

**SEMMELWEIS EGYETEM  
DOKTORI ISKOLA**

**Ph.D. értekezések**

**2044.**

**PULAY ATTILA JÓZSEF**

**Pszichiátria**

című program

Programvezető: Dr. Tringer László, egyetemi tanár

Témavezető: Dr. Réthelyi János, egyetemi docens

# Multilókusz asszociációs elemzések a szuicid magatartás és szkizofrénia genetikai vizsgálatában

Doktori értekezés

**Dr. Pulay Attila József**

Semmelweis Egyetem

Mentális Egészségtudományok Doktori Iskola



**Témavezető:** Dr. Réthelyi János, Ph.D., egyetemi docens

**Hivatalos bírálók:** Dr. Antal Péter, Ph.D., egyetemi docens  
Dr. Juhász Gabriella, Ph.D., egyetemi docens

**Szigorlati bizottság elnöke:**  
Dr. Bereczki Dániel, Ph.D., egyetemi tanár

**Szigorlati bizottság tagjai:**  
Dr. Barta Csaba, Ph.D., egyetemi adjunktus  
Dr. Szekeres György, Ph.D.

Budapest  
2016

**Tartalomjegyzék**

<b>Rövidítések jegyzéke.....</b>	<b>3</b>
<b>Táblázatok jegyzéke .....</b>	<b>4</b>
<b>1. Bevezető .....</b>	<b>6</b>
1.1. A pszichiátriai genetika általános kérdései.....	6
1.2. Multilokusz vizsgálati módszerek a pszichiátriai genetikában.....	8
1.2.1. Haplotípus alapú elemzések .....	11
1.2.2. Régió-alapú asszociációs vizsgálatok .....	17
1.2.3. Génszett- és génhálózat-alapú vizsgálatok .....	22
1.2.4. A genomikus heritabilitást és poligénes additív hatást vizsgáló módszerek	28
1.3. A szkizofrénia és a szuicid magatartás teljes-genom vizsgálatai .....	34
1.3.1. Teljes-genom asszociációs és multilokusz vizsgálatok szkizofréniaiban .....	34
1.3.2. Teljes-genom asszociációs vizsgálatok szuicid magatartásban.....	36
<b>2. Célkitűzések.....</b>	<b>44</b>
2.1. Az első vizsgálat célkitűzései és hipotézisei .....	44
2.2. A második vizsgálat célkitűzése és hipotézisei .....	45
<b>3. Módszerek .....</b>	<b>47</b>
3.1. Az első vizsgálat módszerei.....	47
3.1.1. Minta, fenotípus .....	47
3.1.2. Genotípezés, imputáció .....	49
3.1.3. Génszelekció, statisztikai elemzés .....	52
3.2. A második vizsgálat módszerei .....	54
3.2.1. Minta, fenotípus .....	54
3.2.2. Genotípezés, exom szekvenálás, imputáció .....	55
3.2.3. Statisztikai elemzés, annotáció, variáns- és génprioritizáció.....	57
<b>4. Eredmények.....</b>	<b>61</b>
4.1. Az első vizsgálat eredményei .....	61
4.1.1. Régió-alapú statisztikai erő.....	61
4.1.2. Gén-alapú asszociációk és a gén-régiókhoz tartozó genomikus heritabilitás	63
4.1.3. Poligénes rizikópont predikció (PRS) .....	77
4.2. A második vizsgálat eredményei.....	82
4.2.1. Gén-alapú asszociációk .....	82
4.2.2. Génszett elemzések: kanonikus útvonal és pozicionális génszett tesztek ....	82

4.2.3. Funkcionális annotációs klaszterelemzés .....	86
4.2.4. Replikációs valószínűség elemzés .....	88
<b>5. Megbeszélés .....</b>	<b>89</b>
5.1. Az eredmények áttekintése .....	89
5.2. Multilokusz asszociációk szuicid magatartásban .....	90
5.3. Multilokusz asszociációk szkizofréniában .....	98
<b>6. Következtetések.....</b>	<b>103</b>
<b>7. Összefoglalás .....</b>	<b>104</b>
<b>8. Summary .....</b>	<b>105</b>
<b>9. Irodalomjegyzék .....</b>	<b>106</b>
<b>10. Saját közlemények .....</b>	<b>138</b>
<b>11. Köszönetnyilvánítás .....</b>	<b>141</b>

**Rövidítések jegyzéke**

BIP:	bipoláris zavar
CNV:	copy number variation, kópiaszám variáns
dbGaP:	database of Genotypes and Phenotypes
DHS:	DNaseI hypersensitivity site, DNaseI hiperszenzitív lókus
DLPFC:	dorzolaterális prefrontális kéreg
DSM-5:	Diagnostic and Statistical Manual of Mental Disorders 5th Edition
ES:	enrichment score, feldúsulási pont
GAIN:	Genetic Association Information Network
GATES:	kiterjesztett Simes'-teszt, régió-alapú asszociációs teszt
GO:	génontológiai annotáció
GPRS:	genomic profile risk scoring, genomikus profil rizikópont
GWAS:	genomewide association study, teljes-genom asszociációs vizsgálat
$h^2$ :	genomikus heritabilitás
$H^2$ :	teljes additív genetikai heritabilitás
HWE:	Hardy-Weinberg egyensúly
HYST:	hibrid gén-szett asszociációs teszt
IBD:	identity-by-descent, leszármazási azonosság
indel:	insertion/deletion variant, inzerció/deleció variáns
LD:	linkage disequilibrium, kapcsoltsági egyensúlytalanság
MAF	minor allél frekvencia
MDD:	major depressive disorder, major depresszív zavar
miRNS:	mikroRNS
NCBI:	National Center for Biotechnology Information
OR:	odds ratio, esélyhányad
PGC:	Psychiatric Genomics Consortium
PRS:	poligénes rizikópont
SCZ:	szkizofrénia
SNP:	single nucleotide polymorphism, egyedi nukleotid polimorfizmus
SNV:	single nucleotide variant, egyedi nukleotid variáns

## Táblázatok jegyzéke

<b>1. táblázat.</b> A szuicid magatartás kandidáns génjei a teljes-genom asszociációs és expressziós vizsgálatok alapján. ....	39
<b>2. táblázat.</b> Vizsgálatainkban elemzett minták jellemzői. ....	51
<b>3. táblázat.</b> A szuicid magatartáshoz kapcsolható gének régió-alapú asszociációja és fogékonysági skálán kifejezett genomikus heritabilitása a bipoláris (BIP), major depresszív (MDD), és a kombinált mintán. ....	65
<b>4. táblázat.</b> A DICER1 gén nominálisan szignifikáns markereinek epigenetikai funkcionális annotációja a GAIN major depresszív zavar (MDD) és bipoláris zavar (BIP) mintákon. ....	66
<b>5. táblázat.</b> Az NTRK2 gén nominálisan szignifikáns markereinek epigenetikai funkcionális annotációja a GAIN major depresszív zavar (MDD) és bipoláris zavar (BIP) mintákon. ....	69
<b>6. táblázat.</b> A szuicid magatartás poligénes predikciójának vizsgálata a GAIN bipoláris mintán az LDpred bayesi megközelítésű, információt maximalizáló modellje szerint. .	81
<b>7. táblázat.</b> A p-érték rangsorban első 10 gén és transzkript régió-alapú asszociációja az explorációs és a replikációs mintán .....	83
<b>8. táblázat.</b> Nominálisan replikált kanonikus útvonalak és átfedő útvonalakból összeállított, szuperútvonalak asszociációs (HYST) illetve feldúsulási (Wilcoxon) p-értékei az explorációs és a replikációs mintán. ....	84
<b>9. táblázat.</b> Nominálisan replikált pozicionális génszettek asszociációs (HYST) illetve feldúsulási (Wilcoxon) p-értékei az explorációs és a replikációs mintán. ....	85
<b>10. táblázat.</b> Statisztikailag szignifikánsan replikált, funkcionális annotációs klaszterek feldúsulási pontjai (ES), és terminusaik feldúsulási p_EASE értékei az explorációs és replikációs minta prioritizált génjei szerint. ....	87

## Ábrák jegyzéke

<b>1. ábra.</b> Az asszociációs vizsgálatok problémája a pszichiátriai genetikában. ....	7
<b>2. ábra.</b> A BDNF gén nominális régió-alapú asszociációjának megfelelő szimulált poszterior statisztikai erő a GAIN bipoláris (A) és major depresszív zavar (B) mintákban a feltételezett DSL lókuszek száma, esélyhányada, és a statisztikai teszt függvényében.....	62
<b>3. ábra.</b> A szuicid magatartás PRS predikciója a korrelációs génszett (Génlista1) poligénes rizikó profiljai szerint. ....	78
<b>4. ábra.</b> A szuicid magatartás PRS predikciója a prefrontális pályákon kifejeződő miRNS gének (Génlista2) poligénes rizikó profiljai szerint. ....	79
<b>5. ábra.</b> A szuicid magatartás PRS predikciója a prefrontális pályákon kifejeződő miRNS-ek és célgénjeik (Génlista3) poligénes rizikó profiljai szerint. ....	80
<b>6. ábra.</b> A multilokus elemzések nominális, és többszörös összehasonlításra korrigált replikációs valószínűsége (P_rep) DSM-IV szkizofrénia kétlépcsős elemzésében. ....	88
<b>7. ábra.</b> Az NTRK2 gén SNP asszociációi, epigenetikai és genomikai funkcionális annotációi az MDD, a bipoláris és a kombinált mintákban. ....	92
<b>8. ábra.</b> A DICER1 gén SNP asszociációi, epigenetikai és genomikai funkcionális annotációi az MDD, a bipoláris és a kombinált mintákban. ....	95

## 1. Bevezető

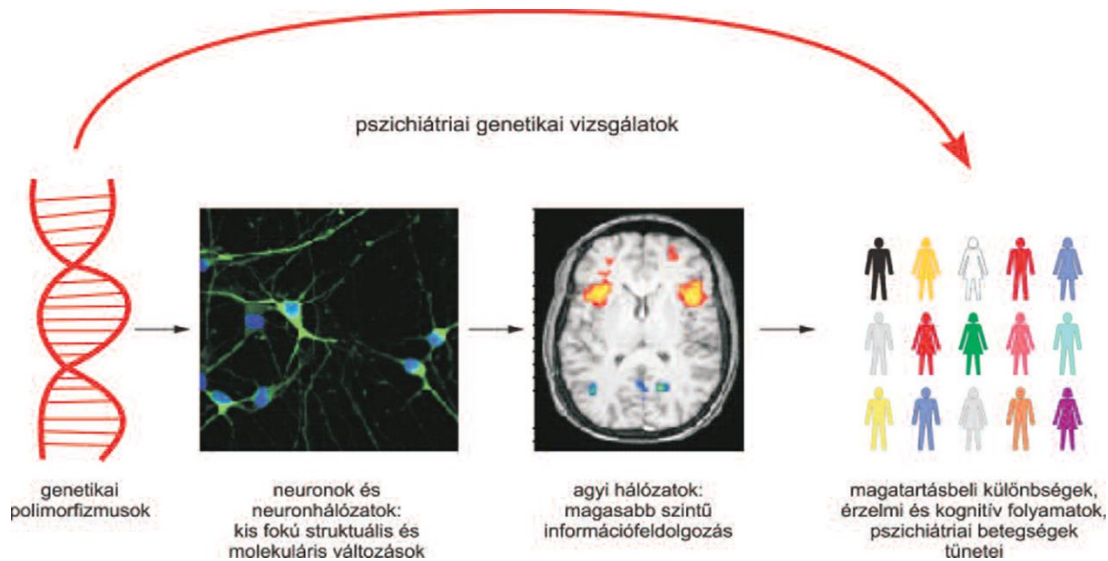
### 1.1. A pszichiátriai genetika általános kérdései

A pszichiátriai betegségek genetikai alapjaival foglalkozó tudományág a pszichiátriai genetika. Régi megfigyelés, hogy egyes személyiségjellemzők, vagy akár egyes pszichiátriai betegségek gyakrabban fordulnak elő vérrokonok között. A családi halmozódás mértéke, pszichiátriai zavarok öröklésmenetének felderítése állt hagyományosan a pszichiátriai genetikai vizsgálatok középpontjában. A család- és ikervizsgálatok a pszichiátriai betegségek mennyiségi-poligénes jellegét igazolták, és több zavar (pl szkizofrénia, bipoláris zavar, autizmus) esetében magas örökölhetőséget (heritabilitás,  $H^2$ ) jeleztek. A heritabilitás és öröklésmenet megismerésével a pszichiátriai genetikai vizsgálatok fókuszába a genom esetleges kóroki, vagy biomarkerként használható variációinak feltárása került.

A pszichiátriai betegségek komplex fenotípusok, melyek diagnózisa viselkedési tünetcsoportok fennállásán alapul. A komplex fenotípusokra jellemzően a viselkedési tünetek háttérében kvantitatív vonások állnak, amelyeket konvencionális küszöbök szerint választunk szét egészségesre vagy kórosra. A közelmúltban bevezetett új klasszifikációs rendszerek, a DSM-5 és a Research Domain Criteria (RDoC) közös célja volt az önkényes konvenciók helyett, az idegtudományi, genetikai és neurobiológiai ismeretekre és biomarkerekre alapozott nozológia és diagnosztika kidolgozása. A tudomány mai állása alapján azonban erre még nincs lehetőség, így az új diagnosztikai rendszer sem eszerint épül még fel.

A pszichiátriai diagnózisok elemi alapjául szolgáló viselkedési tünetek is bonyolult neuroimmunoendokrin hálózatok működésének megváltozását jelentik, de még ezen folyamatok is magasan a genetikai hatások szintje felett vannak (**1. ábra**). A sok közbeeső lépcsőnek köszönhetően, a tünetekből nagyon nehéz az egyedi genetikai variánsok hatására pontosan következtetni. Az azonosított asszociációk értelmezésekor figyelembe kell vennünk a köztes szintek szabályozó, módosító hatását, a gén-gén és a gén-környezeti kölcsönhatások lehetőségét.





**1. ábra.** Az asszociációs vizsgálatok problémája a pszichiátriai genetikában.

(Réthelyi, 2015 [1] a szerző engedélyével)

A fenotípus és genotípus közötti nagy távolság mellett a genetikai faktorok azonosítását a poligénes, kvantitatív öröklődésre jellemző, nagyfokú **genetikai heterogenitás** és **pleiotrópia** is nehezíti. A **genetikai heterogenitás** fennállásakor egy adott fenotípus háttérében több, különböző genotípus is állhat, míg a **pleiotrópiáról** akkor beszélünk, ha egy adott genotípus több fenotípushoz is társulhat. Mindkét jelenség a pszichiátriai zavarok poligénes jellegén alapul, vagyis egy-egy variáns önmagában nagyon kis hatású, a fenotípus sok allél összeadó hatásából ered, vagyis sokféle kombináció eredményezheti ugyanazt a fenotípust. Ennek köszönhető a fent említett folyamatos átmenet az egészséges és patológiás viselkedés között, és a háttérben álló genetikai sokféleség, ami megnehezíti a felfedezett egyedi asszociációk validálását.

Jelen dolgozatban két pszichiátriai fenotípust, a szuicid magatartást és a szkizofréniát vizsgáljuk, melyek klinikai szempontból nagy népegészségügyi jelentőségű betegségek. Arra teszünk kísérletet, hogy a bioinformatika fejlődésének köszönhető, új típusú multilókusz elemzésekkel átfogóbban ragadjuk meg ezen komplex fenotípusok genetikai hátterét

## 1.2. Multilókusz vizsgálati módszerek a pszichiátriai genetikában

A pszichiátriai zavarok genetikai hátterét vizsgáló módszerek fejlődését az elérhető genotípezálási eljárások és az előző fejezetben ismertetett, összetett etiológiájuk egyaránt meghatározták. A teljes genom analízisét lehetővé tévő technológiák megjelenése előtti, ún. „pregenomikus” korszakot, a genotípezálási eljárások (polimeráz láncreakció (plomerase chain reaction, PCR) alapú technológiák, restriktációs fragment hosszúság polimorfizmus (RFLP), kapilláris szekvenálás stb.) magas időigénye és alacsony költséghatékonysága miatt, elsősorban genetikai epidemiológiai vizsgálatok és hipotézis alapú, kandidáns gén vizsgálatok jellemezték. Ez utóbbi vizsgálatok fókuszában a rendelkezésre álló biológiai, klinikai ismeretek, vagy korábban közölt eredmények szerint válogatott, korlátozott számú marker (SNP, mikroszatellita vagy variable number of tandem repeats, VNTR) asszociációjának, transzmissziójának vagy transzkripciójának elemzése állt. Az eredmények többsége a későbbiekben nem volt replikálható, amely a pszichiátriai zavarokat jellemző lókusz- és fenotípus heterogenitás, valamint poligénes öröklésment ismeretében nem meglepő. A biológiailag plauzibilis, funkcionális variációk konzisztens replikálhatósága ugyanis alacsony genetikai heterogenitást feltételez, amely a mendeli öröklésmentű, mono- vagy oligogénes etiológia jellemzője.

A genetikai vizsgálatok korszakváltása, a genomika korának kezdete, 2003. április 14-re, az National Human Genome Research Institute (NHGRI) Human Genome (HUGO) Projektének [2] hivatalos lezárására, és az első teljes emberi referenciagenom közzétételére datálható. A következő mérföldkőnek a szintén az NHGRI által elindított, International HapMap Project [3] tekinthető, amely 2005-ben, majd 2007-ben az emberi alapítópopulációk haplotípus és kapcsoltsági térképeit [4,5] publikálta. A váltás legfontosabb facilitáló tényezőjének az egyszerre több százezer, későbbiekben akár 2 millió, gyakori (MAF>0.01) nukleotid polimorfizmus gyors és viszonylag olcsó genotípezálására, vagy a génexpresszió vizsgálatára alkalmas DNS-chip technológia megjelenése tekinthető. Elterjedésének köszönhetően megindult a pszichiátriai zavarok teljes-genom asszociációs térképezése, emellett a sztenderd genotípezáló platform jelentősen javította a publikált eredmények összehasonlíthatóságát is. A teljes-genom adatok maximális kiaknázásának elősegítése érdekében amerikai részről az NHGRI Genetic Association and Information Network (GAIN) [6] és a National Center for Biotechnology Information (NCBI) által fenntartott database of Genotypes and

Phenotypes (dbGaP) [7,8], európai részről pedig a European Bioinformatics Institute, European Geno-phenome Archive (EGA) [9] biztosítja a vizsgálatok adatainak archiválását és hozzáférését. Az archivált teljes-genom asszociációs vizsgálatok (genome-wide association study, GWAS) vizsgálatok elérhetősége mellett a biológiai információk genomikus annotációit biztosító adatbázisok is egyre részletesebbé váltak, lehetővé téve az ismeretanyag szintézisét. Elindult a HapMap Project küldetését folytató és kibővítő 1000 Genomes Project [10] is, amely a kontinentális populációk mellett átkevert populációkból is származó, 1094 emberi referenciagenom és kapcsolódó gyakori és ritka variációk leírását tűzte ki célul.

A pszichiátriai zavarok GWAS vizsgálatok első hullámát [11–14] azonban kezdetben csalódás fogadta. A nagyszámú marker tesztelése ugyan biztosította a hipotézismentességet, ugyanakkor a többszörös tesztelésből eredő álpozitív asszociációk kiszűrésére végzett, genomikus p-érték korrekció után csak kevés asszociáció maradt szignifikáns. A szignifikáns vagy szuggesztív asszociációk pedig jellemzően nem replikálódtak a későbbi GWAS vizsgálatokban [15]. Emellett a publikált asszociációk a pszichiátriai zavarok ikervizsgálatokban becsült, jelentős heritabilitásának csupán elhanyagolható töredékszázalékát fejezték ki, ez a „hiányzó heritabilitás” (missing heritability) [16] néven elhíresült probléma inspirálta a poligénes módszerek fejlesztését [17–21]. A GWAS chipek a gyakori, sok más szomszédos nukleotiddal erős genetikai korrelációt mutató, ún. jelölő (tagging) SNP markereket tartalmazták, a ritka variánsokról legfeljebb közvetett információt hordoztak. A nagy teljesítményű, teljes genom, vagy exom gyors szekvenálására képes, új generációs szekvenálási technológia olcsóbbá válásával, a ritka variánsok és esetleges de novo mutációk felfedezése érdekében [22] elindult, és jelenleg is bővül a szekvenált és a NCBI dbGaP-en keresztül hozzáférhető pszichiátriai minták száma. A teljes genom és exom szekvenálás módszere tovább bővítette az annotációs és referencia adatbázisokat is (pl., NCBI dbSNP, dbVar, ClinVar, 1000 Genomes Phase3, GENCODE [23,24]). Végül, a genom funkcionális elemeit feltérképező ENCODE Projekt [25] és a humán génexpressziós adatbázisok (NCBI eQTL Browser, GTEx, Geuvadis, GeneVar [26–29]) megnyitották az utat a genom és transzkriptom integrálása előtt.

A ritka variánsok keresése mellett a GWAS továbbra sem került ki a képből, hiszen a technológia alapjául szolgáló „gyakori betegség – gyakori marker” elv [30] által

definiált, részleges penetranciájú, a fenotípushoz önmagukban nagyon kis hatáserősséggel asszociálódó variánsok jól illeszkednek a pszichiátriai zavarok poligénes, nagy lókuszt heterogenitással és pleiotrópiával jellemzett genetikai modelljéhez. A GWAS vizsgálatok fejlesztésének egyik iránya az extrém gyenge hatáserősség észlelésére is képes statisztikai erő elérése, a minta növelésével [31]. A Psychiatric Genomics Consortium által képviselt megközelítés egyik kritikája, hogy az így létrejövő „mega-analízisek” paradox módon tovább fokozhatják a lókuszt heterogenitást, és a multicentrikus gyűjtés szükségessége miatt, a populációs felrétegződést is.

A GWAS elemzések másik lehetséges iránya az egyedi SNP asszociációk minél pontosabb észlelése helyett az összeadó hatásokkal számoló, és a nukleotidok lókuszt heterogenitását megengedő, multilókuszt elemzések alkalmazása. A nagyszámú multilókuszt elemzés közös jellemzője, hogy az egyedi markereket valamilyen szempont szerint csoportosítja, és az így képzett összetett „markereket” teszteli az adott módszernek megfelelő statisztikai eljárással [32–37]. Ennek hatására egyrészt nagymértékben csökkenhet a tesztek száma és a következményes p-érték korrekció mértéke, másrészt a csoportosító elv biológiai információt is integrál (pl. haplotípus blokk – kapcsoltág/diverzitás, régió-alapú tesztek – gén, transzkript, génszett), utalva a funkcionalitásra. Végül, de nem utolsósorban, az utóbbi évek gyarapodó számú multilókuszt vizsgálatai megerősítették Ayalew és munkatársai eredményeit [38], miszerint a valid, funkcionális a priori ismeretek alapján meghatározott multilókuszt „markerek” részleges megoldást jelenthetnek az SNP heterogenitásból eredő replikációs nehézségekre. A szövetspecifikus transzkriptom térképet integrálni képes módszerekkel [39,40] pedig a funkcionális prioritizálás a gyakorlatban is egyszerűen megvalósítható.

A publikált módszerek száma az utóbbi években ugrásszerűen növekedett, a fejezetben a vizsgálataink szempontjából is releváns haplotípus-, régió-, és génszett-és génhálózat alapú asszociációs elemzéseket, valamint a genomikus heritabilitás és a poligénes additív rizikó becslésén alapuló módszereket mutatjuk be.

### 1.2.1. Haplotípus alapú elemzések

A haplotípusok nagy jelentőséggel bírnak a genetikai vizsgálatokban. Felhasználhatók jelölő (tagging) markerek meghatározására, vagy multilokusz markerként asszociációs vagy populációgenetikai vizsgálatokban, a kiterjesztett haplotípusok pedig jelentősen gyorsíthatják a hiányzó markereket pótló imputációt, vagy kapcsoltsági referenciaként szolgálhatnak régió-alapú tesztekben. A haplotípus elemzések két legfontosabb szempontja a vizsgált haplotípus (blokk) határainak és alléljainak meghatározása.

#### 1.2.1.1. Haplotípus allélok rekonstrukciója: gamétafázis-elemzés

A haplotípus a lokuszok (nukleotidok, strukturális variánsok) alléljainak sorrendje az adott szülői kromoszómán, mérete két nukleotidtól akár a teljes kromoszóma szekvenciájáig terjedhet. A humán kromoszómakészlet diploid, vagyis a vizsgált lokuszok haplotípusai tekintetében két homológ alléllal rendelkezünk (ez alól csak a férfiak haploid nemi kromoszómái jelentenek kivételt). Egyes kísérleti jellegű eljárásokat [41–43] leszámítva, a genotípezálás az allélok szülői eredettől független, diploid genotípusát adja meg, a haplotípusra pedig csak következtetni tudunk bizonyos algoritmusokkal a megfigyelhető genotípusokból. A haplotípus rekonstrukció, másnéven az allélok gamétikus fázisának meghatározása (phasing) [44,45], alapvetően három módon történhet: molekuláris, transzmissziós (családi), valamint populációs fáziselemzéssel.

**Molekuláris fáziselemzés:** a metafázisban leállított sejtekből, vagy fluoreszcens aktivált sejtszétválasztással szeparált, haploid kromoszómák genotípezálásával [42], vagy fosmid könyvtárakkal végzett új generációs szekvenálással [41] akár teljes-genom haplotípus is genotípezálható, de túl drága és lassú ahhoz, hogy széles körben alkalmazható legyen. Ugyanakkor az új generációs szekvenálás „paired-end read” technológiájú változatainál a leolvasási szakasz (read) informatív a fázisra, bár önmagában túl rövid a haplotípus allél rekonstrukciójához, a populációs frekvencián vagy leszármazási azonosságon (identity by descent, IBD) alapuló algoritmusokkal kombinálva azok pontosságát jelentősen növeli (pl. SHAPEIT2, [46]).

**Transzmissziós (családi) fázis elemzés:** triók vagy összetett családi minta esetén, a komplett, informatív szülői genotípusok birtokában a gaméta fázis a transzmisszióból könnyen megállapítható, a nem-informatív transzmisszió (mindkét szülő heterozigóta) esetén pedig a fázis valószínűsítésére a populációs frekvencián vagy IBD számításra alapuló algoritmusok használhatók. Előnye, hogy ritka variánsok haplotípusa is kiszámítható, és a populáció stratifikáció nem torzítja. Hátránya, hogy 50%-kal magasabb genotípusozási költséget jelent, valamint a szülői minta nem mindig elérhető [45].

**Populációs fázis elemzés:** rokonságban nem álló egyéneket feltételezve, a lokusz allélok frekvenciáját és genetikai kapcsoltságát (kapcsoltsági egyensúlytalanság, linkage disequilibrium, LD), kiterjesztett haplotípus (pl. teljes kromoszóma) esetén pedig referencia rekombinációs térképet is alapul véve becsülik meg a legvalószínűbb gamétikus fázist. A populációs fázis elemzés a legolcsóbb és leggyorsabb a három fázis elemző eljárás között, ugyanakkor csak valószínűsíteni képes a fázist, pontossága a mintanagyságtól és az allélfrekvenciáktól függ, emiatt érzékeny a populáció stratifikációra is. A fázis valószínűsítésére számos algoritmust fejlesztettek. A korai modellek, pl. a kombinatorikai parszimónia elvet követő Clark-algoritmus [47], és az Expectation-Maximisation (E-M) iteratív, maximális-valószínűség alapú fázis elemzések [48], a rekombináció hiányát feltételezve, szorosan kapcsolt, maximum néhány tíz variánst tartalmazó haplotípus allélok fázisának kiszámítására alkalmasak. A rekombinációt vagy mutációt a koaleszcens priori eloszlással modellező PHASE [49], a minta lehetséges haplotípusaiból visszavezethető ancesztrális allélok poszteriori valószínűségével számítja a fázist. A kalkulációhoz használt MCMC modell miatt pontos, de kb. 100 lokusznál hosszabb haplotípusoknál már túlságosan lassú. A modernebb, fázis számításához rejtett Markov modellt (Hidden Markov Model, HMM) és E-M algoritmust használó fastPHASE [50], MACH, IMPUTE2 [51], SHAPEIT1-2 [52,53] és az IBD kalkulációt BEAGLE [54] már teljes kromoszóma haplotípusok fázis számítására is alkalmasak, a rekombináció modellezéséhez az International HapMap Project által publikált humán haplotípus térkép [5] rekombinációs valószínűségeit veszik alapul.

## 1.2.1.2. Haplotípus blokkstruktúra, kapcsoltság, rekombináció

A gamétafázis rekonstrukciós problémáján túl, a haplotípus vizsgálatok másik fontos kérdése a haplotípus blokkok feltérképezése. Haplotípus blokknak egymással szoros kapcsoltságban álló lókuszok csoportját nevezzük, ahol a rekombináció valószínűsége alacsony [4,55,56]. A genetikai kapcsoltságot a populációgenetika terminológiájából származó kapcsoltsági egyensúllyal (linkage equilibrium), vagy annak hiányával (linkage disequilibrium, LD) jellemezzük. A genetikai kapcsoltság mérésére az alábbi két mutató használata terjedt el [4,57].

A **sztemdizált kapcsoltsági együttható ( $D'$ )** értéke 0 és 1 között változhat, Aa és Bb lókuszok esetén  $D_{AB} = p_{AB} \cdot p_{ab} - p_{Ab} \cdot p_{aB}$  és  $D' = D_{AB} / D_{max}$ , ahol  $p_{AB}$ ,  $p_{ab}$ ,  $p_{Ab}$ ,  $p_{aB}$  a haplotípus allélok frekvenciáit jelölik.  $D' = 0$  esetén a vizsgált két allél haplotípusának frekvenciája megegyezik a gyakoriságuk szorzatával, függetlenek egymástól,  $D' = 1$  esetén a két allél között tökéletes a kapcsoltság, a lehetséges 4 haplotípusból csak maximum 3 észlelhető.

A másik, újabban gyakrabban használt változó, a **Pearson -féle korrelációs együttható négyzete ( $r^2$ )**. E mutató levezethető a sztemdizált kapcsoltsági mutatóból:  $r^2 = D^2 / p_A \cdot p_a \cdot p_B \cdot p_b$ , a fenti két lókusz példája alapján. Tökéletes kapcsoltság  $r^2 = 1$  esetén áll fent, a lókuszok alléljai ekkor mindig ugyanabban a fázisban vannak, míg  $r^2 = 0$  esetén az allélok függetlenek egymástól.

A két mutató bár egymásból levezethető és a szélső értékeik is megegyeznek, mégsem teljesen ugyanúgy értelmezendők. A  $D'$  az allélok frekvenciától függetlenül méri a köztük lévő kapcsoltságot, és elsősorban a kapcsoltsági egyensúlyt megváltoztató folyamatokat jelzi: rekombináció hiánya, szelekció, sodródás, palacknyak (bottleneck) hatás, populációs felrétegződés (stratifikáció), genotípuszálási hiba. Ezzel szemben az  $r^2$  az allélok kölcsönös predikciós erejét mutatja, vagyis, hogy mennyire függetlenek, egyediek a lókuszok asszociációs jelei. A  $D'$  biológiailag informatív, ugyanakkor alacsony mintanagyságnál értéke pontatlan, inflált, és csak az adott vizsgálat markereire értelmezhető. Ezzel szemben az  $r^2$  a markerek korrelációja szempontjából, vagyis statisztikailag informatív, ismeretében meghatározhatók a független asszociációs jelek, vizsgálatok adatait összevetve is értelmezhető, viszont érvényessége megkérdőjelezhető, ha a két allél frekvenciája nagymértékben különbözik egymástól [58].

A rekombinációs valószínűség és haplotípus blokkok partícionálására többféle módszer is ismert [45,55–57,59–61]. A kapcsoltság alapú definíciók a rekombináció minimalizálásával és a  $D'$  maximalizálásával határozzák meg a blokkhatárokat. Példaként a Gabriel-féle algoritmus [55] említendő, amely a blokkhatárokat képező markerek között  $D'$  konfidenciahatárok minimális értékével, és a blokk lókuszi közötti rekombináció maximális arányával definiálja (min.  $D'$   $CI_{\text{alsó}} > 0,7 - CI_{\text{felső}} > 0,95$ , rekombináció  $< 0,05$ ). A kapcsoltság mérésén kívül a blokkhatárok meghatározhatók a haplotípus allélok csökkent diverzitása alapján [59], vagy információ elméleten alapuló módszerek [62] alapján is. A blokkok jelentőségét elsősorban a blokkok haplotípus allélja inak variációját leképező, ún. haplotípus jelölő SNP (htSNP) szett meghatározása adta, amelytől jelentős dimenzióredukció volt remélhető az asszociációs vizsgálatokban [56]. Szimulációs vizsgálatok alapján, a különböző módszerekkel definiált blokkok számuk és pozícióik alapján nem voltak erős konszenzusban, a blokkhatárok pedig nagymértékben függtek a markerek denzitásától és frekvenciájától, ugyanakkor a lefedett régiók tekintetében nagy átfedést mutattak [56,63]. A klasszikus partícionáló algoritmusok további hátulütője, hogy uniformizálják az LD felmérését, miközben számos genomikus régióban az algoritmusok merev paramétereivel, pl. az 500 kb távolság betartásával információt veszítenek. Hosszabb szakaszok, vagy sűrűn genotípezált, ritka variánsokat is tartalmazó adatok elemzésére nem alkalmasak, számításigényük az elemzett markerek számának négyzetével növekszik [57].



## 1.2.1.3. Haplotípus elemzések jellemzői: asszociációs tesztek, imputáció

A haplotípus blokk-alapú asszociációs vizsgálatok - a blokkszerkezet particionálás korlátai miatt – kizárólag a kandidáns gén vizsgálatok eszköztárát gazdagították, a GWAS vizsgálatok a haplotípus blokkokat elsősorban a páros és multiallélikus jelölést biztosító SNP szett kiválasztására használták. A számításigény négyzetes növekedése mellett az eltérő lefedettségű GWAS adatok particionálása gyakorlatilag összehasonlíthatatlan szerkezetű blokkokat eredményezne [64]. Egyedül a rögzített hosszúságú, ún. csúszóablak módszer vezetne összehasonlítható haplotípusokhoz, de számítási igénye miatt, még a csúszóablak módszerrel is csak az Alzheimer-kór esetében végeztek teljes-genom haplotípus asszociációs vizsgálatot a neuropszichiátriai fenotípusok közül [65]. A teljes genom szignifikanciával asszociáló haplotípus a FRMD4A génrégióban helyezkedett el. Az asszociáció validitását a replikációk mellett későbbi transzkriptom elemzés is megerősítette, amely egyéb gének mellett a FRMD4A gén differenciált expresszióját is meghatározó fontosságúnak találta az Alzheimer-kór egyik fő patomechanizmusában, a neurofibrillum képződésben [66].

A nehézségek ellenére, a particionált blokkok biológiai információt is hordoznak a véletlenszerűen kiválasztott lokuszok, vagy csúszóablak haplotípusok elemzéséhez képest, vagyis jó eséllyel jelölnek funkcionális - akár nem is genotípezált – variánst. Az asszociációs tesztek a haplotípus markereket multiallélikusként kezelik, ahol az  $n$  allél biállélikus statisztikájának kiszámítása mellett multiallélikus omnibusz khi-négyzet statisztika is generálható  $n-1$  szabadságfokkal [34,36,60]. A blokkokat jellemző fokozott kapcsoltság és csökkent alléldiverzitás miatt e markereknél kevesebb allél várható, amely kisebb szabadságfokot, tehát nagyobb statisztikai erőt jelenthet, mint például a csúszóablak protokoll esetén.

A haplotípusok azonban nemcsak asszociációs markerként, hanem kapcsoltsági referenciaként is használatosak, többnyire a genotípus imputációhoz [67], vagy a multilokus asszociációt aggregált eredményfájlból számító tesztekhez szükségesek [68–71]. Mind az imputációhoz, mind egyes régió-alapú asszociációs számítások paraméterezéséhez jellemzően a kromoszóma teljes szekvenciájára rekonstruált haplotípusok használatosak. Az imputáció során a nagy markerlefedettségű referenciaminta haplotípus alléljainak illesztése történik az imputálni kívánt minta

genotípusából rekonstruált haplotípus allélokhoz. Az illeszkedő referencia haplotípus allélok alapján számítható a mintában nem szereplő lókuszok alléljainak valószínűsége, amely tükrözi a haplotípus fázis elemzésből és az illesztés minőségéből eredő imputációs bizonytalanságot. Az imputáció számítás- és időigényes lépéseit jelentősen felgyorsítja, ha fázis elemzett genotípust, azaz rekonstruált haplotípusú mintát imputálunk [72]. A fent említett multilókusz tesztek pedig, SNP asszociációs statisztikák és referencia haplotípusokból számított LD alapján, genotípus hiányában is képesek multilókusz asszociáció statisztikát számítani, ami különösen hasznos publikált asszociációs vizsgálatok metaanalízise esetén.

Összefoglalva, a haplotípus, azaz a gaméta-fázis-informált genotípus, a linkage-vizsgálatok mellett a legkorábbi multilókusz vizsgálatnak tekinthető. Az általánosan elterjedt genotípező eljárásokkal a fázis nem figyelhető meg közvetlenül, hanem vagy a szülői genotípusból, vagy hiányos adatok kiegészítésére alkalmas valószínűség-számítási algoritmusokkal következtethető ki. A HapMap projekt haplotípus térképeiből ismert, hogy az egymáshoz közel elhelyezkedő markerek nagy valószínűséggel szorosan kapcsolatosak, tehát a kromoszómális haplotípus egymástól nagy rekombinációs valószínűségű régiókkal elválasztott blokkokból épül fel, amelyek a rekombináció mellett egyéb populációgenetikai jelenségekre (palacknyak-hatás, sodródás, szelekciós söprés) is informatívak. A blokkokat leképező algoritmusok azonban nem alkalmasak teljes genom minta elemzésére, ezért GWAS vizsgálatokban a legkisebb számú, htSNP és tSNP jelölő markerkészlet meghatározásához használták, a tesztekben részt vevő markerek számát csökkentve. Mindemellett, a teljes kromoszóma haplotípus allélok felgyorsíthatják az imputációt, és LD referenciaként használva, lehetővé tehetik bizonyos multilókusz asszociációk számítását genotípus hiányában is.

## 1.2.2. Régió-alapú asszociációs vizsgálatok

A haplotípusok mellett egyéb multilokusz markerek meghatározásának egyszerű, de érvényes módja az egyedi polimorfizmusok valamilyen régió szerinti csoportosítása, és a régió közös asszociációjának számítása. A haplotípusokkal szemben azonban e módszerek nem igényelnek fázisinformációt, gyakran genotípust sem, csak összesített SNP asszociációs statisztikát [32,35,68,69,71,73,74]. Kézenfekvő csoportosító régió a gén, hiszen az SNP alapú asszociációk elemzésekor is az asszociált markert tartalmazó (esetleg a közelében található) gén(ek) alapján került sor a szignifikáns eredmények értelmezésére [32,33,35,75]. A régió határai ugyanakkor könnyedén leszűkíthetők a transzkript határokra, szabályozó régiókra, vagy kiterjeszhetők a kódolt fehérjével azonos molekuláris funkciót betöltő fehérjék génjeire is, utóbbi már egylépcsős, önálló génútvonal elemzésnek is tekinthető. A régiók definiálhatók pusztán genomikus koordinátákkal, vagy génhatároktól független biológiai annotációkkal is (pl. az ENCODE projekt által leírt promoterek, a transzkripció fokozott valószínűségére utaló DNase hiperszenzitív helyek (DHS), vagy kromatinállapotok, hiszton markerek [25]). A régió sokrétű értelmezése ellenére, a legtöbb régió-alapú asszociációs vizsgálat praktikus szempontokból vezérelve a génekre fókuszál. A gének funkció és koordináták szerint is részletesen katalogizáltak, ráadásul a meta-analízisük módszertana is részletesen kidolgozott [33,76]. A továbbiakban ezért a gén-alapú vizsgálatokkal foglalkozunk, a részben szintén régió-alapú génútvonal elemzéseket a következő szekcióban tárgyaljuk.

### 1.2.2.1. A gén-alapú asszociációs vizsgálatok általános szempontjai

A gén-alapú asszociáció számítása a gén markereinek genotípusos varianciájának egyesített tesztelésével vagy a markerek egyedi asszociációinak kombinálásával történhet. A többi multilokusz elemzéshez hasonlóan, a teljes-genom génkészleten alapuló teszt egyrészt a jelentős dimenzióredukció révén, a többszörös tesztelés korrekcióját mérsékelve ( $p_{\text{GÉN}} < 1E-06$  és  $p_{\text{SNP}} < 5E-08$ ), másrészt az alkalmazott statisztikai módszertől függően, az összevont markerek kis hatáserősségű asszociációt erősítve növelheti a statisztikai erőt. Emellett, az SNP-k asszociációinak vagy genetikai variációinak egyesítése a minták közti heterogenitást csökkenti, vagyis a replikáció valószínűségét is megnövelheti [33,38]. Azonban éppen a megnövekedett komplexitásuk

miatt a gén-alapú tesztek eredményeinek összevethetősége nem olyan egyértelmű, mint az SNP asszociációké. Az összemérhetőséget és a replikációt többnyire az eltérő annotáció, genotípezálás, továbbá a populációs felrétegződés eltérő kezelése nehezíti meg. [35]

**Annotációs bizonytalanságok:** A génrégióknak a transzkripciót szabályozó, közeli szekvenciák variációt is javasolt lefednie, ezért a vizsgálatokban a génhatárok 5, 10 vagy 50 kilobázis méretű kiterjesztése a gyakorlat, aranystandard érték nincs. A kiterjesztés miatt nagyobb az átfedések és többszörös SNP annotációk száma, fokozva a gének közötti korrelációt, és megnehezítve a későbbi génútvonal elemzések értelmezését. Továbbá a génhatárok és annotációk már több forrásból is nyerhetők, ezek között jelentős eltérések lehetnek (pl. NCBI refSeq vagy EMBL-EBI GENCODE [24,77]) és a genom összeállítási (build) verzió (pl.hg19, GRCh37, GRCh38) jelölésének elmaradása is gátolja az összevethetőséget.

**Genotípezálási különbségek:** A haplotípus blokkok particionálásához hasonlóan, a gének asszociációját és esetleges rangsorolását jelentősen befolyásolhatja, ha az egyedi markerek eloszlása nem egyenletes. A GWAS chipek esetében ez gyakori probléma, különösen, ha több, különböző platformokon genotípezált minta vesz részt az elemzésben. Ilyenkor megoldás lehet egy nagy markersűrűségű, referencia populációval (pl. 1000 Genomes Project) végzett imputáció, viszont ebben az esetben, a megnövekedett idő- és erőforrásigény mellett számolni kell az imputációs bizonytalanságból eredő műtermékekkel is [67].

**Populációs struktúra korrekciója:** a vizsgált populációk eltérő földrajzi eredetéből, vagy a rejtett rokonsági fokból eredő allélfrekvencia különbségek a gén-alapú asszociáció számítását is torzíthatják, erre az SNP statisztikát kombináló, a kapcsoltság hatását nem permutációval korrigáló módszerek [68,69] különösen érzékenyek, mivel a kombinálandó SNP asszociációk inflálódása mellett, a független asszociációk azonosítása is hibás lehet az LD torzulása miatt. Esetükben az LD számítását célszerű a vizsgált mintához legjobban illeszkedő referencia populáción végezni, pl. 1000 Genomes vagy HapMap populációk [10,58].

### 1.2.2.2. Gén-alapú asszociációs elemzések módszerspecifikus jellemzői

A jelenleg elérhető asszociációs tesztek, a régió-alapú statisztika számítási módja szerint alapvetően két nagy csoportba sorolhatók: megkülönböztetünk a genotípusos varianciát transzformáló, valamint az SNP asszociációs statisztikát kombináló módszereket. [35] A két kategória alapjaiban tér el egymástól a bemeneti adatok és optimális alkalmazhatóság tekintetében, ezért az alábbiakban mindkettőt részletesen bemutatjuk.

#### 1.2.2.2.1. A genotípusos varianciát transzformáló eljárások

E módszerek a vizsgált régió egyedi nukleotidainak genotípusát, a minor allélok számát tükröző pontozással folyamatos változókká alakítják, majd kiszámítják a transzformált genotípus főkomponenseit. Az első néhány főkomponenssel a régió genetikai varianciájának 80-90%-a leképezhető, beleértve az LD mintázatot és a populációs rétegeket is. A legkorábbi eljárás, a főkomponens-alapú regressziós teszt [78] ezeket a főkomponenseket logisztikus regresszió prediktoraként együttesen használva, a modell omnibusz valószínűségi hányados tesztjével (likelihood ratio test) fejezi ki a gén asszociációját. A módszer hátránya, hogy az asszociációért felelős SNP-k hozzájárulása nem követhető vissza. A módosított változata, a főkomponens-alapú genotípus klaszterezés [79,80] ezt a hiányosságot pótolta. Módosítása, hogy főkomponensek helyett a főkomponensek rotálásával klaszterekbe rendezett, SNP genotípus-klaszterek a prediktorok. A logisztikus kernel gép regressziót alkalmazó Sequence Kernel Association Test, SKAT [81,82], sokoldalú, különböző peremfeltételeknek megfelelően parametrizálható módszer. Kevert modellként, különböző eloszlású, akár egymással összetett interakcióban is lévő prediktorok egyesített eloszlásának többváltozós asszociációja is kiszámítható. Emiatt a SKAT-teszt alkalmazható ritka variánsok és gyakori polimorfizmusok együttes eloszlásának érvényes tesztelésére is, amely az új generációs szekvenálás növekvő elterjedése miatt egyre hangsúlyosabb szempont. Ide tartoznak még a fenotípusokat együttesen elemző, és ezzel az elvégzett tesztek számát csökkentő regressziós modellek. A többváltozós, többszörös lineáris regresszió [83], és a ritka variánsokra optimalizált, adaptív súlyozású, reverz regresszió [84] említendő.

A csoport utolsó tagja a gén-alapú tesztek új irányát jelentő PrediXcan [39], újdonsága a genotípus variancia által meghatározott transzkriptom „imputálásán” és tesztelésén alapul. A vizsgált gén asszociációját a gén markereinek poligénes varianciája által determinált, expressziós változás (GReX) és a fenotípus korrelációjával fejezi ki, amely a hatáserősség mellett az asszociáció irányát is jelzi. A GReX poligénes predikcióját elasztikus hálóval, vagy LASSO regresszióval számítja a GEUVADIS [85], a Depression Genes and Network (DGN, Battle et al. 2014) és a Genotype-Tissue Expression (GTEx, Melé et al. 2015; Ardlie et al. 2015) és a Braineac [87] projektek expressziós adatai alapján. A SKAT-teszthez hasonlóan, mind a ritka, mind a gyakori variánsokat képes használni, ugyanakkor a fenotípussal a genotípus helyett, a genotípus által determinált transzkripció varianciát korreláltatja, egyedülálló módon, egyszerre teszteli a gén asszociációját és funkcionalitását.

#### 1.2.2.2.2. Asszociációs statisztika kombinálásán alapuló módszerek

A második kategóriába tartozó tesztek a gén-alapú asszociációt az SNP asszociációk p-értékeiből, vagy egyéb statisztikájából számolják [35]. Előnyük, hogy az SNP asszociációs értékeken kívül csak valid LD referenciát igényelnek, a minta genotípusára nincs szükség. Az SNP asszociációk gén-alapúvá alakítása szerint a tesztek további két csoportra oszthatók: a **„legjobb SNP”** algoritmusokra, valamint az **„SNP kombináló”** módszerekre [68,73,74].

A **„legjobb SNP”** csoportba tartozó tesztek a gén legkisebb SNP p-értékét korrigálják a gén méretéből, azaz a független SNP tesztek szabadságfokából eredő fals pozitív asszociációk ellensúlyozására. A korrekció történhet permutációval pl. PLINK set teszt [88], SIMES-teszt, [89], VEGAS2 Best SNP teszt [71], vagy a referencia LD adatok alapján, a független markerek azonosításával, a kiterjesztett SIMES-teszt (GATES-teszt, Li et al. 2011). Előnyük, hogy robusztusak kapcsoltság mértékére, és a nem asszociálódó, neutrális markerek okozta asszociációs zajra. Hátrányuk a markerek statisztikáját kombináló módszerekhez képest elsősorban a közepes méretű, és arányaiban több független, asszociációt tartalmazó gének esetében észlehető.

Az **„SNP kombináló”** módszerek a gén-alapú asszociációt a lehető legtöbb független marker Fisher-kombinációján [90], vagy annak módosított kiterjesztése in

alapulnak: pl. Rank Truncated Product (TP) kombináció [91], VEGAS2 Sum teszt [71], Skálázott Khi-négyzet Szummáció [92,93], és a génszettek régió-alapú asszociációjának elemzésére is alkalmas Hybrid Set-Based Test (HYST, Li et al. 2012). A Fisher-kombináció lényege, hogy a gén p-értékének számításához a régió markereinek p-értékéből számított khi-négyzet statisztikáját összegzi, amely szintén khi-négyzet eloszlást mutat. A tesztek közötti dependencia, vagyis a kapcsoltág, ismerten torzítja az így számított p-értéket, ezért az LD kezelésére alapvetően két stratégiát alkalmaznak. Az egyik megoldás, a markerek közötti LD ismeretében az összegzendő statisztikák skálázása (HYST, Skálázott Khi-négyzet Szummáció), vagy a statisztika összegzését követően, a multivariáns normál disztribúció szimulációjával empirikus p-érték számítása (VEGAS2 Sum teszt). A kombináción alapuló módszerek előnye a több, független, közepes erősségű, de a genomikus korrekcióhoz gyenge SNP asszociációt tartalmazó gének tesztelésekor érvényesül, amely nagyfokú lókus heterogenitás esetén gyakori scenárió. Hátrányuk azonban, hogy nagyszámú, független neutrális markert tartalmazó gének esetén, a neutrális markerek statisztikai csökkentik a gén-alapú asszociációs szignált, továbbá - a permutációt alkalmazó tesztek kivételével - érzékenyek a kapcsoltág hibás paraméterezésére.

Összefoglalva, a régió-alapú asszociáció számítására már számos módszer áll rendelkezésre, különböző előnyökkel és hátrányokkal. A genotípust transzformáló tesztek populációs felrétegződésre robusztusak, de - a SKAT és a PredXcan kivételével – csak populációs adaton alkalmazhatók, emellett genotípus adatokat igényelnek a számításokhoz. Az SNP statisztikát használó módszerek rugalmasabbak, több vizsgálat metaanalíziséhez egyszerűbben használhatók, mivel nem igénylik az elemzett minták genotípusát, részletes LD referencia pedig a 1000 Genomes Project publikus adataiból rendelkezésre áll. A pszichiátriai zavarokat jellemző magas lókus heterogenitás és poligénes öröklésment modelljéhez legjobban az „SNP kombináló” asszociációs tesztek illeszkednek, azonban alacsony statisztikai erővel bíró GWAS minták elemzéséhez, a gén asszociációját a „legjobb SNP” algoritmussal számító tesztek alkalmasak. A SKAT és a PrediXcan segítségével pedig maximalizálható az új generációs szekvenálásból nyert információ, valamint a transzkriptom integrációjával az asszociációk funkcionalitása is vizsgálható.

### 1.2.3. Génszett- és génhálózat-alapú vizsgálatok

A gén-alapú statisztikához hasonlóan, a génszett elemzés koncepciója is az expressziós vizsgálatokból eredeztethető [94,95]. A gének csoportjainak együttes tesztelése egyrészt segíthet a gén-alapú asszociációs mintázat értelmezésben, finomításában, másrészt a komplexitás növelésével tovább csökkenthető a többszörös összehasonlítás mértéke. Gének csoportosítása és elemzése többféle módon történhet: priori hipotézisek alapján, feltételezett élettani vagy kórélettani folyamatokban betöltött szerepüket tükrözve, vagy támaszkodhat empirikus adatokra, lehet mellérendelő, vagy leképezhet egymás közti viszonylatokat, hierarchiákat, ráadásul az alkalmazott statisztikai módszer tekintetében is jelentős különbségek lehetnek. Emiatt téves a sokszor gyűjtőfogalomként használt „útvonal (pathway) elemzés” elnevezés, mivel ez csak a molekuláris vagy sejtfunkció, patomechanizmus, esetleg kísérleti beavatkozásokra adott válasz priori hipotézise alapján csoportosító elemzéseket jelenti. A génhálózat-elemzések pedig hipotézisek helyett, mért kölcsönhatás adatok alapján, a gráfelméleten alapuló módszerekkel az útvonal- vagy ontológiai vizsgálatok új generációját képviselik [94].

#### 1.2.3.1. Általános szempontok

A génszett-alapú elemzések felfoghatók a régió-alapú tesztek kiterjesztésének, az egy-lépcsős elemzések kivételével, a legtöbb módszer a gének asszociációira, esetleg rangsorára támaszkodik[96]. Tehát itt is érvényesek a régió-alapú teszteknel tárgyalt szempontok: az annotációk forrásának és verzióinak pontos rögzítése, a genomikus lefedettség (hálózatoknál az összkonnektivitás) egyenetlenségeinek kezelése és a korreláló annotációk korrekciója. Ugyanakkor e szempontok fokozottan érvényesek, az annotációs források vagy a markersűrűség eltérései és a listák korrelációja már nemcsak a meta-analízisekben való összevethetőséget akadályozzák, hanem a génszett-asszociáció értékét is torzíthatják [94,95].



**Annotációs források jelentősége:** a génszett-alapú vizsgálatokban használt annotációs adatbázisok több módon is befolyásolhatják a génszett asszociáció eredményét. Egyrészt az adatbázisokban szereplő annotációk és az általuk lefedett gének száma jelentősen különbözhet, Mooney és munkatársai elemzése [94] alapján a Pathway Commons adatbázisában [97] a kanonikus útvonalak mintegy 7000-8000 gént fedtek le, szemben a STRING protein-protein interakciós (PPI) hálózat [98] 20000 vagy a génontológiai (GO) [99,100] annotációk „biológiai folyamatok” kategóriája által lefedett 15000 génnel. Továbbá, a hasonló funkcióra utaló annotációk sokszor eltérő génlistát jelölnek. Emellett az annotációk információtartalma is eltérő, például a GO az általános felől az egyre specifikusabb terminológia felé ágazó, hierarchikusan rétegzett adatstruktúrája egészen más elemzést tesz lehetővé, mint a topológiai információkat gráfokkal integráló PPI adatbázisok. Az elemzések összehasonlíthatóságát tovább nehezíti, hogy a génútvonali listák meghatározására nincs sztenderd eljárás, csak ajánlás (Biopax, Demir et al. 2010).

**Genomikus lefedettség hatása:** az SNP markerek száma, sűrűsége az asszociációs teszt sajátosságának függvényében többféle torzítást is eredményezhet. A gének nem egyenletes lefedettségére különösen a kompetitív tesztek érzékenyek, de az önálló asszociációs tesztek is torzítja. A kétlépcsős analízisekben, pusztán az első lépcsőben alkalmazott, gén-alapú teszt típusa alapján rövid vagy hosszú gének feldúsulása is előfordulhat. A markersűrűség kiegyenlítésében használt imputáció ugyanakkor jellemzően a rövid, neutrális SNP-ket tartalmazó gének feldúsulását okozza. A vizsgált génszett mérete is meghatározó lehet, az extrém kevés vagy éppen sok gént lefedő génszettek egyaránt nagyobb eséllyel lesznek szignifikánsak, ezért kizárásuk bevett gyakorlat, ugyanakkor nincsenek evidencia alapú sztenderdek a küszöbértékek megállapítására [94,95,102].

**Korreláló annotációk (geneset crosstalk):** a génszettek közötti korreláció nemcsak az eredmény értelmezésekor okoz problémát, hanem közvetlenül torzítja is a kompetitív típusú tesztek statisztikáit. A korreláció nagyobb részt az átfedő géneknek, kisebb részben az egymáshoz közeli gének között esetleges szoros genetikai kapcsoltságnak köszönhető [103]. Mivel szinte az összes adatbázist érinti, érdemes az átfedő géneket kizárni, vagy a kompetitív mellett önálló tesztek is használni.

### 1.2.3.2. A génszett- és hálózat alapú vizsgálatok fő típusai

**Biológiai útvonal (pathway) elemzések:** biológiai útvonalaknak azon génszettek tekinthetők, amelyek a géneket a priori molekuláris, metabolikus vagy sejtfunkciós modellek, vagy különféle betegségekben feltételezett patomechanizmus modellek, esetleg gyógyszerre, vagy más intervencióra adott válasz feltételezett modellje alapján listázzák. Előnyük, hogy jelentősen segítik a fenotípussal kapcsolatos patomechanizmus megértését, hátrányuk, hogy az ismert gének kis részét fedik csak le, és megalkotásuk önkényes, manuális, a priori modell érvényessége meghatározza az asszociációk validitását is. Számos adatbázis definiál útvonalakat: Pathway Commons [97], Kyoto Encyclopaedia of Genes and Genomes (KEGG, Kanehisa & Goto 2000; Kanehisa et al. 2012), Reactome Pathways [106], Molecular Signature Database (MSigDB, Subramanian et al. 2005).

**Génontológia elemzések (GO):** a génontológia egyfajta annotációs „szótárként” értelmezhető, ahol a terminológiák hierarchikusan, általánostól az egyre specifikusabb jelentés szintjei felé leágazóan rendeződnek, három ontológiai osztály szerint: Biológiai Folyamatok (Biological Process, BP), Molekuláris Funkció (Molecular Function, MF) és a Sejtkomponens (Cell Component, CC). Az útvonalakkal szemben, a GO annotációk nem próbálják modellezni az egy annotációval jelölt gének kapcsolatát. Előnye, hogy a gének nagy részét lefedi, és nem függ a priori modell érvényességétől. Hátránya az egymásra épülő, inkluzív szintekből eredő, információs redundancia, amely klaszterelemzést vagy a túl általános szintű terminológiák kiszűrését teheti szükségessé. A GO adatbázis egységes [99,100], és webes felületen elérhető, de egyéb annotációs adatbázis (pl. MSigDB, NCBI DAVID, Huang et al. 2008; Huang et al. 2009) biztosít szűrt GO adatokat.

**Fenotípus korreláción alapuló génszett elemzések:** szakirodalmi publikációk adatbányászata segítségével, különböző betegségekkel való asszociációk vagy expressziós adatok alapján katalogizált gének. A patomechanizmus szerint tervezett útvonal elemzéstől, a hipotézis hiánya különbözteti meg. [94].

**Genomikai jellemzőn alapuló génszett elemzések:** ebbe a kategóriába tartozó elemzések a géneket közös genomikai jellemzőik mentén, pl. azonos, evolúciósan konzervált, vagy szabályozó szereppel bíró nukleotid motívumaik szerint (pl. ENCODE

promoter, exonátlépést eredményező alternatív splicingra utaló szakasz, miRNS kötőhely, stb), illetve kromoszómán való pozíciójuk szerint csoportosítják.

Adatbázisok. MSiGDB C1, C3, NCBI DAVID

**Génhálózat-alapú elemzések:** a génhálózat-alapú elemzések jellemzője, hogy nem biológiai funkciók és folyamatok hipotézis alapú modellezésével, hanem kísérleti eredményekből, vagy ezek alapján jósolt, PPI, szignál transzdukciós, valamint metabolikus génhálózati adatbázisok alapján állítanak össze és tesztelnek génszettek, vagy a modernebb topológiai módszerek a gráfok elemzésével a gének közötti kapcsolatokat is modellezik [94,96,102,110–112]. Számos PPI adatbázis elérhető, legismertebbek a STRING [98], a MINT [113] és a BioGRID [114].

