni genetikai háttere? Gyűjtünk hozzá két populációt, amelyek közül az egyik tud evőpálcikával enni, a másik nem (kontroll populáció). Abban az esetben, ha az első populáció főleg kínaiakból áll, a másik pedig nem, akkor azt fogjuk találni, hogy a kínaiakban gyakori (és európaiakban ritkább) bizonyos HLA-A2-es allél erős asszociációban áll az evőpálcikával evés képességével. Ez nyilvánvalóan hamis asszociáció, amit a két populáció helyes megválasztásával el lehet kerülni. Azonban ez nem mindig ilyen egyértelmű. Főleg a mai globalizált világban, gyakran élnek együtt különböző etnikai csoportok, részben kevert, sokszor vegyes genetikai háttérrel (pl. USA-ban afroamerikaiak, hispán-amerikaiak, európai eredetű amerikaiak stb., vagy Magyarországon a romák és nem-romák stb.), és az etnikai besorolás, sokszor pl. etikai okok miatt nagyon nehézkes, vagy akár lehetetlen. Amikor két eredetileg más etnikumhoz tartozó populáció genomszinten keveredik egymással, **population admixture**-nek nevezzük, és ezt a jelenséget

egyes genetikai vizsgálatoknál fel is lehet használni (ld. *population admixture mapping* a 9. fejezetben).

A populációrétegződésből adódó hibáknak az elkerülésére számos módszert dolgoztak ki. Pl. a belső kontroll módszerek. Ilyen a **transmission disequilibrium test (TDT)** alkalmazása. Ez ugyan 50%-kal több munkával jár, hiszen beteg + szülők is kellenek hozzá. Azokat a szülőket választják ki, akik heterozigóták a betegséggel asszociáló M1 markerre, és azt vizsgálják, hogy hány szülő adja át az M1 allélt a beteg gyermekébe vs. hány nem. Ha az M1 marker nem asszociál a betegséggel, annak esélye, hogy a beteg megkapja a szülőtől 50 %, ha asszociál, akkor ennél nagyobb. Ld.: http://en.wikipedia.org/wiki/Transmission_disequilibrium_test.

Egy másik módszer a **discordant sib pair** analízis. Itt olyan testvérpárokat vizsgálnak, melyek közül az egyik beteg, a másik nem.

Manapság a genomikai módszerek fejlődésével egyszerre rengeteg markert tudunk vizsgálni. Ezzel kapcsolatban már kifejlesztettek olyan statisztikai módszereket, amelyek korrigálni képesek az eltérő etnikai háttérrel rendelkező populációk összehasonlításából eredő statisztikai torzításokat.

A másik téma, amit itt még ki kell emelni, hogy a vizsgált allél LD-ben van a felelős alléllal. Ez azért fontos, mert pozitív asszociáció esetén a legnagyobb valószínűséggel ez következik be. Az egyik fontos feladat annak megállapítása, hogy a marker direkt hatása, vagy az LD-ben levő allél felelős a kapott asszociációért. A legjobb módszer, ha laboratóriumi körülmények között, *in vitro*, vagy *in vivo* módszerekkel, állatkísérletekkel igazoljuk a funkcionális hatást. Manapság annak is megnőtt az esélye, hogy *in silico* módszerekkel találunk valamit. Az interneten számos olyan adatbázis, szoftver található, amelyekkel egy allélhez funkcionális hatást lehet kapcsolni. Például, a variáció megváltoztatja egy transzkripció faktor kötőhelyet, megváltoztatja a kódolt protein szerkezetét, miRNS szekvenciát, vagy kötőhelyet befolyásol, szabályozó szekvenciát változtat meg, stb. A vizsgált alléllal LD-ben lévő felelős allél azonosítása történhet direkt szekvenálással, vagy a populáció haplotípus térképe alapján keresünk markerünkkel szoros LD-ben lévő másik polimorfizmust. A későbbi fejezetben több példát mutatunk be, amelyek mutatják, hogy mennyire nehéz egy marker és a betegség közötti kapcsolat pontos okát kideríteni. A jelenlegi genomikai vizsgálatokban ez az egyik szűk keresztmetszet.

3.1.7. Mintagyűjtések típusai

A betegségek pathomechanizmusának, genetikai és környezeti tényezőknek a vizsgálatára két alapvető mintagyűjtési módszert különböztetünk meg. Az egyik a retrospektív mintagyűjtés (**retrospective study**), ahol általában két populációt, egy beteget és egy kontrollt gyűjtünk össze. Ezt technikailag viszonylag egyszerű lebonyolítani, akár egyetlen szakorvos is könnyedén elvégezheti, egyszerűen a hozzá járó betegektől, illetve egy kontroll populációtól megfelelő biológiai mintát vesz, és rögzíti (pl. kérdőív kitöltésével) a vizsgálatához szükséges laboratóriumi, klinikai, kezelési, környezeti, viselkedési stb. adatokat. Itt nagyon fontos, hogy az adatok felvétele gondosan, alaposan és előre megtervezetten történjen, hiszen ezeken az adatokon nagyban múlik az értékelések minősége. A könnyű kivitelezhetőség és a gyors elvégezhetőség miatt a genetikai, genomikai vizsgálatok túlnyomó többségénél ilyen vizsgálat folyik. Angolban **case-control study**-nak hívják ezt a vizsgálatot. Számos ilyen nagy vizsgálatot ismerünk. Kifejezetten a genetikai háttér kutatását célozta meg a 2005-ben indult **WTCCC** (Wellcome Trust Case-Control Consortium). Itt 50 kutatócsoport együttműködésével 16 ezer beteg és 3 ezer egészséges ember mintájának segítségével a gyakori variációk (SNP és CNV) és a betegségek összefüggéseinek vizsgálatát tűzték ki maguk elé. A projekt sikeres volt, hiszen 90 új genetikai variációt azonosított, amelyek valamilyen szerepet játszanak gyakori betegségekre való hajlamban. A projekt sikere főleg

abban rejlett, hogy a nagyszámú minta és a szigorú statisztikai elemzések következtében a talált új variációk mindegyike nagy valószínűséggel valós asszociációt mutat, szemben a korábbi eredményekkel, ahol az eredmények túlnyomó többségét nem tudták egyértelműen igazolni. Az eredmények másik nagy jelentőségét az adja, hogy mivel nem célzott géneket, variációkat vizsgáltak, hanem hipotézis mentes vizsgálatokat folytattak (GWAS, ld. később) a talált gének nagy része a betegségekkel kapcsolatos új anyagcsereutakat, és pathomechanizmusokat tárt fel, megadva a lehetőségét új típusú kezelések, gyógyszerek kifejlesztéséhez (3).

A WTCCC sikere nyomán 2008. áprilisában létrehozták a nemzetközi WTCCC2-t, amelyben összesen 120 ezer minta mérését és elemzését tűzték ki célul (teljes genom asszociációs vizsgálatok segítségével, ld. később) különböző gyakori betegségekben, de olyan poligénes jellegekben is, mint a matematikai és az olvasási képességek, vagy a statin kezelésre adott válasz.

A prospektív mintagyűjtésnél (**prospective study**) egészséges populációtól gyűjtünk mintát, majd sorsukat (pl. visszahívásokkal) sokszor évtizedekig nyomon követjük, és összefüggéseket keresünk a bennük kialakuló betegségek és a különböző vizsgált, pl., genetikai, laboratóriumi és környezeti tényezők között. Ez jóval nagyobb szervezési munkát, több gyűjtött beteget és hosszabb időt igényel, így drágább, mint a retrospektív, viszont számos előnye van. A retrospektív vizsgálatoknál számos torzítás lehetséges pl. a mintagyűjtésnél. Így a betegségben meghaltak mintái nyilván alul-reprezentáltak.

Több híres prospektív vizsgálatot ismerünk. Az egyik a **Framingham heart study**, amely 1948-ban indult az USA-beli Framingham városában 5209 férfi és nő részvételével, és ma már a 3. generáció vizsgálatával azóta is folyik. Mai tudásunk nagy része a kardiovaszkuláris betegségek kockázati faktorairól ebből a vizsgálatból ered (ld.

<http://www.framinghamheartstudy.org/>).

Még nagyobb ilyen vizsgálat az **UK biobank** projekt, mely 2007-ben kezdődött és 500 ezer, 40-69 éves ember mintáinak és adatainak összegyűjtését célozta meg, mely azóta teljesült is 2010-ben. A vizsgálat fő célja a 21. század betegségeinek kutatása. Részleteket ld. a UK biobank honlapján: <http://www.ukbiobank.ac.uk/>.

3.1.8. Kockázatszámítás

Asszociációs vizsgálatoknál számszerűsíteni szokták a talált összefüggések erősségét. Az egyik ilyen jellemező a p érték, amely azt mutatja, hogy mekkora a valószínűsége a hamis asszociációnak. A szignifikancia határ általában $p = 0,05$. Az ennél kisebb értékeket fogadjuk el szignifikáns, azaz statisztikailag igazolt asszociációnak. Ld. még Bonferroni korrekció (2. fejezet).

A kockázat számolásnál használt p értékkel összefüggésbe hozható fogalom retrospektív vizsgálatoknál az **odds ratio** vagy **OR** érték. OR jelentése: az esély, hogy az illető allél (vagy lókus) asszociál a betegséggel a betegekben osztva az eséllyel a kontroll csoportban.

Részletesebben: http://en.wikipedia.org/wiki/Odds_ratio .

Ezzel rokon a prospektív vizsgálatoknál használt **relative risk**, vagy **RR** érték. RR jelentése: az asszociáció valószínűsége a betegcsoportban osztva az asszociáció valószínűségével a kontrollcsoportban. Ld.: http://en.wikipedia.org/wiki/Relative_risk.

Mindkét érték azt mutatja, hogy az adott genetikai variáns hordozása hány szorosára növeli meg az illető kockázatát a betegség kialakulására. Az 1-nél nagyobb érték kockázat növekedést, a kisebb érték kockázat csökkenést jelent. Fontos még megadni az érték 95%-os konfidencia (95% CI) határát is. Ez azokat az értékhatárokat mutatja, amelyeken belül az összefüggés 95%-os valószínűséggel igaz. Az összefüggés akkor fogadható el általában, ha a két szám közötti érték nem lépi át az 1-et. Pl. az $OR = 3,2$ (2,4-4,8) nem lépi át, így

elfogadható az összefüggés, míg pl., $OR = 1,8$ (0,8-3,6) átlépi, így nem valós az összefüggés. Egy marker vizsgálatánál $p = 0,05$ az a határ, amely fölött átlépi, alatta nem lépi át az OR 95%CI-je az 1-et. Ez mutatja, hogy a szignifikancia határ és a kockázatértékek között összefüggés van.

3.1.9. Genetikai markerek

Minden olyan genetikai variáció használható genetikai markernek, amellyel jellemezni lehet egy lókuszt, és jó eséllyel, két homológ kromoszómán levő lókuszt között különbséget lehet tenni.

