

Skóre polygénneho rizika a jeho využitie pre odhad kumulatívneho rizika ochorení diabetu mellitu a COVID-19

Patrik Krumpolec, Klaudia Babišová, Michaela Hýblová, Gabriel Minárik
MEDIREX GROUP ACADEMY, n. o., Nitra

Stanovenie miery rizika výskytu ochorení je aktuálnou témou už po desaťročia. Veľká nádej sa vkladala do programu Human Genome Project, od ktorého sa očakávalo, že poodhalí genetické pozadie ochorení a uľahčí predikciu výskytu ochorení. Ako sa však ukázalo, bola to len ilúzia. Namiesto priamych vzťahov medzi génmi, resp. skupinami génov a s nimi spájanými ochoreniami bol odhalený celý komplex interakcií naprieč genómom, zahŕňajúc stovky až milióny jednonukleotidových polymorfizmov. Navyše, s nimi asociované ochorenia majú často polygénny charakter a jednotlivé varianty majú samostatne len malý vplyv na výsledný fenotypový prejav ochorenia. Výskum v tejto oblasti sa teda vrátil k pôvodnému zámeru predikcie na základe skórovania, ktoré by zohľadnilo čo najviac faktorov vplývajúcich na rozvoj ochorenia. Neustále sa rozširujúca sieť genetických databáz v spojení s moderným hardvérovým vybavením umožňujúcim využitie náročných štatistických analytických nástrojov umožnilo výpočet spoločného genetického rizika pre špecifické ochorenia, resp. pre kombináciu ochorení. Jednou z takýchto metód je skóre polygénneho rizika.

Kľúčové slová: skóre polygénneho rizika, jednonukleotidový polymorfizmus, COVID-19, diabetes

Polygenic risk score and their utilization for cumulative risk assessment of diabetes mellitus and COVID-19 diseases

Determination of disease risk prediction has been an actual topic over the last decades. A great hope was put in the Human Genome Project, which was expected to reach insight into the genetic background of diseases that allow us to predict disease risk. However, these ideas were only illusions. Instead of simple relations between genes or a set of genes and associated diseases, there was a complex of interactions across the whole genome involving hundreds to millions of single nucleotide polymorphisms. Moreover, such diseases have a polygenic character when the genetic variants have only a tiny effect on the disease development. Thus, research was forced to backtrack to the original idea of scoring, which would incorporate as many factors affecting the disease process as possible. Expanding genetic-database network with modern hardware equipment for data-intensive statistical analytical approaches has enabled the calculation of common genetic risks for specific diseases and their combination. One such method is the polygenic risk score.

Keywords: polygenic risk score, single nucleotide polymorphism, COVID-19, diabetes

NewsLab, 2023; roč. 14 (1): 49 – 51

Pojmom genetický variant sa označuje rozdiel v nukleotidovej sekvencii na úrovni DNA reťazca medzi jednotlivcami v populácii. Rozlišujeme 5 základných typov genetických zmien: a) štruktúrne varianty; b) tzv. jednonukleotidové varianty; c) varianty typu inzercii a delécií; d) varianty v počte kópií (z angl. copy number variations); a e) varianty typu translokácií a inverzií.

Zámena jednej z báz v bázevom páre sa označuje ako jednonukleotidový variant (single-nucleotide variation, SNV) na úrovni jednotlivca, resp. jednonukleotidový polymorfizmus (single-nucleotide polymorphism, SNP) na úrovni celej populácie. Odhaduje sa, že každý človek je nositeľom niekoľkých miliónov takýchto SNP polymorfizmov, pričom priemerná nukleotidová diverzita π (z angl. average nucleotide diversity) definovaná ako priemerné zastúpenie rozdielov v nukleotidoch medzi náhodne vybranými jedincami sa pohybuje niekde medzi 1 : 1 000 až 1 : 1 500⁽¹⁻²⁾.

Tieto zmeny na úrovni genómu môžu byť jednak dedičné, ale môžu vzniknúť aj spontánne počas bunkového delenia,

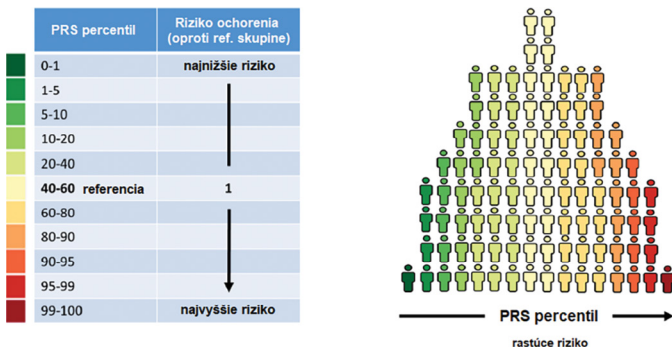
prípadne v dôsledku pôsobenia faktorov vonkajšieho prostredia. Väčšina variantov má minimálny, resp. žiadny fenotypový prejav, a teda neovplyvňujú zdravotný stav ich nositeľa, avšak viaceré z nich sa môžu prejaviť aj na úrovni fenotypu nevynímajúc ani zmeny patologického charakteru manifestovateľné až ako ochorenie.

