

Hatóanyagok és terápiás célpontok vizsgálata adatmérnöki eszközökkel

Doktori tézisek

Temesi Gergely Botond

Semmelweis Egyetem
Molekuláris orvostudományok Doktori Iskola



Témavezető:

Dr. Szalai Csaba, az MTA doktora, egyetemi tanár

Hivatalos bírálók:

Dr. Gulyás-Kovács Attila, Ph.D.

Dr. Rónai Zsolt, Ph.D., egyetemi adjunktus

Szigorlati bizottság elnöke:

Dr. Vásárhelyi Barna, az MTA doktora, egyetemi tanár

Szigorlati bizottság tagjai:

Dr. Cserző Miklós, Ph.D., tudományos főmunkatárs

Dr. Szakács Orsolya, Ph.D.

Budapest
2014

Bevezetés

Az új évezred valóban egy új korszakot is nyitott számos tudományos, technológiai és gazdasági területen. A nyugati társadalmak öregedésének egészségügyi kiadásai évtizedek óta egyre növekvő gazdasági nyomást okoznak. A gyógyszeripar a szigorodó szabályzások hatására az utóbbi évtizedben évről-évre egyre drasztikusabb költségnövekedésről számol be, miközben az éves szinten engedélyezett hatóanyagok száma az elmúlt húsz év alatt csak csökkent. Ezek a folyamatok nemzetgazdasági és iparági szinten is fenntarthatatlanok. Mindeközben az élettudományok, és elsősorban a molekuláris élettudományok az elmúlt 50 évben hatalmas fejlődésen mentek keresztül. Bár a genomika eredményeinek translációja a klinikum szintjére a vártnál lassabban halad, a nagy áteresztőképességű műszeres módszerek fejlődése soha nem látott mennyiségű adatot szolgáltat a biológiai rendszerekről. A számítástechnika elmúlt ötven éves fejlődése még talán az élettudományoknál is nagyobb jelentőségű volt, a világon elérhető digitális adat- és információtömeg a hatvanas évek óta minden területen exponenciálisan növekszik („Big Data”).

Az egészségügy és gyógyszeripar fenntarthatatlan modelljeiből való kitörés egyik legfontosabb eszköze lehet az információ technológiák fokozott kiaknázása. Ezeket a trendeket felismerve, korábbi Mérnök Informatikus MSc és Orvosbiológiai Mérnök MSc tanulmányaimra építve PhD kutatásaim során olyan információ technológiai eszközök kutatását és fejlesztését tűztem ki célul, amelyek a világ küszöbön álló egészségügyi és gyógyszeripari változásaira reagálnak, illetve képesek segíteni azt.

Az első kutatási témában („Asztma genetika”) az asztma és gyulladásos megbetegedések molekuláris patomechanizmusát, elsősorban genetikai hátterét vizsgáltam. Ebben az esetben egy teljes orvosbiológiai felfedező kutatás kivitelezése volt célom: kezdve a korábbi állatmodell kutatások alapján a kísérlet megtervezésétől, a humán biobank gyűjtésen és a több szintű omikai vizsgálatok elvégzésén át, modern információ technológiai módszerekkel történő kiértékelésig. Ebben az esetben az információ technológiai módszerek eszközként jelentek meg a kísérlet megtervezésében és a tárgyterület adatainak rendszerszintű modellezésében, elsősorban korábbi fejlesztési eredmények felhasználásával.

A második téma („Feldúsulás elemzés”) fókuszusa ezzel szemben egy új információ technológiai módszertan fejlesztése volt farmakológiai területen. Ebben az esetben is egy komplex tárgyterület adatainak integrálása és elemzése volt a cél, de itt az új módszertan fejlesztésén volt a hangsúly, amelyet egy tárgyterületi problémán teszteltem is. A technikával ismert hatóanyagok heterogén tulajdonságai és mérési eredményei alapján előre lehet jelezni egy új vegyület különböző tulajdonságait, így például lehetséges indikációkat. A módszer teszteléséhez az amantadinet

vizsgáltam meg, mely egy jól ismert, több indikációval rendelkező hatóanyag, így jól tudtuk ellenőrizni a módszer hatékonyságát és korlátait.