Az egyik legnépszerűbb marker a **mikroszatellita**, amelyet **short tandem repeat**-nek vagy **STR**-nek is szoktak hívni. Ezek 1-6 ismétlődésből (más definíció szerint 1-4 ismétlődésből) álló szekvencia szakaszok, melyek általában neutrálisak. Ilyen pl. a TA ismétlődés, pl. TATATATA, amelyet $(TA)_n$ -nel szokás jelölni (itt $n=4$). Itt a polimorfizmus abból adódik, hogy az ismétlődések száma két homológ kromoszóma között nagymértékben különbözhet. Azaz, pl. $(TA)_n$ -nél az n értéke széles skálán mozoghat. Egy STR-nél szokás megadni a diverzitás fokát, azaz a **heterozigótaság valószínűségét**. Ez alapján a **diverzitás foka** egy markernél annak a valószínűsége, hogy egy marker egy véletlenszerűen kiválasztott emberben heterozigóta; 0 ha nem variábilis, 1 ha nagyon variábilis. Ez utóbbi értéket nyilván nem veheti fel, mert ez azt jelentené, hogy a marker minden emberben más. Informatív a marker ha a diverzitás foka $\geq 0,7$, azaz $>70\%$ a valószínűsége, hogy egy véletlenszerűen kiválasztott emberben a marker heterozigóta. Az STR-eknek az előnye az SNP-kel szemben, hogy egy STR-nek több allélja is lehet (akár >10), szemben az SNP-knél általános 2-vel. Az emberi genom kezdeti feltérképezéséhez, a monogénes betegségek genomikai hátterének felderítéséhez, teljes genom szűrésekhez mikroszatellita markereket használtak, és sokszor használnak még ma is. Sőt bizonyos esetekben, pl. emberek (bűnözők, áldozatok, rokonok stb.) azonosítására még ma is ezek a hivatalos markerek. A humán genomban kb. 30 ezer STR markert ismerünk. A humán genom, vagy egér genom szűréseket általában 5, vagy 10 cM-os szettel végezték, ami azt jelenti, hogy két marker egymástól való távolsága a humán genomban 5, illetve 10 cM. Ez utóbbi a humán genomban 811 markert jelentett.

Fő előnyei:

- (1) a nagyfokú polimorfizmusból adódó diverzitás. Két ember között jóval kevesebb STR-rel tudunk különbséget tenni, mint a biallélikus SNP-kel. Pl. igazságügyi orvosszakértői azonosításnál használt 10-15 STR-rel összehasonlítva, 20-50 SNP ad azonos szintű információt.
- (2) Régebben használják, ezért pl. az igazságügyi adatbankok mind STR alapúak, így ezek használatát preferálják.

Hátrányai:

- (1) bonyolult a kimutatása (ld. [http://en.wikipedia.org/wiki/Microsatellite_\(genetics\)](http://en.wikipedia.org/wiki/Microsatellite_(genetics))). Az SNP-knél az utóbbi években óriási fejlődés történt, amellyel az STR-ek nehézkes kimutatása nem tudott lépést tartani.
- (2) Jóval kevesebb van belőlük, mint az SNP-kből (kb. 30 ezer, vs. >30 millió)
- (3) Mutációs rátája 100 ezerszer magasabb, mint az SNP-ké, így a rokonsági vizsgálatoknál hátrányban van.
- (4) Eloszlása nem olyan egyenletes, mint az SNP-ké, ezért vannak nem lefedett genomterületek.

Manapság az STR-ek hátrányai messze meghaladják előnyeiket, ezért humán genomikai vizsgálatokban egyre inkább SNP-ket használnak (ld. lent GWAS). De pl., már készül a 45

SNP-ből álló emberi azonosításra tesztelt univerzális teszt, amelyet a világ 44 populációjában teszteltek, és átveheti az igazságügyi vizsgálatokban használt mikroszatelliták szerepét. Az SNP-k előnyeit, mint markerek, tartalmazza az előző felsorolás. Hátránya abból származik, amelyik egyben az előnye is, hogy olyan mennyiségben lehet egy mérésben meghatározni őket, amely egy sor új, elemzési, statisztikai problémát vet fel (ld. előző fejezet).

3.1.10. Power analízis

Nagyon sokszor szükség van arra, hogy megmondjuk, hogy egy adott eredmény mennyire megbízható, illetve fordítva, hogy pl. egy populáció genetikai vizsgálatban hány embert kell gyűjtenünk és genotipizálnunk, hogy megbízható eredményt kapjunk. Ezeket a tesztek az ún. „power” analízis segítségével tudjuk elvégezni, amelyre *on line* programok állnak rendelkezésünkre, (pl.: (<http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html> és <http://pngu.mgh.harvard.edu/~purcell/gpc/>). Egy ilyen eredmény megbízhatósága több mindentől függ. Ha az a kérdésünk, hogy egy beteg-egészséges (case-control) összehasonlító vizsgálatban, ahol azt nézzük, hogy egy adott genetikai variáció milyen kockázati értékkel (pl. OR) rendelkezik a betegségre vonatkozóan, akkor a megállapításunk erejét, a „power”, azaz megbízhatóságát növeli a populációk nagysága, az allélfrekvencia, és az OR nagysága. A power analízis lehet „*a priori*”, ilyenkor a vizsgálat előtt szeretnénk megbecsülni a kívánt mintaszámot, és „*post hoc*”, ilyenkor a vizsgálat után elemezzük, hogy milyen megbízható az eredményünk.

Bár nincs standard értéke a kívánt megbízhatóságnak (power-nek), genetikai vizsgálatokban a 0,8 (vagy 80%) értéket szokták megfelelőnek elfogadni. Ez a szám a β és az α kockázat 4:1 arányára utal. Itt a β kockázat a II-es típusú hiba statisztikai valószínűsége, azaz a hamis negatív eredményé (általában 0,2), az α az I-es típusú statisztikai hiba, azaz a hamis pozitív eredményé (általában 0,05). Részletesebben ld. http://en.wikipedia.org/wiki/Statistical_power. A különböző faktorok befolyását mutatják a következő példák. Ha OR = 1,1, és a kockázati allél frekvenciája, $q = 10\%$, akkor 6-6.000 beteg és kontroll kell a 80% powerhez. 5% allélfrekvencia mellett, ugyanehhez több 10 ezer beteg és kontroll szükséges, sőt bizonyos kérdések megválaszolásához akár 100-100 ezernél is több. OR = 3-nál, és $q = 5\%$ -nál 150-150 körüli mintaszám már elég, OR = 3 és $q = 10\%$ -nál pedig 50-50 mintánál is megfelelően megbízható eredményeket kaphatunk.

Sajnos, a multifaktoriális betegségeknél néhány kivételtől eltekintve az OR = 1,05-1,2 körüli értékek a jellemzőek, így általában több 10 ezres mintaszámnál várható megbízható eredmény. Ezért van nagy jelentősége a > 100 ezres mintaszámot gyűjtő nagy projekteknek, mint az UK Biobank projekt, vagy a WTCCC2.

3.2. Betegségek genomikai hátterének vizsgálati módszerei

3.2.1. Genetikai variációk kimutatása

Két fő típust különböztetünk meg:

- **Hipotézis, preconcepció által irányított;** pl. jelölt gén asszociációs vizsgálat, gének szekvenálása, génexpresszió mérés stb. (általában a **genetikai módszerek**)
- **Hipotézismentes;** pl. teljes genomszűrés, teljes genom asszociációs vizsgálat (**genome wide association study = GWAS**), teljes genomszekvenálás, mikroarray mérések, (**genomikai módszerek**)

A betegségek genetikai hátterének kiderítésében, új gyógyszercélpontok keresésében, a személyreszabott gyógyászatban, azaz a gyógyszerkutatásban nagyon fontosak a genetikai variációk, hiszen számos kutatás igazolta, hogy a különböző betegségekre való hajlamban, vagy a gyógyszerekre való válaszban döntő jelentőségűek. A molekuláris genetikai fejlődésével lehetőség nyílt ezek vizsgálatára is, hiszen már a HGP-ben is rengeteg variációt azonosítottak, majd több nagy, nemzetközi projekt is ezek felderítését tűzte ki célul (pl. Human Variome Project, HapMap, 1000 Genome project).

Kezdetben a legnépszerűbbek az ún. jelölt (vagy kandidáns) gén asszociációs vizsgálatok voltak. Ezekben a vizsgálatokban olyan géneket választottak ki, amelyekről tudták, hogy szerepet játszanak a betegség kialakulásában, bennük genetikai variációkat kerestek, majd összehasonlították azok gyakoriságát beteg és egészséges populációban. Rengeteg ilyen vizsgálat történt a különböző betegségeken. Ezekkel a vizsgálatokkal azonban számos probléma volt. Először is ismerni kellett hozzá a betegség patomechanizmusát, azaz új gének (új gyógyszercélpontok, új patomechanizmusok) felfedezésére nem volt alkalmas. Másodszor, különböző okok miatt számos hamis pozitív eredmény született, és az eredményeket később más vizsgálok, más populációkban nem tudták reprodukálni. Végül, például az is problémát jelentett, hogy a korai módszerekkel általában egyszerre egy, vagy csak néhány variációt tudtak vizsgálni, és már akkoriban ismert volt, hogy az ún. multifaktoriális, vagy gyakori betegségeken (pl. cukorbetegség, atherosclerosis, magas vérnyomás, asztma stb.) több száz genetikai variáció egymásra hatásából alakul ki az a genetikai hajlam, amely a környezeti faktorok hatására végül betegséghez vezet.

Az első és az utolsó probléma megoldására a **teljes genomszűrés** módszerét fejlesztették ki. Ebben általában beteg testvérpárral (**ASP**) rendelkező családokban a genom teljes területén nagyjából egyenletesen elosztva variábilis mikroszatellita markereket határoztak meg, és azt nézték, hogy a betegekben milyen genomterületeken tér el a markerek eloszlása a várttól. A markerekhez valószínűségi értékeket rendeltek (LOD score), azaz, hogy milyen eséllyel asszociálnak a betegséggel (3. ábra). A módszer számos értékes és hasznos eredményt hozott, új gének felfedezéséhez vezetett, de számos probléma volt vele. Az egyik, hogy a mikroszatelliták meghatározása, ahol az ismétlődések pontos számát kell meghatározni, nagyon munka- és időigényes és drága is. Ezért egy-egy vizsgálatba csak néhány száz mikroszatellitát tudtak bevonni, ami azt jelentette, hogy az egyes mikroszatelliták egymástól nagyon távol (általában kb. 10 cM-ra) estek. Ebből kifolyólag viszonylag kicsi volt az esély, hogy egy betegség lókusszal egy mikroszatellita kapcsolatban legyen (ld. Founder populációk), azaz vele egyszerre öröklődjön. Ebből kifolyólag a legtöbb betegséghez kapcsolható variáció elveszett. A drágaságon és a nagy munkaigényen kívül még olyan problémák is voltak pl., hogy nehéz volt a mintagyűjtés (nehéz megfelelő számú, együttműködésre is hajlamos szülőkkal rendelkező beteg testvérpárt gyűjteni).

