

SMJERNICE U POVEZIVANJU SEKVENCIJSKIH VARIJANTI KANDIDATNIH GENA S KLINIČKIM FENOTIPOM NOVIH MONOGENSKIH BOLESTI U DJECE

BERNARDA LOZIĆ*

Sekvenciranje egzona ili cijelog genoma postaje sve učestaliji molekularno dijagnostički test za osobe s rijetkim genetskim poremećajima. Većina tih slučajeva su klinički heterogena i zahtijevaju široku dijagnostičku obradu. U dijagnostičkoj obradi za uzročnim varijantama većina bolesnika bila je više od 5 godina, a preko spektra genetski raznorodnih Mendelskih poremećaja. Nakon provedene molekularne dijagnostike sekvenciranja svih egzona više od 25% bolesnika dobije točnu dijagnozu. Do sada su istraživanja pokazala dva 2 pika u dobnoj skupini djece: prvi u 4-5 godini što odražava bolesnika s prirođenim poremećajima, a drugi u 16-18 godini, a predstavljaju poremećaje kod kojih simptomi nastaju u ranoj školskoj dobi. Klinička primjenom sekvenciranja egzona ujedno otkrivamo i nove bolesti koje će osigurati medicinski relevantne uvide u biokemiju, fiziologiju i biologiju stanice.

Deskriptori: SEKVENCIRANJE, VARIJANTE GENA, MONOGENSKE RIJETKE I NOVE BOLESTI

Skraćenice:

NIH (engl. National Institutes of Health) = Nacionalni Institut Zdravlja; UDP (engl. Undiagnosed Diseases Program) = Program Nedijagnosticiranih bolesti; Array-SNP (engl. array- Single Nucleotide Polymorphism) = polimorfizmi jednog nukleotida na čipu; NSG (engl. Next Generation Sequencing) = sekvenciranje sljedeće generacije; MAF (engl. Minor Allele Frequency) = učestalosti rjedeg alela; ES (engl. exome sequencing) = sekvenciranje egzona; CES (engl. Clinical Exome Sequencing) = Kliničko sekvenciranje egzona

Uvod

Nove tehnologije sekvenciranja cijelog genoma ili samo egzona omogućile su identifikaciju mnogih rijetkih varijanti koje se uzročno povezuju s rijetkim i novim bolestima dječjoj dobi. U humanom genomu je mapirano oko 23.000 gena, a procjenjuje se da je do danas otkriveno svega oko 7000 rijetkih bolesti pa se očekuje da će se mnoge genetičke bolesti tek biti otkrivene. Poznavanje

ljudskog genoma stoga otvara nove dijagnostičke izazove za otkrivanje iznimno rijetkih poremećaja (1). Od strane američkog Nacionalnog Instituta za Zdravlje, NIH (engl. *National Institutes of Health*, NIH) uvodi se program za nedijagnosticirane bolesnike (engl. *Undiagnosed Diseases Program*, UDP), a s jednim ciljem postavljanja ispravne dijagnoze (2). Mogući razlozi zbog kojih bolesnici unatoč intenzivnom prethodnom ispitivanju mogu ostati nedijagnosticirani su:

- Genetska mutacija nije prethodno bila povezana s fenotipom bolesti.
- Postojanje alelske heterogenost (isti gen, ali različite mutacije stvaraju različit fenotip).
- Postojanje lokusne heterogenost (različiti geni stvaraju sličan fenotip).
- Prisutnost jednog simptoma ili neobičnih kliničkih značajki u višesimptomatske ili rijetke bolesti (1).

Klinička evaluacija bolesnika

Pojam nedijagnosticiran odnosi na bolesnike koji su nakon opsežnog obrade ostali bez konačne dijagnoze. Dijelom je to posljedica zbog toga što svaki poje-

dinac ima jedinstveni genom, a dijelom zbog i jedinstvenih čimbenika okoliša koji utječu na varijabilnu manifestaciju bolesti. Nedijagnosticirani uključuju i one koji nikada nisu viđeni od strane dijagnostičar, imaju neobičnu prezentaciju, inače prepoznatljivih bolesti. Putem bazičnog istraživanja, savjetovanja, kliničkog ispitivanja i razmatranje atipičnih prezentacija ranije poznate bolesti proširuje se diferencijalna dijagnoza. Za bolju prezentaciju kliničke slike prilaže se detaljna fenotipizacija, fotografije i drugi testovi koji se daju na raspolaganju genetičkim bazama u svrhu otkrivanja nove bolesti i genetske varijante (1).

