

Ročník XIV.

newslab

časopis laboratórnej medicíny



2023

Clinically relevant RNA-virus whole genome assembly testing on metatranscriptomic data gained directly from COVID-19 clinical samples

Skladanie genómu klinicky relevantných RNA vírusov z klinických vzoriek pochádzajúcich od COVID-19 pacientov

Hodnotenie početnosti, pozitivity a Ct hodnôt SARS-CoV-2 detekcie na Slovensku v čase pandémie COVID-19 – deti verzus dospelí jedinci

Evaluation of the frequency, positivity and Ct values of SARS-CoV-2 detection in Slovakia during the COVID-19 pandemic – children versus adults

Urothelial carcinoma of the bladder with adamantinoid (ameloblastoma-like) features. Case report

Urotelový karcinóm močového mechúra s adamantinoidnou morfológiou. Kazuistika

Tandem repeat motif characterization for precision medicine: A brief overview of conventional methods and massive parallel sequencing approaches

Charakterizácia tandemových repetitívnych motívov v precíznej medicíne: Stručný prehľad konvenčných metód a masívne paralelného sekvenovania

newslab – recenzovaný časopis laboratórnej medicíny ponúka autorom priestor na uverejnenie výsledkov ich vedecko-výskumnej práce formou prehľadových článkov, pôvodných prác, noviniek v odbore alebo kazuistík.

newslab už štrnásť rokov prináša čitateľom najnovšie poznatky z oblasti vedy a výskumu. Jeho hlavnou cieľovou skupinou sú lekári, vedecko-výskumní pracovníci na univerzitách a laboratórni pracovníci. Je distribuovaný aj do všetkých dôležitých vedeckých knižníc.

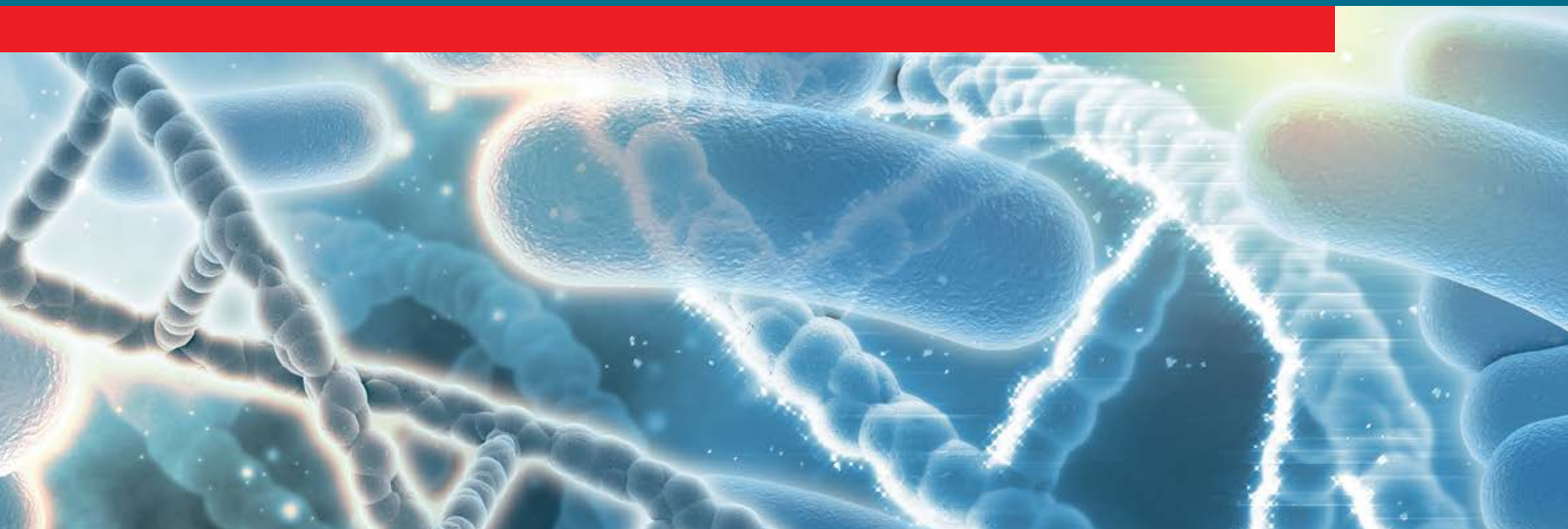
Časopis má pridelenú citačnú skratku, čo v praxi znamená, že autorom sa ich príspevky započítavajú do publikačnej činnosti. Zároveň je indexovaný v Bibliographia Medica Slovaca (BMS) a citácie sú spracované v CiBaMed.

Všetky uverejnené príspevky autorov sú recenzované.

Vydavateľ: MEDIREX GROUP ACADEMY n. o.
Redakcia: MEDIREX GROUP ACADEMY n. o.
Periodicita: 2x ročne
Počet strán: 70 – 100
Náklad: 5 000 ks

Aktuálne vydanie nájdete na stránke www.newslab.sk.

V prípade záujmu o publikovanie nás kontaktujte na e-mailovej adrese: redakcia@newslab.sk.



Milí čitatelia,

v súčasnosti žijeme fascinujúcu dobu, ktorá významným spôsobom ovplyvní všetky oblasti života, tak ako ich dnes poznáme. Stojíme doslova na prahu štvrtej technologickej revolúcie, ktorá zmení život, akým žijeme, pracujeme, komunikujeme. Príchod umelej inteligencie (angl. artificial intelligence, AI), jej rozvoj, aplikácia budú znamenať významný posun pre celú populáciu. Hoci si momentálne nevieme ani len predstaviť všetky možné aplikácie AI, už teraz je jasné, že reakcia rôznych sektorov spoločnosti musí byť integrovaná a komplexná. Ako príchod AI ovplyvní vedu a výskum v biomedicíne? Biomedicínsky výskum je kľúčový nástroj pre získanie kvalitných biologických vzoriek a dát nevyhnutných na realizáciu experimentálnych a dátových analýz, ktorých výsledky prispievajú k novým objavom a vzniku inovácií. Biobanky sú podporné platformy pre výskum, zaručujú prístup týchto vzoriek v dostatočnej kvalite, množstve, zabezpečujú ich adekvátne fyzické uskladnenie a vďaka využívaniu informačných systémov ukladanie informácií o vzorkách a údajov o pacientoch, čím sa stávajú kľúčovým nástrojom úspechu translačného výskumu. Biobanky sú zvyčajne súčasťou kooperatívnych pracovných sietí, čím sa výrazne zvyšuje ich schopnosť poskytovať biologické zdroje. Slovensko patrí medzi posledné krajiny Európskej únie, ktoré nemajú zriadenú vlastnú národnú biobanku. Hoci existujú viaceré lokálne projektové biobanky alebo nemocničné biobanky, tie nepostačujú, pretože sú väčšinou úzko zamerané na niektoré konkrétne typy biologického materiálu alebo na pacientov s vybranými diagnózami a nemajú k dispozícii potrebnú technickú infraštruktúru na ukladanie asociovaných klinických alebo experimentálnych dát. Vybudovanie národnej biobanky je preto veľmi potrebné a je dlhodobým cieľom viacerých štátnych, ale aj súkromných subjektov na Slovensku. V krajinách, kde výskumná platforma v podobe sietí biobáň existuje, je vďaka dostupnosti obrovského objemu klinických údajov a biomedicínskych dát jednoduchšie zaviesť metódy AI. Aplikácia metód AI môže výrazne prispieť k zlepšeniu klinického rozhodovania a personalizovanej zdravotnej starostlivosti. Najmä prístupy strojového (angl. machine learning, ML) a hĺbkového učenia (angl. deep learning, DL) môžu pomôcť identifikovať mechanizmy vzniku ochorení, vyhodnocovať prognózu ochorení, zamerať sa na zlepšenie terapeutických stratégií, predvídanie klinických výsledkov, monitorovanie pacientov v reálnom čase, odhadovanie rizikových faktorov a znižovanie chýb lekárov v zdravotníckych zariadeniach. Napríklad pomocou informácií z elektronických zdravotných záznamov, ako je anamnéza, liečba alebo laboratórne výsledky, môže AI predpovedať, či by pre pacienta bol prínosný konkrétny lekárske postup (napr. chirurgický zákrok) alebo terapia (napr. chemoterapia) a odhadnúť potenciálne riziko pooperačných komplikácií, rozvoja špecifických ochorení, prežitia septického šoku a potreby opätovného prijatia do nemocnice alebo dokonca predpovedať úmrtnosť hospitalizovaných pacientov. Aj keď sme len v počiatočnom štádiu využívania AI v biologických vedách, mnohé vyspelé krajiny budú schopné túto technológiu aplikovať prakticky ihneď vďaka vybudovanej výskumnej infraštruktúre a dostupnosti potrebných dát. Slovensko ešte stále stojí pred výzvou túto štruktúru dobudovať, preto je aktuálne skvelou investičnou príležitosťou plán obnovy, ktorého hlavným cieľom je podporiť reformy a investície, ktoré Slovensku umožnia opäť dobiehať životnú úroveň priemeru EÚ. Cieľom je podporiť transformáciu kľúčových sektorov ekonomiky s vysokým inovačným potenciálom, ktorým sektor biomedicíny jednoznačne je. Osobne preto verím a dúfam, že táto finančná podpora bude využitá aj na vznik a budovanie siete biobáň, čo umožní vedcom realizovať kvalitnejší výskum, dúfajme v budúcnosti využívajúci AI, aj na Slovensku.

Prajem vám príjemné čítanie.



Mgr. Petra Hirjaková
MEDIREX GROUP ACADEMY n. o.

Pôvodné práce

- Analýza lipofilných kvasiniek *Malassezia* spp. v klinických vzorkách** 5
Martina Čupajová, Martina Sládeková, Veronika Kadličeková, Silvia Bokorová, Miroslava Póczová, Tomáš Szemes
- Mikrobiálna biodiverzita z nazofaryngálnych výterov na podklade metatranskriptomovej analýzy: pilotná štúdia** 10
Michaela Hýblová, Dominik Hadzega, Klaudia Babišová, Patrik Krumpolec, Andrej Gnip, Peter Sabaka, Gabriel Minárik
- Hodnotenie početnosti, pozitivity a Ct hodnôt SARS-CoV-2 detekcie na Slovensku v čase pandémie COVID-19 – deti verus dospelí jedinci** 17
Nikola Janoštiaková, Andrej Gnip, Dominik Kodada, Gabriela Bľandová, Emília Miková, Elena Tibenská, Vanda Repiská, Gabriel Minárik
- Zber klinických údajov a biologických vzoriek na využitie v základnom a aplikovanom výskume** 21
Petra Hirjaková, Marianna Jagelková, Michal Lichvár, Ľubomír Arnold, Gabriel Minárik
- Clinically relevant RNA-virus whole genome assembly testing on metatranscriptomic data gained directly from COVID-19 clinical samples** 28
Dominik Hadzega, Klaudia Babisova, Patrik Krumpolec, Michaela Hyblova, Oliver Petrovic, Gabriel Minarik

Prehľadové práce

- Genetika vybraných zriedkavých atypických syndrémov parkinsonizmu** 33
Natália Forgáčová, Juraj Gazdarica, Monika Buchalová, Jozef Martiš, Tomáš Szemes
- Neinvasívny skrining onkologických ochorení pomocou biomarkerov vo voľne cirkulujúcej DNA** 36
Zuzana Holešová, Juraj Gazdarica, Jaroslav Budiš, Ondrej Pös, Tomáš Szemes
- Tandem repeat motif characterization for precision medicine: A brief overview of conventional methods and massive parallel sequencing approaches** 40
Ingrid Lojova, Jaroslav Budis, Tomas Szemes, Monika Buchalova, Jan Radvanszky
- Zmena expresie génov spojených s imunitnou odpoveďou u pacientov s COVID-19** 45
Klaudia Babišová, Patrik Krumpolec, Michaela Hýblová, Gabriel Minárik
- Skóre polygénneho rizika a jeho využitie pre odhad kumulatívneho rizika ochorení diabetu mellitu a COVID-19** 49
Patrik Krumpolec, Klaudia Babišová, Michaela Hýblová, Gabriel Minárik

Kazuistiky

- Urothelial carcinoma of the bladder with adamantinoid (ameloblastoma-like) features. Case report** 52
Michal Zámečník, Peter Kohútek

Ročník 14, 2023, číslo 1
Vychádza 3x ročne
Dátum vydania: jún 2023

Redakčná rada:

Šéfredaktorka:

RNDr. Elena Tibenská, PhD.

Členovia:

Dr. Domenico Pangallo, DrSc.
Dr. Boris Klempa, DrSc.
RNDr. Martin Radina
prof. MUDr. Milan Buc, DrSc.
prof. RNDr. Ján Turňa, CSc.
Ing. Pavol Sulo, CSc.
Dr. Bálint Nagy, PhD.
Dr. Branislav Zagrapan, PhD.
Dr. Tomas Kovacic
Dr. Gabriela Gubo
Dr. Štefan Lukáč
Mgr. Peter Baráth, PhD.
RNDr. Ján Radvánszky, PhD.
Dr. Jaroslav Bojňanský
Dr. Katarína Vlniešková

Vydavateľ:

MEDIREX GROUP ACADEMY n. o.

Sídlo vydavateľa a adresa redakcie:

MEDIREX GROUP ACADEMY n. o.
Novozámocká 1/67, 949 05 Nitra
redakcia@newslab.sk, www.newslab.sk
IČO: 37 98 68 05

Výkonná redaktorka:

Mgr. Martina Šoltésová

Technická redaktorka:

Ing. Danica Paulenová

Grafická úprava:

Ing. Lukrécia Schiller

Všetky články sú plne recenzované.

Registrácia MK SR pod číslom EV 5164/15

ISSN 1338-9661 (tlačené vydanie)

ISSN 2454-0021 (online)

Časopis je indexovaný v Bibliographia medica Slovaca (BMS).

Citácie sú spracované v CiBaMed.

Citacia skratka: Newslab

Na informačné a nekomerčné účely je dovolené články v publikácii newslab voľne a bezplatne sťahovať, kopírovať, distribuovať, tlačiť, prehľadávať a odkazovať na ne. Je však potrebné, ak je to podľa spôsobu použitia možné, uvádzať ich autora a zdroj. Použitie článkov na iné účely a iným spôsobom je možné len so súhlasom vydavateľa.

Vydavateľ nenesie zodpovednosť za údaje a názory autorov jednotlivých článkov či inzerátov.

Analýza lipofilných kvasiniek *Malassezia* spp. v klinických vzorkách

Martina Čupajová¹, Martina Sládeková¹, Veronika Kadličeková¹, Silvia Bokorová^{2,3},
Miroslava Póczová¹, Tomáš Szemes^{2,3}

¹Medirex Group, a. s., člen MEDIREX GROUP, úsek mykológie, Bratislava

²Geneton, s. r. o., Bratislava

³Vedecký park Univerzity Komenského, Bratislava

Lipofilné kvasinky *Malassezia* spp. patria medzi komenzálne mikroorganizmy na povrchu kože zdravého človeka aj zvierat. Pre ich rast sú esenciálne lipidy, ktoré musia prijímať z vonkajšieho prostredia, preto kolonizujú najmä tie časti kože, kde je obsah mastných kyselín vyšší. Vplyvom endo- a exogénnych faktorov často dochádza k ich premnoženiu, prípadne až k ochoreniu *Pityriasis versicolor*. Identifikácia spočíva najmä v mikroskopick kej analýze vzoriek a ich následnej kultivácii. Cieľom práce bol náš záujem a pokus o určenie druhového zastúpenia lipofilných kvasiniek v 48 náhodne vybraných vzorkách. Klinické vzorky pochádzali od pacientov z celého Slovenska. Metódou MALDI-TOF sa nám z vybraných izolátov podarilo druhovo určiť len 1 vzorku, sekvenovaním podľa Sangeru to bolo 84 % izolátov. K najčastejšie izolovaným druhom patrili: *M. sympodialis*, *M. globosa*, *M. slooffiae*, *M. restricta* a *M. furfur*.

Kľúčové slová: *Malassezia* spp., *Pityriasis versicolor*, MALDI-TOF, sekvenovanie

The analysis of the lipophilic yeasts *Malassezia* spp. in the clinical samples

Lipophilic yeasts *Malassezia* spp. are normal inhabitants of the superficial epidermis of healthy humans and animals. The external lipids are essential for their growth. Therefore, they mainly colonize those parts of the skin where the content of fatty acids is higher. Endo and exogenous factors can cause their proliferation, even *Pityriasis versicolor* disease. The main part of identification is the microscopic analysis of the samples and their cultivation. The aim of the work was our interest and the attempt to determine the species of lipophilic yeasts in 48 randomly selected samples. The clinical samples were isolated from patients from almost the whole of Slovakia. Using the MALDI-TOF method, we determined the species of only 1 sample from the selected isolates. By Sanger sequencing, it was 84 % of the isolates. The most frequently isolated species included: *M. sympodialis*, *M. globosa*, *M. slooffiae*, *M. restricta* and *M. furfur*.

Keywords: *Malassezia* spp., *Pityriasis versicolor*, MALDI-TOF, sequencing

NewsLab, 2023; roč. 14 (1): 5 – 9

Úvod

Povrchové mykózy sú bežne rozšírené na celom svete. Predpokladá sa, že postihujú 20 – 25 % svetovej populácie, pričom sa ich výskyt stále zvyšuje⁽¹⁾. Sú to ochorenia kože a kožných adnexov. Vyvolávatelmi týchto dermatomykóz sú 3 hlavné skupiny mikromycét: dermatofyty, oportúnne hyfomycéty a kvasinky vrátane lipofilných. Lipofilné kvasinky *Malassezia* spp. sú súčasťou normálneho kožného mikrobiómu. Vyskytujú sa prirodzene na povrchu zdravej kože nielen človeka, ale aj mnohých zvierat⁽²⁾. Kožná kolonizácia do určitej miery varíruje v závislosti od veku. Vplyvom mnohých endo a exogénnych faktorov môže tak dôjsť k ich premnoženiu a následnému rozvoju ochorenia. Sú to lipofilné kvasinky, čo znamená závislé od exogénnych lipidov, pretože im chýbajú gény syntézy mastných kyselín, okrem druhu *M. pachydermatis*^(3,4). V súčasnosti je do rodu zaradených 18 druhov, ktoré boli izolované zo zdravej aj chorej ľudskej a zvieracej kože⁽⁵⁾.

Bunky *Malassezia* spp. sa vyznačujú rôznym tvarom – guľovitými, oválnymi alebo cylindrickými. Pučanie je zvyčajne monopolárne. Bunková stena je viacvrstvová, pomáha chrániť pred rôznymi environmentálnymi vplyvmi a tiež vyháňať sa

fagocytóze. Hlavnými zložkami bunkovej steny sú sacharidy (70 %), proteíny (10 %), lipidy (15 – 20 %) a malé množstvá dusíka a síry⁽⁶⁾. Patofyziológia kožných ochorení spôsobených *Malassezia* spp. je do značnej miery neznáma vzhľadom na komplexné interakcie tohto komenzála s kožou. V zdravej pokožke sa nachádzajú kvasinky *Malassezia* spp. bez klinickej zmeny. Keď dôjde k pôsobeniu exo- alebo endogénnych faktorov na rast kvasiniek, tie sa prispôbujú modifikáciou expresie enzýmov podieľajúcich sa na získavaní energie (lipázy, fosfolipázy)⁽⁷⁾.

Hodnotenie antifungálnych mikrobiologických profilov pre tieto kvasinky je náročné, a doteraz nebola vyvinutá referenčná metóda. Veľkosť inokula, inkubačný čas a kritériá použité na určenie minimálnej inhibičnej koncentrácie (MIC) sa v jednotlivých štúdiách líšia. Vo všeobecnosti, *Malassezia* spp. sú citlivé na lokálne a perorálne prípravky s keratolytickými vlastnosťami, ako aj antifungálnym účinkom. Medzi prvé a často účinné liečebné preparáty patria šampóny a krémy so soľami selénu a zinku, s propylénglykolom a so sírovými zlúčeninami. Topická liečba môže tiež obsahovať špeciálne antimykotiká vrátane azolov a terbinafínu. Všetky druhy

Obrázok 1. Klinické prejavy *Pityriasis versicolor* (i1)



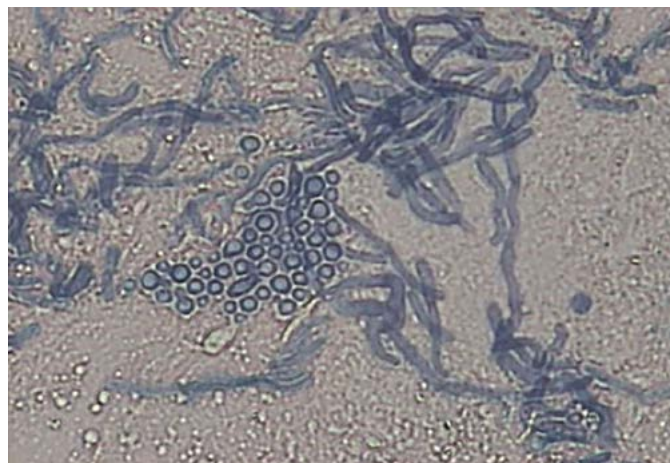
tohto rodu sú prirodzene rezistentné proti echinokandínom a griseofulvínu. Niektoré druhy tiež vykazujú vysoké hodnoty MIC pre ketokonazol, ktorý je odporúčaný ako liečivo pri kožných ochoreniach⁽⁸⁾.

Klinické infekcie spôsobené lipofilnými kvasinkami postihujú prevažne kožu, pričom najčastejším prejavom je ochorenie *Pityriasis versicolor* (PV). Ďalšie kožné ochorenia, pri ktorých hrajú úlohu druhy *Malassezia* spp., sú seboroická dermatitída, atopická dermatitída, rôzne folikulitídy, spolupôsobia aj pri prejavoch akné, rozacei^(9,10,11). Vo všeobecnosti platí, že kožné infekcie vyvolané *Malassezia* spp. sú často chronické a u citlivejších jedincov sa vyskytujú opakovane. Imunosupresívni pacienti majú vyššiu mieru recidívy. Špecifické environmentálne expozície môžu rovnako zlepšiť alebo zhoršiť symptómy⁽¹²⁾.

Pityriasis versicolor (PV) je chronická superficiálna kožná infekcia postihujúca rohovú vrstvu kože – *stratum corneum*. V patogenéze sa ako predispozičné faktory uplatňujú najmä zvýšená potivosť, natieranie kože masťnými prípravkami (oleje, krémy...), celková liečba kortikoidmi a imunosupresívami, nedostatočná výživa a mnohé ďalšie. Ochorenie je prenosné z človeka na človeka buď priamym kontaktom, alebo aj nepriamo. Kontagióznosť je však nízka, pre vznik a rozvoj infekcie je potrebná aj určitá individuálna vnímavosť⁽¹³⁾. Oblasť výskytu ochorenia je od mierneho klimatického pásma až po tropické podnebie, pričom v teplom a vlhkom prostredí sa PV vyskytuje častejšie. Uvádza sa, že v týchto oblastiach na ochorenie trpí až 50 % populácie. Nie sú pozorované výraznejšie rozdiely medzi pohlaviami. Podľa vekového rozdelenia sú častejšie postihnutí dospelí a mladí dospelí, ale ochorenie sa môže vyskytnúť v akomkoľvek veku^(13,14). PV sa vo všeobecnosti vyskytuje aj u zdravých jedincov. Okrem toho existujú údaje o asociácii medzi PV a ďalšími ochoreniami, napr. *diabetes mellitus*, poruchy činnosti štítnej žľazy a iné. PV sa vyznačuje veľmi typickým mikroskopickým obrazom šupín z kožných lézií⁽¹⁵⁾.