3.2.2. GWAS

A 2000-es évek fejlesztései során kiderültek, hogy az SNP-k kimutatása automatizálható, és óriási mennyiségben lehet őket egyszerre meghatározni. Párhuzamosan több cég is olyan rendszereket fejlesztett ki, amelyekben egyetlen chippel, vagy array-vel több 100 ezer SNP-t lehet meghatározni. A versenyben a meghatározások árát is sikerült erősen leszorítani. Ma már a több 100 ezer variációt kimutatni képes chippek ára 100\$ nagyságrendű. A chip fejlesztések során derült ki, hogy az SNP-k mellett nagy jelentőségük van a CNV-knek is. A cégek gyorsan reagáltak. Az időközben felfedezett CNV-kre jellemző markerek pillanatok alatt belekerültek a vizsgálatokba.

Jelenleg két vezető cég van a piacon. Az egyik az Affymetrix, melynek a 6.0 array-e 906 ezer SNP-t és 946 ezer CNV-re jellemző markert tartalmaz. A másik az Illumina, melynek jelenleg legfejlettebb terméke a HumanOmni2.5-8 BeadChip, amely 2,379 millió lókuszt jellemez, mely SNP-eket és CNV-eket is tartalmaz, ráadásul egy chippel, egyszerre 8 mintát, mindössze 200 nanogramnyi DNS-ből képes mérni. Közben más technikákban, így például a DNS szekvenálásban is óriásit fejlődött a tudomány. Ennek folyományaként olyan projektek indultak, mint pl. az 1000 genom projekt, amely különböző etnikumú populációkhoz tartozó, összesen 2500 ember megszekvenálását tervezi (4). A projekt már rengeteg új genetikai variációt tárt fel (4. ábra). Ezeknek a felhasználásával készült a fent említett Illumina chip, amely gyakorlatilag minden olyan variációról tartalmaz információt, melynek a populációs gyakorisága (a megszekvenált afrikai, ázsiai és európai populációkban) nagyobb, mint 2,5%. De, már folynak a fejlesztések mindkét cégnél a meglévő termékeik továbbfejlesztéséért. Így az Illuminánál készül az 5 millió markert tartalmazó chip, mely a >1% gyakoriságú variációkat képes kimutatni, vagy az Affymetrix etnikum specifikus chipje, mellyel ki lehet küszöbölni a történelmileg régen elvált embercsoportok különböző genetikai variációiból adódó nehézségeket.

A GWAS vizsgálatok általában úgy történnek, hogy a fenotípusos jellel (pl. beteg) rendelkező, illetve kontroll populációt genotipizálnak pl. a fent említett chippek segítségével, majd megkeresik, hogy melyik marker (SNP) gyakorisága különbözött a két populáció között. A statisztikai módszerek fejlődésével már folytonos változókat is fel lehet használni GWAS-okban. Például a vércukorszinttel, vagy vérnyomással asszociáló genetikai markerek keresésénél már nincs szükség két jól elkülöníthető populációra. Ha statisztikailag szignifikáns összefüggést találnak, akkor azt mondják, hogy a marker kapcsolt a jelleghez. Ez legtöbbször két dolgot jelenthet: (1) a marker a funkcionálisan befolyásolja a jelleget (pl. fehérje struktúrát, genom 3D struktúrát, génexpressziót); (2) a markerhez genetikailag kapcsolt (vele általában együtt öröklődő) szekvencia-variáció a jelleg funkcionális oka (ld. még asszociációs vizsgálatok). A módszer a **hipotézismentes** módszerek közé tartozik, azaz végrehajtása nem igényel előzetes genetikai tudást a jellegről.

Napjainkban a GWAS óriási lehetőségeket nyújt a multifaktoriális jellegek genomikai hátterének tisztázására, amit a különböző kutatócsoportok gyakorlatilag rögtön ki is használtak. Mivel ezek az eredmények nagy valószínűséggel valós összefüggéseket mutatnak, pl. a betegségek patomechanizmusának új aspektusait mutatják be, létrehoztak egy web oldalt is, amelyre összegyűjtve felkerülnek az eredmények (A Catalog of Published Genome-Wide Association Studies (<http://www.genome.gov/gwastudies/>)). Erre olyan vizsgálatoknak az eredményei kerülhetnek fel, amelyekben minimum 100 ezer SNP-t vizsgáltak, és csak olyan SNP-k, melyekre a p érték $9,5 \times 10^{-6}$, azaz ekkora a tévedés esélye. Az eredmények mennyiségére jellemző, hogy 2008. november 25. és 2011. novembere között 1062 publikáció és 5267 SNP került fel az oldalra.

3.2.3. GWAS eredmények értékelése

A GWAS eredmények értékelése különösen nagy kihívás elé állította a bioinformatikusokat. Mint ahogy már korábban is írtuk, a hamis pozitív állítás elkerülése miatt nagyon szigorú statisztikai szabályokat kellett bevezetni az eredmények értékelésénél. A legegyszerűbb, és legtöbbször használt a Bonferroni korrekció, amikor a 0,05-ös p értéket osztjuk a vizsgált SNP-k számával, és akkor mondjuk, hogy egy SNP asszociál, ha a rá kapott p érték ennél az értéknél kisebb. Ennek teljesülésekor azt mondjuk, hogy az SNP **genomszintű asszociációt** mutatott. Ez pl., 1 millió SNP esetén 5×10^{-8} . Ennek eléréséhez, viszont a multifaktoriális betegségekre jellemző gyenge hatású variánsok miatt, sokszor 100 ezres nagyságrendű populációkat kell vizsgálni (ld. power analízis), amelyet egyes betegségeknél nagyon nehéz, vagy nem is lehet elérni. Ezen probléma elkerülésének az egyik módszere, hogy több kisebb populációt külön-külön vizsgálnak, pl. úgy, hogy az első teljes GWAS-ban kiválasztják a legjobban asszociáló x darab ($x = 25-120$) SNP-t, és az ismétlődő populációkban (replication cohorts) már csak ezeket nézik. Mivel az egyes populációkban kapott p értékek összeszoródnak, ráadásul a Bonferroni korrekciót is csak x-szel (mondjuk 100-zal) kell végezni, így megnő az esélye, hogy pozitív asszociációt találjanak. Ezekben a vizsgálatokban az első elemzés a kritikus, hiszen ilyenkor kell „elkapni” az asszociáló SNP-eket, és a nagy mennyiségű adat miatt itt merülnek fel azok a nehézségek amelyekről a 2. fejezetben szóltunk, azaz pl. a többszörös tesztelés problémája (Bonferroni korrekció), vagy az SNP-SNP kölcsönhatások óriási mennyisége. Ennek a vizsgálatnak az is az előnye, hogy így két vagy több független populáción is validáljuk az eredményeinket, így nagymértékben lecsökkenthetjük a hamis pozitív eredményeket. Hátránya, hogy a hamis negatív eredmények száma is nőhet. Az előny viszont túlkompenzálja a hátrányt, hiszen egy hamis pozitív eredmény sokszor több év felesleges kutatásait vonja maga után, annak minden költség, munka és idővonzatával együtt. Emiatt a jelentősebb újságok ma már megkövetelik, hogy az asszociációs vizsgálatok eredményeit legalább egy független populáción validálni kell. Egy másik lehetséges megoldás, az **útvonal analízis** (5). Itt a géneket aszerint, hogy az általuk kódolt fehérjék milyen anyagcsereútvonalban szerepelnek funkcionális kategóriákba sorolják. Ebben két fontos adatbázis áll rendelkezésre: a **GO**, azaz **Gene Ontology** (<http://www.geneontology.org/>), illetve a **KEGG**, azaz **Kyoto Encyclopedia of Genes and Genomes** (<http://www.genome.jp/kegg/>). Ezután, egy megfelelő szoftverrel (pl. ALIGATOR) azt nézik, hogy az útvonalon levő egyes géneknél genotipizált szignifikáns variánsok feldúsulnak-e az egyes útvonalon. Azaz azt vizsgálja a szoftver, hogy az egyes útvonalakon talált asszociált SNP-k eloszlása hogyan tér el a várt eloszlástól. Ha egy útvonalon sok olyan gén van, amelyben szignifikánsan asszociáló SNP-k vannak, akkor az az útvonal szerepet játszhat a betegségben, vagy egyéb jellegben. Fontos persze, hogy itt az egyes SNP-kre a szignifikancia határ nem a genomszintű szignifikancia érték, hanem annál nagyságrendekkel magasabb. A Bonferroni korrekciót itt általában a tanulmányozott útvonalak számával végzik. Hasonló a **Gene set enrichment analysis (GSEA)**, amelyet először a génexpressziós eredmények értékelésére fejlesztettek ki. Itt előre meghatározott „gén szettek” elemeznek, és azt nézik, hogy az egyes gén szettekben hány gén asszociál az adott fenotípussal, és eszerint rangsorolják az egyes gén szetteket. Ezeknél a vizsgálatoknál felmerül, hogy az olyan SNP-k, amelyek nem génben találhatóak, hová sorolhatók? Itt az egyes statisztikusok más-más stratégiát követnek. Az egyik ilyen, hogy a nem génben található SNP ahhoz a génhez tartozik, amelyik az SNP-től 20 kb-on belül található. Ha több ilyen gén is van, akkor mindegyikhez tartozik. Ha nincs ilyen gén, akkor az SNP kiesik ebből az elemzésből. Ezekkel a módszerekkel több, már a hagyományos módszerrel elemzett GWAS-t értékelték újra, és pl. magas vérnyomásban számos addig ismeretlen, a vérnyomás szabályozásában addig nem ismert anyagcsereútvonalat fedeztek fel.

3.2.4. *Parciális genomszűrések*

A teljes genomszűrések mellett ezek egyszerűbb változatai is gyakran kerülnek alkalmazásra. Ilyen pl. a kapcsoltsági analízisek után elvégzett parciális genomszűrés, vagy jelölt régió asszociációs vizsgálat. Itt a korábbi vizsgálatokban LOD csúcsokat adó markerek közelében nagyobb sűrűségű újabb markerekkel, általában SNP-kel végeznek asszociációs vizsgálatot (**partial genome association study, vagy PGAS**). Előnye, hogy jóval kevesebb adatot kell elemezni, így a statisztikai nehézségek jó része megoldódik.

Egy másik lehetőség, hogy a betegségben szerepet játszó szövetben, esetleg állat modellek segítségével elvégzett teljes mikroarray génexpressziós mérés után, a kiválasztott génekben végeznek SNP-asszociációs vizsgálatot.

3.2.5. *Személyre szabott genomika*

A szigorúan vett tudományos kutatások mellett, sokan, hamar felismerték, hogy a GWAS kapacitását, eredményeit profit orientált módon is értékesíteni lehet. Hamarosan megjelentek a személyre szabott genomika (**personal genomics**) jelszavát a zászlójukra tűző, az eredményeket közvetlenül a felhasználó számára rendelkezésre bocsátó cégek (angolul *direct to consumer*, vagy DTC szolgáltatás; 1. táblázat). Ennek az egyik legismertebb példája a 23andMe cég (6). A cégnél a feltételektől függően 99-399 \$ között lehet a szolgáltatást megrendelni. Ekkor elküldenek egy kisebb tartályt egy részletes útmutatóval. A tartályban a nyálunkat kell visszaküldeni, amelyből DNS-t vonnak ki, majd az Illumina egyik chipje segítségével kb. 1 millió SNP-t határoznak meg benne. Az eredményeket kielemezve 6-8 hét múlva kapjuk meg. Olyan információkat kaphatunk, mint pl.: a genomunk milyen etnikumú genomok keveredéséből származik; ha családunk több tagjának a mintáját beküldtük, meghatározzák, hogy a genetikai hátterük milyen arányban egyezik, különbözik (pl. testvérek unokatestvérek, távolabbi rokonok stb.). Értesítést kaphatunk, ha velünk rokonságban állókat találtak; különböző híres emberekkel hasonlítják össze a genomunkat; a jelenlegi tudományos ismeretek birtokában 198 betegségről becslést kapunk, hogy az átlag populációhoz képest nagyobb vagy kisebb kockázattal rendelkezünk; vagy a genomunk hogyan befolyásolja különböző, gyógyszerekre való reakciónkat, stb.