Asztma genetika

Az asztma a légutak krónikus, összetett multifaktoriális betegsége, genetikai és környezeti hatások együtt alakíthatják ki. Az allergia és az asztma együtt több embert érint, mint az összes többi krónikus betegség összesen. A betegség tünetei vagy a kezelésekre adott válaszok nagyon heterogének, számos endofenotípust különböztetnek meg. Az utóbbi évtizedek rohamos molekuláris biológiai fejlődése mellett számos asszociált gént és patomechanizmust vizsgáltak, de az egyre növekvő ismeretanyag mellett sem tisztázottak a pontos kiváltó okok. Eddig több mint 120 gént hoztak összefüggésbe asztmával illetve atópiával, de csak kb. 10 olyan gén van, amelyet tíznél is több vizsgálat tudott igazolni. A multifaktoriális jellegek és betegségek esetén többnyire hasonló a helyzet, a sikertelenségre számos hipotézis és ezzel együtt új kutatási stratégia született („hiányzó örökletesség”, „genetika sötét anyaga”).

A molekuláris orvostudományokban az 1980-as évek végéig a technológiai lehetőségekből adódóan szinte kizárólag hipotézis-vezérelt felfedező módszereket alkalmaztak: a korszakot elsősorban a kis adatmennyiséget szolgáltató, drága mérések, és hipotézis-tesztelésen alapuló, úgy nevezett frekventista statisztikai kiértékelés jellemezte. A 1990-as évek elejétől a műszertechnikai (és információ technológiai) fejlődéssel a nagy áteresztőképességű technikákkal teljes omikai szintek (pl. az összes gén expressziója) váltak egyszerre mérhetővé, ezzel a hangsúly áthelyeződött a hipotézismentes tervezésre és vizsgálatokra. Az új módszerek által szolgáltatott adatok dimenziója a hagyományos statisztikai módszerekkel igen nehezen kezelhető, ezért érdemes egy alternatív normatív keretrendszert, a Bayesi következtetést bevezetni. Kutatócsoportunk részletesen tanulmányozta, kidolgozta és számos helyen sikerrel alkalmazta a Bayes-háló alapú Bayesi többszintű relevancia (BN-BMLA) elemzés módszertanát. Segítségével lehetséges az egyszerű statisztikai asszociáció fogalmának specifikusabb karakterizálása, a gyengébb hatások kimutatása, és így a multifaktoriális betegségek pontosabb megismerése.

Feldúsulás elemzés

Sok biztató jel utal arra, hogy a gyógyszeriparnak sikerül elkerülnie a fenntarthatatlan üzleti modell katasztrófáját („patent cliff disaster”), de az ipar továbbra is intenzíven keresi a hatékonyságnövelő stratégiákat. Az egyik ilyen innovatív ipari törekvés azt célozza meg, hogy különböző módszerekkel hatóanyagok indikációit vagy alkalmazásait kiterjessze, újrapozicionálja. Egy újrapozicionált hatóanyag piacra juttatása az originális fejlesztéshez képest jelentősen gyorsabb

és olcsóbb, ráadásul a gyógyszeripari cégek polcain heverő bukott hatóanyagok számát 2000 és 30 000 közé teszik. A hatóanyagok újrapozicionálását két tudományosan jól leírt koncepció is lehetővé teszi. Az első szerint a legtöbb hatóanyag több molekuláris célpontra hat, így gyakran érdemes a másodlagos célpontokat keresni („off-target repositioning”). A másik koncepció szerint a legtöbb molekuláris célpont több biológiai szabályozási mechanizmusban is részt vesz, így a célpont eltérő szabályozási útvonalaira fókuszálva új indikációk is felvetődhetnek („on-target repositioning”).

A kémiai és vegyészeti számítógépes technikák fejlődésében a számítógépes biológiához hasonló folyamatokat figyelhetünk meg. A korai szimuláció-intenzív számítógépes kémia („computational chemistry”) mellett a kilencvenes évek végétől a nagy áteresztő képességű vizsgáló módszerek terjedésével egyre nagyobb szerep jutott az adatoknak és az adatintegrációs technikáknak („chemoinformatika”). Az ismert, hasonló molekulák alapján jósolható a biológiai aktivitás, és sok esetben pl. a potenciális célpontok, a mellékhatások, és az indikációk is. A vegyületek hasonlóság alapú keresése és sorrendezése egy régóta tanulmányozott probléma, melynek a számítógépes hatóanyag újrapozicionálás területén is léteznek alkalmazásai. A különböző módszereknek megvannak a maguk erősségei, jelenleg a legfontosabb kérdés a heterogén eredmények integrálása és az új technikák megfelelő beillesztése a klasszikus farmakológiai szaktudás mellé.