**Funkcionális annotációs klaszterelemzések:** több adatbázis kombinációjakor, de már önmagában a GO elemzésekor egyre fokozódó annotációs redundancia keletkezik. Az annotációk klaszterezésével maximalizálható a génszettekkel és hálózatokkal nyerhető információ és kezelhető a korreláció [108,109]. Az NCBI DAVID egy paramétereztető, heurisztikus algoritmus segítségével, a vizsgálati génlista és háttérgének megoszlása alapján klaszterekbe rendezi a közös gének által felülreprezentált annotációkat. Az egyes klasztereket jellemző feldúsulási pont (Enrichment Score, ES) az adott klaszterbe tartozó annotációk p-értékeinek negatív logaritmikus skálán kifejezett mértani közepe. Vagyis a nominálisan szignifikáns  $\alpha=0,05$  küszöbértéknek  $ES = -\log_{10}(0,05)=1,3$  feldúsulási pont feleltethető meg.

### 1.2.3.3. A génszett- és hálózat-alapú asszociációs tesztek típusai

A génszettek csoportosító jellemzőn kívül, a génszett- és hálózat alapú tesztek kategorizálhatók az alkalmazott statisztikai elemzés alapján is. Az elemzés lehet egylépcsős, ha az SNP asszociációkat közvetlenül génszett-asszociációvá kombinálja, de többnyire kétlépcsős, vagyis a génszett-asszociáció az első lépésben számított, gén-alapú asszociációs statisztikákon alapul [94,96].

A második lépésben használt tesztek szerint a génszett-asszociáció lehet **önálló (self-contained)**, amelyek a vizsgált génszett saját asszociációját kizárólag a listázott gén statisztikák kombinációjával (pl. HYST), vagy genotípus adat birtokában, permutációval számítják. Előnyük, hogy robusztusak az egyetlen lefedettség és a génszettek közti

korrelációk torzító hatására. Hátrányuk, hogy érzékenyek a populációs felrétegződés vagy rejtett rokonság torzító hatására. [102].

A második csoportot a **kompetitív tesztek** alkotják, amelyek közös jellemzője, hogy a génszinten belüli, gén-alapú asszociációkat génszinten kívüli asszociációkhoz viszonyítják, ezáltal kevésbé érzékenyek a populációs struktúrából eredő torzításra. A kompetitív tesztek legkorábbi típusa, a génszett-alapú tesztek „első generációs” változataként is ismert, felülreprezentációs elemzés (overrepresentation analysis, ORA). Az ORA teszt a paraméterként megadott p-érték küszöb alapján szignifikáns és nem szignifikáns csoportba rendezi a géneket, majd a hipergeometrikus eloszlás segítségével meghatározza, hogy a génszinten belüli szignifikáns gének aránya, a génszinten kívüli szignifikáns gének aránya ismeretében, mekkora eséllyel lehet a véletlen műve. Az ORA teszt előnye az egyszerű kivitelezhetősége, hátránya viszont, hogy önkényesen megválasztott szignifikancia küszöböt használ [102].

A „második generációs” kompetitív tesztek, génszett funkcionális pontozásán alapulnak, ide tartozik az expressziós vizsgálatok koordinált feldúsulás-elemzéséből (Gene Set Enrichment Analysis, GSEA) derivált génszett-alapú teszt (Gene Set based Association, GSA), az SNP arány teszt (SNP ratio test, SRT). A funkcionális pontozás lényege, hogy az expresszió „fold change” mutatójához hasonlóan, a gének asszociációját súlyponttá transzformálja, és a génszett pontjainak eloszlását vagy a saját permutált eloszlásával szemben (önálló teszt), vagy a listán kívüli pontok eloszlásával szemben (kompetitív teszt) teszteli (pl. Wilcoxon előjeles rangösszeg tesztel, vagy a Kolgomorov-Smirnov tesztel). Előnye az ORA tesztekhez képest, hogy nincs önkényesen megszabott küszöb, az összes gén asszociációját figyelembe veszi, és nem feltételezi a gének azonos hatását sem [95,96,102,115].

A tesztek „harmadik generációjának” a topológiai információt is integráló, a gének egymás közötti kölcsönhatásait, a biológiai kontextus szerint gráfokkal reprezentáló hálózati elemzések is tartoznak. A gének súlypontjai az asszociáció vagy az expressziós „fold change” mellett a strukturális, sejtfunkciónak megfelelő kontextust is tartalmazzák. Előnyük a lehető legnagyobb információs komplexitás integrálása, ugyanakkor hátrányuk is pontosan ebből eredeztethető: nagyon érzékenyek az annotációs pontatlanságra. [96,110,111]

Megemlítenők még a multivariáns, regresszió alapuló tesztek, amelyek a gén-alapú asszociációs változataik szerves kiterjesztése. Ide tartozik a főkomponens regressziós elemzés és a SKAT génszett elemzésre módosított változata [96,116,117]. E módszerek előnye, hogy a populációs felrétegződésből eredő torzítást egyszerűen korrigálják, és lehetőséget adnak a gyakori és ritka variánsok együttes elemzésére is.

Összefoglalva, az eredetileg expressziós adatok feldolgozására tervezett, génszett-alapú tesztek, bizonyos megkötésekkel és megfelelő értelmezéssel, a GWAS vizsgálatokban is használhatók. A többféle annotációból kiválasztható a vizsgálat hipotéziséhez legközelebb álló, több annotációs adatbázis vagy ontológia elemzés esetén dimenzióredukció szükségessé válhat. A statisztikai módszer kiválasztásában a genomikus lefedettség és az annotációs adatok típusa és a rendelkezésre álló bemeneti adat egyaránt számít: a genotípusból számított, önálló asszociációk erősebbek, a kétlépcsős, genotípust nem igénylők viszont rugalmasabban használható több minta kombinálására. Az egyetlen genomlefedettség és a nem kezelt génszett korreláció a kompetitív tesztek torzítja, ugyanakkor az önálló asszociációs tesztekkel szemben, nem érzékenyek a populációs felrétegződésből eredő torzításra. Mivel a génszett-asszociációk eredményét az annotáció típusa, a gén- illetve génszett-alapú statisztika számítása, és a korreláció kezelése is nagymértékben befolyásolja, ezért a PGC Network and Pathway Analysis Group ajánlása [95] szerint érdemes több, különböző génszett elemzési módszer eredményeit kombinálni.

#### 1.2.4. A genomikus heritabilitást és poligénes additív hatást vizsgáló módszerek

A pszichiátriai zavarok multifaktoriális, poligénes öröklésmenetét és viszonylag magas örökölhetőségét a fenotípus szegregációján alapuló, családfa- és ikervizsgálatok már közel 50 éve valószínűsítették [118]. Azonban először csak a GWAS vizsgálatok tették lehetővé a genomikus heritabilitás számítását, és a genomikus rizikó profil meghatározását (genomic profiling, Khoury et al. 2004). A pszichiátriai GWAS vizsgálatok első hulláma [11–14] azonban nem hozta meg a várt eredményeket. A „gyakori betegség – gyakori variáns” elv alapján, az  $OR \geq 1,5$  és  $MAF \geq 0,2$  SNP asszociációk észlelésére tervezett vizsgálatok csak néhány, teljes-genom szignifikancia szintet elérő asszociációt találtak, amelyek a családfa- és ikervizsgálatokban becsült heritabilitás töredékét fejezték csak ki. A probléma később „hiányzó heritabilitás” (missing heritability) néven vált ismertté [16]. Az ellentmondásos eredmény alapvetően két okra vezethető vissza. Egyrészt a GWAS elemzések statisztikai erejét a mendeli betegségek alapján, vagyis az inkább ritka variánsokra jellemző hatáserősség alapján becsülték ( $OR: 1,5-3,0$ ), szemben a gyakori polimorfizmusokra valóban jellemző értékekkel ( $OR: 1,05-1,2$ ), azaz a valódi asszociációk döntő többségét nem tudták statisztikailag szignifikánsként azonosítani.[19]. Másfelől, a DNS chipeket a gyakori polimorfizmusok genotipizálására tervezték, ezért az általuk kifejezett genomikus heritabilitás ( $h^2$ ) – amelyet „SNP heritabilitásnak” vagy „chip heritabilitásnak” nevezünk – még optimális statisztikai erő mellett sem fejezi ki a szegregációs vizsgálatokban mérhető, teljes additív genetikai heritabilitást ( $H^2$ ).

A pszichiátriai zavarok genetikai architektúráját jellemző, poligénes additív hatás kimutatása, és prediktorként használó statisztikai eljárás, a genomikus rizikópontozás (genomic profile risk scoring, GPRS) kidolgozása Purcell és munkatársai vizsgálatához köthető [18]. A szkizofrénia és bipoláris zavar multicentrikus, GWAS mintáinak elemzésekor, a teszt statisztika inflációját mutatták ki, amely a populációs struktúra torzító hatásának korrekciója ellenére is fennállt. A genomikus infláció hátterében a gyenge ( $OR < 1,05-1,1$ ) SNP asszociációk felülreprezentációját észlelték, amely megfelelt a poligénes öröklődésnek. Az SNP p-értékek szerint genomikus profilokat definiáltak, majd egy független mintán az egyes profilokhoz tartozó, összesített genomikus rizikópontok predikcióját vizsgálták. A később GPRS-teszt néven ismert módszerrel igazolták, hogy legtöbb SNP-t tartalmazó genomikus profil ( $p_t < 0,5$ )

predikciós modellje fejezi ki a legmagasabb fenotípusos varianciát, bizonyítva a többségében nominálisan sem szignifikáns markerek poligénes hatásban betöltött szerepét.

A GPRS-teszt egyszerűsége és flexibilitása miatt széles körben elterjedt, érvényességét és a genomikus heritabilitással való kapcsolatát későbbi vizsgálatok [120] is igazolták. Mindazonáltal, diszkrét fenotípusok (pl. diagnózisok) esetén, a GPRS által számított kifejezett fenotípusos variancia jelentősen eltér a betegségre való fogékonyság genomikus heritabilitásától. Az eltérés oka a regressziós modellek mutatóiban keresendő: míg a lineáris regresszió determinációs együtthatója, az  $R^2$  valóban a modell által kifejezett információt jelenti, a logisztikus regresszió esetében nincs valódi információs mutató. A jobb híján használt „pszeudo $R^2$ ” mutató, pl. a Nagelkerke  $R^2$ , közös jellemzője, hogy csak a mért adatban értelmezhető, relatív információt adják meg, vagyis két különböző vizsgálat Nagelkerke  $R^2$  értékét nem lehet érvényesen összehasonlítani. Ezért a genomikus heritabilitás mérésére, a lineáris kevert modellek közé tartozó, genetikus kapcsolati mátrixra korlátozott maximális valószínűség (genetic-relationship-matrix restricted maximal likelihood, GREML), illetve az LD pont regresszió alapuló módszerek kerültek előtérbe [17,19]. Mindhárom poligénes elemző módszerben közös, hogy a genetikai variánsok egyedi asszociációi helyett, azokat valamely módon összesítő, mennyiségi változókat definiálnak és tesztelnek. A GPRS teszt során a poligénes rizikópontok, a GREML esetén a genotípusból számított genetikai távolság, az LD pont regresszió esetén pedig az LD pontok és a fenotípus regressziója képezi a teszt alapját. Az alábbiakban röviden ismertetjük a tesztek sajátosságait és alkalmazhatóságukat.

#### 1.2.4.1. Genomikus profil rizikóponozás (GPRS)

A GPRS-teszt lényege, hogy amennyiben a vizsgált fenotípus poligénes öröklése feltételezhető, akkor egy adott célminta (target sample) genomjaiban egy adott genomikus profilba tartozó allélok súlyozott összegzésével kiszámítható a genomikus profil rizikópontja (GPRS), amellyel a fenotípus predikciója regressziós modellben tesztelhető. A torzítás elkerülése érdekében a független rizikóallélokat és súlyként használt, genetikai hatásukat, a célmintával nem átfedő fejlesztőmintán (training sample) határozzuk meg. Első lépésként a fejlesztőmintán kiszűrjük az információt duplikáló,

kapcsolt ( $r^2 > 0,2$  500 kb régióban) markereket, majd az allélikus  $p$ -értékek alapján genomikus profilokba rendezzük (pl.  $p_t < 0,5$ ,  $p_t < 0,4$ ,  $p_t < 0,3$ ,  $p_t < 0,2$ ,  $p_t < 0,1$ ,  $p_t < 0,05$ ), és megállapítjuk a markerek rizikóalléljaihoz tartozó súlyok értékét. A súly leggyakrabban regressziós együttható, vagy súlyozatlan analízis esetén, az asszociáció irányát kifejező  $+1$  és  $-1$  értékeivel definiált. A második lépcsőben a célmintán kiszámítjuk a genomokhoz tartozó GPRS pontot, a vizsgált genomikus profilhoz tartozó rizikóallélok súlyozott számának összegzésével vagy átlagával. Végezetül a kimeneti fenotípus függvényében, lineáris vagy logisztikus regresszió prediktoraként teszteljük a GPRS pontokat, a szignifikáns predikció bizonyítja a vizsgált markerek által közvetített poligénes hatást. A modell determinációs együtthatója ( $R^2$  vagy Nagelkerke pszeudo  $R^2$ ) vagy egyéb illeszkedési paraméterek segítségével pedig kiválasztható a legnagyobb fenotípusos varianciát kifejező genomikus profil.

A GPRS-teszt számos előnnyel bír: egyszerűen és gyorsan kiszámolható, valamint a fejlesztő- és célminta fenotípusa, illetve a vizsgált markerek alapján többféle típusú elemzésre is alkalmazható. A fejlesztő- és célminta megegyező fenotípusa esetén a vizsgált markerek poligénes hatása validálható, különböző fenotípusokkal pedig a transzdiagnosztikus genetikai átfedés tesztelhető. GPRS predikcióval egyszerűen vizsgálható különböző pszichiátriai diagnózisok (pl. szkizofrénia-autizmus, MDD-bipoláris zavar), diagnózisok és altípusok (pl. szkizofrénia-bipoláris spektrum zavarok), vagy (neuro)pszichológiai jellemzők és diagnózisok (neuroticizmus-MDD, ADHD-figyelemzavar általános populáción) közötti etiológiai kapcsolat.

A módszer hátránya, hogy legalább két, nem átfedő mintát igényel. Emellett fenotípusként diagnózisokat elemző modellek esetén, a modell determináns együtthatójával kifejezett fenotípusos variancia, az ún. megfigyelhető heritabilitást (observed heritability) jelenti. A megfigyelhető heritabilitás értéke függ a minta esetkontroll arányától és a populációban mért prevalenciától, vagyis még azonos fenotípus esetén is csak korlátozottan vethető össze más minták eredményeivel. Továbbá, a jel-duplikáció elkerülése érdekében, kiszűrésre kerülnek már a mérsékelten kapcsolt markerek is, tehát elkerülhetetlenül is információvesztés lép fel. Ez utóbbi problémára megoldást jelenthet Vilhjálmsson és mtsai módszere, az LDpred, amely kapcsolt markerek kiszűrése helyett a súlyokat módosítja a kapcsoltsági szerkezetnek megfelelően [70]. A módszerük lényege, hogy a fejlesztőminta kapcsoltsági szerkezetének és a GPRS



$p_t$  értékeihez hasonló, ún. kauzális markerek arányának ismeretében, kiszámítja a markerek hatáserősségéhez tartozó poszterior PRS értéket. A jelduplikálás elkerülésére nem ritkítja a markereket, hanem az minta markerszámából számított LD-rádiusz ( $N_{\text{SNP}}/3000$ ) érték alapján súlyozza, tehát valamilyen mértékben minden SNP részt vesz a tesztben. A módszer további előnye, hogy a konkrét predikció előtt kiválasztható a legjobb kalibrációt biztosító kauzális frakció, tehát nem kell az összes tesztet elvégezni, vagyis minimalizálható a többszörös összehasonlítás korrekciója. Hasonlóan az LD-pont regresszióhoz, ez a módszer is elsősorban nagy elemszámú mintákon előnyösebb a többi poligén teszténél.

#### 1.2.4.2. Genetikai-kapcsolati-mátrixra korlátozott maximális valószínűségeen (GREML) alapuló tesztek

A lineáris kevert modellek közé tartozó GREML tesztek lényege, hogy a minta alanyai között fennálló fenotípusos hasonlóság mennyire jósolja be az SNP eloszlásból számított genetikai hasonlóságukat, tehát közvetlenül mérik a genomikus heritabilitást. Első lépésként a teszt az összes SNP együttes illesztésével meghatározza a genomok közötti páronkénti genetikai távolságot, létrehozva a minta genetikai kapcsolati mátrixát. Ezt követően a genetikai kapcsolati mátrix alapján, korlátozott maximális valószínűség (REML) analízissel, random hatásként számítható az SNP variancia által kifejezett fenotípusos variancia, a genomikus heritabilitás. Eset-kontroll vizsgálatok elemzésekor, a diagnózis kategórikus fenotípusa helyett, a prevalencia alapján számított, zavarra való fogékonyság (liability) mennyiségi fenotípusa segítségével számolható a genomikus heritabilitás.

A GREML analízis előnye, hogy GPRS-teszttel szemben egy mintát is képes elemezni és robusztus a kapcsoltságra, a genomikus heritabilitás torzítatlan becslését végzi, valamint képes a genom egy részére particionált heritabilitás számítására is. Kétféle változós modellje két minta elemzésével, azonos fenotípusok esetén az SNP korreláció, különböző fenotípusok esetén pedig a koheritabilitás becslésére is használható.

A GREML analízis hátránya, hogy erőforrásigényes, és viszonylag magas elemszámot igényel a heritabilitás pontos becsléséhez ( $N > 3000$ ). Emellett a populációs

dizájn miatt érzékeny a rejtett rokonság, illetve a populációs stratifikáció torzító hatására, továbbá a genotípezési hibákra, ezért a minőség-ellenőrzés különösen fontos.

#### 1.2.4.3. Kapcsoltsági egyensúlytalanság pontozásán alapuló regresszió (LD pont regresszió)

A poligénes statisztikai módszerek harmadik típusát a kapcsoltsági egyensúlytalanság pontozáson alapuló regresszió (LD-pont regresszió) képviseli. A módszer alapja, hogy minél több más variánsal kapcsolt egy nukleotid, annál nagyobb eséllyel van a kapcsolt variánsok között kauzális polimorfizmus is, tehát a magasabb LD-ponttal bíró SNP erősebb asszociációval bír, míg a zavaró tényezőkből eredő, fals pozitív asszociációk nem járnak magas LD ponttal. Az SNP asszociációs teszt statisztika ( $X^2$ ) az SNP kapcsoltság alapján számított LD-ponttal szembeni regressziós együtthatója ( $\beta$ ), vagyis a függvény meredeksége a genomikus heritabilitás, az intercept pedig a zavaró tényezők torzító hatását mutatja (pl. populációs stratifikáció, rejtett rokonság). Az LD-pont regresszióval tehát egyszerre lehet mérni a genomikus heritabilitást és a populációs stratifikáció hatását.

Az LD-pont regresszió előnye, hogy genotípus hiányában is használható eredményfájlok elemzésére, amennyiben a vizsgált populációnak megfelelő referenciagenom (pl. 1000 Genomes Project) is a rendelkezésünkre áll az LD-pontok számításához. Hasonlóan a GREML és GPRS módszerekhez, különböző fenotípusokat elemezve, kiszámítható a két fenotípus közti genetikai korreláció is, ráadásul tesztelés közben a populációs stratifikáció hatása is egyszerűen korrigálható. Hátránya, hogy még a GREML elemzésekénél is nagyobb mintaszámot igényel ( $N > 5000$ ) a poligénes hatás pontos észleléséhez.

Összefoglalva, a poligénes additív hatás és a genomikus heritabilitás statisztikai vizsgálatára alkalmas módszerek az allélok összeadó hatását és az általuk kifejezett mennyiségi jelleget elemzik, amely jól illeszkedik a pszichiátriai zavarok öröklésmenetéhez. E módszerekkel könnyen és gyorsan becsülhető a vizsgált fenotípus genetikai meghatározottsága, illetve lehetővé teszik különböző fenotípusok közös genetikai rizikójának felmérését, emiatt jól használhatók a pszichiátriai zavarok transzdiagnosztikai vizsgálatában. Hátrányuk, hogy viszonylag nagy mintát igényelnek,

és - a partícionált elemzéseken kívül - nem informatívak a genetikai rizikó genomikus eloszlására, az egyes gének, régiók, útvonalak szerepére. A közös jellemzőkön túlmenően, a tesztek módszerspecifikus előnyei és hátrányai figyelembevételével megválaszthatók a vizsgálat mintáinak és hipotézisének leginkább megfelelő teszt. A GPRS alacsonyabb elemszámot, viszont legalább két mintát igényel, amely ideálissá teszi a komparatív vagy transzdiagnosztikai vizsgálatokhoz. Az LD-pont regresszió multicentrikus vizsgálatok nagy elemszámú, de nagymértékben heterogén mintái elemzéséhez különösen előnyös, mivel a populációs felrétegződést a poligénes hatás torzítása nélkül is képes kontrollálni. A GREML/REML elemzések pedig jól használhatók genomikus régió szerint partícionált heritabilitás vizsgálatában, akár nagyfokú kapcsoltságban álló régiók esetén is.

### 1.3. A szkizofrénia és a szuicid magatartás teljes-genom vizsgálatai

#### 1.3.1. Teljes-genom asszociációs és multilokus vizsgálatok szkizofréniában

A szkizofrénia egy viszonylag gyakori, komplex neuropszichiátriai zavar, amelyet fiatal felnőttkorban induló, és többnyire visszatérő, pszichotikus epizódok jellemeznek, a pszichés funkciók és az életminőség jelentős, gyakran progresszív károsodását vonva maguk után. Változatos tüneti képe ellenére prevalenciája világszerte 0,5-1,2% között mérhető, amely döntően kulturális hatásoktól független, biológiai eredetre utal. Ezzel egybecsengtek a szegregációs vizsgálatok közel 50 éve ismert eredményei [118], amelyek a szkizofrénia magas heritabilitását ( $H^2 \sim 0,8$ ) és a pszichiátriai zavarokra általánosan jellemző, poligénes jellegét valószínűsítették.

Erős genetikai meghatározottsága és népegészségügyi jelentősége miatt a szkizofrénia a pszichiátriai genetikai vizsgálatok egyik legintenzívebben kutatott fenotípusává vált. Már a GWAS vizsgálatok részletes bemutatása is meghaladná jelen értekezés kereteit, ezért az alábbiakban csak a legkorszerűbb és legszélesebb konszenzust élvező eredményeket foglaljuk össze.

##### 1.3.1.1. Gyakori SNP és ritka SNV/indel teljes-genom asszociációk szkizofréniában

A kezdeti, alultervezett elemszámú GWAS vizsgálatokat követően, a Schizophrenia Working Group of the Psychiatric Genomics Consortium korábbi vizsgálatokat is inkorporáló szkizofrénia mega-analízise ( $N_{SCZ} = 36.989$ ,  $N_{KONT} = 113.075$ ) rendelkezett megfelelő statisztikai erővel a kis hatáserősségű ( $OR < 1,2$ ), gyakori SNP asszociációk megbízható észleléséhez [121]. A számos kutatócsoport mintáiból összeállított, nemzetközi projekt 128 független SNP asszociációt észlelt genomikus szignifikanciával ( $p < 5E-08$ ). Az SNP-k 108, javarészt fehérjekódoló génen helyezkedtek el, de csak 10, exonikus SNP kódolt aminosavcserét. Mindazonáltal, a nem kódoló SNP-k pozíciói is ideg- vagy immunsejtek eQTL régióinak szignifikáns feldúsulását mutatták. Az azonosított lokuszok és kapcsolt markerek replikálták 25, az NHGRI GWAS catalog adatbázisában szkizofréniához kapcsolt SNP asszociációját, a fennmaradó 83 lokusz közül pedig számos a szkizofrénia feltételezett etiológiájához kapcsolható. Elsősorban megemlítendő az antipszichotikumok legfontosabb

célmolekulája, a dopamin 2 receptor gén (DRD2), vagy a szintén lehetséges terápiás célmolekulaként vizsgált, metabotróp glutamát receptorok génjei (GRM2, GRM3). A 83 azonosított lókuszt közé tartoznak például glutamát-jelátvitelhez köthető gének (GRIN2a, GRIA1, SRR, CLCN3, SLC38A7), neuronális kalcium-jelátvitelhez kapcsolt gének (CACNA1I, RIMS1), szinaptikus plaszticitáshoz köthető gének (KCNT1, NLGN4X, PTN, IGSF9B, CNTN4, MEF2C, PAK6, SNAP91), nikotinos acetil-kolin receptorgének (CHRNA3, CHRNA5, CHRNB4), idegfejlődésben szerepet játszó gének (FXR1, SATB2, PODXL, BCL11B, FAM5B, TLE1, TLE3) [121].

A ritka egyedi nukleotid illetve indel variánsok ( $MAF < 0,001$ ) egyedi asszociációi még a szkizofrénia eddigi legnagyobb elemszámú teljes-exom szekvenálásban ( $N_{SCZ} = 2536$ ,  $N_{KONT} = 2543$  Purcell et al. 2014) sem bizonyultak szignifikánsnak. Ugyanakkor, terhelés (burden) analízisük 2546 génben, köztük az idegi aktivitás-szabályozott sejtvezetékhez kapcsolt fehérjekomplex (activity-regulated cytoskeleton-associated protein complex, ARC) csoportban, a fragilis X mentális retardáció fehérje (fragile X mental retardation protein, FMRP) RNS-kötése által szabályozott gének csoportjában és az N-metil-D-aszpartát receptor (NMDAR) rendszer génjeiben szignifikáns SNV/indel terhelést/felülreprezentáltságot mutatott ki [123].

### 1.3.1.2. Multilókuszt teljes-genom asszociációs vizsgálatok szkizofréniaiban

Az SNP/SNV asszociációk felderítése mellett számos teljes-genom asszociációs és expressziós vizsgálat végzett génszett-elemzéseket, vagy poligén predikciós tesztet szkizofréniaiban. A génszett-asszociációs elemzések a szinaptikus plaszticitásban szerepet játszó neuronális kalcium-jelátvitel, glutamát-transzmisszió, miR-137 célgének, az ARC és FMRP-hez kapcsolt gének, az asztrocita és oligodendroglia transzkripcióhoz, sejtciklus-szabályozáshoz, kromatin-átrendeződéshez, valamint idegfejlődéshez, axonnövekedéshez, neuronális migrációhoz és immunfunkciókhoz köthető gének feldúsulását illetve asszociációját észlelték [121,124,125]. Az elemszám növelése a poligén predikció hatékonyságát is megnövelte, debütálásakor a GPRS 0,032 Nagelkerke  $R^2$  értéket mért szkizofréniaiban [18], ezzel szemben a mega-analízis 0,18 Nagelkerke  $R^2$  értékével [121]. A transzdiagnosztikus elemzések jelentős átfedést mutattak a szkizofrénia és a bipoláris zavar gyakori SNP markerei által kifejezett

fenotípusos variancia között, és kevesebbet a szkizofrénia és depresszió, illetve szkizofrénia és ADHD között. Az autizmus spektrum zavarral (ASD) érdekes módon egyedül a szkizofrénia mutatott, jöllehet nem túl jelentős poligénes átfedést a vizsgált 5 pszichiátriai kórkép közül [17,19,126].

### 1.3.2. Teljes-genom asszociációs vizsgálatok szuicid magatartásban

Az öngyilkosság a pszichiátriai zavarok lehetséges legpusztítóbb kimenetele. A WHO 2000-2012-es kiadású Global Health Estimates jelentésében ez a második vezető halálok a 15-29 éves korosztályban, és az első 20 halálok között szerepel világszerte. Az öngyilkosság és a mentális zavarok között szoros kapcsolat áll fenn, az öngyilkossági kísérletek körülbelül 90%-a pszichiátriai betegségben szenvedők körében fordul elő.

A WHO World Mental Health Survey Initiative című, Nock és munkatársai [127] által készített átfogó tanulmánya feltárta, hogy a hangulatzavarok különösen erős, független előrejelzői mind az öngyilkossági kísérletnek (OR: 4,7-5,9), mind az öngyilkossági gondolatoknak (szuicid ideáció) (OR: 3,4-4,7), még akkor is, ha a szociodemográfiai változókat figyelembe vették. Bár az öngyilkossági gondolatok és tervek gyakoribbak, mint a kísérletek (egész élettartam alatti előfordulása 9,2%, a kísérletek 2,7%-ával szemben), azt találták, hogy a válaszadók közül az öngyilkossági gondolatokkal foglalkozók 29%-a és a határozott öngyilkossági terveket szövőők 59%-a meg is kísérelte az öngyilkosságot. Ezen kívül a szerzők leírták, hogy a pszichiátriai betegségek és az öngyilkossági kísérletek közötti erős korreláció eltűnt, miután az öngyilkossági gondolatokra kontrolláltak regressziós modellekben. Ezek az eredmények azt is jelzik, hogy önmagukban a pszichiátriai betegségeken kívül más faktorok is közvetíthetik az átmenetet az öngyilkossági gondolatoktól a kísérletig.

Más pszichiátriai betegségekhez hasonlóan, a szuicid viselkedésnek is feltételezhetően komplex, soktényezős etiológiája van. Bár genetikai epidemiológiai tanulmányok [128–130] arra utalnak, hogy a genetikai faktorok bizonyosan szerepet játszanak a szuicid magatartásban, az öngyilkosság genetikai háttere még tisztázásra vár.

A központi idegrendszer szerotonerg deficitje és az öngyilkosság között észlelt összefüggésekre építő, korai, kandidáns-gén tanulmányok a szerotoninerg rendszert vették célba, különösen a TPH2 és a SLC6A4 géneket [131]. A későbbi kandidáns-gén,

GWAS vizsgálatok és szisztematikus metaanalízisek [132]azonban nem erősítették meg a kezdeti asszociációkat.

A hipotézis-vezérelt kutatások fókuszja fokozatosan eltolódott más neurokémiai rendszerek génjeire, mint a dopaminerg transzmisszió (DRD2, COMT, MAOA) [133], a poliaminok (SAT1) [134], stressz-válasz (FKBP5, HPA tengely gének) [133,135], az elmúlt években pedig a neurotrofinok, főleg a BDNF-NTRK2-CREB útvonal génjeire (Dwivedi 2010; Dwivedi et al. 2003; Dwivedi et al. 2009; Kohli et al. 2010; Kim et al. 2008; Juhasz et al. 2011; Ernst et al. 2009). Ezen kívül, az öngyilkossági gondolatok és kísérletek során tapasztalható, ezen állapotokhoz tartozó, a végrehajtó funkciók állapot jellegű károsodása miatt a génexpressziós tanulmányokat végeztek azokon a területeken, amelyek felelősek ezekért a kognitív doménekért, például a prefrontális kortexben. Ezek a vizsgálatok a microRNS (miRNS) rendszer lehetséges szerepét vetették fel az öngyilkos viselkedésben [144,145].

Különösen a hsa-miR-491-3p, hsa-miR-185, hsa-miR-195 és hsa-mir30a szerepét kell kiemelnünk, mivel a hsa-miR-491-3p és hsa-miR-185 emelkedett kifejeződését találták öngyilkosság áldozatainak prefrontális kortexében [144], ezen kívül a hsa-miR-185, hsa-miR-195 és hsa-mir30a fontos szabályozó szerepet játszik a BDNF szignál útvonalban (Serafini et al. 2014; Maussion et al. 2012) [144,146].

Sajnos a GWAS tanulmányok [147–155], amelyek az öngyilkos viselkedést kutatták, napjainkig nem tisztázták a fent említett összetettséget. Az egyetlen ACP1 gén közeli intergenikus SNP kivételével, melyet Willour és munkatársai írtak le [155], a tanulmányok legjobb esetben is csak szuggesztív, tendencia értékű eredményeket hoztak. Továbbá ezen asszociációk génjeinek legtöbbszörét nem sikerült replikálni két független GWAS elemzésben. A szignifikáns asszociációk és replikációk hiánya legnagyobb valószínűséggel elsősorban a nem megfelelő statisztikai erőnek [147,148,153], és a fenotípusos heterogenitásnak köszönhető.

Három tanulmány [147,148,153] vizsgálta az öngyilkossági gondolatok megjelenését vagy rosszabbodását antidepresszáns kezelést követően, négy tanulmány [149,150,154,155] az öngyilkossági kísérletet, mint kimeneti változót használta, míg egy tanulmány, Galfalvy és munkatársai tollából, befejezett öngyilkosság áldozatainak mintáit elemezte [152]. Továbbá a fenti okokon kívül a genetikai heterogenitás és a feltételezett poligenikus öröklődés is hozzájárulhatott ahhoz, hogy az eredmények

megismétlése sikertelen maradt. Mint Ayalew és munkatársai leírták, a heterogenitás különösen az SNP szinten erős, de az adatbázisok közötti átfedést jelentősen meg lehetne növelni, ha összetettebb genetikai „markereket” is figyelembe vennénk, mint például gén-régiók, kanonikus útvonalak vagy génhálózatok [38]. A szuicid magatartás teljes-genom asszociációs és expressziós vizsgálataiban szuggesztív p-értékű SNP-t tartalmazó vagy szuggesztív expressziós p-értékű géneket az **1. táblázat** mutatja be.



**1. táblázat.** A szuicid magatartás kandidáns génjei a teljes-genom asszociációs és expressziós vizsgálatok alapján.

<b>Gén</b>	<b>Vizsgálat (szerző, név)</b>	<b>Biológiai szerep</b>
ACP1	Willour et al. 2012	A kódolt fehérje egy tirozin foszfatáz, amely gátolja a neurogenézist a sejt plazma $\beta$ -catenin koncentrációja csökkentésével.
ANK3	Perroud et al. 2012	A sejt hártya és a citoskeleton integráns fehérje, szerepe van az idegfejlődésben, axon növekedésben, szignál transzdukcióban és akciós potenciálban.
BDNF	Dwivedi et al. 2003 Kim et al. 2008	Neurotróf faktor, alapvető fontosságú a kérgi és striatális idegsejtek túlélésében, stressre adott válaszban és hangulat zavarokban implicál
CACNA1C	Ferreira et al. 2008 Lett et al. 2011	Feszültségfüggő kalcium csatorna, az axon növekedéshez, szinaptikus transzmisszióhoz, és kalciumfüggő jelátvitelhez kapcsolható
CD44	Galfalvy et al. 2011	Sejtfelszíni glikoproteint kódol, amely sejtheadhézióban, limfocita aktivációban és immunszabályozásban is szerepet játszik
CYTH3 (GRP1)	Perroud et al. 2012	A Golgi szerv szerkezetének és funkciójának szabályozásához, valamint a fehérjecsomagoláshoz és -transzportoz is kapcsolható

<b>Gén</b>	<b>Vizsgálat (szerző, név)</b>	<b>Biológiai szerep</b>
DICER1	Smalheiser et al. 2012	A miRNS turnover egyik legfontosabb enzime, onkogén szerepe mellett az idegfejlődéshez, végrehajtó funkciókhoz, munkamemóriához, stresszre adott válaszhoz kapcsolható.
DSC2	Galfalvy et al. 2011	Cadherin szupercsaládba tartozó membránfehérje, sejtadhézióban, dezmoszóma képződésben játszik szerepet.
FKBP5	Roy et al 2012 Schosser et al. 2011 Menke et al. 2012	Fontos szereppel bír a fehérjetranszportban, -csomagolásban, immunszabályozásban, stresszre adott válaszhoz kapcsolható. Az intracelluláris glukokortikoid választ modulálja.
FOXN3	Galfalvy et al 2011	Mitózis ellenőrző pontokon szupresszorként hat, fontos szerepet játszik a DNS javításban
GDA	Perroud et al. 2012 Menke et al 2012	Guanin-deamináz enzimet kódolja, az idegrendszer fejlődéséhez köthető
GFRA1	Schosser et al 2011	Glia sejt vonal eredetű neurotróf faktor, szerepe van a neuronok túlélésében, differenciálódásában és fejlődésében
GRIA3	Laje et al. 2009; Laje et al. 2007	AMPA3 glutamát receptor, neuropszichiátriai zavarok széles skálájához kapcsolható
GRIK2	Laje et al 2007 Laje et al 2009	Kainát2 glutamát receptor, neuropszichiátriai zavarok széles skálájához kapcsolható

<b>Gén</b>	<b>Vizsgálat (szerző, név)</b>	<b>Biológiai szerep</b>
hsa-miR-185	Maussion et al. 2012	A tropomiozin receptor kináz B enzim expresszióját szabályozza, idegfejlődésben, neuron képződésben és szinaptikus plaszticitásban implikált
hsa-miR-195	Mellios et al. 2008	A BDNF és HTR2A gének expresszióját szabályozza, idegfejlődésben és szinaptikus plaszticitásban is implikált
hsa-miR-30a	Mellios et al 2008	BDNF expresszióját szabályozza, idegfejlődésben és szinaptikus plaszticitásban is implikált
hsa-miR-491	Maussion et al. 2012	Transzkripció szabályozó, szuicid magatartásban szignifikánsan megnövekedett expresszióját észlelték
IFNLR1	Laje et al 2009	Interleukin 10B receptor, amely az IL-28A, IL-28B és IL-29 interleukinokkal is kölcsönhatásban van, az immunszabályozásban játszik szerepet
KIAA1244	Schosser et al 2011	A transzGolgi hálózat és sejtmembrán komponens, funkciója a vezikuláris transzportoz kapcsolódik
MARCH1	Galfalvy et al 2011	Membrán ubikvitin E3 ligáz enzim, fontos szereppel bír az immunválaszban és az MHC fehérje kötésben
MBNL2	Galfalvy et al 2011	Izomkapcsolt splicing regulátor 2 gén az alternatív splicing szabályozásával az agyfejlődésben és myotoniás izomdisztrófiában implikált

<b>Gén</b>	<b>Vizsgálat (szerző, név)</b>	<b>Biológiai szerep</b>
NTRK2	Perroud et al 2012 Schosser et al 2011 Kohli et al 2010	Neurotróf tirozin kináz 2, a BDNF nagy affinitású receptora, hangulatzavarokhoz és szuicid magatartáshoz asszociált
NXPH1	Perroud et al 2012	Szoros komplexet képez a neurexinekkel, meghatározó szerepű a dendrit-axon adhézióban és szinapszis képzésben
PAPLN	Laje et al 2009	Metallopeptidáz, amely az SSRI kezeléshez kapcsolt szuicid ideációhoz asszociált
PRKCE	Schosser et al 2011 Dwivedi et al 2003	Szignál transzdukciós fehérje, amelyet a DAG és kalcium aktivál, szerepe van a neuroncsatornák aktiválásában és a stresszre adott válaszban
RGS18	Schosser et al 2011	GTPáz aktivátor, G-protein receptorok aktiválásában van szerepe
RGS2	Schosser et al 2011	GTPáz aktivátor, G-protein receptorok szabályozásában és a szorongás kialakulásában van szerepe
SAT1	Fiori et al. 2010	Poliaminok lebontási sebességét meghatározó enzimet kódoló gén
SLC4A4	Schosser et al 2011	Nátrium bikarbonát kotranszporter, az intracelluláris pH szabályozásában vesz részt
SPTLC1	Galfalvy et al 2011	A szfingolipid bioszintézis egyik kulcsfontosságú enzimét kódolja
SRSF11	Galfalvy et al 2011	Splicing faktor, mRNS szabályozásában implikált

<b>Gén</b>	<b>Vizsgálat (szerző, név)</b>	<b>Biológiai szerep</b>
TMEM132C	Willour et al. 2012.	Transzmembrán protein foszfatáz fehérje, amely a szuicid magatartáshoz asszociált
TUBGCP3	Galfalvy et al 2011	gamma tubulin asszociált proteint kódol, fontos szereppel bír a mitotikus G2/M fázis átmenetben
ZNF804A	Lett et al 2011	Zink-ujj fehérje, transzkripció szabályozó, implikált, szkizofréniában, bipoláris zavarban és öngyilkossági kísérletben

## 2. Célkitűzések

Vizsgálataink általános célkitűzése volt, hogy az általánosan használt statisztikai módszerek helyett, a vizsgált markerek együttes hatását elemző, multilokusz asszociációs modellekkel térképezzük fel a pszichiátriai zavarok genetikai szerkezetét. A funkcionális annotáció alapján meghatározott, multilokusz asszociációs modellekkel a biológiai információt is integráló, komplex genetikai struktúrák (gén, útvonál, funkcionális génklaszter, stb) asszociációját vizsgáltuk különböző pszichiátriai fenotípusokkal, illetve replikációjuk valószínűségét elemeztük azonos fenotípusú minták esetén. Első vizsgálatunkban öngyilkossági viselkedés transzdiagnosztikai jellegű, régió-alapú asszociációs, és poligénes predikciós (GPRS), kandidáns-gén elemzését végeztük, DSM-IV major depresszióval, illetve DSM-IV bipoláris zavarral diagnosztizált betegek teljes-genom mintáin. Második vizsgálatunkban pedig különböző komplexitású, multilokusz asszociációs módszerekkel a DSM-IV szkizofrénia genetikájának hipotézismentes feltárását, továbbá a replikációs valószínűségre gyakorolt hatásuk felmérését tűztük ki célul, ezúttal két azonos fenotípusú, de egymástól jelentősen eltérő mintát használva.

### 2.1. Az első vizsgálat célkitűzései és hipotézisei

Első vizsgálatunkban az öngyilkossági viselkedés transzdiagnosztikus, kandidáns-gén elemzését végeztük, két mintánk a National Center of Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP) kutatható adatbázisából DSM-IV major depresszív zavarral diagnosztizált betegek, valamint DSM-IV bipoláris zavarral diagnosztizált betegek teljes-genom asszociációs vizsgálatából származott. Ahogyan a bevezetőben részletesen kifejtettük, a pszichiátriai zavarokra való hajlam mennyiségi jellegként öröklődik, tehát a génhatást sok allél összeadó, poligénes hatása hozza létre, az egyes allélok saját hatása gyenge. Ennek megfelelően, az öngyilkossági viselkedés SNP alapú, teljes-genom asszociációs vizsgálatai nem találtak teljes-genom szignifikanciával bíró nukleotid asszociációt. Ugyanakkor, a poligénes modell érvényessége ellenére, viszonylag kevés multilokusz vizsgálati adat állt rendelkezésre. Elsődleges célunk ezért a szuicid magatartás genetikai faktorainak multilokusz asszociációs elemzése volt, neuropszichológiai, génexpressziós és SNP asszociációs vizsgálatok priori ismereteit integráló, kandidáns génszettekkel.

Feltételezésünk szerint, a nukleotidok funkcionális információt integráló prioritizálása, és csoportosított elemzése – az allélikus asszociációk várható feldúsulása és kombinációjukból eredő felerősödése miatt - azonosíthat az SNP alapú tesztekben a zajtól nem elkülöníthető asszociációkat. Hipotézisünk érvényességét különböző megközelítésű elemzésekkel és kandidáns génszettekkel is teszteltük a két GWAS mintán. Az első elemzésben vizsgált génszettet irodalomkutatás segítségével állítottuk össze, a teljes-genom asszociációs és expressziós vizsgálatok publikációiban legjobbnak bizonyult SNP asszociációk alapján. A korrelációs génszettet a második elemzésben kiegészítettük további két kandidáns génlistával, ezáltal a géneket szuicid viselkedésben implikált biológiai funkciók szerint választottuk. A három kandidáns génszett nem kapcsolt markereiből poligénes rizikópontot számítottunk a depressziós GWAS mintán, és predikciós erejét teszteltük a bipoláris mintán. Fent említett hipotézisünk mellett vizsgálatunk mindkét elemzése egyúttal a bipoláris zavar és a depresszió közös genetikai faktorait is elemezte, a szuicid magatartás transzdiagnosztikai fenotípusán keresztül.

## **2.2. A második vizsgálat célkitűzése és hipotézisei**

Második vizsgálatunkban a multilokus asszociációs tesztek a DSM-IV szkizofrénia genetikai hátterének feltérképezésére használtuk, alapvetően két cél elérésére törekedve. Vizsgálatunk elsődleges célja a szkizofrénia genetikai rizikófaktorainak hipotézismentes explorációja volt, a komplex fenotípusok poligénes öröklődéséhez legjobban illeszkedő, és biológiai funkciót is integráló, multilokus statisztikai módszerekkel. Ennek érdekében gén-, transzkript-, kanonikus útvonal- és pozicionális génszett-alapú asszociációját, valamint kanonikus útvonalak, pozicionális génszettek és prioritizált génekhez kapcsolt annotációs klaszterek feldúsulását teszteltük. Statisztikai elemzésünket két lépcsős szerkezetben, egy hipotézisgeneráló explorációs, és egy hipotézisvizsgáló replikációs lépcsőben végeztük. Az explorációs lépésben a SCHIZOBANK projekt [159] eset-trió mintáján lefuttattuk a teljes-exom elemzéseket, majd a legalább szuggesztív p-értékkel generált hipotéziseket a két szkizofrénia GWAS adatbázisból összevont replikációs mintán teszteltük. A vizsgálatunk kétlépcsős szerkezetéhez kapcsolódik másodlagos célkitűzésünk, a multilokus elemzési módszerek replikációs valószínűsége gyakorolt hatásának felmérése eltérő szerkezetű minták

között. A funkcionális információt is integráló, gén- és útvonal-alapú GWAS elemzések SNP-alapú tesztekénél jóval magasabb arányú replikációs valószínűsége ismert tény [38], de egymástól eltérő szerkezetű minták közötti replikáció arányáról nincs szakirodalmi adat. E hiányt pótoltuk a SCHIZOBANK kis elemszámú, exomszekvenált, családi mintáján generált hipotéziseinek nagy elemszámú, imputált populációs GWAS mintán tesztelésével, és a növekvő komplexitású, multilokusz tesztek replikációs arányának összehasonlításával.



### 3. Módszerek

#### 3.1. Az első vizsgálat módszerei

##### 3.1.1 Minta, fenotípus

Az első vizsgálatunk mintái az National Center for Biotechnology Information (NCBI) és a Genetic Association and Information Network (GAIN) által fenntartott database of Genotypes and Phenotypes (dbGaP) adatbázisából származnak, a vizsgálat a munkacsoportunk a GAIN és a Joint Addiction, Aging, and Mental Health (JAAMH) Data Access Committee (DAC) által elfogadott, „Project#2862: Endophenotypes in affective and psychotic disorder spectra: a cross-disorder analysis” kutatási projektje részeként valósult meg. Vizsgálatunkban két teljes-genom asszociációs vizsgálati mintát elemeztünk: a GAIN Whole Genome Association Study of Bipolar Disorder adatbázisának európai-amerikai származású bipoláris zavar mintáját ( $n_{BIP} = 1044$ ,  $n_{KONT} = 1068$ , elérési azonosító: phs000017.v3.p1; [160–162], valamint a Major Depression: Stage 1 Genomewide Association in Population-Based Samples európai származású, major depresszív zavar (MDD) mintáját ( $n_{MDD} = 1821$ ,  $n_{KONT} = 1822$ , elérési azonosító: phs000020.v2.p1; [163,164]. A teljes statisztikai elemzést mindkét mintában az esetcsoportra korlátoztuk, de az imputációt, és a nominálisan szignifikáns asszociációk specificitásának validálását a teljes mintán (eset + kontroll) végeztük el. A két vizsgálatunk mintáit a **2. táblázat** mutatja be.

A bipoláris zavar mintáját a Bipolar Consortium kapcsoltsági vizsgálatainak 1-4. hullámában gyűjtött családi mintáiból, az 5. hullám populációs mintájának a DSM-IV bipoláris I, II zavarral vagy szkizoaffektív zavar bipoláris típusával diagnosztizált, egymással rokonságban nem álló, európai származású alanyai alkották. A diagnózist a vizsgálatot vezető pszichiáter, a felvett Diagnostic Interview for Genetic Studies (DIGS) és Family Instrument for Genetic Studies (FIGS) interjúk mániás, hipomániás, depresszív és kevert hangulati epizódjaira, szerhasználatlaltal vagy pszichózissal való kapcsolatukra, az epizód bizonyosságára, illetve családi anamnézisre vonatkozó adatai alapján, a legvalószínűbb diagnózis elve szerint állapította meg. Az egészséges kontrollcsoport telefonos interjú alapján került kiválasztásra, kizárva a major depresszív zavarral, bipoláris I vagy II zavarral, illetve pszichotikus zavarral valaha diagnosztizálható

jelentkezőket. Az eset és kontrollcsoport európai származását önbevallás alapján, négy európai származású nagyszülővel definiálták [165].

Vizsgálatunkba a bipoláris GWAS adatbázisból, az egészséges kontroll személyek, szkizoaffektív zavarral diagnosztizált alanyok, duplikált vagy alacsony minőséggel genotípezált minták kizárását követően, 999 mintát vontunk be. A szuicid magatartást legalább egy, ambivalens vagy komoly öngyilkossági szándékkal elkövetett, öngyilkossági kísérlet meglétével definiáltuk a DIGS interjú alapján. (Kérdőív: „Have you ever tried to kill yourself?” + „Interviewer: rate the intent of the most serious attempt”; változók: DIGS2: I14670 > 0 + I14750 > 1, DIGS3: Suicide\_Attempt > 0, Intent\_of\_Attempt\_1 > 1, DIGS4: V1822 > 0, V1860 > 1). Fenti definíciónk alapján a bipoláris mintán 358 szuicid, 616 nem szuicid és 25 hiányzó fenotípusú alanyt azonosítottunk.

A major depresszív zavar GWAS adatbázisa a holland Netherlands Study on Depression and Anxiety (NESDA) és Netherlands Twin Registry (NTR) vizsgálatokból származott. Az esetcsoport beválasztási kritériumai a következők voltak: 1, DSM-IV MDD élettartam diagnózisa a Composite International Diagnostic Interview (CIDI) szerint, 2, 18 és 65 év közötti életkor, 3, a holland nyelv megfelelő szintű ismerete, 4, észak-európai leszármazás. Az esetcsoportból kizárást a pszichotikus zavar, kényszerbetegség, bipoláris zavar vagy súlyos szerhasználati zavar elsődleges diagnózisa, vagy a holland nyelv elégtelen ismerete jelentett. A vizsgálat egészséges kontrolljai közé az NTR adatbázisában 1, kor és nem szerint illesztett, 2, alacsony szorongásos-depresszív faktor ponttal (< 0,65) rendelkezők, 3, kórelőzményben major depresszív epizódtól mentes, 4, 18 és 65 év közötti, 5, észak-európai származású alanyok kerültek [163,164].

Vizsgálatunkba az MDD GWAS adatbázisból, az egészséges kontrollok és a duplikált, illetve alacsony minőséggel genotípezált minták kizárását követően, 1753 alanyt vontunk be. Mivel a CIDI változóiban nem állt rendelkezésre kifejezetten öngyilkossági kísérletet mérő változó, ezért a szuicid magatartást a legalább hetente többszöri öngyilkossági gondolat, konkrét terv, esetleg kísérlet, vagy az öngyilkossági szándék elismerésével határoztuk meg, a rendelkezésre álló Inventory of Depressive Symptomatology (IDS) és a rövidített Beck Depression Inventory (BDI) alapján. (Kérdőív: ISD item 16: „az „I think of suicide or death several times a week for several minutes”, vagy a „I think of suicide or death several times a day in some detail, or I have

made specific plans for suicide or have actually tried to take my life” válaszok; BDI item 7 (g): az „I have definite plans about committing suicide” vagy a „I would kill myself if I had the chance” válaszok). E meghatározás alapján az 1753 MDD esetből 245 szuicid, 1360 nem szuicid és 150 hiányzó fenotípussal volt jellemezhető.

### 3.1.2. Genotípezálás, imputáció

A bipoláris minta genotípezálását az amerikai Broad Institute Center for Genotyping and Analysis végezte, vérből izolált, 50 ng/ul koncentrációjú DNS mintákon, az Affymetrix Genome-Wide Human SNP Array 6.0 1 millió SNP-t hibridizáló DNS chip segítségével. A DNS mennyiségét PicoGreen fluorometriával, a minták minőségét pedig és az alany nemét 24 SNP-t hibridizáló, Sequenom iPLEX platform segítségével ellenőrizték. Az allélok mérését az Affymetrix Power Tools BirdSeed algoritmusával végezték. A minőségellenőrző panel 85%-nál alacsonyabb mérési arányt mutató mintákat újra genotípezálták.

A genotípezálást követő minőségellenőrzésen kiszűrésre került minden minta, amelynél 1, a sikeresen mért SNP-k aránya 98,5% alatt volt, 2, a heterozigóta lokuszok aránya extrém alacsony vagy magas arányt ért el ( $< 0,344$  vagy  $0,363 <$ ), 3, az IBD analízis rejtett rokonságon, vagy azonosságot mutatott valamelyik másik mintával, illetve 4, a genotípezált nem eltért a minta rögzített nemétől. A minták szűrését követően kiszűrésre kerültek azon markerek, 1, amelyek mérési aránya 95%-ot nem érte el, 2, a túl ritka markerek ( $MAF < 0,01$ ), 3, a háromnál több mendeli hibával bíró markerek, 4, a kontroll mintákon a Hardy-Weinberg egyensúlytól szignifikánsan deviáló ( $p_{H-W} < 10^{-6}$ ) markerek, és 5, a szignifikáns lemez-specifikus hatást (bármelyik két lemez között  $p < 10^{-6}$  vagy több,  $p < 10^{-4}$ ) mutató markerek is. A minőség-ellenőrzés kritériumainak végül 729.087 SNP felelt meg.

Az MDD GWAS minta genotípezálása is vérből izolált DNS mintákon történt, az izolációt Qiagen FlexiGene® DNA AGF3000 Kit, GENTRA Puregene® Kit és AutoGenFlex 3000 izolációs munkaállomás segítségével végezték. A genotípezálást Perlegen 600K DNS chippel végezték, a genotípezálást követő minőségellenőrzés során eltávolításra került minden: 1, a 85% alatti sikeres SNP mérési aránnyal bíró minta, 2, az extrém alacsony vagy magas arányú heterozigóta lokusszal bíró minta ( $< 0,325$  vagy

0.363<). 3, szélsőséges IBS vagy IBD mutatókkal bíró minta (eltérő földrajzi eredet, vagy rejtett rokonság, duplikátum), 4, a túl ritka SNP-k ( $MAF < 0,01$ ), 5, a 95% alatti mérési rátával bíró markerek, és 6, a kontroll mintákon a Hardy-Weinberg egyensúlytól túlzottan eltérő eloszlású markerek ( $p_{H-W} < 10^{-6}$ ). A minőségellenőrzési kritériumoknak 1753 eset and 1800 kontroll és 437.114 marker felelt meg.

A szűrést követően a két genotípezált minta 175.373 közös markerrel rendelkezett, a két minta 24% illetve 40%-os átfedését jelölve, amely a tervezett összehasonlító elemzésben rendszerszintű torzításhoz vezetett volna. Az átfedő markerek arányának növelése érdekében ezért mindkét mintánkat azonos referenciagenom, a 1000 Genomes [10] Phase1 version3 kiadása alapján imputáltuk. Mintáink GRCH36/hg18 genom szerinti bázispozícióit a liftOver program segítségével konvertáltuk a referenciagenom GRCH37/hg19 pozícióira, majd a SHAPEIT [52,166] programmal ún. „best guess” fáziselemzést végeztünk. Az imputációt az IMPUTE2 [67] program segítségével hajtottuk végre mintáink haplotípus alléljain. Az imputáció elvét és lépéseit részletesen ismertettük a módszertani bevezető haplotípus fejezetében.

Az imputációs bizonytalanság okozta torzítások csökkentése érdekében konzervatív szűrést végeztünk, imputáció előtt kiszűrve a túl ritka és problémás eloszlást mutató markereket ( $MAF < 0,01$ ,  $p_{H-W} < 10^{-6}$ ), imputációt követően pedig a bizonytalan valószínűségű, alacsony információs mutatóval bíró, vagy túl ritka imputált markereket (imputációs valószínűség  $< 0,9$ , információs mutató  $< 0,9$ ,  $MAF < 0,01$ ). Imputációt követően mindkét mintában 6.706.910 közös biallélikus marker (SNP és indel) képezte az összehasonlító elemzés későbbi lépéseinek alapját.

**2. táblázat.** Vizsgálatainkban elemzett minták jellemzői.

<b>Jellemző</b>	<b>GAIN<sup>a</sup> MDD</b>	<b>GAIN BIP</b>	<b>SCHIZOBANK</b>	<b>GAIN SCZ</b>	<b>MGS<sup>b</sup> SCZ</b>
elérési kód	phs000020.v2.p1	phs000017.v3.p1	-	phs000021.v3.p2	phs000167.v1
teljes minta (eset:kontroll:hiány)	N = 3643 (1753:1822:68)	N = 2112 (999:1068:45)	N = 16 (trió)	N = 2787 (1318:1469)	N = 2935 (1405:1530)
elemzett minta (eset:kontroll:hiány)	n = 1753 (245:1360:150)	n = 999 (358:616:25)	n = 16	n = 5337 (2632:2705)	
genotípezálás	GWAS 600K	GWAS 1M	teljes-exom szekvenálás	GWAS 1M	
genotípezált SNV/SNP száma QC szűrés után	437.114	729.087	120.719	716.923	
SNP száma imputáció után	6.706.910	6.706.910	-	7.942.021	

a: GAIN: Genetics Information and Association Network

b: MGS: Molecular Genetics of Schizophrenia

### 3.1.3. Génszelekció, statisztikai elemzés

Vizsgálatunkban három kandidáns génlistát elemeztünk, amelyek eltérő hipotézisek szerint állítottunk össze. Az első listát (Génlista1) irodalomkutatással, legnagyobb részben az öngyilkossági viselkedés (szuicid ideáció vagy kísérlet) vizsgálati mintáinkkal legfeljebb csak részlegesen átfedő, teljes-genom asszociációs és expressziós profil vizsgálatok eredménye alapján állítottuk össze. A publikációkban legalább szuggesztív asszociációt, vagy expressziós változást elérő SNP-t tartalmazó géneket, valamint a mintáinkkal nem átfedő vizsgálatok szuggesztív gén-alapú asszociációt mutató génjeit gyűjtöttük össze. A 32 gént tartalmazó listát második lépésben kiegészítettük a bipoláris zavar GWAS vizsgálataiban szignifikáns SNP asszociációt mutató 3 génnel (CACNA1C, ANK3, and ZFN804A) [156,157], a bipoláris zavar és szuicid viselkedés klinikai és epidemiológiai vizsgálatokból ismert erős összefüggése miatt (Génlista1, gének száma: 35). Második kandidáns génlistánk (Génlista2), a végrehajtó funkciók zavarát a szuicid magatartás endofenotípusaként [167] kezelő, a végrehajtó funkciókhoz feltételezetten kapcsolható miRNS géneket tartalmazta. E listát a végrehajtó funkciók szervezésében meghatározó neurobiológiai rendszer, a dorsolaterális és az orbitofrontális prefrontális körök régióiban (prefrontális kéreg, cinguláris kéreg, nucleus caudatus, putamen, corpus striatum, globus pallidus, substantia nigra, thalamus, nucleus subthalamicus) expresszálódó miRNS génekből állítottuk össze, az European Bioinformatics Institute Gene Expression Atlas [27] segítségével. Az expresszió szerint meghatározott miRNS géneket kiegészítettük a bioszintézisükben, és a translációt gátló hatásukban kulcsfontosságú, RNA-induced-silencing-complex (RISC) génekkel, azaz a DICER1, DROSHA, DGCR8, EIF2C2, TARBP, TRNC6A, és XPO5 génjeivel (Génlista2, gének száma: 68). A harmadik kandidáns génlista (Génlista3) a Génlista2 gének mellett a fent említett anatómiai régiókban is expresszálódó célgénjeiket is tartalmazta (Génlista3, gének száma: 11259). Az miRNS-ek jóslott célgénjeit a miRanda és a mirSVR magas predikciós ponttal bíró adatbázisából határoztuk meg.