Amint az várható is volt, a közvetlenül a megrendelőnek, sőt „laikusoknak” eladott genetikai teszt eredmények számos jogi és etikai problémát vetettek fel. Ez kifejezetten igaz akkor, amikor a különböző betegségekre való hajlamról van szó. Ilyen problémák például, hogy joga van-e egy profitorientált magáncégnek ilyen típusú orvosi jellegű diagnózist kiadni, illetve, hogy mit kezd egy ember egy olyan információval, hogy bizonyos betegségre az átlagosnál egy kicsit emelkedett hajlama van? A váratlan, gyors technikai fejlődés, és a hozzá kapcsolódó szolgáltatás a törvényhozókat is váratlan helyzet elé állította. Erre jellemző, hogy például Kalifornia 2008. júniusában levelet küldött 13 ilyen cégnek, amelyben felszólította a cégeket, hogy állítsák le ezeknek a teszteknek az eladását kaliforniai lakosok számára, és bizonyítékokat kért arra vonatkozólag, hogy a cégek megfelelő szabályozással és engedélyekkel rendelkeznek (7). Az elmúlt években a cégek idomultak a meglévő szabályozásokhoz, és pl. Kaliforniában is engedélyezték működésüket, bár hozzá kell tenni, hogy a törvényhozás, illetve szabályozások egyelőre nem tudtak lépést tartani az új kihívásokkal. Pl. az USA-ban egyetlen törvény született ezekkel a tesztekkel kapcsolatban, amely azt mondja, hogy eredményeiket biztosító cégek, illetve munkaadók semmilyen módon nem vehetik figyelembe.

A cégek hangsúlyozzák, hogy az általuk adott információ nem helyettesíti a szakorvosi tanácsot, diagnózist vagy kezelést. Például a 23andMe olyan dokumentumot irat alá a

megrendelővel, amely tartalmazza, hogy a kapott információ nem szolgál betegség vagy más állapot diagnózisára, megakadályozása, vagy kezelésére, és a szolgáltatás kizárólag oktatási és kutatási célokat szolgál.

Sokan tartottak attól, hogy az olyan információk, hogy az illetőnek valamilyen súlyos, esetleg kezelhetetlen betegségre az átlagosnál nagyobb hajlama van, károsan befolyásolja az illető pszichés állapotát, például depressziós lesz. Érdekes módon, az ezzel kapcsolatos felmérések semmi ilyesmit nem mutattak ki. Valószínűleg ez kicsit hasonlít ahhoz, mint amikor a dohányos, vagy elhízott ember is tudja, hogy számos betegségre megnő a hajlama. Sőt, mivel feltehetőleg az átlagosnál egészségtudatosabb emberek kérnek ilyen információkat, inkább pozitív hatásokat tapasztaltak. Azaz, a kapott eredmények hatására egyes emberek tudatosan megelőző lépéseket tettek, vagy egészségesebb életformára tértek át.

3.2.6. Újgenerációs szekvenálás (NGS)

Részleteket ld. az 1. fejezetben.

Az NGS jelenleg még túl költséges több alkalmazáshoz és nehezen kezelhető, nagy mennyiségű adatot ad. Ezért sokszor ennek redukált formáit használják. Ilyen pl. az **exom szekvenálás**, amikor csak a fehérjét kódoló gének exonjait szekvenálják meg, ami kb. 30 Mb, a teljes genom 1%-a. Ld.: http://en.wikipedia.org/wiki/Exome_sequencing. Becslések szerint a monogénes betegségeket okozó mutációk 85%-a található itt. Hátránya persze, hogy más funkcionális szekvenciákról nem kapunk információt, és becslések szerint a multifaktoriális jellegekért (pl. betegség) felelős variációk 80%-a a protein-kódoló régiókon kívül esik.

3.2.7. Génexpresszió mérés

A harmadik nagy jelentőségű módszer, amely alkalmas a genom vizsgálatára, fontos szerepe van a gyógyszerkutatásban, és itt röviden ismertetünk, az a génexpresszió mérés. Mint tudjuk a különböző sejteink genomja megegyezik egymással. Amiben mégis különböznek az az expresszált, vagyis működő gének halmaza. A sejtek működéséről sok információt kaphatunk, ha ismerjük a bennük működő géneket, azaz melyik gén, milyen arányban íródik át, expresszálódik. Ez egy adott szövet esetén is eltérhet attól függően, hogy éppen milyen állapotban van, milyen hatások érik. Például egy asztmás tüdőben, vagy egy atherosclerotikus plakkban más a génexpressziós mintázat, mint egy egészséges szövetben. Ugyanígy a sejtekre ható külső hatások, pl. a gyógyszerek is, megváltoztatják a sejt génexpressziós mintázatát. A fejlődés itt is azt az utat járta be, mint a DNS variációknál. Először egyesével, elég bonyolult módon próbálták meg kvantifikálni a gének működését, majd a magyar származású Stephen Fodor révén az Affymetrix cég 1996-ban kezdte el forgalmazni génexpressziós chipjeit. Azóta más cégek, (pl. az Agilent, Illumina) is megjelentek a piacon, és a módszer egyre inkább kezdi levetkőzni gyermekbetegségeit. Kezdetben ugyanis sok problémát jelentett a nagy mérési pontatlanság, nehezen reprodukálható és értékelhető eredmények, valamint a módszerek magas ára. Jelenleg az árak a kiindulásiakhoz képest drasztikusan lecsökkentek, a folyamatos fejlesztésekkel a mérések pontossága, reprodukálhatósága is sokat javult, és az adatok értékelése is rengeteget fejlődött az elmúlt években. Egy Agilent chip 44 ezer transzkriptumot tud mérni, sőt egyes termékeken (pl. 4x44K chip) párhuzamosan több különböző mintát is lehet vizsgálni. Szerre a világon, így Magyarországon is, több szolgáltató labor működik ahol viszonylag olcsón, teljes génexpressziós mintázatot lehet, ma már elég pontosan mérni, azaz nem is kell feltétlenül megvenni a drága leolvasót. A bioinformatika fejlődésével pedig a nem matematikus vénával rendelkező kutatók is értékes eredményeket

tudnak kihozni a több 10 ezer, elsőre igen kaotikusnak tűnő számadatból, amely egyetlen méréshez tartozik.

A nagyáteresztő képességű módszerek közé az utóbbi időben felzárkózott az NGS technikát felhasználó RNS szekvenálás (<http://en.wikipedia.org/wiki/RNA-Seq>) is. A módszer előnye, hogy a transzkriptum mennyiségének mérése mellett olyan információkat is kaphatunk, mint pl., hogyan expresszálódnak a gén különböző alléljai, detektálhatóak a poszt-transzkripció mutációk, vagy a génfúziók.

3.2.8. Egyéb mikroarray alapú módszerek

A technika, illetve a tudásunk fejlődésével számos egyéb termék nőtt ki a génexpressziós chip módszeréből. A miRNS-ek felfedezése után rövidesen megjelentek a piacon a miRNS chipek, a komparatív genom hibridizációval (CGH), pedig pl. CNV-ket lehet mérni, vagy vizsgálni lehet a tumorokban keletkező nagyobb genomikai átrendeződéseket. A ChIP-on-chip array termékkel monitorozni lehet, hogy a génexpresszió szabályozásában résztvevő fehérjék hová kötődnek. A metilációs array-vel, pedig a genom metilációs mintázatát lehet tanulmányozni; a transzkriptom térképezés (*mapping*) technikával, pedig lokalizálni lehet az expresszáló géneket a genomban. Ez utóbbi módszerek a „*tiling array*” módszer családba tartoznak, de ugyanazzal a leolvasóval lehet őket értékelni, és hasonló elven működnek, mint a hagyományos génexpressziós chipek.

3.3. Állatmodellek

3.3.1. Állatmodellek előnyei

A multifaktoriális betegségek genomikai hátterének, patomechanizmusának megismerésében nagyon fontos szerepet töltenek be az állatkísérletek, állatmodellek. Vannak olyan betegségek (pl. asztma, atherosclerosis), amelyek patomechanizmusában a molekuláris és sejt-szintű tudásunk nagy részét állatkísérletek révén nyertük.

Nézzük meg, milyen előnyei vannak ezekben a vizsgálatokban az állatok használatának:

- A kísérleti állatok általában szabadon keresztezhetőek. Az emberrel szemben, itt lehetőségünk nyílik több generáción keresztül a genetikai háttér, és a fenotípus közötti összefüggéseket vizsgálni, ezzel kapcsolatban sokkal könnyebb QTL vizsgálatokat elvégezni. Az egérnek pl. 2 hónap a generációs ideje, szemben az ember 20-30 évével. Irányított keresztezésekkel nyomon lehet pl. követni, hogy egy genetikai marker milyen QT-val, fenotípussal asszociál, szegregál stb.
- Különböző egértörzsek állnak rendelkezésre, amelyek fenotípusban (pl. betegségre való hajlamban) különböznek egymástól. Ezeket egyrészt fel lehet használni betegségmodellként, vagy keresztezésekkel vizsgálhatjuk a fenotípus és a genotípus szegregációjának összefüggését. Ezeket az állatokat szigorúan kontrollált körülmények között tarthatjuk, szemben az emberrel, ahol a vizsgálatban résztvevők előéletét, környezetét, táplálkozását stb. nem lehet pontosan ismerni, kontrollálni.
- Gén-szinten az ember nem különbözik jelentősen a többi élőlénytől. A legfontosabb alap-gének gyakorlatilag minden élőlényben megtalálhatók. Az emlősökhöz különösen hasonlít az ember gén-szinten, pl., az egér genom csak 300 génben különbözik az emberétől, de sok vizsgálatban alacsonyabb rendű állatokat is eredményesen használhatunk, pl. gyümölcslegyet (*Drosophila melanogaster*), vagy a *Caenorhabditis elegans* nevű fonalférget elterjedten használják genetikai vizsgálatokban. Érdekesség, hogy ez a kis állat „kétszeres Nobel-díjas”, hiszen mind a

szervfejlődés és a programozott sejthalál genetikai szabályozásával, mind az RNS interferencia leírásával kapcsolatban tanulmányozóit, Nobel-díjjal tüntették ki.