Medicinska dokumentacija uključuje kliničke nalaze iz prenatalnog, neonatalnog razdoblja, neurorazvojni obrazac, somatski rast i razvoj, nastanak i progresiju simptoma, čimbenike okoliša, reakciju na lijekove i naposljetku obiteljsko stablo u kojem se izdvoje pogođeni članovi obitelji. Klinički nalazi uključuju: opis dizmorfologije, prisutnost organomegalija, neurološki deficit i promjene na koži i kostima. Budući da je riječ o rijetkim i novim poremećajima, koji su nerijetko multisistemske, ključnu ulogu u dijagnostičkoj evaluaciji imaju speci-

*Klinika za pedijatriju
Klinički Bolnički Centar Split

Adresa za dopisivanje:
Doc. dr. sc. Bernarda Lozić
Klinika za pedijatriju
Klinički bolnički centar Split
21000 Split, Spinčićeva 1
E-mail: blozic@kbsplit.hr

Tablica 1.
Početno ispitivanje za postavljanje novih mogućih dijagnoza

Test	Povezani poremećaji/grupa poremećaja
Elektroliti, laktat, piruvat	Greške energetskog metabolizma, uključuje i mitohondrijske poremećaje
Aminokiseline u plazmi	Bubrežni poremećaji, poremećaji aminokiselina
Organske kiseline u urinu	Bubrežni poremećaji, poremećaji organskih kiselina, greške energetskog metabolizma, vitaminski deficiti
Aldolaza, kreatin fosfokinaza	Mišićne bolesti
Karnitin (slobodni, ukupni, profil)	Poremećaji oksidacije masnih kiselina i metabolizma karnitina
Analiza likvora	Poremećaji neurotransmitera, prirodnih grešaka metabolizma, specifičnih poremećaja SŽS-a
MR mozga/MR sa spektroskopijom	Strukturni poremećaji SŽS-a
Masena spektrometrija za otkrivanje N-i O- povezane abnormalnosti proteoglikana	Prirodni poremećaji glikozilacije
Testiranje lizosomnih enzima	Lizosomske bolesti taloženja
Elektronska mikroskopija kože	Lizosomske bolesti taloženja; neuronalna lipofuscinoza
Patohistološka procjena zahvaćenog tkiva s posebnim bojenjem, DNA hibridizacija	Svi poremećaji
Echo- /elektrokardiogram	Srčane greške
Brzina provodljivosti živaca Elektromiogram	Neurološki deficiti
Stanična linija fibroblasta	Svi poremećaji
Kariotip/SNP/egzon/genom analize	Svi poremećaji
Brzina sedimentacije eritrocita, C-reaktivni protein	Upalne bolesti

SŽS-središnji živčani sustav; MR- magnetska rezonanca; SNP-polimorfizam jednog nukleotida
Prerađeno prema referenci: Nelson Gahl WA, et al. in Nelson Textbook of Pediatrics; 2015: 629-33.

jalisti različitih profila. Ispitivanja koja je potrebno provesti prije postavljanja moguće dijagnoze, a s ciljem da se otkrije nova bolest prikazani su Tablici 1. Za evaluaciju neuropedijatrijskih slučajeva uključuju se intenzivnija ispitivanja (1).

Raspoloživi laboratorijski genetički testovi

Nakon učinjene fenotipizacije, određuje se diferencijalna dijagnoza genetičkog poremećaja. Sve veći spektar poremećaja ima na raspolaganju panele molekularnih testova. U većini slučajevima takvi paneli uključuje i nekoliko srodnih bolesti. Kao primjeri navode se X-vezana kognitivna oštećenja, nasljedna spastična paraplegija, spinocerebralne ataksije, distonije i mitohondrijske bolesti. Ponekad su pojedinačni testovi skupi i prelaze troškove sekvenciranja egzona (1, 2).

Dvema tehnikama moguće je na brz i jeftin način ispitati cijeli genom s rezo-

lucijom na razini jednog para baza. To su polimorfizmi jednog nukleotida na čipu (array-SNP) (engl. *array-SNP*) i sekvenciranje sljedeće generacije (eng. *next generation sequencing*, NGS). U jednoj laboratorijskoj pretrazi, array-SNP istovremeno se ispituje nekoliko stotina tisuća do nekoliko milijuna SNP-ova (1).