Ochorenie sa prejavuje splyvajúcimi, šupinatými, tmavými alebo depigmentovanými škvrnami nachádzajúcimi sa v hornej časti trupu siahajúcimi na krk, brucho, chrbát a ramená, zriedka inde⁽²⁾. PV sa môže prejavovať aj v axilách, na slabínach, stehnách či genitáliách a takisto môže ochorenie postihovať predlaktia a chrbty rúk. Primárne lézie sú dobre ohraničené

Obrázok 2. *Pityriasis versicolor*, vzorka DE 21702/2022 x400 (autor)



makuly, ktoré môžu byť mierne erytematózne a pokryté jemnými šupinami. Tie môžu byť však viditeľné až po škriabnutí povrchu lézie. Môžu sa spájať a vytvárať rozptýlené škvrny hypo- alebo hyperpigmentácie alebo splyvať do väčších mapovitých ložísk (**obrázok 1**)⁽¹⁶⁾.

Lézie môžu pod UV (Woodovým) svetlom vykazovať svetložltú fluorescenciu. Základom je priama mikroskopia. V preparáte možno pozorovať mycélium, krátke hýfy spolu s guľovitými hrubostennými kvasinkami (**obrázok 2**). Zriedka môžu byť pozorované len oválne kvasinky. Charakteristický vzhľad pod mikroskopom bol opísaný ako „meatballs and spaghetti“. Druhá identifikácia *Malassezia* spp. kultivačnými alebo molekulárnymi metódami nie je súčasťou bežného diagnostického vyšetrenia PV⁽¹⁷⁾.

Materiál a metódy

V práci sme použili klinické vzorky pacientov pochádzajúcich takmer z celého Slovenska. Vzorky z kožných šupín (najmä z oblastí trupu, hlavy, ramien atď.) boli zozbierané v období jún až december 2022.

Všetky vzorky sme v prvom kroku analyzovali mikroskopicky – natívnym lúhovým preparátom s 20 % hydroxidom draselným. Následne sme vzorky kultivovali na selektívnej pôde ŽOT (obsahujúca žlč, olivový olej a TWEEN) pri 30 °C počas 7 – 10 dní. Z pozitívnych kultivačných nálezov potvrdených mikroskopicky laktofenolovým preparátom sme náhodne vybrali 48 izolátov, ktoré sme použili na druhovú identifikáciu metódami MALDI-TOF a sekvenovaním podľa Sangera.

MALDI-TOF – je metóda identifikácie mikroorganizmov pomocou laserovej desorpcie/ionizácie za účasti matrice. Princíp metódy spočíva v generovaní hmotnostných spektrálnych profilov ribozómových proteínov mikroorganizmov, ktoré sa následne porovnávajú s hmotnostnými spektrami v databáze⁽¹⁸⁾. Pri analýze mikromycét je okrem štandardného postupu potrebný medzikrok s použitím kyseliny mravčej⁽¹⁹⁾.

Sekvenácia enzymatickou metódou podľa Sangera – metóda je založená na selektívnej inkorporácii dideoxynukleotidov pomocou DNA polymerázy v priebehu replikácie DNA, čím dochádza k terminácii polymerizačnej reakcie⁽²⁰⁾. Na izoláciu DNA zo vzoriek sme použili komerčne dostupný kit – Fungi/Yeast Genomic DNA isolation kit (NORGEN, Biotec Corporation). Z metód molekulárnej biológie sme aplikova-

Obrázok 3. *Malassezia* spp. na ŽOT 10 dní/30 °C (autor)

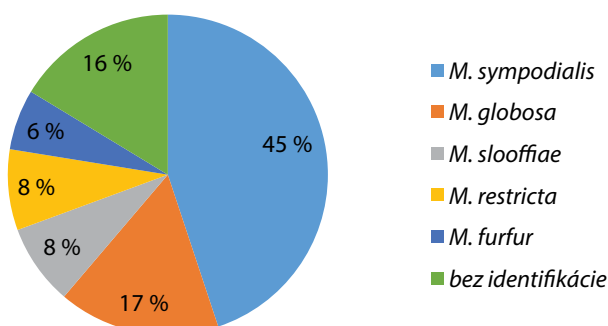
li klasickú polymerázovú reťazovú reakciu (PCR) s použitím primerov špecifických na konzervované úseky DNA – ITS gény. Následne sme PCR produkty elektroforeticky separovali v 1,5 % agarózovom géli pri napätí 120 V 20 – 30 min s cieľom zistiť veľkosť produktu s použitím fluorescenčnej farbičky Serva DNA stain G. Vzorky sme po prečistení ExoSapom fluorescenčne kvantifikovali.

Na Sanger PCR sme použili BigDye™ Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific) modifikovaný Vedeckým parkom UK.

Získané .ab1 súbory sme analyzovali v programe Chromas Lite a získané sekvencie sme porovnávali s dostupnými známymi sekvenciami pomocou databázy BLAST (i2).

Výsledky

V natívnom lúhovom preparáte s 20 % hydroxidom draselným sme pozorovali v 17 vzorkách obraz typický pre *Pityriasis versicolor*. V 4 vzorkách sme nepozorovali žiadne konídie *Malassezia* spp. a v ostatných vzorkách boli prítomné len konídie. Všetky kultivačné nálezy sme overili mikroskopickou analýzou v laktofenolovom preparáte s 200- a 400-násobným zväčšením. Makroskopicky sa kolónie izolátov líšili štruktúrou (hladké, zvrásnené), veľkosťou (1 – 5 mm), farebnou (svetlokrémové až tmavšie) (obrázok 3). Bunky *Malassezia* spp. tiež vykazovali rozdielnu veľkosť a tvar. Veľkosti boli v rozmedzí 2,5 – 4,5 μm a ich tvar sme opísali ako guľovitý, oválny až cylindrický. Bežne sme pozorovali aj monopólarne pučanie.

Graf 1. Výsledky druhového zastúpenia *Malassezia* spp.

Tabuľka 1. Výsledky sekvenovania

Vzorka	Výsledok	Referenčný kmeň	Pokrytie sekvencie (%)	% identifikácie
DE 10866	<i>M. sympodialis</i>	ATCC 42132	100	99,9
DE 10934	Bez identifikácie			
DE 11090	<i>M. sympodialis</i>	KS292	100	99,89
DE 11238	<i>M. sympodialis</i>	ATCC 42132	100	99,9
DE 11460	<i>M. sympodialis</i>	KS292	100	99,89
DE 11646	<i>M. sympodialis</i>	KS292	100	100
DE 12164	<i>M. slooffiae</i>	CBS 7956	100	95,14
DE 12452	<i>M. sympodialis</i>	ATCC 42132	100	100
DE 12816	<i>M. furfur</i>	CBS14141	99	98,64
DE 13218	<i>M. furfur</i>	CBS14141	100	99,9
DE 19724	<i>M. restricta</i>	KCTC 27527	99	98,82
DE 19786	<i>M. globosa</i>	CBS7966	100	98,73
DE 19860	<i>M. globosa</i>	CBS7966	100	98,1
DE 19862	<i>M. globosa</i>	CBS7966	100	98
DE 19952	<i>M. sympodialis</i>	KS292	100	99,56
DE 20012	<i>M. slooffiae</i>	CBS:7956	81	99,87
DE 20090	<i>M. sympodialis</i>	KS292	100	99,69
DE 20178	<i>M. globosa</i>	CBS7966	100	96,7
DE 20240	<i>M. globosa</i>	CBS7966	100	98,02
DE 20406	<i>M. sympodialis</i>	KS292	100	99,89
DE 20444	<i>M. restricta/ M. globosa</i>	KCTC 27527/ TN1512	99/ 60	88,24/ 81,54
DE 20498	<i>M. globosa</i>	CBS7966	100	97,16
DE 20506	<i>M. sympodialis</i>	ATCC 42132	100	100
DE 20554	Bez identifikácie			
DE 20612	<i>M. slooffiae</i>	CBS:7956	78	99,73
DE 20614	<i>M. sympodialis</i>	ATCC 42132	100	99,78
DE 20644	<i>M. restricta</i>	CBS 7877	99	99,65
DE 20774	<i>M. slooffiae</i>	CBS:7956	78	93,76
DE 20890	<i>M. restricta</i>	CBS 7877	100	99,78
DE 20896	<i>M. sympodialis</i>	KS292	100	99,89
DE 21130	Bez identifikácie			
DE 21302	<i>M. sympodialis</i>	ATCC 42132	100	99,49
DE 21456	<i>M. sympodialis</i>	KS292	100	99,9
DE 21560	<i>M. globosa</i>	CBS7966	100	97,93
DE 21574	<i>M. sympodialis</i>	KS292	100	99,89
DE 21690	<i>M. sympodialis</i>	KS292	100	99,9
DE 21702	<i>M. sympodialis</i>	ATCC 42132	100	100
DE 21704	<i>M. sympodialis</i>	KS292	100	99,59
DE 23066	<i>M. sympodialis</i>	KS292	99	94,85
DE 23166	Bez identifikácie			
DE 23310	Bez identifikácie			
DE 23318	Bez identifikácie			
DE 23382	<i>M. sympodialis</i>	KS292	100	99,69
DE 23546	<i>M. sympodialis</i>	KS292	100	99,69
DE 23942	<i>M. sympodialis</i>	ATCC 42132	100	99,58
DE 24242	Bez identifikácie			
DE 24376	<i>M. furfur</i>	CBS14141	100	99,36
DE 24748	Bez identifikácie			

Metódou MALDI-TOF sme druhovo identifikovali len 1 izolát. Bola to vzorka šupín z oblasti trupu a mikroskopicky sme v natívnom lúhovom preparáte pozorovali obraz typický pre PV. Výsledkom identifikácie bol druh *M. furfur* s identifikačným skóre 1,83, čo predstavuje úspešnú identifikáciu, keďže tento interval sa pri kvasinkách pohybuje v rozmedzí 1,7 a vyššie⁽²¹⁾.

Výsledky sekvenčnej metódy podľa Sangera sú zobrazené v **tabuľke 1**. V našom súbore pacientov sme izolovali 5 rôznych druhov *Malassezia* spp.

Percentuálne zastúpenie jednotlivých druhov môžeme vidieť na **grafe 1**. Najpočetnejším druhom v našom súbore bola *M. sympodialis*, tvorí takmer polovicu všetkých kmeňov. Za ňou nasleduje *M. globosa* vyskytujúca sa v pätine izolátov. Druhy *M. slooffiae*, *M. restricta* a *M. furfur* sú zastúpené približne rovnako v 6 – 8 %. Z grafu je zrejmé, že až 16 % izolátov sa nám nepodarilo určiť.

Diskusia a záver

V tejto práci sme sa zamerali na analýzu lipofilných kvasiniek *Malassezia* spp. v klinických vzorkách. Hoci sú tieto kvasinky bežnou súčasťou kože, vlasov, chlpov, už ich rodová identifikácia je pomerne náročná. Bežná identifikácia spočíva v dvoch krokoch – mikroskopia a kultivácia. Prvým bola mikroskopická analýza kožných šupín lúhovým preparátom. Preparáty sme hodnotili pri 200-, 400- a 1 000-násobnom zväčšení. V mikroskopickom obraze sme pozorovali buď len modrasté často aj pučiacie konídie *Malassezia* spp., alebo aj krátke hýfy. Typický obraz PV sme vyhodnotili v 35 % vzoriek. Naopak, negatívnu mikroskopiu sme mali približne v 8 % vzoriek. Podľa Rhodoplu a kol. (2014) je miera pozitivity priamej mikroskopickej analýzy v rozmedzí 46,65 – 100 %. Naša miera pozitivity 92 % sa nachádzala v danom rozmedzí. Kultiváciu sme hodnotili po 7 – 10 dňoch a pozitívitu potvrdili laktófenolovým preparátom prítomnosťou konídií.

Cieľom tejto práce bol náš záujem a pokus o druhové zastúpenie lipofilných kvasiniek, ktoré sa v rutinnej klinickej diagnostike nevykonáva. Pokúsili sme sa o ňu dvomi metodikami. Metódou MALDI-TOF, bežne využívanou na identifikáciu baktérií a čiastočne aj kvasiniek a metódou sekvenovania podľa Sangera uskutočnenou vo Vedeckom parku Univerzity Komenského.

Výsledky metódy MALDI-TOF sú založené na porovnávaní hmotnostných spektier ribozómových proteínov mikroorganizmov s databázou. Druhové spektrum lipofilných kvasiniek v tejto databáze je však veľmi úzke, obsahuje len dva druhy – *M. furfur* a lipidovo nezávislý druh *M. pachydermatis*. Z tohto dôvodu sme metódou MALDI-TOF významne určili len jeden izolát.

Molekulárnobiologickou metódou sekvenovania podľa Sangera sa nám nepodarilo určiť ani všetkých 48 izolátov. Sanger PCR sme druhovo určili 84 % izolátov. Najpočetnejšie zastúpenie mala *M. sympodialis* v 45 % izolátov. Romano a kol. (2013) vo svojej práci uvádza, že najčastejšie izolovaným druhom v Taliansku je *M. globosa* približne v 65 % vzoriek. Rozdiely v druhovom zastúpení môžu súvisieť s klimatickými rozdielmi, metódou odberu (zoškrab, ster), zložením kultivačného média, prípadne spôsobom identifikácie. V 47 prípadoch sme izolovali 1 druh, iba v jednej vzorke bola izolovaná zmesová kultúra 2 druhov *Malassezia* spp. Tento údaj nám nekoreluje so štúdiou Romano a kol. (2013), ktorá opisuje až v 38 % výskyt dvoch druhov *Malassezia* spp. izolovaných z rovnakej lézie. Neúspešná druhová identifikácia bola pri 8 izolátoch, čo predstavuje približne pätinu vzoriek. Pri týchto izolátoch sme sa pokúsili jednotlivé metódy modifikovať: opakovaná izolácia DNA, použitie iných master mixov, opätovné prečisťovanie vzoriek a využitie vysokošpecifickej DNA polymerázy. Kmene sa nám však napriek týmto snahám nepodarilo určiť. Dôvodom môže byť, že kvasinky ako eukaryotické mikroorganizmy majú zložitejšiu štruktúru bunkovej steny. Z tohto dôvodu môže byť problematická už prvotná izolácia DNA, najmä v procese lýzy bunky.

Pre vyššiu úspešnosť identifikácie by bolo potrebné optimalizovať podmienky reakcií. Keďže je druhové určenie lipofilných kvasiniek časovo aj finančne veľmi náročné, v rutinnej klinickej praxi nie je potrebné ani sa bežne nevyužíva. Touto prácou sa nám potvrdil fakt, že molekulárne metódy sú v druhovej detekcii *Malassezia* spp. jednoznačne citlivejšie ako kultivačné (fenotypové). Na základe našich výsledkov sme získali aspoň orientačnú predstavu, aké druhy *Malassezia* spp. sú v našich klimatických podmienkach prevládajúce.

Podakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekty ITMS: 313011ATL7 (PanClinCov) a ITMS: 313011W428 (BIOMEDIRES II) spolufinancované zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

- Havlickova B, Czaika VA, Friedrich M. Epidemiological trends in skin mycoses worldwide. *Mycoses*. 2008; 51: 2–15.
- Saunte DML, Gaitanis G, Hay RJ. Malassezia-Associated Skin Diseases, the Use of Diagnostics and Treatment. *Front Cell Infect Microbiol*. 2020; 10: 112.
- Glatz M, Bosshard PP, Hoetzenecker W, et al. The role of Malassezia spp. in atopic dermatitis. *J. Clin. Med*. 2015; 4: 1217–1228.
- Mahmoudabadi AZ, Zarrin M, Azish M. Detection of Malassezia species isolated from patients with pityriasis versicolor and seborrheic dermatitis using Nested-PCR. *Jentashapir J Health Res*. 2014; 5(6): e26683.
- Guillot J, Bond R. Malassezia yeasts in veterinary dermatology: an updated overview. *Front Cell Infect. Microbiol*. 2020; 10: 79.
- Celis AM, Wösten HAB, Triana S, et al. Malassezia spp. beyond the mycobiota. *SM Dermatol. J*. 2017; 3: 1019-1–1019-10.
- Sun S, Hagen F, Xu J, et al. Ecogenomics of human and animal Basidiomycetous yeast pathogens. *The Ecological Genomics of Fungi*. 2013; 215–242.
- Leong C, Buttafuoco A, Glatz M, et al. Antifungal susceptibility testing of Malassezia spp. with an optimized colorimetric broth microdilution method. *J Clin Microbiol*. 2017; 55(6): 1883–1893.
- Jorgensen JH, Pfaller MA, Carroll KC, et al. *Manual of Clinical Microbiology*. Printed in Canada. 2015; 2: 194. ISBN: 978-1-55581-737-4.
- Annisette EJ, McGinnis MR, Pfaller MA. *Clinical Mycology*. Churchill Livingstone. 2003; 1: 260–263. ISBN: 0-443-07937-4.
- Buchvald J, Buchvald D. *Dermatovenerológia. SAP – Slovak Academic Press*. 2002; str. 391–394. ISBN: 80-89104-03-7.
- VesT BE, Krauland K. *Malassezia Furfur*. Treasure Island (FL). 2022.
- Buchvald J, Sinka L, Péč J. *Dermatovenerológia*. Vydavateľstvo Osveta. 1993; str. 145. ISBN: 80-217-0498-5.
- Santana JO, De Azevedo FLA, Campos Filho PC. Pityriasis versicolor: clinical-epidemiological characterization of patients in the urban area of Buerarema-BA, Brazil. *An. Bras. Dermatol*. 2013; 88: 216–221.
- Aghaei Gharebolagh S, Kordbacheh P, Hashemi SJ, et al. MGL_3741 gene contributes to pathogenicity of Malassezia globosa in pityriasis versicolor. *Mycoses*. 2018; 61: 938–944.

16. Tellechea Ó, Cravo M, Brinca A, et al. Pityriasis versicolor atrophicans. *Eur. J. Dermatol.* 2012; 22: 287–288.
17. Mathur M, Acharya P, Karki A, et al. Dermoscopic pattern of pityriasis versicolor. *Clin. Cosmet. Investig. Dermatol.* 2019; 12: 303–309.
18. Suarez S, Ferroni A, Lotz A, et al. Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. *J. Microbiol. Methods.* 2013; 94 (3), 390–396.
19. Gorka J, Bah U, Karas M. Graphite supported preparation (GSP) of alpha-cyano-4-hydroxycinnamic acid (CHCA) for matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) for peptides and proteins. *J. Am. Soc. Mass Spectrom.* 2012; 23 (11), 1949–1954.
20. Shendure J, Ji H. „Next-generation DNA sequencing“. *Nature Biotechnology.* 2008; 26 (10), 1135–1145.
21. Becker PT, de Bel A, Martiny D, et al. Identification of filamentous fungi isolates y MALD-TOF mass spectrometry: clinical evaluation of an extended referenc spectra library. *Med Mycol.* 2014; 52: 826–834.
22. Rhodoplu G, Saracli MA, Gumral R, et al. Distribution of Malassezia species in patients with pityriasis versicolor in Turkey. *Journal de Mycologie Medicale.* 2014; 24(2): 117–123.
23. Romano C, Mancianti F, Nardoni S, et al. Identification of Malassezia species isolation from patients with extensive forms of pityriasis versicolor in Siena, Italy. *Revista Iberoamericana de Micología.* 2013; 30(4): 23–234.

INTERNETOVÉ ZDROJE

24. (i1) – <https://www.msmanuals.com/home/quick-facts-skin-disorders/fungal-skin-infections/tinea-versicolor>
25. (i2) – [https://BLAST: Basic Local Alignment Search Tool \(nih.gov\)](https://BLAST: Basic Local Alignment Search Tool (nih.gov))

RNDr. Martina Čupajová

Kozínska 92

027 05 Zázrivá

e-mail: martina.cupajova@medirex.sk

Mikrobiálna biodiverzita z nazofaryngálnych výterov na podklade metatranskriptomovej analýzy: pilotná štúdia

Michaela Hýblová¹, Dominik Hadzega¹, Klaudia Babišová¹, Patrik Krumpolec¹, Andrej Gnip¹, Peter Sabaka², Gabriel Minárik¹

¹MEDIREX GROUP ACADEMY, Nitra, Slovakia

²Department of Infectology and Geographical Medicine, Faculty of Medicine, Comenius University in Bratislava, Bratislava, Slovakia

Celosvetový výskyt pandémie SARS-CoV-2 urýchlil výskum vo viacerých oblastiach vedy, ktorého cenné výstupy a zistenia pomôžu v budúcnosti riešiť zdravotné výzvy v prípade výskytu nových vírusov. Spravili sme metatranskriptomovú analýzu z nazofaryngálnych výterov pacientov s rôznou závažnosťou ochorenia COVID-19 a negatívneho kontrolného súboru na porovnanie rozdielov v bakteriálnom spektre a prítomnosti vírusov sekvenovaním RNA nezávislého od kultivácie. Pozorovali sme výrazné zvýšenie diverzity mikrobiálnych druhov u pacientov s COVID-19 bez ohľadu na závažnosť ochorenia. Bakteriálny kmeň *Actinobacteria* je výrazne viac zastúpený u zdravých ľudí ako u infikovaných, zatiaľ čo kmeň *Bacteroides* je zastúpený výrazne menej. Infikovaní ľudia bez ohľadu na závažnosť a príznaky majú v rovnakom pomere zastúpené kmene *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* a *Fusobacteriales*. Okrem SARS-CoV-2 a početných skupín fágov sme v niekoľkých vzorkách identifikovali sekvencie klinicky významných vírusov, ako je *Human Herpes Virus 1*, *Human Mastadenovirus D*, *Molluscum Contagiosum Virus* a *Rhinovirus A*. Analýzy boli realizované retrospektívne, preto sú v prípade SARS-CoV-2 zastúpené rôzne varianty (WHO klasifikácia) ako je *alpha* (B.1.1.7), *delta* (B.1.617.2), *omicron* (B.1.1.529) a 20C.

Kľúčové slová: metatranskriptom, mikrobióm, SARS-CoV-2, viróm

Microbial biodiversity from nasopharyngeal swabs based on metatranscriptome analysis: a pilot study

The global emergence of the SARS-CoV-2 pandemic has accelerated research in several areas of science, the valuable outputs and findings of which will help to address future health challenges in the case of emerging viruses. We performed metatranscriptome analysis of nasopharyngeal swabs from patients with varying severity of COVID-19 disease and a negative control cohort to compare differences in the bacterial spectrum and viral presence via culture-independent RNA sequencing. We observed a significant increase in microbial species diversity in COVID-19 patients regardless of disease severity. Some bacterial *phyla*, such as *Actinobacteria*, are significantly more abundant in healthy subjects than in infected subjects, whereas the *Bacteroides* *phyla* are less abundant. Infected people, regardless of severity and symptoms, have *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Fusobacteriales* represented in equal proportions. In addition to SARS-CoV-2 and numerous phage groups, we identified sequences of clinically significant viruses such as *Human Herpes Virus 1*, *Human Mastadenovirus D*, *Molluscum Contagiosum Virus*, and *Rhinovirus A* in several samples. The analyses were performed retrospectively; therefore, different variants (WHO classification) are represented in SARS-CoV-2, such as *Alpha* (B.1.1.7), *Delta* (B.1.617.2), *Omicron* (B.1.1.529) and 20C.