- Embereken nyilvánvaló etikai okokból csak nagyon korlátozottan lehet kísérletezni. Állatokkal, ez sokkal szélesebb körben megtehető.
- Az előzőek folytatásaként az állatokból sokkal könnyebb szövetekhez jutni (pl. tüdő, agy stb.), és pl. génextpressziós vizsgálatokat végezni, könnyebb, pontosabb a betegségek diagnózisa.
- Rengeteg állat-specifikus reagens, anyag áll már rendelkezésünkre, pl. egérspecifikus ellenanyagok, genetikai próbák. Ismert már több állat géntérképe.
- Legtöbb betegségünkre kidolgozható állatmodell, amely segítségével részletesen vizsgálható a betegségek molekuláris háttere (ld. az egyes betegségeknél).
- Genetikailag módosított állatok lehetősége.

Ez utóbbi két pontot, témánk miatt kicsit részletesebben tárgyaljuk.

Témánk szempontjából két típusú genetikailag módosított állatot kell megemlíteni. Az egyik a génkiütött, **knockout, vagy KO állat**. Közülük is magasan kiemelkedik témánk szempontjából a KO egér, melynek kialakításáért 2007-ben Nobel díjat is adtak. Röviden, itt az egér egy génjét (vagy esetleg a genom egy másik funkcionális egységét) valamilyen módszerrel inaktívvá tesznek. Részletesebben: http://en.wikipedia.org/wiki/Knockout_mouse. 2006-ban 9 országban 33 kutatóközpont megalakította az **International Knockout Mouse Consortium-ot** (IKMC), majd 2011 júniusában az **International Mouse Phenotyping Consortium-ot** (IMPC), amelyek azt a célt tűzték ki maguk elé, hogy az összes génre előállítsanak KO egeret, majd fenotipizálják őket.

KO állatok segítségével vizsgálni lehet, hogy egy gén hiányának milyen fenotípusos jegyei vannak, így a gén funkciójáról jóval többet megtudhatunk, mintha csak *in vitro* körülmények között tanulmányoznánk. Az ismert gének többségéről már kapható KO egér, illetve meg is rendelhetünk cégektől ilyen egereket. Így derült ki pl., hogy az **APOA5** gén hiánya magas szérum triglicerid szintet okoz. Egy másik lehetőség, hogy gén-KO segítségével, elő lehet állítani az egéren valamilyen betegséget. Pl., ismert, hogy a „vad” egéren nem alakul ki atherosclerosis. Azonban, ha az LDL receptor (**LDLR**), vagy az **APOE** génjüket kiütjük, és megfelelő, ún. Western-típusú diétával etetjük őket, az emberéhez igen hasonló atherosclerotikus léziók alakulnak ki az ereik falában.

A másik típusú genetikailag módosított állat, ahol főleg egeret használnak, a **transzgenikus egér**. Itt szemben az előzőekkel, egyes gének működését felerősítik, túl-expresszáltatják.

Ennek végső céljai ugyanazok, mint a KO egérnél elmondottak.

Mindkét technikát akár szövet-specifikusan is alkalmazhatjuk. Pl. transzgenikus esetben a gént egy olyan gén promótere után illesztjük be, ami csak az egyik szövetben expresszálódik. Pl., a **SCGB1A1**, korábban CC16 gén főleg a tüdőben expresszálódik, ott viszont nagyon erősen. Amikor egérbe az **IL5** génjét rakták mögé, az IL5 a tüdőben mutatott igen erős expressziót, és az állatban tüdő eozinofília, és asztma alakult ki. Ezzel bizonyították, hogy az IL-5 citokin az eozinofilek működésében fontos szerepet játszik, és a tüdő eozinofília közvetlenül asztmát okozhat.

Sokáig problémát okozott a magzati életben életfontosságú gének tanulmányozása, hiszen azok hiánya, vagy esetleg túlexpressziója magzati korban letális. Ezért fejlesztették ki a Cre-lox rekombináció segítségével (ld. http://en.wikipedia.org/wiki/Cre-Lox_recombination) az olyan KO állatokat, ahol, pl. valamilyen indukció (pl. antibiotikum-evés) segítségével időzítetten lehet *in vivo* kivágni a kérdéses gént, vagy genom-szakaszt. Ezek az ún.

kondicionális KO állatok.

Más, *in vivo* technikákkal is lehet még géneket időlegesen inaktíválni. Ezek közül a legjelentősebb az RNS interferencia jelenségét felhasználó technikák. Pl. a fonalféreg

Caenorhabditis elegans-ban minden génnek tanulmányozták az RNSi-vel inaktivált hatását. De a módszer pl. egérben is működik.

3.3.2. *Állatmodellek hátrányai*

Ha gén-szinten kicsi is a különbség pl. ember és egér között, genom-szinten nagyobb. A vizsgálatok során kiderült, hogy egyes folyamatok egerekben teljesen máshogy működhetnek, mint emberben. Ebből az következik, hogy az egéren kapott eredmények csak kiindulópontok lehetnek, ha emberre akarunk következtetni. Ezután, még az összefüggéseket emberben is meg kell erősíteni, ami persze nem mindig könnyű.

Hasonló a helyzet a betegségeknél is. Sok betegség nem modellezhető tökéletesen egérben, mások a vezető patomechanizmusok, sokszor pont a legfontosabb tünetek különböznek, vagy egyes gének hiányai emberben igen, egérben nem okoznak betegséget, vagy fordítva. Egyes, a betegségben az emberben jelentős gének az egérben más sejtekben, vagy szövetekben expresszálódnak, más a funkciójuk. Például, az inzulin rezisztenciát fokozó hatású, így a 2-es típusú cukorbetegségben fontos resistin gén egérben főleg a zsírsejtekben, emberben a makrofágokban expresszálódik.

3.3.3. *Kísérleti betegsémodellek*

Mint ahogy majd az egyes betegségek tárgyalásánál látni fogjuk, a betegségben fontos gének manipulálásával az emberéhez hasonló betegségeket lehet egereken előidézni. Ezt lehet közvetlenül az ismert gének kiütésével, vagy túlexpressziójával, de lehet keresztezésekkel is elérni, amikor több nemzedéken keresztül mindig a betegség tüneteit leginkább hordozó állatokat szaporítunk tovább. Ez utóbbi esetben genomikai szempontból a cél, hogy megállapítsuk a betegséget okozó variációkat. Ráadásul ilyenkor az emberi multifaktoriális kórképekhez sokkal hasonlóbb kórformákat kapunk, hiszen itt, szemben az egy gén módosításával okozott betegséggel, általában a többféle egértörzsben jelenlevő, amúgy gyenge hatású variációk keveredhetnek össze, halmozódhatnak fel. Ezt a folyamatot mesterségesen is fel lehet gyorsítani. Ez történik pl. a „**The Jackson Laboratory**”-ban, ahol a hím egerekbe N-ethyl-N-nitrosourea (**ENU**)-t juttatnak, amely 1000x-sére növeli a spermatogenezisben a mutációs rátát. Ezután megfelelő keresztezésekkel továbbszaporítják az egereket. Nagyon részletes fenotipizálás segítségével különböző betegség-tüneteket hordozó egereket katalogizálnak, amelyekre rá lehet keresni, és meg lehet vásárolni őket. Az egész a „**Mouse phenome project**”-ből nőtt ki, és az adatokat a **Mouse Phenome Database**-ban tárolják. Részletek: <http://phenome.jax.org/>.

3.3.4. *QTL analízis egérben*

A módszer alapja, hogy a beltenyésztett egértörzsek közötti mérhető fenotípusos jellegekben különbségek vannak. A QTL analízis során keresztezésekkel és visszakereszte�ésekkel ezeket nyomon követik, majd a különbséget okozó genetikai hátteret tisztázzák.

Két típusú egértörzs bizonyult igen hasznosnak ezekben a vizsgálatokban. Az egyik a **kongenikus törzs (congenic strain)**: olyan (általában egér)törzs amely a fenotípusra „nem hajlamos” törzstől csak a hajlamosító genomikai régióban különbözik. Mindig a QT-t hordozóállatokat kell 5-10 generáción keresztül visszakeresztezeni a QT-t nem-hordozó szülői egértörzsszel (5. ábra).

Jó példa erre egy atherosclerosis genomikai hátterében elvégzett QTL analízis (6. ábra). Korábban ismert volt, hogy a C57BL/6J (B6)-nak nevezett egértörzsben az LDLR gén kiütése segítségével az emberi atherosclerosis-hoz hasonló betegséget, léziókat lehetett előidézni.

Ugyanezzel szemben a CAST/Ei (Cast) egértörzs védett volt. Ezzel a vizsgálattal azt akarták megállapítani, hogy mi a genomikai oka ennek a védettségnek. A 6. ábrán látható módon keresztezték a két törzset, és az F2 generációban teljes genomszűrést végeztek. Itt, a LOD score görbe segítségével megállapították, hogy a 6-os kromoszómán van néhány mikroszatellita marker, amely az aorta lézió nagyságával és inzulin rezisztenciával asszociált (a LOD score csúcsot adó markerek). Ilyenkor a genomszűrést úgy végzik, hogy olyan markereket használnak, amelyekkel egyértelműen különbséget lehet tenni a két egértörzs között (azaz a mikroszatellita ismétlődés-számban különbözik ezekben a markerekben a két törzs egymástól). Azt nézik, hogy melyik marker mutat asszociációt a kérdéses fenotípussal az F2 generációban. Itt azt is kihasználják, hogy az F2 generációban a rekombináció miatt már olyan kromoszómák találhatók, amelyekben a két nagyszülő genetikai anyaga keveredett egymással. Itt akár 100-nál több egeret is használhatnak, hogy egyértelmű LOD score csúcsot kapjanak. Ezután az F2 generációból kiválasztották azokat az egereket, melyek hordozták a csúcsot adó markereket, és a kérdéses tüneteket is, és visszakeresztezték, ebben az esetben a B6 egérrel. Ezek utódai közül azokat keresztezték vissza megint csak a B6 egérrel, amelyek hordozták a genomikai markereket. Kongenikus egértörzs előállításánál ezt kell 5-10-szer megismételni. A végén olyan egeret kaptak, amelyik genomja kb. 99%-ban a B6 egértől, és 1%-ban a Cast egértől származott, viszont ez az 1% az, amely felelős az atherosclerosis-sal szembeni védelemért. Ezt az utolsó generációnál fenotipizálással megerősítették, majd valamilyen módszerrel (általában irodalmi, vagy adatbázis kutatás *in silico*) kiválasztották a leszűkített genom régióból a potenciális, a tüneteket okozó gént, itt az 5-lipoxigenázt (5-LO). Ezt itt megszekvenálták és találtak benne egy funkció-vesztést (*lost of function*) okozó mutációt. Ezután előállítottak egy 5-LO KO egeret és megállapították, hogy ez is védett atherosclerosis-sal szemben. Ezzel bizonyították, hogy az 5-LO-nak fontos szerepe van az atherosclerosis patomechanizmusában. Később emberben a homológ **ALOX5** génben SNP-eket kerestek, és asszociációs vizsgálattal igazolták, hogy vannak olyan variációk, amelyek emberben befolyásolják a betegségre való hajlamot (8,9).