Célkitűzés

Az asztma kutatásaim célja elsősorban új, asztmára hajlamosító gének, esetleges terápiás célpontok azonosítása volt kandidáns gén asszociációs vizsgálatok tervezésével egy korábbi, ovalbumin indukált egér asztmamodellen végzett teljes genom expressziós tanulmány alapján. A kutatásaim ezen belül három kérdésre fókuszáltak:

1. A genetikai asszociációs vizsgálatok tervezése, különösképpen a mérendő genetikai polimorfizmusok mérés technikailag és információelméletileg optimális halmazának kiválasztása hogyan automatizálható számítógépes eszközökkel?
2. Egy modern eszközökkel megtervezett, többszintű validációra építő, asztmával kapcsolatos kutatás során milyen új genetikai variánsokat, oki tényezőket, terápiás célpontokat lehet azonosítani?
3. A hagyományos frekventista statisztika eszköztára mellett egy rendszerszintű modellezés, a Bayes-háló alapú Bayesi több szintű relevancia elemzés eszköztára mennyiben segítheti az oki tényezők feltárását?

A farmakológiában felhasznált vegyületekről hatalmas mennyiségű heterogén adat- és információtömeg érhető el. A tanulmányaim során azt vizsgáltam meg, hogy egy molekuláris biológiai adathalmazok elemzésében bevált matematikai módszert, a feldúsulás elemzést, milyen módon lehet átültetni farmakológiai területre, és az alábbi két kérdésre kerestem a választ.

1. A feldúsulás elemzés módszertana hogyan alkalmazható a farmakológiai információtömeg integrálására, a gyenge jelzések elemzésére, és melyek a felhasználás korlátai?
2. Milyen funkciókkal segítheti ez a módszertan egy adott hatóanyag fejlesztését, és hogyan ágyazhatóak be ezek a technikák a gyógyszerfejlesztési gyakorlatba?

Módszerek

Asztma genetika

Kutatásaimat kutatócsoportunk egy korábbi, ovalbumin-indukált egér asztmamodell vizsgálatának eredményeire alapoztam. A tanulmány génexpressziós eredményeit használtuk fel 60 olyan kandidáns gén kiválasztására, melyek ortológiai kiinduló pontjai lehetnek egy humán asztma és atópia tanulmánynak.

A kandidáns gének polimorfizmusainak nagy áteresztőképességű mérése több tanulmányunkban is primer extenziós technikával történt. A műveletnek mérés technikai szempontból több korlátozó tényezője van, amely bizonyos polimorfizmusok mérését kizárja (pl. a PCR primereken a rövid ismétlődő szakaszok beépülés tévesztéshez vezethetnek), vagy közös reakció térben történő mérését nem teszi lehetővé (pl. az összes primer olvadáspontjának lehetőleg egyeznie kell). Ezek az egymásnak kissé ellentmondó követelmények egy-egy multiplex mérésbe bevonható polimorfizmusok halmazának egyedi megtervezését igénylik. Ezért létrehoztam és leírtam egy olyan szoftver rendszert és módszertant (TIGER kísérlettervező rendszer), amely a kísérlettervezési folyamathoz szükséges adatbázisokat integrálja és a legtöbb lépését automatizálja.

A kísérlettervezés során 60 kiválasztott génben 90 mérendő polimorfizmust azonosítottunk funkcionális, kapcsoltsági és mérés technikai szempontok alapján. A genotipizálási tanulmányba bevont magyar (kaukázusi) populáció 671 rokonsági viszonyban nem álló személyből állt (311 asztmás gyermek és 360 egészséges kontroll), akikről komplex fenotípusos adatot is gyűjtöttünk. Végül a genotipizálás eredményei alapján indukált köpet expressziós vizsgálatot végeztünk, melybe összesen 34 felnőtt személyt vontunk be.

A genotipizálás eredményein a Hardy-Weinberg egyensúlyt (HWE) χ^2 négyzet próbával ellenőriztem, a hiányzó genotípus adatokat a polimorfizmusok egyváltozós eloszlásából véletlenszerűen mintavételezve pótoltam. A HaploView és IBM SPSS Statistics V20 szoftvereket használtam a Pearson-féle χ^2 négyzet próbák elvégzéséhez, hogy az asztmához való asszociációt és az esélyhányadosokat meghatározzam, mind allélikus, mind domináns-recesszív modellben, illetve haplotípus szinten is. A többváltozós logisztikus regressziós elemzéseket szintén IBM SPSS Statistics V20 szoftverrel végeztem, a kor és a nem minden esetben a modellbe került, hogy elégséges statisztikai adjusztálást végezzünk az esetek és kontrollok közötti eltérés miatt. A többszörös hipotézisvizsgálat hatását permutációs teszttel és Bonferroni korrekcióval is elvégeztem.