Vizsgálatunkban két statisztikai elemzést végeztünk kandidáns génlistáinkkal. Az első elemzésben a szuicid viselkedés és a Génlista1 35 génjének régió-alapú, betegség-specifikus és transzdiagnosztikus asszociációját vizsgáltuk a bipoláris, és a major depresszió mintáin. A gén-alapú asszociációkat két lépcsőben, a variáns asszociációk khi-

négyszet értékét kombináló teszttel számítottuk, ezért első lépésben a variáns asszociációkat határoztuk meg. A refSeq génhatárait mindkét irányban 5 kilobázissal kiterjesztve, a Génlista1 génei 17010 variánst tartalmaztak. Bár érdemi populációs stratifikációra egyik minta genomikus inflációs faktora ( $\lambda$ ), illetve becsült populációs rétegszáma (K) sem utalt (MDD: genotípezált  $\lambda = 1$ , imputált  $\lambda = 1,013$ ,  $K = 1$ ; bipoláris: genotípezált  $\lambda = 1,006$ , imputált  $\lambda = 1,001$ ,  $K = 1$ ), de a minták közti heterogenitás miatt a kombinációt egyszerű összevonás helyett metaanalízisben végeztük el. A genomikus inflációs faktort az R statisztikai programhoz írt GENABEL kiegészítővel [168] számítottuk, a populációs rétegszerkezetet pedig az ADMIXTURE programmal [169], 100.000 nem kapcsolt SNP segítségével kalkuláltuk. A két minta variáns alapú asszociációját a PLINK[88] program „dosage data analysis” funkciójával számítottuk ki, a p-értékeket a METASOFT program bináris-hatás metaanalízisével [170] kombináltuk, amely a random-hatás metaanalízis egy nagyobb statisztikai erővel bíró változata.

A gén-alapú asszociációkat a variáns statisztikákból az „SNP kombináló” HYST [69] módszerrel kalkuláltuk, amelyet a KGG [171] program implementált, és amelyet részletesen ismertettünk a módszertani bevezetőben. Az asszociációk hatáserősségének becslésére, a génrégiókra particionált, fogékonysági skálán kifejezett, genomikus heritabilitást is kiszámítottuk a GCTA [20,172] algoritmusával. Az eredmények jobb értelmezése érdekében, a mintáinkon elemzett gén-alapú asszociációk poszteriori statisztikai erejének határértékeit is megbecsültük a KGG programmal, a HYST teszt tekintetében optimális és szuboptimális hipotetikus génszerkezettel, valamint a BDNF gén valós genotípusa alapján.

A vizsgálatunk második elemzése a három kandidáns lista klasszikus GPRS, és egy bayesi megközelítésű poligénes predikciós módszer összehasonlítása volt, a nagyobb elemszámú MDD mintát fejlesztő, és a kisebb bipoláris mintát validáló mintaként használva. A klasszikus GPRS elemzést a PRSice program [173] alapbeállításának paramétereivel: 250 kilobázis csúszó ablakkal és  $r^2 < 0,1$  kapcsoltsággal futtattuk, az optimális  $p_t$  küszöbértéket a PRSice algoritmus határozta meg, a modell illeszkedési mutatói szerint. Ezzel hasonlítottuk össze az LDpred program [70] bayesi statisztikán alapuló predikcióját, amelynek alapjait részletesen ismertettünk a módszertani bevezetőben. Az LDpred elemzés első lépésében, a fejlesztő és validáló minták koordinációja során, kizártuk a nem egyértelmű alléllal bíró markereket, ezzel 13313,

1953, és 973758 markerre csökkentve a három Génlista méretét, amely 1, 4 és 324 LD-rádiusznak felelt meg az  $N/3000$  formula alapján. Az LD-rádiuszok ismeretében, az MDD mintán meghatároztuk a kauzális frakciók egyenletesen eloszló küszöbértékeihez ( $P < 0,001, < 0,003, < 0,005, < 0,01, < 0,03, < 0,05, < 0,1, < 0,3, < 0,5, P < 1$ ) tartozó poszterior PRS értékeket, majd a bipoláris mintán a legjobb kalibrációjú és pontosságú modellt validáltuk.

## 3.2. A második vizsgálat módszerei

### 3.2.1. Minta, fenotípus

Második vizsgálatunkba három független mintát vontunk be, amelyeket kétlépcsős (feltáró minta– validáló minta) szerkezetben elemeztünk. Kutatásunk feltáró mintáját a SCHIZOBANK vizsgálat 16 eset - triója alkotta, a triókat egy-egy DSM-IV szkizofréniával diagnosztizált utód és a szkizofréniával nem diagnosztizált szülők alkották. A SCHIZOBANK vizsgálat 18 és 60 év közötti, szkizofréniával diagnosztizált beteg és egészséges kontroll személyek mintájának biobankolására jött létre, a kizárást súlyos komorbid neurológiai betegség, mentális retardáció, koponyatrauma, bevasztáskor fennálló, aktív droghasználat és alkoholabúzus és alkoholfüggőség jelentette. A SCHIZOBANK teljes mintájának vizsgálati alanyai öt centrumból (Debreceni Egyetem Pszichiátriai Klinika, Szegedi Tudományegyetem Pszichiátria Klinika, a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinika, a budapesti Nyíró Gyula Kórház I. Pszichiátriai Osztálya és a miskolci Borsod-Abaúj-Zemplén Megyei Kórház és Egyetemi Oktatókórház Pszichiátriai Osztálya) kerültek bevonásra, vizsgálatunkban elemzett 16 trió döntő többsége ( $N = 12$ ) azonban a SE Pszichiátriai és Pszichoterápiás Klinika szkizofréniá miatt gondozott páciensei és szülei közül került ki. A DSM-IV szkizofréniá diagnózisát a Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) strukturált diagnosztikus interjú segítségével állítottuk fel, leggyakoribb diagnózis a nem differenciálható szkizofréniá volt ( $N = 10$ ). A vizsgálati személyek teljeskörű tájékoztatás után, írásos beleegyezést követően vettek részt a vizsgálatokban. A vizsgálatok az Egészségügyi Tudományos Tanács Tudományos és Kutatásetikai Bizottság (ETT TUKÉB) és a Semmelweis Egyetem Etikai Bizottságának



jóváhagyásával történtek. A kutatást az Egészségtudományi Tanács Tudományos és Kutatásügyi Bizottsága engedélyezte.

Vizsgálatunk replikációs mintája, az NCBI dbGaP két szkizofrénia eset-kontroll, teljes-genom asszociációs vizsgálatából származott. A minták bevonása a munkacsoportunk GAIN és JAAMH DAC által jóváhagyott, „Project#2862: Endophenotypes in affective and psychotic disorder spectra: a cross-disorder analysis” kutatási projektje részeként valósult meg. A GAIN Genome-Wide Association Study on Schizophrenia (GAIN SCZ, azonosító: phs000021.v3.p2, N = 2787) és a Molecular Genetics of Schizophrenia (MGS SCZ, azonosító: phs000167.v1.p1, N = 2935) [6,174–176] európai származású résztvevőiből vontuk össze a teljes replikációs mintát (N = 5722, N<sub>SCZ</sub> = 2723, N<sub>KONT</sub> = 2725, és 274 ismeretlen fenotípusú egyén). A szkizofrénia diagnózisának felállítására, mindkét mintában, legalább két gyakorlott klinikus egybehangzó vizsgálati lelete alapján került sor. A beválasztási és kizárási kritériumok mindkét vizsgálatban megegyeztek. Beválasztásra azon 18. életévüket betöltött, szkizofrénia vagy szkizoaffektív zavar DSM-IV kritériumait teljesítő alanyok kerültek, akiket a vizsgálat szervezői legalább 2 éve ismertek. Kizárási kritériumok közé az angol nyelv nem elégséges ismerete, EEG abnormalitások, súlyos mentális retardáció, illetve ha a pszichotikus tünetek kizárólag szerhasználat vagy neurológiai betegség következtében léptek fel.

### 3.2.2. Genotípezés, exom szekvenálás, imputáció

Az explorációs mintánk genomjain teljes-exom szekvenálást (whole-exome sequencing, WES) hajtottunk végre, új generációs szekvenálást biztosító Illumina HiScanSQ platformon, a hídamplifikációhoz NimbleGen SeqCap EZ Human Exome Library v3.0 könyvtárakat használtunk. A szekvenálás során 100 bázispár hosszú, párosított végű adaptereket használtunk, az olvasási szakaszokat a Q30 sztenderdek szerint szűrtük, és a hg19 referencia humán genomhoz illesztettük. A 48 minta olvasási szakaszait a Burrows-Wheeler Aligner szoftver [177] segítségével építettük össze. A variánsok méréséhez és újrakalibrálásához a Samtools szoftvert [178] használtuk, a variánsok annotációját az ANNOVAR [179] és a KGGseq [180], minőségkontrollját pedig a KGGseq programmal végeztük. Az exomok átlagos lefedettsége 27 és 46 között

változott, a teljes mintán 246.245 variánst mértünk. A minőségellenőrzés lépései során kiszűrtük: 1. a 20 alatti genotípus Phred-ponttal, vagy 5 alatti lefedettséggel bíró genotípusokat, 2. a mendeli hibát jelentő genotípusokat, 3. a 20 alatti variáns Phred-ponttal, vagy 40 alatti mapping minőségponttal bíró variánsokat, 4. a singleton variánsokat és a csak szülői genotípussal rendelkező variánsokat, 5. a kettőnél több alléllal rendelkező variánsokat, 6. a variánsok átlagánál 2 sztenderd deviációval több hiányzó genotípusú variánsokat, 7. a minta átlagánál 2 sztenderd deviációval több hiányzó genotípusú alanyokat. A minőség-ellenőrzés kritériumait 120.719 SNV vagy indel variáns teljesítette.

A vizsgálatunk replikációs mintáját alkotó szkizofrénia GWAS minták (GAIN és MGS SCZ) gyűjtése és genotípezálása azonos módszerekkel történt, ezért a továbbiakban összevontan tárgyaljuk őket. Az izolált DNS mintát Affymetrics Genome-Wide 6 chippel genotípezálták, amely megközelítőleg 1 millió markert hibridizál. Az SNP mérést az Affymetrix Power Tools Birdseed algoritmusával végezték el. A genotípezálást követő minőségellenőrzés során kiszűrésre kerültek: 1. az extrém alacsony vagy magas arányban heterozigóta lókuszt hordozó minták ( $<0,26$ , vagy  $0,285<$ ), 2. 97% alatti mérési aránnyal bíró minták, 3. túl ritka markerek ( $MAF < 0,01$ ), 4. 95% alatti mérési arányt mutató SNP-k, 5. a kontroll mintákon a Hardy-Weinberg egyensúlytól túlzottan deviáló markerek ( $p < 10^{-6}$ ), 6. szignifikáns lemezhatást mutató SNP-k (egy lemez esetén  $p < 10^{-8}$ , vagy több lemez esetén  $p < 10^{-4}$ ), 7. duplikált minták között kettőnél több eltérést mutató SNP-k. A szűrési feltételeknek a GAIN és MGS SCZ adatbázisban 5722 minta és 729454 SNP felelt meg.

Az adattisztítás lépései során kizártuk: 1. a 98% alatti mérési arányt mutató markereket, 2. a 98% alatti mérési aránnyal bíró mintákat, 3. a hiányzó fenotípusú mintákat, 4. az ADMIXTURE populációs szerkezet-elemzésével nem európai származásúként azonosított mintákat, és 5. a PLINK IBD analízis szerint rejtett rokonságban álló egyéneket. Az adattisztítás kritériumainak 5337 minta ( $N_{SCZ} = 2632$ ,  $N_{KONT} = 2705$ ) és 716.923 SNP felelt meg.

Az explorációs minta és a replikációs minta jelentősen különböző genotípezálása miatt kevés közös markerrel rendelkeztek (9366 SNP), ezért tervezett összehasonlító elemzésük érvényességét biztosítandó, a GWAS minták imputációjával emeltük meg a minták közötti átfedést. Mintáinkat az IMPUTE2 programmal imputáltuk, a 1000

Genomes Phase1v3 referenciagenom alapján. Az imputációhoz mintáink genomikus koordinátáit a genotípezált hg18-ról hg19-re konvertáltuk a liftOver program alkalmazásával, és a SHAPEIT szoftver segítségével meghatároztuk kromoszomális haplotípusalléljaikat. Az imputációt követően csak a legjobb minőségű, biallélikus variánsokat tartottuk meg (imputációs valószínűség  $> 0,9$ , információmutató  $> 0,9$ , MAF  $> 0,01$ ). A minőség kritériumoknak 7.942.021 SNV vagy indel variáns felelt meg, amellyel a minták közti lefedettség 22.279 markerre bővült.

### 3.2.3. Statisztikai elemzés, annotáció, variáns- és génprioritizáció

Kétlépcsős szerkezetű (explorációs és replikációs) vizsgálatunk explorációs lépcsőjében, a teljes-genom hipotézismentes tesztelésével meghatároztuk a replikálandó, legalább szuggesztív p-értéket ( $p < 0,1$ ) elérő géneket és génszetteket. A második lépcsőben már csak a replikációra jelölt géneket és génszetteket teszteltük, korrigálva p-értékeiket az adott hipotéziscsoport összes elvégzett explorációs és replikációs tesztjének számára. A vizsgálatunk három, részben egymásra épülő elemzésre tagozódott: 1. a multilokusz tesztekhez szükséges teljes-genom variáns asszociációk kiszámítása mindkét tesztmintán az additív modell szerint, explorációs lépésben transzmissziós elemzést, a replikációs lépcsőben a fő zavaró hatásokra kontrollált logisztikus regressziót alkalmaztunk.

2. a variáns asszociációk kombinációjából, teljes-genom, gén- és génszett-alapú asszociációk és feldúsulások elemzése az explorációs mintán, és a szuggesztív asszociációk hipotéziscsoportra korrigált validálása a replikációs mintán, 3. predikciós algoritmusokkal prioritizált variánsok és gényeik meghatározása, és funkcionális annotációik klaszterezése mindkét mintán. Tekintettel az explorációs minta alacsony elemszámára, a kétlépcsős vizsgálatokban jellemző nominális  $\alpha = 0,05$  helyett, az első lépcsőben már  $\alpha = 0,1$  küszöbértéket elérő asszociációk is a replikációs lépcsőbe kerültek. A többszörös összehasonlítás korrekcióját minden esetben a replikációs lépésben végeztük, az egyes multilokusz módszerek szerinti hipotéziscsoportoknak megfelelően, a csoportba tartozó explorációs és replikációs tesztek összesített száma szerinti Bonferroni-korrekciónal. A gén-alapú hipotéziscsoport 17.600, a génszett-

asszociációk 1107, a génszett-feldúsulás 1107, a funkcionális annotációs klaszterelemzés pedig 927 tesztelt hipotézist tartalmazott.

A variánsok asszociációs elemzését mindkét mintán additív öröklési modell szerint végeztük. Az explorációs, eset-trió alapú mintában, a minta alacsony elemszámára tekintettel, az inkomplett genotípusú triókat is használni képes, Family-Based Association Test (FBAT, Laird et al. 2000) elemzést végeztünk. Az alacsony elemszám miatt, az FBAT statisztika számításához minimálisan szükséges, informatív triók számát a javasolt 10-ről 4-re csökkentettük, a teszt érvényességét maximálisan 100.000 ciklusú, adaptív permutációval biztosítottuk. A minta 120.719 biallélikus SNV vagy indel variánsából 66.737 mutatott additív transzmissziót minimálisan 1 informatív családban, legalább 4 informatív családban pedig csak 29.821 variáns fordult elő.

A minta megfelelő nagysága miatt, a validálás során a bemeneti SNP asszociációkat a PLINK szoftver additív öröklésmentet modellező, logisztikus regressziójával számítottuk, kovariánsokkal kontrollálva a nem, az életkor és a populációs stratifikáció zavaró hatását. A populációs szerkezet torzító hatását a GWAS minta genetikai varianciájának főkomponenseivel korrigáltuk. A főkomponens elemzést a FlashPCA [182] szoftverrel, a populációtól szélsőségesen deviáló minták kizárásával, a genotípezált, független ( $r^2 < 0,2$ ) markerek varianciájából számított genetikai távolsági mátrixon végeztük. A sajátértékfüggvény és a kumulatív kifejezett varianciafüggvény inflexiós pontja alapján választottuk ki a populációs kovariánsként használandó főkomponenseket, e módszer az első 5 főkomponenst azonosította. A replikációs minta SNV-alapú asszociációit a PLINK szoftverrel számítottuk.

A gén- és génszett-alapú analíziseket a KGG program 3.5 verziójával végeztük, amely a variánsok egyedi asszociációs statisztikáit a kapcsoltsági mutatók alapján kombinálva számítja a régió-alapú asszociációt és a feldúsulási teszteket. A gén-alapú asszociációkat a KGG GATES tesztjével elemeztük, amelyet részletesen ismertettünk a módszertani bevezetőben. A vizsgált gének határait a GENCODEv19 kiadásában rögzített koordinátákhoz képest, mindkét irányban 5 kilobázis méretben kiterjesztve határoztuk meg. A GENCODE annotációk az explorációs mintán 15.795, a replikációs mintán pedig 51.366 gént és transzkriptet azonosítottak.

A kanonikus útvonalak és a pozicionális génszettek asszociációját és feldúsulását az MSiGDB v5.2 [107] c2.cp illetve c1 adatbázisai alapján elemeztük. Az útvonalak

méretéből eredő torzítás minimalizálása céljából az elemzést 10-500 gén méretű útvonalakra korlátoztuk. A tesztekhez szükséges kapcsoltsági paramétereket a 1000 Genomes Phase1v3 kiadásának európai leszármazású (N = 379) haplotípusaiból számítottuk. Az MSiGDB v5.2 c1 pozicionális adatbázis 326 nem átfedő génszettjéből 248 felelt meg a méretbeli kritériumoknak. A kanonikus útvonalak jelentős mértékű átfedései miatt, a redundanciát a ReCiPa [183] algoritmusával kontrolláltuk, 0,85 maximális és 0,1 minimális átfedési paraméterekkel. Az MSiGDB 1329 kanonikus útvonalából a ReCiPa algoritmus 541 nem átfedő kanonikus útvonalat, és az átfedők kombinációjával 122 ún. „szuperútvonalat” (Superpathway), vagyis összesen 663 útvonalat definiált. A génszett-elemzés metodológiai ajánlásait követve, két különböző statisztikai módszert alkalmaztunk: a KGG önálló (nem-kompetitív), régió-alapú asszociációs HYST tesztjét, valamint a kompetitív jellegű, Wilcoxon előjeles rangszámösszeg feldúsulási tesztet. Mindkét tesztet részleteiben ismertettük a módszertani bevezetőben.

A funkcionális annotációs klaszterek meghatározását az online elérhető Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8 verziójával [109] végeztük, az információ maximalizálása érdekében az alapbeállítás annotációs forrásait kiegészítettük: 1. az ontológiai kategóriából „leggyakrabban használt terminológia” (Frequently Accessed Terms: GOTERM\_FAT) adataival, 2. az általános annotációs csoportból a kromoszóma, a citogenetikai sáv és a hivatalos gén szimbólum annotációival, 3. az útvonalak kategóriából pedig a REACTOME kanonikus útvonalaiival. Az annotációs klaszterelemzés alapját jelentő felülreprezentációs (ORA) teszt bemeneti génlistáját, mindkét mintában azonos algoritmus szerint prioritizált „rizikó gének” alkották, az ORA teszt háttéréként pedig a mintákban lefedett gének szolgáltak. Rizikó géneként a KGGseq mendeli [184], illetve regulátoros [185] predikciója alapján patogén, vagy az EIGEN [186] spektrális algoritmusával funkcionális (EIGEN-pont > 1,25), illetve a KGGseq alternatív splicing annotációja alapján splicing módosítást valószínűsítő, emellett legalább szuggesztív p-értékkel asszociálódó (explorációs minta:  $p < 0,1$ , replikációs minta:  $p < 0,05$ ) variánsok génjeit definiáltuk. Az explorációs mintában, a kombinált predikciós algoritmussal 441 valószínűsíthetően funkcionális variánst és 402 gént választottunk ki, míg a replikációs GAIN-MGS mintában 3941 potenciónalis funkcionális variánst, és ezeket tartalmazó 1785 prioritizált gént

határoztunk meg. A prioritizált variánsok és gének tekintetében egyaránt alacsony volt a minták közti átfedés (közös „rizikóvariáns”: 23, „rizikógén”:75), amely az explorációs minta prioritizált variánsainak 5,21%-át és génjeinek 18,66%-át jelentette. Az ORA teszt háttéré az explorációs mintában 14.618 gént tartalmazott, a replikációs mintában pedig 38.940 gént. A DAVID klaszterező algoritmusát az alapértelmezett (közepes) szigorúsági fokon használtuk. A klaszterek nominálisan szignifikáns küszöbértékét a  $ES = 1,3 (-\log_{10} 0,05)$  pontban határoztuk meg, de a szerző ajánlásait [108] követve, a kis elemszámú explorációs mintánkon az  $ES > 1,2$  értékkel rendelkező klasztereket is exploráltuk.

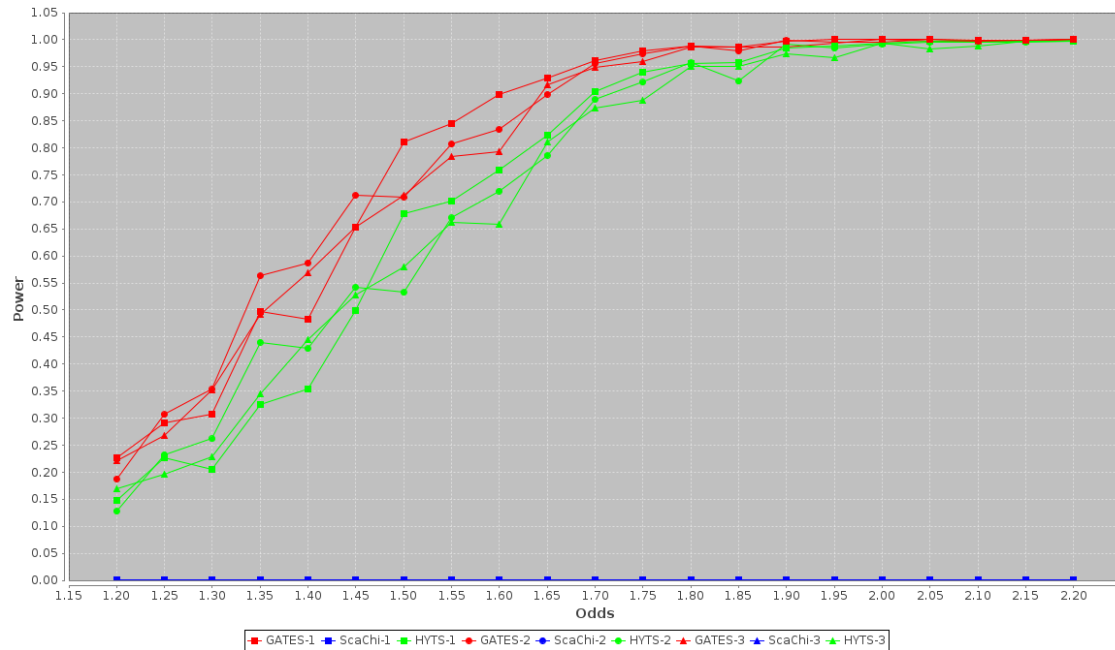
## 4. Eredmények

### 4.1. Az első vizsgálat eredményei

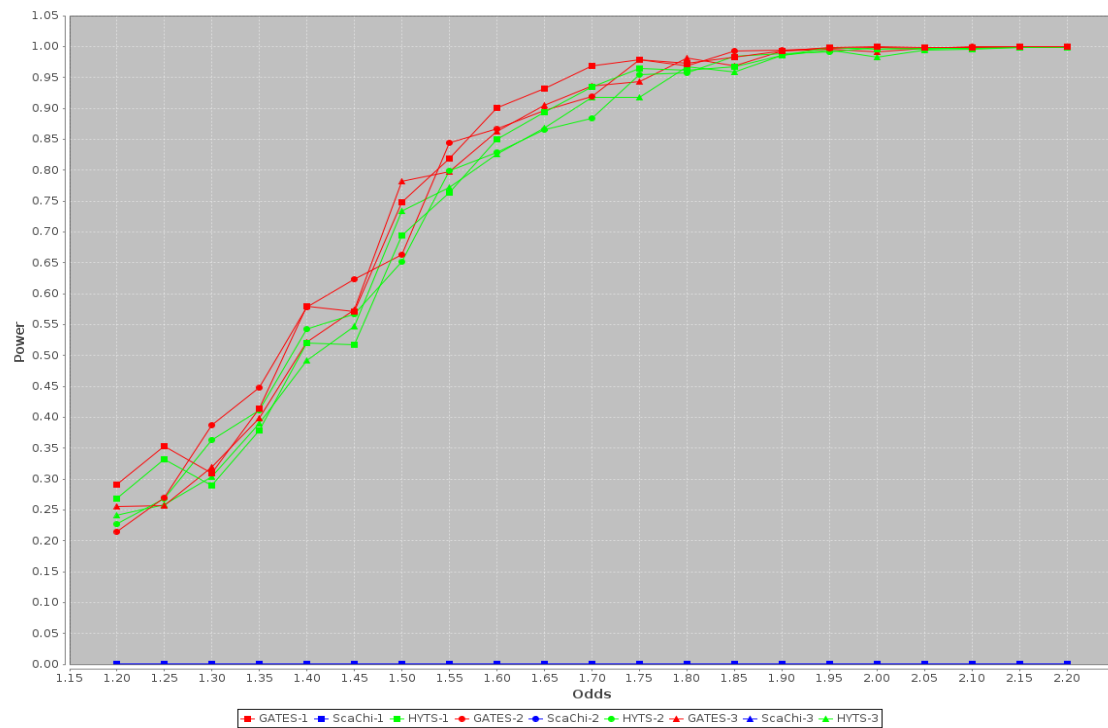
#### 4.1.1. Régió-alapú statisztikai erő

Vizsgálatunk régió-alapú szimulált statisztikai ereje, a HYST-teszt szempontjából optimális ( $n = 12$  DSL lókuszt,  $MAF > 0,05$ , LD-blokk  $r^2 = 0,8$ ) és kedvezőtlen ( $n = 1$  DSL lókuszt,  $MAF < 0,01$ , LD-blokk  $r^2 = 0,25$ ) paraméterek függvényében, széles skálán mozgott. A statisztikai erő szimulált határértékei  $OR = 1,2$  DSL lókuszt tartalmazó gének esetében  $0,05-0,70$  között változott a GAIN bipoláris mintán,  $0,08-0,74$  között a GAIN MDD mintán és  $0,06-0,94$  között a metaanalízisben kombinált mintán. A BDNF gén nominális asszociációjának kimutatásának megfelelő poszterior statisztikai erő,  $OR = 1,2$  kauzativ lókuszt feltételezve  $0,16-0,29$  között változott a GAIN bipoláris mintán (**2A. ábra**) és  $0,22-0,29$  között a GAIN MDD mintán (**2B. ábra**) a DSL lókusztok számától függően ( $n = 1-3$ ).

A



B



**2. ábra.** A BDNF gén nominális régió-alapú asszociációjának megfelelő szimulált poszterior statisztikai erő a GAIN bipoláris (A) és major depresszív zavar (B) mintákban a feltételezett DSL lókuszok száma, esélyhányada, és a statisztikai teszt függvényében.



## 4.1.2. Gén-alapú asszociációk és a gén-régiókhöz tartozó genomikus heritabilitás

A korrelációs génlista (Génlista1) régió-alapú asszociációi és a gének szerint particionált genomikus heritabilitás ( $h^2$ ) értékei a **3. táblázatban** láthatók. A GAIN MDD mintán az NTRK2 gén asszociációja volt a legerősebb ( $p=0,0002$ ,  $h^2=0,01$ ), ezt követte az NXPH1 ( $p=0,012$ ,  $h^2=0,014$ ), a GRIA3 ( $p=0,017$ ,  $h^2=0,016$ ), a DICER1 ( $p=0,032$ ,  $h^2=0,006$ ) és az SPTLC1 ( $p=0,041$ ,  $h^2=0,006$ ) gének asszociációi.

A GAIN bipoláris mintában nominálisan szignifikáns asszociációkat észleltünk a hsa-miR-195 ( $p=0,017$ ,  $h^2=0,003$ ), a CD44 ( $p=0,019$ ,  $h^2=0,009$ ) és a DICER1 ( $p=0,034$ ,  $h^2=0,004$ ) gének esetében. Ugyanakkor a kombinált elemzésben ismét az NTRK2 ( $p=0,002$ ), a GRIA3 ( $p=0,011$ ) és a DICER1 ( $p=0,046$ ) gének  $p$ -értékei lépték át a nominális küszöböt ( $\alpha=0,05$ ). A többszörös összehasonlítás korrekciója után egyik gén asszociációja sem maradt szignifikáns, kivéve az NTRK2 gént az MDD mintában ( $p_{\text{corr}}=0,014$ ).

Mivel a szuicid magatartás a GWAS minták fő fenotípusának részalmazát képezi, ezért az szignifikáns asszociációk validitásának bizonyítása is szükségesé vált. Ennek érdekében, a nominálisan szignifikáns gének, és a minták fő fenotípusai közötti asszociációit is teszteltük, az egészséges kontrollokat is tartalmazó, teljes GWAS mintán. Feltételeztük, hogy ha a gének valójában a szuicid magatartást is tartalmazó hangulatzavarral korrelálnak, akkor a minta megnövelésével felerősödnek az asszociációk. Ezzel szemben, a minták fő fenotípusaival egyik gén sem mutatott szignifikáns asszociációt az MDD és az NTRK2 gén asszociációján kívül, azonban itt is az asszociáció jelentős gyengülését tapasztaltuk ( $p_{\text{MDD}}=0,04$  vs  $p_{\text{szuic}}=0,0002$ ), amely az eredmények érvényességét igazolja

A DICER1 és az NTRK2 SNP markereinek funkcionális annotációja a **4-5. táblázatban** látható. Bár a nominálisan szignifikáns markerek között egyik sem fehérjekódoló, számos SNP esetében valószínűsíthető szabályozó funkció. Az ENCODE projekt [187] 6 immortalizált sejtvonalaának, konzisztens kromatin szegmentációja, illetve DNaseI Hiperszenzitivitás (DHS) csúcsai [188], és a humán asztrocita kromatin állapota alapján promoter, enhanszer vagy DHS lókuszként annotálhatók. A 6 sejtvonala konzisztens jelei a szövetektől független, transzkripciót szabályozó elemek, míg a humán asztrocita kromatin-állapota a neuron-specifikus szabályozó lókuszek azonosításában informatív. A DICER1 gén markerei közül, két, egymással szorosan kapcsolt DHS SNP

(rs1209904 és rs1187655) mindkét mintában azonos irányú, nominálisan szignifikáns asszociációját észleltük, miközben rs1187646, rs11851369 és rs7157322 SNP-k csak az egyik mintában bizonyultak szignifikánsnak, de a másik mintában is hasonló OR értékkel bírtak. Az NTRK2 szignifikáns markerei többnyire mintaszpecifikusak voltak, egy közös szignifikáns variáns-alapú asszociációt sem észleltünk, ugyanakkor a nominálisan szignifikáns markerek közel negyede ( $35/149 = 23\%$ ) epigenetikai szabályozó funkcióval rendelkezhet.

**3. táblázat.** A szuicid magatartáshoz kapcsolható gének régió-alapú asszociációja és fogékonysági skálán kifejezett genomikus heritabilitása a bipoláris (BIP), major depresszív (MDD), és a kombinált mintán.

Gén	Lokalizáció	#SNP	BIP p-érték (h <sup>2</sup> )	MDD p-érték (h <sup>2</sup> )	Kombinált p-érték
ACPI	2p25	43	0,319 (0,002)	0,713 (0,000)	0,483
ANK3	10q21	2040	1,000 (0,000)	0,687 (0,000)	0,981
BDNF	11p13	107	0,850 (0,000)	0,907 (0,000)	0,957
CACNA1C	12p13.3	1228	0,796 (0,000)	0,689 (0,003)	0,711
CD44	11p13	269	0,019 (0,008)	0,749 (0,000)	0,137
CYTH3	7p22.1	199	0,725 (0,001)	0,925 (0,000)	0,919
DICER1	14q32.13	88	0,035 (0,004)	0,032 (0,006)	0,025
DSC2	18q12.1	46	0,488 (0,001)	0,604 (0,000)	0,570
FKBP5	6p21.31	289	0,992 (0,000)	0,158 (0,002)	0,930
FOXP3	14q31.3	961	0,643 (0,000)	0,998 (0,000)	1,000
GDA	9q21.13	395	0,082 (0,009) <sup>a</sup>	0,958 (0,000)	0,239
GFRA1	10q26.11	532	0,353 (0,001)	0,625 (0,000)	0,502
GRIA3	Xq25	387	0,883 (0,000)	0,017 (0,016) <sup>a</sup>	0,015
GRIK2	6q16.3	1736	0,061 (0,024) <sup>a</sup>	0,663 (0,000)	0,164
hsa-miR-185	22q11.21	24	0,814 (0,000)	0,852 (0,000)	1,000
hsa-miR-195	17p13.1	14	0,017 (0,003) <sup>a</sup>	0,999 (0,000)	0,054
hsa-miR-30a	6p13	22	0,534 (0,000)	0,305 (0,000)	0,158
hsa-miR-491	9p21.3	17	0,986 (0,000)	0,336 (0,002)	0,597
IFNLR1	1p36.11	74	0,813 (0,000)	0,707 (0,000)	0,965
KIAA1244	6q23.3	446	0,300 (0,001)	0,320 (0,004)	0,348
MARCH1	4q32.2	1532	0,766 (0,000)	0,971 (0,000)	0,965
MBNL2	13q32.1	324	0,158 (0,002)	0,779 (0,000)	0,576
NTRK2	9q22.1	838	0,963 (0,000)	0,0002* (0,01) <sup>a</sup>	0,002
NXPH1	7p22	1032	0,959 (0,000)	0,012 (0,014) <sup>a</sup>	0,192
PAPLN	14q24.2	129	0,911 (0,000)	0,763 (0,000)	0,981
PRKCE	2p21	1527	0,961 (0,000)	0,241 (0,018)	0,694
RGS18	1q31.2	44	0,437 (0,001)	0,546 (0,000)	0,801
RGS2	1q31	14	0,332 (0,002)	0,163 (0,000)	0,351
SAT1	Xp22.1	29	0,420 (0,000)	0,196 (0,000)	0,275
SLC4A4	4q21	540	0,707 (0,000)	0,930 (0,000)	0,901
SPTLC1	9q22.2	296	0,503 (0,000)	0,041 (0,005)	0,106
SRSF11	1p31	67	0,912 (0,000)	0,484 (0,000)	0,650
TMEM132C	12q24.32	1089	0,287 (0,000)	0,807 (0,001)	0,829
TUBGCP3	13q34	284	0,876 (0,000)	0,974 (0,000)	0,968
ZNF804A	2q32.1	1027	0,897 (0,000)	0,944 (0,000)	1,000

\* p-érték szignifikáns hipotéziscsoport szerinti korrekció után

a a génhez tartozó genomikus heritabilitás szignifikánsan nagyobb nullánál (h<sup>2</sup> > 0)

jegyzet: h<sup>2</sup> = (V(G)/Vp\_L : a fogékonyság (liability) skálán kifejezett genomikus heritabilitás)

**4. táblázat.** A DICER1 gén nominálisan szignifikáns markereinek epigenetikai funkcionális annotációja a GAIN major depresszív zavar (MDD) és bipoláris zavar (BIP) mintákon.

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrócita)
rs12882291	95626054	0,223	1,108	3,40E-01	0,241	1,358	4,83E-03	-	Gyenge Enhanszer
rs11851369	95624926	0,205	1,115	3,30E-01	0,224	1,34	9,38E-03	PROMOTER	Gyenge Enhanszer
rs74076802	95589285	0,197	1,214	8,65E-02	0,215	1,337	1,07E-02	DHS	Gyenge Enhanszer
<b>rs1187655</b>	<b>95565556</b>	<b>0,224</b>	<b>1,282</b>	<b>2,12E-02</b>	<b>0,244</b>	<b>1,306</b>	<b>1,59E-02</b>	<b>DHS</b>	<b>Gyenge Enhanszer</b>
<b>rs1209904</b>	<b>95563712</b>	<b>0,224</b>	<b>1,282</b>	<b>2,11E-02</b>	<b>0,244</b>	<b>1,306</b>	<b>1,59E-02</b>	<b>DHS</b>	<b>Gyenge Enhanszer</b>
rs7156553	95612231	0,195	1,186	1,33E-01	0,216	1,311	1,73E-02	-	Gyenge Enhanszer
rs17784006	95561608	0,087	0,868	4,27E-01	0,099	0,63	1,87E-02	-	-
rs79824073	95576968	0,087	0,873	4,46E-01	0,099	0,631	1,89E-02	-	-
rs34487929	95589261	0,194	1,197	1,12E-01	0,214	1,303	1,98E-02	DHS	Gyenge Enhanszer
rs2275182	95581899	0,195	1,203	1,03E-01	0,214	1,303	1,98E-02	-	Erős Enhanszer
rs77538246	95565200	0,087	0,873	4,44E-01	0,099	0,634	1,99E-02	-	-
rs12890506	95606785	0,194	1,189	1,27E-01	0,214	1,302	2,07E-02	DHS	Gyenge Enhanszer
rs12890241	95597540	0,195	1,189	1,27E-01	0,214	1,301	2,09E-02	-	Erős Enhanszer
rs6575499	95622200	0,195	1,179	1,46E-01	0,214	1,301	2,11E-02	PROMOTER	Inaktív Promoter
rs11160231	95611297	0,195	1,195	1,16E-01	0,217	1,298	2,18E-02	-	Gyenge Enhanszer
rs75901131	95626714	0,174	1,111	3,77E-01	0,185	1,309	2,45E-02	-	Gyenge Enhanszer
rs76051421	95551554	0,198	1,202	1,00E-01	0,22	1,282	2,78E-02	PROMOTER	Erős Enhanszer
rs148499550	95623594	0,095	1,215	2,25E-01	0,087	0,641	2,96E-02	DHS_PROMOTER	Aktív Promoter
rs66633849	95567026	0,089	1,262	1,55E-01	0,082	0,64	3,28E-02	-	Gyenge Enhanszer
rs17091818	95567299	0,089	1,262	1,55E-01	0,082	0,64	3,28E-02	-	Gyenge Enhanszer
rs55819369	95569019	0,09	1,27	1,44E-01	0,083	0,643	3,29E-02	-	Erős Enhanszer
rs10498642	95588539	0,088	1,268	1,45E-01	0,081	0,641	3,34E-02	DHS	Gyenge Enhanszer
rs67737119	95591070	0,088	1,257	1,63E-01	0,081	0,643	3,48E-02	-	Erős Enhanszer
rs2282265	95559858	0,089	1,264	1,52E-01	0,081	0,642	3,50E-02	-	Erős Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs1953535	95596279	0,088	1,259	1,61E-01	0,081	0,644	3,60E-02	-	Erős Enhanszer
rs7146983	95610878	0,087	1,26	1,60E-01	0,081	0,644	3,60E-02	-	Gyenge Enhanszer
rs8006416	95617062	0,087	1,26	1,61E-01	0,081	0,644	3,62E-02	-	Gyenge Enhanszer
rs3742330	95553362	0,09	1,263	1,49E-01	0,082	0,647	3,63E-02	PROMOTER	Erős Enhanszer
rs55874297	95593957	0,088	1,257	1,63E-01	0,081	0,645	3,64E-02	DHS	Gyenge Enhanszer
rs11629124	95600072	0,088	1,259	1,61E-01	0,081	0,645	3,65E-02	DHS	Gyenge Enhanszer
rs4905282	95598665	0,087	1,258	1,64E-01	0,081	0,645	3,66E-02	-	Gyenge Enhanszer
rs2297730	95578675	0,09	1,301	1,05E-01	0,083	0,654	3,84E-02	-	Erős Enhanszer
rs17091810	95564538	0,09	1,302	1,05E-01	0,083	0,654	3,84E-02	-	Gyenge Enhanszer
rs67981103	95549402	0,091	1,285	1,20E-01	0,082	0,652	3,88E-02	-	Gyenge Enhanszer
rs67822993	95587355	0,089	1,277	1,34E-01	0,083	0,655	3,91E-02	-	Gyenge Enhanszer
rs66707045	95548066	0,09	1,265	1,45E-01	0,083	0,653	3,95E-02	-	Gyenge Enhanszer
rs17091808	95564233	0,09	1,264	1,52E-01	0,082	0,654	4,06E-02	-	Gyenge Enhanszer
rs17091855	95619101	0,09	1,305	1,05E-01	0,083	0,66	4,32E-02	-	Gyenge Enhanszer
rs55758041	95594818	0,088	1,278	1,35E-01	0,083	0,66	4,33E-02	DHS	Gyenge Enhanszer
rs66997818	95566108	0,198	1,251	5,42E-02	0,209	1,273	4,52E-02	-	Gyenge Enhanszer
rs1187646	95548723	0,315	1,377	1,36E-03	0,328	1,114	2,96E-01	DHS	Gyenge Enhanszer
rs10131275	95579992	0,31	1,37	1,85E-03	0,321	1,119	2,77E-01	-	Erős Enhanszer
rs7157322	95582544	0,31	1,37	1,85E-03	0,321	1,123	2,63E-01	DHS	Erős Enhanszer
rs8007729	95580470	0,31	1,368	1,91E-03	0,321	1,123	2,63E-01	-	Erős Enhanszer
rs12892375	95585553	0,309	1,367	1,96E-03	0,321	1,119	2,78E-01	-	Gyenge Enhanszer
rs201176701	95592420	0,308	1,366	2,05E-03	0,32	1,118	2,84E-01	-	Gyenge Enhanszer
rs4513027	95575414	0,314	1,362	2,08E-03	0,328	1,106	3,30E-01	-	Erős Enhanszer
rs1187634	95627753	0,391	1,259	1,63E-02	0,385	1,152	1,58E-01	-	Gyenge Enhanszer
rs1187633	95627797	0,391	1,259	1,63E-02	0,385	1,152	1,58E-01	-	Gyenge Enhanszer
rs10139161	95600006	0,283	1,284	1,64E-02	0,297	1,096	3,88E-01	DHS	Gyenge Enhanszer
rs1953536	95608450	0,283	1,284	1,65E-02	0,297	1,096	3,86E-01	-	Gyenge Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs7145897	95610324	0,283	1,283	1,65E-02	0,297	1,096	3,86E-01	DHS	Gyenge Enhanszer
rs8012909	95602659	0,283	1,283	1,65E-02	0,297	1,093	3,98E-01	PROMOTER	Gyenge Enhanszer
rs12434687	95613471	0,283	1,283	1,68E-02	0,297	1,096	3,87E-01	-	Gyenge Enhanszer
rs12323635	95625711	0,391	1,257	1,71E-02	0,385	1,152	1,57E-01	PROMOTER	Gyenge Enhanszer
rs7157100	95619475	0,282	1,279	1,80E-02	0,297	1,095	3,90E-01	-	Gyenge Enhanszer
rs75784129	95621132	0,282	1,279	1,80E-02	0,297	1,095	3,90E-01	-	Gyenge Enhanszer
rs10587098	95621926	0,282	1,279	1,81E-02	0,297	1,095	3,90E-01	-	Inaktív Promoter
rs6575500	95622232	0,282	1,279	1,81E-02	0,297	1,095	3,90E-01	PROMOTER	Gyenge Promoter
rs56200135	95594687	0,284	1,277	1,84E-02	0,297	1,097	3,82E-01	DHS	Gyenge Enhanszer
rs11620843	95595654	0,284	1,274	1,99E-02	0,297	1,096	3,85E-01	DHS	Erős Enhanszer
rs28564498	95626503	0,168	1,299	3,86E-02	0,144	0,792	1,25E-01	-	Gyenge Enhanszer
rs3742331	95625254	0,17	1,297	3,98E-02	0,146	0,797	1,35E-01	PROMOTER	Gyenge Enhanszer

**Jegyzet:** a két mintában azonos irányú, nominálisan szignifikáns asszociációt mutató, SNP félkövérrel emeltük ki

**5. táblázat.** Az NTRK2 gén nominálisan szignifikáns markereinek epigenetikai funkcionális annotációja a GAIN major depresszív zavar (MDD) és bipoláris zavar (BIP) mintákon.

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs34756138	87279007	0,034	0,959	8,70E-01	0,026	2,528	1,10E-04	-	Gyenge Enhanszer
rs13292902	87279230	0,034	0,959	8,70E-01	0,026	2,528	1,10E-04	-	Gyenge Enhanszer
rs13292909	87279242	0,034	0,96	8,70E-01	0,026	2,53	1,00E-04	-	Gyenge Enhanszer
rs13294053	87279291	0,034	0,959	8,70E-01	0,026	2,528	1,10E-04	-	Gyenge Enhanszer
rs78109009	87279668	0,034	0,959	8,70E-01	0,026	2,529	1,10E-04	-	Gyenge Enhanszer
rs13288682	87281538	0,034	0,959	8,70E-01	0,026	2,53	1,00E-04	PROMOTER	Gyenge Enhanszer
rs3758316	87282568	0,034	0,959	8,70E-01	0,026	2,53	1,00E-04	DHS_PROMOTER	Gyenge Enhanszer
rs3758318	87284998	0,035	0,956	8,60E-01	0,026	2,533	1,00E-04	DHS_PROMOTER	Aktív Promoter
rs45554739	87285237	0,035	0,997	9,90E-01	0,027	2,41	2,10E-04	PROMOTER	Gyenge Promoter
rs3824517	87287171	0,035	0,949	8,40E-01	0,026	2,538	1,00E-04	DHS	Gyenge Enhanszer
rs35127806	87288173	0,035	0,94	8,10E-01	0,026	2,545	9,70E-05	-	Gyenge Enhanszer
rs35087952	87290058	0,035	0,925	7,60E-01	0,026	2,561	9,00E-05	-	Gyenge Enhanszer
rs35012916	87291558	0,035	0,908	7,10E-01	0,026	2,511	1,20E-04	-	Gyenge Enhanszer
rs75654661	87291645	0,035	0,908	7,10E-01	0,026	2,511	1,20E-04	-	Gyenge Enhanszer
rs55811269	87292328	0,06	0,94	7,60E-01	0,052	1,677	8,50E-03	DHS	Gyenge Enhanszer
rs55923049	87292690	0,06	0,938	7,60E-01	0,052	1,677	8,50E-03	-	Gyenge Enhanszer
rs36004043	87293143	0,034	0,899	6,90E-01	0,026	2,513	1,20E-04	-	Gyenge Enhanszer
rs67725649	87293781	0,06	0,936	7,50E-01	0,052	1,677	8,50E-03	-	Gyenge Enhanszer
rs34572907	87294675	0,034	0,897	6,80E-01	0,026	2,513	1,20E-04	-	Gyenge Enhanszer
rs34389193	87295257	0,034	0,896	6,80E-01	0,026	2,514	1,20E-04	-	Gyenge Enhanszer
rs75170711	87296294	0,034	0,899	6,80E-01	0,026	2,517	1,10E-04	-	Gyenge Enhanszer
rs13293643	87296476	0,034	0,9	6,90E-01	0,026	2,518	1,10E-04	-	Gyenge Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs36118278	87296969	0,035	0,902	6,90E-01	0,026	2,52	1,10E-04	DHS	Gyenge Enhanszer
rs36117205	87297157	0,035	0,903	7,00E-01	0,026	2,521	1,10E-04	-	Gyenge Enhanszer
rs56134320	87298614	0,035	0,909	7,20E-01	0,026	2,528	1,10E-04	-	Gyenge Enhanszer
rs56102454	87299187	0,034	0,927	7,70E-01	0,026	2,544	1,10E-04	-	Gyenge Enhanszer
rs35085611	87299715	0,035	0,914	7,30E-01	0,026	2,532	1,00E-04	-	Gyenge Enhanszer
rs13302051	87301865	0,035	0,924	7,60E-01	0,026	2,54	1,00E-04	-	Gyenge Enhanszer
rs67624169	87302115	0,06	0,956	8,20E-01	0,052	1,669	9,00E-03	-	Gyenge Enhanszer
rs35042863	87304165	0,035	0,932	7,90E-01	0,026	2,5	1,50E-04	-	Gyenge Enhanszer
rs34698491	87305601	0,035	0,937	8,00E-01	0,025	2,552	1,10E-04	-	Gyenge Enhanszer
rs113414991	87308053	0,035	0,937	8,00E-01	0,025	2,552	1,10E-04	-	Gyenge Enhanszer
rs34712720	87310166	0,035	0,938	8,00E-01	0,025	2,551	1,20E-04	-	Gyenge Enhanszer
rs13283613	87311038	0,035	0,916	7,30E-01	0,026	2,415	2,50E-04	-	Gyenge Enhanszer
rs67387913	87311853	0,035	0,916	7,30E-01	0,026	2,415	2,50E-04	-	Gyenge Enhanszer
rs35725744	87314580	0,039	0,915	7,20E-01	0,03	2,132	1,50E-03	DHS	Gyenge Enhanszer
rs13301519	87315834	0,035	0,934	7,90E-01	0,025	2,55	1,20E-04	-	Gyenge Enhanszer
rs13283268	87316085	0,035	0,938	8,00E-01	0,025	2,55	1,20E-04	-	Gyenge Enhanszer
rs7041260	87318528	0,036	0,954	8,50E-01	0,025	2,483	1,80E-04	-	Gyenge Enhanszer
rs34121083	87319018	0,035	0,938	8,00E-01	0,025	2,538	1,30E-04	DHS	Gyenge Enhanszer
rs13290917	87321367	0,036	0,955	8,60E-01	0,026	2,365	3,50E-04	-	Gyenge Enhanszer
rs13288194	87323985	0,035	0,938	8,00E-01	0,025	2,537	1,30E-04	-	Gyenge Enhanszer
rs13293362	87325380	0,035	0,937	8,00E-01	0,025	2,537	1,30E-04	-	Inaktív Promoter
rs55881710	87327190	0,035	0,937	8,00E-01	0,025	2,537	1,30E-04	DHS	Inaktív Promoter
rs34978260	87328615	0,035	0,937	8,00E-01	0,025	2,537	1,30E-04	-	Gyenge Enhanszer
rs10081630	87328898	0,036	0,955	8,60E-01	0,026	2,365	3,50E-04	-	Gyenge Enhanszer



SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs13292542	87330260	0,035	0,937	8,00E-01	0,025	2,525	1,40E-04	-	Gyenge Enhanszer
rs55888263	87330966	0,035	0,937	8,00E-01	0,025	2,541	1,30E-04	-	Gyenge Enhanszer
rs35205599	87331074	0,035	0,937	8,00E-01	0,025	2,542	1,20E-04	-	Gyenge Enhanszer
rs13293356	87333948	0,035	0,938	8,00E-01	0,025	2,548	1,20E-04	DHS	Gyenge Enhanszer
rs13284421	87338999	0,035	0,954	8,60E-01	0,026	2,478	2,20E-04	-	Gyenge Enhanszer
rs71506660	87339760	0,035	0,956	8,60E-01	0,025	2,589	1,10E-04	-	Gyenge Enhanszer
rs35766327	87340404	0,034	0,957	8,70E-01	0,025	2,592	1,10E-04	-	Gyenge Enhanszer
rs35147532	87340613	0,034	0,957	8,70E-01	0,025	2,593	1,10E-04	-	Gyenge Enhanszer
rs200952203	87341411	0,035	0,955	8,60E-01	0,025	2,561	1,40E-04	-	Gyenge Enhanszer
rs34926972	87341569	0,035	0,942	8,20E-01	0,025	2,573	1,20E-04	-	Gyenge Enhanszer
rs67293254	87342041	0,034	0,957	8,70E-01	0,025	2,597	1,10E-04	-	Gyenge Enhanszer
rs4877286	87343940	0,034	0,957	8,70E-01	0,025	2,601	1,10E-04	-	Gyenge Enhanszer
rs4877287	87344022	0,034	0,957	8,70E-01	0,025	2,601	1,10E-04	-	Gyenge Enhanszer
rs13293214	87345799	0,037	0,935	7,90E-01	0,026	2,473	2,20E-04	-	Gyenge Enhanszer
rs17379500	87346520	0,034	0,957	8,70E-01	0,025	2,606	1,10E-04	-	Gyenge Enhanszer
rs73651112	87347484	0,036	0,948	8,30E-01	0,026	2,497	2,00E-04	-	Gyenge Enhanszer
rs17087580	87347756	0,039	0,854	5,30E-01	0,026	2,476	2,20E-04	-	Gyenge Enhanszer
rs71506661	87348062	0,036	0,931	7,80E-01	0,026	2,476	2,20E-04	-	Gyenge Enhanszer
rs71506662	87348084	0,036	0,931	7,80E-01	0,026	2,476	2,20E-04	-	Gyenge Enhanszer
rs71506663	87348391	0,034	0,954	8,60E-01	0,025	2,608	1,10E-04	-	Gyenge Enhanszer
rs7018602	87350385	0,036	0,916	7,30E-01	0,026	2,424	3,00E-04	-	Gyenge Enhanszer
rs7039299	87350878	0,036	0,91	7,10E-01	0,026	2,484	2,10E-04	-	Gyenge Enhanszer
rs7040077	87351434	0,034	0,93	7,80E-01	0,025	2,611	1,10E-04	-	Gyenge Enhanszer
rs71506664	87352027	0,034	0,927	7,70E-01	0,025	2,617	1,00E-04	-	Gyenge Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs17379611	87353874	0,034	0,922	7,60E-01	0,025	2,618	1,00E-04	-	Gyenge Enhanszer
rs71506666	87353946	0,036	0,899	6,80E-01	0,026	2,488	2,10E-04	-	Gyenge Enhanszer
rs17326762	87357451	0,034	0,917	7,40E-01	0,024	2,621	1,00E-04	DHS	Gyenge Enhanszer
rs75250684	87361440	0,034	0,917	7,40E-01	0,024	2,622	1,00E-04	-	Gyenge Enhanszer
rs2889962	87362039	0,036	0,894	6,60E-01	0,026	2,492	2,10E-04	-	Gyenge Enhanszer
rs1891979	87365866	0,034	0,917	7,40E-01	0,024	2,624	1,00E-04	-	Gyenge Enhanszer
rs36060009	87368417	0,036	0,893	6,60E-01	0,026	2,465	2,40E-04	-	Gyenge Enhanszer
rs57189532	87369508	0,036	0,892	6,60E-01	0,026	2,465	2,40E-04	DHS	Gyenge Enhanszer
rs71506668	87369568	0,034	0,917	7,40E-01	0,024	2,625	1,00E-04	DHS	Gyenge Enhanszer
rs34847678	87370565	0,034	0,917	7,40E-01	0,024	2,625	1,00E-04	DHS	Inaktív Promoter
rs10512153	87371365	0,034	0,917	7,40E-01	0,024	2,625	1,00E-04	DHS	Gyenge Enhanszer
rs11140752	87373318	0,065	1,522	2,20E-02	0,054	1,235	3,00E-01	-	Gyenge Enhanszer
rs34632776	87373375	0,034	0,917	7,40E-01	0,024	2,627	1,00E-04	-	Gyenge Enhanszer
rs17379953	87376728	0,034	0,917	7,40E-01	0,024	2,627	1,00E-04	-	Gyenge Enhanszer
rs17327132	87383571	0,034	0,917	7,40E-01	0,024	2,633	9,90E-05	-	Gyenge Enhanszer
rs12555679	87384174	0,036	0,889	6,50E-01	0,026	2,537	1,60E-04	-	Gyenge Enhanszer
rs201701210	87384507	0,033	0,968	9,00E-01	0,024	2,706	8,10E-05	-	Gyenge Enhanszer
rs71506670	87385603	0,034	0,905	7,00E-01	0,025	2,653	8,70E-05	DHS	Gyenge Enhanszer
rs17087628	87386499	0,036	0,888	6,40E-01	0,026	2,539	1,60E-04	-	Gyenge Enhanszer
rs34811816	87387192	0,034	0,917	7,40E-01	0,024	2,635	9,80E-05	-	Gyenge Enhanszer
rs34184820	87390378	0,036	0,887	6,40E-01	0,026	2,543	1,60E-04	DHS	Gyenge Enhanszer
rs17327368	87390418	0,034	0,917	7,40E-01	0,024	2,636	9,80E-05	DHS	Gyenge Enhanszer
rs71506671	87390805	0,036	0,887	6,40E-01	0,026	2,543	1,60E-04	-	Gyenge Enhanszer
rs34199097	87391272	0,036	0,887	6,40E-01	0,026	2,543	1,60E-04	-	Gyenge Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs34995551	87391445	0,036	0,887	6,40E-01	0,026	2,543	1,60E-04	-	Gyenge Enhanszer
rs56284394	87391777	0,036	0,887	6,40E-01	0,026	2,543	1,60E-04	-	Gyenge Enhanszer
rs56355720	87391887	0,034	0,917	7,40E-01	0,024	2,636	9,80E-05	-	Gyenge Enhanszer
rs12555159	87394832	0,036	0,895	6,70E-01	0,026	2,442	3,00E-04	DHS	Gyenge Enhanszer
rs12555182	87395111	0,036	0,895	6,70E-01	0,025	2,522	1,90E-04	DHS	Gyenge Enhanszer
rs13287600	87395778	0,036	0,895	6,70E-01	0,025	2,522	1,90E-04	-	Gyenge Enhanszer
rs61468746	87395919	0,036	0,895	6,70E-01	0,025	2,522	1,90E-04	-	Gyenge Enhanszer
rs13287010	87396208	0,036	0,894	6,60E-01	0,025	2,522	1,90E-04	-	Gyenge Enhanszer
rs10649715	87396326	0,036	0,903	6,90E-01	0,026	2,456	2,70E-04	-	Gyenge Enhanszer
rs7020597	87396407	0,036	0,893	6,60E-01	0,026	2,443	3,00E-04	DHS	Gyenge Enhanszer
rs13292400	87397008	0,036	0,893	6,60E-01	0,026	2,443	3,00E-04	DHS	Inaktív Promoter
rs17080621	87398104	0,036	0,893	6,60E-01	0,026	2,443	3,00E-04	-	Gyenge Enhanszer
rs35092125	87398796	0,036	0,893	6,60E-01	0,025	2,523	1,90E-04	-	Gyenge Enhanszer
rs17327806	87398906	0,034	0,917	7,40E-01	0,024	2,639	9,80E-05	-	Gyenge Enhanszer
rs67208166	87399760	0,034	0,917	7,40E-01	0,024	2,639	9,80E-05	-	Gyenge Enhanszer
rs13289538	87400217	0,036	0,893	6,60E-01	0,025	2,523	1,90E-04	-	Gyenge Enhanszer
rs12552902	87401314	0,036	0,873	6,00E-01	0,026	2,443	3,00E-04	-	Gyenge Enhanszer
rs35138510	87401351	0,034	0,917	7,40E-01	0,024	2,639	9,80E-05	-	Gyenge Enhanszer
rs13301912	87401977	0,036	0,893	6,60E-01	0,026	2,443	3,00E-04	-	Gyenge Enhanszer
rs13302094	87402136	0,034	0,917	7,40E-01	0,024	2,639	9,80E-05	-	Gyenge Enhanszer
rs12552381	87402990	0,036	0,893	6,60E-01	0,026	2,443	3,00E-04	-	Gyenge Enhanszer
rs12552425	87403399	0,036	0,893	6,60E-01	0,026	2,443	3,00E-04	-	Gyenge Enhanszer
rs13292008	87404070	0,034	0,917	7,40E-01	0,024	2,64	9,80E-05	DHS	Gyenge Enhanszer
rs35565890	87404815	0,034	0,917	7,40E-01	0,025	2,556	1,60E-04	-	Gyenge Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs111529362	87408037	0,073	0,99	9,60E-01	0,067	1,719	1,20E-03	-	Gyenge Enhanszer
rs78459715	87408053	0,034	0,934	8,00E-01	0,024	2,655	9,70E-05	-	Gyenge Enhanszer
rs58425616	87409314	0,077	0,895	5,50E-01	0,069	1,722	1,40E-03	-	Gyenge Enhanszer
rs112857149	87409715	0,034	0,97	9,10E-01	0,024	2,765	5,10E-05	-	Gyenge Enhanszer
rs79903316	87420415	0,036	1,093	7,20E-01	0,027	2,67	4,90E-05	DHS	-
rs112837959	87423279	0,036	1,091	7,20E-01	0,027	2,664	5,00E-05	DHS	-
rs201203220	87426282	0,035	1,085	7,40E-01	0,027	2,657	5,30E-05	-	-
rs45596934	87428619	0,038	1,123	6,30E-01	0,029	2,481	1,40E-04	-	-
rs138535351	87428807	0,035	1,091	7,20E-01	0,027	2,646	5,20E-05	-	Gyenge Enhanszer
rs74356179	87428808	0,035	1,091	7,20E-01	0,027	2,646	5,20E-05	DHS	Gyenge Enhanszer
rs76133834	87432137	0,035	1,089	7,30E-01	0,027	2,637	5,20E-05	-	Gyenge Enhanszer
rs55769032	87432715	0,035	1,088	7,30E-01	0,027	2,636	5,20E-05	-	Gyenge Enhanszer
rs41511747	87434256	0,035	1,064	8,00E-01	0,027	2,636	5,30E-05	DHS	Gyenge Enhanszer
rs17087726	87434460	0,038	1,01	9,70E-01	0,028	2,508	1,10E-04	DHS	Gyenge Enhanszer
rs77956199	87434782	0,037	1,016	9,50E-01	0,028	2,509	1,10E-04	DHS	Gyenge Enhanszer
rs78343279	87436081	0,037	1,015	9,50E-01	0,028	2,508	1,10E-04	-	Gyenge Enhanszer
rs17087734	87437055	0,037	1,014	9,50E-01	0,028	2,504	1,20E-04	-	Gyenge Enhanszer
rs11140768	87437451	0,037	1,013	9,60E-01	0,028	2,503	1,20E-04	DHS	Gyenge Enhanszer
rs7049053	87438018	0,037	1,012	9,60E-01	0,028	2,508	1,10E-04	-	Gyenge Enhanszer
rs112151991	87438199	0,035	1,089	7,30E-01	0,027	2,634	5,30E-05	-	Gyenge Enhanszer
rs4877882	87439911	0,037	1,032	9,00E-01	0,028	2,604	5,60E-05	-	Gyenge Enhanszer
rs10125445	87441607	0,037	1,025	9,20E-01	0,028	2,564	6,50E-05	-	Gyenge Enhanszer
rs141337456	87442000	0,037	1,026	9,20E-01	0,028	2,561	6,50E-05	-	Gyenge Enhanszer
rs77493522	87442207	0,037	1,026	9,20E-01	0,028	2,558	6,60E-05	DHS	Gyenge Enhanszer