V porovnaní s mendelistickou dedičnosťou, pri ktorej dochádza k zmenám na úrovni jedného génu, prípadne malého počtu génov veľkého účinku, pri genetických variantoch ide o polygénny typ dedičnosti zahŕňajúci zmeny veľkého množstva génov malého, prípadne stredného účinku.

Hoci v súčasnosti sú už jednonukleotidové polymorfizmy využívané v genetických analýzach, výzvou do budúcnosti, stále ostáva stanovenie rizika dedičných ochorení zohľadnením viacerých jednonukleotidových polymorfizmov vzhľadom na polygénny typ dedičnosti mnohých ochorení.

Rozsiahle projekty mapovania genomických variantov typu HapMap či 1000 Genome Project s podporou menších platforiem umožnili realizáciu tzv. Genome-wide association

Obrázok 1. Riziko polygénneho skóre (PRS) vyjadrené ako percentil genetického rizika. Jedinci s PRS blízky populáčnemu priemeru majú odhadované genetické riziko podobné populačnému riziku. Jedinci s PRS v 91. – 100. percentile majú najvyššie genetické riziko. Jedinci s PRS do 10. percentilu majú najnižšie genetické riziko. Upravené podľa(5).



studies (GWAS) na veľkých kohortách⁽³⁻⁴⁾. GWAS umožnili doposiaľ identifikáciu tisícok genetických variantov (predovšetkým SNP), ktoré sú úzko prepojené s určitým fenotypovým prejavom, prípadne priamo s ochorením. Získané poznatky o génoch, ich vzájomných vzťahoch a s nimi asociovaných dráhach sa podarilo využiť pri zostavení modelu na odhad genetického rizika. Takýmto odhadom vyjadrujúcim vplyv genetických variantov naprieč ľudským genómom do jedného skóre je tzv. skóre polygénneho rizika (z angl. polygenic risk score, PRS).

Skóre polygénneho rizika možno v jednom z najjednoduchších vyjadrení vypočítať ako súčet n jednonukleotidových polymorfizmov (SNP) vážených mierou ich účinku β :

$$\sum_i^n x_{ij} \beta_i$$

kde n je počet zahrnutých SNP, x_{ij} je počet kópií i -tého SNP v genotype j -tého jedinca a β vyjadruje mieru účinku príslušného polymorfizmu vypočítaného z GWAS analýzy.

Riziko polygénneho skóre možno vyjadriť aj ako percentil genetického rizika⁽⁵⁾. Jedinci s PRS blízky populačnému priemeru majú odhadované genetické riziko podobné populačnému riziku. Jedinci s PRS do 10. percentilu majú najnižšie genetické riziko, zatiaľ čo osoby s PRS v 91. – 100. percentile majú genetické riziko najvyššie (**obrázok 1**).

Na stanovenie PRS sa štandardne využívajú stovky až tisíce SNP, avšak treba si uvedomiť, že so zvyšujúcim sa počtom zahrnutých SNP v analýze rastie nielen jej štatistická sila, ale aj miera šumu. Preto pri navrhovaní PRS treba nájsť rovnováhu medzi menším počtom analyzovaných SNP s celkovo vyššou presnosťou na jednej strane a vyšším počtom zahrnutých SNP v analýze, ale s nižšou presnosťou⁽⁶⁾. Navyše stanovenie takejto rovnováhy je komplikované aj ďalšími faktormi, ako napríklad genetické pozadie ochorenia, hustota genotypizácie či veľkosť populácie. Navrhnutý model PRS si preto vyžaduje optimalizáciu na nezávislom testovacom súbore, čím sa eliminuje riziko „preučenia“ (z angl. overfitting). Komplikáciu predstavuje aj väzbová nerovnováha súvi-

siaca s tesnou blízkosťou identifikovaných génov, ktorá môže byť príčinou falošne pozitívnych výsledkov. Významným parametrom, na ktorý netreba zabúdať pri zostavovaní aj využívaní PRS, je rasová homogenita populácie. Približne 85 – 90 % genetických variantov sa vyskytuje v rámci troch základných kontinentálnych skupín (európskej, ázijskej a africkej) a len zvyšných 10 – 15 % variantov možno nájsť medzi nimi⁽⁷⁾. A hoci situácia sa v poslednom období zlepšila, stále platí, že väčšina modelov PRS je konštruovaná pre európsku populáciu. Napríklad u Európanov sa riziko diabetu zvyšuje s obezitou, zatiaľ čo u obyvateľov východnej Ázie riziko diabetu rastie už pri nadváhe, pretože je potrebné zohľadniť aj rozdiely v sekrécii inzulínu.