Varijante gena

Ljudski genom ima 3,2 milijarde parova nukleotida u molekuli DNA. Protein-kodirajuće sekvence (egzoni) čine manje od 1,7% ljudskog genoma (3). Ljudi se međusobno razlikuju u oko 0,1% genoma (4). Sekvenciranjem cijelog genoma ili samo egzona otkriven je veliki broj rijetkih varijanti u genomu koje se nazivaju sekvencijske varijante. Uobičajeno su varijante svrstane u tri skupine prema učestalosti rjeđeg alela (engl. *Minor Allele Frequency*, MAF) u populaciji. Varijanta je učestala u populaciju ako

je njen MAF veći od 5%, a rijetka ako je manji od 1%. Velika većina sekvencijske varijante u genomu su SNP-ovi, a ima ih oko 3,5 milijuna u svakom genomu. U SNP-ovoj baza podataka (dbSNP) navodi se više od 37 milijuna varijanti među ljudima. SNP-ovi unutar kodirajućeg područja gena dijele se u dvije glavne skupine prema utjecaju na polipeptidni slijed kodirajućeg proteina, na sinonimne SNP-ove (sSNP) (engl. *synonymous*) ili nesinonimne (nsSNP) (engl. *non-synonymous*) (5). Svega 10.000 SNP-ova su nesinonimni SNP-ovi i smatraju se da uzrokuje promjene u kodirajućem polipeptidu te imaju potencijalni efekt na fenotip. SNP-ovi mogu biti locirani u intronima i regulatornim područjima gena pa mogu utjecati na razinu genske ekspresije, transkripcijsku aktivnost gena i prekradanje mRNA (5).

Sekvenciranjem egzona određuje se slijed sekvenci samo u kodirajućim

dijelovima gotovo svih gena. Budući da to uključuje 1,7% od 3,2 milijarde baza u ljudskom genomu to je oko 60 milijuna baza (3). Ovom tehnikom otkrivamo poremećaje u 80-85% poznatih gena. Prosječno egzoni imaju oko 20.000 baze (0,03%), koji se razlikuju od "referentnog" u jednoj sekvenci, a obično se radi nevažnim varijantama povezanih s različitim etničkim skupinama. Uloga svake od ovih 20.000 varijanta nije poznata i predstavlja mogući uzrok bolesti (engl. *disease-causing variant*). Međutim samo jedna od varijanata uzrokuje bolest, a 2-3 dodatna lokusa moguće modificiraju težinu bolesti (7-9).

Zadatak kliničara je reducirati vjerojatnost varijanti sa 20.000 na vjerojatnih 5 koje će potencijalno povezati s bolešću. Varijante s MAF većom od 1% uklone se iz istraživanja i smatraju se benignima, a pri tom se najčešće koriste javno dostupne baze podataka kao što su dbSNP i 1000 Genomes Projekt (7).

Za ovaj proces potrebno je uključiti programe filtriranja (filters) ili programe koje eliminiraju lažno pozitivne varijante bez otklanjanja prave varijante (10). Jedan od najboljih filtra koji se koristi pri sekvenciranju egzona, ES (engl. *exome sequencing*, ES) su članovi uže obitelji (roditelja, braće i sestara). Ako, na primjer, kod zahvaćenog probanda nađemo u jednom genu dvije varijante na suprotnim alelima, a iste nađemo i kod zdravog brata ili sestre, onda se one eliminiraju jer nisu uzrok bolesti. Pažljiva evaluacija svih varijanti u oba roditelja, braće ili sestara daje dovoljnu snagu za reduciranje kandidatnih varijanata gena na razuman broj za sve Mendelove (Mendelian) modele nasljeđivanja, pod pretpostavkom potpune penetracije (1). U slučajevima trio-CES (engl. *Clinical exome sequencing*, CES) simultano se sekvenciraju egzoni bolesnika i oba roditelja, a njihove varijante se filtriraju u 4 kategorije:

- de novo (nove varijante nisu pronađene ni u jednog roditelja, obično su heterozigotne u bolesnika, a potencijalno uzrokuje autosomno dominantnu bolest);
- homozigotne (oba roditelja su heterozigotni za istu varijantu, dijete na-

sljeđuje rijetke alele od oba roditelja, uzrokuju potencijalno recesivnu bolest);

- složeni heterozigoti (bolesnik ima jednu rijetku varijantu koju je naslijedio od majke, a drugu rijetku varijantu je naslijedio od oca, uzrokuju potencijalno recesivnu bolest);
- naslijeđene varijante (ovo je najveća skupina varijanti, koja je naslijeđena od roditelja i obično ne uzrokuju bolest (11).

Evaluacija kandidatnih gena nakon filtriranja

Softverski programi, koji uključuju PolyPhen-2, SIFT, i MutationTaster daju udio patogenosti aminokiselinske promjene. Promjene nukleotidnih baza koji mijenjaju aminokiselinske sekvence (*missense mutations*) se evaluiraju od strane programa kao patogene promjene. Ovi programi uključuju i procjenu evolucijske konzistentnosti baze, a ona može biti dokaz za patogenost te baze čak i ako nije uključena u promjenu aminokiseline već se nalazi na vežućim regijama promotora i metilacijskim mjestima. Međutim u većini slučajeva do sada nije poznato da li promjene genetskog koda ostavljaju posljedice na funkciju proteina (12, 13).