A genotipizálási adatok elemzését rendszerszintű modellezéssel is elvégeztem, ehhez a kutatócsoportunk által kifejlesztett Bayes-háló alapú Bayesi többszintű relevancia elemzést (BN-BMLA) használtam. Az adat oksági összefüggéseit legjobban leképező Bayes-háló gráfokat összegezve a BN-BMLA módszerrel meghatározható egy-egy adott gráf jegy poszteriori valószínűsége pl. direkt, áttételes, vagy egyéb zavaró hatás okozza-e a statisztikai asszociációt a betegséggel. A nagy számítási kapacitást igénylő rendszer szintű elemzéseket a GenaGrid Konzorcium 512 processzoros SGI Altix ICE szuperszámítógépén végeztük, az eredmények megjelenítéséhez a BayesEye kliens programot használtam.

Az indukált köpet expressziós vizsgálatot real time quantitative PCR méréssel végeztük, normalizálásához a delta-delta-CT algoritmust használtam (a housekeeping génhez és a kontroll mintákhoz normalizálva). A normalizált transzkriptom változások statisztikai értékelését Student-féle t-próbával végeztem.

Feldúsulás elemzés

A számítógépes hatóanyag újrapozicionálás során a legegyszerűbb esetben egy hatóanyaghoz keresünk új indikációkat, így a rendszer bemenete általában egy vizsgált hatóanyag. A kutatócsoportunk által kifejlesztett QDF² módszertan az ismert vegyületekről elérhető adattömege épít, amely lehet zajos vagy hiányos, az adatok széles skáláját képes feldolgozni a fizikai, kémiai leíróktól, tetszőleges farmakológiai tesztek eredményén át bármilyen kvantitatív klinikai jellemzőig. A QDF² módszertan a vizsgált hatóanyaghoz való hasonlóságok szerint egy kernel fúziós módszerrel sorrendezi a referencia adatbázisban szereplő vegyületeket. Az eredmény farmakológiai értelmezése a hatóanyag tulajdonságaival kapcsolatosan új hipotéziseket vethet fel (pl. új indikáció, célpont, anyagsere útvonal, stb.). A sorrend vizsgálatára léteznek különböző módszerek (szűrések, hálózatelemzés, stb.), de ezek már komoly háttértudást igényelnek, és sok esetben

nehezen összeegyeztethetőek a klasszikus farmakológiai háttértudással. Saját munkáim során erre a problémára kísértem meg egy új eszközt kidolgozni.

A géncsoport feldúsulás elemzés (Gene Set Enrichment Analysis, GSEA) génexpressziós mintázatok elemzésében és értelmezésében már jól bevált technika. A módszertant alapvetően nagyszámú entitásból álló sorrendezett listák absztrakt fogalmi szinteken történő elemzésére és értelmezésére fejlesztették ki, így a módszertan a megfelelő módosításokkal alkalmas vegyület-adatbázisok elemzésére is. Munkám során a feldúsulás elemzésnek a génsorrendek helyett vegyületsorrendek elemzésére módosított változatát fejlesztettem ki (Compound Set Enrichment Analysis, CSEA). Ez egy olyan újszerű alkalmazása a bevált feldúsulás elemzési technikának ezen a területen, amellyel például a következő kijelentést tehetnénk: „ha az összes engedélyezett hatóanyagot a vizsgált vegyülethez való hasonlóság szerint sorrendezzük, akkor az ismert dopaminerg agonisták többnyire előre rangsorolódnak, tehát valószínűsíthető a hatóanyag dopaminerg hatása”.