Keywords: metatranscriptome, microbiome, SARS-CoV-2, virome

NewsLab, 2023; roč. 14 (1): 10 – 16

Úvod

Denne do seba vdýchneme asi milión mikrobiálnych častíc. Horné dýchacie cesty sú hlavnou vstupnou bránou pre veľké množstvo vírusov, a to nielen respiračných. Sliznica nosovej dutiny je zároveň kolonizovaná širokým spektrom mikroorganizmov, ktorých špecifický význam nie je dobre objasnený. Pod vplyvom nedávnej pandémie COVID-19 sa do popredia dostal záujem o to, ako vírusová infekcia ovplyvňuje celkový mikrobióm dýchacích ciest. V súčasnosti sa intenzívne študuje mikrobiálna diverzita v kontexte závaž-

nosti ochorenia COVID-19, žiaľ, často s kontroverznými výsledkami⁽¹⁻³⁾.

Spomedzi baktérií prevládajú vo väčšine skúmaných anatomických oblastí vrátane horných dýchacích ciest (HDC) kmene *Bacteroidetes* a *Firmicutes*, ktoré sú najviac skúmanými zložkami ľudského mikrobiómu⁽⁴⁾.

Mikrobióm nosovej dutiny zdravého človeka je zložený predovšetkým z kmeňov *Actinobacteria*, *Bacteroidetes*, *Firmicutes* a *Proteobacteria* s prevahou zástupcov rodov *Bifido-*

Tabuľka 1. A. študovaná kohorta, charakteristika vek a pohlavie; **B.** Ct (len ťažkí, mierni, asymptomatickí pacienti)

A	počet	M	Ž	vek (medián)	SD	M	SD	Ž	SD
negatívne kontroly	72	26	46	37 (25-75)	10,59	42 (25-58)	9,05	35 (25-75)	11,45
ťažkí	30	16	14	68 (32-90)	13	69.5 (32-90)	13,6	64 (41-77)	11,7
mierni	25	13	12	37 (17-57)	11,83	38 (17-57)	12,77	32.5 (19-57)	10,54
asymptomatickí	24	14	10	42 (20-49)	7,86	39 (21-49)	9,35	43 (36-48)	3,72
	151	69	82	39,5		40,5		39	
		45,70 %	54,30 %						

B	Ct hodnoty E gén (medián)	SD
ťažkí	28.06 (13.9-37.3)	7,34
mierni	24.89 (15-38.77)	5,32
asymptomatickí	28.56 (19.24-33.59)	4,72

M – muži, Ž – ženy, SD – štandardná odchýlka (standard deviation), Ct – threshold cyklus v RT-PCR

bacterium, Corynebacterium, Staphylococcus, Streptococcus, Dolosigranulum a Moraxella^(5,6).

O niečo menej známym faktom je, že ľudia sú tiež kolonizovaní pozoruhodným množstvom DNA a RNA vírusov, ktoré možno označiť ako viróm. Mnohé z nich sú schopné spôsobovať akútnu, perzistentnú alebo latentnú infekciu hostiteľa. Ľudský viróm pozostáva hlavne z bakteriofágov (fágov), ktoré infikujú baktérie, archeálnych vírusov, eukaryotických vírusov, ktoré infikujú ľudské bunky, a vírusov prechodne sa vyskytujúcich v potravinách^(7,8). Spomedzi identifikovaných vírusových genómov sa v horných dýchacích cestách najviac vyskytujú pikornavírusy, anellovírusy a bakteriofágy z čeľade *Siphoviridae*⁽⁹⁾.

Vo všeobecnosti existujú na identifikáciu mikroorganizmov dva hlavné sekvenčné prístupy: amplikónové sekvenovanie 16S rRNA a metagenomické shotgun sekvenovanie. Prvé z nich je atraktívne nižšími nákladmi, uprednostňuje sa v prostredí s nižšou mikrobiálnou diverzitou, a identifikuje najmä dominantné mikroorganizmy⁽¹⁰⁾. Druhá metóda sa vyznačuje robustnejším rozlíšením a pokrýva metagenómovú komunitu vrátane ľudského genómu, vírusov a húb, ktorú nie je možné zachytiť amplikónovým sekvenovaním⁽¹¹⁾. V našej štúdiu sme nahradili analýzu metagenómu metatranskriptomom, aby sme pokryli RNA vírusy, medzi ktoré patrí aj SARS-CoV-2. Nesekvenovali sme celkovú RNA, ale v počiatočných krokoch prípravy knižnice sme odstránili dominantnú ľudskú 18S rRNA, aby sme zvýšili relatívne množstvo reziduálnej ľudskej a mikrobiálnej, resp. vírusovej RNA. Vo všeobecnosti sa odhaduje, že 18S rRNA tvorí približne 80 – 90 % celkovej RNA v typickej bunke cicavcov⁽¹²⁾.

Materiál a metódy

Pacienti a vzorky

Štúdia sa uskutočnila so súhlasom Etickej komisie Bratislavského samosprávneho kraja pod identifikačným číslom 03228/2021/HF zo dňa 12. januára 2021. Všetci pacienti podpísali informovaný súhlas a dostali dotazníky s relevantnými otázkami týkajúcimi sa ich anamnézy a zdravotného stavu vo vzťahu ku COVID-19. Všetkých 151 pacientov bolo zo Slovenska a boli rozdelení do štyroch skupín: asymptomatickí (n = 24), pacienti s miernym/so stredne ťažkým (n = 25), s ťažkým/so závažným COVID-19 (n = 30) a negatívna kontrolná skupina (n = 72). Skupina pacientov so závažnou formou COVID-19 pochádzala z dvoch univerzitných nemocníc v Bratislave: Ružinov a Kramáre (Slovensko). Charakteristika

kohorty ohľadom veku a pohlavia účastníkov štúdie je zhrnutá v **tabuľke 1A**. Výtery z nosohltana boli odoberané od marca 2021 do októbra 2022 do transportného média viRNAtrap (GeneSpector, Česká republika) a do spracovania boli uchovávané pri teplote 4 °C v chladničke.

Izolácia RNA, RT-PCR a príprava knižnice

RNA bola izolovaná retrospektívne zo zberného média (z nosohltanového výteru) pomocou súpravy Sera-Xtracta™ virus/Pathogen Kit (Cytiva, UK) na stolnom automatizovanom zariadení KingFisher Flex (ThermoFisher Scientific, UK). Množstvo RNA sa meralo fluorometricky pomocou Qubit™ RNA High sensitivity (Invitrogen, USA). Izoláty RNA sa až do spracovania skladovali pri teplote –80 °C. Genomické knižnice sa pripravili pomocou súpravy Kapa HyperPrep s RiboErase Kit (Roche, USA) s depléciou eukaryotickej RNA pre 18S rRNA podľa odporúčaní výrobcu. Súčasťou prípravy knižnice je reverzný prepis RNA do cDNA v dvoch krokoch: 1. syntéza prvého vlákna cDNA pomocou náhodných primerov, 2. syntéza komplementárneho vlákna na ds cDNA pomocou dUDP. Na indexovanie vzoriek sa použili duálne adaptéry TruSeq CD spoločnosti Illumina. Výsledné knižnice sa kvantifikovali fluorometrickou analýzou pomocou súpravy Qubit™ dsDNA HS Assay (Invitrogen, USA) a fragmentovou analýzou pomocou súpravy čidiel High Sensitivity DNA (Agilent Technologies, Litva).

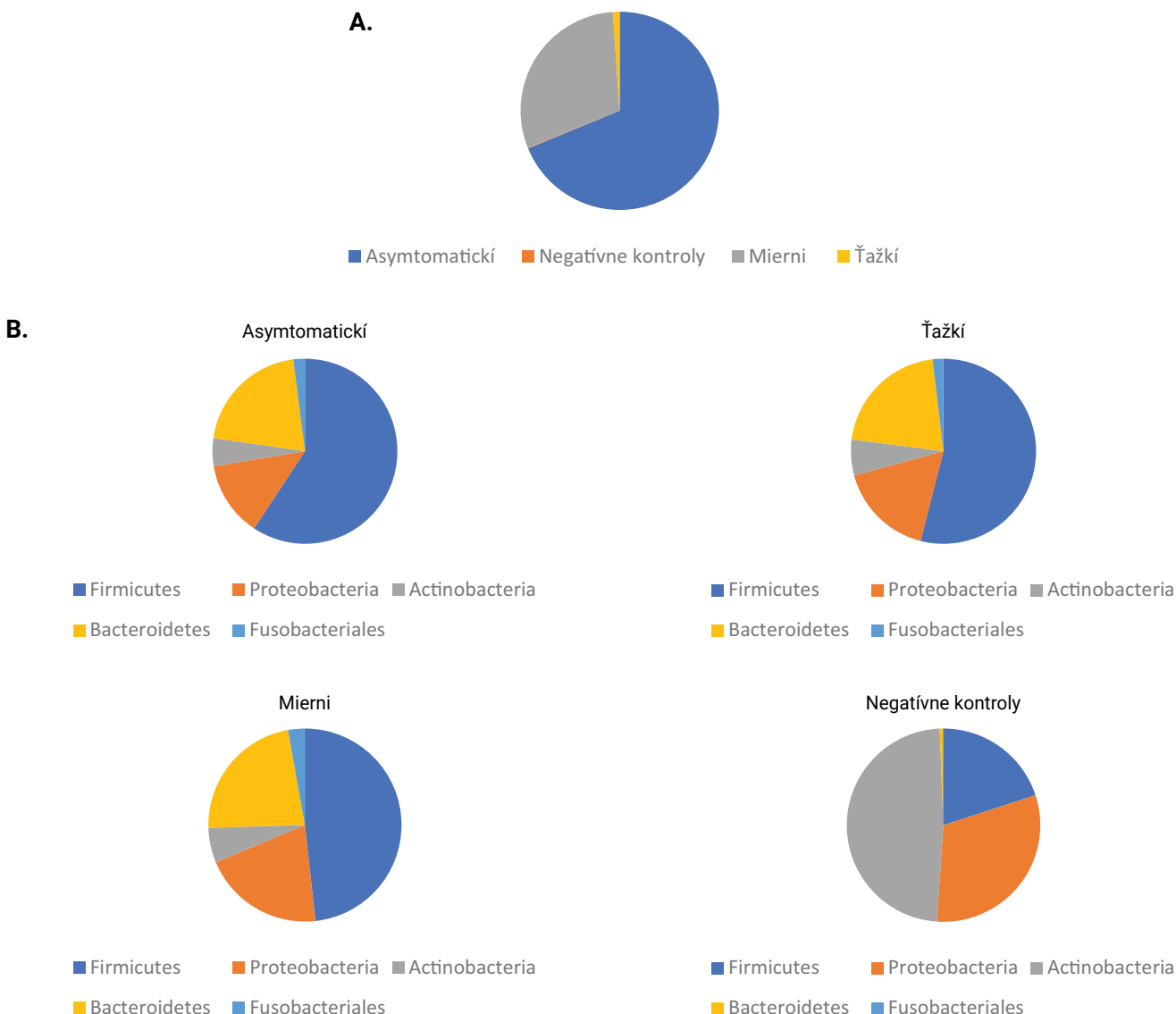
Všetky RNA sa podrobili RT-PCR na potvrdenie alebo vylúčenie pozitivity SARS-CoV-2 pomocou súpravy COVID-19 Real-Time Multiplex RT-PCR Kit (Labsystems Diagnostics, Fínsko) a platformy ABI QuantStudio 6 Real-Time PCR System RT-qPCR (ThermoFisher, USA) s použitím originálnych protokolov výrobcu. Na vyhodnotenie vzorky ako pozitívnej sa vyžadovala hodnota Ct < 40.

Sekvenovanie a bioinformatická analýza

Párové sekvenovanie – 2x 75 a 2x 100) sa realizovalo na platformách NextSeq500/550 a NextSeq2000 (Illumina, USA). Výsledné čítania sa mapovali na ľudskú referenciu hg38 (GRCh38) pomocou algoritmu BWA-MEM⁽¹³⁾. Nemapované čítania dlhšie ako 50 bp boli ďalej analyzované s cieľom identifikácie mikrobiómu. Na identifikáciu a kvantifikáciu bakteriálnych druhov sme použili Kraken 2 v2.1.2⁽¹⁴⁾.

Na skladanie vírusu SARS-CoV-2 sme použili nástroj *de novo assembly* coronaspades.py zo Spades v3.15.5⁽¹⁵⁾. Priraďovanie línií SARS-CoV-2 sa uskutočnilo podľa Maier a Batut, 2023⁽¹⁶⁾.

Obrázok 1. Koláčové grafy relatívnej početnosti baktérií vo všetkých 4 skupinách **A.** Zastúpenie hlavných bakteriálnych taxónov naprieč všetkými 4 skupinami **B.**



Štatistická analýza

Počty čítaní sa normalizovali delením celkovým počtom čítaní v danej analýze a následne sa vynásobili priemerným počtom čítaní. Rozdiel v počte zistených bakteriálnych transkriptov medzi skupinami sa štatisticky analyzoval pomocou nástroja LefSE (Galaxy verzia 1.0)⁽¹⁷⁾. Na porovnanie distribúcie bakteriálnych rodov a druhov medzi jednotlivými skupinami sme použili štatistický nástroj PERMANOVA. Mannov-Whitneyho U test (Wilcoxonov test) sa použil na porovnanie počtov čítaní zhodujúcich sa so sekvenciami z databázy Kraken2 medzi jednotlivými dvojicami skupín. Tento test sa použil aj na porovnanie hodnôt Ct medzi každou dvojicou pozitívnych skupín.

Výsledky

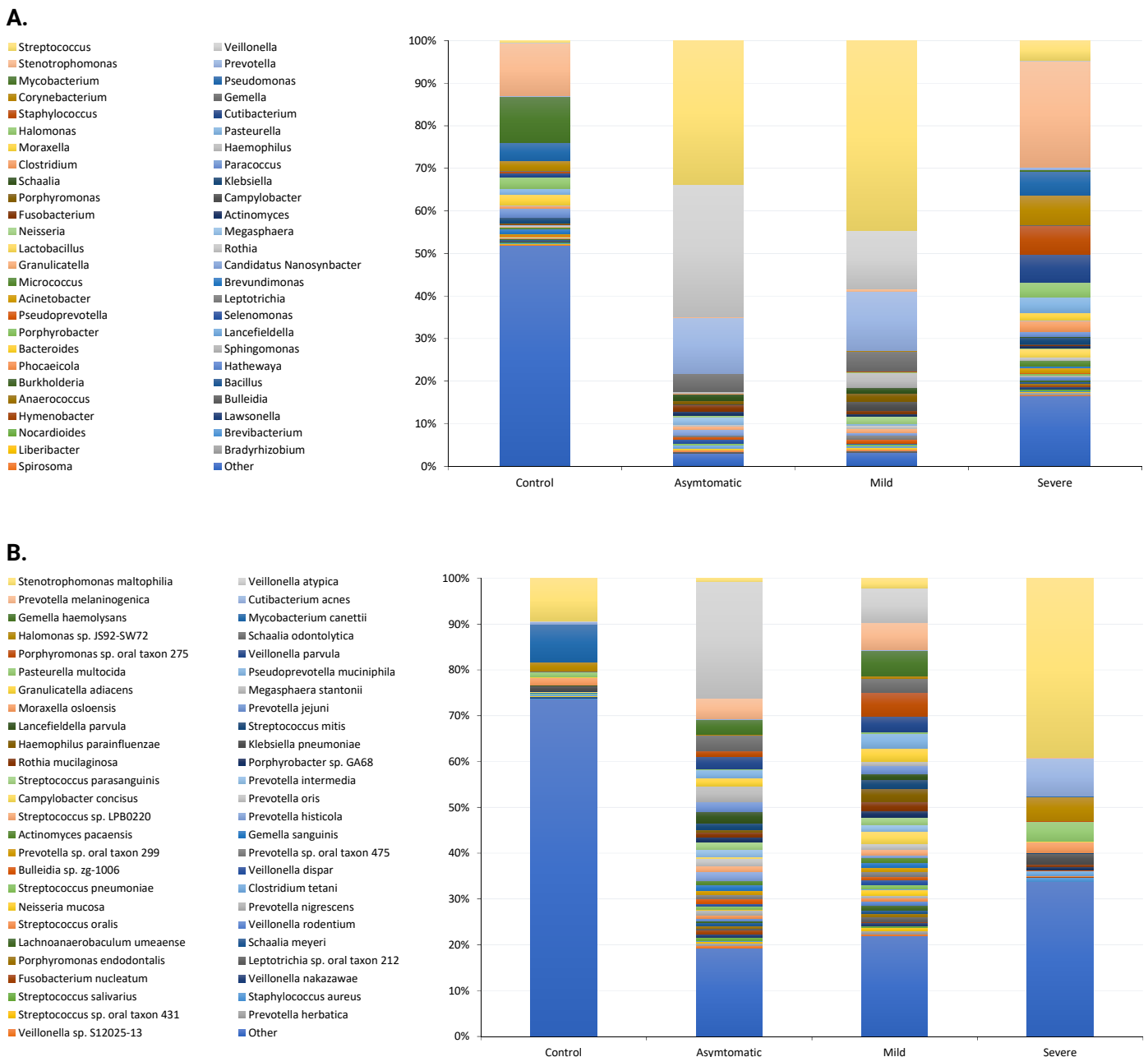
Prokaryotický mikrobióm

Metatranskriptómovým sekvenovaním sme analyzovali 151 vzoriek rozdelených do 4 hlavných skupín: asymptomatickí (24, A), mierni (25, M), ťažkí (30, Ť) a kontrolná skupina

(72, NK). Priemerný počet čítaní na vzorku bol 45,3 M (27,7 – 133 M). Podiel mikrobiómu, distribúcia a zastúpenie bakteriálnych taxónov boli vo všetkých 4 skupinách odlišné. Relatívna početnosť baktérií podľa mapovaných sekvenčných čítaní bola výrazne vyššia u asymptomatických a miernych pacientov, zatiaľ čo v ťažkých a negatívnych skupinách bola len 1,1 %, resp. 0,1 % (**obrázok 1A**). *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Actinobacteria* a *Fusobacteriales* boli najpočetnejšie kmene medzi COVID pozitívnymi vzorkami. Naopak, v kontrolnej skupine boli najpočetnejšou skupinou *Actinobacteria*, *Proteobacteria* a *Firmicutes*, zatiaľ čo *Bacteroides* a *Fusobacteriales* mali len zanedbateľný podiel (**obrázok 1B**). V tejto súvislosti je dôležité spomenúť, že u značnej časti pacientov so závažným COVID-19 z nemocníc už bola podávaná antibiotická liečba (ATB), čo malo za následok zníženie celkového množstva baktérií, ale pomer jednotlivých bakteriálnych taxónov sa napriek tomu zachoval.

Identifikovali sme celkovo 944 druhov, 531 rodov, 218 čeľadí, 110 radov, 28 tried a 25 kmeňov. Najvyšší počet druhov bol v skupine s ťažkým (725), asymptomatickým (692)

Obrázok 2. Stĺpcové grafy znázorňujúce relatívnu početnosť vo všetkých 4 skupinách **A.** bakteriálnych rodov **B.** bakteriálnych druhov



a miernym (574) ochorením COVID-19 a viac ako desaťnásobne nižším počtom v skupine negatívnych kontrol (58). Ďalej sme profilovali mikrobiálne zloženie na úrovni rodov a druhov v zdravých kontrolách a rôznych skupinách COVID-19. Najpočetnejšími rodmi vo vzorkách nazofaryngálnych výterov v miernej a asymptomatickej skupine je *Streptococcus*, *Prevotella* a *Veillonella*, (**obrázok 2A**). V skupine s ťažkým COVID-19 (ŤP) dominujú rody *Stenotrophomonas*, *Staphylococcus* a *Corynebacterium*. V COVID negatívnej skupine je opäť prekvapivo najpočetnejší rod *Stenotrophomonas*, ďalej *Mycobacterium* a *Pseudomonas*. Na úrovni druhov boli pri ťažkých najviac zastúpené druhy *Stenotrophomonas maltophilia*, *Cutibacterium acnes* a *Halomonas sp. JS92-SW72*. *Veillonella atypica* a *Prevotella melaninogenica* boli najviac zastúpené pri miernych a asymptomatických a celkové zastúpenie os-

tatných druhov rodu *Prevotella* (*P. jejuni*, *P. histicola*, *P. intermedia*, *P. oris*) a *Streptococcus* (*S. parasanguinis*, *S. mitis*, *S. sp. LPB0220*) bolo u oboch veľmi podobné (**obrázok 2B**). V kontrolnej skupine zdravých, ako aj v skupine s ťažkým priebehom boli najviac zastúpené *Stenotrophomonas maltophilia*, *Halomonas sp. JS92-SW72* a *Mycobacterium canettii*. Z klinicky relevantných oportúnnych bakteriálnych druhov boli vo všetkých štyroch skupinách prítomné napr. baktérie *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Streptococcus pneumoniae* a *Haemophilus parainfluenzae*, hoci v žiadnej zo skupín výrazne nedominovali a rozdiely medzi skupinami neboli štatisticky významné; ich výskyt bol skôr podobný v dvojiciach ťažká a negatívni verzus mierni a asymptomatickí. Mimoriadne zaujímavá bola prítomnosť bakteriálnych druhov vo všetkých skupinách, ktoré nie sú úplne typické pre mik-

robióm horných dýchacích ciest človeka. *Moraxella osloensis* sa z klinických vzoriek izolujú len veľmi zriedkavo a *Pasteurella multocida* je súčasťou normálnej flóry nosohltana mnohých divých a domácich zvierat vrátane mačiek a psov.

Analýza virómu

Zo 79 RT-PCR pozitívnych vzoriek bolo metatranskriptomovým sekvenovaním zostavených 41 kompletných a 19 čiastočne kompletných genómov SARS-CoV-2, čo predstavuje 62 % (51/79) zo všetkých. Okrem toho sme neidentifikovali sekvencie SARS-CoV-2 v žiadnej zo vzoriek kontrolnej skupiny (SARS-CoV-2 RT-PCR negatívny). V asymptomatickej, miernej a ťažkej skupine sme identifikovali varianty SARS-CoV-2: *alpha* (B.1.1.7) – 20I, *delta* (AY.4; AY.43; AY.43.9; AY.122; AY.9.2;) – 21I a 21J, *omicron* (BA.1.1; BA.2.9; BA.2; BA.2.67; BA.5) – 21L a 22B a v jednej vzorke variant 20C. Varianty *alpha* a *delta* prevládali v skupine s ťažkým ochorením, avšak v 30 % analýz sa nepodarilo zostaviť genóm z dôvodu nedostatočného počtu čítaní mapovaných na referenčný genóm SARS-CoV-2. V skupine s miernym priebehom dominoval variant *alpha*, zatiaľ čo v asymptomatickej skupine SARS-CoV varianty *omicron* a *delta*.

Okrem SARS-CoV-2 sme identifikovali sekvencie ďalších ľudských RNA vírusov, napr. z čeľade *Picornaviridae* (*Rhinovirus A*). Podľa očakávania to boli hlavne DNA fágy z čeľadi *Pedoviridae*, *Rountreeviridae* a *Siphoviridae*. V niekoľkých jednotlivých vzorkách sme identifikovali sekvencie ľudského herpesvírusu 1, *HSV-1* (*Alphaherpesviridae*), ľudského masta-denovírusu D (*Adenoviridae*) a vírusu *Molluscum Contagiosum* (*Poxviridae*).