SNP	BP	BIP MAF	BIP OR	BIP p-érték	MDD MAF	MDD OR	MDD p-érték	Konzisztens DHS csúcs és kromatin szegmentáció	Kromatin-állapot (asztrocita)
rs148157946	87442663	0,039	0,987	9,60E-01	0,029	2,533	7,60E-05	-	Gyenge Enhanszer
rs10114165	87442791	0,037	1,026	9,20E-01	0,028	2,543	7,20E-05	DHS	Gyenge Enhanszer
rs10114276	87443187	0,037	1,026	9,20E-01	0,028	2,535	7,50E-05	DHS	Gyenge Enhanszer
rs10114941	87443528	0,037	1,026	9,20E-01	0,028	2,537	7,20E-05	-	Gyenge Enhanszer
rs200375099	87443908	0,037	1,026	9,20E-01	0,028	2,535	7,30E-05	-	Gyenge Enhanszer
rs75619049	87445132	0,037	1,026	9,20E-01	0,028	2,532	7,50E-05	DHS	Gyenge Enhanszer
rs4877883	87445806	0,037	1,026	9,20E-01	0,028	2,527	7,70E-05	-	Gyenge Enhanszer
rs10512155	87445916	0,037	1,027	9,10E-01	0,028	2,527	7,70E-05	-	Gyenge Enhanszer
rs60337135	87449367	0,086	1,549	9,50E-03	0,104	1,248	1,40E-01	-	Gyenge Enhanszer
rs58748420	87449747	0,087	1,537	1,10E-02	0,106	1,233	1,60E-01	-	Gyenge Enhanszer
rs17087753	87450180	0,086	1,554	8,90E-03	0,104	1,258	1,30E-01	-	Gyenge Enhanszer
rs149120026	87450451	0,086	1,537	1,10E-02	0,104	1,26	1,30E-01	-	Gyenge Enhanszer
rs11140780	87452780	0,085	1,525	1,20E-02	0,104	1,256	1,30E-01	-	Gyenge Enhanszer
rs11140783	87454817	0,083	1,486	2,00E-02	0,101	1,256	1,40E-01	DHS	Gyenge Enhanszer
rs74899846	87515998	0,077	1,435	4,40E-02	0,088	1,311	1,00E-01	-	Gyenge Enhanszer
rs920776	87538336	0,145	0,722	2,10E-02	0,136	0,963	8,00E-01	-	Gyenge Enhanszer
rs7023976	87544689	0,142	0,714	1,80E-02	0,132	0,931	6,20E-01	-	Gyenge Enhanszer
rs12238534	87547350	0,14	0,719	2,10E-02	0,13	0,924	5,90E-01	-	Gyenge Enhanszer
rs12340212	87551964	0,141	0,716	1,90E-02	0,132	0,929	6,20E-01	DHS	Gyenge Enhanszer
rs6559838	87553563	0,214	0,789	4,80E-02	0,216	0,877	2,80E-01	-	Gyenge Enhanszer
rs12342902	87554440	0,14	0,718	1,90E-02	0,133	0,921	5,70E-01	-	Gyenge Enhanszer
rs12348710	87554605	0,14	0,721	2,10E-02	0,131	0,914	5,40E-01	-	Gyenge Enhanszer
rs2586567	87561978	0,22	0,776	3,40E-02	0,22	0,89	3,30E-01	-	Gyenge Enhanszer
rs1078947	87563252	0,146	0,709	1,40E-02	0,138	0,941	6,70E-01	-	Gyenge Enhanszer

<b>SNP</b>	<b>BP</b>	<b>BIP MAF</b>	<b>BIP OR</b>	<b>BIP p-érték</b>	<b>MDD MAF</b>	<b>MDD OR</b>	<b>MDD p-érték</b>	<b>Konzisztens DHS csúcs és kromatin szegmentáció</b>	<b>Kromatin-állapot (asztrocita)</b>
rs2808713	87564451	0,221	0,773	3,10E-02	0,221	0,897	3,60E-01	-	Gyenge Enhanszer
rs200629680	87565508	0,147	0,705	1,20E-02	0,138	0,94	6,60E-01	-	Gyenge Enhanszer
rs1006446	87566393	0,147	0,702	1,10E-02	0,138	0,94	6,60E-01	-	Gyenge Enhanszer
rs2808706	87567365	0,152	0,713	1,40E-02	0,143	0,975	8,60E-01	DHS	Gyenge Enhanszer

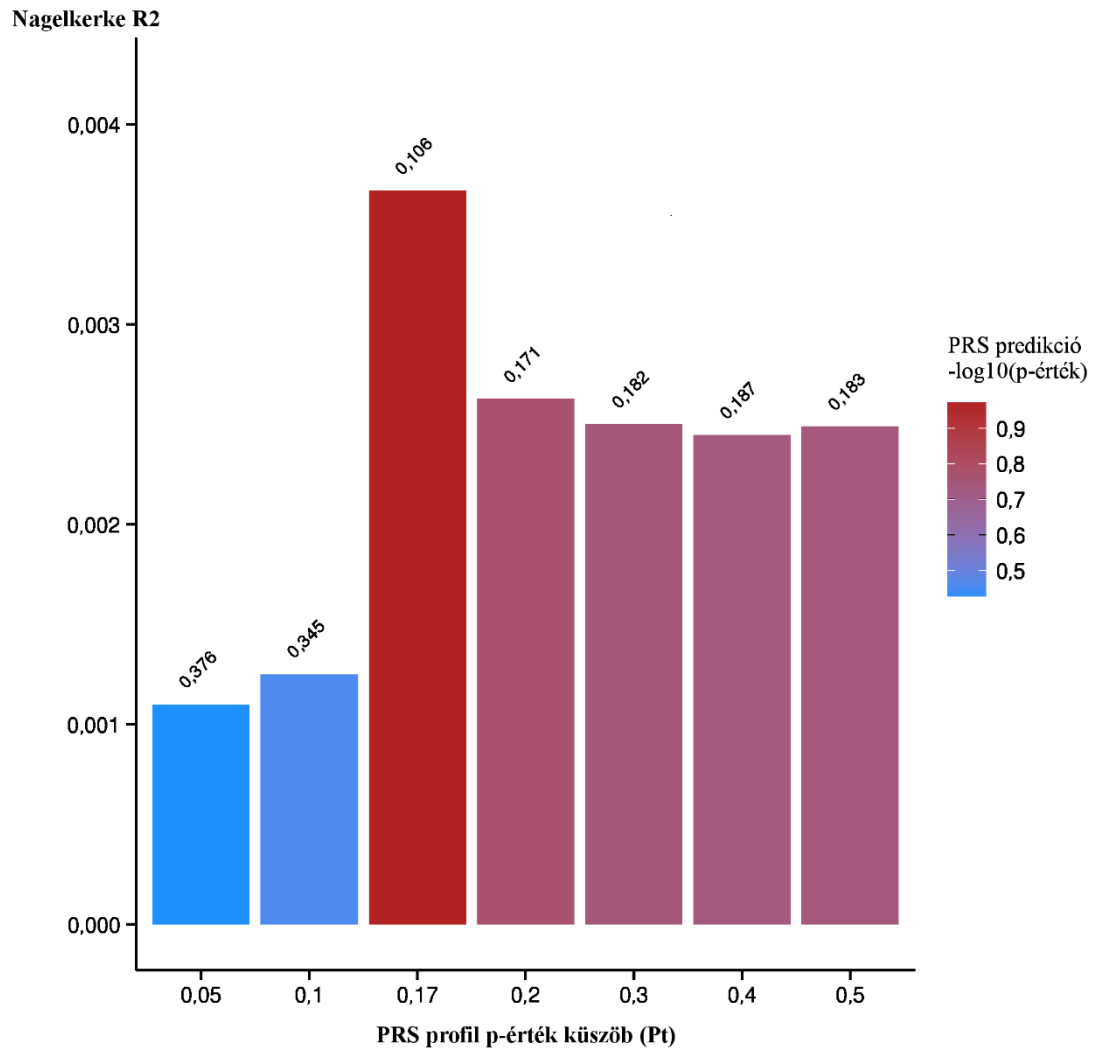
#### 4.1.3. Poligénes rizikópont predikció (PRS)

A három kandidáns génlista LDpred bayesi PRS elemzésének eredményét az **6. táblázat** foglalja össze, míg a PRSice által a kandidáns génlisták markerkészletére optimalizált PRS értékek predikciós elemzésének eredményei a **3-5. ábrákon** láthatók.

Az irodalomkutatáson alapuló korrelációs génlista (Génlista1) esetében a bayesi pontossági és kalibrációs mutatók az infinitezimális modell súlypontjait indikálták, amely az összes markert a kauzális frakcióba sorolja ( $p < 1$ ). A modell a bipoláris mintán Nagelkerke  $R^2 = 0,0028$  varianciát fejez ki  $p < 0,156$  szignifikancia mellett. A PRSice is hasonló, nem szignifikáns ( $p < 0,106$ ) predikciót jelzett a Génlista1  $p_t < 0,17$  küszöbvel definiált, 232 markere által kifejezett variációval (Nagelkerke  $R^2 = 0,0036$ , **3. ábra**).

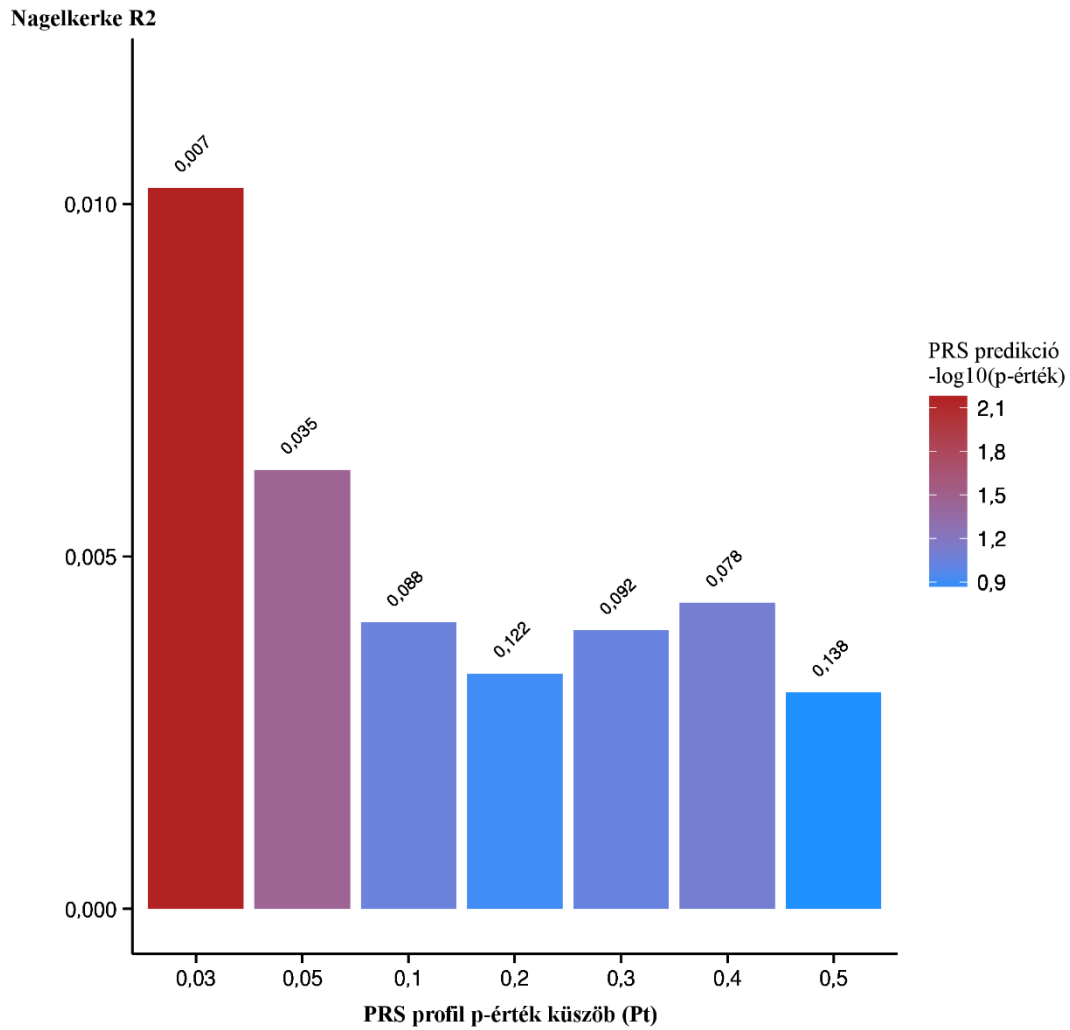
Az miRNS expressziós génlista (Génlista2)  $p < 0,01$  kauzális frakció szerint megállapított poszterior PRS pontjai biztosították a legjobb kalibrációt és pontosságot, amellyel a bipoláris mintában  $p = 0,019$  szignifikanciával jósolták be az öngyilkossági viselkedést (Nagelkerke  $R^2 = 0,0076$  and  $\beta = -1,52$ ). PRSice elemzés ismét nagyon hasonló modellt eredményezett, az optimális PRS pontot  $p_t < 0,03$  küszöbvel definiált 21 markerrel érte el, a bipoláris mintában  $p < 0,007$  szignifikancia mellett a szuicid viselkedés variációjának 1%-át kifejezve (Nagelkerke  $R^2 = 0,01$ , **4. ábra**).

Az miRNS célgéneket is tartalmazó Génlista3 esetében egyik módszer sem jelzett szignifikáns predikciót, az LDpred pontossági és kalibrációs mutatói a  $p < 0,003$  kauzális frakciót indikálták, azonban ez sem jósolta be szignifikánsan a szuicid magatartást a bipoláris mintán (Nagelkerke  $R^2 = 0,0012$ ,  $\beta = -1,14$  és  $p < 0,357$ ). A PRSice megint nagyon hasonló optimumot talált (**5. ábra**), a  $p_t < 0,02$  küszöbvel definiált, 2444 független PRS marker, majdnem megegyező értékekkel jellemzett, nem szignifikáns predikciót mutatott (Nagelkerke  $R^2 = 0,0016$  és  $p < 0,262$ ).

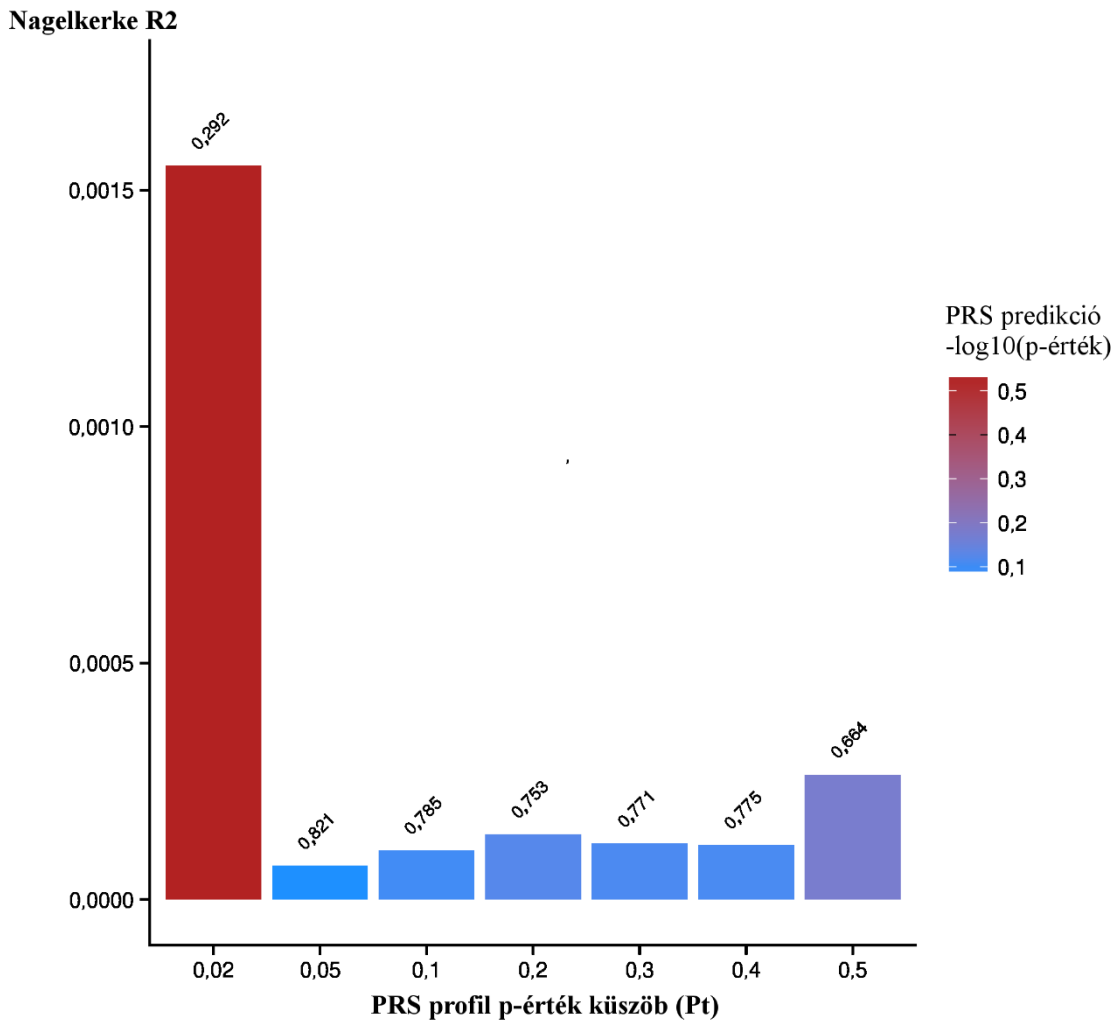


**3. ábra.** A szuicid magatartás PRS predikciója a korrelációs génszett (Génlista1) poligén rizikó profiljai szerint.





**4. ábra.** A szuicid magatartás PRS predikciója a prefrontális pályákon kifejeződő miRNS gének (**Génlista2**) poligénes rizikó profiljai szerint.



**5. ábra.** A szuicid magatartás PRs predikciója a prefrontális pályákon kifejeződő miRNS-ek és célgénjeik (**Génlista3**) poligén rizikó profiljai szerint.

**6. táblázat.** A szuicid magatartás poligénes predikciójának vizsgálata a GAIN bipoláris mintán az LDpred bayesi megközelítésű, információt maximalizáló modellje szerint.

Vizsgált kandidáns génlisták: **Génlista1:** szuicid magatartás korrelációs gészettje, **Génlista2:** a prefrontális pályákon expresszáldó miRNS gének, és **Génlista3:** prefrontális pályákon expresszáldó miRNS gének és feltételezett célgénjeik.

Kauzális markerek frakciója	Génlista1				Génlista2				Génlista3			
	Jóslási		Jósolt		Jóslási		Jósolt		Jóslási		Jósolt	
	Pont. R <sup>2</sup>	Kalib. Slope	Var. R <sup>2</sup>	Hatás β (p-érték)	Pont.s R <sup>2</sup>	Kalib. Slope	Var. R <sup>2</sup>	Hatás β (p-érték)	Pont. R <sup>2</sup>	Kalib. Slope	Var. R <sup>2</sup>	Hatás β (p-érték)
p<0,001	0,0001	0,0001			0,0003	0,271			0,0004	0,0036		
p<0,003	0,0001	<0,0001			0,0005	0,317			<b>0,0009</b>	<b>0,5751</b>	<b>0,0012</b>	<b>-1,14 (0,357)</b>
p<0,005	0,0001	0,0006			0,0000	0,004			<0,0001	0,0025		
p<0,01	0,0001	<0,0001			<b>0,0015</b>	<b>0,326</b>	<b>0,0076</b>	<b>-1,52 (0,019)*</b>	0,0009	0,0681		
p<0,03	0,0001	0,0004			0,0022	0,146			0,0002	0,0680		
p<0,05	0,0001	<0,0001			0,0048	0,032			0,0009	0,0510		
p<0,1	<0,0001	0,0001			0,0002	0,069			0,0006	0,3431		
p<0,3	0,0006	0,0004			0,0040	0,017			0,0009	0,0484		
p<0,5	<0,0001	0,0010			0,0002	0,547			0,0007	0,0007		
p<1,0	<b>0,0008</b>	<b>0,0006</b>	<b>0,0028</b>	<b>-4.58 (0,156)</b>	0,0039	0,016			0,0009	0,0481		

**Jegyzet:** A pontosságot (Pont.) a jóslási R<sup>2</sup>-vel, a kalibrációt (Kalib.) a meredekséggel (Slope), a jósolt varianciát (Var.) a Nagelkerke R<sup>2</sup> alapján számítjuk.

\* p-érték nominálisan szignifikáns (p < 0,05)

## 4.2. A második vizsgálat eredményei

### 4.2.1. Gén-alapú asszociációk

A gén-alapú asszociációs GATES-teszt 1805 gén esetében eredményezett szuggesztív p-értéket az explorációs mintán, közülük 145 gén asszociálódott legalább nominálisan szignifikáns p-értékkel a replikációs mintán is. Nem találtunk átfedést a két minta első 10 génje között a gének p-értékek szerinti rangsorában, és az első helyen álló PTPN21 gén p-értéke sem maradt szignifikáns a hipotéziscsoport szerinti korrekció után ( $p_{\text{exp}} = 0,026$ ,  $p_{\text{rep}} = 6,42\text{E-}06$ ,  $p_{\text{corr}} = 0,113$ ). Az első 10 replikált gén asszociációt az 6. táblázat mutatja.

### 4.2.2. Génszett elemzések: kanonikus útvonal és pozicionális génszett tesztek

A kanonikus útvonalak explorációs lépcsőjében a HYST asszociációs teszttel 60 útvonalat, a Wilcoxon feldúsulási teszttel pedig további 44 útvonalat jelöltünk a replikációs elemzésbe. A replikációs lépcsőben nem találtunk nominálisan feldúsult útvonalat, 6 útvonal ért el nominális p-értéket a HYST asszociációs tesztben, de egyik sem maradt szignifikáns a hipotéziscsoport korrekcióját követően. A nominális szignifikanciával replikált kanonikus útvonalakat a 7. táblázat mutatja be

A pozicionális génszettek elemzésének explorációs lépcsője 38 génszettet azonosított szuggesztív HYST p-értékkel, és 89 génszettet szuggesztív Wilcoxon p-értékkel, összesen 91 génszettet jelölve replikációra. A replikációs mintán 5 kromoszóma szegmens asszociációjának, és velük átfedő 3 szegmens feldúsulásának nominálisan szignifikáns replikációját észleltük, ebből 1 szegmens asszociációja és 2 szegmens feldúsulása a többszörös összehasonlítás korrekcióját követően is szignifikáns maradt (14q31:  $p_{\text{exp\_HYST}} = 0,098$ ,  $p_{\text{rep\_HYST}} = 1,5\text{E-}06$ ,  $p_{\text{rep\_HYST\_corr}} = 0,002$ ; 5q31:  $p_{\text{exp\_Wilcoxon}} = 7,2\text{E-}04$ ,  $p_{\text{rep\_Wilcoxon}} = 5,6\text{E-}06$ ,  $p_{\text{rep\_Wilcoxon\_corr}} = 0,0062$ ; Xq13:  $p_{\text{exp\_Wilcoxon}} = 0,018$ ,  $p_{\text{rep\_Wilcoxon}} = 4,5\text{E-}05$ ,  $p_{\text{rep\_Wilcoxon\_corr}} = 0,048$ ). A nominális szignifikanciával replikált pozicionális génszettek a 8. táblázatban láthatók.

**7. táblázat.** A p-érték rangsorban első 10 gén és transzkript régió-alapú asszociációja az explorációs és a replikációs mintán

<b>Gén (HGNC szimbólum)</b>	<b>Genomikus koordináták (kromoszóma: bázispár)</b>	<b>Explorációs minta (n = 16) p_GATES*</b>	<b>Replikációs minta (n = 5337) p_GATES (korrigált)</b>
Protein tirozin foszfatáz nonreceptor 21 (PTPN21)	14:88932121-89016966	0,0260	6,42E-06 (0,113)
Kromoszóma strukturális fenntartó 1B (RP1-102D24.5)	22:45831743-45844624	0,0455	2,00E-04 (1,00)
Mátrix remodelling asszociált 5 (MXRA5)	X:3226605-3264682	0,0614	5,00E-04 (1,00)
C2H2 típusú cink ujj (ZFP42)	4:188916924-188926199	0,0325	0,0010 (1,00)
RIB43A coiled-coil domén 2	22:45809571-45828376	0,0455	0,0013 (1,00)
Feszültségfüggő Ca <sup>2+</sup> -csatorna alfa2delta4 alegység (CACNA2D4)	12:1901122-2028002	0,0650	0,0015 (1,00)
BCL2 asszociált athanogén 6 (BAG6)	6:31606804-31607565	0,0348	0,0017 (1,00)
Szinaptonémás komplex központi elem fehérje 2(SYCE2)	19:13009599-13030090	0,0707	0,0019 (1,00)
XK asszociált 4 (XKR4)	8:56014948-56454613	0,0252	0,0022 (1,00)
MACC1 antiszenz RNS 1 (MACC1-AS1)	7:20181538-20193154	0,0143	0,0022 (1,00)

p\_GATES: Kiterjesztett Simes'-teszt alapján számított régió-alapú asszociációs p-érték

**8. táblázat.** Nominálisan replikált kanonikus útvonalak és átfedő útvonalakból összeállított, szuperútvonalak asszociációs (HYST) illetve feldúsulási (Wilcoxon) p-értékei az explorációs és a replikációs mintán.

Kanonikus útvonal	Explorációs minta (n = 16)			Replikációs minta (n = 5337)		
	p <sub>HYST</sub>	p <sub>Wilcoxon</sub>	Gének száma	p <sub>HYST</sub> (korrigált)	p <sub>Wilcoxon</sub> (korrigált)	Gének száma
Superpathway 75	0,0086	0,0328	57	0,0039 (1.00)	0,5625 (1.00)	121
Superpathway 106	0,0304	0,0030	61	0,0047 (1.00)	0,7163 (1.00)	99
Superpathway 69	0,0329	0,2020	13	0,0071 (1.00)	0,4661 (1.00)	18
PID_HNF3B_PATHWAY	0,0875	0,0928	23	0,0149 (1.00)	0,5711 (1.00)	44
Superpathway 107	0,0292	0,0130	52	0,0166 (1.00)	0,5723 (1.00)	123
PID_P38_ALPHA_BETA_PATHWAY	0,0769	0,2305	18	0,0168 (1.00)	0,2107 (1.00)	31

**HYST:** hybrid set-based association test, génszett-alapú asszociációs teszt

**Wilcoxon:** Wilcoxon rangösszeg feldúsulási teszt

**Superpathway 69:** REACTOME PD1 SIGNALING + REACTOME PHOSPHORYLATION OF CD3 AND TCR ZETA CHAINS

**Superpathway 75:** KEGG CELL CYCLE + BIOCARTA CELLCYCLE PATHWAY + BIOCARTA G1 PATHWAY

**Superpathway 107:** KEGG OXIDATIVE PHOSPHORYLATION + REACTOME RESPIRATORY ELECTRON TRANSPORT + REACTOME INSULIN RECEPTOR RECYCLING

**Superpathway 106:** REACTOME SIGNALING BY INSULIN RECEPTOR + REACTOME INSULIN RECEPTOR RECYCLING + REACTOME INSULIN RECEPTOR SIGNALLING CASCADE + REACTOME ACTIVATED POINT MUTANTS OF FGFR2 + REACTOME SIGNALING BY ACTIVATED POINT MUTANTS OF FGFR1 + REACTOME SIGNALING BY FGFR3 MUTANTS + REACTOME REGULATION OF AMPK ACTIVITY VIA LKB1 + REACTOME SHC MEDIATED CASCADE + REACTOME ENERGY DEPENDENT REGULATION OF MTOR BY LKB1 AMPK + REACTOME REGULATION OF RHEB GTPASE ACTIVITY BY AMPK + REACTOME PKB MEDIATED EVENTS + REACTOME SIGNAL ATTENUATION + REACTOME SOS MEDIATED SIGNALLING + REACTOME RAF MAP KINASE CASCADE + REACTOME SHC MEDIATED SIGNALLING + REACTOME MTORC1 MEDIATED SIGNALLING + REACTOME FGFR LIGAND BINDING AND ACTIVATION + REACTOME FGFR2C LIGAND BINDING AND ACTIVATION + REACTOME FGFR4 LIGAND BINDING AND ACTIVATION + REACTOME FGFR1 LIGAND BINDING AND ACTIVATION + REACTOME PI3K CASCADE + REACTOME SHC RELATED EVENTS + REACTOME FRS2 MEDIATED CASCADE

**9. táblázat.** Nominálisan replikált pozicionális génszettek asszociációs (HYST) illetve feldúsulási (Wilcoxon) p-értékei az explorációs és a replikációs mintán.

Kromoszómasáv	Explorációs minta (n = 16)			Replikációs minta (n = 5337)		
	p <sub>HYST</sub>	p <sub>Wilcoxon</sub>	Gének száma	p <sub>HYST</sub> (korrigált)	p <sub>Wilcoxon</sub> (korrigált)	Gének száma
14q31	0,0977	0,0374	16	1,54E-06 (0.002)*	0,0436 (1.00)	32
1p22	0,0320	0,0013	50	0,0112 (1.00)	0,1827 (1.00)	81
Xq13	0,0736	0,0181	12	0,0171 (1.00)	4,32E-05 (0,048)*	78
5q31	0,4765	7,20E-04	134	0,0302 (1.00)	5,75E-06 (0,006)*	194
10q23	0,0123	2,87E-06	55	0,0443 (1.00)	0,4366 (1.00)	91
5q12	0,0751	0,0150	29	0,0453 (1.00)	0,2533 (1.00)	43

\*: a p-érték hipotéziscsoportra korrigálva is szignifikáns

**HYST:** hybrid set-based association test, génszett-alapú asszociációs teszt

**Wilcoxon:** Wilcoxon rangösszeg feldúsulási teszt

#### 4.2.3. Funkcionális annotációs klaszterelemzés

Az explorációs minta 402 rizikógénjéből 386, míg a háttérként meghatározott 14.618 gén közül 13.647 rendelkezett DAVID azonosítóval. A replikációs mintában az 1785 rizikógénből 1543, a 38.940 háttér génből pedig 24.517 szerepelt az adatbázisban. Az explorációs minta 306 klaszteréből 12 annotációs klaszter bírt az  $ES > 1,3$  küszöb feletti feldúsulási mutatóval, további 2 annotációs klaszter volt jellemezhető közvetlenül küszöbérték alatti, de még szuggesztívnek tekinthető feldúsulási ponttal ( $ES > 1,2$ ). Azonos annotációs források és szigorúsági kritérium mellett, a replikációs minta rizikógénjeivel a lehetséges 621 klaszterből 126 annotációs klasztert azonosítottunk  $ES > 1,3$  feldúsulással, amelyek közül 7 megegyezett az explorációs mintán  $ES > 1,2$  küszöbértékkel azonosított klaszterekkel. Legnagyobb feldúsulási mutatóval mindkét mintában a splicing és alternatív splicing annotációs terminusok klasztere bírt ( $ES_{exp}: 2,85, ES_{rep}: 23,93$ ). Továbbá, az explorációs mintától eltérő rangsorban, de a replikációs mintában is feldúsultak az endoszómális transzport ( $ES_{exp}: 1,56, ES_{rep}: 1,90$ ), a neuro- és axongenezis ( $ES_{exp}: 1,44, ES_{rep}: 2,25$ ), az izom- és vázrendszer fejlődése ( $ES_{exp}: 1,40, ES_{rep}: 1,95$ ), az embrionális fejlődés ( $ES_{exp}: 1,34, ES_{rep}: 4,77$ ), a motoros viselkedés ( $ES_{exp}: 1,33, ES_{rep}: 2,30$ ) és az agyfejlődés ( $ES_{exp}: 1,23, ES_{rep}: 4,55$ ) funkcionális annotációs terminusaival kapcsolatos klaszterei. Mivel a szuggesztív klaszterek szelektíven nem vizsgálhatók, a klaszteranalízis replikációját a klaszterek összesített számára korrigáltuk ( $\alpha = 0,05/927$ ), ennek megfelelően a  $ES > 4,27$  feldúsulású klaszterek, azaz a splicing, az agyfejlődés és embrionális fejlődés klasztere, korrekció után is szignifikáns maradt.



**10. táblázat.** Statisztikailag szignifikánsan replikált, funkcionális annotációs klaszterek feldúsulási pontjai (ES), és terminusaik feldúsulási p\_EASE értékei az explorációs és replikációs minta prioritizált génjei szerint.

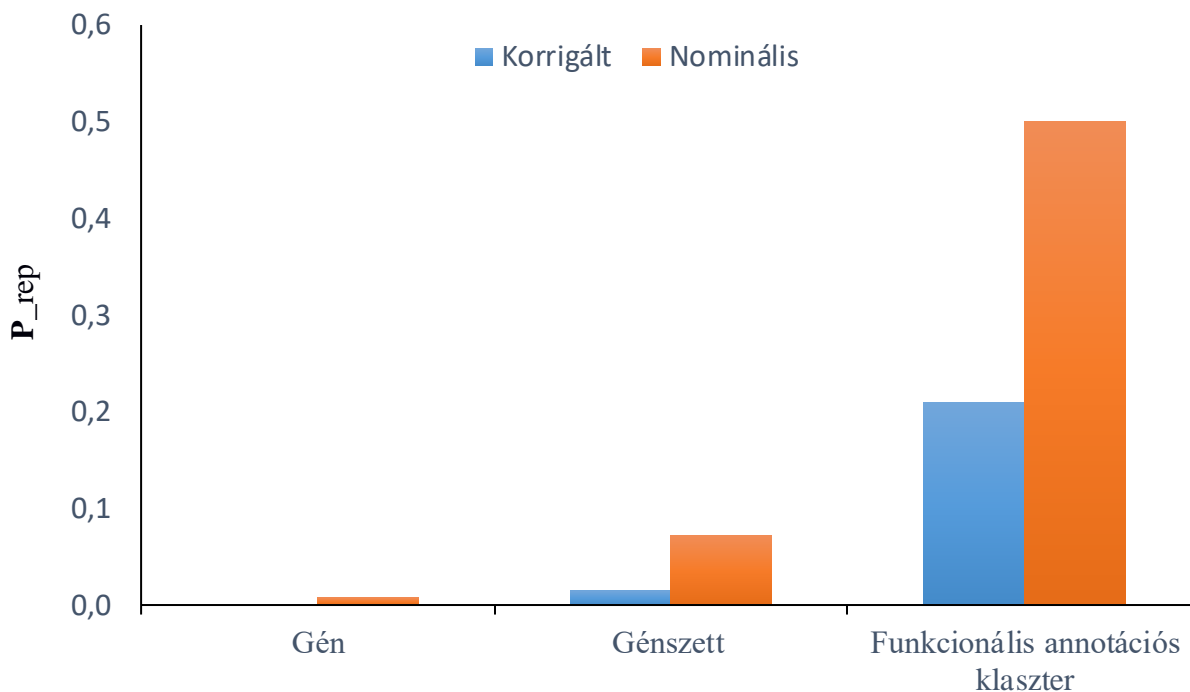
Klaszter / Terminus	Explorációs minta (n <sub>gén</sub> = 386)			Replikációs minta (n <sub>gén</sub> = 1543)		
	ES p_EASE	FDR	Gének száma (%)	ES p_EASE	FDR	Gének száma (%)
<b>1. Klaszter</b>	<b>ES: 2,85</b>			<b>ES: 23,93</b>		
splice variant	2,43E-04	0,0041	200 (51,81%)	2,57E-26	4,80E-25	783 (50,75%)
alternative products	0,0016	0,0132	192 (49,74%)	1,53E-25	2,20E-24	1006 (65,2%)
alternative splicing	0,0071	0,0926	248 (64,25%)	4,10E-22	3,29E-21	755 (48,93%)
<b>2. Klaszter</b>	<b>ES: 1,34</b>			<b>ES: 4,77</b>		
GO:0009790: embryo development	0,0068	0,1209	34 (8,81%)	1,34E-08	2,67E-07	131 (8,49%)
GO:0048598: embryonic morphogenesis	0,0532	0,6445	20 (5,18%)	2,91E-05	5,80E-04	77 (4,99%)
GO:0048568: embryonic organ development	0,0875	0,8230	15 (3,89%)	0,0084	0,1550	51 (3,31%)
<b>3. Klaszter</b>	<b>ES: 1,24</b>			<b>ES: 4,55</b>		
GO:0007417:central nervous system develop	0,0492	0,6147	29 (7,51%)	2,33E-09	4,64E-08	132 (8,55%)
GO:0007420:brain development	0,0630	0,7080	23 (5,96%)	6,05E-07	1,21E-05	99 (6,41%)
GO:0060322: head development	0,0631	0,7085	24 (6,22%)	9,19E-08	1,83E-06	106 (6,87%)
GO:0030900:forebrain development	0,0828	0,8048	14 (3,63%)	1,09E-06	2,18E-05	61 (3,95%)

p\_EASE: a terminus felülreprezentációs elemzésének p-értéke; ES: enrichment score, a klaszter feldúsultsági mutatója, ES =  $-\log_{10}(p_{EASE})$

FDR: false discovery rate, Benjamini-Hochberg szerint

## 4.2.4. Replikációs valószínűségelemzés

Vizsgálatunk másodlagos célkitűzése tekintetében, a gén-alapú tesztek  $P_{rep\_nom\_g} = 0,008$  nominális replikációs rátát és  $P_{rep\_corr\_g} = 0$  többszörös összehasonlításra korrigált replikációt eredményeztek. A génszett-tesztek összesített nominális replikációja  $P_{rep\_nom\_s} = 0,072$ , a többszörös összehasonlításra korrigált pedig  $P_{rep\_corr\_s} = 0,015$ . A funkcionális annotációs klaszterezés tehát  $P_{rep\_nom\_ac} = 0,5$  nominális, illetve  $P_{rep\_corr\_ac} = 0,21$  többszörös összehasonlításra korrigált replikációs valószínűséget eredményezett (6.ábra).



**6. ábra.** A multilokus elemzések nominális, és többszörös összehasonlításra korrigált replikációs valószínűsége ( $P_{rep}$ ) DSM-IV szkizofrénia kétlépcsős elemzésében.

## 5. Megbeszélés

### 5.1. Az eredmények áttekintése

Vizsgálati eredményeink a következő pontokban foglalhatók össze:

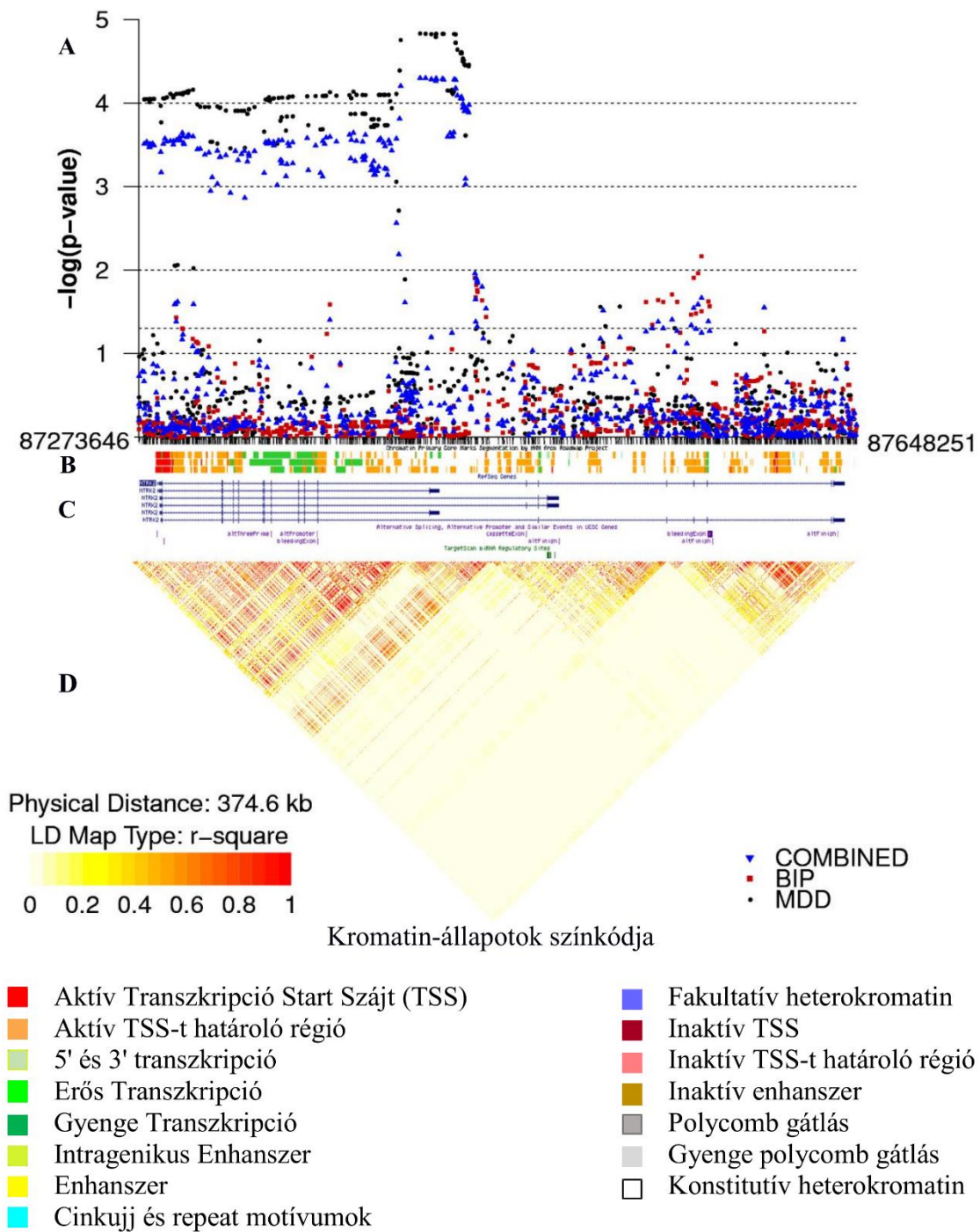
1. A szuicid magatartás transzdiagnostikus, régió-alapú elemzése többszörös összehasonlításra korrigált, statisztikailag szignifikáns asszociációt igazolt a szuicid magatartás és az NTRK2 gén között a major depresszív zavar mintájában.
2. A korrelációs kandidáns gének közül egyedül a DICER1 gén esetében találtunk mindkét mintában legalább nominálisan szignifikáns régió-alapú asszociációt a szuicid magatartással.
3. A három kandidáns génszett közül, a végrehajtó funkciók szervezését végző agyi régiókban expresszálandó miRNS gének poligénes rizikópontjai jósolták be legalább nominális szignifikanciával a szuicid magatartást a validáló mintán. Ugyanakkor a bejósolt változás mindegyik szignifikáns predikciós modellben ellenkező irányú volt.
4. A szkizofrénia multilokus elemzésében a 14q31, Xq13 és a 5q31 kromoszómasávok pozicionális génszettjei és a DSM-IV szkizofrénia között többszörös összehasonlításra korrigáltan is szignifikáns asszociációt találtunk. Ugyanakkor sem a gén-alapú tesztek, sem a kanonikus útvonalak elemzése nem mutatott ki szignifikáns összefüggést.
5. A funkcionális prioritizációs algoritmusunkkal azonosított gének annotációs klaszterelemzése mindkét mintánkban a szkizofrénia jelenlegi etiológiai modelljébe illeszthető klasztereket eredményezett.
6. A kétlépcsős szerkezetű elemzéssel igazoltuk, hogy a multilokus asszociációs elemzések egyértelműen javítják a replikáció esélyét, akár gyökeresen eltérő minták esetén is.

## 5.2. Multilókusz asszociációk szuicid magatartásban

Ebben a vizsgálatunkban a bipoláris zavarhoz és a major depresszióhoz társuló öngyilkossági viselkedés transzdiagnosztikus, kandidáns gén elemzését végeztük, multilókusz régió-alapú asszociációs tesztekkel és poligénes rizikópont predikciós analízissel. A régió-alapú asszociációs tesztek során összesen 7 gén p-értéke bizonyult nominálisan szignifikánsnak, de a többszörös összehasonlítás hipotéziscsoport szerinti korrekcióját követően csak az NTRK2 gén MDD mintában mért p-értéke maradt szignifikáns. Továbbá, a 35 gén közül egyedül a DICER1 gén társult legalább nominális szignifikanciával az öngyilkos viselkedéshez mindkét vizsgálati mintában. A dorzolaterális és orbitofrontális prefrontális pályák régióiban expresszáldó miRNS gének MDD mintán számított, poligénes rizikópontjai, mindkét statisztikai módszer szerint, szignifikánsan bejósolták a szuicid magatartást a bipoláris mintában, bár inverz irányban, és csak szerény kifejezett fenotípusos varianciával (legjobb Nagelkerke  $R^2=0,01$ ). Egyik PRS elemzéssel sem találtunk szignifikáns predikciót sem a korrelációs kandidáns génlistával, sem az miRNS-ek célgénjeit tartalmazó, kiterjesztett miRNS készlettel kapcsolatban.

Az NTRK2 gén és az öngyilkossági viselkedés vizsgálatunkban észlelt, statisztikailag szignifikáns asszociációja megfelel a biológiai funkciójáról szerzett ismereteinknek és a korábbi vizsgálatok eredményeinek is. Az NTRK2 gén a BDNF (brain derived neurotrophic factor) egyik nagy-affinitású tirozin-kináz receptorát kódolja (TrkB). A neurotrofinok – különösen a BDNF – létfontosságúak az idegsejtek növekedéshez, differenciálódáshoz, éréshez és túléléshez, továbbá a szinaptikus plaszticitás alapvető modulátorai [189]. A neurotrofinoknak, különösen a CREB1-BDNF-NTRK2 útvonalnak, több, független asszociációs vizsgálat [140–142,190] és génexpressziós tanulmány [137–139,191] is kiemelt szerepet tulajdonított a stresszre és a terápiára adott válaszban, a hangulatzavarok patofiziológiájában, és az öngyilkossági viselkedésben is. Kohli és munkatársai, major depresszióban szenvedő vizsgálati alanyok öngyilkossági kísérleteit elemezve, az NTRK2 gén szignifikáns régió-alapú asszociációjáról, és négy jelölő (tagging) markerének szignifikáns kombinált asszociációjáról számoltak be [140]. Az erős, a fogékonysági skálán 0,01 genomikus heritabilitást kifejező, gén-alapú asszociáció, amelyet az NTRK2 és a szuicid magatartás között találtunk az MDD mintában, egybevág a fenti eredményekkel. Bár nem

ugyanazokat az SNP asszociációkat replikáltuk, az MDD mintában a legerősebb asszociációs jeleket a 16. exon közelében, a gén 3' végéhez közel észleltük, a bipoláris mintában pedig kisebb intenzitással a 20-22. exonokhoz közel (**7. ábra**). Ebben a régióban az alternatív splicing a TrkB fehérje csonka, TrkB-Shc és TrkB-T1 izoformáit eredményezi, melyekről hiányzik a tirozin-kináz domén és a foszfolipáz C kötőhely, sőt, a TrkB-T1 esetében még az Shc kötőhely is [136,139]. A csonkolt TrkB izoformák fontos szabályozó szerepet tulajdonítanak a neurotróf jelátvitelben [192], ugyanis a prefrontális kéregben megváltozott mRNS koncentrációjukat észlelték szkizofréniában [193], hangulatzavarokban [193,194] és befejezett öngyilkosságokban [143,146]. Az MDD mintában emellett számos, mérsékelt erősségű asszociációs jelet észleltünk a gén 5' végén is, ahol több, alternatív splice variáns és alternatív promoter is ismert, tovább valószínűsítve a TrkB izoformák lehetséges szerepét a szuicid magatartás kialakulásában. Végül a nominálisan asszociálódó markerek nagyarányú (23%) annotációja a neuron-specifikus kromatin állapotokkal, illetve az uniform DHS, promoter és enhanszer térképpel [188] (**4. táblázat**), is az NTRK2 asszociációjának érvényességét erősíti. Emellett megerősíti a multilokusz tesztek használhatóságát is, hiszen egy SNP alapú elemzésben egyik marker sem maradt volna statisztikailag szignifikáns. Ugyanakkor az epigenetikai annotációk ismeretében, a zajtól pusztán p-értékük alapján nem elkülöníthető SNP asszociációk kiszűrése nem kevés valószínű asszociáció elvesztésével járna.



**7. ábra.** Az NTRK2 gén SNP asszociációi, epigenetikai és genomikai funkcionális annotációi az MDD, a bipoláris és a kombinált mintákban.

**A:** SNP asszociációk  $-\log_{10}(p\text{-érték})$ , **B:** 15 szintű kromatin-állapot annotációja 3 releváns agyi régióból: 1. n. caudatus anterior, 2. gyrus cingularis, 3. frontális kéreg, **C:** mRNS izoformák, splice variánsok és miRNS kötőhelyek annotációja, **D:** SNP kapcsoltsági térkép  $r^2$  alapján.

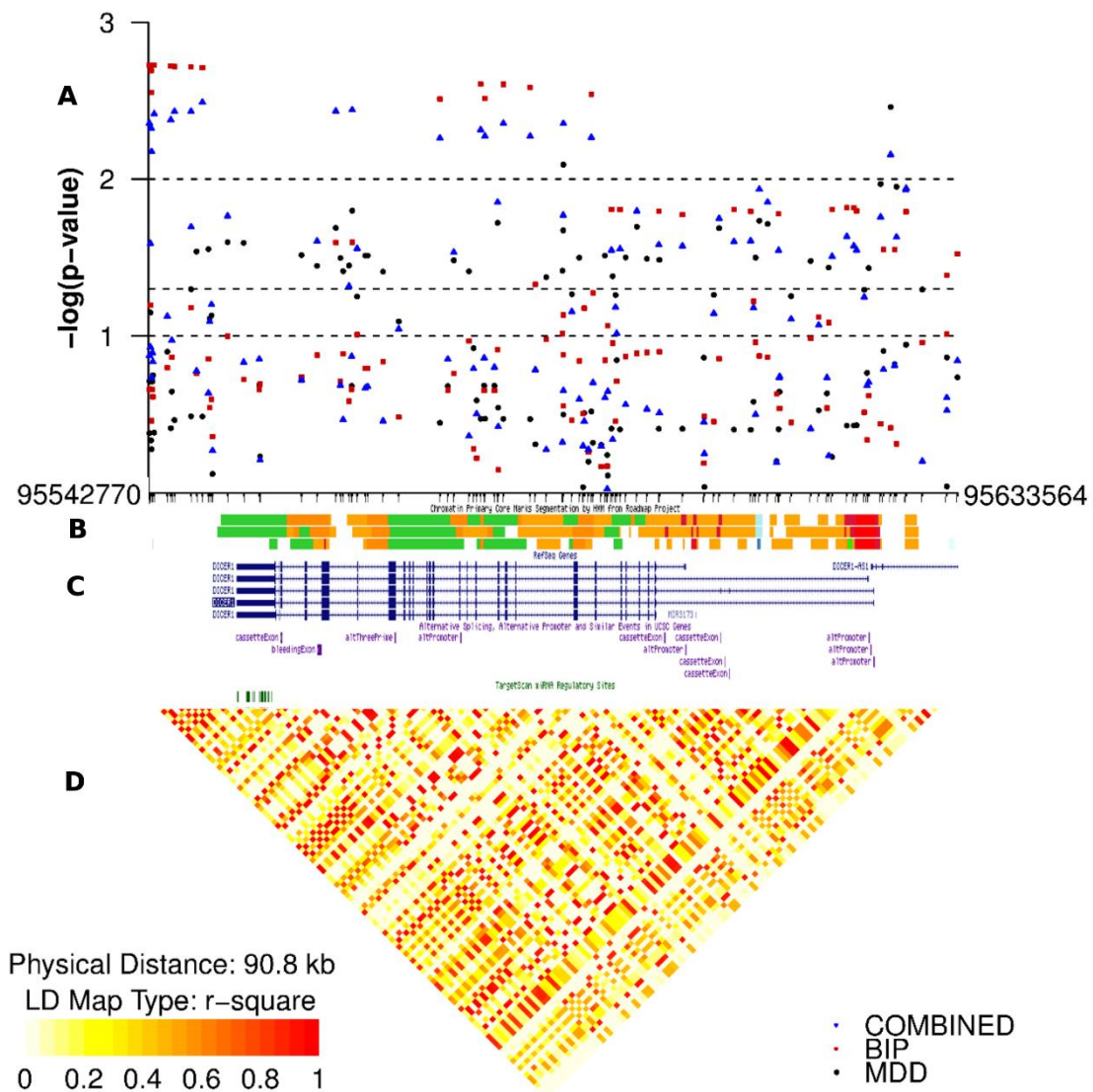
A bipoláris mintában viszont még névlegesen szignifikáns, gén-alapú asszociációt sem tudunk kimutatni az NTRK2 és a szuicid viselkedés között, bár nominálisan szignifikáns SNP asszociációkat észleltünk a fent említett alternatív splicing helyek közelében. Az asszociáció hiánya feltehetően a minták elsődleges vizsgálatainak módszertani különbségére, vagy a jelen vizsgálatunkban is fennálló, fenotípusos heterogenitásra vezethető vissza. Lehetséges okként még a mintavételt meghatározó, elsődleges fenotípusok eltérő etiológiája is felmerül, a környezeti hatások valószínűleg különböző arányú előfordulása, pl. a gyermekkori traumák és az NTRK2 gén-környezeti kölcsönhatása depresszióban [142]. Végezetül, a bipoláris minta kisebb elemszáma, következésképpen kisebb statisztikai ereje is hozzájárult az asszociációk elmaradásához.

A kandidáns gének legtöbbször, például a BDNF-nél, mindkét mintában tapasztalható asszociáció hiánya érdekes, de nem meglepő. Az elemzésünkben használt HYST teszt az „SNP kombináló” jellegének köszönhetően, a sok nominálisan szignifikáns, és mérsékelten független asszociációs jelet tartalmazó régiók asszociációs jelet felerősíti, így ellensúlyozva a helyi lókus heterogenitás hatását. Ugyanakkor nagyszámú neutrális variánst és csak kevés DSL lókuszt tartalmazó régiók (pl. BDNF) esetében a tesztől csak gyenge régiós asszociáció várható. A BDNF gén kapcsoltsági szerkezetén alapuló szimulációnk mindkét mintában alacsony statisztikai erőt jelzett (0,16-0,29) 1-3 DSL lókuszt esetén. Tehát a szignifikáns régió-alapú asszociáció hiánya nem érvényteleníti a korábbi SNP-alapú, vagy „legjobb SNP” technikájú gén-alapú vizsgálatok pozitív eredményét.