Cieľom pri navrhovaní modelu PRS preto je, aby bolo možné takýto model využiť na čo najširšom populačnom výbere.

Napriek značnému rozšíreniu PRS v praxi je k dispozícii stále len obmedzené množstvo príručiek na zostavenie optimálneho PRS, ako aj na interpretáciu výstupných dát z analýzy⁽⁸⁾.

S rastúcim záujmom o PRS sa zdokonaľujú nielen modely výpočtu, ale aj dostupnosť údajov z už existujúcich databáz. V súčasnosti je možné už aj zverejňovať dáta potrebné pre analýzu napr. prostredníctvom platforiem Polygenic Score Catalog alebo Cancer PRS-Web⁽⁹⁻¹⁰⁾

Súčasnosť je ovplyvnená na jednej strane pandémiou koronavírusu, ale vzhľadom na viac ako desaťpercentnú prevalenciu diabetu ju už možno charakterizovať aj ako obdobie pandémie diabetu. Do popredia sa preto dostáva potreba stanoviť mieru kumulatívneho rizika pre obidve ochorenia, pretože pribúda čoraz viac a viac dôkazov o vzájomnom prepojení obidvoch ochorení.

Prakticky už na začiatku pandémie COVID-19 sa pri predikcii vývoja ochorenia zvažoval vek, pohlavie a pridružené komorbidity ako rizikové faktory. V tom čase však neboli dostupné relevantné dáta, ktoré by umožnili definovať mieru ich vplyvu na závažnosť ochorenia, čo viedlo k nesprávne odhadu rizika samotnej infekcie, ako aj závažného priebehu ochorenia.

Z laboratórnych hlásení bolo pomerne skoro známe, že takmer všetci kriticky chorí pacienti vykazujú závažnú hyperglykémiu, ktorá sa začala využívať ako marker závažnosti priebehu ochorenia. A hoci ešte ani dnes nie je jasné, či infekcia SARS-CoV-2 indukuje diabetogénny stav podobným mechanizmom, aký sa uplatňuje v patogenéze diabetu 1. a 2. typu alebo ide o atypickú formu diabetu⁽¹¹⁾, je známych viacero mechanizmov spoločných pre obidve ochorenia. Patrí medzi ne chronický zápalový stav manifestovateľný zvýšenou koncentráciou prozápalových cytokínov IL-11, IL-6 a TNF- α , ako aj zvýšenou hladinou C-reaktívneho proteínu (CRP) a adhéziou monocytov k cievnemu endotelu⁽¹²⁻¹³⁾. Pozorované sú aj zvýšené hladiny chemokínov CCL1, CCL2, CCL4 a CXCL10. Norouzi a kol. opísali u pacientov s ochorením COVID-19 signifikantne zvýšené hladiny zápalových cytokínov a chemokínov v stave akútnej hyperglykémie s následným zvýšeným rizikom multiorgánového zlyhania⁽¹⁴⁾.

Bránou vstupu vírusu SARS-CoV-2 do organizmu sú receptory enzýmu konvertujúceho angiotenzín (ACE 2), ktoré sú hojne rozšírené vo viacerých typoch tkanív a ktoré zohrávajú významnú úlohu aj pri diabete. Po infekcii sa vírus replikuje v endokrinných a exokrinných sekréčných bunkách pankrea-

su s následným poškodením pankreatických β -buniek. Uvažuje sa o vzájomnom vzťahu systémovej zápalovej odpovede a imunitnej dysfunkcie na jednej strane a hyperglykémie a inzulinovej rezistencie v dôsledku poruchy funkcie pankreatických β -buniek na druhej strane⁽¹⁴⁻¹⁵⁾. Recipročne v porovnaní so zdravými jedincami majú diabetici vyššie riziko koronavírusovej infekcie, a to s celkovo horšou prognózou a vyššou mierou úmrtnosti⁽¹⁶⁾.