Z hľadiska distribúcie veku boli rozdiely najmä medzi pacientmi s ťažkým COVID-19 z nemocnice (medián 68 rokov) a ostatnými skupinami (M-37, A-42, N-37) (**tabuľka 1**).

Diskusia

V našej štúdii sme porovnávali mikrobióm vrátane virómu nosohltana medzi štyrmi skupinami na základe závažnosti vírusového ochorenia dýchacích ciest a pozitivity na SARS-CoV-2: negatívna (bez príznakov ochorenia, RT-PCR negatívny), mierna (nevýrazné príznaky, nevyžadujúce hospitalizáciu), asymptomatická (bez príznakov ochorenia, RT-PCR pozitívny) a skupina ťažko chorých pacientov s COVID-19 (dýchacie ťažkosti, pneumónia, hospitalizovaní) pomocou metatranskriptomového sekvenovania. V texte na zjednodušenie používame početnosť baktérií alebo vírusov, hoci v skutočnosti stále hovoríme o transkriptoch, ktoré však vyjadrujú zastúpenie aktívnej časti mikrobiómu. Relatívna početnosť baktérií je výrazne vyššia v miernej a asymptomatickej skupine, zatiaľ čo v ťažkej a negatívnej kontrole je prítomná len malá časť (**obrázok 1A**). Nízka početnosť baktérií u ťažkých pacientov je ľahko vysvetliteľná užívaním antibiotík. Používanie antibiotík bez jasnej indikácie bakteriálnej superinfekcie je kontroverzné, zdá sa však, že v niektorých zdravotníckych centrách bolo bežnou praxou bez ohľadu na riziká vzniku rezistencie. Podľa závažnosti ochorenia sú *Firmicutes* najčastejšie u asymptomatických, miernych a ťažkých pacientov, zatiaľ čo u zdravých kontrol sú výrazne menej početné a proporčne zastúpené. U ťažkých pacientov sme pozorovali výrazný pokles početnosti, nie však diverzity bakteriálnych taxónov a ich zastúpenia.

Rozdelenie na dve skupiny COVID pozitívnych miernych a asymptomatických bolo motivované cieľom zistiť, či existuje nejaký rozdiel medzi mikrobiómom ľudí s príznakmi respiračných ochorení, ako je nádcha, horúčka, kašeľ, a ľudí, ktorí sú síce infikovaní (RT-PCR pozitívni), ale nevykazujú žiadne príznaky. Naše výsledky naznačujú, že početnosť baktérií aj pomerné zastúpenie jednotlivých taxónov sú v oboch skupinách veľmi podobné. Hoci sa mediánové hodnoty Ct v miernej (24,89) a asymptomatickej skupine (28,56) zdali dosť odlišné, neboli štatisticky významné (Wilcoxonov test, $p = 0,017$) (**tabuľka 1B**). Veľkosť oboch skupín (25 vs 24) a vekové rozloženie (37 vs 42) boli inak podobné (**tabuľka 1A**). Na úrovni rodov a druhov sme nezaznamenali žiadne významné rozdiely, čo potvrdila aj štatistická analýza (PERMANOVA; $p = 0,053$ a $0,024$, vypočítaná hraničná hodnota p bola $0,0023$). Najviac zastúpenými rodmi v oboch skupinách boli *Streptococcus*, *Veillonella* (oba kmeň *Firmicutes*) a *Prevotella* (kmeň *Bacteroidetes*), pričom najpočetnejšími druhmi boli *Veillonella atypica* a *Prevotella melaninogenica* (**bakteriálne druhy a rody, obrázky 2A,B**). Toto pozorovanie je v súlade so zisteniami iných štúdií mikrobiómu horných dýchacích ciest^(18–20). *P. melaninogenica* je gramnegatívny obligátne anaeróbny kokobacil, ktorý môže pôsobiť ako oportúnny patogén, pričom existujú protichodné hypotézy o jeho vplyve na dýchacie cesty. Napríklad sa zistilo, že *P. melaninogenica* je „prospešným“ členom mikrobiómu dýchacích ciest, pretože zvyšuje ochranu pred bakteriálnou pneumóniou spôsobenou najmä *Streptococcus pneumoniae* (*S. pneumoniae*). Ďalej bola *P. melaninogenica* zaradená medzi najvýraznejšie odlišujúce bakteriálne druhy u pacientov s pneumóniou spôsobenou *S. pneumoniae* (s menším množstvom *P. melaninogenica*) a zdravými kontrolami (s väčším množstvom *P. melaninogenica*)⁽²¹⁾. Zaujímavé je, že závažná skupina so získanou pneumóniou neznámeho bakteriálneho pôvodu v našom súbore mala významne menej *P. melaninogenica* (**obrázok 2B**).

Analyzovali sme 72 negatívnych zdravých kontrol, všetky boli dodatočne overené/testované pomocou RT-PCR (gény E, ORF1ab, N) a vyhodnotené ako negatívne. Relatívna početnosť baktérií na základe počtu namapovaných čítaní a transkriptov predstavovala u nich najnižší podiel (0,1 %) zo všetkých skupín (asymptomatická 68,8 %; mierna 30 % a ťažká 1,1 %) (**obrázok 1A**). Nie je to až také prekvapujúce, keďže každá vírusová infekcia vrátane infekcie SARS-CoV-2 zvyšuje pravdepodobnosť sekundárnej bakteriálnej superinfekcie. U zdravých dospelých ľudí sú v nosovej dutine zvyčajne baktérie bežne prítomné na koži, prevažne zástupcovia rodu *Actinobacterium* (napr. *Corynebacterium spp.*, *Propionibacterium spp.*), ďalej *Firmicutes* (napr. *Staphylococcus spp.*) a *Proteobacteria*⁽⁵⁾. V negatívnych kontrolách sme identifikovali 58 druhov z viac ako 900 druhov, čo zodpovedá na úrovni mikrobiálnej diverzity len približne 6 %.

Zastúpenie rôznych variantov SARS-CoV-2 z nazofaryngálnych výterov u asymptomatických, miernych a ťažkých pacientov korelovalo s obdobím výskytu jednotlivých kmeňov na Slovensku. Senzitivita detekcie SARS-CoV-2 prostredníctvom sekvenovania metatranskriptomu bola len približne 60 %, na druhej strane súbor SARS-CoV-2 pozitívnych vzoriek pozostával z rôznych hodnôt Ct zodpovedajúcich rôznym úrovniam vírusovej nálože. Na individuálnej úrovni sme

zo všetkých vzoriek, ktoré mali Ct 20 alebo menej, čo zodpovedá vyššej vírusovej náloži, dokázali zostaviť celý genóm vírusu SARS-CoV-2. Vzorky od ťažkých pacientov v nemocnici boli odobrané často 3 až 7 dní po zaradení do štúdie a títo pacienti boli liečení širokou škálou liekov (antibiotiká, kortikoterapia, imunomodulačná liečba, antivirotiká), ktoré mohli medzitým ovplyvniť vírusovú nálož. Jednoznačnou výhodou transkriptómového prístupu je jeho schopnosť zachytiť nielen RNA, ale aj DNA vírusy (ich transkripty). Okrem SARS-CoV-2 boli neprekvapivo najčastejšie identifikované sekvencie pochádzajúce z DNA vírusov baktérií, t. j. fágov z čeľadí *Pedoviridae*, *Rountreeviridae* a *Siphoviridae*.

Sekvencie mapované na ľudský transkriptóm tvorili nezanedbateľnú časť analýzy, ale nie sú predmetom tohto článku a zasluhujú si osobitnú pozornosť.

Záver

Identifikovali sme rozdiely v mikrobióme medzi rôznymi skupinami COVID pozitívnych a negatívnu skupinou. Napriek rozdielom v početnosti bola relatívna početnosť jednotlivých taxónov v skupinách pozitívnych na COVID-19 podobná bez ohľadu na závažnosť a použitú liečbu. Skupinu zdravých negatívnych kontrol definoval najmenší počet mikroorganizmov. Sekvenovanie metatranskriptómu je nástroj

na širokospektrálne štúdium aktívnej časti prokaryotického mikrobiómu, virómu, ale aj ľudského transkriptómu z rôznych dostupných zdrojov. Hoci jeho diagnostický potenciál z hľadiska citlivosti nie je optimálny, môže doplniť skladačku mozaiky širokej druhovej rozmanitosti a pomôcť pri pochopení zložitých vzťahov medzi mikrobiómom a imunitou človeka v podmienkach konkrétneho ochorenia, jeho závažnosti a na rôznych miestach ľudského tela. V našej štúdii bolo modelové ochorenie COVID-19 s rôznou závažnosťou, od asymptomatických až po ťažké prejavy vyžadujúce hospitalizáciu a intenzívnu liečbu. Nosohltan je zasa miestom, ktoré je dobre prístupné na odber vzoriek, je kolonizované širokou škálou mikroorganizmov, ktoré ovplyvňujú lokálnu imunitnú odpoveď, a je tiež miestom vstupu a počiatočného množenia širokého spektra respiračných vírusov.

Podakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Závažné civilizačné ochorenia a COVID-19, kód ITMS: 313011AVH7, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

1. BAI, Xiangning, Aswathy NARAYANAN, Magdalena SKAGERBERG, Rafael CEÑA-DIEZ, Christian G. GISKE, Kristoffer STRÄLIN a Anders SÖNERBERG. Characterization of the Upper Respiratory Bacterial Microbiome in Critically Ill COVID-19 Patients. *Biomedicines* [online]. 2022, 10(5), 982. ISSN 2227-9059. Dostupné z: doi:10.3390/biomedicines10050982
2. CANDEL, Sergio, Sylwia D. TYRKALSKA, Carmen ÁLVAREZ-SANTACRUZ a Victoriano MULERO. The nasopharyngeal microbiome in COVID-19. *Emerging Microbes & Infections* [online]. 2023, 12(1). ISSN 2222-1751. Dostupné z: doi: 10.1080/22221751.2023.2165970
3. ZHU, Taiping, Jun JIN, Minhua CHEN a Yingjun CHEN. The impact of infection with COVID-19 on the respiratory microbiome: A narrative review. *Virulence* [online]. 2022, 13(1), 1076–1087. ISSN 2150-5594. Dostupné z: doi: 10.1080/21505594.2022.2090071
4. NEMERGUT, Diana R., Steven K. SCHMIDT, Tadashi FUKAMI, Sean P. O'NEILL, Teresa M. BILINSKI, Lee F. STANISH, Joseph E. KNELMAN, John L. DARCY, Ryan C. LYNCH, Phillip WICKEY a Scott FERRENBERG. Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* [online]. 2013, 77(3), 342–356. ISSN 1092-2172. Dostupné z: doi: 10.1128/MMBR.00051-12
5. BASSIS, Christine M, Alice L TANG, Vincent B YOUNG a Melissa A PYNNONEN. The nasal cavity microbiota of healthy adults. *Microbiome* [online]. 2014, 2(1), 27. ISSN 2049-2618. Dostupné z: doi:10.1186/2049-2618-2-27
6. STEARNS, Jennifer C, Carla J DAVIDSON, Suzanne MCKEON, Fiona J WHELAN, Michelle E FONTES, Anthony B SCHRYVERS, Dawn M E BOWDISH, James D KELLNER a Michael G SURETTE. Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. *The ISME Journal* [online]. 2015, 9(5), 1246–1259. ISSN 1751-7362. Dostupné z: doi:10.1038/ismej.2014.250
7. MINOT, Samuel, Alexandra BRYSON, Christel CHEHOUD, Gary D. WU, James D. LEWIS a Frederic D. BUSHMAN. Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences* [online]. 2013, 110(30), 12450–12455. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.1300833110
8. RAJAGOPALA, Seesandra V., Nicole G. BAKHOUM, Suman B. PAKALA, Meghan H. SHILTS, Christian ROSAS-SALAZAR, Annie MAI, Helen H. BOONE, Rendie MCHENRY, Shibu YOUSEPH, Natasha HALASA a Suman R. DAS. Metatranscriptomics to characterize respiratory virome, microbiome, and host response directly from clinical samples. *Cell Re-*

ports Methods [online]. 2021, 1(6), 100091. ISSN 26672375. Dostupné z: doi:10.1016/j.crmeth.2021.100091

9. ROVIRA RUBIÓ, Judit, Spyridon MEGREMIS, Maria PASIOTI, John LAKOUMENTAS, Bede CONSTANTINIDES, Paraskevi XEPAPADAKI, Claus BACHERT, Susetta FINOTTO, Tuomas JARTTI, Evangelos ANDREAKOS, Barbara STANIC, Cezmi A. AKDIS, Mübecel AKDIS a Nikolaos G. PAPA-DOPOULOS. Respiratory virome profiles reflect antiviral immune responses. *Allergy* [online]. 2023. ISSN 0105-4538. Dostupné z: doi:10.1111/all.15634
10. TESSLER, Michael, Johannes S. NEUMANN, Ebrahim AFSHINNEKOO, Michael PINEDA, Rebecca HERSCH, Luiz Felipe M. VELHO, Bianca T. SEGOVIA, Fabio A. LANSAC-TOHA, Michael LEMKE, Rob DESALLE, Christopher E. MASON a Mercer R. BRUGLER. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Scientific Reports* [online]. 2017, 7(1), 6589. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-017-06665-3
11. BRUMFIELD, Kyle D., Anwar HUQ, Rita R. COLWELL, James L. OLDS a Menu B. LEDDY. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLOS ONE* [online]. 2020, 15(2), e0228899. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0228899
12. MORTAZAVI, Ali, Brian A WILLIAMS, Kenneth MCCUE, Lorian SCHAEFFER a Barbara WOLD. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* [online]. 2008, 5(7), 621–628. ISSN 1548-7091. Dostupné z: doi:10.1038/nmeth.1226
13. HENG LI. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [online]. nedatováno. Dostupné z: doi:10.48550/arXiv.1303.3997
14. WOOD, Derrick E a Steven L. SALZBERG. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* [online]. 2014, 15(3), R46. ISSN 1474-760X. Dostupné z: doi:10.1186/gb-2014-15-3-r46
15. MELESHKO, Dmitry, Iman HAJIRASOULIHAM a Anton KOROBENIKOV. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. *bioRxiv*. 2021.
16. MAIER, Wolfgang a Bérénice BATUT. *Mutation calling, viral genome reconstruction and lineage/clade assignment from SARS-CoV-2 sequencing data*. 2023.

17. SEGATA, Nicola, Jacques IZARD, Levi WALRON, Dirk GEVERS, Larisa MIROPOLSKY, Wendy GARRETT a Curtis HUTTENHOWER. *Metagenomic Biomarker Discovery and Explanation*. 24. červen 2011
18. LIU, Jing, Sheng LIU, Zhao ZHANG, Xuejun LEE, Wenxuan WU, Zhanlian HUANG, Ziyang LEI, Wenxiong XU, Dabiao CHEN, Xing WU, Yang GUO, Liang PENG, Bingliang LIN, Yutian CHONG, Xiangyu MOU, Mang SHI, Ping LAN, Tao CHEN, Wenjing ZHAO a Zhiliang GAO. Association between the nasopharyngeal microbiome and metabolome in patients with COVID-19. *Synthetic and Systems Biotechnology* [online]. 2021, 6(3), 135–143. ISSN 2405805X. Dostupné z: doi:10.1016/j.synbio.2021.06.002
19. RATTANABURI, Somruthai, Vorthon SAWASWONG, Suwalak CHITCHAROEN, Pavaret SIVAPORNNUKUL, Pattaraporn NIMSAMER, Nungruthai SUNTRONWONG, Jiratchaya PUENPA, Yong POOVORAWAN a Sunchai PAYUNGPORN. Bacterial microbiota in upper respiratory tract of COVID-19 and influenza patients. *Experimental Biology and Medicine* [online]. 2022, 247(5), 409–415. ISSN 1535-3702. Dostupné z: doi:10.1177/15353702211057473
20. VENTERO, Maria Paz, Rafael R. C. CUADRAT, Inmaculada VIDAL, Bruno G. N. ANDRADE, Carmen MOLINA-PARDINES, Jose M. HARO-MORENO, Felipe H. COUTINHO, Esperanza MERINO, Luciana C. A. REGITANO, Cynthia B. SILVEIRA, Haithem AFLI, Mario LÓPEZ-PÉREZ a Juan Carlos RODRÍGUEZ. Nasopharyngeal Microbial Communities of Patients Infected With SARS-CoV-2 That Developed COVID-19. *Frontiers in Microbiology* [online]. 2021, 12. ISSN 1664-302X. Dostupné z: doi:10.3389/fmicb.2021.637430
21. HORN, Kadi J., Melissa A. SCHOPPER, Zoe G. DRIGOT a Sarah E. CLARK. Airway Prevotella promote TLR2-dependent neutrophil activation and rapid clearance of *Streptococcus pneumoniae* from the lung. *Nature Communications* [online]. 2022, 13(1), 3321. ISSN 2041-1723. Dostupné z: doi:10.1038/s41467-022-31074-0

RNDr. Michaela Hýblová, PhD.

MEDIREX GROUP ACADEMY, Bratislava
Galvaniho 17/C, 821 04 Ružinov
e-mail: michaela.hyblova@medirex.sk

Hodnotenie početnosti, pozitivity a Ct hodnôt SARS-CoV-2 detekcie na Slovensku v čase pandémie COVID-19 – deti verzus dospelí jedinci

Nikola Janoštiaková¹, Andrej Gnip², Dominik Kodada¹, Gabriela B'andová¹, Emília Miková², Elena Tibenská², Vanda Repiská¹, Gabriel Minárik³

¹Univerzita Komenského v Bratislave Lekárska fakulta, Ústav lekárskej biológie, genetiky a klinickej genetiky, Bratislava

²Medirex, a. s., Pezinok

³MEDIREX GROUP ACADEMY, n. o., Nitra

Vplyv infekcie SARS-CoV-2 na detskú populáciu v porovnaní s dospelými nie je doteraz presne objasnený. Detegované prahové hodnoty cyklu (Ct) pri RT-qPCR SARS-CoV-2 diagnostike sú nepriamoúmerné vírusovej záťaži – test nepriamo kvantifikuje počet kópií vírusovej RNA vo vzorke. Cieľom tejto štúdie bolo determinovať a štatisticky podložiť možnú koreláciu medzi rozdielnymi priemernými Ct hodnotami a vekom pozitívne testovaných jedincov v kategorizovaných skupinách maloletých a dospelých.

Kľúčové slová: SARS-CoV-2, COVID-19, pandémia, Slovensko, Ct hodnoty, deti, vírusová záťaž

Evaluation of the frequency, positivity and Ct values of SARS-CoV-2 detection in Slovakia during the COVID-19 pandemic – children versus adults

The impact of SARS-CoV-2 infection on the pediatric population compared to adults is not yet clearly understood. The detected cycle threshold values (Ct) in RT-qPCR SARS-CoV-2 diagnostics are inversely proportional to the viral load – the test indirectly quantifies and determines the number of viral RNA copies in the sample. This study aimed to determine and statistically substantiate a possible correlation between different average Ct values and the age of positively tested individuals in the categorized groups of children, adolescents and adults.

Keywords: SARS-CoV-2, COVID-19, pandemic, Slovakia, Ct values, children, viral load

NewsLab, 2023; roč. 14 (1): 17 – 20

Úvod

Ako je už v súčasnosti známe, ochorenie COVID-19 vyvolané vírusom SARS-CoV-2 spôsobilo celosvetovú pandémiu, ktorá viedla ku globálnej núdzovej situácii v oblasti verejného zdravotníctva z dôvodu vysokej rýchlosti šírenia infekcie, širokého spektra klinických prejavov vo všetkých vekových skupinách a vysokej úmrtnosti najmä u dospelých osôb⁽¹⁾.

Takisto je známe, že prvotne boli kontrolné opatrenia na Slovensku aj vo svete zamerané na detekciu a následnú izoláciu potvrdených prípadov a taktiež karanténu kontaktných osôb pozitívnych prípadov⁽²⁾. Sekvenovanie genómu SARS-CoV-2 umožnilo vývoj testu reverznej transkripčnej polymerázovej reťazovej reakcie v reálnom čase (RT-qPCR) na detekciu a relatívnu kvantifikáciu vírusovej nálože. Táto technika je aktuálne stále zlatým štandardom pre diagnostiku ochorenia COVID-19. RT-qPCR diagnostika deteguje rôzne štrukturálne oblasti kódované génmi: S (*Spike*), E (*Envelope*), N (*Nucleocapsid*), RdRp (*RNA-dependent RNA polymerase*) a ORF1ab (*polyprotein*). Vo všeobecnosti boli RT-qPCR diagnostické súpravy primárne navrhnuté pre kvalitatívnu detekciu SARS-CoV-2 („detegovaná“ alebo „nedetegovaná“ prítomnosť SARS-CoV-2 transkriptov), no uvádzané prahové hodnoty cyklu (Ct) reprezentujúce počet cyklov amplifikácie potrebných na to, aby fluorescenčný signál prekročil úroveň pozadia, je možné využiť na semikvantitatívny odhad vírusovej nálože⁽³⁾.

Ct hodnoty sú nepriamoúmerné vírusovej náloži a poskytujú možnosť kvantifikácie počtu kópií vírusovej nukleovej kyseliny (RNA) vo vzorkách^(4,5). Hodnoty sú ovplyvnené preanalytickými, analytickými a postanalytickými premennými. Správna interpretácia týchto faktorov je rozhodujúca pre odhad korelácie medzi vírusovou záťažou a závažnosťou ochorenia. Systematický prehľad klinickej užitočnosti prahových hodnôt cyklu naznačil, že Ct hodnoty sú užitočné na predpovedanie klinického výsledku u pacientov s ochorením COVID-19^(5,6).

Už na začiatku pandémie existoval predpoklad, že dospelí jedinci budú mať závažnejší priebeh SARS-CoV-2 infekcie ako deti, informácie o etiológii choroby u detí však boli obmedzené. To prinieslo dlhé obdobia karantén, zatvárania škôl, prechodov na online výučbu, čoho následkom boli/sú početné psychologické a sociálno-emocionálne komplikácie u maloletých^(7,8). Okrem problémov s duševným zdravím boli u žiakov s distančnou výučbou zaznamenané zmeny týkajúce sa nezdravého priberania na váhe, oneskorenej imunizácie či prudkého nárastu počtu novovzniknutého diabetu 2. typu. Tieto priame dôsledky zatvorenia škôl môžu byť predzvesťou dlhodobých komplikácií vrátane zníženej priemernej dĺžky života školákov⁽⁸⁾. Pochopenie charakteristík a klinických prejavov ochorenia COVID-19 u detí je preto kľúčové.

Nedávny systematický prehľad a metaanalýza zahŕňajúca 37 štúdií o ochorení COVID-19 u detí potvrdila predchádzajúce tvrdenia o relatívne miernom priebehu ochorenia u mladších detí, ktorých symptómy zahŕňajú najčastejšie horúčku (48,5 %) a kašeľ (40,6 %) (9), čo pravdepodobne súvisí s nižšou vírusovou záťažou u detských pacientov v porovnaní s dospelými jedincami (10). Už v úvode pandémie bolo preukázané, že SARS-CoV-2 využíva na vstup do hostiteľskej bunky proteín prítomný na jej povrchu, enzým konvertujúci angiotenzín 2 (ACE2). Štúdie naznačujú, že expresia ACE2 je závislá od tkaniva aj veku, pričom najvyššia expresia ACE2 je v nosovom epiteli zdravých dospelých jedincov. Nižšia expresia ACE2 je, naopak, detegovaná v horných a dolných dýchacích cestách u detí, na základe čoho sa predpokladá, že tieto rozdiely prispievajú k zníženiu závažnosti ochorenia u detí. Autori Yoshida a kol. výsledkami ich nedávneho výskumu naznačujú, že za obmedzenú replikáciu vírusu a progresiu ochorenia COVID-19 u detí zodpovedajú vyššie detské vrodené interferónové reakcie (11). Doteraz však nie je úplne jasné, prečo má ochorenie COVID-19 u detí zvyčajne miernejší charakter.