A kandidáns gének közül csak a DICER1 gén társult az öngyilkos viselkedéssel mindkét mintában, habár egyik mintában sem haladta meg az asszociációk erőssége a nominális szintet. A gén által kódolt dicer fehérje a klasszikus miRNS bioszintézis sebességkorlátozó enzime. A miRNS-ek, azáltal, hogy kötődnek a cél mRNS-hez és a RISC komplexen vagy a transzláció közvetlen gátlásán keresztül hozzájárulnak annak lebomlásához, módosítják vagy csökkentik a génextpressziót a transzkripció utáni szinten. Ezáltal a miRNS-ek kritikus szerepet játszanak a fejlődési folyamatokban, ide értve az idegrendszeri fejlődést és a szinaptikus plaszticitást is, és a dicer fehérjével együtt számos különböző neuropszichiátriai betegség [144,145,195,196] és prefrontális kortikális funkciók [197–199] patomechanizmusával hozták őket összefüggésbe. A szuicid viselkedés és végrehajtó funkciók zavara közötti szoros kapcsolattal egybehangzóan,

Smalheiser és munkatársai posztmortem expressziós vizsgálata az miRNS-ek globálisan csökkent expresszióját mutatta ki depresszós, öngyilkoságot elkövetett személyek prefrontális agykérgében [200]. A fenotípussal való asszociáció mellett, Huang és munkatársai in vitro vizsgálata [201] egy lehetséges molekuláris patomechanizmust is felvázolt. Eredményeik szerint, a fehérjeszintézis BDNF-hez kapcsolódó gyors változása, és ennek hatása a szinaptikus plaszticitásra a hippocampusz idegsejtjeiben a dicer enzim poszttranszkripció aktiválásához és az miRNS biogenesis következményes fokozódásához köthető. Az aktivált BDNF receptor, TrkB aktiválja a MAPK/ERK jelátviteli utat is, majd a HIV-1 TAR RNS kötő fehérje (TARBP) következményes, Erk-függő foszforilációja stabilizálja a dicer komplexet és az RNS-t processzáló P-testeket. A DICER1 gén expressziójának kiütése teljesen megszüntette a BDNF által kiváltott P-test növekedését, míg a DICER1 transzkripció gátlása nem volt hatással a BDNF válaszra. A DICER1 által közvetített BDNF válasz, dominánsan poszttranszkripciónak tűnő szabályozása fényében érdekes, hogy a legerősebb SNP asszociációs jeleket mindkét mintánkban az miRNS dokkoló helyeket is magába foglaló haplotípus blokkban azonosítottuk (**8. ábra**). Elképzelhető, hogy a szignifikáns markerek nem genotípezált variánsokat jelölnek, amelyek az miRNS kötőhelyeket, és ezzel a dicer transzlációját megváltoztatják. Továbbá, számos nominálisan szignifikáns markert azonosítottunk lehetséges szabályozó funkcióval az ENCODE annotációk alapján is. Ám ezeknek a variánsok pontos funkcionális jelentősége még mindig nem világos, és mivel csak nominálisan szignifikánsak, nem vethető el a véletlen találat valószínűsége, ezért további vizsgálatok szükségesek, hogy megerősítsék eredményeinket.





## Kromatin-állapotok színekódja

- |   |                                |
|---|--------------------------------|
| ■ Aktív Transzkripció Start Szájt (TSS) | ■ Fakultatív heterokromatin    |
| ■ Aktív TSS-t határoló régió            | ■ Inaktív TSS                  |
| ■ 5' és 3' transzkripció                | ■ Inaktív TSS-t határoló régió |
| ■ Erős Transzkripció                    | ■ Inaktív enhanszer            |
| ■ Gyenge Transzkripció                  | ■ Polycomb gátlás              |
| ■ Intragenikus Enhanszer                | ■ Gyenge polycomb gátlás       |
| ■ Enhanszer                             | ■ Konstitutív heterokromatin   |
| ■ Cinkujj és repeat motívumok           |                                |

**8. ábra.** A DICER1 gén SNP asszociációi, epigenetikai és genomikai funkcionális annotációi az MDD, a bipoláris és a kombinált mintákban.

**A:** SNP asszociációk  $-\log_{10}(p\text{-érték})$ , **B:** 15-szintű kromatin-állapot annotációja 3 releváns agyi régióból: 1. n. caudatus anterior, 2. gyrus cingularis, 3. frontális kéreg, **C:** mRNS izoformák, splice variánsok és miRNS kötőhelyek annotációja, **D:** SNP kapcsoltsági térkép  $r^2$  alapján.

A DICER1 gén lehetséges szintetikus asszociációin kívül, az miRNS rendszer szuicid magatartásban betöltött szerepét a poligénes profil elemzés is alátámaszthatja. Mint az az **5. táblázatban** és a **3-5. ábrákon** látható, a bipoláris mintán az MDD mintából a kiválasztott miRNS készletre végzett poligenikus profil elemzés szignifikánsan jósolta az szuicid magatartást - habár negatív bétákkal. Az miRNS génszettet a végrehajtó funkcióhoz tartozó neuroanatómiai képletek expressziós profilja alapján választottuk. Végrehajtó-funkciós deficitet mind az öngyilkossági gondolatok, mind az öngyilkossági viselkedés során kimutattak, ahogy azt is, hogy a deficit független a háttérben húzódó pszichiátriai betegségtől. Újabb keletű állatkísérletes bizonyítékokból arra következtethetünk, hogy a fő miRNS processzálo enzimek (DGCR8, DROSHA and DICER1) elégtelensége megváltoztatja a kognitív és végrehajtó funkciókat [197,198]. Ezért lehetséges, hogy a vizsgálatunkban azonosított, szignifikáns PRS predikciók, és az miRNS-ek gén-alapú asszociációi valójában a végrehajtó funkciózavarral áll kapcsolatban. Mindazonáltal, a szerény hatáserősség és kifejezett fenotípusos variancia, valamint a negatív predikciók utalhatnak, az öngyilkossági gondolatok és kísérletek komplex jellegére, azaz a neuropszichológiai és genetikai háttér nagyfokú különbségeire. A korrelációs és a kiterjesztett miRNS célgén készlet hiányzó predikciói, valószínűleg ebből a fenotipikus heterogenitásból és a viszonylag kis mintaszámból eredhet.

Vizsgálatunk nem mentes a korlátoktól sem. Elemzésünk talán legfontosabb gyengesége a viszonylag kis elemszám, különösen a bipoláris mintában. A régió-alapú asszociációk statisztikai erejének poszterior elemzése megerősítette, hogy vizsgálatunknak a szuboptimális paraméterekkel bíró géne (pl. BDNF) asszociációs elemzéséhez nincs elegendő statisztikai ereje. Mindazonáltal, az elemzésből szintén következik, hogy viszonylag független kockázati lokuszokkal rendelkező géneknél az vizsgálat ereje elfogadható volt, még kis hatáserősségű DSL variánsok esetében is. Mindkét mintához az adatokat visszamenőleg gyűjtöttük, a páciensek mintáit értékelve. Ez a mintavételi mód elkerülhetetlenül torzítja az eredményeket, mert így csak az öngyilkosságot túlélők kerülnek be a vizsgálatba, és az egyik a legfontosabb populációba tartozók, a beféjezett öngyilkosságot elkövetők maradnak ki. Sajnos az adatelemzés során már nem lehetséges a szerkezetből eredő torzítás kijavítása, ezért különös figyelmet kell szentelnünk azon a tanulmányok eredményeinek, amelyek öngyilkosság következtében elhunyt személyekkel foglalkoznak.

Másik probléma volt a szuicid viselkedés eltérő meghatározása, amely fenotipizálásra használt eszközök különbségéből eredt. Míg a DIGS interjú az aktuális öngyilkossági kísérleteket vette fel és értékelte a bipoláris mintán, az IDS kérdőív és a BDI skála, amelyet az MDD adatbázison használtunk, a szuicid ideáció és szándék súlyosságát becsülte meg, közöttük a konkrét öngyilkossági terveket és kísérleteket is, melyek a legsúlyosabb forma kritériumai között szerepeltek. A fenotípusos definíciók különbsége miatt a hiányzó asszociációk értelmezése (például az NTRK2 gén asszociációja) igen problémás. Nagyobb elemszámmal és egységesebb szuicid fenotípussal rendelkező további vizsgálatok szükségesek, hogy megerősítsék, vagy elvessék az asszociáció betegség-specifikus jellegét.

Összefoglalva fő eredményeinket, vizsgálatunk régió-alapú, multilokusz asszociációs teszthei és poligén rizikó-pont predikciós elemzése megerősítették a neurotrofinok szerepét a szuicid viselkedés patogenezisében. A neurobiológiai ismereteken alapuló, miRNS génszett nominálisan szignifikáns, de inverz transzdiagnosztikus poligenikus predikciói, és a DICER1 mindkét mintában nominálisan szignifikáns asszociója azt sugallják, hogy a végrehajtó funkciózavar szerepet játszhat mind az öngyilkossági gondolatok, mind a kísérletek, előfordulásában. Azonban a miRNS rendszer és a szuicid magatartás kapcsolata összetett, és valószínűleg a gén-környezet interakciók, a genetikai heterogenitás és a pleiotrópia nagyban befolyásolja. Bár az eredményeink értelmezését a fenti megszorítások korlátozzák, plauzibilis biológiai hátterük további replikációt indokol.

### 5.3. Multilókusz asszociációk szkizofréniában

Második vizsgálatunkban a DSM-IV szkizofréniás multilókusz markereinek, régió-alapú asszociációját, valamint génszett-alapú feldúsulását, és a feldúsult funkcionális annotációk klasztereit elemeztük, a GENCODE annotáció alapján, kétlépcsős felépítésben. A SCHIZOBANK eset-trió mintán exploratív elemzéssel azonosított gének és kanonikus útvonalak közül, számos asszociációt, valamint feldúsulást replikáltunk nominális p-értékkel a GAIN-MGS GWAS mintán, azonban a többszörös összehasonlítás korrekcióját követően ezek egyike sem szignifikáns, a legerősebb korrigált p-értékkel ( $p = 0,113$ ) a PTPN21 gén rendelkezett. A pozicionális génszettek elemzésekor, az explorációs mintán a 10q23 kromoszómasáv feldúsulása, a replikációs mintán pedig a PTPN21 gént is tartalmazó, 14q31 szegmens régió-alapú asszociációja, valamint a 5q31 és az Xq13 kromoszómaszegmensek feldúsulási tesztje esetében észleltünk - többszörös összehasonlításra korrigáltan is - szignifikáns p-értéket. A két mintán azonos paraméterű predikciós algoritmusokkal prioritizált génlisták funkcionális annotációs klaszterelemzése, mindkét mintán a szkizofréniás feltételezett neurobiológiai hátterére utalhat, pl. sejtosztódás és transzkripció szabályozásával, neuron- és axonfejlődéssel, agy- és embrionális fejlődéssel, illetve szinaptikus plaszticitással kapcsolatos annotációs klaszterek feldúsulását mutatta. Az explorációs mintán azonosított, 14 annotációs klaszter fele a replikációs mintán is nominális feldúsulást mutatott, közülük három, a szkizofréniás etiológiájához jól illeszkedő splicing/alternatív splicing, az agyfejlődés és az embrionális fejlődés klasztere, statisztikailag szignifikáns maradt a teljes hipotéziscsoportnak megfelelő p-érték korrekció után is. A vizsgált multilókusz „markerek” komplexitásának növekedésével mind a nominális, mind a korrigált replikációs valószínűség növekedését észleltük, amely megerősíti a multilókusz asszociációs módszerek genetikai heterogenitást ellensúlyozó hatását.

Az explorációs lépcsőben szuggesztív p-értékkel bíró gének közül, a 14q31 szakaszon elhelyezkedő PTPN21, protein tirozin foszfátáz gén asszociálódott a szkizofréniához a legalacsonyabb p-értékkel ( $p_{rep} = 6,42E-06$ ) a replikációs mintában. Bár gén-alapú tesztek hipotéziscsoportjára korrigált p-értéke csak szuggesztív övezetbe sorolható ( $p_{rep\_corr} = 0,113$ ), az asszociáció érvényessége mellett szól, hogy a gént tartalmazó 14q31 pozicionális génszett asszociációja a korrekció után is szignifikáns

maradt ( $p_{\text{rep\_HYST}}=1,5E-06$ ,  $p_{\text{rep\_HYST\_corr}}=0,002$ ). A PTPN21 és a szkizofrénia kapcsolatát korábbi kutatások is felvetették. Chen és mtsai [202] a szkizofrénia és a PTPN21 és a közeli EML5 gének, potenciálisan funkcionális nukleotid polimorfizmusok szignifikáns asszociációját közölték. Jóllehet az explorációs mintájuk részben átfed jelen vizsgálat replikációs mintáival, a PTPN21 SNP asszociációit meta-analízissel összevont, 13 független mintán is sikeresen replikálták. Bár a PTPN21 gén funkciója még nem teljesen tisztázott, a PTPN gének által kódolt protein tirozin foszfátázok a T-sejtes immunválasz, valamint sejtnövekedés fontos szabályozó faktorai. Továbbá, Plani-Lam és mtsai [203] szerint a PTPN21 meghatározó a neuregulin3 (NRGN3) neurotróf faktor aktiválásában, amely meghosszabbítja a kérgi neuronok túlélését, és javítja a szinaptikus plaszticitást is. Tehát a PTPN21 a szkizofrénia több patomechanizmusában is szerepet játszhat, pl. az idegfejlődésfejlődés, szinaptikus plaszticitás zavarában, vagy a neuroinflammációban, azaz a gén asszociációjának érvényességét a kromoszómarégió asszociációja mellett etiológiai plauzibilitása is támogatja.

Az MSiG v5.2 adatbázis független kanonikus útvonalai és az átfedőkből összeállított „szuperútvonalak” sem az explorációs lépésben, sem replikációkor nem értek el többszörös összehasonlításra korrigált, statisztikailag szignifikáns asszociációt vagy felmásulást, hasonlóan a Psychiatric Genomics Consortium útvonalelemzéséhez. Hat útvonal esetében találtunk nominálisan szignifikáns replikációt, legalacsonyabb asszociációs p-értékkel mindkét lépcsőben a KEGG és BIOCARTA sejtciklus és BIOCARTA G1 útvonalakat tömörítő Superpathway 75 rendelkezett. A sejtciklus, DNS replikáció és transzkripció szabályozásával kapcsolatos útvonalak, a sejt szintű stresszorokra adott válaszhoz, az apoptózis, növekedés és sejt differenciálódás folyamataihoz is szorosan kapcsolódnak, közös pontot jelentve a szkizofrénia idegfejlődési illetve immunológiai eredetelméleteiben [204]. A további nominális útvonalak közül, a REACTOME programozott sejthalál, CD3 és T-sejtek zeta lánc foszforilását tömörítő Superpathway 69, valamint p38 $\alpha$  és p38 $\beta$  MAPK szignál-transzdukciós útvonal asszociációi szintén az apoptózis, neurondifferenciálódás és gyulladós immunválasz érintettségére utalnak. A HNF3B (FOXA2) útvonal asszociációjának biológiai jelentőségét pedig a FOXA1/A2 transzkripció faktorok adják, amelyek központi szabályozó szereppel bírnak a közepagy dopaminerg neuronok fejlődésének minden fázisában [205]. Végezetül, az inzulin szignál-transzdukció és

szabályozás útvonalait aggregáló Superpathway 106, illetve a bioenergetikában meghatározó jelentőségű oxidatív foszforiláció és elektron transzport útvonalait összegző Superpathway 107 nominális asszociációja összhangban van a szkizofrénia és metabolikus szindrómák, pl. cukorbetegség korrelációjával [204,206,207]. Mindazonáltal, a többszörös összehasonlítás korrekciója után, egyik kanonikus útvonal asszociációja sem maradt szignifikáns, ezért a plauzibilisnek tűnő asszociációk esetében is további replikációk szükségesek.

A kromoszómaszegmensek által definiált pozicionális génszettek esetében viszont a 14q31 szegmens asszociációját, valamint a 5q31 és Xq13 szegmensek feldúsulását korrekció után is szignifikáns  $p$ -értékkel replikáltuk. A 14q31 szegmensben, a korábban tárgyalt PTPN21 gén mellett, az EML5, ZC3H14, SPATA7, TTC8 és SEL1L gének esetében mértünk nominálisan szignifikáns replikációt. Az SZGR2 adatbázis [208] szerint, az SEL1L gén kivételével, mindegyik gén esetében publikáltak gyakori SNP asszociációt a szkizofréniaival, emellett a ZC3H14 és a SPATA7 géneket differenciált metiláció, a TTC8 gént pedig differenciált expresszió is a szkizofréniahoz kapcsolja. Funkcióik alapján, az TTC8, EML5 és a SPATA7 a citoskeleton megfelelő működésnek fenntartásával, TTC8 az axogenezis, a ZC3H14 pedig mRNA stabilizáló hatásával szükségesek idegrendszer fejlődéshez. A régióban található, szkizofréniaival korábban összefüggésbe hozott, STON2 és NRXN3 gének asszociációja egyik mintában sem érte el a  $p < 0,1$  küszöböt.

A 5q31 és Xq13 kromoszómaregiók esetében a feldúsulást mérő Wilcoxon  $p$ -értékük maradt a korrekció után is szignifikáns. Mivel a feldúsulással a génszeten belüli és kívüli génasszociációk eloszlásának különbségét teszteljük, a szisztematikus jellege miatt nem lehet csak néhány gént kiemelni. Ezért az egész génszett korrelációját vizsgáltuk az MSiGDB v5.2 génontológiai (GO) annotációival. Az 5q31 szegmens génei a sejtadhézió, szinapszis szerveződés, szinaptikus jelátadás, kalcium kötés és kalciumfüggő sejtadhézió GO terminusaival korrelált  $FDR < 0,05$  korrekció mellett, összhangban a PGC szkizofrénia mega-analízis eredményeivel. Az Xq13 sáv génszettje pedig az RNS polimeráz II transzkripció regulációja és acetyl-transzferáz funkció terminusaival korrelált ( $FDR < 0,05$ ).

A statisztikailag szignifikáns replikációk mellett még a 10q23, 5p12 és 1p22 szegmensek asszociációi replikálódtak nominális szignifikanciával. Bár a 10q23 és az 1p22

szegmensek a korábbi, teljes-genom kapcsoltsági vizsgálatok metaanalízisével kijelölt régiók közelében helyezkednek el, sem kiemelkedő asszociációjú gént, sem szignifikáns ontológiai terminust észleltünk, ezért esetükben a véletlen asszociáció esélye nem elhanyagolható.

A funkcionális annotációk elemzése előtt, a KGGseq program gépi tanuláson és mélytanuláson alapuló predikcióit és az EIGEN adatbázis annotációját kombinálva, a funkcionálisnak jósló kódoló (mendeli predikció [184]), vagy nem-kódoló (regulátoros predikció [185]) variánsok alapján prioritizáltuk a minták genomjait. A prioritizált génlistákhoz kapcsolódó, funkcionális annotációk klaszterelemzésekor, az explorációs minta nominális feldúsulást mutató klaszterei közül, három esetben regisztráltunk többszörös összehasonlítás korrekciójának is megfelelő feldúsulási pontot: a splicing/alternatív splicing, az embrionális fejlődés és az ideg/agyfejlődés annotációs klaszternél. Mindhárom klaszter feldúsulása illeszkedik a szkizofrénia neuronális fejlődéssel kapcsolatos betegségmodelljébe. A splicing/alternatív splicing a transzkripció kromatinstátusztól független szabályozó mechanizmusa, amely neuronális fejlődés mellett a gyulladásos citokinek szabályozó faktora is, tehát a szkizofrénia neuroinflammációs etiológiáját is erősítheti. Ugyanakkor, nem zárhatjuk ki, hogy a magas ES pont ellenére a klaszter csak műtermék, a prioritizációs algoritmusok a regulátoros szájtók felismeréséhez a splicing szájtók gépi tanuláson alapuló annotációját is tartalmazták, a mely vezethetett a splicing/alternatív splicing feldúsulásához. Végezetül megemlítenéd, hogy a szkizofrénia etiológiájában jelen ismereteink szerint központi szerepet játszó neurotranszmisszió, szinaptikus funkciók és plaszticitás zavarára utaló klaszterek csak az explorációs mintából hiányoztak, a replikációs minta 126 nominális klasztere között jelen voltak, amely főként a két minta közötti metodológiai különbségeknek tudható be.

Vizsgálatunk másik fő célja a multilókusz elemzések replikációs valószínűsége gyakorolt hatásának felmérése volt, amelyhez mind a nominális, mind a többszörös összehasonlításra korrigált replikációk arányát regisztráltuk. A p-értékek korrekciójától függetlenül, a multilókusz elemzések egyértelműen növelik a replikáció valószínűségét, méghozzá a definiált „marker” komplexitásának megfelelően. A replikáció aránya valószínűleg a genetikai heterogenitás csökkenése miatt növekszik, de nem vethetjük el

a lehetőségét, hogy e multilokusz markerek egyszerűen csak több gént annotálnak méretüknél fogva.

Vizsgálatunk nem mentes a korlátoktól sem. A SCHIZOBANK túlságosan kis elemszámú és alacsony lefedettségű mintája explorációs célokra sem ideális. A statisztikai erő növelése érdekében, az exploráció során szuggesztív p-értékkel ( $\alpha=0,1$ ) bíró asszociációkat is elfogadtunk, a fals pozitív asszociációk ellensúlyozására a második lépcsőben konzervatív p-érték korrekciót végeztünk. A közös variánsok gyarapítása érdekében a 1000 Genomes ph1v3 mintája szerint imputáltuk a GWAS mintát. Az imputáció az átfedést javította, azonban részben torzítja is az adatot: a hosszabb gének esetében több marker akad fenn a minőségkontroll szűrőn, ezért a rövidebb gének viszonylagos lefedettségét megnöveli. A rövidebb gének feldúsulása torzíthatja a kompetitív génszett teszteket, pl. Wilcoxon-teszt, felülreprezentációs teszt stb. E torzítást mérséklendő, a génszettek vizsgálatakor a kompetitív feldúsulási tesztek mellett régió-alapú, HYST génszett-asszociációs tesztet is végeztünk. Továbbá, az eset-trió minta elemzésekor nem tudtuk a replikációs minta kovariánsait használni, az FBAT nem kezel kovariánst, regressziós TDT-teszthez pedig túl sok a hiányzó genotípus. Az eltérő lefedettség, vagy a génszettek korrelációja, átfedése, torzíthatja a kompetitív feldúsulási tesztek eredményét, ennek csökkentését célozva elemeztünk a ReCiPa programmal függetlenné alakított kanonikus- és szuperútvonalakat, és pozicionális génszetteket.

Összefoglalva vizsgálatunk eredményeit, a biológiai és molekuláris funkciók, jellemzők alapján definiált, multilokusz elemzések a szkizofrénia etiológiájában szerepet játszó szinaptikus plaszticitás, neuron- és axon- és agyfejlődés, valamint immunológiai és transzkripció szabályozó annotációknak megfelelő gének, génszettek asszociációját, illetve feldúsulását eredményezték. Emellett a replikáció folyamatát – komplexitásukkal arányosan – egyértelműen elősegítették, még a vizsgált, egymástól jelentősen eltérő mintáink esetében is.



## 6. Következtetések

A genetikai asszociációs vizsgálatok általános problémája a valós és véletlen asszociációk megkülönböztetése, „arany standard” módszerének jelenleg is a független mintákon történő replikáció tekinthető. A pszichiátriai fenotípusok elemzésekor azonban a replikáció különösen problémás, poligénes jellegükből eredő nagyfokú genetikai heterogenitás miatt. A tézisekben ezért nemcsak a multilókusz asszociációs módszerekkel azonosított rizikófaktorokat, hanem a módszerekhez köthető replikációs valószínűséget is feltérképeztük, eltérő fenotípusú GWAS minták, illetve különböző vizsgálati szerkezet szerint genotípezált, de azonos fenotípusú minták asszociációit elemezve. A szkizofrénia etiológiájához illeszkedő eredményeink replikációja, illetve a szuicid magatartás vizsgálatában, a statisztikailag is szignifikáns replikációk és azonos irányú PRS predikciók hiánya, a multilókusz elemző módszerek differenciált hatékonyságát támasztja alá fenotípusos heterogenitás esetén. Egyúttal eredményeink rámutattak az öngyilkossági gondolatok és kísérlet fenotípusainak alapvető különbözőségére is, amely további vizsgálatok tekintetében megkérdőjelezi egyesített elemzésük érvényességét. Mindazonáltal, a nominálisan szignifikáns asszociációjú SNP-k nagyarányú annotációja az epigenetikai szabályozó lókuszokon, e markerek jelentős hányadában is valós asszociációkat valószínűsít. Tehát az inkonzisztens irányú PRS predikciók, és csak nominális p-értékek ellenére, feltételezhető a DICER1 gén és a miRNS rendszer szerepe is a neurotrofinok mellett, a szuicid magatartás kialakulásában. Mindez újfent megerősíti, hogy az asszociációs vizsgálati eredmények értelmezése a markerek genetikai, epigenetikai, expressziós, vagy egyéb funkcionális annotációinak integrációjára épüljön, a p-értékek mechanikus, matematikai rangsorolása helyett.

## 7. Összefoglalás

**Célkitűzés:** Első vizsgálatunkban a major depresszióhoz, illetve bipoláris zavarhoz társuló öngyilkossági viselkedés kandidáns gén alapú, multilokusz asszociációit elemeztük. Második vizsgálatunkban a szkizofrénia multilokusz asszociációit és replikációra gyakorolt hatásukat exploráltuk.

**Módszerek:** első vizsgálatunkban az NCBI GAIN bipoláris (N=999, szuicid n=358) és major depresszív (N=1753, szuicid n=245) mintáit elemeztük. A régió-alapú asszociációkat korrelációs (Génlista1), míg a PRS predikciókat korrelációs, prefrontális körökben kifejeződő miRNS (Génlista2) gének és célgénjeik (Génlista3) szerint elemeztük. Második, kétlépcsős vizsgálatunk a SCHIZOBANK projekt eset-trióin (n=16) explorációkor a szuggesztív multilokusz asszociációkat a GAIN szkizofrénia GWAS mintáin (N=5337) validáltuk.

**Eredmények:** statisztikailag szignifikáns gén-alapú asszociációt csak az NTRK2 gén mutatott, az MDD mintán, és csak a DICER1 p-értéke volt legalább nominális mindkét mintán. A Génlista2 PRS pontjai nominálisan szignifikáns, de negatív korrelációs együtthatóval jósolták a szuicid magatartást, legnagyobb varianciát a PRSice  $p < 0,03$  modellje fejezett ki. Második vizsgálatunkban a 14q31, 5q31 és Xq13 pozicionális génszettek, továbbá a splicing/alternatív splicing, idegfejlődés és embrionális fejlődés funkcionális annotációs klaszterek asszociációi maradtak szignifikánsnak a korrekció után is. A szignifikáns replikációk valószínűsége a teszt összetettségével nőtt.

**Következtetések:** első vizsgálatunk igazolta a BDNF-NTRK2-CREB útvonal, és a miRNS rendszer és végrehajtó funkciók lehetséges mediáló szerepét a szuicid ideációban, de a szuicid fenotípusok heterogenitását is. Második vizsgálatunk megerősítette az idegfejlődés, szinaptikus plaszticitás és immunrendszer érintettségét szkizofréniaiban, és igazolta a multilokusz elemzések előnyeit nagy genetikai heterogenitás esetén is.

## 8. Summary

**Objectives:** Our first study aimed to explore multilocus candidate-gene associations with suicidal behavior in major depressive and bipolar disorders. The goals of our second study were to explore the multilocus associations in schizophrenia and their replication.

**Methods:** Our first study analyzed samples of the GAIN bipolar disorder (N=999, suicidal n=358) and the GAIN major depressive disorder (N=1,753, suicidal n=245). Candidate genes were selected by literature search (Geneset1), gene expression of microRNA genes (Geneset2) and their targets (Geneset3). Gene-based associations were analyzed for Geneset1, while PRS score predictions of suicidal behavior were performed for all genesets. The second study had a two-staged design, associations with suggestive p-values at the discovery set of case trios (n=16) from the SCHIZOBANK study were tested on GAIN schizophrenia GWAS datasets (N=5,337) during the validation step.

**Results:** Only NTRK2 in the MDD sample associated statistically significantly with suicidal behavior, whereas only DICER1 had at least nominally significant association in both samples. Nominally significant, but negative predictions of suicidal behavior were found for Geneset2, with the best model by PRSice with  $p_t < 0.03$ . In the schizophrenia sample, genesets 14q31, 5q31 and Xq13 and functional annotation clusters of splicing, neurodevelopment and embryonic development had statistically significant p-values. Replication probability increased parallel with test complexity.

**Conclusions:** Our results supported the involvement of the BDNF-NTRK2-CREB pathway in suicidal ideation and suggested a potential mediating role of microRNA system and executive functions. Our second study corroborated the pathogenetic role of neurodevelopment, synaptic plasticity and immune system functions in schizophrenia and the utility of multilocus methods in cases of large genetic heterogeneity.

## 9. Irodalomjegyzék

1. Réthelyi J. (2015). A pszichiátriai betegségek genetikai háttere. In: Füredi J, Németh A, editors. A pszichiátria magyar kézikönyve. 5.bővített. Budapest: Medicina Könyvkiadó Zrt.; p. 42–64.
2. Collins FS, Green ED, Guttmacher AE, Guyer MS, US National Human Genome Research Institute. (2003). A vision for the future of genomics research. *Nature*. 422(6934):835–47.
3. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'ang L-Y, Huang W, Liu B, Shen Y, Tam PK-H, Tsui L-C, Waye MMY, Wong JT-F, Zeng C, Zhang Q, Chee MS, Galver LM, Kruglyak S, Murray SS, Oliphant AR, Montpetit A, Hudson TJ, Chagnon F, Ferretti V, Leboeuf M, Phillips MS, Verner A, Kwok P-Y, Duan S, Lind DL, Miller RD, Rice JP, Saccone NL, Taillon-Miller P, Xiao M, Nakamura Y, Sekine A, Sorimachi K, Tanaka T, Tanaka Y, Tsunoda T, Yoshino E, Bentley DR, Deloukas P, Hunt S, Powell D, Altshuler D, Gabriel SB, Zhang H, Zeng C, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Aniagwu T, Marshall PA, Matthew O, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Stein LD, Cunningham F, Kanani A, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Donnelly P, Marchini J, McVean GAT, Myers SR, Cardon LR, Abecasis GR, Morris A, Weir BS, Mullikin JC, Sherry ST, Feolo M, Altshuler D, Daly MJ, Schaffner SF, Qiu R, Kent A, Dunston GM, Kato K, Niikawa N, Knoppers BM, Foster MW, Clayton EW, Wang VO, Watkin J, Gibbs RA, Belmont JW, Sodergren E, Weinstock GM, Wilson RK, Fulton LL, Rogers J, Birren BW, Han H, Wang H, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Todani K, Fujita T, Tanaka S, Holden AL, Lai EH, Collins FS, Brooks LD, McEwen JE, Guyer MS, Jordan E, Peterson JL, Spiegel J, Sung LM, Zacharia LF, Kennedy K, Dunn MG, Seabrook R, Shillito M, Skene B, Stewart JG, Valle DL, Clayton EW, Jorde LB, Belmont JW, Chakravarti A, Cho MK, Duster T, Foster MW, Jasperse M, Knoppers BM, Kwok P-Y, Licinio J, Long JC, Marshall PA, Ossorio PN, Wang VO, Rotimi CN, Royal CDM, Spallone P, Terry SF, Lander ES, Lai EH, Nickerson DA, Abecasis GR, Altshuler D, Bentley DR, Boehnke M, Cardon LR, Daly MJ, Deloukas P, Douglas JA, Gabriel SB, Hudson RR, Hudson TJ,

- Kruglyak L, Kwok P-Y, Nakamura Y, Nussbaum RL, Royal CDM, Schaffner SF, Sherry ST, Stein LD, Tanaka T. (2003). The International HapMap Project. *Nature*. 426(6968):789–96.
4. The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*. 437(7063):1299–320.
  5. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MMY, Tsui SKW, Xue H, Wong JT-F, Galver LM, Fan J-B, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier J-F, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok P-Y, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui L-C, Mak W, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K,

- Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 449(7164):851–61.
6. GAIN Collaborative Research Group, Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Collaborative Association Study of Psoriasis, Ballinger D, Daly M, Donnelly P, Faraone S V, International Multi-Center ADHD Genetics Project, Frazer K, Gabriel S, Gejman P, Molecular Genetics of Schizophrenia Collaboration, Guttmacher A, Harris EL, Insel T, Kelsoe JR, Bipolar Genome Study, Lander E, McCowin N, Mailman MD, Nabel E, Ostell J, Pugh E, Sherry S, Sullivan PF, Major Depression Stage 1 Genomewide Association in Population-Based Samples Study, Thompson JF, Warram J, Genetics of Kidneys in Diabetes (GoKinD) Study, Wholley D, Milos PM, Collins FS. (2007). New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet*. 39(9):1045–51.
  7. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 39(10):1181–6.
  8. Tryka K a., Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, Feolo M. (2014). NCBI's database of genotypes and phenotypes: DbGaP. *Nucleic Acids Res*. 42(D1):1–5.
  9. Lappalainen I, Almeida-King J, Kumanduri V, Senf A, Spalding JD, Ur-Rehman S, Saunders G, Kandasamy J, Caccamo M, Leinonen R, Vaughan B, Laurent T,

- Rowland F, Marin-Garcia P, Barker J, Jokinen P, Torres AC, de Argila JR, Llobet OM, Medina I, Puy MS, Alberich M, de la Torre S, Navarro A, Paschall J, Flicek P. (2015). The European Genome-phenome Archive of human data consented for biomedical research. *Nat Genet.* 47(7):692–5.
10. McVean GA, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs RA, Dinh H, Kovar C, Lee S, Lewis L, Muzny D, Reid J, Wang M, Wang J, Fang X, Guo X, Jian M, Jiang H, Jin X, Li G, Li J, Li Y, Li Z, Liu X, Lu Y, Ma X, Su Z, Tai S, Tang M, Wang B, Wang G, Wu H, Wu R, Yin Y, Zhang W, Zhao J, Zhao M, Zheng X, Zhou Y, Lander ES, Altshuler DM, Gabriel SB, Gupta N, Flicek P, Clarke L, Leinonen R, Smith RE, Zheng-Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H, Sudbrak (Project Leader) R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Sherry ST, McVean GA, Mardis (Co-Principal Investigator) ER, Wilson (Co-Principal Investigator) RK, Fulton L, Fulton R, Weinstock GM, Durbin RM, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt JP, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Auton A, Gibbs RA, Yu (Project Leader) F, Bainbridge M, Challis D, Evani US, Lu J, Muzny D, Nagaswamy U, Reid J, Sabo A, Wang Y, Yu J, Wang J, Coin LJM, Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Z, Lin H, Liu B, Luo R, Qin N, Shao H, Wang B, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Marth GT, Garrison EP, Kural D, Lee W-P, Fung Leong W, Ward AN, Wu J, Zhang M, Lee C, Griffin L, Hsieh C-H, Mills RE, Shi X, von Grotthuss M, Zhang C, Daly MJ, DePristo (Project Leader) MA, Altshuler DM, Banks E, Bhatia G, Carneiro MO, del Angel G, Gabriel SB, Genovese G, Gupta N, Handsaker RE, Hartl C, Lander ES, McCarroll SA, Nemesh JC, Poplin RE, Schaffner SF, Shakir K, Yoon SC, Lihm J, Makarov V, Jin H, Kim W, Cheol Kim K, Korbel JO, Rausch T, Flicek P, Beal K, Clarke L, Cunningham F, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zheng-Bradley X, Clark AG,

Gottipati S, Keinan A, Rodriguez-Flores JL, Sabeti PC, Grossman SR, Tabrizi S, Tariyal R, Cooper DN, Ball E V., Stenson PD, Bentley DR, Barnes B, Bauer M, Keira Cheetham R, Cox T, Eberle M, Humphray S, Kahn S, Murray L, Peden J, Shaw R, Ye K, Batzer MA, Konkel MK, Walker JA, MacArthur DG, Lek M, Sudbrak (Project Leader), Amstislavskiy VS, Herwig R, Shriver MD, Bustamante CD, Byrnes JK, De La Vega FM, Gravel S, Kenny EE, Kidd JM, Lacroute P, Maples BK, Moreno-Estrada A, Zakharia F, Halperin E, Baran Y, Craig DW, Christoforides A, Homer N, Izatt T, Kurdoglu AA, Sinari SA, Squire K, Sherry ST, Xiao C, Sebat J, Bafna V, Ye K, Burchard EG, Hernandez RD, Gignoux CR, Haussler D, Katzman SJ, James Kent W, Howie B, Ruiz-Linares A, Dermitzakis ET, Lappalainen T, Devine SE, Liu X, Maroo A, Tallon LJ, Rosenfeld JA, Michelson LP, Abecasis GR, Min Kang (Project Leader) H, Anderson P, Angius A, Bigham A, Blackwell T, Busonero F, Cucca F, Fuchsberger C, Jones C, Jun G, Li Y, Lyons R, Maschio A, Porcu E, Reinier F, Sanna S, Schlessinger D, Sidore C, Tan A, Kate Trost M, Awadalla P, Hodgkinson A, Lunter G, McVean GA, Marchini JL, Myers S, Churchhouse C, Delaneau O, Gupta-Hinch A, Iqbal Z, Mathieson I, Rimmer A, Xifara DK, Oleksyk TK, Fu Y, Liu X, Xiong M, Jorde L, Witherspoon D, Xing J, Eichler EE, Browning BL, Alkan C, Hajirasouliha I, Hormozdiari F, Ko A, Sudmant PH, Mardis (Co-Principal Investigator) ER, Chen K, Chinwalla A, Ding L, Dooling D, Koboldt DC, McLellan MD, Wallis JW, Wendl MC, Zhang Q, Durbin RM, Hurles ME, Albers CA, Ayub Q, Balasubramaniam S, Chen Y, Coffey AJ, Colonna V, Danecek P, Huang N, Jostins L, Keane TM, Li H, McCarthy S, Scally A, Stalker J, Walter K, Xue Y, Zhang Y, Gerstein MB, Abyzov A, Balasubramaniam S, Chen J, Clarke D, Fu Y, Habegger L, Harmanci AO, Jin M, Khurana E, Jasmine Mu X, Sisu C, Li Y, Luo R, Zhu H, Lee C, Griffin L, Hsieh C-H, Mills RE, Shi X, von Grotthuss M, Zhang C, Marth GT, Garrison EP, Kural D, Lee W-P, Ward AN, Wu J, Zhang M, McCarroll (Project Leader) SA, Altshuler DM, Banks E, del Angel G, Genovese G, Handsaker RE, Hartl C, Nemesh JC, Shakir K, Yoon SC, Lihm J, Makarov V, Degenhardt J, Flicek P, Clarke L, Smith RE, Zheng-Bradley X, Korbel JO, Rausch T, Stütz AM, Bentley DR, Barnes B, Keira Cheetham R, Eberle M, Humphray S, Kahn S, Murray L,



Shaw R, Ye K, Batzer MA, Konkel MK, Walker JA, Lacroute P, Craig DW, Homer N, Church D, Xiao C, Sebat J, Bafna V, Michaelson JJ, Ye K, Devine SE, Liu X, Maroo A, Tallon LJ, Lunter G, McVean GA, Iqbal Z, Witherspoon D, Xing J, Eichler EE, Alkan C, Hajirasouliha I, Hormozdiari F, Ko A, Sudmant PH, Chen K, Chinwalla A, Ding L, McLellan MD, Wallis JW, Hurles ME, Blackburne B, Li H, Lindsay SJ, Ning Z, Scally A, Walter K, Zhang Y, Gerstein MB, Abyzov A, Chen J, Clarke D, Khurana E, Jasmine Mu X, Sisu C, Gibbs RA, Yu (Project Leader) F, Bainbridge M, Challis D, Evani US, Kovar C, Lewis L, Lu J, Muzny D, Nagaswamy U, Reid J, Sabo A, Yu J, Guo X, Li Y, Wu R, Marth GT, Garrison EP, Fung Leong W, Ward AN, del Angel G, DePristo MA, Gabriel SB, Gupta N, Hartl C, Poplin RE, Clark AG, Rodriguez-Flores JL, Flicek P, Clarke L, Smith RE, Zheng-Bradley X, MacArthur DG, Bustamante CD, Gravel S, Craig DW, Christoforides A, Homer N, Izatt T, Sherry ST, Xiao C, Dermitzakis ET, Abecasis GR, Min Kang H, McVean GA, Mardis ER, Dooling D, Fulton L, Fulton R, Koboldt DC, Durbin RM, Balasubramanian S, Keane TM, McCarthy S, Stalker J, Gerstein MB, Balasubramanian S, Habegger L, Garrison EP, Gibbs RA, Bainbridge M, Muzny D, Yu F, Yu J, del Angel G, Handsaker RE, Makarov V, Rodriguez-Flores JL, Jin H, Kim W, Cheol Kim K, Flicek P, Beal K, Clarke L, Cunningham F, Herrero J, McLaren WM, Ritchie GRS, Zheng-Bradley X, Tabrizi S, MacArthur DG, Lek M, Bustamante CD, De La Vega FM, Craig DW, Kurdoglu AA, Lappalainen T, Rosenfeld JA, Michelson LP, Awadalla P, Hodgkinson A, McVean GA, Chen K, Chen Y, Colonna V, Frankish A, Harrow J, Xue Y, Gerstein MB, Abyzov A, Balasubramanian S, Chen J, Clarke D, Fu Y, Harmanci AO, Jin M, Khurana E, Jasmine Mu X, Sisu C, Gibbs RA, Fowler G, Hale W, Kalra D, Kovar C, Muzny D, Reid J, Wang J, Guo X, Li G, Li Y, Zheng X, Altshuler DM, Flicek P, Clarke (Project Leader) L, Barker J, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Toneva I, Vaughan B, Zheng-Bradley X, Bentley DR, Cox T, Humphray S, Kahn S, Sudbrak (Project Leader) R, Albrecht MW, Lienhard M, Craig DW, Izatt T, Kurdoglu AA, Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P,

- O’Sullivan C, Ostapchuk Y, Phan L, Ponomarov S, Schneider V, Shekhtman E, Sirotkin K, Slotta D, Xiao C, Zhang H, Haussler D, Abecasis GR, McVean GA, Alkan C, Ko A, Dooling D, Durbin RM, Balasubramaniam S, Keane TM, McCarthy S, Stalker J, Chakravarti A, Knoppers BM, Abecasis GR, Barnes KC, Beiswanger C, Burchard EG, Bustamante CD, Cai H, Cao H, Durbin RM, Gharani N, Gibbs RA, Gignoux CR, Gravel S, Henn B, Jones D, Jorde L, Kaye JS, Keinan A, Kent A, Kerasidou A, Li Y, Mathias R, McVean GA, Moreno-Estrada A, Ossorio PN, Parker M, Reich D, Rotimi CN, Royal CD, Sandoval K, Su Y, Sudbrak R, Tian Z, Timmermann B, Tishkoff S, Toji LH, Tyler Smith C, Via M, Wang Y, Yang H, Yang L, Zhu J, Bodmer W, Bedoya G, Ruiz-Linares A, Zhi Ming C, Yang G, Jia You C, Peltonen L, Garcia-Montero A, Orfao A, Dutil J, Martinez-Cruzado JC, Oleksyk TK, Brooks LD, Felsenfeld AL, McEwen JE, Clemm NC, Duncanson A, Dunn M, Green ED, Guyer MS, Peterson JL, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Min Kang H, Marth GT, McVean GA. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491(7422):56–65.
11. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe’er I, Dudbridge F, Holmans PA, Whitemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman P V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 460(7256):753–7.
12. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OPH, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Suvisaari J, Lonnqvist J, Paunio T, Børglum AD, Hartmann A, Fink-Jensen A, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Böttcher Y, Olesen J, Breuer R, Möller H-J, Giegling I, Rasmussen HB, Timm S, Mattheisen M, Bitter I, Réthelyi JM, Magnusdottir BB, Sigmundsson T, Olason P, Masson G, Gulcher JR, Haraldsson M, Fossdal R, Thorgeirsson TE, Thorsteinsdottir U, Ruggeri M, Tosato S, Franke B, Strengman E, Kiemenev LA, Genetic Risk and Outcome in Psychosis (GROUP), Melle I, Djurovic S, Abramova L, Kaleda V, Sanjuan J, de Frutos R, Bramon E, Vassos E, Fraser G, Ettinger U, Picchioni M, Walker N, Touloupoulou T, Need AC, Ge

- D, Yoon JL, Shianna K V, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Carracedo A, Arango C, Costas J, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Matthews PM, Muglia P, Peltonen L, St Clair D, Goldstein DB, Stefansson K, Collier DA. (2009). Common variants conferring risk of schizophrenia. *Nature*. 460(7256):744–7.
13. Sullivan PF, de Geus EJC, Willemsen G, James MR, Smit JH, Zandbelt T, Arolt V, Baune BT, Blackwood D, Cichon S, Coventry WL, Domschke K, Farmer A, Fava M, Gordon SD, He Q, Heath AC, Heutink P, Holsboer F, Hoogendijk WJ, Hottenga JJ, Hu Y, Kohli M, Lin D, Lucae S, Macintyre DJ, Maier W, McGhee KA, McGuffin P, Montgomery GW, Muir WJ, Nolen WA, Nöthen MM, Perlis RH, Pirlo K, Posthuma D, Rietschel M, Rizzu P, Schosser A, Smit AB, Smoller JW, Tzeng J-Y, van Dyck R, Verhage M, Zitman FG, Martin NG, Wray NR, Boomsma DI, Penninx BWJH. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry*. 14(4):359–75.
  14. Sklar P, Smoller JW, Fan J, Ferreira MAR, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone S V, Kirby A, de Bakker PIW, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. (2008). Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 13(6):558–69.
  15. Psychiatric GWAS Consortium Coordinating Committee, Cichon S, Craddock N, Daly M, Faraone S V, Gejman P V, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF. (2009). Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 166(5):540–56.
  16. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. (2009). Finding the missing heritability of complex

- diseases. *Nature*. 461(7265):747–53.
17. Dudbridge F. (2016). Polygenic Epidemiology. *Genet Epidemiol*. 40(4):268–72.
  18. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, Purcell (Leader) SM, Stone JL, Sullivan PF, Ruderfer DM, McQuillin A, Morris DW, O'Dushlaine CT, Corvin A, Holmans PA, O'Donovan MC, Sklar P, Wray NR, Macgregor S, Sklar P, Sullivan PF, O'Donovan MC, Visscher PM, Gurling H, Blackwood DHR, Corvin A, Craddock NJ, Gill M, Hultman CM, Kirov GK, Lichtenstein P, McQuillin A, Muir WJ, O'Donovan MC, Owen MJ, Pato CN, Purcell SM, Scolnick EM, St Clair D, Stone JL, Sullivan PF, Sklar (Leader) P, O'Donovan MC, Kirov GK, Craddock NJ, Holmans PA, Williams NM, Georgieva L, Nikolov I, Norton N, Williams H, Toncheva D, Milanova V, Owen MJ, Hultman CM, Lichtenstein P, Thelander EF, Sullivan P, Morris DW, O'Dushlaine CT, Kenny E, Quinn EM, Gill M, Corvin A, McQuillin A, Choudhury K, Datta S, Pimm J, Thirumalai S, Puri V, Krasucki R, Lawrence J, Quested D, Bass N, Gurling H, Crombie C, Fraser G, Leh Kuan S, Walker N, St Clair D, Blackwood DHR, Muir WJ, McGhee KA, Pickard B, Malloy P, Maclean AW, Van Beck M, Wray NR, Macgregor S, Visscher PM, Pato MT, Medeiros H, Middleton F, Carvalho C, Morley C, Fanous A, Conti D, Knowles JA, Paz Ferreira C, Macedo A, Helena Azevedo M, Pato CN, Stone JL, Ruderfer DM, Kirby AN, Ferreira MAR, Daly MJ, Purcell SM, Sklar P, Purcell SM, Stone JL, Chambert K, Ruderfer DM, Kuruvilla F, Gabriel SB, Ardlie K, Moran JL, Daly MJ, Scolnick EM, Sklar P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 460(7256):748–52.
  19. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. (2014). Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*. 55(10):1068–87.
  20. Yang J, Lee SH, Goddard ME, Visscher PM. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 88(1):76–82.
  21. Lee SH, Wray NR, Goddard ME, Visscher PM. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*. 88(3):294–305.
  22. Goldman D, Domschke K, Abecasis G, Auton A, Brooks L, DePristo M, Durbin

- R, Handsaker R, Kang H, Marth G, McVean G, Aberg K, McClay J, Nerella S, Clark S, Kumar G, Chen W, Khachane A, Xie L, Hudson A, Gao G, Harada A, Hultman C, Sullivan P, Magnusson P, Oord E van den, Bevilacqua L, Doly S, Kaprio J, Yuan Q, Tikkanen R, Paunio T, Zhou Z, Wedenoja J, Maroteaux L, Diaz S, Belmer A, Hodgkinson C, Dell'osso L, Suvisaari J, Cocco E, Rose R, Peltonen L, Virkkunen M, Goldman D, Biesecker L, Biesecker B, Peay H, Domschke K, Tidow N, Kuithan H, Schwarte K, Klauke B, Ambrée O, Reif A, Schmidt H, Arolt V, Kersting A, Zwanzger P, Deckert J, Girard S, Hu X, Zhang B, Liu W, Paciga S, He W, Lanz T, Kleiman R, Dougherty B, Hall S, McIntosh A, Lawrie S, Power A, John S, Blackwood D, Clair DS, Brandon N, Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner J, Pariante C, Pace T, Mercer K, Mayberg H, Bradley B, Nemeroff C, Holsboer F, Heim C, Ressler K, Rein T, Binder E, Ley T, McPherson J, Michaelson J, Neale B, Ng S, Buckingham K, Lee C, Bigham A, Tabor H, Dent K, Huff C, Shannon P, Jabs E, Nickerson D, Shendure J, Bamshad M, O'Roak B, Deriziotis P, Lee C, Vives L, Schwartz J, Girirajan S, Karakoc E, Mackenzie A, Ng S, Baker C, Rieder M, Nickerson D, Bernier R, Fisher S, Shendure J, Eichler E, Sanger F, Nicklen S, Coulson A, Voineagu I, Wang X, Johnston P, Lowe J, Tian Y, Horvath S, Mill J, Cantor R, Blencowe B, Geschwind D, Xu B, Roos J, Dexheimer P, Boone B, Plummer B, Levy S, Gogos J, Karayiorgou M, Zhou Z, Zhou Z, Karlsson C, Liang T, Xiong W, Kimura M, Tapocik J, Yuan Q, Barbier E, Feng A, Flanigan M, Augier E, Enoch M, Hodgkinson C, Shen P, Lovinger D, Edenberg H, Heilig M, Goldman D. (2014). Making sense of deep sequencing. *Int J Neuropsychopharmacol*. 17(10):1717–25.
23. Harrow J, Denoeud F, Frankish A, Reymond A, Chen C-K, Chrast J, Lagarde J, Gilbert JGR, Storey R, Swarbreck D, Rossier C, Ucla C, Hubbard T, Antonarakis SE, Guigo R. (2006). GENCODE: producing a reference annotation for ENCODE. *Genome Biol*. 7 Suppl 1(Suppl 1):S4.1-9.
24. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N,

- Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22(9):1760–74.
25. ENCODE Project Consortium. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 306(5696):636–40.
  26. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segrè A V, Djebali S, Niarchou A, GTEx Consortium, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigó R. (2015). Human genomics. The human transcriptome across tissues and individuals. *Science.* 348(6235):660–5.
  27. Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, Malone J, Rustici G, Williams E, Parkinson H, Brazma A. (2009). Gene expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* 38(SUPPL.1):690–8.
  28. Yang T-P, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics.* 26(19):2474–6.
  29. Ardlie KG, Deluca DS, Segre A V., Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, Ward LD, Kheradpour P, Iriarte B, Meng Y, Palmer CD, Esko T, Winckler W, Hirschhorn JN, Kellis M, MacArthur DG, Getz G, Shabalin AA, Li G, Zhou Y-H, Nobel AB, Rusyn I, Wright FA, Lappalainen T, Ferreira PG, Ongen H, Rivas MA, Battle A, Mostafavi S, Monlong J, Sammeth M, Mele M, Reverter F, Goldmann JM, Koller D, Guigo R, McCarthy MI, Dermitzakis ET, Gamazon ER, Im HK, Konkashbaev A, Nicolae DL, Cox NJ, Flutre T, Wen X, Stephens M, Pritchard JK, Tu Z, Zhang B, Huang T, Long Q, Lin L, Yang J, Zhu J, Liu J, Brown A, Mestichelli B, Tidwell D, Lo E, Salvatore M, Shad S, Thomas JA, Lonsdale JT, Moser MT, Gillard BM, Karasik E, Ramsey K, Choi C, Foster BA, Syron J, Fleming J, Magazine H, Hasz R, Walters GD, Bridge JP, Miklos M, Sullivan S, Barker LK, Traino HM, Mosavel M, Siminoff LA, Valley DR, Rohrer DC, Jewell SD, Branton PA, Sobin LH, Barcus M, Qi L, McLean J, Hariharan P, Um

- KS, Wu S, Tabor D, Shive C, Smith AM, Buia SA, Undale AH, Robinson KL, Roche N, Valentino KM, Britton A, Burges R, Bradbury D, Hambright KW, Seleski J, Korzeniewski GE, Erickson K, Marcus Y, Tejada J, Taherian M, Lu C, Basile M, Mash DC, Volpi S, Struewing JP, Temple GF, Boyer J, Colantuoni D, Little R, Koester S, Carithers LJ, Moore HM, Guan P, Compton C, Sawyer SJ, Demchok JP, Vaught JB, Rabiner CA, Lockhart NC, Ardlie KG, Getz G, Wright FA, Kellis M, Volpi S, Dermitzakis ET. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* (80- ). 348(6235):648–60.
30. Psychiatric GWAS Consortium Steering Committee. (2009). A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry*. 14(1):10–7.
  31. Sullivan P. (2012). Don't give up on GWAS. *Mol Psychiatry*. 17(1):2–3.
  32. Zeng P, Zhao Y, Qian C, Zhang L, Zhang R, Gou J, Liu J, Liu L, Chen F. (2015). Statistical analysis for genome-wide association study. *J Biomed Res*. 29(4):285–97.
  33. Neale BM, Sham PC. (2004). PERSPECTIVE The Future of Association Studies: Gene-Based Analysis and Replication. *Am J Hum Genet*. 75:353–62.
  34. Montana G. (2006). Statistical methods in genetics. *Brief Bioinform*. 7(3):297–308.
  35. Beyene J, Tritchler D, Asimit JL, Hamid JS. (2009). Gene- or region-based analysis of genome-wide association studies. *Genet Epidemiol*. 33 Suppl 1:S105–10.
  36. Hauser E, Cremer N, Hein R, Deshmukh H. (2009). Haplotype-based analysis: a summary of GAW16 Group 4 analysis. *Genet Epidemiol*. 33 Suppl 1(Suppl 1):S24–8.
  37. Moore JH, Asselbergs FW, Williams SM. (2010). Bioinformatics challenges for genome-wide association studies. *Bioinformatics*. 26(4):445–55.
  38. Ayalew M, Le-Niculescu H, Levey DF, Jain N, Changala B, Patel SD, Winiger E, Breier A, Shekhar A, Amdur R, Koller D, Nurnberger JI, Corvin A, Geyer M, Tsuang MT, Salomon D, Schork NJ, Fanous a. H, O'Donovan MC, Niculescu a. B. (2012). Convergent functional genomics of schizophrenia: from

- comprehensive understanding to genetic risk prediction. *Mol Psychiatry*. 17(9):887–905.
39. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V, Aquino-Michaels K, Carroll RJ, Eyler AE, Denny JC, GTEx Consortium D, Nicolae DL, Cox NJ, Im HK. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 47(9):1091–8.
  40. Chung D, Yang C, Li C, Gelernter J, Zhao H. (2014). GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. Tang H, editor. *PLoS Genet*. 10(11):e1004787.
  41. Burgtorf C, Kepper P, Hoehe M, Schmitt C, Reinhardt R, Lehrach H, Sauer S. (2003). Clone-based systematic haplotyping (CSH): a procedure for physical haplotyping of whole genomes. *Genome Res*. 13(12):2717–24.
  42. Yang H, Chen X, Wong WH. (2011). Completely phased genome sequencing through chromosome sorting. *Proc Natl Acad Sci U S A*. 108(1):12–7.
  43. Snyder MW, Adey A, Kitzman JO, Shendure J. (2015). Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet*. 16(6):344–58.
  44. Browning SR, Browning BL. (2011). Haplotype phasing: existing methods and new developments. *Nat Rev Genet*. 12(10):703–14.
  45. Glusman G, Cox HC, Roach JC. (2014). Whole-genome haplotyping approaches and genomic medicine. *Genome Med*. 6(9):73.
  46. Delaneau O, Howie B, Cox AJ, Zagury J-FF, Marchini J. (2013). Haplotype estimation using sequencing reads. *Am J Hum Genet*. 93(4):687–96.
  47. Clark AG. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol*. 7(2):111–22.
  48. Excoffier L, Slatkin M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*. 12(5):921–7.
  49. Stephens M, Scheet P. (2005). Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. *Am J Hum Genet*. 76(3):449–62.
  50. Scheet P, Stephens M. (2006). A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes



- and Haplotypic Phase. *Am J Hum Genet.* 78(4):629–44.
51. Li N, Stephens M. (2003). Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. *Genetics.* 165(4):2213–33.
  52. Delaneau O, Coulonges C, Zagury J-F. (2008). Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics.* 9(1):540.
  53. Delaneau O, Zagury J-F, Marchini J. (2012). Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 10(1):5–6.
  54. Browning SR, Browning BL. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 81(5):1084–97.
  55. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002). The structure of haplotype blocks in the human genome. *Science.* 296(5576):2225–9.
  56. Pattaro C, Ruczinski I, Fallin DM, Parmigiani G. (2008). Haplotype block partitioning as a tool for dimensionality reduction in SNP association studies. *BMC Genomics.* 9(1):405.
  57. Taliun D, Gamper J, Pattaro C. (2014). Efficient haplotype block recognition of very long and dense genetic sequences. *BMC Bioinformatics.* 15(1):10.
  58. Barnes MR. (2006). Navigating the HapMap. *Brief Bioinform.* 7(3):211–24.
  59. Chen W-P, Hung C-L, Lin Y-L, Chen W-P, Hung C-L, Lin Y-L. (2013). Efficient Haplotype Block Partitioning and Tag SNP Selection Algorithms under Various Constraints. *Biomed Res Int.* 2013:1–13.
  60. Liu N, Zhang K, Zhao H. (2008). Haplotype-association analysis. *Adv Genet.* 60:335–405.
  61. Salem RM, Wessel J, Schork NJ. (2005). A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. *Hum Genomics.* 2(1):39–66.
  62. Zhang L, Liu J, Deng H-W. (2009). A multilocus linkage disequilibrium measure based on mutual information theory and its applications. *Genetica.* 137(3):355–64.