Az elemzésekhez ismert hatóanyagokból egy referencia adatbázist építettünk. A vegyületek kémiai profiljához három (Molconn-Z, MACCS kulcsok, 3D Pharmacophore) leírót használtunk, mindhárom vektort a Schrödinger 2012 Suite szoftver csomag segítségével állítottuk elő. A vegyületek fehérje célpont profiljait a DrugBank adatbázis segítségével állítottuk elő. Az egységes mellékhatás profilokat a DailyMed adatbázisban található gyógyszer betegtájékoztatók szövegbányászatával nyertük ki. A CMAP expressziós adatbázist használtuk fel a hatóanyagokhoz tartozó expressziós profilok előállításához. Ezzel létrehoztunk egy 1730 hatóanyagból álló referencia adatbázist, ez az FDA által engedélyezett hatóanyagok közel háromnegyede.

Eredmények

Asztma genetika

Egy vizsgálatra alkalmas, optimális polimorfizmus készlet kiválasztása manuálisan végrehajtva akár több hónapos, meglehetősen körülményes munka. Az általam kifejlesztett TIGER kísérlettervező szoftver rendszer segítségével a felhasználók súlyozott polimorfizmus szűrési feltételeket definiálhatnak (pl. genom pozíció, gén, génen belüli funkcionális régió, polimorfizmus típus, MAF); a rendszer elvégzi a mérés technikai megfelelés ellenőrzését; esetleges helyettesítő polimorfizmusokat javasol és támogatja a mérés technikai szempontból optimális halmaz kiválasztását. A TIGER kísérlettervező rendszer segítségével több tanulmányunkban is sikerült az előkészítő munka időtartamát közel egy nagyságrenddel leshorítani, miközben a hibázási lehetőségek is lényegesen csökkentek.

Az asztma tanulmányunkban a 90 mért polimorfizmus közül több esetén tapasztaltunk eltérő genotípus eloszlást a beteg és az egészséges populáció között. Ezek közül két génben 4 polimorfizmus asszociációja minden statisztikai teszttel erősen szignifikáns volt (a két gén legerősebb polimorfizmusai: *SCIN* gén rs2240572, $p=0,00007$, $OR=0,637$; *PPARGC1B* gén rs32588, $p=0,00012$, $OR=0,563$). Hat másik gén asszociációja mutatott határeset közeli szignifikanciát klasszikus frekventista statisztikai módszerekkel (*ITLNI*, *FABP3*, *MAT1A*, *OSGIN*, *LY9*, *LGMN*). További vizsgálatokkal sikerült kimutatni, hogy a *SCIN* és a *PPARGC1B* gének védő hatása nők között lényegesen erősebb; míg az *ITLNI* gén az allergiás csoportban, az *LGMN* gén pedig domináns-recesszív modellben vizsgálva szignifikáns asszociációt mutat.

A genotipizálási eredményeket tovább elemeztem Bayes-háló alapú Bayesi többszintű relevancia elemzéssel (BN-BMLA), hogy részletesebb képet kapjak a genetikai polimorfizmusok és fenotípus jegyek összefüggéseiről. A BN-BMLA a frekventista elemzéshez hasonlóan egyértelműen azonosította a *SCIN* és *PPARGC1B* gén hatását. A *SCIN* gén esetén erős interakciót is azonosított a *TFF1* génnel, a *PPARGC1B* esetén gyenge kölcsönhatást a terheléses asztmával. Ugyanakkor a módszer az *ITLNI* gén hatását csak infekciós asztma esetén találta kiemelkedőnek, míg a frekventista módszerekkel alig kimutatható *LGMN* hatását minden elemzésben kiemelkedőnek találta, különösen allergiával együtt vizsgálva. Az intrinsic asztma endotípust vizsgálva egyértelműen kirajzolódott annak minden más asztma endotípustól eltérő genetikai háttere, így feltételezhető a merőben eltérő patomechanizmus is. Továbbá, a BN-BMLA elemzés azt is kimutatta, hogy az intrinsic asztma esetén a genetikai háttér szerepe jelentősen erősebb a kor és a nem hatásánál; a többi asztma (és főleg az allergiás) endotípus esetén ez utóbbiaknak is erős a szerepe a genetikai háttérrel együtt; míg az asztma súlyossága (GINA) és a vizsgált polimorfizmusok között nem találtam közvetlen összefüggést.