Cieľom tejto štúdie bolo posúdiť koreláciu medzi epidemiologickými znakmi u detí a dospelých jedincov. Výsledky zahŕňajú dáta z diagnostického testovania SARS-CoV-2 na Slovensku s rozsahom definovaným mierou vykonaných testov v laboratóriách Medirex, a. s. Počet údajov z tejto štúdie je robustný a predstavuje viac ako 1,4 milióna hodnotených RT-qPCR testov. Dátumový výber údajov zahŕňa celú časovú os masívnej SARS-CoV-2 detekcie v rámci Slovenskej republiky. Zaujímali nás najmä parametre z hľadiska rozdielov v pozitívite a vo vírusovej záťaži (odhadované pomocou hodnôt Ct) medzi charakterizovanými vekovými podskupinami. Naše poznatky o trendoch pandémie v kontexte vekových podskupín v slovenskej populácii možno využiť na efektívnejšie nastavenie diagnostiky aj manažmentu opatrení pri prírodných budúcich epidémiách/ pandemiách.

Metódy

Štúdia analyzovala dáta poskytnuté akreditovaným laboratóriom Medirex, a. s. Údaje predstavujú výsledky 1 420 572 RT-qPCR testov jedincov pred plánovanou hospitalizáciou, jedincov s príznakmi ochorenia a/alebo kontaktom s pozitívne testovanou osobou, jedincov po pozitívnom výsledku antigénového testu (12), samoplatcov bez selekcie; z dátumového obdobia od marca 2020 do septembra 2022.

Automatizovaná izolácia nukleovej kyseliny (RNA) na magnetických časticach bola uskutočnená pomocou súprav Sera-Xtracta Virus /Pathogen Kit (Cytiva) a Zybio Nucleic Acid Extraction Kit (Zybio) s využitím systému KingFisher™ Flex

Purification System (Thermo Scientific) a Zybio EXM 3000 Nucleic Acid Isolation System/ Zybio EXM 6000 Nucleic Acid Isolation System (Zybio). Testovanie sa uskutočnilo metódou RT-qPCR s použitím súprav COVID-19 Real Time Multiplex RT-PCR Kit (Labsystems Diagnostics), SARS-CoV-2 Nucleic Acid Detection Kit (Zybio) a Real Time Multiplex RT-PCR Kit (Liferiver) pomocou qPCR platforiem ABI 7500 (Fast) Real-Time PCR System (Applied Biosystems), QuantStudio 5 a QuantStudio 6 Real-Time PCR System (ThermoFisher). Údaje získané pre každý test zahŕňali denné číslo vzorky, dátum odberu, ID pacienta, vek, pohlavie, miesto odberu vzorky, výsledok testu a v prípade pozitívneho výsledku aj hodnotu Ct vírusového génu E – hraničná hodnota do 40 (pre kit SARS-CoV-2 Nucleic Acid Detection Kit (Zybio) alebo 41 (pre kity Real Time Multiplex RT-PCR Kit (Labsystems Diagnostics) a Real Time Multiplex RT-PCR Kit (Liferiver) – t. j. sekvencia nukleovej kyseliny bola identifikovaná v čase, keď PCR prešla 39 alebo 40 cyklami. Hodnota Ct bola spolu s pozitívnym výsledkom testu laboratóriom Medirex, a. s., uvádzaná a archivovaná od 3. decembra 2020, čo vysvetľuje nižšiu početnosť pozitívnych PCR testov s uvedenou Ct hodnotou v porovnaní s celkovým počtom pozitívnych PCR testov v rámci tejto štúdie.

Údaje boli kategorizované do 5 skupín podľa veku, resp. úrovne navštevovanej školy: novorodenci + batolata + deti predškolského veku (0 – 5 r.), žiaci ZŠ (6 – 13 r.), žiaci SŠ (14 – 19 r.), študenti VŠ + dospelí (20 – 65 r.) a seniori (> 65 r.).

Na porovnanie kategórií údajov a kontrolu korelácií údajov boli využité štatistické testy. Na porovnanie hodnôt pozitívnosti bol použitý Fisherov exaktný test a na porovnanie hodnôt Ct Welchov t-test medzi vekovými kategóriami. Spolu bolo vykonaných 20 štatistických testov, na základe tohto počtu bola aplikovaná Bonferroniho korekcia pre úpravu prahu významnosti p-hodnoty – ako prahová hodnota bola využitá hodnota 0,0025 namiesto 0,05.

Výsledky

Celková pozitivita analyzovaných testov predstavovala 24,64 % (350 067 testov), pričom najvyššiu mieru pozitívnosti (nad 30 %) dosiahli vekové kategórie: 6 – 13 r., 14 – 19 r. (tabuľka 1). Na základe Fisherovho exaktného testu bolo preukázané, že rozdiely v pozitívite medzi všetkými dvojicami vekových skupín boli signifikantne významné (p-hodnota < 0,0025).

Obrázok 1 znázorňuje rozdelenie Ct hodnôt v závislosti od veku jedincov. Husľový graf preukazuje postupné znižovanie hodnôt mediánu Ct spolu so zvyšujúcim sa vekom kategorizovaných skupín, výnimkou je veková skupina seniorov > 65 r. Najvyššia hodnota mediánu Ct bola zistená vo veko-

Tabuľka 1. Počet pozitívnych/negatívnych RT-qPCR testov a celková pozitivita pre jednotlivé vekové kategórie.

Veková skupina	Počet pozitívnych testov	Počet negatívnych testov	Celkový počet testov	Pozitivita (%)
0 – 5 r.	6 563	23 474	11,079	21,85 %
6 – 13 r.	36 882	63 656	18,958	36,69 %
14 – 19 r.	28 822	55 511	31,801	34,18 %
20 – 65 r.	244 965	793 982	68,737	23,58 %
> 65 r.	32 835	133 882	27,951	19,70 %
Celkovo	350 067	1 070 505	1 420 572	24,64 %

vej skupine 0 – 5 r., rovná sa číslu 29,79. Najnižšia hodnota mediánu Ct bola zistená v skupine 20 – 65 r., rovná sa číslu 26,67. Vo vekovej skupine seniorov > 65 r. bola preukázaná hodnota mediánu Ct 26,70. Signifikantné rozdiely v mediáne Ct boli preukázané na základe Welchovho t-testu medzi všetkými dvojicami vekových skupín (p-hodnota < 0,0025) s výnimkou dvojice 20 – 65 r. a > 65 r. (p-hodnota 0,17).

Diskusia

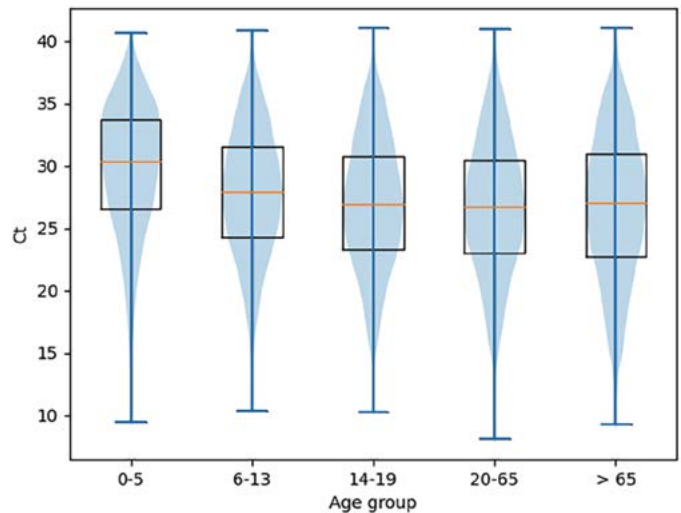
Existuje niekoľko faktorov, ktoré môžu potenciálne ovplyvniť zistené Ct hodnoty vrátane typu vzorky (napr. nosohltanová, predná sliznica nosa, sliny, spútum) alebo účinnosti odberu (neinvasívne odbery môžu viesť k nepresnostiam výsledkov)⁽¹³⁾. Detekcia nižších hodnôt Ct u detí môže súvisieť s kvalitou odberu⁽¹⁴⁾. Odbery slín sa zaviedli do praxe práve pre komplikácie odberov z nosohltana v určitých skupinách jedincov – výraznú bolesť pri výteroch z nosohltana opísalo 58 % mladých ľudí oproti žiadnej pri odbere slín, pričom 90 % detí preferuje odber vzoriek slín⁽¹⁵⁾. Výsledky viacerých štúdií demonštrujú, že detekcia SARS-CoV-2 zo vzoriek slín funguje rovnako alebo dokonca lepšie ako výtery z nosa a hltana⁽¹⁶⁻¹⁹⁾, no priemerná Ct hodnota v slinách môže byť vyššia v porovnaní s priemernou Ct hodnotou z nazofaryngálnych výterov⁽²⁰⁾. Preto sa ako vhodnejšia vzorka na diagnostiku RT-qPCR zvyčajne odporúča kombinovaný výter z nosa a hrdla⁽¹³⁾. Keďže sme do tejto štúdie chceli zahrnúť všetky vzorky od marca 2020 do septembra 2022 poskytnuté laboratóriom Medirex s priradeným pozitívnym/negatívnym výsledkom, údaje zahŕňajú nazofaryngálne výtery aj vzorky slín. Výsledky reprezentujúce detekciu SARS-CoV-2 zo vzoriek slín však predstavujú menej ako 5 % všetkých vzoriek, a preto nepredpokladáme, že by táto skutočnosť mohla ovplyvniť celkovú výpovednú hodnotu priemerných hodnôt Ct pre údaje tejto štúdie.

Predbežné dôkazy naznačujú, že u detí infikovaných SARS-CoV-2 je nižšia pravdepodobnosť, že sa u nich rozvinú symptómy alebo závažný priebeh ochorenia COVID-19⁽²¹⁾. Z hľadiska vírusovej záťaže odhadovanej na základe detegovanej Ct hodnoty boli preukázané signifikantne nižšie Ct hodnoty u dospievajúcich a dospelých jedincov v porovnaní s deťmi mladšími ako 10 rokov⁽²²⁾, čo naznačuje, ktoré vekové kategórie môžu viac, resp. menej prispievať k šíreniu infekcie. V našej štúdií sme analyzovali, či a ako stredné Ct hodnoty RT-qPCR korelujú s vekom a výskytom infekcie SARS-CoV-2 v subpopulácii slovenských obyvateľov testovaných v jednom akreditovanom laboratóriu.

Skutočná lokálna prevalencia a počet vykonaných testov ovplyvňujú hodnotu miery pozitivity. Vo všeobecnosti, čím viac testov je zrealizovaných, tým sa miera pozitivity bude približovať ku skutočnej prevalencii⁽²³⁾. Analyzované testy tejto štúdie predstavujú > 19 % všetkých diagnostických PCR testov na Slovensku (za daný dátumový interval), takže by mali byť reprezentatívne pre populáciu.

Biomedicínsky výskum sa aktuálne spolieha na p-hodnoty ako deterministickú mieru pri rozhodovaní založenom na údajoch, kde sa zo vzoriek dát vypočítava jedna p-hodnota, aby sa určili štatisticky významné rozdiely medzi pozorovanými skupinami. Pretože odhadovaná p-hodnota má tendenciu klesať so zvyšujúcou sa veľkosťou vzorky, aplikácia tejto metódy na súbory údajov s veľkou početnosťou vzoriek

Obrázok 1. Distribúcia Ct pozitívnych diagnostických testov v kategorizovaných vekových skupinách. Údaje sú zobrazené ako husľové grafy, jednotlivé boxploty sú doplnené o príslušné mediány Ct (oranžové úsečky).



vedie k vyššej pravdepodobnosti zamietnutia nulovej hypotézy⁽²⁴⁾, čo sa odrazilo aj v našich údajoch (viac ako 1,4 mil. RT-qPCR testov). Táto skutočnosť pravdepodobne prispela k preukázaniu výrazných rozdielov v pozitívite a priemerných hodnotách Ct medzi takmer všetkými kategorizovanými vekovými skupinami. To však neznamená, že pozorovaný význam nie je skutočný. Naopak, nízke p-hodnoty sú prítomné pri veľkých veľkostiach vzoriek, pretože robustnosť údajov znamená, že aj malý rozdiel s väčšou pravdepodobnosťou odráža skutočný základný rozdiel medzi dvoma súbormi údajov. Graf na **obrázku 1** determinuje, že pokles mediánu Ct bol v súlade so zvyšujúcim sa vekom v kategorizovaných podskupinách, čo je podložené aj zistením, že SARS-CoV-2 pozitívne deti sú zriedkavo „prípados indexu“ v prenosových reťazcoch v rodinách, resp. domácnostiach⁽²⁵⁾. Hoci naše výsledky nie je možné priamo použiť pri analýze efektu zatvárania škôl ako potenciálnych ohnísk komunitného šírenia COVID-19, poukazujú na to, že na šírení pandémie sa deti školského veku podieľali menšou mierou ako dospelí. K porovnateľným výsledkom dospeli autori z partnerských laboratórií Spadia, a. s., pričom v ich prípade bolo možné túto domnienku s ohľadom na dôsledné sledovanie šírenia pandémie práve so zameraním na otváranie a zatváranie škôl v Českej republike potvrdiť⁽²⁶⁾.

Podakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum progresívnych metód diagnostiky COVID-19 a biomarkerov umožňujúcich skorú detekciu jedincov so zvýšeným rizikom ťažkého priebehu ochorenia, kód ITMS: 313011ATA2, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja a s podporou Agentúry na podporu výskumu a vývoja na základe zmluvy č. PP-COVID-20-0056.

Doplnkové informácie

Tabuľka so všetkými vzorkami zahrnutými do analýzy je dostupná v podobe .xls súboru (ST – Supplementary table) na vyžiadanie od autora.

LITERATÚRA

1. Wu P, Hao X, Lau EHY, et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Euro Surveill.* 2020; 25: 2000044.
2. Rattan A, Ahmad H. Can quantitative RT-PCR for SARS-CoV-2 help in better management of patients and control of coronavirus disease 2019 pandemic. *Indian J Med Microbiol.* 2020; 38: 284–287.
3. Bustin SA, Mueller R. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin Sci (Lond).* 2005; 109: 365–379.
4. Mackay IM. Real-time PCR in the microbiology laboratory. *Clin Microbiol Infect.* 2004; 10(3): 190–212.
5. Rao SN, Manissero D, Steele VR, et al. A Systematic Review of the Clinical Utility of Cycle Threshold Values in the Context of COVID-19. *Infect Dis Ther* 2020; 9(3): 573–586.
6. Tom MR, Mina MJ. To Interpret the SARS-CoV-2 Test, Consider the Cycle Threshold Value. *Clin Infect Dis* 2020; 71(16): 2252–2254.
7. Hen M, Shenaar-Golan V, Yatzker U. Children and Adolescents' Mental Health Following COVID-19: The Possible Role of Difficulty in Emotional Regulation. *Front Psychiatry.* 2022; 13: 865435.
8. Dooley DG, Rhodes H, Bandealy A. Pandemic Recovery for Children—Beyond Reopening Schools. *JAMA Pediatr.* 2022; 176(4): 347–348.
9. Qi K, Zeng, W, Ye, M, et al. Clinical, laboratory, and imaging features of pediatric COVID-19: a systematic review and meta-analysis. *Medicine (Baltimore)* 2021; 100(15): e25230.
10. Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science.* 2021; 373(6551): ea-bi5273.
11. Yoshida M, Worlock KB, Huang N, et al. NU SCRIPT Study Investigators; Reynolds G, Haniffa M, Bowyer GS, Coates M, Clatworthy MR, Calero-Nieto FJ, Göttgens B, O'Callaghan C, Sebire NJ, Jolly C, De Coppi P, Smith CM, Misharin AV, Janes SM, Teichmann SA, Nikolić MZ, Meyer KB. Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature.* 2022; 602(7896): 321–327.
12. <https://www.health.gov.sk/Clanok?test-pcr-ochorenie> dostupné dňa 5. 4. 2023
13. Yoo HM, Kim IH, Kim S. Nucleic Acid Testing of SARS-CoV-2. *Int. J. Mol. Sci.* 2021; 22, 6150.
14. <https://www.aacc.org/science-and-research/covid-19-resources/statements-on-covid-19-testing/aacc-recommendation-for-reporting-sars-cov-2-cycle-threshold-ct-values> dostupné dňa 24. 3. 2023
15. Gagnon F, Bhatt M, Zemek R, et al. Nasopharyngeal swabs vs. saliva sampling for SARS-CoV-2 detection: A cross-sectional survey of acceptability for caregivers and children after experiencing both methods. *PLoS one.* 2020; 7(7), e0270929.
16. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020; 395(10223): 497–506.
17. Ponti G, Maccaferri M, Ruini C, et al. Biomarker associated with COVID-19 disease progression. *Crit Rev Clin Lab Sci.* 2020: 1–11.
18. Zaninotto M, Mion M, Cosma C, et al. Presepsin in risk stratification of SARS-CoV-2 patients. *Clin Chim Acta.* 2020; 507: 161–163.
19. Hansen CB, Bayarri-Olmos R, Kristensen MK, et al. Complement related pattern recognition molecules as markers of short-term mortality in intensive care patients. *J Infect.* 2020; 80(4): 378–387.
20. Torres M, Collins K, Corbit M, et al. Comparison of saliva and nasopharyngeal swab SARS-CoV-2 RT-qPCR testing in a community setting. *J infect.* 2021; 82(4): 84-123.
21. <https://stacks.cdc.gov/view/cdc/89980> dostupné dňa 25. 4. 2023.
22. Bullard J, Funk D, Dust K, et al. Infectivity of severe acute respiratory syndrome coronavirus 2 in children compared with adults. *CMAJ.* 2021; 193: E601–E606.
23. Montesinos-López JC, Daza-Torres ML, García YE, et al. The Role of SARS-CoV-2 Testing on Hospitalizations in California. *Life* 2021; 11(12): 1336.
24. Gómez-de-Mariscal E, Guerrero V, Sneider A. et al. Use of the p-values as a size-dependent function to address practical differences when analyzing large datasets. *Sci Rep.* 2021; 11, 20942.
25. Zhu YS, Bloxham CJ, Hulme KD, et al. Meta-analysis on the Role of Children in Severe Acute Respiratory Syndrome Coronavirus 2 in Household Transmission Clusters. *Clin. Infect. Dis.* 2021; 72(12): E1146-E53.
26. Musalkova D, Piherova L, Kwasny O, et al. Trends in SARS-CoV-2 cycle threshold values in the Czech Republic from April 2020 to April 2022. *Sci Rep.* 2023; 13(1): 6156.

Mgr. Nikola Janoštiaková

Ústav lekárskej biológie, genetiky a klinickej genetiky, LF UK Bratislava

Špitálska 24, 811 08 Bratislava

e-mail: n.janostiakova@gmail.com

Zber klinických údajov a biologických vzoriek na využitie v základnom a aplikovanom výskume

Petra Hirjaková, Marianna Jagelková, Michal Lichvár, Ľubomír Arnold, Gabriel Minárik
MEDIREX GROUP ACADEMY, n. o., Nitra

Biobanking je neoddeliteľnou súčasťou mnohých vedeckých alebo klinických štúdií. Cieľom článku je informovať o aktuálnom stave realizovaných štúdií, súbore údajov a biologických vzoriek získaných v rámci projektov Medirex Group Academy, n. o. (BIOMEDIRES II., PreveLynch, Promedicov, DIACOVID) na použitie v súčasnom, ale aj budúcom základnom a aplikovanom výskume, ako aj definovať nové perspektívy a rozvojové stratégie pre vzniknutú biobanku biologických vzoriek a klinických údajov. V článku uvádzame príklady zbieraných anamnestických a klinických dát a diskutujeme výhody, resp. nevýhody súčasne využívaného prístupu zberu vzoriek a údajov pre lepšie nastavenie realizácie klinických štúdií v budúcnosti.

Kľúčové slová: klinická štúdia, výskum, biobanka, biologické vzorky

Collection of clinical data and biological samples for use in basic and applied research

Biobanking is an integral part of many scientific or clinical studies. The article aims to inform about the current state of studies, a set of data and results obtained thanks to the Medirex Group Academy projects (BIOMEDIRES II., PreveLynch, Promedicov, DIACOVID) for use in current and future basic and applied research and to define future perspectives and development strategies for the established biobank of samples and clinical data. In the article, we present examples of collected anamnestic and clinical data and discuss various advantages or disadvantages of the currently used approach of collecting samples and data for a better adjustment of the results of studies in the future.

Keywords: clinical study, research, biobank, biological samples

NewsLab, 2023; roč. 14 (1): 21 – 27

Úvod

Výber dizajnu štúdie je najdôležitejším krokom v metodológii výskumu^(1,2). Vhodný dizajn štúdie zaručuje dosiahnutie cieľov výskumu, preto je pri výbere dizajnu štúdie potrebné zväžiť niekoľko faktorov: formulovaná výskumná hypotéza, spôsob výberu účastníkov, spôsob zberu dát a odberu vzoriek^(1,3). Najpoužívanejší dizajn štúdie na skúmanie genetického a environmentálneho základu komplexného ochorenia je prípadová kontrolná štúdia. Takéto typy štúdií zahŕňajú vyšetrenie všetkých prípadov ochorenia (prípady) v porovnaní s reprezentatívnou vzorkou bez ochorenia (kontrolou). Prípady a kontroly sa zvyčajne skúmajú retrospektívne na dôkaz genetických a iných rizikových faktorov, ktoré existovali pred vypuknutím choroby, a tak pravdepodobne prispeli k rozvoju tohto ochorenia⁽⁴⁾. Viaceré štúdie tohto typu sme realizovali aj v neziskovej organizácii Medirex Group Academy (MGA, n. o.) v priebehu rokov 2020 – 2023. Tie sú súčasťou vedeckovýskumných projektov zameraných predovšetkým na oblasť onkológie (BIOMEDIRES II, PreveLynch) a oblasť infekčných a civilizačných ochorení (Promedicov, DIACOVID).

Projekt BIOMEDIRES II. je špecificky zameraný na problematiku karcinómu endometria a endometriálnych patológií vedúcich k vzniku karcinómu endometria. Cieľom projektu je nájsť potenciálne nové diagnostické, prognostické alebo terapeutické biomarkery detegovateľné pomocou tekutej biopsie, poukazujúce na riziko vzniku/rozvoja rakoviny endometria. Pri hľadaní biomarkerov sa plánuje využitie

umelej inteligencie, ktorá bude analyzovať výsledky získané viacerými laboratórnymi metódami predovšetkým z oblasti genomiky, proteomiky alebo glykomiky a bude ich porovnávať s dostupnými anamnestickými, klinickými a histologickými údajmi. Súčasťou projektu je klinická štúdia zameraná na získanie klinických vzoriek a dát pacientok s endometriálnou patológiou a karcinómom endometria⁽⁵⁾.