63. Indap AR, Marth GT, Struble CA, Tonellato P, Olivier M. (2005). Analysis of concordance of different haplotype block partitioning algorithms. *BMC Bioinformatics*. 6(1):303.
64. Bagos PG. (2011). Meta-analysis of haplotype-association studies: comparison of methods and empirical evaluation of the literature. *BMC Genet*. 12:8.
65. Lambert J-C, Grenier-Boley B, Harold D, Zelenika D, Chouraki V, Kamatani Y, Sleegers K, Ikram MA, Hiltunen M, Reitz C, Mateo I, Feulner T, Bullido M, Galimberti D, Concari L, Alvarez V, Sims R, Gerrish A, Chapman J, Deniz-Naranjo C, Solfrizzi V, Sorbi S, Arosio B, Spalletta G, Siciliano G, Epelbaum J, Hannequin D, Dartigues J-F, Tzourio C, Berr C, Schrijvers EMC, Rogers R, Tosto G, Pasquier F, Bettens K, Van Cauwenberghe C, Fratiglioni L, Graff C, Delepine M, Ferri R, Reynolds CA, Lannfelt L, Ingelsson M, Prince JA, Chillotti C, Pilotto A, Seripa D, Boland A, Mancuso M, Bossù P, Annoni G, Nacmias B, Bosco P, Panza F, Sanchez-Garcia F, Del Zompo M, Coto E, Owen M, O'Donovan M, Valdivieso F, Caffarra P, Caffara P, Scarpini E, Combarros O, Buée L, Champion D, Soininen H, Breteler M, Riemenschneider M, Van Broeckhoven C, Alperovitch A, Lathrop M, Trégouët D-A, Williams J, Amouyel P. (2013). Genome-wide haplotype association study identifies the *FRMD4A* gene as a risk locus for Alzheimer's disease. *Mol Psychiatry*. 18(4):461–70.
66. Martiskainen H, Viswanathan J, Nykänen N-P, Kurki M, Helisalmi S, Natunen T, Sarajärvi T, Kurkinen KMA, Pursiheimo J-P, Rauramaa T, Alafuzoff I, Jääskeläinen JE, Leinonen V, Soininen H, Haapasalo A, Huttunen HJ, Hiltunen M. (2015). Transcriptomics and mechanistic elucidation of Alzheimer's disease risk genes in the brain and in vitro models. *Neurobiol Aging*. 36(2):1221.e15-28.
67. Howie BN, Donnelly P, Marchini J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. Schork NJ, editor. *PLoS Genet*. 5(6):e1000529.
68. Li M-X, Gui H-S, Kwan JSH, Sham PC. (2011). GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure. *Am J Hum Genet*. 88(3):283–93.
69. Li MX, Kwan JSH, Sham PC. (2012). HYST: A hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-

- based association analysis. *Am J Hum Genet.* 91(3):478–88.
70. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, Hayeck T, Won H-H, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology and R of IV in BC (DRIVE) study, Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet.* 97(4):576–92.
  71. Mishra A, Macgregor S. (2014). VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet.* 18(1).
  72. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 44(8):955–9.
  73. Ballard DH, Cho J, Zhao H. (2010). Comparisons of multi-marker association methods to detect association between a candidate region and disease. *Genet Epidemiol.* 34(3):201–12.
  74. Wojcik GL, Kao WHL, Duggal P. (2015). Relative performance of gene- and pathway-level methods as secondary analyses for genome-wide association studies. *BMC Genet.* 16(1):34.
  75. Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, Zhao J, Zhou X, Reveille JD, Jin L, Amos CI, Xiong M. (2010). Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet.* 18(1):111–7.
  76. Edwards AC, Bigdeli TB, Docherty AR, Bacanu S, Lee D, de Candia TR, Moscati A, Thiselton DL, Maher BS, Wormley BK, Molecular Genetics of Schizophrenia Collaboration (MGS), Walsh D, O’Neill FA, Kendler KS, Riley BP, Fanous AH. (2016). Meta-analysis of Positive and Negative Symptoms Reveals Schizophrenia Modifier Genes. *Schizophr Bull.* 42(2):279–87.
  77. Frankish A, Uszczyńska B, Ritchie GRS, Gonzalez JM, Pervouchine D, Petryszak R, Mudge JM, Fonseca N, Brazma A, Guigo R, Harrow J. (2015). Comparison of GENCODE and RefSeq gene annotation and the impact of

- reference geneset on variant effect prediction. *BMC Genomics*. 16 Suppl 8(Suppl 8):S2.
78. Gauderman WJ, Murcray C, Gilliland F, Conti D V. (2007). Testing association between disease and multiple SNPs in a candidate gene. *Genet Epidemiol*. 31(5):383–95.
  79. Black MH, Watanabe RM. (2011). A principal components-based clustering method to identify variants associated with complex traits. *Hum Hered*. 71(1):50–8.
  80. Black MH, Watanabe RM. (2009). A principal-components-based clustering method to identify multiple variants associated with rheumatoid arthritis and arthritis-related autoantibodies. *BMC Proc*. 3 Suppl 7(Suppl 7):S129.
  81. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X. (2010). Powerful SNP-Set Analysis for Case-Control Genome-wide Association Studies. *Am J Hum Genet*. 86:929–42.
  82. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *Am J Hum Genet*. 89(1):82–93.
  83. Basu S, Zhang Y, Ray D, Miller MB, Iacono WG, McGue M. (2013). A rapid gene-based genome-wide association test with multivariate traits. *Hum Hered*. 76(2):53–63.
  84. Wang Z, Wang X, Sha Q, Zhang S. (2016). Joint Analysis of Multiple Traits in Rare Variant Association Studies. *Ann Hum Genet*. 80(3):162–71.
  85. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, González-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlöf J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HPJ, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O, Pirinen M, Montgomery SB, Donnelly P, McCarthy MI, Flicek P, Strom TM, Geuvaris Consortium, Lehrach H, Schreiber S, Sudbrak R, Carracedo A, Antonarakis SE, Häsler R, Syvänen A-C, van Ommen G-J, Brazma A, Meitinger T, Rosenstiel P, Guigó R, Gut IG, Estivill X, Dermitzakis ET. (2013). Transcriptome and genome

- sequencing uncovers functional variation in humans. *Nature*. 501(7468):506–11.
86. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, Urban AE, Montgomery SB, Levinson DF, Koller D. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res*. 24(1):14–24.
  87. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, Hardy J, Ryten M, Trabzuni D, Guelfi S, Weale ME, Ramasamy A, Forabosco P, Smith C, Walker R, Arepalli S, Cookson MR, Dillman A, Gibbs JR, Hernandez DG, Nalls MA, Singleton AB, Traynor B, van der Brug M, Ferrucci L, Johnson R, Zielke R, Longo DL, Troncoso J, Zonderman A, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci*. 17(10):1418–28.
  88. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 81(3):559–75.
  89. Simes RJ. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 73(3):751–4.
  90. Curtis D, Vine AE, Knight J. (2008). A simple method for assessing the strength of evidence for association at the level of the whole gene. *Adv Appl Bioinform Chem*. 1:115–20.
  91. Dudbridge F, Koeleman BPC. (2003). Rank truncated product of P-values, with application to genomewide association scans. *Genet Epidemiol*. 25(4):360–6.
  92. Brown MB. (1975). A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics*. 31(4):987.
  93. Moskvina V, O’Dushlaine C, Purcell S, Craddock N, Holmans P, O’Donovan MC. (2011). Evaluation of an approximation method for assessment of overall significance of multiple-dependent tests in a genomewide association study. *Genet Epidemiol*. 35(8):861–6.
  94. Mooney MA, Nigg JT, McWeeney SK, Wilmot B. (2014). Functional and

- genomic context in pathway analysis of GWAS data. *Trends Genet.* 30(9):390–400.
95. Mooney MA, Wilmot B. (2015). Gene set analysis: A step-by-step guide. *Am J Med Genet B Neuropsychiatr Genet.* 168(7):517–27.
  96. Jin L, Zuo X-Y, Su W-Y, Zhao X-L, Yuan M-Q, Han L-Z, Zhao X, Chen Y-D, Rao S-Q. (2014). Pathway-based Analysis Tools for Complex Diseases: A Review. *Genomics Proteomics Bioinformatics.* 12(5):210–20.
  97. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. (2011). Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39(Database):D685–90.
  98. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43(Database issue):D447-52.
  99. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25(1):25–9.
  100. Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43(D1):D1049–56.
  101. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur Ö, Blinov M, Brauner E, Corwin D, Donaldson S, Gibbons F, Goldberg R, Hornbeck P, Luna A, Murray-Rust P, Neumann E, Reubenacker O, Samwald M, van Iersel M, Wimalaratne S, Allen K, Braun B, Whirl-Carrillo M, Cheung K-H, Dahlquist K, Finney A, Gillespie M, Glass E, Gong L, Haw R, Honig M, Hubaut O, Kane D, Krupa S, Kutmon M, Leonard J, Marks D, Merberg D, Petri V, Pico A, Ravenscroft D, Ren L, Shah N, Sunshine M, Tang R, Whaley R, Letovksy S, Buetow KH,

- Rzhetsky A, Schachter V, Sobral BS, Dogrusoz U, McWeeney S, Aladjem M, Birney E, Collado-Vides J, Goto S, Hucka M, Novère N Le, Maltsev N, Pandey A, Thomas P, Wingender E, Karp PD, Sander C, Bader GD. (2010). The BioPAX community standard for pathway data sharing. *Nat Biotechnol.* 28(9):935–42.
102. Khatri P, Sirota M, Butte AJ. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. Ouzounis CA, editor. *PLoS Comput Biol.* 8(2):e1002375.
103. Donato M, Xu Z, Tomoiaga A, Granneman JG, Mackenzie RG, Bao R, Than NG, Westfall PH, Romero R, Draghici S. (2013). Analysis and correction of crosstalk effects in pathway analysis. *Genome Res.* 23(11):1885–93.
104. Kanehisa M, Goto S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1):27–30.
105. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40(Database issue):D109-14.
106. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, D'Eustachio P. (2016). The Reactome pathway Knowledgebase. *Nucleic Acids Res.* 44(D1):D481-7.
107. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 102(43):15545–50.
108. Huang DW, Sherman BT, Lempicki RA. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44–57.
109. Huang DW, Sherman BT, Lempicki RA. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37(1):1–13.
110. Mitrea C, Taghavi Z, Bokanizad B, Hanoudi S, Tagett R, Donato M, Voichița C,

- Drăghici S. (2013). Methods and approaches in the topology-based analysis of biological pathways. *Front Physiol.* 4:278.
111. Jia P, Zheng S, Long J, Zheng W, Zhao Z. (2011). dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics.* 27(1):95–102.
  112. Leiserson MDM, Eldridge J V, Ramachandran S, Raphael BJ. (2013). Network analysis of GWAS data. *Curr Opin Genet Dev.* 23(6):602–10.
  113. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G. (2007). MINT: the Molecular INTERaction database. *Nucleic Acids Res.* 35(Database):D572–4.
  114. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34(Database issue):D535-9.
  115. Medina I, Montaner D, Bonifaci N, Pujana MA, Carbonell J, Tarraga J, Al-Shahrour F, Dopazo J. (2009). Gene set-based analysis of polymorphisms: finding pathways or biological processes associated to traits in genome-wide association studies. *Nucleic Acids Res.* 37(Web Server):W340–4.
  116. Wu G, Zhi D. (2013). Pathway-based approaches for sequencing-based genome-wide association studies. *Genet Epidemiol.* 37(5):478–94.
  117. Wang X, Liu X, Sim X, Xu H, Khor C-C, Ong RT-H, Tay W-T, Suo C, Poh W-T, Ng DP-K, Liu J, Aung T, Chia K-S, Wong T-Y, Tai E-S, Teo Y-Y. (2012). A statistical method for region-based meta-analysis of genome-wide association studies in genetically diverse populations. *Eur J Hum Genet.* 20(4):469–75.
  118. Gottesman II, Shields J. (1967). A POLYGENIC THEORY OF SCHIZOPHRENIA. *Proc Natl Acad Sci.* 58:199–205.
  119. Khoury MJ, Yang Q, Gwinn M, Little J, Flanders WD. (2004). An epidemiologic assessment of genomic profiling for measuring susceptibility to common diseases and targeting interventions. *Genet Med.* 6(1):38–47.
  120. Dudbridge F. (2013). Power and predictive accuracy of polygenic risk scores. Wray NR, editor. *PLoS Genet.* 9(3):e1003348.
  121. Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature.*



- 511(7510):421–7.
122. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O’Dushlaine C, Chambert K, Bergen SE, Kähler A, Duncan L, Stahl E, Genovese G, Fernández E, Collins MO, Komiyama NH, Choudhary JS, Magnusson PKE, Banks E, Shakir K, Garimella K, Fennell T, DePristo M, Grant SGN, Haggarty SJ, Gabriel S, Scolnick EM, Lander ES, Hultman CM, Sullivan PF, McCarroll SA, Sklar P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 506(7487):185–90.
  123. Rees E, O’Donovan MC, Owen MJ. (2015). Genetics of schizophrenia. *Curr Opin Behav Sci*. 2:8–14.
  124. Sullivan PF, Posthuma D. (2015). Biological pathways and networks implicated in psychiatric disorders. *Curr Opin Behav Sci*. 2:58–68.
  125. Hall J, Trent S, Thomas KL, O’Donovan MC, Owen MJ. (2015). Genetic risk for schizophrenia: Convergence on synaptic pathways involved in plasticity. *Biol Psychiatry*. 77(1):52–8.
  126. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, Duncan L, Perry JRB, Patterson N, Robinson EB, Daly MJ, Price AL, Neale BM. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 47(11):1236–41.
  127. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, Bruffaerts R, Wai TC, De Girolamo G, Gluzman S, De Graaf R, Gureje O, Haro JM, Huang Y, Karam E, Kessler RC, Lepine JP, Levinson D, Medina-Mora ME, Ono Y, Posada-Villa J, Williams D. (2008). Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry*. 192(2):98–105.
  128. Petersen L, Sørensen TI a, Kragh Andersen P, Bo Mortensen P, Hawton K. (2014). Genetic and familial environmental effects on suicide attempts: A study of Danish adoptees and their biological and adoptive siblings. *J Affect Disord*. 155(1):273–7.
  129. Tidemalm D, Runeson B, Waern M, Frisell T, Carlström E, Lichtenstein P, Långström N. (2011). Familial clustering of suicide risk: a total population study of 11.4 million individuals. *Psychol Med*. 41(12):2527–34.
  130. Voracek M, Loibl LM. (2007). Genetics of suicide: A systematic review of twin

- studies. *Wien Klin Wochenschr.* 119(15–16):463–75.
131. Antypa N, Serretti A, Rujescu D. (2013). Serotonergic genes and suicide: A systematic review. *Eur Neuropsychopharmacol.* 23(10):1125–42.
  132. Schild AHE, Pietschnig J, Tran US, Voracek M. (2013). Genetic association studies between SNPs and suicidal behavior: A meta-analytical field synopsis. *Prog Neuro-Psychopharmacology Biol Psychiatry.* 46:36–42.
  133. Tsai SJ, Hong CJ, Liou YJ. (2011). Recent molecular genetic studies and methodological issues in suicide research. *Prog Neuro-Psychopharmacology Biol Psychiatry.* 35(4):809–17.
  134. Fiori LM, Wanner B, Jomphe V, Croteau J, Vitaro F, Tremblay RE, Bureau A, Turecki G. (2010). Association of polyaminergic loci with anxiety, mood disorders, and attempted suicide. *PLoS One.* 5(11):e15146.
  135. Roy A, Hodgkinson CA, DeLuca V, Goldman D, Enoch MA. (2012). Two HPA axis genes, CRHBP and FKBP5, interact with childhood trauma to increase the risk for suicidal behavior. *J Psychiatr Res.* 46(1):72–9.
  136. Dwivedi Y. (2010). Brain-derived neurotrophic factor and suicide pathogenesis. *Ann Med.* 42(2):87–96.
  137. Dwivedi Y, Rizavi HSH, Conley RR, Roberts RC, Tamminga CA, Pandey GN. (2003). Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Arch Gen Psychiatry.* 60:804–15.
  138. Dwivedi Y, Rizavi HS, Zhang H, Mondal AC, Roberts RC, Conley RR, Pandey GN. (2009). Neurotrophin Receptor Activation and Expression in Human Postmortem Brain: Effect of Suicide. *Biol Psychiatry.* 65(4):319–28.
  139. Easton JB, Royer AR, Middlemas DS. (2006). The protein tyrosine phosphatase, Shp2, is required for the complete activation of the RAS/MAPK pathway by brain-derived neurotrophic factor. *J Neurochem.* 97(3):834–45.
  140. Kohli MA, Salyakina D, Pfennig A, Lucae S, Horstmann S, Menke A, Kloiber S, Hennings J, Bradley BB, Ressler KJ, Uhr M, Müller-Myhsok B, Holsboer F, Binder EB, Article O. (2010). Association of genetic variants in the neurotrophic receptor-encoding gene NTRK2 and a lifetime history of suicide attempts in depressed patients. *Arch Gen Psychiatry.* 67(4):348–59.

141. Kim B, Kim CY, Hong JP, Kim SY, Lee C, Joo YH. (2008). Brain-derived neurotrophic factor Val/Met polymorphism and bipolar disorder. Association of the Met allele with suicidal behavior of bipolar patients. *Neuropsychobiology*. 58(2):97–103.
142. Juhasz G, Dunham JS, McKie S, Thomas E, Downey D, Chase D, Lloyd-Williams K, Toth ZG, Platt H, Mekli K, Payton A, Elliott R, Williams SR, Anderson IM, Deakin JFW. (2011). The CREB1-BDNF-NTRK2 pathway in depression: multiple gene-cognition-environment interactions. *Biol Psychiatry*. 69(8):762–71.
143. Ernst C, Deleva V, Deng X, Sequeira A, Pomarenski A, Klempan T, Ernst N, Quirion R, Gratton A, Szyf M, Turecki G. (2009). Alternative splicing, methylation state, and expression profile of tropomyosin-related kinase B in the frontal cortex of suicide completers. *Arch Gen Psychiatry*. 66(1):22–32.
144. Serafini G, Pompili M, Hansen KF, Obrietan K, Dwivedi Y, Shomron N, Girardi P. (2014). The Involvement of MicroRNAs in Major Depression, Suicidal Behavior, and Related Disorders: A Focus on miR-185 and miR-491-3p. *Cell Mol Neurobiol*. 34(1):17–30.
145. Smalheiser NR, Lugli G, Zhang H, Rizavi H, Cook EH, Dwivedi Y. (2014). Expression of microRNAs and other small RNAs in prefrontal cortex in schizophrenia, bipolar disorder and depressed subjects. *PLoS One*. 9(1):e86469.
146. Maussion G, Yang J, Yerko V, Barker P, Mechawar N, Ernst C, Turecki G. (2012). Regulation of a truncated form of tropomyosin-related kinase B (TrkB) by Hsa-miR-185\* in frontal cortex of suicide completers. Hoheisel JD, editor. *PLoS One*. 7(6):e39301.
147. Menke A, Domschke K, Czamara D, Klengel T, Hennings J, Lucae S, Baune BT, Arolt V, Müller-Myhsok B, Holsboer F, Binder EB. (2012). Genome-Wide Association Study of Antidepressant Treatment-Emergent Suicidal Ideation. *Neuropsychopharmacology*. 37(3):797–807.
148. Perroud N, Uher R, Ng MYM, Guipponi M, Hauser J, Henigsberg N, Maier W, Mors O, Gennarelli M, Rietschel M, Souery D, Dernovsek MZ, Stamp AS, Lathrop M, Farmer A, Breen G, Aitchison KJ, Lewis CM, Craig IW, McGuffin P. (2012). Genome-wide association study of increasing suicidal ideation during

- antidepressant treatment in the GENDEP project. *Pharmacogenomics J.* 12(1):68–77.
149. Mullins N, Perroud N, Uher R, Butler AW, Cohen-Woods S, Rivera M, Malki K, Euesden J, Power R a., Tansey KE, Jones L, Jones I, Craddock N, Owen MJ, Korszun A, Gill M, Mors O, Preisig M, Maier W, Rietschel M, Rice JP, Müller-Myhsok B, Binder EB, Lucae S, Ising M, Craig IW, Farmer AE, McGuffin P, Breen G, Lewis CM. (2014). Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: A genome-wide association and polygenic scoring study. *Am J Med Genet Part B Neuropsychiatr Genet.* 165(5):428–37.
  150. Perlis RH, Huang J, Purcell S, Fava M, Rush AJ, Sullivan PF, Hamilton SP, McMahon FJ, Schulze T, Potash JB, Zandi PP, Willour VL, Penninx BW, Boomsma DI, Vogelzangs N, Middeldorp CM, Rietschel M, Nöthen M, Cichon S, Gurling H, Bass N, McQuillin A, Hamshere M, Craddock N, Sklar P, Smoller JW. (2010). Genome-Wide Association Study of Suicide Attempts in Mood Disorder Patients. *Am J Psychiatry.* 167(12):1499–507.
  151. Galfalvy H, Haghighi F, Hodgkinson C, Goldman D, Oquendo MA, Burke A, Huang Y, Giegling I, Rujescu D, Bureau A, Turecki G, Mann JJ. (2015). A genome-wide association study of suicidal behavior. *Am J Med Genet B Neuropsychiatr Genet.* 168(7):557–63.
  152. Galfalvy H, Zalsman G, Huang Y-Y, Murphy L, Rosoklija G, Dwork AJ, Haghighi F, Arango V, Mann JJ. (2011). A pilot genome wide association and gene expression array study of suicide with and without major depression. *World J Biol Psychiatry.* 14(8):1–9.
  153. Laje G, Allen AS, Akula N, Manji H, John Rush A, McMahon FJ. (2009). Genome-wide association study of suicidal ideation emerging during citalopram treatment of depressed outpatients. *Pharmacogenet Genomics.* 19(9):666–74.
  154. Schosser A, Butler AW, Ising M, Perroud N, Uher R, Ng MY, Cohen-Woods S, Craddock N, Owen MJ, Korszun A, Jones L, Jones I, Gill M, Rice JP, Maier W, Mors O, Rietschel M, Lucae S, Binder EB, Preisig M, Perry J, Tozzi F, Muglia P, Aitchison KJ, Breen G, Craig IW, Farmer AE, Müller-Myhsok B, McGuffin P, Lewis CM. (2011). Genomewide association scan of suicidal thoughts and

- behaviour in major depression. *PLoS One*. 6(7):e20690.
155. Willour V, Seifuddin F, Mahon P, Jancic D, Pirooznia M, Steele J, Schweizer B, Goes F, Mondimore F, Mackinnon D, Consortium BGS, Perlis R, Lee P, Huang J, Kelsoe J, Shilling P, Rietschel M, Nöthen M, Cichon S, Gurling H, Purcell S, Smoller J, Craddock N, DePaulo J, Schulze T, McMahon F, Zandi P, Potash J. (2012). A genome-wide association study of attempted suicide. *Mol Psychiatry*. 17(4):433–77.
  156. Ferreira M a R, O’Donovan MC, Meng Y a, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee K a, Williamson R, MacIntyre DJ, MacLean AW, St CD, Robinson M, Van Beck M, Pereira ACP, Kandaswamy R, McQuillin A, Collier D a, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans P a, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 40(9):1056–8.
  157. Lett TAP, Zai CC, Tiwari AK, Shaikh SA, Likhodi O, Kennedy JL, Muller DJ. (2011). ANK3, CACNA1C and ZNF804A gene variants in bipolar disorders and psychosis subphenotype. *World J Biol Psychiatry*. 12(5):392–7.
  158. Laje G, Paddock S, Manji H, Rush AJ, Wilson AF, Charney D, McMahon FJ. (2007). Genetic markers of suicidal ideation emerging during citalopram treatment of major depression. *Am J Psychiatry*. 164(10):1530–8.
  159. Inczédy-Farkas G, Benkovits J, Balogh N, Álmos P, Scholtz B, Zahuczky G, Török Z, Nagy K, Réthelyi J, Makkos Z, Kassai-Farkas Á, Égerházy A, Túzkó J, Janka Z, Bitter I, Németh G, Nagy L, Molnár MJ. (2010). SCHIZOBANK – The Hungarian national schizophrenia biobank and its role in schizophrenia research. *Orv Hetil*. 151(35):1403–8.
  160. Dick DM, Foroud T, Flury L, Bowman ES, Miller MJ, Rau NL, Moe PR, Samavedy N, El-Mallakh R, Manji H, Glitz DA, Meyer ET, Smiley C, Hahn R,

- Widmark C, McKinney R, Sutton L, Ballas C, Grice D, Berrettini W, Byerley W, Coryell W, DePaulo R, MacKinnon DF, Gershon ES, Kelsoe JR, McMahon FJ, McInnis M, Murphy DL, Reich T, Scheftner W, Nurnberger JIJ. (2003). Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet.* 73(1):107–14.
161. McInnis MG, Dick DM, Willour VL, Avramopoulos D, MacKinnon DF, Simpson SG, Potash JB, Edenberg HJ, Bowman ES, McMahon FJ, Smiley C, Chellis JL, Huo Y, Diggs T, Meyer ET, Miller M, Matteini AT, Rau NL, DePaulo JR, Gershon ES, Badner JA, Rice JP, Goate AM, Detera-Wadleigh SD, Nurnberger JI, Reich T, Zandi PP, Foroud TM. (2003). Genome-wide scan and conditional analysis in bipolar disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. *Biol Psychiatry.* 54(11):1265–73.
162. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zöllner S, Schork NJ, Kelsoe JR. (2009). Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry.* 14(8):755–63.
163. Boomsma DI, Beem AL, van den Berg M, Dolan C V, Koopmans JR, Vink JM, de Geus EJ, Slagboom PE. (2000). Netherlands twin family study of anxious depression (NETSAD). *Twin Res.* 3(4):323–34.
164. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, Kluff C, Smit G, Nolen WA, Zitman FG, Smit JH, Hoogendijk WJ, van Dyck R, de Geus EJC, Penninx BWJH. (2008). Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* 16(3):335–42.
165. Smith DJ, Escott-Price V, Davies G, Bailey MES, Colodro-Conde L, Ward J, Vedernikov A, Marioni R, Cullen B, Lyall D, Hagenaars SP, Liewald DCM,

- Luciano M, Gale CR, Ritchie SJ, Hayward C, Nicholl B, Bulik-Sullivan B, Adams M, Couvy-Duchesne B, Graham N, Mackay D, Evans J, Smith BH, Porteous DJ, Medland SE, Martin NG, Holmans P, McIntosh AM, Pell JP, Deary IJ, O'Donovan MC. (2016). Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry*. 21(11):1644.
166. Delaneau O, Marchini J, Zagury J-F. (2012). A linear complexity phasing method for thousands of genomes. *Nat Methods*. 9(2):179–81.
167. Mann JJ, Arango V a., Avenevoli S, Brent D a., Champagne F a., Clayton P, Currier D, Dougherty DM, Haghighi F, Hodge SE, Kleinman J, Lehner T, McMahon F, Mościcki EK, Oquendo M a., Pandey GN, Pearson J, Stanley B, Terwilliger J, Wenzel A. (2009). Candidate Endophenotypes for Genetic Studies of Suicidal Behavior. *Biol Psychiatry*. 65(7):556–63.
168. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 23(10):1294–6.
169. Alexander DH, Novembre J, Lange K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 19(9):1655–64.
170. Han B, Eskin E. (2012). Interpreting meta-analyses of genome-wide association studies. *PLoS Genet*. 8(3):e1002555.
171. Li M-X, Sham PC, Cherny SS, Song Y-Q. (2010). A knowledge-based weighting framework to boost the power of genome-wide association studies. Mailund T, editor. *PLoS One*. 5(12):e14480.
172. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet*. 43(6):519–25.
173. Euesden J, Lewis CM, O'Reilly PF. (2014). PRSice: Polygenic Risk Score software. *Bioinformatics*. 31(9):1466–8.
174. Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, Burrell GJ, Rice JP, Nertney DA, Olincy A, Rozic P, Vinogradov S, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Crowe RR,

- Cloninger CR, Martinez M, Gejman P V. (2008). No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry*. 165(4):497–506.
175. Suarez BK, Duan J, Sanders AR, Hinrichs AL, Jin CH, Hou C, Buccola NG, Hale N, Weilbaecher AN, Nertney DA, Olincy A, Green S, Schaffer AW, Smith CJ, Hannah DE, Rice JP, Cox NJ, Martinez M, Mowry BJ, Amin F, Silverman JM, Black DW, Byerley WF, Crowe RR, Freedman R, Cloninger CR, Levinson DF, Gejman P V. (2006). Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. *Am J Hum Genet*. 78(2):315–33.
176. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CCA, Howie B, Leung H-T, Hartmann AM, Möller H-J, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman P V, Cichon S, Nöthen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Cloninger CR, Molecular Genetics of Schizophrenia Collaboration. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 40(9):1053–5.
177. Li H, Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26(5):589–95.
178. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078–9.
179. Wang K, Li M, Hakonarson H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 38(16):e164.
180. Li M-X, Gui H-S, Kwan JSH, Bao S-Y, Sham PC. (2012). A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian



- diseases. *Nucleic Acids Res.* 40(7):e53.
181. Laird NM, Horvath S, Xu X. (2000). Implementing a unified approach to family-based tests of association. *Genet Epidemiol.* 19 Suppl 1(S1):S36-42.
  182. Abraham G, Inouye M, Zhang Y. (2014). Fast Principal Component Analysis of Large-Scale Genome-Wide Data. *PLoS One.* 9(4).
  183. Vivar JC, Pemu P, McPherson R, Ghosh S. (2013). Redundancy control in pathway databases (ReCiPa): an application for improving gene-set enrichment analysis in Omics studies and “Big data” biology. *OMICS.* 17(8):414–22.
  184. Li M-X, Kwan JSH, Bao S-Y, Yang W, Ho S-L, Song Y-Q, Sham PC. (2013). Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet.* 9(1):e1003143.
  185. Li MJ, Pan Z, Liu Z, Wu J, Wang P, Zhu Y, Xu F, Xia Z, Sham PC, Kocher J-PA, Li M, Liu JS, Wang J. (2016). Predicting regulatory variants with composite statistic. *Bioinformatics.* 32(18):2729–36.
  186. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat Genet.* 48(2):214–20.
  187. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature.* 489(7414):57–74.
  188. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E, Kähler AK, Hultman CM, Purcell SM, McCarroll SA, Daly M, Pasaniuc B, Sullivan PF, Neale BM, Wray NR, Raychaudhuri S, Price AL. (2014). Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *Am J Hum Genet.* 95(5):535–52.
  189. Gupta VK, You Y, Gupta VB, Klistorner A, Graham SL. (2013). TrkB receptor signalling: Implications in neurodegenerative, psychiatric and proliferative disorders. *Int J Mol Sci.* 14(5):10122–42.
  190. Wang Z, Fan J, Gao K, Li Z, Yi Z, Wang L, Huang J, Yuan C, Hong W, Yu S, Fang Y. (2013). Neurotrophic tyrosine kinase receptor type 2 (NTRK2) gene associated with treatment response to mood stabilizers in patients with bipolar I disorder. *J Mol Neurosci.* 50(2):305–10.
  191. Cattaneo A, Gennarelli M, Uher R, Breen G, Farmer A, Aitchison KJ, Craig IW,

- Anacker C, Zunsztain P a, McGuffin P, Pariante CM. (2013). Candidate genes expression profile associated with antidepressants response in the GENDEP study: differentiating between baseline “predictors” and longitudinal “targets”. *Neuropsychopharmacology*. 38(3):377–85.
192. Wong J, Garner B. (2012). Evidence that truncated TrkB isoform, TrkB-Shc can regulate phosphorylated TrkB protein levels. *Biochem Biophys Res Commun*. 420(2):331–5.
193. Ray MT, Shannon Weickert C, Webster MJ. (2014). Decreased BDNF and TrkB mRNA expression in multiple cortical areas of patients with schizophrenia and mood disorders. *Transl Psychiatry*. 4(5):e389.
194. Qi X-R, Zhao J, Liu J, Fang H, Swaab DF, Zhou J-N. (2015). Abnormal retinoid and TrkB signaling in the prefrontal cortex in mood disorders. *Cereb Cortex*. 25(1):75–83.
195. Miller BH, Wahlestedt C. (2010). MicroRNA dysregulation in psychiatric disease. *Brain Res*. 1338:89–99.
196. Xu B, Hsu PK, Karayiorgou M, Gogos J a. (2012). MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction. *Neurobiol Dis*. 46(2):291–301.
197. Stark KL, Xu B, Bagchi A, Lai W-S, Liu H, Hsu R, Wan X, Pavlidis P, Mills A a, Karayiorgou M, Gogos J a. (2008). Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat Genet*. 40(6):751–60.
198. Konopka W, Kiryk A, Novak M, Herwerth M, Parkitna JR, Wawrzyniak M, Kowarsch A, Michaluk P, Dzwonek J, Arnsperger T, Wilczynski G, Merkschlager M, Theis FJ, Köhr G, Kaczmarek L, Schütz G. (2010). MicroRNA loss enhances learning and memory in mice. *J Neurosci*. 30(44):14835–42.
199. Beveridge NJ, Santarelli DM, Wang X, Tooney P a, Webster MJ, Weickert CS, Cairns MJ. (2013). Maturation of the Human Dorsolateral Prefrontal Cortex Coincides With a Dynamic Shift in MicroRNA Expression. *Schizophr Bull*. 40(2):1–11.
200. Smalheiser NR, Lugli G, Rizavi HS, Torvik VI, Turecki G, Dwivedi Y. (2012).

- MicroRNA expression is down-regulated and reorganized in prefrontal cortex of depressed suicide subjects. *PLoS One*. 7(3):e33201.
201. Huang YW a, Ruiz CR, Eyley ECH, Lin K, Meffert MK. (2012). Dual regulation of miRNA biogenesis generates target specificity in neurotrophin-induced protein synthesis. *Cell*. 148(5):933–46.
  202. Chen J, Lee G, Fanous AH, Zhao Z, Jia P, O'Neill A, Walsh D, Kendler KS, Chen X, International Schizophrenia Consortium. (2011). Two non-synonymous markers in PTPN21, identified by genome-wide association study data-mining and replication, are associated with schizophrenia. *Schizophr Res*. 131(1–3):43–51.
  203. Plani-Lam JH-C, Chow T-C, Siu K-L, Chau WH, Ng M-HJ, Bao S, Ng CT, Sham P, Shum DK-Y, Ingleby E, Jin D-Y, Song Y-Q. (2015). PTPN21 exerts pro-neuronal survival and neuritic elongation via ErbB4/NGF signaling. *Int J Biochem Cell Biol*. 61:53–62.
  204. Hess JL, Tylee DS, Barve R, de Jong S, Ophoff RA, Kumarasinghe N, Tooney P, Schall U, Gardiner E, Beveridge NJ, Scott RJ, Yasawardene S, Perera A, Mendis J, Carr V, Kelly B, Cairns M, Tsuang MT, Glatt SJ. (2016). Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia. *Schizophr Res*. 176(2–3):114–24.
  205. Arenas E, Denham M, Villaescusa JC. (2015). How to make a midbrain dopaminergic neuron. *Development*. 142(11):1918–36.
  206. Liu Y, Li Z, Zhang M, Deng Y, Yi Z, Shi T. (2013). Exploring the pathogenetic association between schizophrenia and type 2 diabetes mellitus diseases based on pathway analysis. *BMC Med Genomics*. 6 Suppl 1(Suppl 1):S17.
  207. Gagnoli C, Reeves GM, Reazer J, Postolache TT. (2016). Dopamine-prolactin pathway potentially contributes to the schizophrenia and type 2 diabetes comorbidity. *Transl Psychiatry*. 6(4):e785.
  208. Jia P, Han G, Zhao J, Lu P, Zhao Z. (2016). SZGR 2.0: a one-stop shop of schizophrenia candidate genes. *Nucleic Acids Res*. :gkw902.

## 10. Saját közlemények

I. Az értekezés témájában megjelent eredeti közlemények:

1. **Pulay AJ**, Rethelyi JM. (2016) Multimarker analysis suggests the involvement of BDNF signaling and microRNA biosynthesis in suicidal behavior. **AMERICAN JOURNAL OF MEDICAL GENETICS PART B-NEUROPSYCHIATRIC GENETICS** 171:(6) pp. 763-776. **IF: 3,391**

2. **Pulay AJ**, Koller J, Nagy L, Molnár MJ, Réthelyi J, Magyar SCHIZOBANK Konzorcium munkatársai. (2017) A szkizofrénia multilókusz genetikai vizsgálata az idegfejlődés és az immunrendszer zavarának oki szerepére utal(hat). **IDEGGYOGYASZATI SZEMLE-CLINICAL NEUROSCIENCE** 70:(3-4) pp. 115-126. **IF: 0,376**

II. Egyéb – nem az értekezés témájában megjelent – eredeti közlemények:

1. Benkovits J, Magyarosi S, **Pulay AJ**, Makkos Z, Egerhazi A, Balogh N, Almos P, Liko I, Schizobank Consortium H, Nemeth G, Molnar JM, Nagy L, Rethelyi JM- (2016) CNTF, COMT, DDR1, DISC1, DRD2, DRD3 es DTNBP1 kandidáns gének vizsgálata szkizofréniaában: eredmények a Magyar SCHIZOBANK Konzorcium vizsgálatából. **NEUROPSYCHOPHARMACOLOGIA HUNGARICA** 18:(4) pp. 181-187.

2. **Pulay AJ**, Bitter I, Papp S, Gulácsi L, Péntek M, Brodszky V, Hevér NV, Rencz F, Baji P. (2016) Exploring the Relationship between Quality of Life (EQ-5D) and Clinical Measures in Adult Attention Deficit Hyperactivity Disorder (ADHD) **APPLIED RESEARCH IN QUALITY OF LIFE** First online: 20 April 2016: pp. 1-16. **IF: 0,894**

3. Szkultecka-Debek M, Walczak J, Augustynska J, Miernik K, Stelmachowski J, Pieniazek I, Obrzut G, Pogroszewska A, Paulic G, Damir M, Antolic S, Tavcar R, Indrikson A, Aadamsoo K, Jankovic S, **Pulay AJ**, Rimay J, Varga M, Sulkova I, Verzun P. (2015) Epidemiology and Treatment Guidelines of Negative Symptoms in Schizo-phrenia in Central and Eastern Europe: A Literature Review. **CLINICAL PRACTICE AND EPIDEMIOLOGY IN MENTAL HEALTH** 11: pp. 158-165.

4. Kerridge BT, Saha TD, Smith S, Chou PS, Pickering RP, Huang B, Ruan JW, **Pulay AJ**. (2011) Dimensionality of hallucinogen and inhalant/solvent abuse and dependence criteria: implications for the diagnostic and statistical manual of mental disorders-Fifth edition. **ADDICTIVE BEHAVIORS** 36:(9) pp. 912-918. **IF: 2,085**

5. Pickering RP, Goldstein RB, Hasin DS, Blanco C, Smith SM, Huang B, **Pulay AJ**, Ruan WJ, Saha TD, Stinson FS, Dawson DA, Chou SP, Grant BF. (2011) Temporal relationships between overweight and obesity and DSM-IV substance use, mood, and anxiety disorders: results from a prospective study, the National Epidemiologic Survey on Alcohol and Related Conditions. **JOURNAL OF CLINICAL PSYCHIATRY** 72:(11) pp. 1492-1502. **IF: 5,799**

6. Dawson DA, **Pulay AJ**, Grant BF. (2010) A comparison of two single-item screeners for hazardous drinking and alcohol use disorder. **ALCOHOLISM-CLINICAL AND EXPERIMENTAL RESEARCH** 34:(2) pp. 364-374. **IF: 3,468**
7. **Pulay AJ**, Stinson FS, Ruan WJ, Smith SM, Pickering RP, Dawson DA, Grant BF. (2010) The relationship of DSM-IV personality disorders to nicotine dependence-results from a national survey. **DRUG AND ALCOHOL DEPENDENCE** 108:(1-2) pp. 141-145. **IF: 3,365**
8. Saha TD, Compton WM, **Pulay AJ**, Stinson FS, Ruan WJ, Smith SM, Grant BF. (2010) Dimensionality of DSM-IV nicotine dependence in a national sample: an item response theory application. **DRUG AND ALCOHOL DEPENDENCE** 108:(1-2) pp. 21-28. **IF: 3,365**
9. Srivastava V, Buzas B, Momenan R, Oroszi G, **Pulay AJ**, Enoch MA, Hommer DW, Goldman D. (2010) Association of SOD2, a mitochondrial antioxidant enzyme, with gray matter volume shrinkage in alcoholics. **NEUROPSYCHOPHARMACOLOGY** 35:(5) pp. 1120-1128. **IF: 6,685**
10. Grant BF, Goldstein RB, Chou SP, Huang B, Stinson FS, Dawson DA, Saha TD, Smith SM, **Pulay AJ**, Pickering RP, Ruan WJ, Compton WM. (2009) Sociodemographic and psychopathologic predictors of first incidence of DSM-IV substance use, mood and anxiety disorders: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. **MOLECULAR PSYCHIATRY** 14:(11) pp. 1051-1066. **IF: 15,049**
11. **Pulay AJ**, Stinson FS, Dawson DA, Goldstein RB, Chou SP, Huang B, Saha TD, Smith SM, Pickering RP, Ruan WJ, Hasin DS, Grant BF. (2009) Prevalence, Correlates, Disability, and Comorbidity of DSM-IV Schizotypal Personality Disorder: Results From the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. **PRIMARY CARE COMPANION TO THE JOURNAL OF CLINICAL PSYCHIATRY** 11:(2) pp. 53-67.
12. Grant BF, Chou SP, Goldstein RB, Huang B, Stinson FS, Saha TD, Smith SM, Dawson DA, **Pulay AJ**, Pickering RP, Ruan WJ. (2008) Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. **JOURNAL OF CLINICAL PSYCHIATRY** 69:(4) pp. 533-545. **IF: 5,053**
13. Gundy S, Szekely G, Farkas G, **Pulay A**, Remenar E. (2008) Biomarkerek alkalmazása során felmerülő problémák malignus és nem malignus betegségben szenvedő alkoholisták esetében: Problems occurring in the application of cytogenetic biomarkers for alcoholics with and without malignant diseases. **MAGYAR ONKOLÓGIA** 52:(2) pp. 153-161.
14. **Pulay AJ**, Dawson DA, Ruan WJ, Pickering RP, Huang B, Chou SP, Grant BF. (2008) The relationship of impairment to personality disorder severity among individuals with specific axis I disorders: results from the National Epidemiologic

Survey on Alcohol and Related Conditions. **JOURNAL OF PERSONALITY DISORDERS** 22:(4) pp. 405-417. **IF: 2,034**

15. **Pulay AJ**, Dawson DA, Hasin DS, Goldstein RB, Ruan WJ, Pickering RP, Huang B, Chou SP, Grant BF. (2008) Violent behavior and DSM-IV psychiatric disorders: results from the national epidemiologic survey on alcohol and related conditions. **JOURNAL OF CLINICAL PSYCHIATRY** 69:(1) pp. 12-22. **IF: 5,053**

16. Stinson FS, Dawson DA, Goldstein RB, Chou SP, Huang B, Smith SM, Ruan WJ, **Pulay AJ**, Saha TD, Pickering RP, Grant BF. (2008) Prevalence, correlates, disability, and comorbidity of DSM-IV narcissistic personality disorder: results from the wave 2 national epidemiologic survey on alcohol and related conditions. **JOURNAL OF CLINICAL PSYCHIATRY** 69:(7) pp. 1033-1045. **IF: 5,053**

17. Goldstein RB, Compton WM, **Pulay AJ**, Ruan WJ, Pickering RP, Stinson FS, Grant BF. (2007) Antisocial behavioral syndromes and DSM-IV drug use disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. **DRUG AND ALCOHOL DEPENDENCE** 90:(2-3) pp. 145-158. **IF: 3,222**

## **11. Köszönetnyilvánítás**

Értekezésem lezárásakor szeretnék köszönetet mondani minden segítségért. Elsősorban feleségemnek, Erzsinek, a végtelen hosszú listát kitevő támogatásáért, és témavezetőmnek, Réthelyi Jánosnak, a kifogyhatatlan türelméért, iránymutatásaiért és a szükséges építő kritikákért. Köszönöm minden kollégáimnak, de különösen Erikának, Líviának és Eszternek, hogy lehetővé tették, hogy csak a kutatásomra fókuszáljak, és pácienseimnek, akik türelmesen alkalmazkodtak a megritkult időpontokhoz!

# Multimarker Analysis Suggests the Involvement of BDNF Signaling and microRNA Biosynthesis in Suicidal Behavior

Attila J. Pulay<sup>1\*</sup> and János M. Réthelyi<sup>1,2</sup>

<sup>1</sup>Department of Psychiatry and Psychotherapy, Semmelweis University, Budapest, Hungary

<sup>2</sup>Molecular Psychiatry Research Group, MTA-SE NAP-B, Hungarian Academy of Sciences and Semmelweis University, Budapest, Hungary

Manuscript Received: 17 July 2015; Manuscript Accepted: 4 February 2016

Despite moderate heritability estimates the genetics of suicidal behavior remains unclear, genome-wide association and candidate gene studies focusing on single nucleotide associations reported inconsistent findings. Our study explored biologically informed, multimarker candidate gene associations with suicidal behavior in mood disorders. We analyzed the GAIN Whole Genome Association Study of Bipolar Disorder version 3 ( $n = 999$ , suicidal  $n = 358$ ) and the GAIN Major Depression: Stage 1 Genomewide Association in Population-Based Samples ( $n = 1,753$ , suicidal  $n = 245$ ) datasets. Suicidal behavior was defined as severe suicidal ideation or attempt. Candidate genes were selected based on literature search (Geneset1,  $n = 35$ ), gene expression data of microRNA genes, (Geneset2,  $n = 68$ ) and their target genes (Geneset3,  $n = 11,259$ ). Quality control, dosage analyses were carried out with PLINK. Gene-based associations of Geneset1 were analyzed with KGG. Polygenic profile scores of suicidal behavior were computed in the major depression dataset both with PRSice and LDpred and validated in the bipolar disorder data. Several nominally significant gene-based associations were detected, but only DICER1 associated with suicidal behavior in both samples, while only the associations of NTRK2 in the depression sample reached family wise and experiment wise significance. Polygenic profile scores negatively predicted suicidal behavior in the bipolar sample for only Geneset2, with the strongest prediction by PRSice at  $P_t < 0.03$  (Nagelkerke  $R^2 = 0.01$ ,  $P < 0.007$ ). Gene-based association results confirmed the potential involvement of the BDNF-NTRK2-CREB pathway in the pathogenesis of suicide and the cross-disorder association of DICER1. Polygenic risk prediction of the selected miRNA genes indicates that the miRNA system may play a mediating role, but with considerable pleiotropy. © 2016 Wiley Periodicals, Inc.

**Key words:** executive functions; polygenic risk score; suicidal behavior; Bayesian PRS; region-based heritability

## INTRODUCTION

Suicide is the most devastating outcome of psychiatric disorders and according to the 2000–2012 Global Health Estimates report on mortality by the World Health Organization (WHO) [Organization,

### How to Cite this Article:

Pulay AJ, Réthelyi JM. 2016. Multimarker Analysis Suggests the Involvement of BDNF Signaling and microRNA Biosynthesis in Suicidal Behavior.

Am J Med Genet Part B 9999:1–14.

2014], it is the second leading cause of death among 15–29 year olds and listed in the top 20 cause of death worldwide. The link between suicide and mental health problems is strong, approximately 90% of suicide attempts occur among individuals with psychiatric disorders. Based on the data from the WHO World Mental Health Survey Initiative, a cross-national study by Nock et al. [2008] revealed that mood disorders are particularly strong, independent predictors of both suicide attempts (OR: 4.7–5.9) and ideation (OR: 3.4–4.7) even if sociodemographic correlates were taken into account. Although suicidal ideation is arguably more common than suicide attempts (lifetime prevalence: 9.2% and 2.7%, respectively), they found that 29% of respondents with suicidal ideation and 59% of those who made definite plans also tried to commit suicide. Moreover, authors reported that the strong association between psychiatric disorders and suicide attempts substantially diminished after suicidal ideation was controlled for in the regression models. These results also indicate that factors other than the psychiatric disorders per se also mediate the transition from suicidal ideation to attempt. One of these potential mediators might be the impairment in executive

Authors have no conflict of interests to declare.

Grant sponsor: Hungarian National Research Fund; Grant sponsor: Hungarian Brain Research Program; Grant numbers: OTKA-PD 83876, NAP-B KTIA\_NAP\_13-2014-0011.

\*Correspondence to

Attila J. Pulay, M.D., Department of Psychiatry and Psychotherapy, Semmelweis University, H-1083 Budapest, Balassa u. 6. Hungary.

E-mail: pulay.attila@med.semmelweis-univ.hu

Article first published online in Wiley Online Library (wileyonlinelibrary.com): 00 Month 2016

DOI 10.1002/ajmg.b.32433



functions, for example, cognitive inflexibility, impaired decision making, attention and working memory deficits, and weak impulse control, that have been repeatedly linked to both suicide ideation and attempt in various psychiatric disorders [Marzuk et al., 2005; LeGris and van Reekum, 2006; Westheide et al., 2008; Malloy-Diniz et al., 2009; Keilp et al., 2012; Miranda and Gallagher, 2012; Richard-Devantoy et al., 2012]. In concert with the neuropsychological findings, structural and functional imaging studies of suicidal behavior showed alterations in the volumes and activity of the non-motor neural networks of the prefrontal cortex and basal ganglia, particularly the dorsolateral and orbitofrontal circuits which are responsible for organizing the aforementioned executive functions [Van Heeringen, 2003].

Similar to other psychiatric conditions, suicidal behavior is suspected to have a complex, multifactorial etiology. Although genetic epidemiologic studies [Voracek and Loibl, 2007; Tidemalm et al., 2011; Petersen et al., 2014] suggest that genetic factors have definitive role in suicidal behavior, the genetic architecture of suicide is yet to be clarified. Based on the observed link between low CNS serotonin level and suicide [Van Heeringen, 2003], the early candidate gene studies targeted the serotonergic system, especially the TPH2 and SLC6A4 genes [Antypa et al., 2013]. However, subsequent candidate and whole-genome association studies (GWAS), and systematic meta-analyses [Schild et al., 2013] failed to replicate the initial associations. The focus of the hypothesis driven studies gradually shifted toward genes from other neurochemical systems, such as dopaminergic transmission (DRD2, COMT, MAOA) [Tsai et al., 2011], polyamines (SAT1) [Fiori et al., 2010], stress response genes (FKBP5, HPA axis genes) [Tsai et al., 2011; Roy et al., 2012] and recently to neurotrophins, especially the genes of the BDNF-NTRK2-CREB pathway [Dwivedi et al., 2003; Kim et al., 2008; Dwivedi et al., 2009; Ernst et al., 2009; Dwivedi, 2010; Kohli et al., 2010; Juhasz et al., 2011]. In addition, because of the apparently state-like impairment of executive functioning in both suicidal ideation and attempt, gene expression studies in areas responsible for these cognitive domains, that is, the prefrontal cortex, also indicated the potential role of the microRNA (miRNA) system in suicidal behavior [Serafini et al., 2014; Smalheiser et al., 2014]. Particularly hsa-miR-491-3p, hsa-miR-185, hsa-miR-195, and hsa-mir30a should be highlighted, as hsa-miR-491-3p and hsa-miR-185 were found upregulated in the prefrontal cortex of suicide victims [Serafini et al., 2014], furthermore hsa-miR-185, hsa-miR-195, and hsa-mir30a play important regulatory role in the BDNF signaling pathway [MauSSION et al., 2012; Serafini et al., 2014].

Unfortunately, the GWAS studies [Laje et al., 2009; Perlis et al., 2010; Galfalvy et al., 2011; Schosser et al., 2011; Menke et al., 2012; Perroud et al., 2012; Willour et al., 2012; Mullins et al., 2014; Galfalvy et al., 2015] exploring suicidal behavior to date did not clarify the aforementioned complexity. With the exception of an intergenic SNP near to the ACP1 gene reported by Willour et al., all of these studies resulted in only suggestive findings at best. Moreover, the genes of these several suggestive associations failed to replicate in two independent GWAS analyses. Potential reasons of these inconsistency include most importantly insufficient power to detect associations with small effect size [Laje et al., 2009; Menke et al., 2012; Perroud et al., 2012], and phenotypic heterogeneity.

Three studies [Laje et al., 2009; Menke et al., 2012; Perroud et al., 2012] analyzed suicidal ideation emerging or worsening after antidepressant treatment, four studies [Perlis et al., 2010; Schosser et al., 2011; Willour et al., 2012; Mullins et al., 2014] used suicide attempt as outcome measure, whereas only one study by Galfalvy et al. [2011] analyzed samples of victims of completed suicide. In addition, further methodologic issues arose, such as low coverage of genotype array, lack of control for psychiatric disorders [Galfalvy et al., 2011].

Besides the aforementioned reasons, the genetic heterogeneity and the assumed polygenic inheritance might have been likely to contribute to the missing replications. As Ayalew et al. [2012] reported, the heterogeneity is particularly strong at the SNP level, but the overlap between datasets could considerably be increased by taking into account more complex genetic markers, such as gene regions, pathways or gene networks. Region-based tests can amplify individually weak SNP signals that are enriched in specific locations, for example, in a gene or in the promoter and have performed well in several psychiatric genetic analyses. Another multimarker test, the polygenic risk scoring (PRS) approach proposed by Purcell et al. [2009] also aims to measure the additive contribution of the genetic markers, however, it differs from the former tests by requiring at least two independent datasets—a training and a validation sample for a proper, unbiased analysis. Although these methods fit the hypothesized polygenic inheritance model of suicidal behavior much more than the SNP-based analyses, such data—especially from cross-disorder studies of suicidal behavior—are still far from complete. Therefore, our study aimed to provide region-based association results in a cross-disorder fashion, between suicidal behavior and those genes that showed suggestive SNP associations in GWAS or expression studies or were clearly implicated by previous candidate gene studies. In addition, we also conducted targeted PRS analyses to assess the additive, polygenic contribution of the literature search based candidate gene set and of those miRNA genes, which might be relevant in executive functioning by using the Genetic Association Information Network (GAIN) major depressive disorder dataset as training sample and the GAIN bipolar disorder dataset as validation sample.

## MATERIALS AND METHODS

### Study Samples

The two non-overlapping groups examined in this study comprised individuals with European ancestry diagnosed with bipolar affective (BD) and major depressive disorder (MDD). Both study groups' data were accessed from the Database of Phenotypes and Genotypes (dbGaP) [Mailman et al., 2007; Tryka et al., 2014] at the National Center for Biotechnology Information (NCBI). Subjects with bipolar disorder were drawn from the GAIN Whole Genome Association Study of Bipolar Disorder [Dick et al., 2003; McInnis et al., 2003] (dbGaP accession number: phs000017.v3.p1) whereas study population with MDD were included from the GAIN Major Depression: Stage 1 Genomewide Association in Population-Based Samples, which was derived from the Netherlands Study on Depression and Anxiety (NESDA) and from the Netherlands Twin Registry (NTR) [Boomsma et al., 2000, 2008]. In the

GAIN bipolar sample, 999 European-American individuals were diagnosed with bipolar I or II disorders by using the Diagnostic Interview for Genetics Studies (DIGS) versions 2, 3, and 4. One thousand seven hundred fifty-three patients with MDD were included from the NESDA-NTR studies, diagnosis of MDD were given by using the Composite International Diagnostic Interview (CIDI). Data collection, inclusion, and exclusion criteria were described in details elsewhere [Boomsma et al., 2000, 2008; Dick et al., 2003; McInnis et al., 2003].

## Suicidal Behavior

Suicidal behavior was defined as (i) 1 or more suicide attempt rated with an at least ambiguous reason by the interviewer and (ii) having severe suicidal ideation (thinking of committing suicide several times a week, or making definite plans of suicide, or stating that he or she would commit suicide if it is possible). Suicidal behavior was assessed among the bipolar cases with the DIGS and with the Inventory of Depressive Symptoms (IDS) and the short form of Beck Depression Inventory (BDI) in the MDD sample, resulting  $n = 245$  MDD cases and  $n = 358$  bipolar cases with suicidal behavior, and  $n = 616$  bipolar and  $n = 1,360$  MDD cases without suicidal behavior, whereas  $n = 25$  bipolar and  $n = 150$  MDD cases were excluded from the analysis due to missing phenotypic information. Except for imputation or cross-validation purposes, control individuals were dropped from the subsequent analyses in both study samples. The rationale of this decision was that he complete lack of assessment for suicidal behavior among the control subjects in the GAIN bipolar sample, and the low number of control subjects in the GAIN MDD sample identified with suicidal behavior ( $n = 2$ ), did not enable the analysis of these subsets without confounding the associations of suicidal behavior with MDD or bipolar disorder. Given the important relationship between younger age and suicide, we assessed the distribution of age and particularly the high-risk group (15–29 year olds) to detect potential confounding or modifier effects. However, neither the mean of age nor the rate of the aforementioned high-risk age group differed significantly by suicidal behavior in both study samples.

## Genotyping and Imputation

Details of GWAS sample collection, processing and quality control for each study groups have been described elsewhere. The GAIN bipolar sample was genotyped with an Affymetrix Genomewide 1 M chip with 729,087 markers that passed the QC filters, whereas the genotyping of the MDD sample was conducted by using a Perlegen 600K platform with 437,114 SNPs passing the QC criteria. To prevent a selection bias stemming from the restriction of the analysis to the shared markers ( $n = 175,373$ ), we imputed both datasets by using a haplotype-based multipoint imputation method implemented in IMPUTE2 [Howie et al., 2009]. For the imputation process both the GAIN bipolar and the MDD datasets were phased by using SHAPEIT [Delaneau et al., 2013] and imputed with IMPUTE2 by using the 1000 Genomes Phase 1 version 3 haplotypes ( $n = 1,092$ ) [McVean et al., 2012] as reference data. Both datasets were lifted to hg19 positions by using liftOver [Hinrichs et al., 2006] and strand flips were corrected according to

the reference data. Although unaffected controls were excluded from the subsequent analyses, they were still phased and imputed together with the cases to increase phasing and imputation certainty. To further improve imputation quality, conservative post-imputation QC filters were set with certainty  $>0.9$  and info metrics  $>0.9$ ,  $MAF \geq 0.01$  and HWE  $P$ -value  $>0.000001$ . After QC filtering, 6,706,910 shared markers (biallelic SNPs and indels) were included in the downstream data processing.

## Gene Selection

We selected 32 candidate genes that had at least one suggestive SNP-based association with suicidal behavior in published genomewide association or expression studies, and also included three more genes (CACNA1C, ANK3, and ZFN804A) due to their SNP-based association with bipolar disorder [Ferreira et al., 2008; Lett et al., 2011], based on the strong link between bipolar disorders and suicidal behavior. This literature search based candidate gene set (Geneset1,  $n = 35$ ) is shown in details in Table I. To avoid biasing the  $P$  values, the literature searches only covered those studies, that either did not share any samples with our present study, or in case of a partial overlap, the gene selection was not based on reported gene-based associations. In addition, to test the hypothesis of the potential involvement of the microRNA system to executive functions and to suicidal behavior, we compiled two sets of candidate genes that are expressed in the neuroanatomical regions of the dorsolateral and orbitofrontal prefrontal circuits. We accessed the gene expression profiles for the prefrontal cortex, cingulate cortex, caudate nucleus, caudate putamen, corpus striatum, globus pallidus, substantia nigra, thalamus, and the subthalamic nucleus by using the Gene Expression Atlas at the European Bioinformatics Institute (EBI) [Kapushesky et al., 2009]. The narrow gene set (Geneset2,  $n = 68$ ) contained the miRNA genes expressed in the aforementioned brain regions and those genes that have critical importance in the miRNA biosynthesis and the RNA-induced silencing complex (RISC) namely, DICER1, DROSHA, DGCR8, EIF2C2, TARBP, TRNC6A, and XPO5. The extended gene set also included the predicted target genes expressed in the same brain regions (Geneset3,  $n = 11,259$ ). The miRNA target predictions were derived from the miRanda and mirSVR high score target site databases and accessed from the microRNA.org website (MicroRNA.org—Targets and Expression n.d.).