Pontokba szedve az előzőekben részletezett QTL analízist:

- F2 generációban LOD score analízissel megállapítani a QTL-t.
- Kongenikus állat létrehozása, mindig a QTL-t hordozó állatot kell visszakereszteni.
- A QTL-t további markerek segítségével finomítani. Elérhető felbontás kb. 1 cM.
- Pozicionális jelölt gének keresése.
- Humán ortológ keresése.
- Asszociációs vizsgálatok emberben.

Az előzőekhez hozzá kell tenni, hogy az esetek többségében megállnak az F2 generációban elvégzett genomszűrésnél, és megadják a LOD score csúcsokat mutató genom-régiókat.

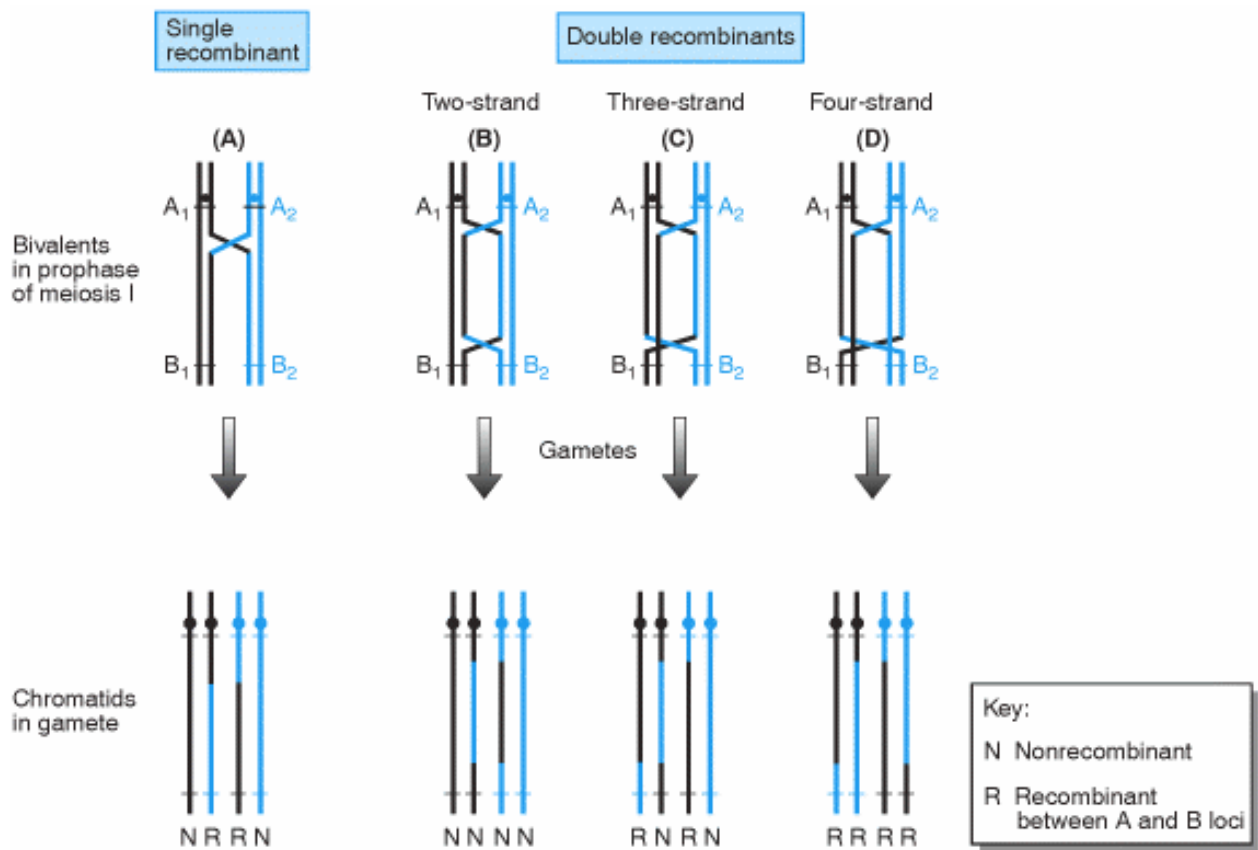
A másik típusú egértörzs a **recombinant inbred** vagy RI egértörzs. Itt, 2 szülői egértörzset kereszteznek a 7. ábrán látható módon 10-20 generáción keresztül (10). Itt testvérkeresztezésekkel stabilizálnak genotípusokat és az ezekhez tartozó fenotípusokat egyszerre több törzsben.

Ezek a módszerek egy nagyságrenddel érzékenyebbek, mint a humán genomszűrés.

Cég	Szolgáltatás	Ár
23andMe	Kb. 1 millió SNP meghatározásából 178 betegség genetikai kockázatáról egy becslés, illetve származás és rokonság elemzés	399 \$, vagy 99 \$ + 9\$/ hónap előfizetés az újabb eredményekre
deCODEme	Illumina Human 1M BeadChip, mely >1 millió SNP-t detektál, 47 betegség, származás	2.000 \$
Knome	Illumina genom szekvenáló platform; az eredmények értelmezése	39.500 \$
Existence Genetics	Saját fejlesztésű chip; 700 betegség és jelleg; javaslatok a betegségek megelőzésére	350 \$
Navigenics	Affymetrix 6.0; 900.000 SNP; szigorú tudományos követelmények, 28 betegség, 12 gyógyszerre való reagálás	999 \$

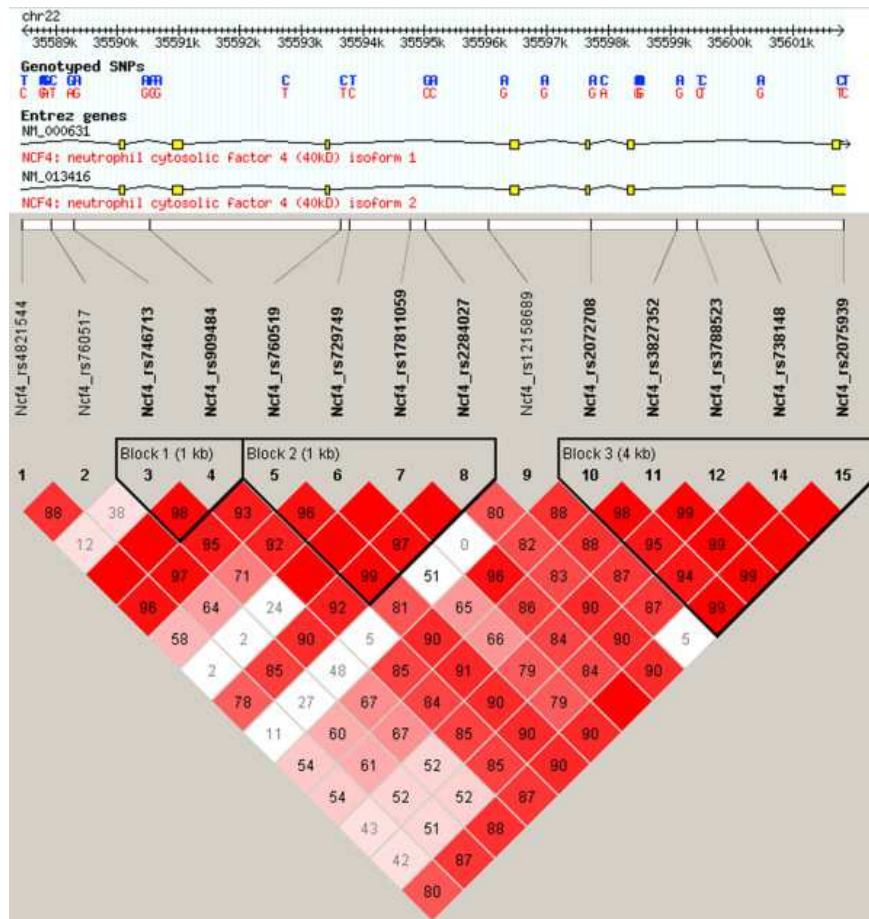
1. táblázat

Jelentősebb cégek, melyek orvosi és más célokból genomikai tesztek elvégzését, és az eredmények értékelését kínálják.



1. ábra

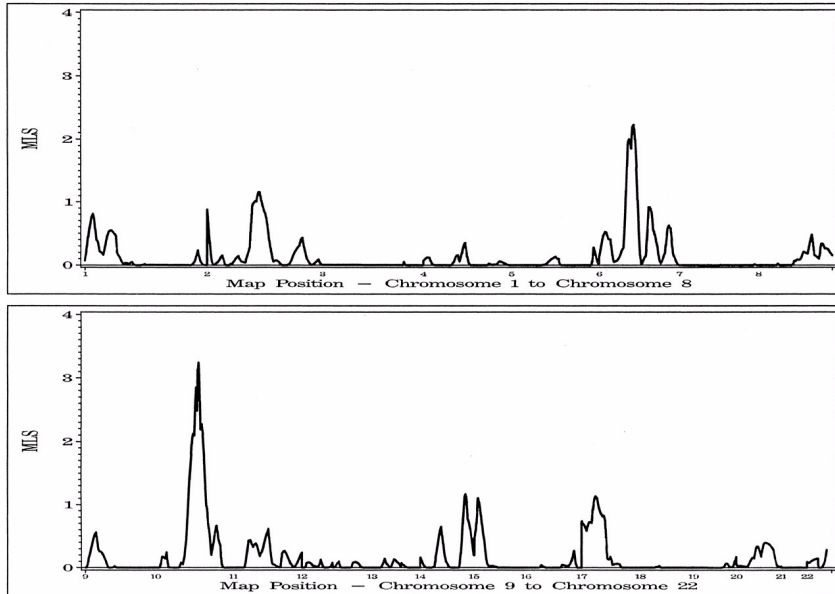
Rekombináció, vagy crossing over meióziskor. Például az emberi férfi meióziskor sejtenként átlag 49 rekombináció történik. A folyamatban a homológ, két szülői kromoszóma genetikai anyaga kicserélődik. Eredményeképpen, az eredetileg egymás mellett található allélok (pl. A_1 és B_1) elkerülhetnek egymás mellől.