Három gén (*SCIN*, *PPARGC1B*, *ITLNI*) expresszióját megvizsgáltuk asztmások és kontroll személyek indukált köpet mintáiban is. Mindhárom gén expressziója szignifikánsan alacsonyabb volt asztmásokban. Ezen eredmények némileg váratlanok voltak, mivel a korábbi egér asztmamodell esetén mindhárom gén expressziója ellentétes irányú összefüggést mutatott. Az eredményeink ellenőrzéséhez megvizsgáltam az NCBI GEO-ban publikusan elérhető asztmás microarray tanulmányok adatait, ezek megerősítették az eredményeinket. Megvizsgáltam a szakirodalmat, példákat keresve humán-egér ellentétes expressziós eredményekre, és a jelenség ismertnek bizonyult, számos tanulmány foglalkozott szisztematikusan is a kérdéssel. Több lehetséges magyarázata is van a jelenségnek. Az első, hogy az egérmodell esetén a teljes tüdőt vizsgáltuk, míg az emberek esetén a köpetet, így a sejtes összetételük eltérése miatt a génexpressziók átlaga különbözhet. Másrészt ismert, hogy a humán asztma egy krónikus betegség,

míg az allergén indukált légúti gyulladás egérben egy akut jelenség, és ennek a folyamatnak a dinamikus expressziós változásait mértük. Bármi is a jelenség valós háttere a gének eltérő expressziója azt jelezheti, hogy azok érintettek a folyamatban.

Feldúsulás elemzés

A CSEA módszertan működésének vizsgálatához létrehoztam egy olyan szoftveres tesztrendszert, amely képes egy új hatóanyagot ismert hatóanyagok adatai alapján farmakológiai fogalmakkal jellemezni, így pl. akár potenciális indikációkat vagy mellékhatásokat jósolni.

A GSEA technikában eredetileg alkalmazott funkcionális gén halmazok helyett az Anatomical Therapeutic Chemical (ATC) Classification System, Level 4 vegyület halmazait használtam fel a feldúsulási jelek definiálására (egy-egy indikációhoz tartozó hatóanyag halmazok). A hatóanyagok hasonlóságának számításához egyszerű cosinus függvényt használtunk a mellékhatások és a célpont leírók esetén, illetve Tanimoto távolságot a kémiai leírók esetén. A CMAP expressziós profilok hasonlóságainak számításához az eszköz alapértelmezett megoldását használtam.

A CSEA módszer tesztjei során kiválasztottam egy-egy vegyületet, majd a különböző metrikák alapján számított hasonlóságok szerint külön-külön sorba rendeztem a referencia adatbázis összes hatóanyagát (kémiai leírók, célpont, expressziós profil, mellékhatás). A következő lépésben sorrendi fúziót végeztem a SumScore algoritmussal, amely sorrendek összefésülésére egy egyszerű, matematikailag torzításmentes normatív megoldásnak tekinthető. Végül azt vizsgáltam meg, hogy az előre definiált hatóanyag halmazok elemeinek megjelenése a rendezett listában véletlenszerű-e. A halmazok (vagyis indikációk) feldúsulási értékeinek számításához a SaddleSum algoritmust használtuk, mely a p-értékeket egy speciális közelítő eljárással (Lugananni-Rice formulával) határozza meg.

A teljes folyamat demonstrálásához egy gyógyszerfejlesztés lépéseit szimuláltam, amikor is először a vizsgált vegyületnek csak a kémiai profilja elérhető, majd később egyre több mérési eredmény. Kiválasztottam az amantadinet, amely egy ismert újrapozicionált vegyület, eredetileg influenza kezelésére fejlesztették, később megkapta az engedélyt Parkinson-kór kezelésére is. Először sorrendeztem a referencia adatbázisunk hatóanyagait az egyes adatforrások szerint az amantadinehoz való hasonlóság alapján. Majd minden lépésben egyre több információforrást felhasználva fuzionáltam sorrendeket, és a megfelelő információ források együttes hatását reprezentáló sorrendeket vizsgáltam feldúsulás elemzéssel. A dopaminerg agonisták, mint a Parkinson-kór kezelésének egyik legfontosabb eszközei, konzisztensen minden lépésben előre rangsorolódtak, és az információ források bővülésével egyre inkább megerősítést nyertek. A sorrendi fúzió és a CSEA együttes alkalmazása jól kezeli a zajos adatokból eredő problémákat, de

egy fontos zavaró hatás is adódik a módszer hasonlóságokra építő működéséből: az eredmények több olyan csoportot emelnek ki, amelyek dopaminerg antagonistákat tartalmaznak. Ez a dopaminerg agonisták és antagonisták hasonló kémiai szerkezetének és célpontjaiknak köszönhető, a hasonlóságon alapuló módszer nem képes ezeket megkülönböztetni. Az anomália olyan információforrások használatával oldható fel, melyek nem érzékenyek erre a hibára, pl. a mellékhatás profil vagy az expressziós profil. Pusztán azt a két információforrást felhasználva a dopaminerg antagonisták nem jelennek meg, míg a dopaminerg agonisták igen. Ehhez hasonlóan (de az előbbinek inverz esete) a kémiai szempontból eltérő, de azonos biológiai folyamatokra ható vegyületek is összekapcsolhatóak.