## Variant and Gene-Based Association Analysis

Gene-based associations with the suicidal phenotype were assessed for the 35 literature search based candidate genes of suicide behavior Geneset 1. To minimize the bias due to between-sample population stratification and the confounding effect of the different sampling phenotype, associations of 17,010 markers located  $\pm 5$  kb up- and downstream of the gene boundaries were analyzed separately in both study samples and were pooled with a binary effect meta-analysis method implemented in the METASOFT software [Han and Eskin, 2012]. Effect of within-sample population strata were assessed and corrected for by computing the genomic inflation factor ( $\lambda$ ) for each datasets with the GenABEL package for R [Aulchenko et al., 2007; R Core Team, 2014],

TABLE I. Candidate Genes Previously Implicated in Suicidal Behavior

Gene symbol	Study (author, year)	Biological implication
ACP1	Willour et al. [2012]	Coded protein is a tyrosine phosphatase that inhibits the neurogenesis by decreasing the cytoplasmic $\beta$ -catenin levels
ANK3	Perroud et al. [2012]	Integral membrane protein to cytoskeleton, plays role in neurodevelopment, axon guidance, signal transduction and action potential
BDNF	Dwivedi et al. [2003]; Kim et al. [2008]	Neurotrophic factor, essential to the survival of cortical and striatal neurons, implicated in stress response and mood disorders
CACNA1C	Ferreira et al. [2008]; Lett et al. [2011]	Voltage dependent calcium channel, implicated in axon guidance, synaptic transmission and calcium mediated signaling
CD44	Galfalvy et al. [2011]	Codes a cell surface glycoprotein that involved in cell adhesion and lymphocyte activation and immunoregulation
CYTH3 (GRP1)	Perroud et al. [2012]	Implicated in the control of Golgi structure and function, as well as in protein folding and membrane trafficking
DICER1	Smalheiser et al. [2012]	Rate limiting enzyme of pre-miRNA processing, implicated in neurodevelopment, working memory and executive functions
DSC2	Galfalvy et al. [2011]	Integral membrane component, belongs to the cadherine superfamily, plays role in cell adhesion and desmosome forming
FKBP5	Roy et al. [2012]; Schosser et al. [2011]; Menke et al. [2012]	Plays important role in immunoregulation and protein folding and trafficking and implicated in stress response
FOXN3	Galfalvy et al. [2011]	Acts as suppressor at mitotic checkpoints, important for repairing DNA damage
GDA	Perroud et al. [2012]; Menke et al. [2012]	Guanine deaminase gene, implicated in nervous system development
GFRA1	Schosser et al. [2011]	Glial cell line derived neurotrophic factor, plays role in neuronal survival, differentiation and development
GRIA3	Laje et al. [2007, 2009]	AMPA 3 glutamate receptor, implicated in wide range of neuropsychiatric disorders
GRIK2	Laje et al. [2007, 2009]	Kainate 2 glutamate receptor, implicated in wide range of neuropsychiatric disorders
hsa-miR-185	Maussion et al. [2012]	Regulates the expression of tropomyosin receptor kinase B, implicated in neurodevelopment, neurogenesis and synaptic plasticity
hsa-miR-195	Mellios et al. [2008]	Regulates BDNF and HTR2A gene transcription, implicated in neurodevelopment
hsa-miR-30a	Mellios et al. [2008]	Regulates BDNF transcription, implicated in neurodevelopment
hsa-miR-491	Maussion et al. [2012]	Transcription regulator, significantly upregulated in suicidal behavior
IFNLR1	Laje et al. [2009]	Interleukine 10B receptor that also interacts with IL-28A, 28B, and 29, involved in immunoregulation
KIAA1244	Schosser et al. [2011]	component of the transGolgi network and cell membrane, implicated in vesicle transport
MARCH1	Galfalvy et al. [2011]	A membrane ubiquitin E3 ligase, plays role in immunoresponse and MHC protein binding
MBNL2	Galfalvy et al. [2011]	RNA splicing regulator, implicated in brain development and myotonic dystrophy
NTRK2	Perroud et al. [2012]; Schosser et al. [2011]; Kohli et al. [2010]	Neurotrophic tyrosine kinase, high affinity receptor of BDNF associated with mood disorders and suicidal behavior
NXPH1	Perroud et al. [2012]	Forms tight complex with neuroligins, implicated in dendrite-axon adhesion and synapsis forming
PAPLN	Laje et al. [2009]	A metalloproteinase, associated with SSRI treatment emerged suicidal ideation
PRKCE	Schosser et al. [2011]; Dwivedi et al. [2003]	A signal transduction protein, activated by DAG and Ca, plays important role in neuron channel activation and stress response
RGS18	Schosser et al. [2011]	GTPase activator, plays role in G-protein receptor regulation,
RGS2	Schosser et al. [2011]	GTPase activator, plays role in G-protein receptor regulation, associated with anxiety phenotypes
SAT1	Fiori et al. [2010]	Gene codes a rate limiting enzyme in polyamine catabolism
SLC4A4	Schosser et al. [2011]	A sodium bicarbonate cotransporter, plays role in intracellular pH regulation
SPTLC1	Galfalvy et al. [2011]	Encodes the key enzyme sphingolipid biosynthesis
SRSF11	Galfalvy et al. [2011]	A splicing factor, plays role in mRNA processing
TMEM132C	Willour et al. [2012]	A transmembrane protein, associated with suicide attempts
TUBGCP3	Galfalvy et al. [2011]	Associated with gamma-tubulins in cells and oocytes
ZNF804A	Lett et al. [2011]	A zinc finger protein, transcription regulator, implicated in psychosis, schizophrenia, bipolar disorder and suicide attempts

including both the genotyped and imputed markers. We also validated the population substructure of both study samples by modeling the  $K$  number of subpopulations with the lowest cross-validation error value with the ADMIXTURE software [Alexander et al., 2009] using a subset of 100,000 genotyped SNPs in low linkage disequilibrium (LD:  $r^2 < 0.2$ ), results are shown in Supplementary Figure S3. Both method suggested that the effect of population stratification is insignificant both in the GAIN MDD cases (genotyped  $\lambda = 1.00$ , imputed  $\lambda = 1.013$ ,  $K = 1$ ) and the GAIN bipolar dataset (genotyped  $\lambda = 1.006$ , imputed  $\lambda = 1.001$ ,  $K = 1$ ; QQ-plots for SNP  $P$  values are shown in Supplementary Fig. S1A–C). The slight increase in  $\lambda$  after the imputation was likely to be attributable to the strict post-imputation QC criteria that filtered out those markers that usually contribute to a genomic deflation because of their low variance

Gene-based  $P$  values were calculated by combining the SNP  $P$  values with the hybrid set based test (HYST) with the KGG software [Li et al., 2010, 2012]. This test combines the advantages of the best-SNP picking and SNP-aggregating methods by performing an extended Simes' test (GATES) among the dependent SNPs with high LD and aggregates the estimated test statistics of the LD blocks with a scaled  $\chi^2$  test [Li et al., 2012]. Although this method fits well the genetic heterogeneity model, it is also very sensitive to LD misspecification or incompleteness when aggregating the LD-block statistics. To minimize the bias due to LD misspecification, we used the European subset of the 1000 Genomes Phase 1 version 3 data ( $n = 379$ ) as reference genome for LD calculations in the sample-specific and combined analysis as well (Supplementary Fig. S2A–C display QQ-plots for gene-based  $P$  values).

Although  $P$  values are easily inferred from the region-wise aggregated  $\chi^2$  sum statistics, estimating a valid effect size for a region-based test is not as straightforward. To provide a valid measure that reflect the strength of association of a genomic region, we assessed the heritability explained by the candidate genes on a liability scale by using the restricted maximum likelihood analysis (REML) [Yang et al., 2011b] implemented in the Genome-wide Complex Trait Analysis (GCTA) statistical software [Yang et al., 2011a]. As this approach requires genotype data for the calculation, gene-specific heritabilities for the combined analyses were not reported. To avoid inflated variances due to a population substructure, we controlled the REML models for the top 20 principal components that were inferred from the whole genome data in both study samples with GCTA.

To account for multiple comparisons,  $P$  values were corrected with the Bonferroni method for the three independent analysis families: sample-specific gene-based tests ( $n = 70$ ), combined gene-based tests ( $n = 35$ ) and for the whole experiment ( $n = 105$ ). QQ-plots were generated with KGG and LD-based SNP plot was created by using `snp.plotter` package in R [Luna and Nicodemus, 2007].

## Statistical Power Calculations for the Gene-Based Association Test

Statistical power of the gene-based analyses was estimated with the SPS tool implemented in KGG. SPS models [Li et al., 2015] the statistical power of GATES, scaled sum of  $\chi^2$  and HYST tests

through simulation under the variable input parameters: size of set, number, OR, and MAF of the disease susceptibility locus (DSL), LD, sample size, genetic model, disease prevalence, types of analysis (single study or meta-analysis). A further advantage of SPS is that it can infer the MAF and LD from a subset of the actual genotype data, so the simulated posterior power would be as close to the real conditions as possible.

As set-based tests may perform quite differently depending on the characteristics of the tested gene-region, therefore, we calibrated the gene and variant parameters to define two marginal conditions: an optimal (gene with 60 variant and multiple LD-blocks ( $n = 6$ ) with high LD ( $r = 0.9$ ) and several ( $n = 6$ – $12$ ) risk loci with MAF = 0.05), and a suboptimal setting (same gene size with 60 variants, but only a few LD-blocks ( $n = 2$ ) with low LD ( $r = 0.5$ ) and few ( $n = 1$ – $3$ ) risk loci with MAF = 0.01). Sample sizes for the simulations were set according to the study samples (case subject with/without suicidal behavior: GAIN MDD 250/1350; GAIN bipolar: 350/650, and their average for the combined meta-analysis: 300/1000)

We calculated the power to detect a gene-based  $P$ -value  $< 0.05$  for risk loci with OR = 1.2 by assuming disease prevalence = 0.03, additive genetic model, and locus heterogeneity at the variant based meta-analysis. In addition, we also assessed the power for detecting BDNF with  $P < 0.05$  by using the actual genotypes of both study samples for the simulations.

## Polygenic Risk Score Test

To maximize the information of the polygenic risk score analysis, we calculated the risk scores by two different approaches. We performed a classical pruning and thresholding based analysis with PRSice [Euesden et al., 2014]. PRSice runs a linkage disequilibrium (LD) based clumping for multiple stepwise  $P$ -value thresholds ( $P_t$ ) to find the optimal  $P_t$  value with the best fit. We used the default settings for the three gene sets (LD window: 250 kb,  $r^2 < 0.1$ ) by using GAIN MDD dataset as training sample and the bipolar sample for validation.

Polygenic risk scores (PRS) were also calculated and weighted by using LDpred [Vilhjálmsón et al., 2015], a Bayesian statistical package, which calculates the posterior mean effects from a GWAS summary statistics conditional to the linkage disequilibrium of the underlying genetic architecture. The resulting weighted  $P$ -value derived mean effects are adjusted for the LD, and can be used as PRS predictors without pruning, therefore, information from all variants are utilized during this analysis.

Similar to PRSice, we set the parameters according to the recommended settings. We selected the MDD dataset as training sample and the bipolar dataset as validation sample. Variants with mismatching or ambiguous alleles were excluded during the dataset coordination step, leaving  $n = 13,313$ , 1,953, and 973758 shared variants and corresponding LD radius values of 1, 4, and 324 for Geneset1, Geneset2, and Geneset3. We set a grid of fractions of causal markers (variants with nonzero effect;  $P < 0.001$ ,  $< 0.003$ ,  $< 0.005$ ,  $< 0.01$ ,  $< 0.03$ ,  $< 0.05$ ,  $< 0.1$ ,  $< 0.3$ ,  $< 0.5$ , and  $< 1$  with the infinitesimal shrink) to find the best calibrated PRS weights. The PRS score with the best prediction accuracy and calibration metrics (area under the curve, prediction accuracy  $R^2$ , and prediction

slope) in each geneset were then validated by predicting suicidal behavior in the bipolar dataset. Effect size of predictions and the explained variance (Nagelkerke  $R^2$ ) were assessed with logistic regression models of the fmsb package in statistical software R [R Core Team, 2014].

## RESULTS

### Set-Based Power Estimation

The set-based statistical power analysis for the HYST tests yielded a range of power adequate for detecting nominal gene-based association signal under various conditions; results are presented in detail in Supplementary Table SI. The simulated marginal values of the set-based statistical power for a gene with DSL  $OR = 1.2$  ranged

between 0.05 and 0.70 for the GAIN bipolar data, 0.08–0.74 for the GAIN MDD sample and between 0.06 and 0.94 for the variant based meta-analysis. Posterior power of the nominal gene-based association of BDNF with risk loci  $OR = 1.2$  ranged between 0.16 and 0.29 in the bipolar sample (Supplementary Fig. S4A) and 0.22–0.29 in the MDD sample (Supplementary Fig. S4B) depending on the assumed number of risk loci (1–3), respectively.

### Gene-Based Association Tests and Heritability Estimation

Table II shows the results of the gene-based association scan and the estimated phenotypic heritability on the liability scale ( $h^2$ ) explained by the corresponding gene region. Among the MDD

TABLE II. Gene-Based Associations and Heritability Explained on Liability Scale of Previously Implicated Genes With Suicidal Behavior

Gene symbol	Location	#SNPs in gene region	BIP $P$ -value [ $h^2$ ]	MDD $P$ -value [ $h^2$ ]	Combined $P$ -value
ACP1	2p25	43	0.319 [0.002]	0.713 (<0.001)	0.483
ANK3	10q21	2,040	1.00 (<0.001)	0.687 (<0.001)	0.981
BDNF	11p13	107	0.850 (<0.001)	0.907 (<0.001)	0.957
CACNA1C	12p13.3	1,228	0.796 (<0.001)	0.689 (0.003)	0.711
CD44	11p13	269	0.019* [0.008]	0.749 (<0.001)	0.137
CYTH3	7p22.1	199	0.725 [0.001]	0.925 (<0.001)	0.919
DICER1	14q32.13	88	0.035* [0.004]	0.032* [0.006]	0.025*
DSC2	18q12.1	46	0.488 [0.001]	0.604 (<0.001)	0.570
FKBP5	6p21.31	289	0.992 (<0.001)	0.158 [0.002]	0.930
FOXP3	14q31.3	961	0.643 (<0.001)	0.998 (<0.001)	1.00
GDA	9q21.13	395	0.082 [0.009] <sup>a</sup>	0.958 (<0.001)	0.239
GFRA1	10q26.11	532	0.353 [0.001]	0.625 (<0.001)	0.502
GRIA3	Xq25	387	0.883 (<0.001)	0.017* [0.016] <sup>a</sup>	0.015*
GRIK2	6q16.3	1,736	0.061 [0.024] <sup>a</sup>	0.663 (<0.001)	0.164
hsa-miR-185	22q11.21	24	0.814 (<0.001)	0.852 (<0.001)	1.00
hsa-miR-195	17p13.1	14	0.017* [0.003] <sup>a</sup>	0.999 (<0.001)	0.054
hsa-miR-30a	6p13	22	0.534 (<0.001)	0.305 (<0.001)	0.158
hsa-miR-491	9p21.3	17	0.986 (<0.001)	0.336 [0.002]	0.597
IFNLR1	1p36.11	74	0.813 (<0.001)	0.707 (<0.001)	0.965
KIAA1244	6q23.3	446	0.30 [0.001]	0.320 [0.004]	0.348
MARCH1	4q32.2	1,532	0.766 (<0.001)	0.971 (<0.001)	0.965
MBNL2	13q32.1	324	0.158 [0.002]	0.779 (<0.001)	0.576
NTRK2	9q22.1	838	0.963 (<0.001)	0.0002** [0.01] <sup>a</sup>	0.002
NXPH1	7p22	1,032	0.959 (<0.001)	0.012* [0.014] <sup>a</sup>	0.192
PAPLN	14q24.2	129	0.911 (<0.001)	0.763 (<0.001)	0.981
PRKCE	2p21	1,527	0.961 (<0.001)	0.241 [0.018]	0.694
RGS18	1q31.2	44	0.437 [0.001]	0.546 (<0.001)	0.801
RGS2	1q31	14	0.332 [0.002]	0.163 (<0.001)	0.351
SAT1	Xp22.1	29	0.420 (<0.001)	0.196 (<0.001)	0.275
SLC4A4	4q21	540	0.707 (<0.001)	0.930 (<0.001)	0.901
SPTLC1	9q22.2	296	0.503 (<0.001)	0.041* [0.005]	0.106
SRSF11	1p31	67	0.912 (<0.001)	0.484 (<0.001)	0.650
TMEM132C	12q24.32	1,089	0.287 (<0.001)	0.807 [0.001]	0.829
TUBGCP3	13q34	284	0.876 (<0.001)	0.974 (<0.001)	0.968
ZNF804A	2q32.1	1,027	0.897 (<0.001)	0.944 (<0.001)	1.00

e is nominally significant.  $h^2$  is calculated as  $(V(G)/Vp_L)$  with the REML analysis in GCTA.

\*  $P$ -value is nominally significant.

\*\*  $P$ -value is significant after familywise correction.

<sup>a</sup>Heritability estimate significantly differs from 0.

subjects, the NTRK2 gene had the strongest association signal ( $P=0.0002$ ,  $h^2=0.01$ ), followed by NXPH1 ( $P=0.012$ ,  $h^2=0.014$ ), GRIA3 ( $P=0.017$ ,  $h^2=0.016$ ), DICER1 ( $P=0.032$ ,  $h^2=0.006$ ), and SPTLC1 ( $P=0.041$ ,  $h^2=0.006$ ) genes.

In the GAIN bipolar sample nominally significant associations were found for hsa-miR-195 ( $P=0.017$ ,  $h^2=0.003$ ), CD44 ( $P=0.019$ ,  $h^2=0.009$ ), and DICER1 ( $P=0.034$ ,  $h^2=0.004$ ), whereas nominally significant associations were observed in the combined analysis for NTRK2 ( $P=0.002$ ), GRIA3 ( $P=0.011$ ), and DICER1 ( $P=0.046$ ) genes. After correction for multiple comparison, none of these gene-based associations remained statistically significant even at the family wise level, with the notable exception of NTRK2 gene in the MDD sample ( $P_{\text{fam}}=0.014$ ,  $P_{\text{exp}}=0.021$ ).

To ensure the specificity of the associations of NTRK2 and DICER1 with suicidal behavior, we cross-validated the nominally significant gene-based associations with the main phenotypes (MDD and bipolar disorder) in both GAIN samples, including the unaffected control subjects as well. With the exception of NTRK2, none of the nominally significant genes were associated with the mood disorder phenotype in the larger samples, and the association between NTRK2 and MDD was even substantially weaker compared to that of suicidal behavior ( $P=0.04$  vs.  $P=0.0002$ ). This cross-validation indicated that the detected gene-based tests genuinely measured associations with suicidal behavior.

Annotation of the variants of DICER1 and NTRK2 genes are shown in Supplementary Tables S2 and S3. Although none of the variants with at least nominal significance are protein coding,

several of them may have regulatory roles, as they are mapped to the DNase I Hypersensitivity Sites (DHS) or promoter sites catalog published by the ENCODE project. [Bernstein et al., 2012; Gusev et al., 2014] In case of DICER1, two DHS SNPs (rs1209904 and rs1187655) were identified with nominally significant association and same direction of association in both datasets, whereas rs1187646, rs11851369, and rs7157322 had significant associations only in one sample but had comparable—albeit not significant—OR estimates in the other dataset. Significantly associated variants in the NTRK2 gene region were more sample specific, no variant based associations were shared between the study samples, but 35 of the 149 nominally significant variants were identified with potential regulatory roles.

## Polygenic Risk Score Tests

Results from the Bayesian PRS analysis with LDpred are shown in Table III, whereas Figure 1A–C present the predictions of the best PRS scores optimized by PRSice for Geneset1, Geneset2, and Geneset3, respectively.

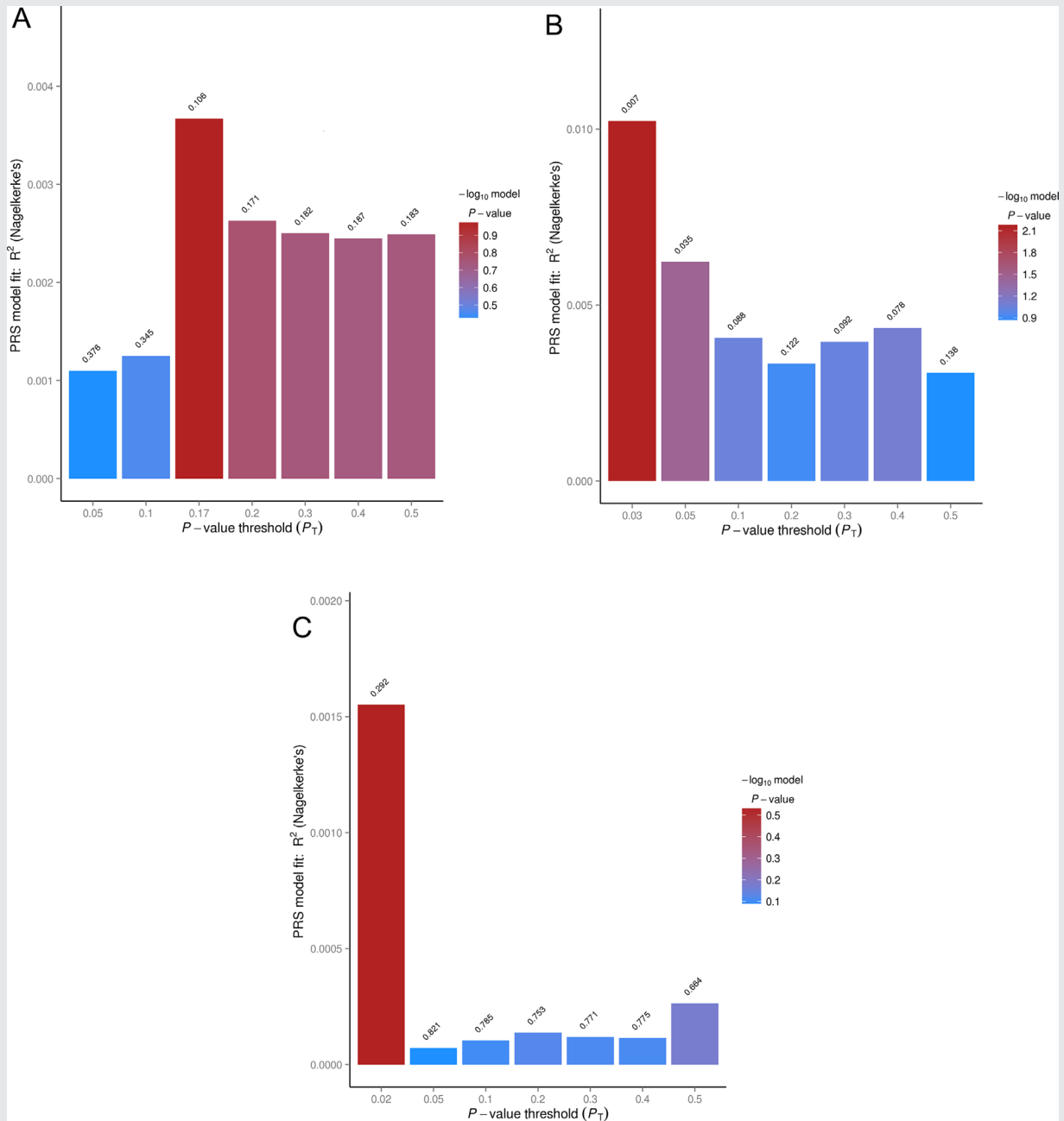
For the literature based candidate genes (Geneset1) the Bayesian accuracy and calibration metrics indicated the infinitesimal model weights (fraction of causal  $P < 1$ ) for the best fit, predicting suicidal behavior in the bipolar sample with Nagelkerke  $R^2 = 0.0028$  and  $P < 0.156$ . PRSice yielded a very similar, not significant prediction of suicidal behavior with Nagelkerke  $R^2 = 0.0036$ , and prediction  $P < 0.106$  by using 232 markers with  $P_i < 0.17$  from Geneset1.

In the miRNA set (Geneset2), posterior weights inferred from the model of causal variants with  $P < 0.01$  provided the best

TABLE III. Targeted Polygenic Risk Score Calibration and Prediction of Suicidal Behavior in the GAIN Bipolar Sample by Using Bayesian Polygenic Scores Calculated in the GAIN MDD Sample for the Three Candidate Gene Sets: Previously Implicated Genes (Geneset1), the Dorsolateral and Orbitofrontal Prefrontal Circuits miRNA Genes (Geneset2), and the Extended DLPCF miRNA Genes + Their Predicted Targets (Geneset3)

	Geneset1				Geneset2				Geneset3			
	Prediction		Predicted		Prediction		Predicted		Prediction		Predicted	
	Accuracy	Calibration	Variance	Effect	Accuracy	Calibration	Variance	Effect	Accuracy	Calibration	Variance	Effect
Fraction of causal variants	$R^2$	Slope	$R^2$	$\beta$ (P-val)	$R^2$	Slope	$R^2$	$\beta$ (P-val)	$R^2$	Slope	$R^2$	$\beta$ (P-val)
$P < 0.001$	0.0001	0.0001			0.0003	0.271			0.0004	0.0036		
$P < 0.003$	0.0001	<0.0001			0.0005	0.317			<b>0.0009</b>	<b>0.5751</b>	0.0012	-1.14 (0.357)
$P < 0.005$	0.0001	0.0006			0.0000	0.004			<0.0001	0.0025		
$P < 0.01$	0.0001	<0.0001			<b>0.0015</b>	<b>0.326</b>	0.0076	-1.52 (0.019)	0.0009	0.0681		
$P < 0.03$	0.0001	0.0004			0.0022	0.146			0.0002	0.0680		
$P < 0.05$	0.0001	<0.0001			0.0048	0.032			0.0009	0.0510		
$P < 0.1$	<0.0001	0.0001			0.0002	0.069			0.0006	0.3431		
$P < 0.3$	0.0006	0.0004			0.0040	0.017			0.0009	0.0484		
$P < 0.5$	<0.0001	0.0010			0.0002	0.547			0.0007	0.0007		
$P < 1.0$	<b>0.0008</b>	<b>0.0006</b>	0.0028	-4.58 (0.156)	0.0039	0.016			0.0009	0.0481		

Accuracy is measured with prediction  $R^2$  and predicted variance by Nagelkerke  $R^2$ . Bolded values refer to the fraction of causal variant  $P$ -value with the best prediction accuracy and calibration in each geneset.



**FIG. 1. A:** PRS score prediction of suicidal behavior by using PRS scores from the literature search based candidate gene set [Geneset1]. **B:** PRS score prediction of suicidal behavior by using PRS scores from the miRNA genes expressed in the dorsolateral prefrontal circuit [Geneset2]. **C:** PRS score prediction of suicidal behavior by using PRS scores from the miRNA target genes expressed in the dorsolateral prefrontal circuit [Geneset3]. [Color Figure can be seen in the online version of this article, available at <http://wileyonlinelibrary.com/journal/ajmgb>].

calibration and accuracy, predicting suicidal behavior in the bipolar sample with  $P = 0.019$ , Nagelkerke  $R^2 = 0.0076$  and  $\beta = -1.52$ . PRSice identified a very similar solution: the PRS score derived from 21 markers with  $P_t < 0.03$  predicted suicidal behavior in the bipolar dataset with Nagelkerke  $R^2 = 0.01$  and  $P < 0.007$ .

Neither methods could find PRS with significant prediction for the miRNA target genes (Geneset3). LDpred model fit metrics

suggested the posterior weights from the model of fraction of causal markers with  $P < 0.003$ , that yielded Nagelkerke  $R^2 = 0.0012$ ,  $\beta = -1.14$  and  $P < 0.357$  by predicting suicidal behavior among the bipolar subjects. Again, PRSice found the optimal threshold of  $P_t < 0.02$  of 2444 PRS markers predicting suicidal behavior in the bipolar sample with an almost identical performance (Nagelkerke  $R^2 = 0.0016$  and  $P < 0.262$ ).

## DISCUSSION

We carried out a comparative, cross-disorder candidate gene study by using region-based and polygenic risk score analyses to test the associations with suicidal behavior among individuals with bipolar or major depressive disorders. In the gene-based association scan seven genes had nominally significant association with suicidal behavior in either dataset, but only *NTRK2* remained statistically significant after the familywise correction in the MDD sample with a decent 1% of the explained phenotypic variance. Most notably, however, among the 35 genes only *DICER1* gene associated with suicidal behavior in both study samples with at least nominal significance. Polygenic risk scores of the miRNA genes expressed in the anatomic regions of the dorsolateral and orbitofrontal prefrontal circuits predicted suicidal behavior both in the LD-score weighting and pruning and thresholding models, although with a modest explained variance (best Nagelkerke  $R^2 = 0.01$ ) and negative predictions. Regardless of the analytic approach, no significant prediction was detected for the previously implicated candidate gene set or the extended miRNA set containing the miRNA target genes.

We found statistically significant gene-based association between *NTRK2* gene and suicidal behavior that remained significant after the familywise Bonferroni correction for multiple comparisons in the MDD sample, and even survived the more strict experimentwise Bonferroni correction in the MDD sample. The *NTRK2* gene encodes a high-affinity tyrosine kinase receptor (TrkB) of the brain derived neurotrophic factor (BDNF). Neurotrophins—particularly BDNF—are essential in neuronal growth, differentiation, maturation and survival and play an important modulatory role in synaptic plasticity [Gupta et al., 2013]. Notably, the involvement of neurotrophins, especially the CREB1-BDNF-NTRK2 pathway in stress response, pathophysiology, and treatment response of mood disorders and suicidal behavior has been implicated by several independent associations [Kim et al., 2008; Kohli et al., 2010; Juhasz et al., 2011; Wang et al., 2013] and expression studies [Dwivedi et al., 2003; Easton et al., 2006; Dwivedi et al., 2009; Cattaneo et al., 2013]. In a recent publication, Kohli et al. [2010] reported a significant gene-based association and a strong combined association between four tagging SNPs of the *NTRK2* gene and suicide attempt among subjects with major depression. The strong gene-based association we detected between the *NTRK2* gene and suicidal behavior in the MDD sample corroborates these results. Although we did not replicate the same SNP associations, the peak association signals in the MDD sample were located in the proximity of the 3' end of exon 16 and in a lesser magnitude in the bipolar sample close to exons 20–22 (Fig. 2). Alternative splicing in this region results in the truncated TrkB-Shc and TrkB-T1 isoforms lacking the tyrosine kinase domain and phospholipase C binding site and in case of the TrkB-T1, the Shc binding site as well [Easton et al., 2006; Dwivedi, 2010]. The truncated TrkB isoforms are assumed to play an important regulatory role in the neurotrophic signaling [Wong and Garner, 2012] and alteration in their mRNA levels in the prefrontal cortex has been found in schizophrenia [Ray et al., 2014], mood disorders [Ray et al., 2014; Qi et al., 2015] and completed suicide [Ernst et al., 2009; Maussion et al., 2012]. In addition, several slightly weaker

hits were found in the MDD sample at the 5' part of the gene, which is known to harbor multiple alternative splicing sites and alternative promoters, further suggesting the potential involvement of the TrkB isoforms in suicidal behavior. Furthermore, by annotating variants with the regulatory sites published by the ENCODE project [Bernstein et al., 2012] and mapped to the 1000 Genomes data by Gusev et al. [2014], we determined that 35 of the 149 nominally significant association of the *NTRK2* gene mapped to these regulatory sites, suggesting potential functionality. It is important to note, however, that none of these markers would remain significant after a conservative correction for multiple comparisons, therefore, extra caution is needed during the interpretations.

Unlike in the MDD sample, we were not able to find even nominally significant gene-based association between the *NTRK2* gene and suicidal behavior in the bipolar sample, however, nominally significant SNP associations were detected close to the regulatory, alternative splicing sites of the aforementioned truncated isoforms. The reason of this discrepancy is intriguing, possible explanations include methodological differences and phenotypic heterogeneity (severe suicidal ideation and/or attempts in the MDD sample vs. suicide attempts in the bipolar sample), the potentially different prevalence of childhood adversities that seems to have a gene-environment interactive effect with *NTRK2* on depression [Juhasz et al., 2011], and the smaller sample size and consequently smaller statistical power of the bipolar sample. Similarly, the lack of association of most of the candidate genes, for instance BDNF, in either sample is interesting, but not surprising. The *HYST* gene-based test used in this analysis is optimized to highlight regions with many nominally significant, albeit moderate independent association signals, balancing the effect of local locus heterogeneity. On the other hand, this method could miss large genomic regions with only one or two disease susceptibility loci associations and many neutral markers. With model parameters inferred from the BDNF region in both datasets, the set-based posterior power analysis indicated low statistical power (0.16–0.29) in case of 1–3 risk loci with small effect sizes ( $OR = 1.2$ ). Therefore, not significant gene-based *P* values by no means invalidate previously identified strong SNP associations, especially if they are potentially functional.

Among the candidate genes only the *DICER1* gene associated with suicidal behavior in both samples, albeit the *P*-value strength did not exceed the nominal level. The phenotypic variance explained by *DICER1* was estimated as 0.6%; however, the confidence intervals of this estimate did include 0 as well, due to the large standard error. The product of this gene, the dicer protein is a type III ribonuclease, a rate limiting enzyme of classical miRNA biosynthesis. MicroRNAs modulate or repress gene expression at the post-transcriptional level by binding to the target mRNA and contributing to its degradation through the RISC complex or by directly inhibiting the translation. As potent gene expression regulators, miRNAs play critical role in developmental processes, including neurodevelopment and synaptic plasticity and along with dicer, they have been implicated in the pathomechanisms of various neuropsychiatric disorders [Miller and Wahlestedt, 2010; Xu et al., 2012; Serafini et al., 2014; Smalheiser et al., 2014] and prefrontal cortical functions [Stark et al., 2008; Konopka et al., 2010; Beveridge et al., 2013]. In concordance with the





crease, whereas transcription inhibition did not affect BDNF response, suggesting the dominant role of post-transcriptional regulation. In the light of the post-transcriptional regulatory nature of miRNAs, it is very intriguing that the strongest SNP association signals were identified in both of our study samples in the same haplotype block that covers all predicted miRNA target sites in the 3' UTR region of the DICER1 gene. One could speculate, that these SNPs might tag untyped rare variations at the miRNA docking sites, possibly altering the translational rate and availability of dicer enzyme. In addition, several nominally significant markers were identified as the DNase I Hypersensitivity Sites or promoter regulatory element based on the ENCODE annotations. However, the functional significance of these variants is still not clear and given their only nominal significance level, the likelihood of finding such associations by chance is quite high, therefore further replications are necessary to validate these findings.

In addition to the potential synthetic associations in the DICER1 gene, the involvement of the miRNA system in suicidal behavior might also be mirrored by the polygenic profile analysis. As shown in Table III and Figure 1A–C, polygenic profile scores inferred from the MDD sample for the selected miRNA set predicted suicidal behavior significantly—albeit with negative betas—in the bipolar sample. The miRNA gene set was selected based on the expression profiles assessed in the relevant neuroanatomic structures of executive functioning, that is, the prefrontal cortex and the dorsolateral (and partially the orbitofrontal) prefrontal circuits. Executive function deficits were described in both suicidal ideation and attempt and were shown to be independent from the underlying psychiatric disorder. Moreover, evidence from recent animal studies suggests that insufficiency of the main miRNA processing enzymes (DGCR8, DROSHA, and DICER1) alters cognitive and executive functions [Stark et al., 2008; Konopka et al., 2010]. Therefore, it is possible that the significant polygenic predictions and gene-based associations of miRNA genes identified in our analysis might in fact be associated with executive function impairment. Nevertheless, the modest association strength and explained variance and the negative predictions suggest that suicidal ideation and attempt are complex phenotypes; therefore, the underlying neuropsychological deficits and genetic architecture might be quite different. The lack of prediction by the literature search based candidate genes and the extended miRNA and target gene set most likely stemmed from this phenotypic heterogeneity and relatively modest sample size.

Limitations of the study are noted. Perhaps the most important weakness of this analysis was the relatively low sample size, mainly that of the bipolar sample. The set-based posterior power analysis confirmed that our study did not have adequate power to detect genes with many neutral variants and only a few risk loci with small effect sizes and in high LD, for example, BDNF. However, the analysis also suggested that for genes with more, relatively independent risk loci the power was acceptable even by considering small effect sizes for the risk loci.

Similar to many other studies in the field, data for both samples were collected retrospectively by assessing patient samples. This sampling design inevitably introduces an ascertainment bias, because only the survivors would be included this way and perhaps the most interesting population, the suicide completers were left

out. Unfortunately, there is no way to correct this bias during the data analysis, therefore, special emphasis should be given to the findings from studies with subjects, who died in suicide.

Another but not less important issue was the different suicidal phenotype definition and phenotyping instrument. Whereas the DIGS interview assessed actual suicide attempts in the bipolar sample, the IDS questionnaire used in the MDD dataset measured degrees of suicidal ideation with suicide attempts included among the criteria for the most severe form. Due to this phenotypic heterogeneity especially the interpretations of missing replications (e.g., the NTRK2 gene association) is quite problematic. Clearly, further analyses with larger sample sizes and more homogeneous suicidal behavior phenotype are required to confirm this disorder specific association.

In summary, by using region-based tests and polygenic prediction analysis for targeted gene sets our study provided additional evidence for the involvement of the neurotrophins in the pathogenesis of suicidal behavior and suggestive findings for the potential mediatory role of the miRNA system. The statistically significant, but inverse cross-disorder polygenic predictions of the neurobiologically informed miRNA gene set indicate that the executive function deficit might be involved both in suicidal ideation and attempt, but the relationship of these phenotypes is complex and probably highly affected by genetic heterogeneity and pleiotropy. Although interpretations of our findings are clearly constricted by the aforementioned limitations, the biological plausibility of our results is intriguing and warrants further investigation and replications.

## ACKNOWLEDGMENTS

The study was supported by the Hungarian National Research Fund and the Hungarian Brain Research Program (grants OTKA-PD 83876 and NAP-B KTIA\_NAP\_13-2014-0011 to JR).

## REFERENCES

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
- Antypa N, Serretti A, Rujescu D. 2013. Serotonergic genes and suicide: A systematic review. *Eur Neuropsychopharmacol* 23(10):1125–1142. Available at <http://www.ncbi.nlm.nih.gov/pubmed/23742855> [Accessed March 28, 2014].
- Aulchenko YS, et al. 2007. GenABEL: An R library for genome-wide association analysis. *Bioinformatics (Oxford, England)* 23(10):1294–1296.
- Ayalew, M. et al. 2012. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction (February). Available at: <http://dx.doi.org/10.1038/mp.2012.37>
- Bernstein BE, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3439153&tool=pmcentrez&rendertype=abstract> [Accessed July 9, 2014].
- Betel D, et al. 2008. The microRNA.org resource: targets and expression. *Nucleic acid Res* 36(Database issue):D149–D153.
- Beveridge NJ, et al. 2013. Maturation of the human dorsolateral prefrontal cortex coincides with a dynamic shift in microRNA expression.

- Schizophr Bull 40(2):1–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23378013> [Accessed March 28, 2014].
- Boomsma DI, et al. 2008. Genome-wide association of major depression: description of samples for the GAIN major depressive disorder study: NTR and NESDA biobank projects. *Eur J Hum Genet* 16(3):335–342.
- Boomsma DI, et al. 2000. Netherlands twin family study of anxious depression (NETSAD). *Twin Res* 3(4):323–334.
- Cattaneo A, et al. 2013. Candidate genes expression profile associated with antidepressants response in the GENDEP study: differentiating between baseline “predictors” and longitudinal “targets”. *Neuropsychopharmacology* 38(3):377–385. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3547188&tool=pmcentrez&rendertype=abstract> [Accessed March 24, 2014].
- Delaneau O, et al. 2013. Haplotype estimation using sequencing reads. *Am J Hum Genet* 93(4):687–696.
- Dick DM, et al. 2003. Genomewide linkage analyses of bipolar disorder: A new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet* 73(1):107–114.
- Dwivedi Y, et al. 2003. Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Arch Gen Psychiatry* 60:804–815. Available at: <http://archpsyc.jamanetwork.com/article.aspx?articleid=207739> [Accessed March 29, 2014].
- Dwivedi Y. 2010. Brain-derived neurotrophic factor and suicide pathogenesis. *Ann Med* 42(2):87–96. Available at: <http://informahealthcare.com/doi/abs/10.3109/07853890903485730> [Accessed March 29, 2014].
- Dwivedi Y, et al. 2009. Neurotrophin receptor activation and expression in human postmortem brain: Effect of suicide. *Biol Psychiatry* 65(4):319–328.
- Easton JB, Royer AR, Middlemas DS. 2006. The protein tyrosine phosphatase, Shp2, is required for the complete activation of the RAS/MAPK pathway by brain-derived neurotrophic factor. *J Neurochem* 97(3):834–845.
- Ernst C, et al. 2009. Alternative splicing, methylation state, and expression profile of tropomyosin-related kinase B in the frontal cortex of suicide completers. *Arch Gen Psychiatry* 66(1):22–32.
- Euesden J, Lewis CM, O'Reilly PF. 2014. PRSice: Polygenic risk score software. *Bioinformatics* 31(9):1466–1468. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4410663&tool=pmcentrez&rendertype=abstract> [Accessed July 23, 2015].
- Ferreira MA, et al. 2008. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 40(9):1056–1058. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2703780&tool=pmcentrez&rendertype=abstract> [Accessed March 28, 2014].
- Fiori LM, et al. 2010. Association of polyaminergic loci with anxiety, mood disorders, and attempted suicide. *PLoS ONE* 5(11):e15146. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2994870&tool=pmcentrez&rendertype=abstract> [Accessed March 28, 2014].
- Galfalvy H, et al. 2015. A genome-wide association study of suicidal behavior. *Am J Med Genet B* 168(7):557–563.
- Galfalvy H, et al. 2011. A pilot genome wide association and gene expression array study of suicide with and without major depression. *World J Biol Psychiatry* 14(8):1–9.
- Gupta VK, et al. 2013. TrkB receptor signalling: Implications in neurodegenerative, psychiatric and proliferative disorders. *Int J Mol Sci* 14(5):10122–10142.
- Gusev A, et al. 2014. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* 95(5):535–552. Available at: <http://www.cell.com/article/S0002929714004261/fulltext> [Accessed November 6, 2014].
- Han B, Eskin E. 2012. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet* 8(3):e1002555. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3291559&tool=pmcentrez&rendertype=abstract> [Accessed March 20, 2014].
- Hinrichs AS, et al. 2006. The UCSC genome browser database: Update 2006. *Nucleic Acids Res* 34(Database issue):D590.
- Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5(6):e1000529.
- Huang YW, et al. 2012. Dual regulation of miRNA biogenesis generates target specificity in neurotrophin-induced protein synthesis. *Cell* 148(5):933–946. Available at: <http://dx.doi.org/10.1016/j.cell.2012.01.036>.
- Juhász G, et al. 2011. The CREB1-BDNF-NTRK2 pathway in depression: Multiple gene-cognition-environment interactions. *Biol Psychiatry* 69(8):762–771. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21215389> [Accessed March 28, 2014].
- Kapushesky M, et al. 2009. Gene expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res* 38(SUPPL.1):690–698.
- Keilp JG, et al. 2012. Neuropsychological function and suicidal behavior: Attention control, memory and executive dysfunction in suicide attempt. *Psychol Med* 42(1):1–13.
- Kim B, et al. 2008. Brain-derived neurotrophic factor Val/Met polymorphism and bipolar disorder. Association of the Met allele with suicidal behavior of bipolar patients. *Neuropsychobiology* 58(2):97–103. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18843195> [Accessed March 28, 2014].
- Kohli MA, et al. 2010. Association of genetic variants in the neurotrophic receptor-encoding gene NTRK2 and a lifetime history of suicide attempts in depressed patients. *Arch Gen Psychiatry* 67(4):348–359.
- Konopka W, et al. 2010. MicroRNA loss enhances learning and memory in mice. *J Neurosci* 30(44):14835–14842.
- Laje G, et al. 2007. Genetic markers of suicidal ideation emerging during citalopram treatment of major depression. *Am J Psychiat* 164(10):1530–1538.
- Laje G, Allen A, Akula N. 2009. Genome-wide association study of suicidal ideation emerging during citalopram treatment of depressed outpatients. *Pharmacogenet Genom* 19(9):666–674.
- LeGris J, van Reekum R. 2006. The neuropsychological correlates of borderline personality disorder and suicidal behaviour. *Can J Psychiatry* 51:131–142.
- Lett TAP, et al. 2011. ANK3, CACNA1C and ZNF804A gene variants in bipolar disorders and psychosis subphenotype. *World J Biol Psychiatry* 12(5):392–397.
- Li J, et al. 2015. SPS: A Simulation Tool for Calculating Power of Set-Based Genetic Association Tests. *Genet Epidemiol* 39(5):395–397. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25995121> [Accessed November 21, 2015].
- Li MX, et al. 2010. A knowledge-based weighting framework to boost the power of genome-wide association studies. *PloS one* 5(12):e14480.
- Li MX, Kwan JSH, Sham PC. 2012. HYST: A hybrid set-based test for genome-wide association studies, with application to protein–protein interaction-based association analysis. *Am J Hum Gen* 91(3):478–488.
- Luna A, Nicodemus KK. 2007. Snp. plotter: An R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics*

- (Oxford, England) 23(6):774–776. Available at: <http://bioinformatics.oxfordjournals.org/content/23/6/774.short> [Accessed July 14, 2015].
- Mailman MD, et al. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 39(10):1181–1186.
- Malloy-Diniz LF, et al. 2009. Suicide behavior and neuropsychological assessment of type I bipolar patients. *J Affect Disord* 112:231–236.
- Marzuk PM, et al. 2005. Executive functioning in depressed patients with suicidal ideation. *Acta Psychiatr Scand* 112(4):294–301. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16156837> [Accessed March 28, 2014].
- MauSSION G, et al. 2012. Regulation of a truncated form of tropomyosin-related kinase b (TrkB) by Hsa-miR-185\* in frontal cortex of suicide completers. *PloS one* 7(6): e39301.
- McInnis MG, et al. 2003. Genome-wide scan and conditional analysis in bipolar disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. *Biol Psychiatry* 54(11):1265–1273.
- McVean GA, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.
- Mellios N, et al. 2008. A set of differentially expressed miRNAs, including miR-30a-5p, act as post-transcriptional inhibitors of BDNF in prefrontal cortex. *Hum Mol Gen* 17(19):3030–3042.
- Menke A, et al. 2012. Genome-wide association study of antidepressant treatment-emergent suicidal ideation. *Neuropsychopharmacology* 37(3):797–807.
- Miller BH, Wahlestedt C. 2010. MicroRNA dysregulation in psychiatric disease. *Brain Res* 1338:89–99. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2891055&tool=pmcentrez&rendertype=abstract> [Accessed March 28, 2014].
- Miranda R, Gallagher M. 2012. Cognitive inflexibility as a prospective predictor of suicidal ideation among young adults with a suicide attempt history. *Depress Anxiety* 29:180–186.
- Mullins N, et al. 2014. Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: A genome-wide association and polygenic scoring study. *Am J Med Genet Part B* 165(5):428–437.
- Nock MK, et al. 2008. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry* 192(2):98–105.
- Organization, W.H., 2014. *Global Health Estimates: Death by cause, age, sex and country, 2000–2012*. WHO. Geneva.
- Perlis R, Huang J, Purcell S. 2010. Genome-wide association study of suicide attempts in mood disorder patients. *Am J Psychiatr* 167(12): 1499–1507.
- Perroud N, et al. 2012. Genome-wide association study of increasing suicidal ideation during antidepressant treatment in the GENDEP project. *Pharmacogenomics J* 12(1):68–77.
- Petersen L, et al. 2014. Genetic and familial environmental effects on suicide attempts: A study of Danish adoptees and their biological and adoptive siblings. *J Affect Disord* 155(1):273–277. Available at: <http://dx.doi.org/10.1016/j.jad.2013.11.012>
- Purcell SM, et al. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460(7256):748–752.
- Qi X-R, et al. 2015. Abnormal retinoid and TrkB signaling in the prefrontal cortex in mood disorders. *Cereb Cortex* 25(1):75–83.
- R Core Team. 2014. *R: A language and environment for statistical computing*.
- Ray MT, Shannon Weickert C, Webster MJ. 2014. Decreased BDNF and TrkB mRNA expression in multiple cortical areas of patients with schizophrenia and mood disorders. *Transl Psychiatry* 4(5):e389. Available at: <http://dx.doi.org/10.1038/tp.2014.26>
- Richard-Devantoy S, et al. 2012. Suicidal behaviours in affective disorders: a deficit of cognitive inhibition? *Can J Psychiatry* 57:254–262. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22480591>
- Roy A, et al. 2012. Two HPA axis genes, CRHBP and FKBP5, interact with childhood trauma to increase the risk for suicidal behavior. *J Psychiatric Res* 46(1):72–79.
- Schild AHE, et al. 2013. Genetic association studies between SNPs and suicidal behavior: A meta-analytical field synopsis. *Prog Neuropsychopharmacol Biol Psychiatry* 46:36–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23831262> [Accessed March 28, 2014].
- Schossor A, et al. 2011. Genomewide association scan of suicidal thoughts and behaviour in major depression. *PLoS ONE* 6(7):e20690. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3130038&tool=pmcentrez&rendertype=abstract> [Accessed March 26, 2014].
- Serafini G, et al. 2014. The involvement of MicroRNAs in major depression, suicidal behavior, and related disorders: A focus on miR-185 and miR-491-3p. *Cell Mol Neurobiol* 34(1):17–30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24213247> [Accessed March 28, 2014].
- Smalheiser NR, et al. 2014. Expression of microRNAs and Other Small RNAs in Prefrontal Cortex in Schizophrenia, Bipolar Disorder and Depressed Subjects. *PLoS ONE* 9(1):e86469. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3903529&tool=pmcentrez&rendertype=abstract> [Accessed March 28, 2014].
- Smalheiser NR, et al. 2012. MicroRNA expression is down-regulated and reorganized in prefrontal cortex of depressed suicide subjects. *PLoS ONE* 7(3):e33201. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3302855&tool=pmcentrez&rendertype=abstract> [Accessed March 28, 2014].
- Stark KL, et al. 2008. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat Genet* 40(6):751–760. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18469815> [Accessed March 21, 2014].
- Tidemalm D, et al. 2011. Familial clustering of suicide risk: A total population study of 11. 4 million individuals. *Psychol Med* 41(12): 2527–2534.
- Tryka KA, et al. 2014. NCBI's database of genotypes and phenotypes: DbGaP. *Nucleic Acids Res* 42(D1):1–5.
- Tsai SJ, Hong CJ, Liou YJ. 2011. Recent molecular genetic studies and methodological issues in suicide research. *Progress Neuropsychopharmacol Biol Psychiatry* 35(4):809–817.
- Van Heeringen K. 2003. The neurobiology of suicide and suicidality. *Can J Psychiatry* 48(5):292–300. Available at: [http://dx.doi.org/10.1016/S2215-0366\(14\)70220-2](http://dx.doi.org/10.1016/S2215-0366(14)70220-2).
- Vilhjálmsdóttir BJ, et al. 2015. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* 97(4):576–592. Available at: <http://www.cell.com/article/S0002929715003651/fulltext> [Accessed October 2, 2015].
- Voracek M, Loibl LM. 2007. Genetics of suicide: A systematic review of twin studies. *Wien Klin Wochenschr* 119(15–16):463–475.
- Wang Z, et al. 2013. Neurotrophic tyrosine kinase receptor type 2 (NTRK2) gene associated with treatment response to mood stabilizers in patients with bipolar I disorder. *J Mol Neurosci* 50(2):305–310. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3658129&tool=pmcentrez&rendertype=abstract> [Accessed March 28, 2014].
- Westheide J, et al. 2008. Executive performance of depressed suicide attempters: The role of suicidal ideation. *Eur Arch Psychiatry Clin*

- Neurosci 258(7):414–421. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18330667> [Accessed March 28, 2014].
- Willour V, et al. 2012. A genome-wide association study of attempted suicide. *Mol Psychiatry* 17(4):433–477. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21423239> [Accessed March 26, 2014].
- Wong J, Garner B. 2012. Evidence that truncated TrkB isoform, TrkB-Shc can regulate phosphorylated TrkB protein levels. *Biochem Biophys Res Commu* 420(2):331–335.
- Xu B, et al. 2012. MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction. *Neurobiol Disease* 46(2):291–301. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3329786&tool=pmcentrez&rendertype=abstract> [Accessed March 21, 2014].
- Yang J, Lee SH, et al. 2011a. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 88(1):76–82. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3014363&tool=pmcentrez&rendertype=abstract> [Accessed July 10, 2014].
- Yang J, Manolio TA, et al. 2011b. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43(6):519–525. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4295936&tool=pmcentrez&rendertype=abstract> [Accessed November 25, 2015].

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.



# A SZKIZOFRÉRIA MULTILÓKUSZ GENETIKAI VIZSGÁLATA AZ IDEGFEJLŐDÉS ÉS AZ IMMUNRENDSZER ZAVARÁNAK OKI SZEREPÉRE UTAL(HAT)

PULAY Attila József<sup>1</sup>, KOLLER Júlia<sup>2</sup>, NAGY László<sup>3</sup>, MOLNÁR Mária Judit<sup>1</sup>, RÉTHELYI János<sup>1,4</sup>,  
a Magyar SCHIZOBANK Konzorcium munkatársai

<sup>1</sup>Semmelweis Egyetem, Pszichiátriai és Pszichoterápiás Klinika, Budapest

<sup>2</sup>Semmelweis Egyetem, Genomikai Medicina és Ritka Betegségek Intézete, Budapest

<sup>3</sup>Debreceni Egyetem, Biokémiai és Molekuláris Biológiai Intézet, Debrecen

<sup>4</sup>MTA-SE NAP-B Molekuláris Pszichiátriai és in vitro Betegségmodellezési Kutatócsoport, Budapest



Hungarian

<http://dx.doi.org/10.18071/isz.70.0115>

[www.elitmed.hu](http://www.elitmed.hu)

## MULTILOCUS GENETIC ANALYSIS IMPLICATES NEURODEVELOPMENT AND IMMUNE SYSTEM IN THE ETIOLOGY OF SCHIZOPHRENIA

Pulay AJ, MD; Koller J, MD; Nagy L, MD, PhD, DSc; Molnár MJ, MD, PhD, DSc; Réthelyi JM, MD, PhD, colleagues of the Hungarian SCHIZOBANK Consortium  
*Ideggyogy Sz* 2017;70(3–4):115–126.

**Háttér** – A szkizofrénia multifaktoriális, csak részlegesen ismert etiológiájú pszichiátriai zavar, melyet magas heritabilitás (örökletesség) és genetikai heterogenitás jellemez. A multilókuszos genetikai vizsgálatok csökkenthetik a heterogenitás hatását, javítva a vizsgálatok közötti replikáció esélyét.

**Célkitűzés** – Vizsgálatunkat kettős céllal végeztük: 1. Megvizsgáltuk a rendelkezésünkre álló mintákban a szkizofrénia genetikai háttérét multilókuszos-asszociációs tesztekkel, valamint 2. meghatároztuk a multilókuszos tesztek replikációs valószínűségét, eltérő szerkezetű mintákat alkalmazva.

**A vizsgálat alanyai** – Explorációs minta: DSM-IV szerint szkizofréniával diagnosztizált betegekből és egészséges szüleikből álló triók (n=16), replikációs minta: szkizofrén betegek és egészséges alanyok eset-kontroll mintája (n=5337).

**Módszerek** – Kétféle elemzésünkben az egyedi polimorfizmusok asszociációjából a régióalapú gén- és génszett-asszociációt, feldúsulást és funkcionális annotációs klaszterek feldúsulását elemeztük. Az explorációs minta p<0,1 asszociációit teszteltük a replikációs mintán, a replikációt többszörös összehasonlításra korrigáltuk.

**Eredmények** – A génalapú tesztek és a kanonikus útvonaltesztek közül nem találtunk p-érték-korrekción után is szignifikáns asszociációt. A pozicionális génszettek közül 14q31, 5q31 és Xq13 szegmensek esetén korrekció után is szignifikáns asszociációt, illetve feldúsulást észleltünk (p<sub>corr</sub>: 0,002, 0,006 és 0,048). A funkcionális annotációs klaszterelemzésben a splicing/alternatív splicing, idegfejlődés és embrionális fejlődés klaszterei bizonyultak korrekció után is

**Background** – Schizophrenia is a severe psychiatric disorder of poorly understood etiology, characterized by high heritability, multifactorial inheritance and high heterogeneity. Multilocus association methods may reduce the genetic heterogeneity and improve the probability of replication between analyses.

**Objectives** – The aims of our study were twofold: 1. To analyse genetic risk factors of schizophrenia by using multilocus genetic tests. 2. To assess the replication probability attributable to the various multilocus tests.

**Subjects** – Discovery set: case-parent trios of unaffected parents and affected probands with a DSM-IV schizophrenia diagnosis (n=16); replication set: schizophrenia cases and unaffected controls (n=5337).

**Methods** – Associations of single nucleotide and indel markers were transferred to gene- and gene-set-based associations, furthermore to gene-set-enrichment tests and functional annotation cluster analyses in a two-staged design. Associations with p<0.1 from the discovery set were tested in the replication sample. Familywise p-value correction for multiple comparisons were performed during the replication step.

**Results** – After correction for multiplicity, no significant association or enrichment were detected for gene-based nor canonical pathway analyses, but significant association of the 14q31 cytoband and enrichments of the 5q31 and Xq13 cytobands were found (p<sub>corr</sub>: 0.002, 0.006 and 0.048, respectively). Functional annotation clustering yielded

Levezelő szerzők (correspondents): dr. PULAY Attila József, dr. RÉTHELYI János, Semmelweis Egyetem, Pszichiátriai és Pszichoterápiás Klinika; 1083 Budapest, Balassa u. 6. Telefon: (06-1) 210-0330, fax: (06-1) 210-0336. E-mail: pulay.attila@med.semmelweis-univ.hu, rethelyi.janos@med.semmelweis-univ.hu  
<http://orcid.org/0000-0003-2602-3027>  
<http://orcid.org/0000-0002-3641-012X>

Érkezett: 2016. november 21. Elfogadva: 2017. január 12.

szignifikánsnak. A szignifikáns replikációk valószínűsége a teszt összetettségével javult ( $P_{rep}$ : 0, 0,015, 0,21).

**Következtetés** – Multilókuszos genetikai elemzésünk megerősítette az idegfejlődés, szinaptikus plaszticitás és immunrendszer érintettségét a szkizofréniában. A multilókuszmódszerekkel növelhető a replikáció valószínűsége, még jelentősen eltérő minták esetében is.

**Kulcsszavak:** szkizofrénia, multilókuszos-szekvencaanalízis, új generációs szekvenálás, teljesgenom-asszociációs vizsgálat

statistically significant enrichment scores for clusters of splicing/alternative splicing, neurodevelopment and embryonic development. Improvements in replication probability were found with increased test complexity ( $P_{rep}$ : 0, 0.015, 0.21).

**Conclusions** – Our results corroborate the involvement of neurodevelopment, synaptic plasticity and immune mechanisms in the etiology of schizophrenia. Also, our findings indicated improvement of replication probability by using multilocus genetic analyses.

**Keywords:** schizophrenia, multilocus sequence analysis, high-throughput sequencing, genomewide association study

A szkizofrénia komplex neuropszichiátriai zavar, amelyet fiatal felnőttkorban induló, és többnyire visszatérő, gyakran krónikussá váló pszichotikus epizódok jellemeznek, a pszichés funkciók és az életminőség jelentős, gyakran progresszív károsodását eredményezve. Változatos tüneti képe ellenére prevalenciája világszerte 0,5–1,2% között mérhető, amely kultúrától független, döntően biológiai eredetre utal. Ezzel egybecsengő a család- és iker-vizsgálatok közel 50 éve ismert eredményei, amelyek a szkizofrénia magas heritabilitását ( $h^2 \sim 0,8$ ) mutatták ki. Mindezek a pszichiátriai zavarokra általánosan jellemző, multifaktoriális, poligénes öröklődést valószínűsítették szkizofréniában is<sup>1</sup>.