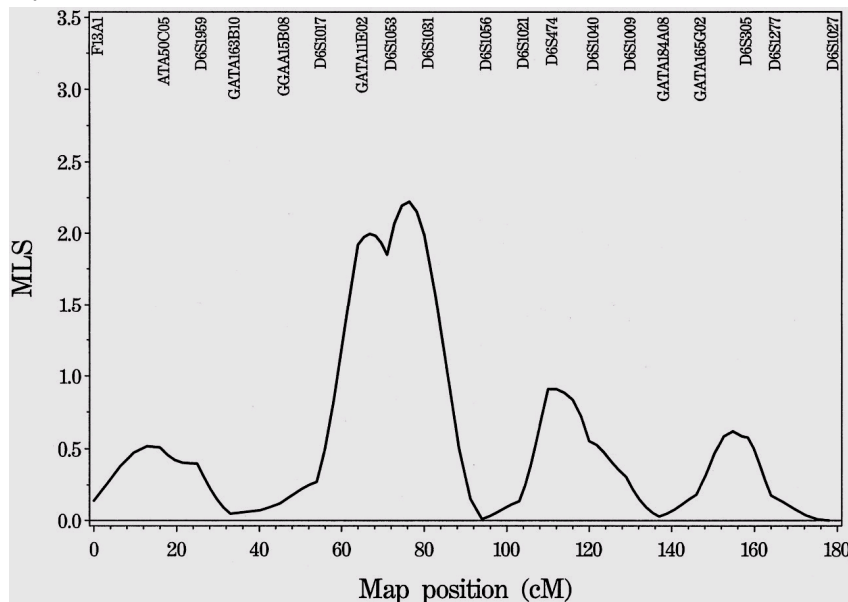
2. ábra

Az LD és haplotípus blokkok legelterjedtebb ábrázolása. A háromszög feletti számok egy-egy allélt jelölnek, itt 15-öt. Minden allélhoz két irány tartozik, amelyet az első négyzet felfelé mutató két oldala jelképez. Az egyes négyzetekbe írt számok LD koefficiens jelentenek, amelyek a négyzet két felső oldala irányában az egyes allélokra vonatkoznak. Például a 11-es és a 8-as allél között az LD koefficiens értéke 83, ami 0,83-at jelent. A jobb vizualizáció miatt a négyzetek színezve vannak. Minél sötétebb piros egy négyzet, annál nagyobb az LD érték a két allél között. A fehér négyzetek azt jelentik, hogy a két allél között nincs kapcsoltság, alacsony az LD koefficiens. Bizonyos allélok, a négyzetekkel együtt, egy ötszögbe vannak rajzolva. Ezek között nagy az LD, és haplotípus blokkokat alkotnak. Ezen az ábrán 3 haplotípus blokkot láthatunk.

A.



B.



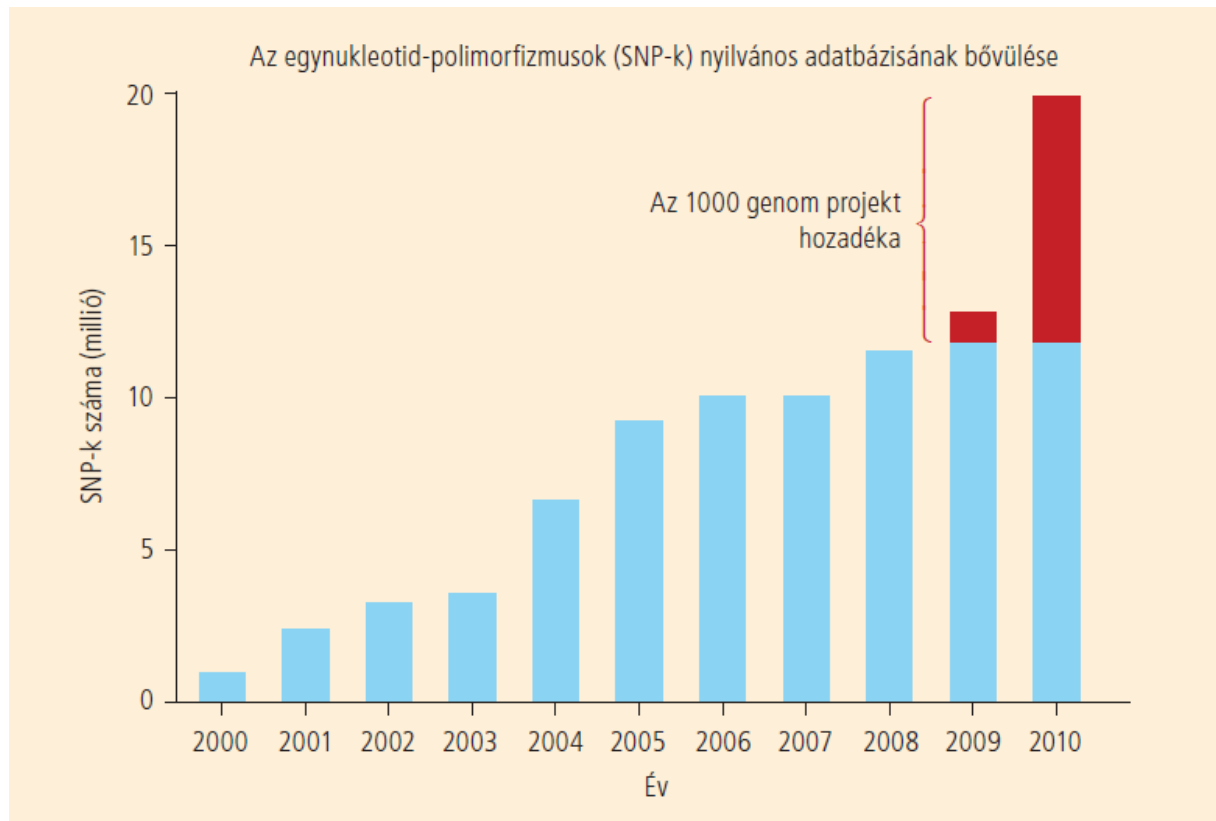
3. ábra

A.

LOD score elemzés eredményének ábrázolása. Az X tengelyen az egyes kromoszómák számai láthatók, amelyeken több markerre is elvégezték a LOD score számolást. Az Y tengelyen az MLS = maximum Lod score. A genomszintű szignifikancia értéket általában $MLS = 3$ -nál szokták meghúzni. Itt a 10-es kromoszómán láthatunk ilyen csúcsot. A csúcs alatti területen lehetségesek olyan szekvencia variációk, allélok, amely a vizsgált fenotípussal asszociálnak, kapcsoltak vele.

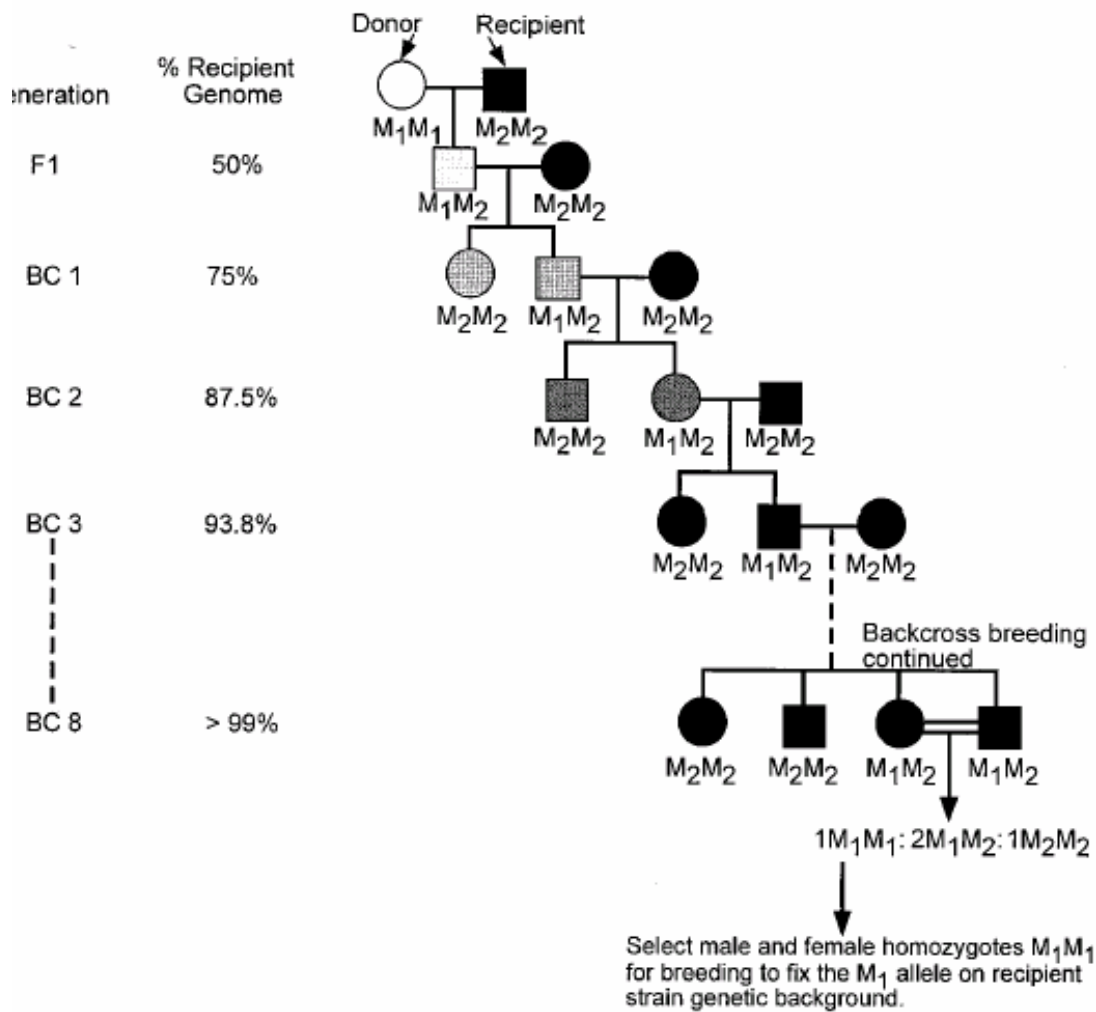
B.

Az egyik LOD score csúcs (a 6-os kromoszómán) felnagyítva. Felül láthatók az elemzett markerek (mikroszatelliták, vagy STR-ek) elnevezései. Pl., D6S1040, ahol a D utáni szám általában a kromoszóma számára utal. Az ábrán kb. a 60 és a 90 cM között genom terület esik a csúcs alá.