A tudás újrahasznosítás informatikai értelemben egy olyan szemlélet, amely azt hangsúlyozza, hogy a teljes felhalmozott tudást szisztematikusan ki kell aknázni ahol csak lehetséges és érdemes, függetlenül annak eredetétől vagy eredeti céljától. Ez a megközelítés motiválta a számos iparágban elterjedt adattárház technológiák kifejlesztését is. A gyógyszerfejlesztés eddigi gyakorlatában a hatóanyag fejlesztés lezárulta után az előállított jelentős mennyiségű adat hasznosítása is lezárul. Tágabb értelemben a CSEA a tudás újrahasznosítás eszköztárával növelheti gyógyszerfejlesztés hatékonyságát. Az esettanulmány bemutatta, hogy a módszer jól beilleszthető a gyógyszerfejlesztés információáramlási folyamatába, ráadásul statisztikailag robosztus, számítástechnikailag jól skálázható, és a farmakológiai fogalomrendszerek szintjén képes a farmakológusok által definiált fontos kérdéseket jelezni.

Következtetések

1. Kísérlettervező rendszer

Egy genetikai asszociációs vizsgálat manuális tervezése akár több hónapos, meglehetősen körülményes munka. Ez a lassú, humán erőforrás igényes elő-optimalizálás több nagyságrenddel csökkentheti a genotipizáló műszerek és az egész folyamat elméleti áteresztőképességét, ráadásul sok a hibázási lehetőség. A kifejlesztett TIGER kísérlettervező rendszerrel az előkészítő munka fontos fázisai automatizálhatóak és annak időtartama közel egy nagyságrenddel leszorítható.

2. *SCIN*

Ez volt az első tanulmány, amely a *SCIN* (Scinderin) gén és az asztma asszociációját vizsgálta, és kimutatta humán populációban. A *SCIN* gén három polimorfizmusának, az első exonon az rs2240572-nek (H61R) a minor, valamint a promoter régióban az rs2240571-nek a major és az rs3735222-nek a minor allélja statisztikailag erősen szignifikáns védő hatást mutat az asztmával szemben. Frekventista statisztikai módszerekkel egyetlen potenciális oki variáns hatását

azonosítottam, mely a nőket külön vizsgálva még erősebbnek bizonyult. A BN-BMLA módszer megerősítette az eredményeket és a *TFPI* gén moduláló hatását sikerült azonosítani, amely illeszkedik a *SCIN* gén ismert funkciójához is: mindkét génnek a mucos termelésben van szerepe.

3. *PPARGC1B*

Jelen tanulmányom volt az első, amely összefüggést talált a *PPARGC1B* (Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Beta) gén 2. exonján lévő rs32588 (L42L) polimorfizmus és az asztma között. Frekventista statisztikai módszerekkel a minor allél erősen szignifikáns védő hatását mutattuk ki, a hatás a nőket külön vizsgálva még erősebbnek bizonyult. A BN-BMLA elemzések megerősítették az eredményeket, és a vizsgált polimorfizmus gyenge kölcsönhatást mutatott terheléses asztmával is.

4. *ITLNI*

A tanulmány során azonosítottam az *ITLNI* (Intelectin-1) gén rs4656958 polimorfizmusának asztmával összefüggő potenciális védő hatását. Az asszociációt több módszerrel is sikerült kimutatni, bár szignifikanciája frekventista módszerekkel elemezve határesetinek tekinthető. A BN-BMLA módszertan kimutatta, hogy az infekciós asztmás csoportban a hatás jóval erősebb, ami összhangban van a gén ismert funkciójával is: a géntermék bakteriális komponensek felismerésében játszik szerepet.

5. *LGMN*

Kutatásom az *LGMN* (Legumain) gén potenciális szerepét elsőként azonosította az asztma patomechanizmusában. A tanulmány klasszikus frekventista elemzéssel, allélikus teszttel gyenge, domináns-recesszív modellben lényegesen erősebb, míg BN-BMLA elemzéssel erős bizonyítékot talált a gén és az asztma összefüggésére, különösen az allergiás csoportban vizsgálva.