Projekt PreveLynch má za cieľ podrobne preskúmať problematiku dedičných nádorových ochorení spôsobených Lynchovým syndrómom a preskúmať predpokladané genetické pozadie vyššej prevalencie vybraných typov nádorových ochorení v slovenskej populácii špecificky so zameraním na kolorektálny karcinóm. Dôležitým cieľom projektu je vývoj menej invazívnych metód skríningu rakoviny založených na tekutej biopsii. Zavedenie neinvasívnych skríninových testov zahŕňajúcich inovatívne metódy by zlepšilo včasný záchyt jedincov s vyšším rizikom vzniku kolorektálneho karcinómu alebo iných typov karcinómov asociovaných s Lynchovým syndrómom, čo by malo za následok celkové zlepšenie zdravotného stavu bežnej populácie⁽⁶⁾.

Štúdia VA-COVID-19 podporuje dva projekty – Promedicov a DIACOVID. Oba projekty sú zamerané na ochorenie COVID-19 a súvisiace infekčné a civilizačné choroby. Hlavným cieľom projektu Promedicov je vytvorenie (univerzálneho) systému pre včasnú a rýchlu detekciu, identifikáciu a diagnostiku nových infekčných ochorení s pandemickým potenciálom, implementovaný v pilotnej fáze v priamej sú-

vislosti s aktuálnou pandémiou COVID-19. Cieľom projektu DIACOVID je prostredníctvom výskumu závažných civilizačných ochorení (diabetes mellitus, zápalové črevné ochorenia) a ich komplikácií spôsobených akútnymi ochoreniami dýchacích ciest zameraných na vírusové ochorenia (ako modelové ochorenie posluži COVID-19) získať také poznatky, ktoré umožnia podporu a rozvoj laboratórných a klinicky aplikovateľných inovatívnych postupov pre personalizovanú diagnostiku a terapiu takýchto pacientov^(7,8).

Do uvedených projektov boli pod odborným dohľadom garantov projektov navrhnuté klinické štúdie s ohľadom na naplnenie stanovených vedeckých cieľov. Získané vzorky pacientov a ich klinické údaje boli uložené v data- a biobanke MGA, z ktorej spolupracujúce výskumné inštitúcie (Vedecký park UK, Slovenská akadémia vied, Geneton, s. r. o.) mohli žiadať o výdaj týchto vzoriek a dať na výskum podľa ich stanovených vedeckovýskumných aktivít v projekte. Celkovo bolo k dnešnému dňu do uvedených štúdií zapojených 4 346 participantov s rôznymi diagnózami (endometriálne patológie, karcinóm endometria, kolorektálny karcinóm, Lynchov syndróm, COVID-19, diabetes mellitus, zápalové črevné ochorenia, kontrolné skupiny zdravých jedincov) a v biobanke uskladnených celkovo 18 958 vzoriek rôzneho typu materiálu (venózna krv, stolica, nazofaryngálny výter, natívne tkanivo, FFPE bločky). V článku opisujeme aktuálny stav realizovaných klinických štúdií, uvádzame príklady získaných anamnestických a klinických dát, biologických vzoriek na použitie v základnom a aplikovanom výskume.

Metódy

Etické komisie a informovaný súhlas pacienta

Klinické štúdie boli pred samotnou realizáciou schválené príslušnou multicentrickou a lokálnou etickou komisiou zapojených klinických pracovísk. Pre všetky štúdie bola multicentrickou etickou komisiou Etická komisia Bratislavského samosprávneho kraja (EK BSK), ktorá vydala kladné stanoviská pod unikátnym číslom pre jednotlivé štúdie (BIOMEDIRES II. – 06196/2020/HF; PreveLynch – 09834/2020/HF; VA-COVID-19 (DIACOVID a Promedicov) – 03228/2021/HF). Medzi zapojené nemocničné pracoviská patrila Fakultná nemocnica Trenčín, Nemocnica Agel, Východoslovenský onkologický ústav, Národný onkologický ústav, Univerzitná nemocnica Bratislava, Nemocnica sv. Michala, Univerzitná nemocnica Milosrdní bratia, Všeobecní fakultní nemocnice Praha). Všetci pacienti pred zapojením do štúdie podpísali informovaný súhlas, v ktorom boli uvedené všetky informácie týkajúce sa výskumu, odberového protokolu a využitia poskytnutých informácií a vzoriek na výskumné účely.

Zber anamnestických a klinických údajov

Od participantov zapojených v štúdiách sme získavali viacero typov informácií. Okrem anamnézy, ktorá pozostávala zo základných fyziologických a antropometrických údajov sa zbierali aj informácie o zdravotnom stave participanta, jeho klinickej histórii, užívaní liečiv, rodinnej anamnéze, rôzne sociodemografické ukazovatele až po výsledky laboratórných a experimentálnych vyšetrení. Všetky anamnestické a klinické údaje boli získavané predovšetkým dotazníkovou formou. Niektoré časti dotazníkov mohli účastníci vyplňať sami, išlo

najmä o časti, ktoré obsahovali otázky zamerané na životný štýl participanta, sociálnodemografické charakteristiky participanta. Ďalšiu časť tvorili otázky zamerané na klinickú anamnézu participanta a tieto časti vyplňali spoločne so zaraďujúcim lekárom, prípadne so sestrou. Pri vybraných diagnózach pacientov, napr. v prípade endometriálneho karcinómu, kolorektálneho karcinómu, sa zbierali aj údaje z biopsie, ktoré boli získané z histopatologického vyšetrenia realizovaného v rámci štandardného diagnosticko-liečebného postupu u pacienta. Niektorým skupinám pacientov boli urobené aj odbery na biochemické, sérologické a imunologické vyšetrenia. Výsledky týchto vyšetrení sa taktiež zbierali a zapisovali do elektronického systému na správu a manažment štúdií.

Odberový protokol

BIOMEDIRES II.

Odberový protokol pre pacientky zapojené do štúdie BIOMEDIRES II. pozostával z venózných odberov krvi a odberu tkaniva endometria vykonaného v rámci štandardných diagnosticko-liečebných postupov na základe stanovenej diagnózy. Od pacientok s patológiou endometria bolo pred kyretážou získaných 3x 10 ml krvi odobranej do EDTA Vacutainer skúmaviek určených na proteomické a genomické analýzy a po kyretáži a ukončení diagnostického procesu boli z patologického pracoviska získané FFPE bločky zdravého a patologického tkaniva. Pacientky s diagnózou, resp. podozrením na karcinóm endometria absolvovali odber pred zákrokom (hysterektómiou) pozostávajúci z venózneho odberu krvi (3x 10 ml EDTA Vacutainer) a odberu krvi po zákroku (3x 10 ml krvi EDTA Vacutainer). Po operácii boli z tkaniva maternice odobrané vzorky natívneho zdravého a nádorového tkaniva endometria do 2 ml skúmaviek (Eppendorf) so stabilizačným roztokom DNA/RNA Shield (Zymo Research) a po ukončení diagnostického procesu boli z patologického pracoviska získané FFPE bločky zdravého a nádorového tkaniva endometria.

PreveLynch

Do štúdie PreveLynch boli zapojené 3 skupiny pacientov – kontrolná skupina zdravých jedincov z populácie, pacienti s primárne diagnostikovaným kolorektálnym karcinómom, pacienti s Lynchovým syndrómom a ich rodinní príslušníci. Odberová schéma sa líšila vzhľadom na danú skupinu pacientov. Kontrolná skupina absolvovala venózny odber krvi 3x 10 ml do skúmaviek EDTA Vacutainer určených na proteomické a genomické analýzy a odber stolice 1x 10 ml do skúmavky so stabilizačným roztokom DNA/RNA Shield (Zymo research). Rovnakú odberovú schému absolvovali aj pacienti s Lynchovým syndrómom a ich rodinní príslušníci. Pacienti s diagnózou kolorektálneho karcinómu absolvovali odber krvi pred operáciou a po nej (resekcii čreva) v množstve 3x 10 ml EDTA Vacutainer, odber stolice 1x 10 ml do skúmavky s roztokom DNA/RNA Shield (Zymo research) pred zákrokom. Zároveň boli od týchto pacientov získané aj vzorky zdravého a nádorového natívneho tkaniva čreva a FFPE bločky.

Promedicov

Pacienti s rozdielnym priebehom COVID-19 boli súčasťou štúdie VA-COVID-19 a projektu Promedicov. V prípade COVID-19 pozitívnych pacientov bez ohľadu na priebeh sa vy-

konali opakované odbery krvi do biobanky (1x 5 ml krvi do sérum gélovej skúmavky, 1x 10 ml krvi do EDTA skúmavky) a nazo- a orofaryngálny výter (1x skúmavka s viRNAtrap médiom) v troch termínoch po diagnostikovaní infekcie. V týchto termínoch im bol zároveň urobený súbor biochemických, sérologických a imunologických vyšetrení na sledovanie vývoja infekcie. Pacientom, u ktorých sa infekcia nepotvrdila, boli odobrané len vzorky výteru, ktoré ostali po diagnostike a boli odoslané do biobanky v skúmavke s viRNAtrap médiom. Pacienti s postkovidovým syndrómom absolvovali odber 1x 10 ml krvi do EDTA skúmavky, ktorá bola uskladnená v biobanke na genomické analýzy.

DIACOVID

Do štúdie DIACOVID boli zaradení pacienti, u ktorých bol primárne diagnostikovaný diabetes mellitus (DM) a/alebo zápalové črevné ochorenie (IBD) (Ulcerózna kolitída alebo Crohnova choroba). Okrem uvedených diagnóz bolo pre týchto pacientov ďalším inklúznym kritériom prekonanie ochorenia COVID-19. Súčasťou štúdie bol aj nábor jedincov do kontrolnej skupiny, u ktorých bol anamnesticky vylúčený DM alebo IBD a ktorí prekonali COVID-19. Všetci pacienti absolvovali venózný odber krvi 3x 10 ml do EDTA Vacutainer skúmavky a odber stolice 1x 10 ml do skúmavky s roztokom DNA/RNA Shield (Zymo Research). Pacienti s IBD absolvovali navyše odber natívneho tkaniva, ktoré bolo získané biopticky pri endoskopickom vyšetrení indikovanom v rámci štandardných diagnosticko-liečebných postupov.

Primárne spracovanie a archivácia vzoriek v biobanke

Po doručení vzoriek do laboratória boli niektoré vzorky primárne spracované podľa štandardizovaného protokolu. Krv-

né vzorky odobraté do EDTA Vacutainer skúmavky určené na proteomické a genomické analýzy boli spracované jednokrokovou (2 200 g, 10 min, 4 °C) a dvojkrokovou (2 200 g, 10 min, 4 °C; 16 000 g, 10 min, 20 °C) centrifugáciou a krvné vzorky odobraté do sérum gélovej skúmavky boli spracované jednokrokovou centrifugáciou (3 000 g, 10 min, 4 °C). Po centrifugácii boli vzorky uchované v podobe alikvot plnej krvi (2-3x 500 µl), plazmy (proteomika – 2x 500 µl, zvyšné alikvoty po 1 800 µl; genomika – 2x 680 µl, zvyšné alikvoty po 1 800 µl) alebo séra (2x 1 ml). Vzorky stolíc a výterov boli pripravené do podoby 1 ml alikvot a uskladnené v biobanke. Natívne tkanivá a FFPE bločky boli uskladnené bez primárneho spracovania.

Výsledky

Aktuálny stav nábora pacientov v klinických štúdiách

BIOMEDIRES II.

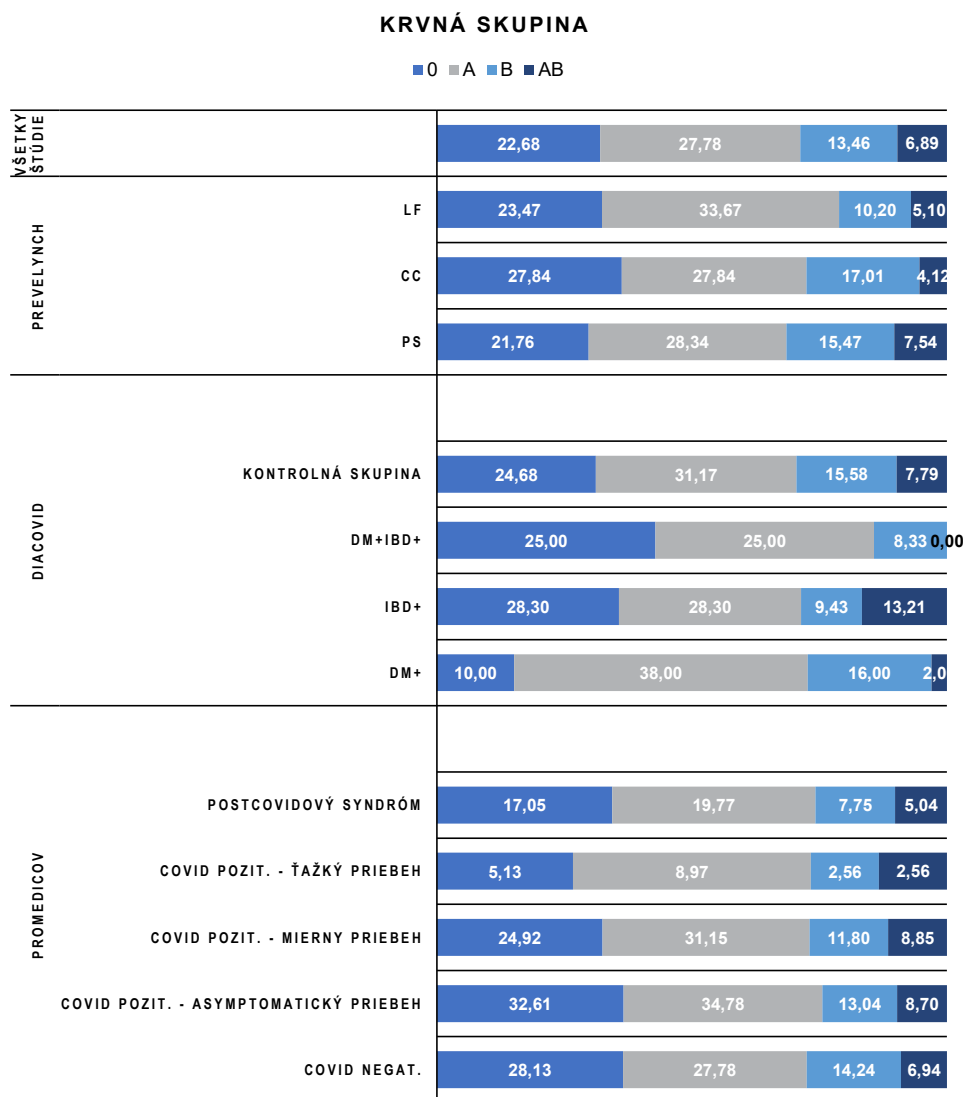
Nábor pacientov do štúdie BIOMEDIRES II. začal v novembri 2020. Od tohto obdobia sme do štúdie postupne zapájali rôzne klinické pracoviská z celého Slovenska. Nábor pacientov do vetvy endometriálne patológie (EP) prebiehal v 6 gynekologických ambulanciách/pracoviskách (Bari group, s. r. o., Gynecare, spol. s r. o., Gyneka, s. r. o., GYN. CENTRUM, spol. s r. o., Nemocnica Agel Košice Šaca, a. s., Nemocničná, a. s. – gynekologická ambulancia, Univerzitná nemocnica Bratislava-Ružinov – onkogynekologická ambulancia). Pacientky s diagnózou endometriálneho karcinómu boli zaradené na špecializovaných onkogynekologických pracoviskách v 5 nemocniciach (Nemocnica Agel Košice-Šaca, Východoslovenský onkologický ústav, Fakultná nemocnica Trenčín,

Tabuľka 1. Súbor vybraných základných fyziologických a antropometrických anamnestických informácií (pohlavie, vek, výška, váha, obvod pásu, BMI) zbieraných u pacientov v realizovaných štúdiách (aktualizované k 31. 3. 2023).

	Skupina pacientov	Počet zaradených pacientov	Počet pacientov s kompletným dotazníkom	Pohlavie				Priemerný vek Medián (cm)	Priemerná výška Medián (cm)	Priemerná váha Medián (kg)	Obvod pásu Medián (cm)	BMI
				Muž		Žena						
		N	N	N	%	N	%					
BIOMEDIRES II.	EP	1508	1503	0	0,00 %	1503	100,00 %	48	165	73	x	x
	EC	268	257	0	0,00 %	257	100,00 %	66	164	83	x	x
Preve-Lynch	PS	1034	998	350	35,07 %	648	64,93 %	29	171,5	65,5	78	22
	CC	194	161	84	52,17 %	77	47,83 %	63	170	78	80	26
	LF	98	97	32	32,99 %	65	67,01 %	46	168	70	80	24
DIACOVID	DM+	50	48	25	52,08 %	23	47,92 %	49,5	174	88,5	94	27,5
	IBD+	53	52	29	55,77 %	23	44,23 %	39	176	77,5	90	24
	DM+IBD+	12	11	5	45,45 %	6	54,55 %	55,5	172	90	100	28
	Kontrolná skupina	154	153	37	24,18 %	116	75,82 %	34	170	68	81	22
Promedico	COVID negat.	288	281	146	51,96 %	135	48,04 %	40	175	78	80	25
	COVID pozit. - asymptomatickí	46	46	26	56,52 %	20	43,48 %	42	176,5	78	89	25
	COVID pozit. - mierny priebeh	305	305	126	41,31 %	179	58,69 %	39,5	172	75	88	24
	COVID pozit. - ťažký priebeh	78	75	43	57,33 %	32	42,67 %	63,5	170	88	0	29
	Postkovidový syndróm	258	256	138	53,91 %	118	46,09 %	58	172	90	107	30
Všetky štúdie		4346	4243	1041	24,53 %	3202	75,47 %	48	171	79	81	26

EP – Endometriálna patológia; EC-Endometriálny karcinóm; PS – Populačná štúdia; CC – kolorektálny karcinóm, LF – Lynchove rodiny; DM + – pacienti s diabetes mellitus; IBD+ – pacienti so zápalovým črevným ochorením; DM+IBD+ – pacienti s diabetes mellitus a zápalovým črevným ochorením

Graf 1. Zastúpenie krvných skupín u pacientov podľa štúdie a zaradenia do skupiny v štúdiu a celkové zastúpenie krvných skupín bez ohľadu na zaradenie do štúdie prípadne skupiny v štúdiu (aktualizované k 31. 3. 2023).



EP – Endometriálna patológia; EC – Endometriálny karcinóm; PS – Populačná štúdia; CC – kolorektálny karcinóm, LF – Lynchove rodiny; DM + – pacienti s diabetes mellitus; IBD+ – pacienti so zápalovým črevným ochorením; DM+IBD+ – pacienti s diabetes mellitus a zápalovým črevným ochorením

Všeobecní fakultní nemocnice Praha, Národní onkologický ústav). Do konca marca 2023 v štúdiu evidujeme zaradených 1 508 pacientok s endometriálnou patológiou a 268 pacientok s karcinómom endometria (**tabuľka 1**).

PreveLynch

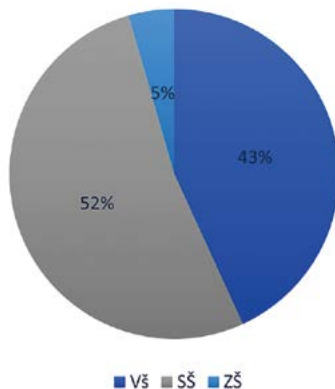
Nábor pacientov do štúdie PreveLynch bol spustený v máji 2021 v gastroenterologickom centre v Bratislave (IBD centrum). Toto centrum zabezpečuje nábor pacientov do populačnej vetvy štúdie a tiež do vetvy pacientov s Lynchovým syndrómom. V populačnej vetve je zaradených 1 034 pacientov a vo vetve Lynchove rodiny 96 pacientov (stav k 31. marcu 2023). Pacienti s diagnózou kolorektálneho karcinómu boli zaradovaní do štúdie na chirurgických klinikách 6 nemocníc (Nemocnica Agel Košice-Šaca, Univerzitná nemocnica s poliklinikou Milosrdní bratia, Národní onkologický ústav, Univerzitná nemocnica sv. Michala, Univerzitná nemocnica

Bratislava-Kramáre, Fakultná nemocnica Trenčín). Súčasťou tejto vetvy štúdie je 194 pacientov (stav k 31. marcu 2023) (**tabuľka 1**).

Promedicov

Nábor pacientov s diagnózou COVID-19 začal v apríli 2021. Pacientov s bezpríznakovým a miernym prejavom ochorenia, prípadne kontrolnú skupinu COVID negatívnych sme získavali prostredníctvom odberových centier Medirex, a. s., 2 nemocničné pracoviská zabezpečovali nábor pacientov s ťažkým priebehom COVID-19 (Univerzitná nemocnica Bratislava-Kramáre, Univerzitná nemocnica Bratislava-Ružinov). Sanatórium Dr. Guhra v Tatranskej Polianke bolo zapojené do nábora pacientov s postkovidovým syndrómom. Nábor pacientov je ukončený a aktuálny stav počtu zaradených pacientov v jednotlivých skupinách predstavuje 288 COVID negatívnych, 48 COVID pozitívnych asymptomatických, 305

Graf 2. Najvyšší typ dosiahnutého vzdelania u pacientov zaradených v štúdiách bez ohľadu a typ štúdie alebo skupinu pacientov v štúdií.



VŠ – vysoká škola; SŠ – stredná škola; ZŠ – základná škola

COVID pozitívnych s miernym priebehom, 78 COVID pozitívnych s ťažkým priebehom a 258 pacientov s postkovidovým syndrómom (**tabuľka 1**).

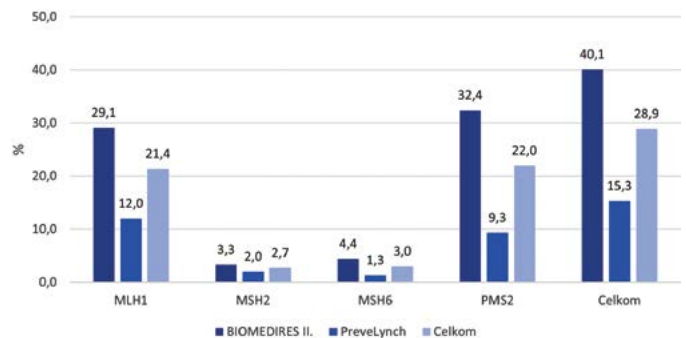
DIACOVID

Nábor pacientov do štúdie DIACOVID sa začal v júni 2022 v gastroenterologickom centre IBD a na internej klinike Univerzitnej nemocnice Bratislava-Ružinov. V tejto štúdií aktuálne evidujeme 50 pacientov s diagnózou DM, 53 pacientov s IBD, 11 pacientov s IBD a DM a 154 pacientov v kontrolnej skupine (stav k 31. marcu 2023) (**tabuľka 1**).