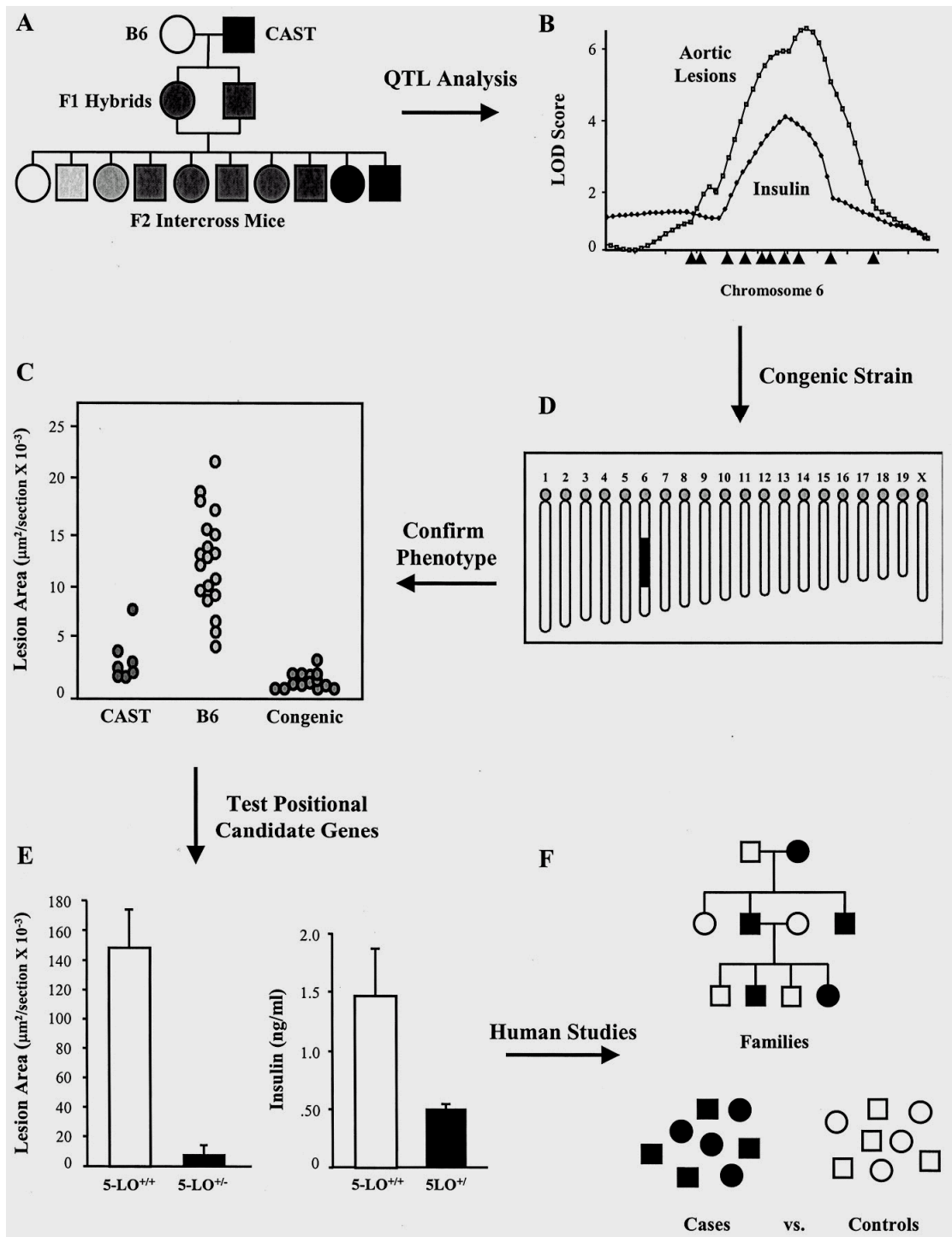
4. ábra

A nyilvános SNP adatbázis növekedése. 2009 és 2010-ben az 1000 genom projekt 8,5 millió új SNP-t talált a megszekvenált, különböző etnikumú genomokban (4).



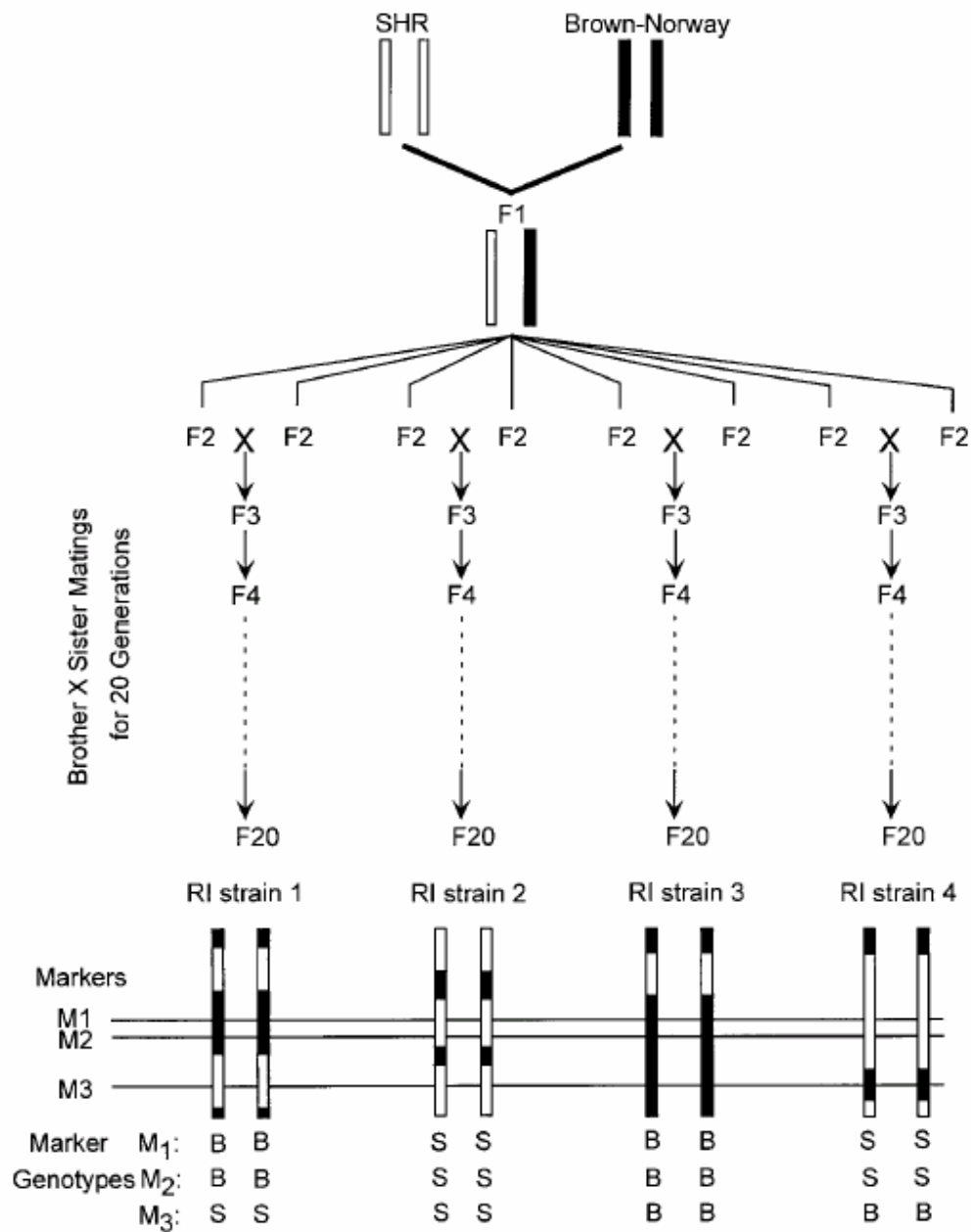
5. ábra

Kongenikus állatok előállítására. Mindig a QT-t (itt M_1 markerral jelölt) hordozóállatokat kell 5-10 generáción keresztül visszakereszteni a QT-t nem-hordozó szülői egértörzzsel. Az utolsó generációban a recipiens állat genomjába lesz benne donor állat genomjának adott része.



6. ábra

QTL vizsgálat atherosclerosisban egértől az emberig (8, 9). Magyarázatot ld. a szövegben.



7. ábra

Recombinant inbred (RI) (rekombináns beltenyésztett) állatok előállítása. Az ábrán az SHR rövidítés „spontaneously hypertensive rat”et jelent. Azaz az esszenciális, vagy állatoknál spontán magas vérnyomás genetikai hátterének vizsgálata céljából végezték el a két patkánytörzs keresztezését (10).

3.4. Irodalom

1. International HapMap 3 Consortium, Altshuler DM, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-8.
2. International HapMap Consortium, Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851-61.
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;7;447:661-78.
4. Pennisi E. 1000 Genomes Project Gives New Map Of Genetic Diversity. *Science* 2010; 330: 574-5.)
5. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet*. 2007 Dec;81(6):1278-83.
6. <https://www.23andMe.com/>
7. Kaye J. The regulation of direct-to-consumer genetic tests. *Hum Mol Genet*. 2008;17:180-3.
8. Allayee H, Ghazalpour A, Lusk AJ. Using mice to dissect genetic factors in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2003 Sep 1;23(9):1501-9.
9. Mehrabian M, et al. Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ Res*. 2002 Jul 26;91(2):120-6.
10. Rapp JP. Genetic analysis of inherited hypertension in the rat. *Physiol Rev*. 2000 Jan;80(1):135-72.

3.5. Fejezethez tartozó kérdések

1. Milyen módszereket lehet használni a multifaktoriális betegségek genomikai hátterének tisztázására?
2. Mi az a Hardy Weinberg eloszlás?
3. Mi lehet az oka a Hardy Weinberg egyensúlytól való eltérésnek?
4. Mi az a haplotípus?
5. Mi az a linkage disequilibrium?
6. Mivel foglalkozik a HapMap project?
7. Mit használunk az LD mérésére, milyen értékei lehetnek és azok mit jelentenek?
8. Genetikai vizsgálatokban mik között lehet kapcsoltság?
9. Mi a 0 hipotézis kapcsoltsági analízisnél?
10. Mi az a LOD érték?
11. Mi a különbség a parametrikus és a nemparametrikus LOD score analízis között?
12. Melyik a legelterjedtebb módszer a nemparametrikus LOD score vizsgálatnál?
13. Mi az a pozicionális klónozás?
14. Mit jelent a „founder populáció” és mi az előnye kapcsoltsági vizsgálatoknál?
15. Mit jelenthet a jelölt gén polimorfizmusának vizsgálatakor az asszociáció?
16. Mik lehetnek a pozitív asszociáció okai SNP vizsgálatoknál?
17. Mi az a „population stratification”?
18. Mi az a population admixture?
19. Milyen módszert lehet használni, hogy asszociációs vizsgálatoknál a kontroll csoport összetétele ne befolyásolja nagyon az asszociációs vizsgálat eredményét?
20. Mi az a prospective és a retrospective vizsgálat? Mondjon példát!
21. Mi az a WTCCC?
22. Kockázatszámolásnál milyen értékeket használhatunk?
23. Mi genetikai vizsgálatokban a power analízis?
24. Mondjon példát a hipotézis által irányított és a hipotézismentes genomikai módszerekre!
25. Multifaktoriális betegségek vizsgálatakor hol használhatják a mikroszatellitákat?
26. Mi a hátránya a kapcsoltsági vizsgálatoknak?
27. Multifaktoriális betegségek vizsgálatakor hol használhatják az SNP-eket?
28. Mi az a jelölt gén asszociációs vizsgálat, és mi az előnye és a hátránya?
29. Milyen markereket használnak a kapcsoltsági vizsgálatoknál?
30. Mit jellemez egy marker esetén a heterozigótaság mértéke?
31. Minek a rövidítése a GWAS?
32. Mi az a teljes genom asszociációs vizsgálat?
33. Mi a hátránya a teljes genom asszociációs vizsgálatoknak?
34. Mi az a genomszintű asszociáció?
35. Mi az az útvonal analízis?
36. Mi az a parciális genomszűrés?
37. Milyen céget ismer, amelyik nagyteljesítményű genotipizáló berendezést forgalmaz, és kb. mekkora ezek teljesítménye?
38. Mit jelent személyreszabott genomika?
39. Mi az az exom szekvenálás?
40. Mi lehet az előnye a microarray génexpressziós méréseknek?
41. Milyen előnyei lehetnek betegségek tanulmányozásánál állatmodellek alkalmazásának?
42. Milyen hátrányai vannak humán betegségek tanulmányozása esetén az állatmodelleknek?
43. Milyen típusú genetikailag módosított állatmodelleket ismer?

44. Mire lehet használni KO állatokat?
45. Mi az a transzgenikus egér?
46. Mi aza kondicionális KO állat?
47. Mi az a kongenikus egér?
48. Hogyan lehet betegségekre genetikailag hajlamos egértörzseket előállítani?
49. Mi az a QTL analízis, hogyan végezhető el?