6. Humán asztma és egér asztmamodell expresszió

Ovalbumin-indukált egér asztmamodellben, a tüdőszövetekben a *Scin*, a *Ppargc1b* és az *Itna* gének emelkedett expresszióját mértünk, míg humán indukált köpet mintákon, az alsó légúti szekrétumban a homológ gének expressziója a krónikus asztmás populációban csökkent. Az eredmények mindhárom esetben statisztikailag szignifikánsak voltak. Így a gének patomechanizmusban betöltött szerepe többszörös megerősítést nyert, de az eltérő irányú expresszió megmutatta a patomechanizmus eltérő dinamikus tulajdonságait és így az egér asztmamodell korlátait.

7. Compound Set Enrichment Analysis (Hatóanyag feldúsulás elemzés)

Ez a tanulmány volt az első, amely a molekuláris biológiában már bizonyított feldúsulás elemzés matematikai megközelítését farmakológiai adatok széles körű integrálására sikerrel alkalmazta. A módszertan a korábbi speciális alkalmazást jelentősen kibővítve tetszőleges hatóanyag könyvtárak változatos mérési adataink alapján, magas szintű farmakológiai tulajdonságok előrejelzését is lehetővé teszi a gyógyszerfejlesztés minden fázisában.

8. Információ újrahasznosítás a CSEA módszertannal

A CSEA módszertan egy statisztikailag robusztus, számítástechnikailag jól skálázható megoldásnak bizonyult a farmakológiai ismeretanyag újrahasznosítására, amely a gyógyszeriparban egy jelentős lehetőségekkel kecsegtető, de eddig mellőzött módszer volt. A módszertan a gyógyszerfejlesztésbe könnyen beilleszthető iteratív módon, heterogén információ források jelzéseinek integrálásával képes lényegesen eltérő vegyületek alapján is következtéseket levonni, és a gyenge hatásokat kiemelni a statisztikai zajból.

Saját publikációk jegyzéke

Az értekezés témájában megjelent közlemények

Temesi G, Bolgár B, Arany Á, Szalai C, Antal P, Mátyus P: Early repositioning through compound set enrichment analysis: a knowledge-recycling strategy. *Future Med Chem* 6(5), 563-575 (2014), IF: 4.000*

Temesi G, Virág V, Hadadi É, Ungvári I, Fodor LE, Bikov A, Nagy A, Gálffy G, Tamási L, Horváth I, Kiss A, Hullám G, Gézsi A, Sárközy P, Antal P, Buzás E, Szalai C: Novel genes in Human Asthma Based on a Mouse Model of Allergic Airway Inflammation and Human Investigations. *Allergy, asthma & immunology research* 6(6), 496-503 (2014), IF: 3.084*

Ungvári I, Hullám G, Antal P, Kiszél Sz P, Gézsi A, Hadadi É, Virág V, Hajós G, Millinghoffer A, Nagy A, Kiss A, Semsei F Á, Temesi G, Melegh B, Kisfali P, Széll M, Bikov A, Gálffy G, Tamási L, Falus A, Szalai C: Evaluation of a partial genome screening of two asthma susceptibility regions using bayesian network based bayesian multilevel analysis of relevance. *PLoS one* 7(3), e33573 (2012), IF: 3.730

Bolgár B, Arany Á, Temesi G, Balogh B, Antal P, Mátyus P: Drug repositioning for treatment of movement disorders: from serendipity to rational discovery strategies. *Current topics in medicinal chemistry* 13(18), 2337-2363 (2013), IF: 3.453

Egyéb témában megjelent közlemények

Gabriella Jobbágy-Óvári, Csilla Páska, Péter Stiedl, Bálint Trimmel, Dorina Hontvári, Borbála Soós, Péter Hermann, Zsuzsanna Tóth, Bernadette Kerekes-Máthé, Dávid Nagy, Ildikó Szántó, Ákos Nagy, Mihály Martonosi, Katalin Nagy, Éva Hadadi, Csaba Szalai, Gábor Hullám, Gergely Temesi, Péter Antal, Gábor Varga, Ildikó Tarján: Complex analysis of multiple single nucleotide polymorphisms as putative risk factors of tooth agenesis in the Hungarian population. *ACTA ODONTOLOGICA SCANDINAVICA* 72:(3) pp. 216-227. (2014), IF: 1.309*