Analýza vybraných anamnestických a klinických údajov

Súčasťou odberovej schémy u všetkých pacientov bolo získanie anamnestických a klinických informácií prostredníctvom anamnestického dotazníka, ktorý vyplňal samotný pacient, prípadne lekár s pacientom, ak išlo o zložitejšie typy otázok ako napríklad klinická alebo rodinná anamnéza. Pri realizácii bolo naším cieľom získať čo najkonzistentnejšie dáta o pacientoch naprieč štúdiami, aby sme v prípade potreby vedeli tieto dáta porovnávať aj medzi pacientmi z rôznych štúdií a získali sme tak viac informácií aj pre prípadné použitie v budúcom výskume. Príkladom rovnakých dát, ktoré sa zbierali naprieč štúdiami, sú základné antropometrické a fyziologické informácie o účastníkovi. Na účely tohto článku sme si vybrali na porovnanie niekoľko základných údajov o participantovi, ako je pohlavie, vek, výška, váha, obvod pása, BMI. Detailné výsledky z analýzy týchto údajov u participantov zaradených v jednotlivých štúdiách a skupinách, ako aj sumárne výsledky za všetky štúdie sú zhrnuté v **tabuľke 1**. Hoci celkovo bolo do štúdií zaradených 4 346 participantov, anamnestický dotazník k 31. marcu 2023 vyplnilo 4 243 pacientov. Z analyzovaných informácií vyplýva, že v súbore zaradených pacientov máme 75,47 % žien a 24,53 % mužov. Priemerný vek pacientov v súbore je 48 rokov. Priemerná výška pacientov je 171 cm, hmotnosť 79 kg, obvod pása 81 cm a BMI 26. V štúdií BIOMEDIRES II. sa údaje ako obvod pása, BMI a krvná skupina nezberali. Tieto údaje plánujeme v budúcnosti využiť pri koreláciách s výsledkami experimentálnych analýz a budú primárne vyhodnocované na úrovni jednotlivých štúdií.

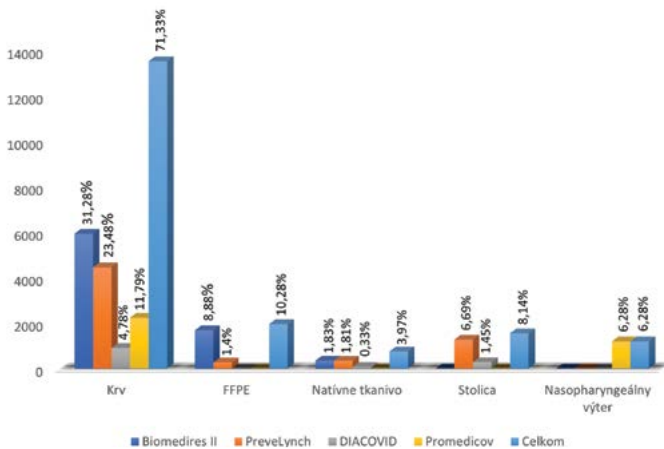
Graf 3. Príklad vybraných sledovaných klinických údajov u pacientov s endometriálnym a kolorektálnym karcinómom zo štúdie BIOMEDIRES II. a PreveLynch – Strata expresie proteínov MMR systémom detegovaná na základe výsledkov imunohistochemického vyšetrenia.



Súčasťou základných fyziologických a antropometrických informácií, ktoré sa u pacientov zisťovali, bola aj krvná skupina a vzdelanie. V súbore máme najvyšší podiel pacientov s krvnou skupinou A (27,78 %) a 0 (22,68 %), menej sa vyskytujú pacienti s krvnou skupinou B (13,46 %) a najmenej so skupinou AB (6,89 %). Podrobné rozdelenie výskytu krvných skupín u pacientov zaradených v jednotlivých vetvách štúdií nájdete v **grafe 1**. Ďalším sledovaným parametrom bolo vzdelanie. Z analýzy vyplýva, že 52 % participantov uviedlo ako svoje najvyššie dosiahnuté vzdelanie stredoškolské, 43 % dosiahlo vysokoškolské vzdelanie a 5 % pacientov ukončilo základné vzdelanie (**graf 2**).

V štúdiách sme zbierali aj vybrané demografické a klinické informácie dotazníkovou formou, medzi ktoré patrila podrobná klinická anamnéza pacienta so zameraním na akútne alebo prekonané ochorenia, absolvovanie transfúzie, transplantácie, užívané liečivá, podrobná rodinná anamnéza. Všetky tieto informácie budú vyhodnocované a korelované s výsledkami jednotlivých štúdií a niektoré z nich môžu byť v budúcnosti vyhodnocované medzi štúdiami. Pri každej zo štúdií sa zbierali aj štúdiivo špecifické dáta, ktoré sa vzťahovali len na určitú skupinu/skupiny participantov s danou diagnózou. Na účely tohto článku sme si vybrali ako príklad zbieraných klinických dát výsledky imunohistochemického (IHC) vyšetrenia *mismatch repair* proteínov (MMR), ktoré bolo realizované u pacientov s kolorektálnym karcinómom a s karcinómom endometria v rámci štandardných diagnosticko-liečebných postupov a výsledky tohto vyšetrenia zapísal na účely štúdie vyšetrujúci patológ formou vyplnenia bioptického nálezu do informačného systému. Imunohistochemické vyšetrenie MMR proteínov slúži na primárnu selekciu potenciálnych prípadov s Lynchovým syndrómom. Najčastejšie sa vyšetruje strata expresie Mlh1, Msh6, Msh2 a Pms2 MMR proteínov. Neprítomná expresia jedného alebo viacerých MMR proteínov môže poukazovať na Lynchov syndróm, a preto je potrebné na potvrdenie/vylúčenie diagnózy Lynchovho syndrómu vykonať genetickú analýzu⁽⁹⁾. V štúdií BIOMEDIRES II. sme mali k dispozícii výsledky imunohistochemického vyšetrenia 182 pacientov s endometriálnym karcinómom a v štúdií PreveLynch boli tieto výsledky dostupné u 150 pacientov

Graf 4. Počet a percentuálne zastúpenie primárnych vzoriek v biobanke MGA podľa typu biologického materiálu zozbieraného v priebehu 4Q/2020 - 1Q/2023 z klinických štúdií BIOMEDIRES II., PreveLynch, DIACOVID, Promedicov.



s kolorektálnym karcinómom. Stratu expresie aspoň jedného z proteínov MMR systému sme detegovali u 73 pacientok zo štúdie BIOMEDIRES II. a u 23 pacientov zo štúdie PreveLynch. Celkovo bola strata expresie niektorého z MMR proteínov detegovaná u 28,9 % pacientov z celkového počtu pacientov s dostupným výsledkom IHC. Najčastejšie bola detegovaná strata expresie proteínu Mlh1 (21,4 %) a proteínu Pms2 (22 %). Menej častá bola strata proteínu Msh6 (3 %) a Msh2 (2,7 %) (**graf 3**). Vzorky týchto pacientov boli vybrané na celogenómovú analýzu, ktorá sa vykonáva v rámci predmetných štúdií. Táto analýza by mala diagnózu Lynchovho syndrómu potvrdiť alebo vylúčiť.

Aktuálny stav vzoriek v biobanke

V súčasnosti (k 31. marcu 2023) je biobanka úložiskom 18 958 rôznych biologických vzoriek (krv, stolica, nazofaryngálny výter, natívne tkanivo, FFPE bločky) uskladnených vo forme alikvot získaných od 4 346 pacientov zúčastňujúcich sa na projektoch BIOMEDIRES II., PreveLynch, Promedicov a DIACOVID. Prvé vzorky sa začali zbierať v novembri 2020 (štúdia BIOMEDIRES II.), v máji 2021 (štúdia PreveLynch), v apríli 2021 (štúdia Promedicov) a v júni 2022 (štúdia DIACOVID). V biobanke je uskladnených 13 523 krvných vzoriek, 1 949 FFPE, 1 544 vzoriek stolíc, 1 190 vzoriek nazo- a orofaryngálnych výterov a 752 vzoriek natívnych tkanív. Percentuálne zastúpenie vzoriek v biobanke je znázornené v **grafe 4**. Zber a uchovávanie vzoriek v biobanke pokračuje počas celého trvania jednotlivých projektov. Archivovaný biologický materiál sa postupne využíva na rôzne typy analýz (genomika, proteomika, glykomika, sérológia, imunológia, histológia). Doteraz bolo pre participujúce partnerské výskumné organizácie (Slovenská akadémia vied, Vedecký park Univerzity Komenského, Medirex Group Academy, Geneton, s. r. o.) uvoľnených viac ako 2 500 alikvot na genomickú a proteomickú analýzu zo vzoriek výterov, plazmy, plnej krvi, natívnych vzo-

riek zdravého a nádorového tkaniva. Výsledky realizovaných experimentálnych analýz sú aktuálne vo fáze spracovania a budú súčasťou odborných publikácií, ktoré budú postupne publikované v priebehu rokov 2023, 2024 a 2025.

Diskusia

Pri realizácii akéhokoľvek klinického výskumu je dôležité zvážiť, aké údaje a vzorky má zmysel zbierať na overenie hypotézy a dosiahnutie cieľov predmetného výskumu. Zbierať príliš veľa údajov je niekedy kontraproduktívne a môže predstavovať zbytočnú záťaž pre účastníkov štúdie, ako aj viesť k zberu údajov, ktoré nebudú použité, analyzované a prezentované vo vedeckej publikácii. Preto je dôležité pred začatím každej štúdie identifikovať skupiny dát, ktoré sú v kontexte skúmanej problematiky zaujímavé, majú význam a prínos pre následnú analýzu, a to výrazne pomôže nastaviť optimálny a efektívny spôsob zberu dát⁽¹⁰⁾. Zo skúseností z realizovaných štúdií sme zistili, že najlepšou formou zberu dát pre klinické štúdie, ktoré najčastejšie realizujeme, je dotazníková forma. Pri niektorých štúdiách boli dotazníky pripravené príliš detailne, zdĺhavo a odborne, čo negatívne ovplyvnilo mieru ich vyplnenia účastníkmi štúdie. Do budúca by sme sa chceli zamerať na skrátenie a zjednodušenie anamnestických a klinických dotazníkov, aby sme získali čo najviac podstatných informácií v čo najlepšej možnej kvalite. Pri realizácii štúdií sa osvedčilo využitie elektronického systému, do ktorého mali prístup lekári, ktorí vedeli dotazník, ale aj iné informácie vyplniť priamo elektronicky, čo umožnilo okamžitý prístup k dátam pre koordinátorov jednotlivých štúdií, ale aj participujúcich vedcov. S využívaním elektronického systému by sme chceli pokračovať aj pri ďalších štúdiách. Pri nastavovaní odberových schém, ktorých súčasťou bol zber biologického materiálu do biobanky, by v budúcnosti bolo prínosné prekonzultovať návrh odberovej schémy s odborníkmi, klinikmi, ktorí budú spolupracovať na nábore pacientov tak, aby boli odbery robené pre výskum čo najviac v súlade so štandardnými diagnosticko-liečebnými postupmi a tak, aby bola fyzická, ale aj časová záťaž pre zapojeného pacienta a lekára čo najmenšia a čo najmenej narúšala ich štandardné postupy. Naším dlhodobým cieľom je pokračovať v budovaní biobanky biologických vzoriek a klinických dát, preto plánujeme v realizácii ďalších klinických štúdií pokračovať a ďalej pracovať na vylepšovaní a optimalizovaní procesov zberu dát, vzoriek a biobankingu.

Podakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum progresívnych metód diagnostiky COVID-19 a biomarkerov umožňujúcich skorú detekciu jedincov so zvýšeným rizikom ťažkého priebehu ochorenia, kód ITMS: 313011ATA2, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

1. Omair A: Selecting the appropriate study design for your research: descriptive study designs. *J Health Spec.* 2015, 3: 153–156. 10.4103/1658-600X.159892
2. Gange SJ, Golub ET: *Study Design. Infectious Disease Epidemiology: Theory and Practice*, 3rd Edition. Nelson KE, Williams CM (ed): Jones & Bartlett Learning, Burlington; 2014. 45–76.
3. Bickman L, Rog DL: *Applied Research Design: A Practical Approach. The SAGE Handbook of Applied Social Research.* Bickman L, Rog DL (ed): SAGE Publications, London; 2009. 3–46. 10.4135/9781483348858.n1
4. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet.* 2006 Oct; 7(10): 812–820. doi: 10.1038/nrg1919. PMID: 16983377.
5. <https://www.medirexgroupacademy.sk/projekt/biomedires-ii/>
6. <https://prevelynch.sk/>
7. <https://www.medirexgroupacademy.sk/projekt/promedicov-19/>
8. <https://diacovid.sk/>
9. Umar, A., et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst.* 2004. 96(4): p. 261–268.
10. Saczynski JS, McManus DD, Goldberg RJ. Commonly used data-collection approaches in clinical research. *Am J Med.* 2013 Nov; 126(11): 946–950. doi: 10.1016/j.amjmed.2013.04.016. Epub 2013 Sep 16. PMID: 24050485; PMCID: PMC3827694.

Mgr. Petra HirjakováMEDIREX GROUP ACADEMY, *n. o.*

Novozámocká 67, 949 05 Nitra

e-mail: Petra.Nemcova@medirexgroup.sk

Clinically relevant RNA-virus whole genome assembly testing on metatranscriptomic data gained directly from COVID-19 clinical samples

Dominik Hadzega, Klaudia Babisova, Patrik Krumpolec, Michaela Hyblova, Oliver Petrovic, Gabriel Minarik

MEDIREX GROUP ACADEMY, n. o., Nitra

Recent pandemics of COVID-19 showed importance of modern methods of molecular biology and bioinformatics in understanding and dealing with such events. In our project, we were sequencing nasopharyngeal swabs from COVID-19 patients. Here in this study, which is a part of mentioned bigger project, we bring brief report of testing RNA virus assembly possibilities for short read Illumina sequencing. We were dealing with metatranscriptomic data from COVID-19 positive patients. There are multiple options of assemblers for RNA genomes. We tested performance of multiple strategies of genome assembly by SPAdes and Trinity. We were evaluating and comparing quality of assemblies with help of Quast tool. We observed, that outcome of assembly is affected by chosen strategy, but we were able to construct genomes only from our short-read sequencing even *de novo* without using reference genome. Additionally, we assigned assembled genomes to SARS-CoV-2 lineage.

Keywords: RNA virus genome assembly, SPAdes, Trinity, SARS-CoV-2, RNA-seq

Skladanie genómu klinicky relevantných RNA vírusov z klinických vzoriek pochádzajúcich od COVID-19 pacientov
Počas pandémie ochorenia COVID-19 sa dostali do popredia viaceré moderné metódy molekulárnej biológie a bioinformatiky, ktoré nám pomáhali lepšie porozumieť pandémie a pomáhali s ňou bojovať. V tejto publikácii, ktorá je súčasťou väčšieho projektu sekvenovania nazofaryngálnych sterov s cieľom štúdie ochorenia COVID-19, sme sa venovali testovaniu skladania genómov RNA vírusov z krátkych čítaní z Illumina sekvenovania. Dáta od pacientov s COVID-19 analyzované v štúdiu mali charakter metatranskriptómového sekvenovania. V súčasnosti je k dispozícii niekoľko softvérov vyvinutých na skladanie RNA genómov. My sme testovali úspešnosť niekoľkých stratégií skladania genómov s nástrojmi SPAdes a Trinity. Vyhodnocovali sme a porovnávali kvalitu poskladaných genómov s pomocou nástroja Quast. Ukázalo sa, že výsledok je naozaj ovplyvnený zvolenou stratégiou, pričom sme však boli z našich krátkych čítaní schopní zostaviť genómy aj na spôsob *de novo* bez použitia referenčného genómu. Poskladané genómy sme priradili variantom vírusu SARS-CoV-2.

Kľúčové slová: skladanie genómu RNA vírusu, SPAdes, Trinity, SARS-CoV-2, RNA sekvenovanie

NewsLab, 2023; roč. 14 (1): 28 – 32

Introduction

In this study we aimed to test possibility of assembling whole genomes of RNA viruses from short reads metatranscriptomic sequencing. There are several genome assemblers that are commonly used for this purpose. Here we tested SPAdes and Trinity^(1,2). However, there are other choices, for example AbySS⁽³⁾, Velvet⁽⁴⁾, IDBA⁽⁵⁾ or Oases⁽⁶⁾. Different assemblers vary depending on the specific characteristics of the RNA sequencing data and the species being studied. Hölzer and Marz (2019) identified several factors that affects quality of transcriptomic assembly, including read length, sequencing depth, and transcriptome complexity⁽⁷⁾. From these results imply it is necessary to evaluate the performance of different assemblers on a case-by-case basis.

There are multiple parameters that can be evaluated quality of the assembly. One is Contiguity, which measures the length of the contigs and scaffolds produced by the assembly. Standard metrics used here are N50, L50 or N90 values.

Other is completeness, measured by the proportion of the genome that has been assembled. Then there is accuracy, or in different words - how well the assembled genome matches the true sequence of the virus.

Here we assembled viral genomes directly from nasopharyngeal swabs samples of COVID-19 patients, evaluated quality of assemblies and assigned them to SARS-CoV-2 lineages.

Methods

Study Approval

Sample collection was performed as part of the clinical study approved by the Ethical committee of Bratislava Self-Governing District under the identifier 03228/2021/HF from January 12, 2021. All patients have filled the questionnaires with relevant information regarding their health status in relation to COVID-19 and signed informed consent.

Table 1. Strategies of SARS-CoV-2 assembly.

Strategy	Reads filtering	Assembler	Mode
1	mapping on SARS-CoV-2 genome	coronaspad.es.py	
2	not mapping on human hg38 genome	coronaspad.es.py	
3	not mapping on human hg38 genome	Spades	--rna
4	not mapping on human hg38 genome	Spades	--rnaviral
5	mapping on SARS-CoV-2 genome	Trinity	

Sample preparation

Samples

Nasopharyngeal swabs from COVID-19 suspected patients were gained in the two basic regimes. The patients hospitalized with middle to severe disease symptoms were recruited to the study in the cooperating hospitals. Patients with mild or any disease symptoms were recruited in the SARS-CoV-2 testing facilities of Medirex Inc. routinely collecting and analyzing samples from the population in the COVID-19 pandemic.

Nucleic acid extraction

Nasopharyngeal swabs specimens were collected from COVID-19 patients and controls and stored in viRNAtrap collection medium (GeneSpector, Czech Republic) at 4 °C. Total nucleic acid was extracted using Sera-Xtracta Virus/Pathogen Kit (Cytiva, UK) according to manufacturer instruction. 400 µl of the nasopharyngeal swab medium was used for nucleic acid extraction with final elution to 50 µl of nuclease-free water. RNA was quantified with the Qubit™ RNA High Sensitivity Assay Kit (Invitrogen). RNA isolates were stored at -80 °C.

RNA library preparation and sequencing

The metatranscriptomic libraries were prepared using KAPA RNA HyperPrep Kit with RiboErase (HMR) (Kapa Biosystems, South Africa) according to the original protocol of manufacturer. For quantity and quality control of prepared libraries a Qubit 1X dsDNA High Sensitivity Assay Kit on Qubit 3.0 (Invitrogen) and Agilent High Sensitivity DNA Kit on Agilent 2100 Bioanalyzer (Agilent) instruments were used. Sequencing of pooled libraries was performed on NextSeq 500 and NextSeq 2000 (Illumina) platforms using 2x75 or 2x100 paired-end sequencing setup, respectively.

Data analysis

Quality Control and Data Preparation for Analysis

First step of any analysis of RNA-seq data is quality control and this step was done by FastQC v0.11.9⁽⁸⁾. Reads were processed by Trimmomatic v0.39 (CROP:96 HEADCROP:10 LEADING:22 TRAILING:22 SLIDINGWINDOW:4:22 MINLEN:25 and our own set of adapter sequences were used in ILLUMINACLIP step)⁽⁹⁾. Parameters were chosen according to FastQC results.

Reads mapping

Reads were mapped to the human genome hg38 by BWA-MEM algorithm of bwa v0.7.17 package⁽¹⁰⁾. Reads were mapped as paired set, otherwise parameters of mapping

were set to default. Same way it was done on SARS-CoV-2 genome.

Genome assembly

We performed an assembly of SARS-CoV-2 genome using coronaspades.py tool from Spades. We tested multiple strategies of assembly. In 2 strategies we proceeded with reads mapping on SARS-CoV-2 genome. Reads mapped by BWA-MEM were extracted using samtools view from samtools v1.6⁽¹¹⁾ and Picard SamToFastq from Picard v2.27.4⁽¹²⁾. Subsequently, Spades and Trinity assemblers were applied^(1,2). In other cases, we tested Spades assembler on reads that didn't map to human genome hg38. All strategies tested are shown on the **Table 1**.

For deeper evaluation and comparison of performance (resemblance to a reference genome, N50 value and other values), we used command line (Conda) instance of Quast software⁽¹³⁾.

Scaffolds taxonomy classification

Scaffolds were classified by Galaxy version 2.10.1. of Blast, with megablast algorithm⁽¹⁴⁾. Databases NCBI NT, RefSeqRNA, SILVA rRNA and Metagenomes & metatranscriptomes DB were used as subject database. Number of hits were limited to 1.

SARS-CoV-2 lineage assignment

To assign assembled scaffolds with SARS-CoV-2 lineage, Nextstrain web-based tool was used (<https://clades.nextstrain.org/>, accessed on 20.04.2023). For confirmation, we used Galaxy pipeline - „Mutation calling, viral genome reconstruction and lineage/clade assignment from SARS-CoV-2 sequencing data“⁽¹⁵⁾, with reads mapping on SARS-CoV-2 genome as an input. First step of the pipeline is variation analysis with key components: BWA-MEM for mapping, Lofreq for variant calling and SnpEff for variants' annotation^(16,17). Then, next step is variant reporting. The third step is to generate consensus sequences and then identify SARS-CoV-2 clades/lineages by Pangolin and Nextclade^(18,19).

Results

We sequenced in total 79 samples from COVID-19 positive human patients. We performed an assembly of the SARS-CoV-2 genome using Spades and Trinity assemblers. We used multiple algorithms and strategies of assembly of Spades and compared it between each other and with Trinity assembler.

First, we tested assembly with all advantages we have – using reference genome and Spades mode for assembly of SARS-CoV-2 genome. Secondly, we observed if there would

be difference if we give up advantage of filtering reads by mapping on SARS-CoV-2, so all reads that didn't map on human genome were used as an input for an assembly. However, to test the ability of assembling genome of unknown RNA viruses, we needed to test another strategies - Spades modes for assembly of transcriptome or RNA viruses (spades with `-rna` and `-rnavirus` modes). We also tried another widely used transcriptome assembler Trinity. To compare its performance for our purpose to performance of Spades, we assembled genomes from reads mapping to SARS-CoV-2.

To evaluate quality of assemblies we used multiple parameters, as it is shown on the Table 2. To complexly evaluate qualities of assemblies, software Quast was used. Results from strategies using human unmapped reads instead of SARS-CoV-2 reads are influenced by RNA from other organisms – some human RNAs left in data and bacterial RNA. Because of this influence, there are additional contigs influencing quality metrics. However, “largest alignment” and “genome fraction” statistic show, that many genomes were assembled as whole or in large fragments, it is only necessary to add one more step of cleaning additional contigs. Also, this is better shown by NGA50 statistic, counting with only reference-aligned contigs. This statistic shows similar values for all strategies used. In total 42 assemblies covered at least 98% of the reference genome (some of them fragmented in more scaffolds). From most of the quality metrics it seems using mapping to the reference genome brings advantage to assembly procedure as expected. Here, some metrics were in favour of Spades, others with Trinity. Spades results were *cleaner* – with high quality and containing fewer small fragments not incorporated into genome and contigs of different origin (using coronaspades.py). It is also shown the assembly with filtering out only host reads can produce complete or almost complete genomes too. Here, it is possible to use modes for coronavirus, RNA or RNA virus. Difference is in cleanliness of the results, since non-specific assembly modes produce contigs unaligned to the reference genome. The possibility of assembly in all strategies is graphically shown in the **Figure 1**.