Funkcionalne studije gena

Kod rijetkih monogenih bolesti, bolesnika je premalo da se dokaže statistički značajna povezanost s varijantom (1). Stoga prije nego se varijanta identificirana sekvenciranjem genoma i interpretira kao uzročna za bolest testirane osobe treba uzeti u obzir određene čimbenike. Dokazi koji se razmatraju pri procjeni kandidatnih patogenih mutacija za pretpostavljenu monogenku bolest su dokazi na razini gena i dokazi na razini varijante.

Dokazi na razini gena: uzročne varijante u gena pronađene su u više članova neovisnih obitelji sa sličnim fenotipom. Za nove kandidatne gene, a za kojeg pretpostavljamo recesivno nasljeđivanje s posljedičnim gubitkom funkcije gena, varijante se ne smije naći u homozigotnom obliku kod zdravih. Gen se ekspri-

mira u tkivima relevantnim za bolest. Ukupni profil ekspresije gena sličan je drugim genima koji su mutirani u istraživanoj bolesti ili fenotipski sličnim bolestima. Protein kodiran kandidatnim genom je u interakciju s proteinima koje kodiraju drugi geni mutirani u istraživanoj ili fenotipski sličnim bolestima (7).

Dokazi na razine varijante

Ovi dokazi uključuje prethodna izvješća o uzročnoj povezanosti varijante s istraživanom bolesti ili fenotipski sličnim bolestima, a fenotip bolesnika je u skladu s fenotipom zabilježenim kod drugih nosioca iste varijante. Segregaciju varijante s obzirom na penetraciju i način nasljeđivanja radi se prema broju pogođenih članova u obitelji. Svi dostupni članovima obitelji, zdravi i bolesni trebaju biti testirani na uzročnu varijantu. Ako je varijanta *de novo* mutacija, roditelje treba isključiti kao moguće nositelje. Učestalost varijante trebala bi biti vrlo niska ili da je nema, a za to se koriste odgovarajuće javne baze podataka ili odredi učestalost na odgovarajućem velikom uzorku zdravih pojedinaca koji potječu iz iste populacije kao i bolesnik. Ako se varijanta nađe u populaciji zdravih ispitanika, njena učestalost bi trebala biti u skladu s pretpostavljenim načinom nasljeđivanja i poznatom incidencijom bolesti kod testirane populacije. Genotip pronađen u bolesnika ne smije biti u zdravoj populaciji. Testiranjem treba isključiti ostale varijante u drugim genima kod bolesnika za koje se zna da su uzročno povezane s bolešću. Ako je već ranije pokazano da mutirani gen ima recesivni obrazac nasljeđivanja, kod bolesnika treba dokazati mutaciju u obje kopije gena. U haplotipu koji nosi varijantu isključuju se druge varijante kojim bi se predvidio funkcionalni utjecaj. Potencijalno štetni utjecaj varijanti kod točkastih mutacija i pomaka okvira čitanja podupiru se funkcionalnim testiranjem kao što su mjerni podaci o njenoj konzerviranosti tijekom evolucije.

Nadalje moguć je i eksperimentalni pristup da se varijanta uzročno poveže s bolešću. Razviju se modeli patogene varijante na staničnim kulturama iz bolesnikova tkiva ili animalni modeli

(miševi, zebaste ribice, vinske mušice, kvasci) i dokažu štetne posljedice patogene sekvencijske varijante u njihovom fenotipu (7).

Konačno, naposljetku je potrebno standardizirati i povezivati fenotipske i genomske podatke i pohraniti u bazu podataka za identifikaciju drugih pojedinaca sa sličnim fenotipom i mutacijama u istom genu. Baze podataka koje se najučestalije koriste pri tome su: US National Center for Biotechnology Information (NCBI)'s www.ncbi.nlm.nih.gov/; ClinVar baza podataka www.ncbi.nlm.nih.gov/clinvar/; OMIM (Online Mendelian Inheritance in Man) <http://omim.org/>; DECIPHER baza podataka (*Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources*) (14).