A szkizofrénia teljesgenom-asszociációs vizsgálatok (genome-wide association study, GWAS) első hulláma azonban rámutatott a komplex fenotípusok teljesgenom asszociációs elemzésének általános problémájára: a genotipizált, gyakori egyedi nukleotid polimorfizmusok (single nucleotide polymorphism, SNP) asszociációi általában nagyon kis hatáserősségűek (OR: 1,05–1,1), jóval alacsonyabban a vizsgálatok tervezésekor becsült  $OR > 1,3$  hatáserősségnél. Az első vizsgálatok ezért csak néhány szignifikáns markert azonosítottak, asszociációikat többnyire nem replikálták a későbbi vizsgálatok, és az általuk kifejezett genomikus heritabilitás a szegregációs elemzések értékének töredékét érte csak el ( $(2-7\%)^2$ ). Ez alól kivételt jelent a 6p21–22 lókuszos, a major hisztokompatibilitási (MHC) régió, melynek asszociációját szkizofréniával minden GWAS-vizsgálat megerősítette<sup>3</sup>.

A gyenge replikációs valószínűség egyik oka a lókuszos (genotípus) -heterogenitásban keresendő, amely az egyedi nukleotid variánsokat sokkal nagyobb mértékben érinti, mint a régiókat, biológiai

útvonalakat, génhálózatokat. A multifaktoriális öröklésmenttel jellemezhető betegségeket – így a pszichiátriai zavarok többségét is – magas lókuszos és fenotípus-heterogenitás jellemzi. A biológiai, molekuláris funkciók, vagy kromoszómán való elhelyezkedés, esetleg a kódolt fehérjék kölcsönhatása alapján csoportosított gének elemzése során alacsonyabb heterogenitást feltételezhetünk, és az ilyen típusú elemzés az eredményt rögtön funkcionális kontextusba helyezi. Azonban e módszerek érvényessége nagyban függ az alkalmazott annotációk minőségétől, azaz a genom egyes szakaszainak génekhez, géncsoportokhoz és génen belüli funkciókhoz (génpromoter, enhancer, aktívan átíródo szakasz) való kapcsolásának, a genom „jelölésének” pontosságától. Az annotációs adatbázisok gyors bővülése mellett a hiányos genomikus lefedettség (üres, ismeretlen szakaszok a genomon belül)<sup>4</sup> és az annotációs redundancia (ismétlődő, átfedő vagy egymásnak ellentmondó annotációs megjelölés)<sup>5</sup> egyszerre fordul elő.

A szkizofrénia esetében is számos GWAS-, expressziós profil- és teljesgenom-kópiaszámvariáció (CNV) -vizsgálat útvonalelemzése történt korábban, esetenként egymásnak ellentmondó eredményekkel<sup>6</sup>. A glutamáttranszmisszió, szinaptikus plaszticitás, axon- és neuronfejlődés, transzkripció- és sejtosztódás-szabályozás útvonalainak eltéréseit több vizsgálat is bemutatta szkizofréniában, bár a legnagyobb elemszámmal rendelkező Psychiatric Genomics Consortium (PGC) ilyen típusú vizsgálata nem talált többszörös összehasonlítás korrekciója után sem szignifikáns útvonal-asszociációt<sup>3</sup>.

Jelen vizsgálatunkban ezekből kifolyólag a szkizofrénia genetikai hátterének vizsgálatára törekedtünk gén- és génszettelapú multilókuszos-asszociáci-

**1. táblázat.** A vizsgálat mintái és az alkalmazott elemzési típusok

Jellemző	Explorációs minta	Replikációs minta	
forrástanulmány	SCHIZOBANK	GAIN SCZ*	MGS SCZ**
elérési kód	–	phs000021.v3.p2	phs000167.v1
teljes minta (eset-kontroll)	n=16 (trió)	n=2787 (1318:1469)	n=2935 (1405:1530)
elemzett minta (eset-kontroll)	n=16	n=5337 (2632:2705)	
genotipizálás	teljes exom szekvenálás	Affymetrix Genomwide 1M csip	
genotipizált SNV/SNP száma minőségkontroll után	120 719	716 923	
SNP száma imputáció után	–	7 942 021	
SNP-alapú elemzés (szoftver)	Family-Based Association Test	logisztikus regresszió***	
génalapú asszociációs teszt (szoftver)	Kiterjesztett Simes' teszt, GATES (KGG v3.5)	Kiterjesztett Simes' teszt, GATES (KGG v3.5)	
génszettelapú asszociációs teszt (szoftver)	Hybrid Set-based Test, HYST (KGG v3.5)	Hybrid Set-based Test, HYST (KGG v3.5)	
génszettelapú feldúsulás teszt (szoftver)	Wilcoxon előjeles rangszámösszeg teszt (KGG v3.5)	Wilcoxon előjeles rangszámösszeg teszt (KGG v3.5)	
funkcionális annotációs klaszterelemzés	DAVID 6.8, közepes szigorúság	DAVID 6.8, közepes szigorúság	
DAVID-variáns és génprioritizáció	KGGseq splicing, mendeli és komplex patogenitás predikció vagy EIGEN > 1,5	KGGseq splicing, mendeli és komplex patogenitás predikció vagy EIGEN > 1,5	

\* GAIN SCZ: Genetics Information and Association Network, Genome-Wide Association Study on Schizophrenia

\*\* MGS SCZ: Molecular Genetics of Schizophrenia

\*\*\* logisztikus regressziós modell életkor, nem és populációs főkomponensekre (PCA) kontrollálva

ós és feldúsulási teszteket végezve. Másodsorban felmértük, hogy milyen mértékben változtatják a replikáció valószínűségét különböző komplexitású multilokusz genetikai módszerek egymástól jelentősen eltérő minták esetében.

## Módszerek

### VIZSGÁLATI MINTÁK

Vizsgálatunkban három, egymással nem átfedő mintát elemeztünk, amelyeket kétlépcsős szerkezetben, explorációs, valamint replikációs mintaként elemeztünk (**1. táblázat**). Kutatásunk explorációs mintáját a SCHIZOBANK vizsgálat<sup>7</sup> (n=16) eset-szülő triója alkotta, mindegyik trió DSM-IV szkizofréniával diagnosztizált utód és a két, szkizofréniával nem diagnosztizált, egészséges szülő mintájából állt. A DSM-IV-TR szerinti szkizofrénia diagnózist a Structured Interview for Clinical Studies (SCID) strukturált diagnosztikus interjú segítségével állítottuk fel, leggyakoribb diagnózis a nem differenciálható szkizofrénia volt (n=10). A SCHIZOBANK multicentrikus vizsgálat 18 és 60 év közötti, szkizofréniával diagnosztizált beteg és egészséges kontrollszemélyek mintájának biobankolására

jött létre, kizárási kritériumot a bevonás idején fennálló, aktív droghasználat vagy alkoholabúzus és -függőség, vagy más neurológiai betegség jelentett.

Vizsgálatunk replikációs mintája a National Center for Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP) tudományos közösség számára hozzáférhető, teljesgenom-adatbázisának két szkizofrénia-GWAS eset-kontroll vizsgálatából származott. A Genetic Association Information Network (GAIN) Genome-Wide Association Study on Schizophrenia<sup>8</sup> (azonosító: phs000021.v3.p2, n=2787) és a Molecular Genetics of Schizophrenia<sup>9</sup> (azonosító: phs000167.v1.p1, n=2935) európai származású résztvevőiből vontuk össze a teljes replikációs mintát (n=5722,  $n_{SCZ}=2723$ ,  $n_{KONT}=2725$ , és 274 ismeretlen fenotípusú egyén). A szkizofrénia diagnózisának felállítására mindkét mintában legalább két, gyakorlott klinikus egybehangzó vizsgálati lelete alapján került sor, beválasztásra azon 18 éves vagy idősebb, szkizofrénia vagy szkizoaffektív zavar DSM-IV-TR kritériumait teljesítő alanyok kerültek, akiket a vizsgálat szervezői legalább két éve ismertek. Kizárási kritériumok közé az angol nyelv nem elégséges ismerete, a pszichotikus tünetek háttérben szerhasználat vagy neurológiai betegség fenn-



állása, illetve súlyos mentális retardáció tartozott. Az adattisztítás lépései során kizártuk a hiányzó fenotípusú, kevert származásúnak, valamint rejtett rokonságban állónak bizonyuló egyéneket ( $n=5337$ ,  $n_{SCZ}=2632$ ,  $n_{KONT}=2705$ ).

#### GENOTIPIZÁLÁS, MINŐSÉG-ELLENŐRZÉS ÉS IMPUTÁCIÓ

A SCHIZOBANK eset-trió mintáit új generációs, teljesexom-szekvenálás módszerével genotipizáltuk, Illumina HiScan (TMSQ) szekvenátor felhasználásával. A könyvtárkészítés NimbleGen SeqCap EZ Human Exome Library v3.0 exome kit segítségével történt. A szekvenálás és polimorfizmusok meghatározásának minőségét térképezési minőségpont (MQ>20), variáns Phred-pont (site Phred score >20), és olvasási lefedettség (read-depth >4) alapján ellenőriztük. A minőségi szűrés és mendeli hibák eltávolítása után 120 719 biallélikus SNP-vagy indelvariáns állt rendelkezésre, ebből 66 737 mutatott additív transzmissziót minimálisan egy informatív családdal, közülük 29 800 allél transzmissziója volt legalább négy informatív családban vizsgálható.

Mind a GAIN-, mind az MGS GWAS-mintáinkat az Affymetrics Genome-Wide 6 csippel genotipizálták, amely 1 000 000 gyakori SNP-t hibridizál. Minőség-ellenőrzés során kiszűrésre kerültek a túl ritka (minor allélfrekvencia:  $MAF < 0,01$ ), a Hardy-Weinberg egyensúlyi eloszlástól túlzottan eltérő ( $p_{H-W} < 1 \cdot 10^{-6}$ ), valamint a 2% genotípushiányt meghaladó variánsok. A szűrést követően 716 923 közös genotipizált SNP maradt mindkét GWAS-mintában, ezekből csak 9366 SNP volt megtalálható SCHIZOBANK teljesexom-mintán is. Az átfedés növelése érdekében a GWAS-mintát az IMPUTE2<sup>10</sup> haplotípus-alapú algoritmusával imputáltuk, a 1000 Genomes<sup>11</sup> Phase1v3 kiadásának teljes populációját használva referenciagenomként. Az imputációs hibák csökkentése érdekében kizártuk a referenciagenom ritka markereit ( $MAF < 0,01$ ) és imputációt követően csak a legjobb minőségű, biallélikus variánsokat tartottuk meg (imputációs valószínűség > 0,9, információmutató > 0,9,  $MAF > 0,01$ ). A minőségkritériumoknak 7 942 021 SNP-vagy indelvariáns felelt meg, amellyel a minták közti lefedettség 22 279 markerre bővült. A két minta annotációját a KGGseq<sup>12</sup> szoftver segítségével végeztük az UCSC hg19 referenciagenom és a GENCODE v19 koordináták alapján.

#### STATISZTIKAI ELEMZÉS

Vizsgálatunk három, egymásra épülő elemzésre tagozódott.

1. A multilokuszesztekhez szükséges teljesgenom-variáns asszociációk kiszámítása mindkét tesztmintán az additív modell szerint, korrigálva a zavaró hatásokra.

2. A variáns asszociációk kombinációja alapján, teljes genom, gén- és génszettelalapú asszociációk és feldúsulások elemzése az explorációs mintán, és a legalább szuggesztív asszociációt vagy feldúsulást mutató ( $p < 0,1$ ) gének és génszettek megvizsgálása a replikációs mintán.

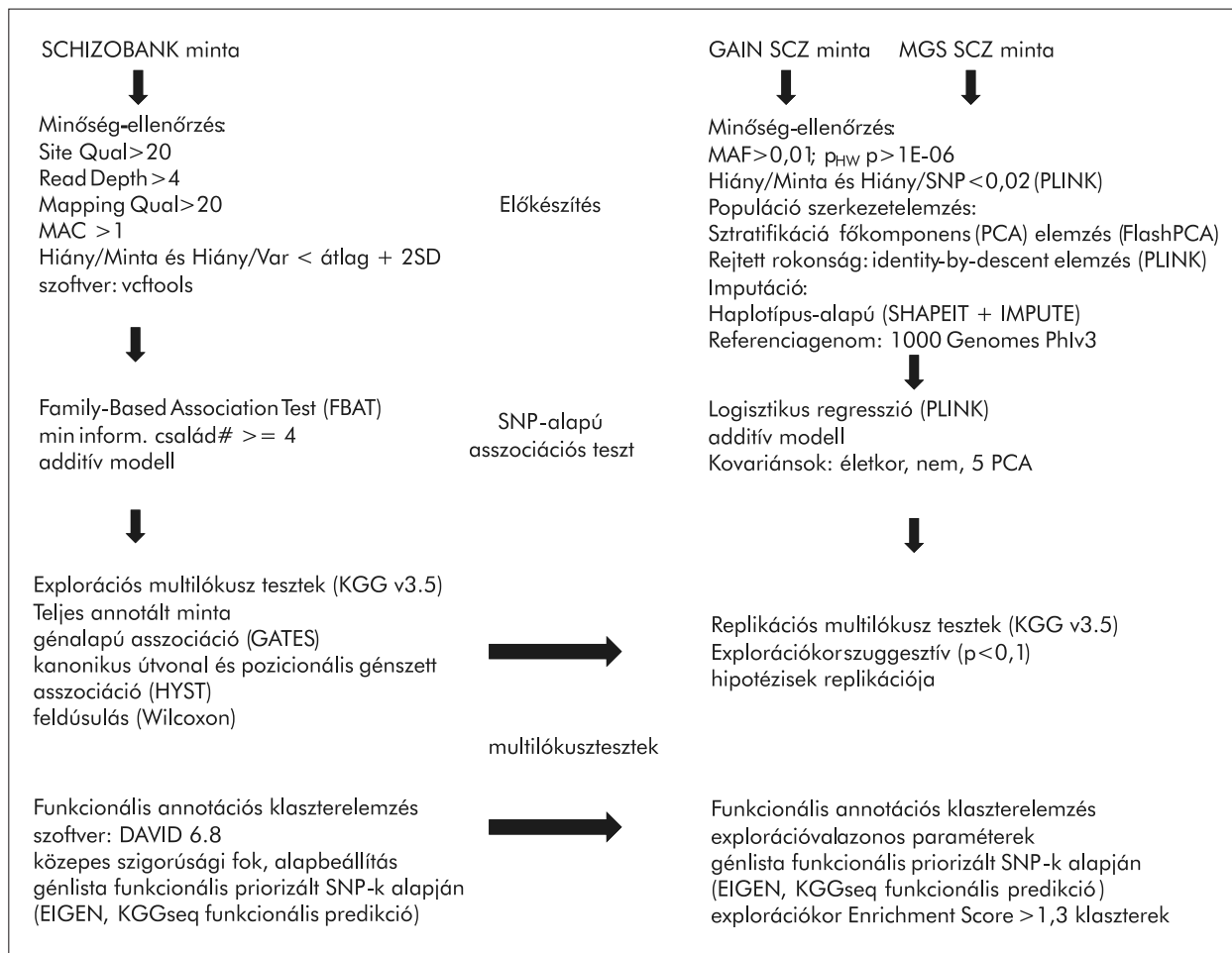
3. Predikációs algoritmusokkal funkcionálisnak jelzett variánsokat hordozó rizikogének meghatározása, és funkcionális annotációik klaszterezése és elemzése mindkét mintán (**1. ábra**). A többszörös összehasonlítás korrekcióját a replikációs lépésben végeztük, a különböző multilokuszmódszerek szerinti hipotéziscsoportoknak megfelelően, a csoportba tartozó explorációs és replikációs tesztek összesített száma szerinti Bonferroni-korrekcióval. A géalapú hipotéziscsoport 17 600, a génszett-asszociációk 1107, a génszettefeldúsulás 1107, a funkcionális annotációs klaszterelemzés pedig 927 tesztelt hipotézist tartalmazott.

#### Az SNP- és indelasszociációk elemzése

A variánsok asszociációs elemzését mindkét mintán additív öröklési modellt feltételezve végeztük. Az explorációs, eset-szülő trió alapú SCHIZOBANK-mintában, a minta alacsony elemszámára tekintettel, az inkomplett genotípusú triókat is használni képes, Family-Based Association Test<sup>13</sup> (FBAT) elemzést végeztük el. Az alacsony elemszám miatt az FBAT-statisztika számításához szükséges informatív triók számát a javasolt 10-ről 4-re csökkenttük, a teszt érvényességét maximálisan 100 000 ciklusban végzett adaptív permutációval biztosítottuk. A replikációs minta megfelelő nagysága miatt logisztikus regressziós modelleket használtunk, amelyek kovariánsként a nem és az életkor mellett, a populációstratifikáció zavaró hatását korrigáló főkomponenseket is tartalmazták. A főkomponenselemzést a genotipizált, egymással nem kapcsolt ( $r^2 < 0,2$ ) markerekre korlátozva, a FlashPCA<sup>14</sup> program segítségével végeztük. A replikációs minta SNP-alapú asszociációit a PLINK<sup>15</sup> program segítségével számítottuk.

#### Gén- és génszettelalapú elemzés

A gén- és génszettelalapú analíziseket a Knowledge-based mining system for Genome-wide Genetic studies (KGG)<sup>16</sup> szoftver 3.5 verziójával végeztük, amely az SNP-k asszociációs statisztikai és kapcsoltsági mutatói alapján számítja a régióalapú



1. ábra. A vizsgálat módszertani összefoglalója

asszociációt és a feldúsulási összefüggéseket. A génalapú asszociációkat a KGG kiterjesztett Simes' tesztjével (GATES)<sup>17</sup> elemeztük, amely a vizsgált gén legjobb SNP-asszociációja alapján határozza meg a régió asszociációját. A módszer előnye, hogy sok neutrális marker esetén sem veszít statisztikai erejéből, és kevésbé érzékeny a kapcsoltság hatására, vagy a markerek számára. A vizsgált gének határait a GENCODE v19 kiadásában rögzített koordinátákhoz képest mindkét irányban 5 kilobázis méretben kiterjesztve határoztuk meg.

A kanonikus útvonalakat és a pozicionális génszetteket az MSiGDB<sup>18</sup> v5.2 c2.cp, illetve c1 adatbázis alapján végeztük, a kanonikus útvonalak ismert átfedései miatt a redundanciát a ReCiPa<sup>19</sup> algoritmusával kontrolláltuk, 0,85 maximális, 0,1 minimális átfedés paraméterekkel. Az MSiGDB 1329 kanonikus útvonalából a ReCiPa algoritmus 541 nem átfedő kanonikus útvonalat, és 122 „superútvonalat” (Superpathway) definiált. Génszettelemzés metodológiai ajánlásait követve, két külön-

böző statisztikai módszert alkalmaztunk: a KGG hybrid set-based test (HYST)<sup>20</sup> nevű, önálló (nem kompetitív), régióalapú asszociációt vizsgáló tesztjét, valamint a kompetitív jellegű, Wilcoxon előjeles rangszámösszeg feldúsulási elemzést. A HYST teszt az adott génszethez tartozó, de egymással nem kapcsolt régiók,  $\chi^2$ -statisztikáit összegezve határozza meg a génszett közös asszociációját. A Wilcoxon-teszt pedig a vizsgált génszett géneihez tartozó GATES p-értékek eloszlását veti össze a génszeten kívüli gének p-értékeinek megoszlásával. Az útvonalak méretéből eredő torzítás minimalizálása érdekében az elemzést 10–500 gént tartalmazó útvonalakra korlátoztuk. A tesztekhez szükséges kapcsoltsági paramétereket az imputációhoz is használt, 1000 Genomes Phase1v3 kiadásának európai leszármazású (n=379) haplotípusaiból számítottuk. Az explorációs mintán szugesztív asszociációt mutató ( $p < 0,1$ ) gének és útvonalak asszociációt azonos paraméterek mellett vizsgáltuk a replikációs mintán.

*Funcionális annotációs klaszterelemzés*

A funkcionális annotációs klaszterek meghatározását az online elérhető Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8 verziójával<sup>5, 21</sup> végeztük. Az információs források növelése érdekében az alapbeállítás annotációs forrásait kiegészítettük az ontológiai kategóriából „leggyakrabban használt terminológia” (GOTERM\_FAT) adatokkal, az általános annotációs csoportból a kromoszóma, a citogenetikai sáv és a hivatalos gén szimbólum annotációival, az útvonalak kategóriából pedig a REACTOME útvonalai-val. A klaszterek meghatározásához használt felül-reprezentációs teszt bemeneti adatait mindkét mintában azonos algoritmus szerint kiválasztott rizikógének listái jelentették. Rizikógénként a KGGseq mendeli és regulátoros predikciós algoritmusai alapján, vagy az EIGEN<sup>22</sup> spektrális algoritmusai alapján funkcionálisnak jelölt, illetve a KGGseq annotációja alapján splicing módosítást valószínűsítő, legalább szuggesztív asszociációval bíró (explorációs minta:  $p < 0,1$ , replikációs minta:  $p < 0,05$ ) variánsokat hordozó géneket definiáltuk. Az explorációs SCHIZOBANK-mintában a kombinált predikciós algoritmussal 441, valószínűsítetten funkcionális variánst hordozó 402 rizikógént azonosítottunk, míg a replikációs GAIN-MGS mintában 3941 funkcionálisnak jósolt variánst és 1785 rizikógént határoztunk meg. Mind a rizikóvariánsok, mind a rizikógének tekintetében alacsony volt a minták közti átfedés (közös rizikóvariáns: 23, közös rizikógén: 75), az explorációs minta rizikóvariánsainak 5,21%-át és rizikógénjeinek 18,66%-át jelentette. A DAVID teszt háttéréként a mintákban a KGGseq által annotált GENCODE géneket vettük alapul, az explorációs minta esetében ez 14 618 gént jelentett, a replikációs minta 38 940 gént fedett le. A DAVID klaszterező algoritmusát az alapértelmezett (közepes) szigorúsági fokon használtuk. A klaszterek feldúsulási mutatója (enrichment score, ES) az annotációs terminusok p-értékei negatív logaritmusának mértani közepe. Ezért a klaszterek nominálisan szignifikáns küszöbértékét  $ES = 1,3$  ( $-\log_{10} 0,05$ ) pontban határoztuk meg, a kis elemszámú explorációs mintákon az  $ES > 1,2$  értékkel rendelkező klasztereket is megvizsgáltuk.

**Eredmények**

## GÉNALAPÚ ASSZOCIÁCIÓK

Az explorációs teljesexom-mintán 29 821 variáns felelt meg a szűrési kritériumoknak, amelyek a

GENCODE alapján 15 795 gént és transzkriptet annotáltak, a replikációs mintán pedig 7 848 329 variáns 51 366 GENCODE gént és transzkriptet annotált. A variáns asszociációkat géalapú asszociációvá kombináló GATES teszt 1805 gén esetében eredményezett szuggesztív p-értéket ( $p < 0,1$ ) az explorációs mintán, közülük 145 gén asszociálódott legalább nominálisan szignifikáns p-értékkel ( $p < 0,05$ ) a replikációs mintán is. Mindazonáltal, a két minta 10 legalacsonyabb p-értékkel rendelkező génjei között nem találtunk átfedést, és a géalapú tesztek teljes száma szerint ( $n = 17\ 600$ ) végzett Bonferroni-korrekción után a legjobb p-érték sem maradt szignifikáns (PTPN21,  $p_{\text{exp}} = 0,026$ ,  $p_{\text{rep}} = 6,42 \cdot 10^{-6}$ ,  $p_{\text{corr}} = 0,113$ ). A géalapú tesztek  $P_{\text{rep\_nom\_g}} = 0,008$ , nominális replikációs rátát és  $P_{\text{rep\_corr\_g}} = 0$  többszörös összehasonlításra korrigált replikációt eredményeztek. A leginkább replikált géalapú asszociációk p-értékei a **2. táblázatban** láthatók.

## GÉNSZETTELEMZÉSEK: KANONIKUS ÚTVONAL ÉS POZICIONÁLIS GÉNSZETT-TESZTEK

Az explorációs mintán a 663 útvonalból 104 mutató szuggesztív ( $p < 0,1$ ) p-értéket a HYST régió-alapú asszociációs tesztje során, vagy a Wilcoxon rangszámösszeg feldúsulási teszt alapján. A 104 szuggesztív útvonalból hat útvonal esetében sikerült, az explorációs p-értékeknek megfelelően, a HYST asszociációs tesztet, vagy pedig a Wilcoxon feldúsulási tesztet legalább nominálisan szignifikáns p-értékkel replikálni a GAIN-MGS GWAS-mintán, azonban egyik asszociáció vagy feldúsulás sem maradt szignifikáns a többszörös összehasonlítás korrekciója után. A nominális szignifikanciával replikált kanonikus útvonalakat a **3. táblázat** mutatja be.

A pozicionális adatbázis 326 kromoszómaszegmenséből az explorációs mintán 248 felelt meg a méretbeli kritériumoknak, amelyek közül 91 szegmens HYST vagy Wilcoxon p-értéke bizonyult szuggesztív mértékben szignifikánsnak. A replikációs mintán nyolc szegmens asszociációjának vagy feldúsulásának nominálisan szignifikáns replikációját észleltük, ebből egy szegmens asszociációja és két szegmens feldúsulása a többszörös összehasonlítás korrekcióját követően is szignifikáns maradt (14q31:  $p_{\text{exp\_HYST}} = 0,098$ ,  $p_{\text{rep\_HYST}} = 1,5 \cdot 10^{-6}$ ,  $p_{\text{rep\_HYST\_corr}} = 0,002$ ; 5q31:  $p_{\text{exp\_Wilcoxon}} = 7,2 \cdot 10^{-4}$ ,  $p_{\text{rep\_Wilcoxon}} = 5,6 \cdot 10^{-6}$ ,  $p_{\text{rep\_Wilcoxon\_corr}} = 0,0062$ ; Xq13:  $p_{\text{exp\_Wilcoxon}} = 0,018$ ,  $p_{\text{rep\_Wilcoxon}} = 4,5 \cdot 10^{-5}$ ,  $p_{\text{rep\_Wilcoxon\_corr}} = 0,048$ ). A génszett-tesztek összesített nominális replikációja  $P_{\text{rep\_nom\_s}} = 0,072$ , a többszörös összehasonlításra korrigált pedig  $P_{\text{rep\_corr:s}} = 0,015$ . A

**2. táblázat.** Legjobb 10 replikációs p-értékkel bíró gén és transzkript régióalapú asszociációja az explorációs és a replikációs mintán

Gén (szimbólum)	Kromoszóma	Pozíció (bázispár)	Explorációs minta (n=16) P_GATES*	Replikációs minta (n=5337) P_GATES (korrigált)
Protein tirozin foszfatáz nonreceptor 21 (PTPN21)	14	88932121-89016966	0,0260	$6,42 \cdot 10^{-6}$ (0,113)
Kromoszóma strukturális fenntartó 1B (RP1-102D24.5)	22	45831743-45844624	0,0455	0,0002 (1,00)
Mátrix remodellelling asszociált 5 (MXRA5)	X	3226605-3264682	0,0614	0,0005 (1,00)
C2H2 típusú cinkujj (ZFP42)	4	188916924-188926199	0,0325	0,0010 (1,00)
RIB43A coiled-coil domén 2	22	45809571-45828376	0,0455	0,0013 (1,00)
Feszültségfüggő Ca <sup>2+</sup> -csatorna $\alpha 2d4$ alegység (CACNA2D4)	12	1901122-2028002	0,0650	0,0015 (1,00)
BCL2 -asszociált athanogén 6 (BAG6)	6	31606804-31607565	0,0348	0,0017 (1,00)
Szinaptonémás komplex központi elem fehérje 2(SYCE2)	19	13009599-13030090	0,0707	0,0019 (1,00)
XK-asszociált 4 (XKR4)	8	56014948-56454613	0,0252	0,0022 (1,00)
MACC1 antiszenz RNS 1 (MACC1-AS1)	7	20181538-20193154	0,0143	0,0022 (1,00)

P\_GATES\*: Kiterjesztett Simes'-teszt alapján számított régióalapú asszociációs p-érték

**3. táblázat.** Nominálisan replikált kanonikus útvonalak és átfedő útvonalakból összeállított „szuperútvonalak” mérete, asszociációs (HYST), illetve feldúsulási (Wilcoxon) p-értékei az explorációs és a replikációs mintán

Útvonal	Explorációs minta (n=16)		Replikációs minta (n=5337)	
	P_HYST	p_wilcoxon	p_HYST (korrigált)	p_wilcoxon (korrigált)
Superpathway 75	0,0086	0,0328	0,0039 (1,00)	0,5625 (1,00)
Superpathway 106	0,0304	0,0030	0,0047 (1,00)	0,7163 (1,00)
Superpathway 69	0,0329	0,2020	0,0071 (1,00)	0,4661 (1,00)
PID_HNF3B_PATHWAY	0,0875	0,0928	0,0149 (1,00)	0,5711 (1,00)
Superpathway 107	0,0292	0,0130	0,0166 (1,00)	0,5723 (1,00)
PID_P38_ALPHA_BETA_PATHWAY	0,0769	0,2305	0,0168 (1,00)	0,2107 (1,00)
			Gének száma	Gének száma
			57	121
			61	99
			13	18
			23	44
			52	123
			18	31

HYST:hybrid set-based association test, génszettelalapú asszociációs teszt

Wilcoxon: Wilcoxon rangösszeg feldúsulási teszt

Superpathway 69: REACTOME PD1 SIGNALING + REACTOME PHOSPHORYLATION OF CD3 AND TCR ZETA CHAINS

Superpathway 75: KEGG CELL CYCLE + BIOCARTEA CELLCYCLE PATHWAY + BIOCARTEA G1 PATHWAY

Superpathway 107: KEGG OXIDATIVE PHOSPHORYLATION + REACTOME RESPIRATORY ELECTRON TRANSPORT + REACTOME INSULIN RECEPTOR RECYCLING

Superpathway 106: REACTOME SIGNALING BY INSULIN RECEPTOR + REACTOME INSULIN RECEPTOR RECYCLING + REACTOME INSULIN RECEPTOR SIGNALING CASCADE + REACTOME ACTIVATED POINT MUTANTS OF FGFR2 + REACTOME SIGNALING BY ACTIVATED POINT MUTANTS OF FGFR1 + REACTOME SIGNALING BY FGFR3 MUTANTS + REACTOME REGULATION OF AMPK ACTIVITY VIA LKB1 + REACTOME SHC MEDIATED CASCADE + REACTOME ENERGY DEPENDENT REGULATION OF MITO BY LKB1 AMPK + REACTOME REGULATION OF RHEB GTPASE ACTIVITY BY AMPK + REACTOME PKB MEDIATED EVENTS + REACTOME SIGNAL ATTENUATION + REACTOME SOS MEDIATED SIGNALING + REACTOME RAF MAP KINASE CASCADE + REACTOME SHC MEDIATED SIGNALING + REACTOME MTORC1 MEDIATED SIGNALING + REACTOME FGFR LIGAND BINDING AND ACTIVATION + REACTOME FGFR2C LIGAND BINDING AND ACTIVATION + REACTOME FGFR4 LIGAND BINDING AND ACTIVATION + REACTOME MTORC2 MEDIATED SIGNALING + REACTOME FGFR1 LIGAND BINDING AND ACTIVATION + REACTOME PI3K CASCADE + REACTOME SHC RELATED EVENTS + REACTOME FRS2 MEDIATED CASCADE

**4. táblázat.** Nominálisan replikált pozicionális génszettek mérete, asszociációs (HYST), illetve feldúsulási (Wilcoxon) p-értékei az explorációs és a replikációs mintán

Kromoszómasáv	Explorációs minta		Gének száma	Replikációs minta (n=5337)		(n=16)
	P <sub>-HYST</sub> (korrigált)	P <sub>-Wilcoxon</sub> P <sub>-Wilcoxon</sub>		P <sub>-HYST</sub> (korrigált)	Gének száma	
14q31	0,0977	0,0374	16	1,54*10 <sup>-6</sup> (0,002)	0,0436 (1,00)	32
1p22	0,0320	0,0013	50	0,0112 (1,00)	0,1827 (1,00)	81
Xq13	0,0736	0,0181	12	0,0171 (1,00)	4,32*10 <sup>-5</sup> (0,048)	78
5q31	0,4765	7,20*10 <sup>-4</sup>	134	0,0302 (1,00)	5,75*10 <sup>-6</sup> (0,006)	194
10q23	0,0123	2,87*10 <sup>-6</sup>	55	0,0443 (1,00)	0,4366 (1,00)	91
5q12	0,0751	0,0150	29	0,0453 (1,00)	0,2533 (1,00)	43

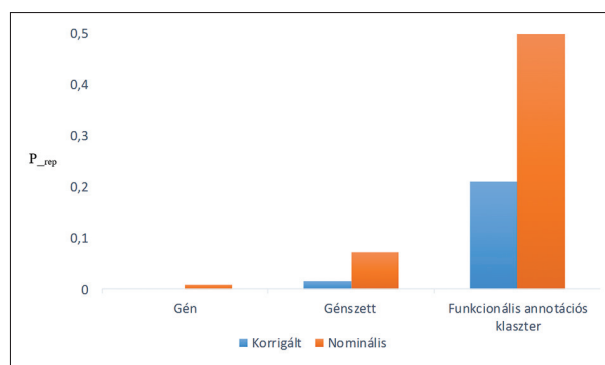
HYST: hybrid set-based association test, génszettelapú asszociációs teszt

Wilcoxon: Wilcoxon rangösszeg feldúsulási teszt

nominálisan szignifikáns replikált pozicionális génszetteket a **4. táblázat** tartalmazza.

#### FUNKCIONÁLIS ANNOTÁCIÓS KLASZTERELEMZÉS

Az explorációs minta 402 rizikógénjéből 386, míg a háttérként meghatározott 14 618 gén közül 13 647 rendelkezett DAVID-azonosítóval, a replikációs mintában pedig az 1785 rizikógénből 1543, a háttérkénekből pedig 24 517 szerepelt az adatbázisban. Az explorációs minta 306 klaszteréből 12 annotációs klaszter bírt az ES>1,3 küszöb feletti feldúsulási mutatóval, további két annotációs klaszter volt jellemezhető közvetlenül küszöbérték alatti, de még szuggesztívnek tekinthető feldúsulási ponttal (ES>1,2). Azonos annotációs források és szigorúsági kritérium mellett, a replikációs minta rizikógénjeivel a lehetséges 621 klaszterből 126 annotációs klasztert azonosítottunk ES>1,3 feldúsulással, amelyek közül hét megegyezett az explorációs mintán ES>1,2 küszöbértékkel azonosított klaszterekkel. Legnagyobb feldúsulási mutatóval mindkét mintában a splicing (pre-mRNS nem átródó részeinek kivágódása és érett mRNS-sé alakulása) és alternatív splicing annotációs terminusok klasztere bírt (ES<sub>exp</sub>: 2,85, ES<sub>rep</sub>: 23,93). Továbbá, az explorációs mintától eltérő rangsorban, de a replikációs mintában is feldúsultak az endoszómális transzport (ES<sub>exp</sub>: 1,56, ES<sub>rep</sub>: 1,90), a neuro- és axongenezis (ES<sub>exp</sub>: 1,44, ES<sub>rep</sub>: 2,25), az izom- és vázrendszer fejlődése (ES<sub>exp</sub>: 1,40, ES<sub>rep</sub>: 1,95), az embrionális fejlődés (ES<sub>exp</sub>: 1,34, ES<sub>rep</sub>: 4,77), a motoros viselkedés (ES<sub>exp</sub>: 1,33, ES<sub>rep</sub>: 2,30) és az agyfejlődés (ES<sub>exp</sub>: 1,23, ES<sub>rep</sub>: 4,55) funkcionális annotációs terminusaival kapcsolatos klaszterei. Mivel a szuggesztív klaszterek szelektíven nem vizsgálhatók, a klaszteranalízis replikációját a klaszterek összesített számára korrigáltuk ( $\alpha=0,05/927$ ), ennek megfelelően az ES>4,27 feldúsulású klaszterek, azaz a splicing, az agyfejlődés és embrionális fejlődés

**2. ábra.** Nominális és többszörös összehasonlításra korrigált replikációs valószínűség (P<sub>-rep</sub>) a növekvő összetettségű multilókus-elemzésekben

klaszttere, korrekció után is szignifikáns maradt. A funkcionális annotációs klaszterezés tehát P<sub>rep\_nom\_ac</sub>=0,5 nominális, illetve P<sub>rep\_corr\_ac</sub>=0,21 többszörös összehasonlításra korrigált replikációs rátát mutatott (**2. ábra**). A replikált klaszterek annotációs terminusaihoz kapcsolódó p<sub>EASE</sub> és FDR-értékek részletesen az **5. táblázatban** szerepelnek.

#### Megbeszélés

Jelen vizsgálatunkban a szkizofrénia multilókus-markereinek régióalapú asszociációját, génszettelapú feldúsulását, valamint a feldúsult funkcionális annotációk klasztereit elemeztük. A SCHIZO-BANK eset-szülő trió mintán exploratív elemzéssel azonosított gének és kanonikus útvonalak közül számos asszociációt, valamint feldúsulást megismételtünk, replikáltunk nominális p-értékkel a GAIN-MGS GWAS mintán, azonban a többszörös összehasonlítás korrekcióját követően ezek egyike sem maradt szignifikáns, korrekció után a legalacsonyabb korrigált p-értékkel (p=0,113) a Protein

tirozin foszfatáz nonreceptor 21 (PTPN21) gén rendelkezett. A pozicionális génszettek elemzésekor, az explorációs mintán a 10q23 kromoszómasáv feldúsulása, a replikációs mintán pedig a PTPN21 gént is tartalmazó, 14q31 szegmens régióalapú asszociációja, valamint az 5q31 és az Xq13 kromoszómaszegmensek feldúsulási tesztje esetében észleltünk – többszörös összehasonlításra korrigáltan is – szignifikáns p-értéket. A két mintán azonos paraméterű predikációs algoritmusokkal priorizált génlisák funkcionális annotációs klaszterelemzése, mindkét mintán a szkizofrénia feltételezett neurobiológiai hátterére utalhat, az elemzés a sejtszótadás és transzkripció szabályozásával, neuron- és axonfejlődéssel, agy- és embrionális fejlődéssel, illetve szinaptikus plaszticitással kapcsolatos annotációs klaszterek feldúsulását mutatta. Az explorációs mintán azonosított, 14 annotációs klaszter fele a replikációs mintán is nominális feldúsulást mutatott, közülük három, a szkizofrénia etiológiájához jól illeszkedő splicing/alternatív splicing, az agyfejlődés és az embrionális fejlődés klasztere, statisztikailag szignifikáns maradt a teljes hipotéziscsoportnak megfelelő p-érték-korrekción után is. A vizsgált multilokusz-„markerek” komplexitásának növekedésével mind a nominális, mind a korrigált replikációs valószínűség növekedését észleltük, amely megerősítette a multilokusz-asszociációs módszerek genetikai heterogenitást ellensúlyozó hatását.

Az explorációs lépcsőben szuggesztív p-értékkel bíró gének közül a 14q31 lokalizációjú PTPN21 asszociálódott a szkizofréniahoz a legalacsonyabb p-értékkel ( $p_{\text{rep}}=6,42 \cdot 10^{-6}$ ) a replikációs mintában. Bár géalapú tesztek hipotéziscsoportjára korrigált p-értéke csak szuggesztív övezetbe sorolható ( $p_{\text{rep\_corr}}=0,113$ ), az asszociáció érvényessége mellett szól, hogy a gént tartalmazó 14q31 pozicionális génszett asszociációja a korrekció után is szignifikáns maradt ( $p_{\text{rep\_HYST}}=1,5 \cdot 10^{-6}$ ,  $p_{\text{rep\_HYST\_corr}}=0,002$ ). A PTPN21 és a szkizofrénia kapcsolatát korábbi kutatások is felvetették. *Chen* és munkatársai<sup>23</sup> a szkizofrénia és a PTPN21 és a közeli EML5 gének, potenciálisan funkcionális nukleotid polimorfizmusok szignifikáns asszociációját közölték. Jóllehet az explorációs mintájuk részben átfed jelen vizsgálat replikációs mintáival, a PTPN21 SNP asszociációit metaanalízissel összevont, 13 független mintán is sikeresen demonstrálták. Bár a PTPN21 gén funkciója még nem teljesen tisztázott, a PTPN gének által kódolt protein tirozin foszfatázok a T-sejtes immunválasz, valamint sejtnövekedés fontos szabályozó faktorai. *Plani-Lam* és munkatársai<sup>24</sup> szerint a PTPN21 meghatározó a neuregulin3 (NRG3) neurotróf faktor aktiválásában.

Az MSiG v5.2 adatbázis független kanonikus

**5. táblázat.** Statisztikailag szignifikánsan replikált, funkcionális annotációs klaszterek feldúsulási pontjai (ES) és terminusaik feldúsulási  $p_{\text{-EASE}}$  értékei az explorációs és replikációs minta priorizált génei szerint

Klaszter / terminus	Explorációs minta ( $n_{\text{gén}}=386$ )		Replikációs minta ( $n_{\text{gén}}=1543$ )		Gének száma (%)
	ES $p_{\text{-EASE}}$	FDR	ES $p_{\text{-EASE}}$	FDR	
1. klaszter	ES: 2,85		ES: 23,93		
splice variant	$2,43 \cdot 10^{-4}$	0,0041	$2,57 \cdot 10^{-26}$	$4,80 \cdot 10^{-25}$	783 (50,75%)
alternative products	0,0016	0,0132	$1,53 \cdot 10^{-25}$	$2,20 \cdot 10^{-24}$	1006 (65,20%)
alternative splicing	0,0071	0,0926	$4,10 \cdot 10^{-22}$	$3,29 \cdot 10^{-21}$	755 (48,93%)
2. klaszter	ES: 1,34		ES: 4,77		
GO:0009790: embryo development	0,0068	0,1209	$1,34 \cdot 10^{-8}$	$2,67 \cdot 10^{-7}$	131 (8,49%)
GO:0048598: embryonic morphogenesis	0,0532	0,6445	$2,91 \cdot 10^{-5}$	$5,80 \cdot 10^{-4}$	77 (4,99%)
GO:0048568: embryonic organ development	0,0875	0,8230	0,0084	0,1550	51 (3,31%)
3.klaszter	ES: 1,24		ES: 4,55		
GO:0007417:central nervous system development	0,0492	0,6147	$2,33 \cdot 10^{-9}$	$4,64 \cdot 10^{-8}$	132 (8,55%)
GO:0007420:brain development	0,0630	0,7080	$6,05 \cdot 10^{-7}$	$1,21 \cdot 10^{-5}$	99 (6,41%)
GO:0060322: head development	0,0631	0,7085	$9,19 \cdot 10^{-8}$	$1,83 \cdot 10^{-6}$	106 (6,87%)
GO:0030900:forebrain development	0,0828	0,8048	$1,09 \cdot 10^{-6}$	$2,18 \cdot 10^{-5}$	61 (3,95%)

ES: enrichment score, a klaszter terminusaik median feldúsultási mutatója, ES= $-\log_{10}(p_{\text{-EASE}})$

$p_{\text{-EASE}}$ : terminus felülreprezentációját mutató, módosított Fisher p-érték

FDR: false discovery rate, Benjamini-Hochberg szerint

útvonalai és az átfedőkből összeállított „szuperútvonalak” sem az explorációs lépésben, sem replikáció során nem értek el többszörös összehasonlításra korrigált, statisztikailag szignifikáns asszociációt vagy feldúsulást, hasonlóan a Psychiatric Genomics Consortium útvonalelemzéséhez<sup>6</sup>. Hat útvonal esetében találtunk nominálisan szignifikáns replikációt, amelyek a szkizofrénia sejtosztódással, apoptózissal, neurondifferenciálódással, stresszre adott válasszal és gyulladással való kapcsolatára utalnak, amely illeszkedik a szakirodalmi adatokhoz<sup>25, 26</sup>. Mindazonáltal, a többszörös összehasonlítás korrigeálása után egyik kanonikus útvonal asszociációja sem maradt szignifikáns, ezért a plauzibilisnek tűnő asszociációk esetében is további vizsgálatok szükségesek.

A kromoszómaszegmensek által definiált pozicionális génszettek esetében viszont a 14q31 szegmens asszociációját, valamint a 5q31 és Xq13 szegmensek feldúsulását korrigeálása után is szignifikáns p-értékkel replikáltuk. A 14q31 szegmensen, a korábban tárgyalt PTPN21 gén mellett, az EML5, ZC3H14, SPATA7, TTC8 és SEL1L gének esetében mértünk nominálisan szignifikáns replikációt. Az SZGR2 adatbázis<sup>27</sup> szerint, az SEL1L gén kivételével, mindegyik gén esetében közöltek korábban szkizofréniával asszociálódó gyakori variánsokat, emellett a ZC3H14 és a SPATA7 gének metilációs mintázatukban, a TTC8 gén pedig expressziójában mutatott eltérést szkizofréniámintákban. Ez utóbbi eredmények az epigenetikai szabályozás lehetséges eltéréseire mutatnak rá szkizofréniában. Funkcióik alapján, a TTC8, EML5 és a SPATA7 a cytoskeleton megfelelő működésének fenntartásával, a TTC8 az axonogenezis, a ZC3H14 pedig mRNS-stabilizáló hatásával szükségesek az idegrendszer fejlődéséhez. A régióban található, szkizofréniában korábban implikált, STON2 és NRXN3 gének asszociációja egyik mintában sem érte el a  $p < 0,1$  küszöböt.

A 5q31 és Xq13 kromoszómarégiók esetében a feldúsulást mérő Wilcoxon p-értékük maradt a korrigeálása után is szignifikáns. A feldúsulási próbával a génszeten belüli és kívüli gének p-értékeinek eloszlását hasonlítjuk össze, a tesztstatisztika értékében nemcsak a legerősebb vagy bizonyos küszöbértéket meghaladó asszociációk, hanem az összes asszociáció trendszerű eltérése is meghatározó lehet. Ezért az asszociációs próbával szemben, a legjobb gének kiemelt tárgyalása helyett, a teljes génszett korrelációját vizsgáltuk az MSiGDB v5.2 géntológiai (GO) annotációival. Az 5q31 szegmens génei a sejtadhézió, szinapsziszerveződés, szinaptikus jelátadás, kalciumkötés és kalciumfüggő sejtadhézió GO terminusaival korreláltak

FDR < 0,05 korrigeálással, összhangban a PGC szkizofrénia megalanálízis eredményeivel<sup>3</sup>. Az Xq13 sáv génszetteje pedig az RNS-polimeráz II transzkripció regulációja és acetyl-transferáz funkció terminusaival korrelált.

A funkcionális annotációk elemzése előtt, a KGGseq program gépi tanuláson és mélytanuláson alapuló predikcióit és az EIGEN adatbázis annotációját kombinálva, a funkcionálisnak jósolt kódoló (mendeli predikció<sup>28</sup>) vagy nem kódoló (regulátoros predikció<sup>29</sup>) variánsok alapján prioritizáltuk a minták genomjait. A prioritizált génlistákhoz kapcsolódó, funkcionális annotációk klaszterelemzésekor, az explorációs minta nominális feldúsulást mutató klaszterei közül, három esetében regisztráltunk többszörös összehasonlítás korrigeálójának is megfelelő feldúsulási pontot: a splicing/alternatív splicing, az embrionális fejlődés és az ideg/agyfejlődés annotációs klaszternél. Mindhárom klaszter feldúsulása illeszkedik a szkizofrénia neuronális fejlődéssel kapcsolatos betegségmodelljébe. A splicing/alternatív splicing a transzkripció kromatinstátusztól független, szabályozó mechanizmusa, melynek segítségével sejt-típus-specifikus mRNS-transzkriptvariánsok jöhetnek létre. Ez a szabályozási lépés a neuronális fejlődés mellett a gyulladáscsökkentő citokinek termelődését is befolyásolja, tehát a szkizofrénia neuroinflammációs etiológiáját is erősítheti. Ugyanakkor, nem zárhatjuk ki, hogy a magas ES-pont ellenére a klaszter csak műtermék, a prioritizációs algoritmusok a regulátoros helyek felismeréséhez a splicing helyek gépi tanuláson alapuló annotációját is tartalmazták, ami vezethetett a splicing/alternatív splicing feldúsulásához. Végezetül megemlítendő, hogy a szkizofrénia etiológiájában jelen ismereteink szerint központi szerepet játszó neurotransmisszió, szinaptikus funkciók és plaszticitás zavarára utaló klaszterek csak az explorációs mintából hiányoztak, a replikációs minta 126 nominális klasztere között jelen voltak, amely főként a két minta közötti metodológiai és méretbeli különbségeknek tudható be.

Vizsgálatunk másik fő célja a multilókuszelemzések replikációs valószínűségekre gyakorolt hatásának felmérése volt, amelyhez mind a nominális, mind a többszörös összehasonlításra korrigált replikációk arányát regisztráltuk. A p-értékek korrigeálójától függetlenül, a multilókuszelemzések egyértelműen növelik a replikáció valószínűségét a definiált „marker” komplexitásának megfelelően. A replikáció aránya valószínűleg a genetikai heterogenitás csökkenése miatt növekszik, de nem vethetjük el a lehetőségét, hogy e multilókuszmerek egyszerűen csak több gént annotálnak méretüknél fogva.

Vizsgálatunk nem mentes a korlátoktól sem. A SCHIZOBANK túlságosan kis elemszámú, és alacsony lefedettségű mintája explorációs célokra sem ideális. A statisztikai erő növelése érdekében az exploráció során szuggesztív p-értékű ( $\alpha=0,1$ ) asszociációkat is elfogadtunk, a fals pozitív asszociációk ellensúlyozására a második lépcsőben konzervatív p-érték-korrekciót végeztünk. A közös variánsok gyarapítása érdekében a 1000 Genomes ph1v3 mintája szerint imputáltuk az GWAS-mintát. Az imputáció az átfedést javította, azonban részben torzítja is az adatot: a hosszabb gének esetében több marker akad fenn a minőségkontroll-szűrőn, ezért a rövidebb gének viszonylagosan lefedettségét megnöveli. A rövidebb gének feldúsulása torzíthatja a kompetitív génszett-teszteket, például a Wilcoxon-tesztet és a felülreprezentációs tesztet. E torzítást mérséklendő, a génszettek vizsgálatakor a kompetitív feldúsulási tesztek mellett régióalapú, HYST génszett-asszociációs tesztet is végeztünk.

Milyen gyakorlati, a klinikai munkában is hasznosítható következtetéseket fogalmazhatunk meg az eredmények alapján? A fenti összefüggések arra utalnak, hogy a szkizofrénia genetikai „architektúrájában”, felépítésében fontosak azok a gyakori, kis hatású variánsok, melyek biológiailag plauzibilis, de nem teljes mértékben átlátott és megértett agyfejlődési és immunológiai folyamatokra hatnak. Ezeket a variánsokat szüleik átadják gyerekeiknek, emiatt lesz a szkizofrén beteg elsőfokú rokonai, gyermekei között magasabb a betegség előfordulása. Fontos ugyanakkor hangsúlyozni, hogy a gyakori variánsok csak a vulnerabilitás kialakulásáért felelősek, a betegség manifesztálódásában rendkívül fontosak a környezeti rizikófaktorok, melyeknek egy része megelőzhető. Hasonlóan fontos klinikai tudás, hogy a bipoláris affektív zavar és a szkizofrénia gyakori variánsai átfedést mutatnak, tehát működik egyfajta „keresztöröklődés” a két betegség között. Szintén van átfedés a szkizofrénia és az autizmus genetikai hátterében, ugyanakkor ezt a ritka genetikai variánsok okozzák, melyeket ebben az elemzésben nem vizsgáltunk<sup>30</sup>.

Összefoglalva vizsgálatunk eredményeit, a biológiai és molekuláris funkciók, jellemzők alapján definiált, multilókusz genetikai elemzések a szkizofrénia etiológiájában implikált szinaptikus plaszticitás, neuron-, axon- és agyfejlődés, valamint immunológiai és transzkripciószabályozó annotációknak megfelelő gének, génszettek asszociációját, illetve feldúsulását eredményezték. Emellett a replikáció valószínűségét – komplexitásukkal arányosan – egyértelműen elősegítették.

**KÖSZÖNETNYILVÁNÍTÁS**  
Köszönetet mondunk a magyar SCHIZOBANK Konzorcium tagjainak és a betegbevonásban, valamint a korábbi elemzésekben részt vett munkatársainak: Álmos Péternek, Balicza Péternek, Bálint L. Bálintnak, Balogh Nóráknak, Barta Endrének, Benkovits Juditnak, Bitter Istvánnak, Égerházy Anikónak, Hársfalvi Viviennek, Horváth Attilának, Inczédy-Farkas Gabriellának, Janka Zoltánnak, Kassai-Farkas Ákosnak, Likó Istvánnak, Magyarósi Szilviának, Makkos Zoltánnak, Nagy Krisztiánnak, Nagy Tibornak, Németh Györgynek, Scholtz Beátának, Török Zsoltnak, Tűzkő Juditnak, Urbán Zoltánnak és Zahuczky Gábornak. Köszönet illeti a kutatásban részt vevő betegeket és családtagjaikat, valamint a mintagyűjtésben részt vevő szakápolókat is. Kutatásunkat az Országos Tudományos Kutatási Alap és a Nemzeti Agykutatási Program támogatták (OTKA-PD 83876 és NAP-B KTIA\_NAP\_13-2014-0011 dr. Réthelyi János számára).

## IRODALOM

1. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a Complex Trait. *Arch Gen Psychiatry* 2003;60:1187. <http://dx.doi.org/doi:10.1001/archpsyc.60.12.1187>
2. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 2014;55:1068-87. <http://dx.doi.org/doi:10.1111/jcpp.12295>
3. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511:421-7. <http://dx.doi.org/doi:10.1038/nature13595>
4. Mooney MA, Nigg JT, McWeeney SK, Wilmot B. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet* 2014;30:390-400. <http://dx.doi.org/doi:10.1016/j.tig.2014.07.004>
5. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1-13. <http://dx.doi.org/doi:10.1093/nar/gkn923>
6. Sullivan PF, Posthuma D. Biological pathways and networks implicated in psychiatric disorders. *Curr Opin Behav Sci* 2015;2:58-68. <http://dx.doi.org/doi:10.1016/j.cobeha.2014.09.003>
7. Inczédy-Farkas G, Benkovits J, Balogh N, Álmos P, Scholtz B, et al. SCHIZOBANK – The Hungarian national schizophrenia biobank and its role in schizophrenia research. *Orv Hetil* 2010;151:1403-8. <http://dx.doi.org/doi:10.1556/OH.2010.28943>
8. Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M, et al. New models of collaboration in



- genome-wide association studies: the Genetic Association Information Network. *Nat Genet* 2007;39:1045-51. <http://dx.doi.org/doi:10.1038/ng2127>
9. Suarez BK, Duan J, Sanders AR, Hinrichs AL, Jin CH, Hou C, et al. Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. *Am J Hum Genet* 2006;78:315-33. <http://dx.doi.org/doi:10.1086/500272>
  10. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. Schork NJ, editor. *PLoS Genet* 2009;5(6):e1000529. <https://doi.org/10.1371/journal.pgen.1000529>
  11. McVean GA, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491(7422):56–65. <https://doi.org/10.1038/nature11632>
  12. Li M-X, Gui H-S, Kwan JSH, Bao S-Y, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res* 2012;40:e53. <http://dx.doi.org/doi:10.1093/nar/gkr1257>
  13. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 2000;19(Suppl 1):S36-42. [http://dx.doi.org/doi:10.1002/1098-2272\(2000\)19:1+::AID-GEPI16>3.0.CO;2-M](http://dx.doi.org/doi:10.1002/1098-2272(2000)19:1+::AID-GEPI16>3.0.CO;2-M)
  14. Abraham G, Inouye M, Zhang Y. Fast principal component analysis of large-scale genome-wide data. *PLoS One* 2014;9. <http://dx.doi.org/10.1371/journal.pone.0093766>
  15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75. <http://dx.doi.org/doi:10.1086/519795>
  16. Li M-X, Sham PC, Cherny SS, Song Y-Q. A knowledge-based weighting framework to boost the power of genome-wide association studies. *PLoS One* 2010;5:e14480. <http://dx.doi.org/doi:10.1371/journal.pone.0014480>
  17. Li M-X, Gui H-S, Kwan JSH, Sham PC. GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure. *Am J Hum Genet* 2011;88:283-93. <http://dx.doi.org/doi:10.1016/j.ajhg.2011.01.019>
  18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50. <http://dx.doi.org/doi:10.1073/pnas.0506580102>
  19. Vivar JC, Pemu P, McPherson R, Ghosh S. Redundancy control in pathway databases (ReCiPa): an application for improving gene-set enrichment analysis in Omics studies and “Big data” biology. *OMICS* 2013;17:414-22. <http://dx.doi.org/doi:10.1089/omi.2012.0083>
  20. Li MX, Kwan JSH, Sham PC. HYST: A hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. *Am J Hum Genet* 2012;91:478-88. <http://dx.doi.org/doi:10.1016/j.ajhg.2012.08.004>
  21. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44-57. <http://dx.doi.org/doi:10.1038/nprot.2008.211>
  22. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat Genet* 2016;48:214-20. <http://dx.doi.org/doi:10.1038/ng.3477>
  23. Chen J, Lee G, Fanous AH, Zhao Z, Jia P, O'Neill A, et al. International Schizophrenia Consortium, Two non-synonymous markers in PTPN21, identified by genome-wide association study data-mining and replication, are associated with schizophrenia. *Schizophr Res* 2011;131:43-51. <http://dx.doi.org/10.1016/j.schres.2011.06.023>
  24. Plani-Lam JH-C, Chow T-C, Siu K-L, Chau WH, Ng M-HJ, Baoet S, et al. PTPN21 exerts pro-neuronal survival and neuritic elongation via ErbB4/NRG3 signaling. *Int J Biochem Cell Biol* 2015;61:53-62. <http://dx.doi.org/doi:10.1016/j.biocel.2015.02.003>
  25. Hess JL, Tylee DS, Barve R, de Jong S, Ophoff RA, Kumasingshe N, et al. Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia. *Schizophr Res* 2016;176:114-24. <http://dx.doi.org/doi:10.1016/j.schres.2016.07.006>
  26. Darby MM, Yolken RH, Sabunciyan S. Consistently altered expression of gene sets in postmortem brains of individuals with major psychiatric disorders. *Transl Psychiatry* 2016; 6:e890. <http://dx.doi.org/doi:10.1038/tp.2016.173>
  27. Jia P, Han G, Zhao J, Lu P, Zhao Z. SZGR 2.0: a one-stop shop of schizophrenia candidate genes. *Nucleic Acids Res* 2016;gkw902. <http://dx.doi.org/doi:10.1093/nar/gkw902>
  28. Li M-X, Kwan JSH, Bao S-Y, Yang W, Ho S-L, Song Y-Q, et al. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet* 2013;9:e1003143. <http://dx.doi.org/doi:10.1371/journal.pgen.1003143>
  29. Li MJ, Pan Z, Liu Z, Wu J, Wang P, Zhu Y, et al. Predicting regulatory variants with composite statistic. *Bioinformatics* 2016;32:2729-36. <http://dx.doi.org/doi:10.1093/bioinformatics/btw288>
  30. Réthelyi JM, Benkovits J, Bitter I. Genes and environments in schizophrenia: The different pieces of a manifold puzzle. *Neurosci Biobehav Rev* 2013;37:2424-37. <http://dx.doi.org/10.1016/j.neubiorev.2013.04.010>