S pribúdajúcimi poznatkami o infekcii SARS-CoV-2 sa rozširujú aj vedomosti o genetickom pozadí ochorenia COVID-19, ktoré by mohli prispieť k odhadu rizika infekcie, resp. odhadu závažnosti priebehu ochorenia. Napríklad Dite a kol. identifikovali 64 jednonukleotidových polymorfizmov a klinicky významné rizikové faktory, ktoré následne využili pri vytvorení modelu na odhad miery rizika závažného priebehu ochorenia COVID-19 pre pacientov infikovaných vírusom SARS-CoV-2 vo veku nad 50 rokov⁽¹⁷⁾. Navyše preukázali, že ich model má v porovnaní s modelmi založenými len na veku a pohlaví vyššiu rozlišovaciu schopnosť pre závažnosť priebehu ochorenia. Hoci počet podobných modelov postupne pribúda napriek nespornej potrebe stanovenia genetického rizika vo forme PRS

spoločného pre COVID-19 aj diabetes 2. typu, v súčasnosti prakticky stále neexistuje žiadna systematická analýza spoločných lokusov charakteristických pre obidve ochorenia.

Téma predikcie priebehu ochorení sa stáva čoraz aktuálnejšou, a to tak z pohľadu pacienta, ako aj z celospoločenského hľadiska vzhľadom na neustále sa zvyšujúce výdavky na zdravotnú starostlivosť uhrádzané z verejných zdrojov. Neustály technologický rozvoj a zavádzanie moderných postupov v genetickom výskume viedlo k vytvoreniu viacerých modelov pre predikciu rizika infekcie a priebehu ochorení. Jedným z nich je metóda skóre polygénneho rizika, ktoré má veľký potenciál uplatnenia v biomedicínskom výskume aj v klinickej praxi.

Podakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Závažné civilizačné choroby a COVID-19, ITMS: 313011AVH7, spolufinancovaný Európskym fondom regionálneho rozvoja.

LITERATÚRA

1. Sachidanandam R, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 2001. 409(6822): p. 928-933.
2. Schneider JA, et al. DNA variability of human genes. *Mech Ageing Dev*, 2003. 124(1): p. 17-25.
3. Genomes Project C, et al. A global reference for human genetic variation. *Nature*, 2015. 526(7571): p. 68-74.
4. International HapMap C, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, 2010. 467(7311): p. 52-58.
5. Lund H. Polygenic risk scores – A useful tool in our risk prediction toolkit? *Journal of the Association of Insurance Medicine of Japan*, 2019. Dostupné online: <https://www.rgare.com/knowledge-center/article/polygenic-risk-scores-a-useful-tool-in-our-risk-prediction-toolkit>.
6. Lambert SA, G Abraham, and M Inouye, Towards clinical utility of polygenic risk scores. *Hum Mol Genet*, 2019. 28(R2): p. R133-R142.
7. Jorde LB and SP Wooding. Genetic variation, classification and 'race'. *Nat Genet*, 2004. 36(11 Suppl): p. S28-33.
8. Choi SW, TS Mak, and PF O'Reilly. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*, 2020. 15(9): p. 2759-2772.
9. Fritsche LG, et al. Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. *Am J Hum Genet*, 2020. 107(5): p. 815-836.
10. Lambert SA, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet*, 2021. 53(4): p. 420-425.
11. Metwally AA, et al. COVID-19-Induced New-Onset Diabetes: Trends and Technologies. *Diabetes*, 2021. 70(12): p. 2733-2744.
12. King GL, The role of inflammatory cytokines in diabetes and its complications. *J Periodontol*, 2008. 79(8 Suppl): p. 1527-1534.
13. Rabaan AA, et al. Role of Inflammatory Cytokines in COVID-19 Patients: A Review on Molecular Mechanisms, Immune Functions, Immunopathology and Immunomodulatory Drugs to Counter Cytokine Storm. *Vaccines (Basel)*, 2021. 9(5).
14. Norouzi M, et al. Type-2 Diabetes as a Risk Factor for Severe COVID-19 Infection. *Microorganisms*, 2021. 9(6).
15. Yang JK, et al. Binding of SARS coronavirus to its receptor damages islets and causes acute diabetes. *Acta Diabetol*, 2010. 47(3): p. 193-199.
16. Roncon L, et al. Diabetic patients with COVID-19 infection are at higher risk of ICU admission and poor short-term outcome. *J Clin Virol*, 2020. 127: p. 104354.
17. Dite GS, NM Murphy, and R Allman. An integrated clinical and genetic model for predicting risk of severe COVID-19: A population-based case-control study. *PLoS One*, 2021. 16(2): p. e0247205.

Mgr. Patrik Krumpolec

MEDIEX GROUP ACADEMY. n. o.

Novozámocká 1/67, 949 05 Nitra

e-mail: Patrik.Krumpolec@medirexgroupacademy.sk