Pedijatrijski slučajevi

Tijekom prve 4 god, UDP na NIH-a obradili su 500 pedijatrijska aplikacija. U više od 10% slučajeva, pokazano je da je na sličan način pogođeno bolešću više od jednog člana obitelji, a obično su to bili braća ili sestre. Istraživanja su pokazala da postoje dva pika bolesnika u dobnim skupinama djece: prvi u dobi od 4-5 godina, predstavlja bolesnike s prirođenim poremećajima, i drugi u dobnj skupini 16-18 godina godine koja predstavlja poremećaje kod kojih su se simptomi pojavili u ranoj školskoj dobi. Većina podnositelja zahtjeva imalo je vrlo dug dijagnostički postupak više od 5 godina. Od obrađenih pedijatrijskih slučajeva njih 25% dobili su dijagnozu (1, 2).

Svaka novoootkrivena varijanta u ljudskom genomu koja se uzročno poveže s specifičnom kliničkim fenotipom treba zadovoljiti stroge biološke stan-

darde, a s ciljem da se spriječi negativni utjecaj genomskih rezultata u biološkom razumijevanja bolesti (7).

NOVČANA POTPORA/*FUNDING*
Nema/*None*

ETIČKO ODOBRENJE/*ETHICAL APPROVAL*
Nije potrebno/*None*

SUKOB INTERESA/*CONFLICT OF INTEREST*
Autori su popunili *the Unified Competing Interest form* na www.icmje.org/coi_disclosure.pdf (dostupno na zahtjev) obrazac i izjavljuju: nemaju potporu niti jedne organizacije za objavljeni rad; nemaju financijsku potporu niti jedne organizacije koja bi mogla imati interes za objavu ovog rada u posljednje 3 godine; nemaju drugih veza ili aktivnosti koje bi mogle utjecati na objavljeni rad./ *All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.*

LITERATURA

1. Nelson Gahl WA, Adams DR, Markello TC, Boerkoel NF, Tiftt CJ. Genetic Approaches to Rare and Undiagnosed Diseases. In: Kliegman RM, Stanton B, St Geme J, Schor NF, editors. Nelson Textbook of Pediatrics. 2-Volume Set, 20th ed.; Elsevier; 2015; 629-33.
2. Gahl WA, Tiftt CJ. The NIH Undiagnosed Diseases Program: lessons learned. JAMA 2011; 305: 1904-5.
3. Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. Nature 2001; 409 (6822): 860-921.
4. Marian AJ. Nature's genetic gradients and the clinical phenotype. Circ Cardiovasc Genet 2009; 2 (6): 537-9.
5. Maurano MT, Humbert R, Rynes E et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012; 337 (6099): 1190-5.
6. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 2010; 467: 1061-73.
7. MacArthur DG, Manolio TA, Dimmock DP et al. Guidelines for investigating causality of sequence variants in human disease. Nature 2014; 508 (7497): 469-76.
8. Duzkale H, Shen J, McLaughlin H et al. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet. 2013; 84 (5): 453-63.
9. Richards CS, Bale S, Bellissimo DB et al. ACMG recommendations for standards for interpretation and reporting of sequence variations. Genet Med 2008; 10 (4): 294-300.
10. Yourshaw M, Taylor SP, Rao AR, Martín MG, Nelson SF. Rich annotation of DNA sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins. Brief Bioinform 2015; 16 (2): 255-64.
11. Lee H, Deignan JL, Dorrani N et al: Clinical exome sequencing for genetic identification of rare mendelian disorders, JAMA 2014; 312: 1880-7.
12. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protoc 2009; 4: 1073-81.
13. Adzhubei IA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations. Nature Methods 2010; 7: 248-9.
14. Firth HV, Richards SM, Bevan AP et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 2009; 84: 524-33.

Summary

GUIDELINES IN LINKING SEQUENTIAL VARIANTS OF CANDIDATE GENES WITH CLINICAL PHENOTYPES
NEW MONOGENIC DISORDERS IN CHILDREN

B. Lozić

Sequencing exomes or whole genome are rapidly becoming a common molecular diagnostic test for individuals with rare genetic disorders. The majority of these cases are clinically heterogeneous and require a broad search for causal variants across the spectrum of genetically heterogeneous Mendelian disorders. Most patients in the diagnostic processing of was more than 5 years. After performed molecular diagnostic more than 25% patients receive accurate diagnosis. Until now research showed two 2 peaks in the age distribution of the children: first at 4-5 year, reflecting patients with congenital disorders, and second at 16-18 year representing disorders with symptoms onset at early school age. Clinical exome sequencing also discover new diseases that will provide medically relevant insights into the biochemistry, physiology and cell biology.

Descriptors: SEQUENCING, GENE VARIANTS, MONOGENIC RARE AND NEW DISEASES

Primljeno/Received: 12. 3. 2016.

Prihvaćeno/Accepted: 10. 4. 2016.