From scaffolds produced by SPAdes from human-unmapped reads, none of the clinically relevant viruses (other than SARS-CoV-2) were identified by Blast (scaffolds

Figure 1. Quast histograms of genome fractions showing portion of reference genome being assembled (A) and Complete genomic features in assembled genomes (B) for all strategies used for genome assembly (1 – Coronaspades on SARS-CoV-2 mapped reads, 2 – Coronaspades on human unmapped reads, 3 – Spades `-rna` on human unmapped reads, 4 – Spades `-rnavirus` on human unmapped reads, 5 – Trinity on SARS-CoV-2 mapped reads) as one of the quality metrics, shown by samples. In the graph, all at least partially assembled samples are shown. Every column in the graph represents completeness of assembly in one of the samples.

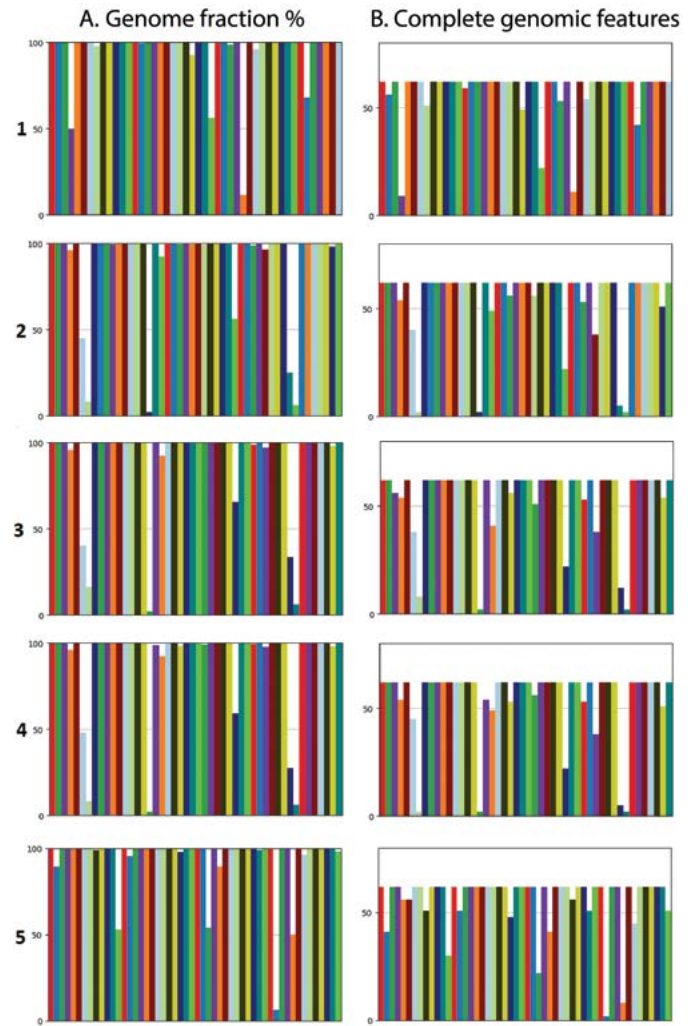
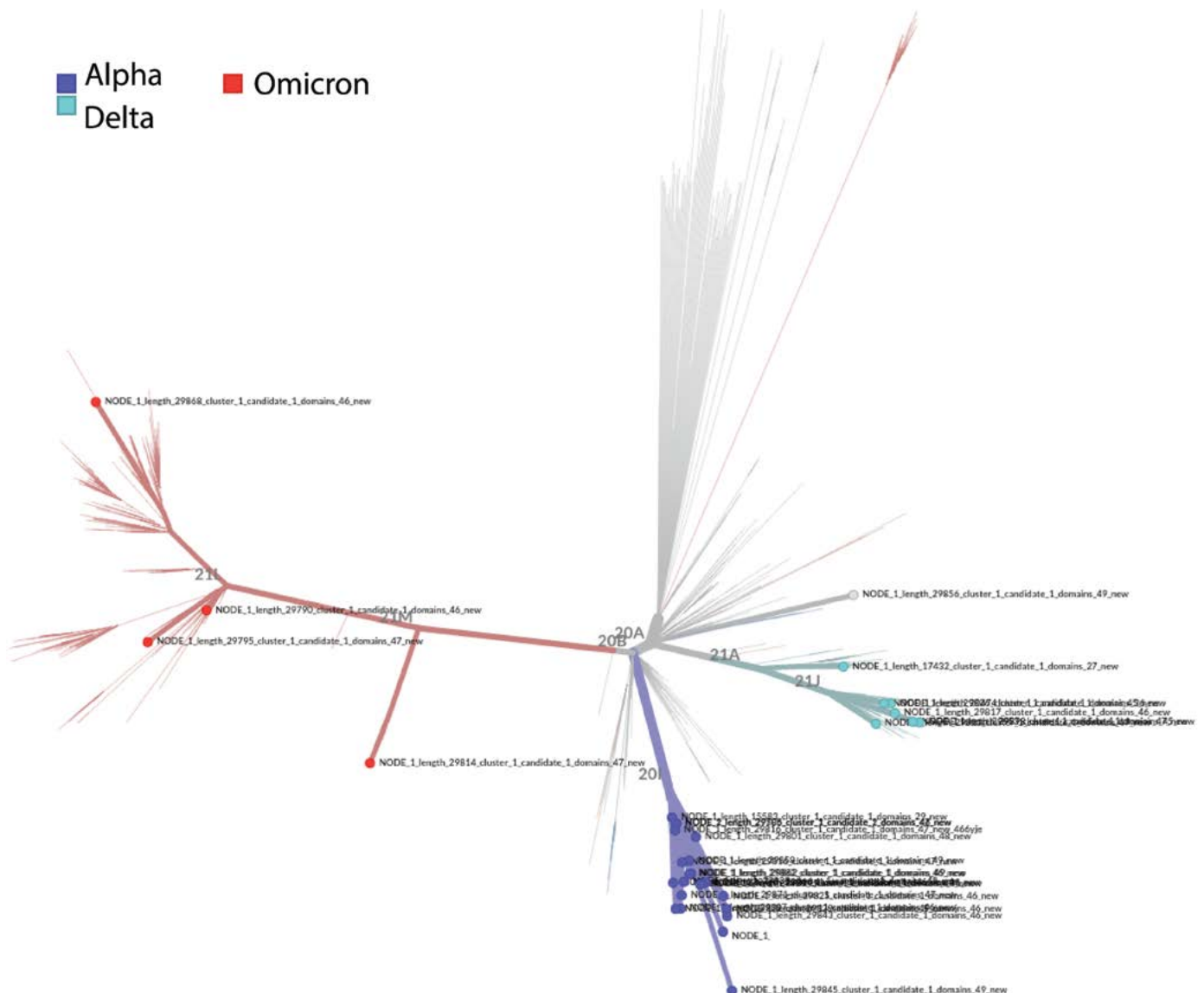


Table 2. Basic stats for assembly qualities for different strategies (As they are numbered in Table 1 in Methods section). For better understanding of statistics used by Quast, please see the authors' paper (Gurevich et al., 2013).

Mean of assembly stats	strategy1	strategy2	strategy3	strategy4	strategy5
N50	24955.39	13924.6	1438.31	1140.62	23271.15
N90	24098.5	4881.32	617.44	574.74	8583
NGA50	25968.11	26133.12	25561.3	24464.98	24282.3
NG90	26334.36	21637.39	20540.79	17102.78	25019.87
Largest alignment	25179.96	23776.4	23362.27	22285.85	23707.04
Genome fraction	94.86	89.91	90.42	90.11	94.68
GC % deviation from reference	0.16	3.05	8	7.25	0.66
Number of samples with values fitting criteria					
Genome fraction > 98	42	39	42	42	46
Genome fraction > 50	44	43	43	43	51
Largest alignment > 28 000	36	35	34	21	37
Largest alignment > 10 000	37	38	38	38	40

Figure 2. Scaffolds assignment to SARS-CoV-2 lineages / WHO variants.



> 1000 nt were classified). Other scaffolds were mostly bacterial and human origin, some tree plants sequences aligned (*Pinus strobus*, *Picea abies* and *Picea asperata*). In few samples there were fragments of clinically irrelevant viruses up to 5000 nt (Pepino mosaic virus, *Gallus gallus* retrovirus, Apple mosaic virus, Picobirnavirus). After genome assembly we were able to look into assembled scaffolds and assign them to SARS-CoV-2 lineages. Most of the complete genomes assembled by coronaspades.py have been successfully assigned. The most frequent were variants alpha (clade 20I), then variants delta (21L) and few omicron variants (21J). Assignment of contigs by nextclade is shown on **Figure 2**.

Discussion

This study is a part of a bigger project about human transcriptome and microbiome of COVID-19 patients. Here, on data generated for mentioned project, we tested possibility of assembling SARS-CoV-2 virus from short Illumina sequencing data with properties of metatranscriptomic sequencing (containing mainly human transcripts, but also various

amounts of SARS-CoV-2 and microbial transcripts). We tested assembly strategies using advantages of available reference genome of SARS-CoV-2 or advantages of assembler mode built for the virus (coronaspades.py from SPAdes package), but also assembly ignoring these advantages to test possibilities of assembling unknown RNA virus without previously assembled genome. Results showed, that filtering reads to keep only those mapping to the virus genome helps to get *cleaner* assemblies (with higher quality, with less contigs from different organisms), so does the use of assembler mode specifically designed for SARS-CoV-2 as expected. More interesting results were about testing assembly without using reference genome to separate SARS-CoV-2 reads and with more general modes of SPAdes (`-rna` and `-rnaviral`). Here, only human reads were filtered out. Although results are messier compared to previous, it was still possible to assemble viral scaffolds covering most of the reference genome. Additional contigs were the most notable problem, although statistic NGA50 ignoring unaligned contigs showed real performance might not be so bad. Filtering final contigs from human and

possibly microbial transcripts would be necessary if there wouldn't be reference for assembled virus. We also tested Trinity assembler with SARS-CoV-2 mapped reads. This assembler was not used with any mode or parameters specifically designed for SARS-CoV-2, so it had disadvantage compared to coronaspades.py, but results were still very reasonable and most parameters were comparable to SPAdes, although Quast detected few missassembled scaffolds and usually small parts of the genome were assembled fragmented.

We also validated our assemblies by subjecting them to variant assignment. We validated findings by using pipeline for galaxy environment, designed by Maier and Batut, 2023⁽¹⁵⁾. Here, we identified mix of multiple clades under WHO variant names alpha, delta, omicron and one sample with clade 20C.

From our results, we argue that de-novo assembly of RNA virus from short-read Illumina sequencing is possible, although adding long reads would be probably ideal strategy for getting more complete and reliable assemblies.

REFERENCES

1. Pribelski A, Antipov D, Meleshko D, et al. Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics*. 2020; 70(1): e102.
2. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011; 15; 29(7): 644–652.
3. Simpson JT, Wong K, Jackman SD, et al. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009; 19(6): 1117–1123.
4. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18(5): 821–829.
5. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012; Volume 28, Issue 11, 1420–1428.
6. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012; 28(8): 1086–1092.
7. Hölzer M, Marz M. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience*. 2019; 1; 8(5): giz039.
8. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>; 2010.
9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15): 2114–2120.
10. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: 1303.3997v2 [q-bio.GN]*; 2013.
11. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. *GigaScience*. 2021; Volume 10, Issue 2.

Acknowledgements

This article was created with the support of the OP Integrated Infrastructure for the project: Research on COVID-19 progressive diagnostic methods and biomarkers useful in early detection of individuals at increased risk of severe disease, ITMS: 313011ATA2, co-financed by the European Regional Development Fund.

Podakovanie

Táto štúdia vznikala za podpory projektu: Výskum progresívnych metód diagnostiky COVID-19 a biomarkerov umožňujúcich skorú detekciu jedincov so zvýšeným rizikom ťažkého priebehu ochorenia, ITMS: 313011ATA2, spolufinancovaná Európskym fondom regionálneho rozvoja.

12. "Picard Toolkit." Broad Institute, GitHub Repository. Available online at: <https://broadinstitute.github.io/picard/>; Broad Institute. 2019.
13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics*. 2013; 15; 29(8): 1072–1075.
14. Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N. NCBI BLAST\mathplus integrated into Galaxy. *GigaScience*. 2015; 4(1).
15. Maier W, Batut B. Mutation calling, viral genome reconstruction and lineage/clade assignment from SARS-CoV-2 sequencing data (Galaxy Training Materials). Available online at: <https://training.galaxyproject.org/training-material/topics/variant-analysis/tutorials/sars-cov-2-variant-discovery/tutorial.html>; accessed Sun Jan 29 2023.
16. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*. 2012; 40(22), 11189–11201.
17. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*. 2012; 6(2), 80–92.
18. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*. 2021; 6(67), 3773.
19. O'Toole A, Scher E, Underwood A, et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. *Virus Evolution*. 2021.

Mgr. Dominik Hadžega

MEDIREX GROUP ACADEMY, n. o.

Novozámocká 67, 949 05 Nitra, Slovensko

e-mail: Dominik.Hadzega@medirexgroupacademy.sk

Genetika vybraných zriedkavých atypických syndrómov parkinsonizmu

Natália Forgáčová^{1,2,3}, Juraj Gazdarica^{1,4,5}, Monika Buchalová^{2,6}, Jozef Martiš², Tomáš Szemes^{1,2,4}

¹Prírodovedecká fakulta, Univerzita Komenského v Bratislave

²Vedecký park Univerzity Komenského, Bratislava

³Ústav klinického a translačného výskumu BMC SAV, Bratislava

⁴Geneton, s. r. o., Bratislava

⁵Centrum vedecko-technických informácií SR, Bratislava

⁶Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave

Atypické parkinsonské syndrómy predstavujú širokú škálu rôznorodých a vzájomne sa fenotypovo prekrývajúcich ochorení, čo výrazne komplikuje stanovenie definitívnej diagnózy. Pokrok a rozvoj čoraz dokonalejších vysokokapacitných sekvenčných metód nám však poskytuje veľké množstvo informácií, vďaka ktorým sa dozvieme čoraz viac o génoch, ktoré by mohli súvisieť s parkinsonizmom. V tejto práci uvádzame súčasný prehľad zriedkavých foriem atypického parkinsonizmu, pričom zdôrazňujeme kľúčové genetické pokroky, výzvy a možné budúce smery vedúce k odhaleniu molekulárnej komplexnosti, ktorá je základom týchto ochorení.

Kľúčové slová: parkinsonizmus, atypické parkinsonské syndrómy, progresívna supranukleárna paralýza, multisystémová atrofia, frontotemporálna demencia

Genetics of selected rare atypical parkinsonism-related syndromes

Atypical parkinsonism-related syndromes represent a wide range of diverse and phenotypically overlapping diseases, which significantly complicates determining a definitive diagnosis. However, the progress and development of increasingly high-throughput sequencing methods provide us with a large amount of information because we learn more about genes that could be related to parkinsonism. The current review summarizes rare forms of atypical parkinsonism, highlighting critical genetic advances, challenges, and possible future directions leading to uncovering the molecular complexity underlying these diseases.

Keywords: parkinsonism, atypical parkinsonism-related syndromes, progressive supranuclear palsy, multiple system atrophy, frontotemporal dementia

Newslab, 2023; roč. 14 (1): 33 – 35

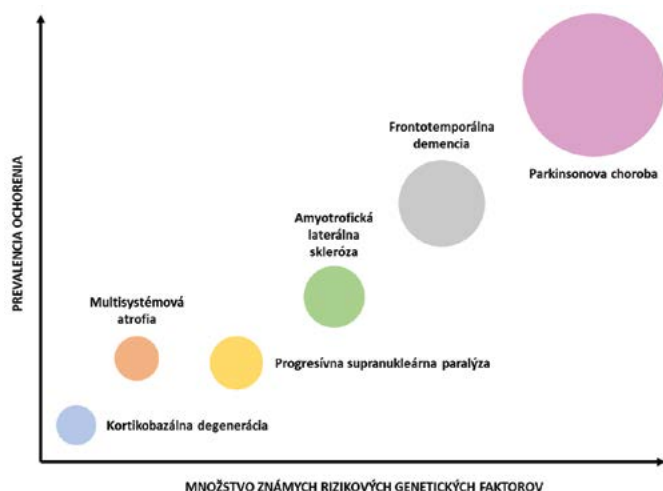
Úvod

Parkinsonizmus je súhrnné označenie pre heterogénnu skupinu ochorení vyznačujúcich sa rôznorodosťou klinických aj patologických charakteristík. Spoločným znakom je však parkinsonský syndróm, zahŕňajúci tetradu motorických príznakov: bradykinézu, rigiditu, pokojový tras a posturálnu nestabilitu. Parkinsonský syndróm je približne v 70 – 80 % asociovaný s Parkinsonovou chorobou, zvyšných 20 – 30 % tvorí tzv. atypický parkinsonizmus. Správna a presná klasifikácia je pri atypickom parkinsonizme zložitá, najčastejšie však do súčasnej klasifikácie patrí multisystémová atrofia (angl. Multiple System Atrophy – MSA), progresívna supranukleárna paralýza (Progressive Supranuclear Palsy – PSP), kortikobazálna degenerácia (Corticobasal Degeneration – CBD), parkinsonské varianty frontotemporálnej demencie (Frontotemporal Dementia – FTD) a iné⁽¹⁾. Za možnú príčinu vzniku sú považované faktory vonkajšieho prostredia aj genetické predispozície alebo kombinácia týchto faktorov. V súvislosti s parkinsonizmom bol opísaný rad génov, či už kauzálnych, alebo tých, ktoré predstavujú možné riziko⁽²⁾.

Progresívna supranukleárna paralýza (PSP)

PSP je zriedkavé neurodegeneratívne ochorenie, ktoré je z patologického hľadiska charakterizované ako 4R – tauopatia s inklúziami tvorenými akumuláciou patologického, hyperfosforylovaného tauproteínu⁽³⁾. Väčšina prípadov PSP sa vyskytuje sporadicky, avšak boli opísané aj familiárne formy s autozomálne dominantnou dedičnosťou. Za kauzálny je najčastejšie považovaný gén *MAPT*⁽⁴⁾, ktorý je lokalizovaný na dlhom ramienku chromozómu 17 v pozícii 21.31. Kóduje proteín tau, ktorý je primárne exprimovaný v mozgu. Dôležitou funkciou tohto proteínu je stabilizácia cytoskeletu a axonálneho transportu v neurónoch. Mutácie *MAPT*, ktoré vedú k neurodegenerácii na základe zmenenej stabilizácie mikrotubulov a zvýšenej náchylnosti na agregáciu proteínov tau, boli identifikované v exóne 10, ale aj v exónoch 1, 12 a 13⁽⁵⁾. V súvislosti s familiárnym výskytom PSP bola opísaná mutácia G2019S v *LRRK2* géne⁽⁶⁾. Okrem tejto mutácie boli s PSP asociované ďalšie štyri mutácie *LRRK2*, v exóne 30 (A1413T)⁽⁷⁾, dve mutácie v exóne 31 (R1441C a R1441H)⁽⁸⁾ a v exóne 41 (T2310M)⁽⁷⁾. Medzi rizikové gény spojené s významným rizikom rozvoja PSP patrí taktiež *DCTN1*, *PARK2*,

Obrázok 1. Vzťah prevalence parkinsonských syndrómov a množstva známych rizikových genetických faktorov. So zväčšujúcim priemerom kruhu sa zvyšuje množstvo známych informácií o rizikových génoch, ktoré sú asociované s danými ochoreniami.



BSN, GRN, C9ORF72, TARDBP, NPC, PGRN a *TBK1*⁽⁵⁾. Nedávna GWAS štúdia opísala gén *TRIM11* ako genetický modifikátor fenotypu PSP. *TRIM11* zohráva dôležitú úlohu pri odstraňovaní nesprávne poskladaných proteínov prostredníctvom ubikvitínovo-proteázového systému (UPS). V budúcnosti sa predpokladá, že práve spojitosť medzi UPS a patológiou proteínu tau môže predstavovať potenciálny cieľ terapií PSP⁽⁹⁾.

Multisystémová atrofia (MSA)

MSA je zriedkavé, rýchlo progredujúce ochorenie, pričom základným morfológickým znakom je prítomnosť typických oligodendrogliaálnych cytoplazmatických inklúzií, ktoré pozostávajú z nesprávne poskladaného, fibrilovaného a agregovaného alfa-synukleínu⁽¹⁰⁾. Keďže MSA je považovaná za alfa-synukleinopatiu, značná pozornosť výskumu rizikových faktorov je venovaná génu *SNCA*, ktorý kóduje proteín alfa-synukleín. Hoci sa zistilo, že gén *SNCA* je jedným z kauzálnych génov pre rozvoj Parkinsonovej choroby, kauzalita s MSA je stále nejasná. Prostredníctvom GWAS bolo opísaných niekoľko jednonukleotidových variantov v *SNCA* géne, ktoré nie sú síce kauzálnymi variantmi, ale sú významne asociované so zvýšeným rizikom rozvoja MSA. Za rizikové boli označené varianty rs11931074, rs3857059, rs3822086 a rs3775444^(11,12). V populáciách východnej Ázie bol identifikovaný variant V393A v géne *COQ2* a v európskej populácii boli zistené najmenej štyri polymorfizmy (*EDN1* rs16872704, *MAPT* rs9303521, *FBXO47* rs78523330, a *ELOVL7* rs7715147) spojené s rizikom rozvoja MSA⁽¹³⁾. Uvažuje sa aj o tom, že varianty počtu kópií DNA (copy number variants – CNV) v nekódujúcich oblastiach môžu prostred-

níctvom modulácie transkripčnej aktivity zvyšovať riziko vzniku MSA⁽¹⁴⁾.

Frontotemporálna demencia (FTD)

FTD zahŕňa skupinu heterogénnych neurodegeneratívnych ochorení s odlišnými klinickými fenotypmi, charakterizovaných progredujúcou atrofiou frontálnych alebo temporálnych lalokov mozgu. Najvýznamnejším rizikovým faktorom FTD je pozitívna rodinná anamnéza. Tá bola zistená až u 30 – 50 % pacientov, z ktorých až v 40 % prípadoch išlo o autozomálne dominantný typ dedičnosti. Väčšina geneticky podmienených FTD je spôsobená mutáciami v troch génoch: *C9orf72*, *GRN* a *MAPT*. Frekvencia týchto mutácií je geograficky variabilná, napríklad v Spojenom kráľovstve existuje približne rovnaký počet mutácií vo všetkých troch génoch. Na druhej strane v Taliansku a Španielsku je najbežnejšia mutácia *GRN* génu, v severnej Európe *C9orf72*⁽¹⁵⁾. Najčastejšou príčinou dedičnej formy FTD je expanzia nekódujúcej hexanukleotidovej repetície GGGGCC v géne *C9orf72*. *C9orf72* je asociovaný nielen s fenotypom FTD, ale aj s fenotypom amyotrofickej laterálnej sklerózy (ALS) a CBD⁽¹⁶⁾. Metaanalýza vykonaná v roku 2017 preukázala genetickú súvislosť medzi PSP, FTD a kortikobazálnou degeneráciou (CBD) pričom identifikovala množstvo nových asociácií vrátane rizikových variantov v génoch *MOBP*, *CXCR4*, *EGFR* a *GLDC*⁽¹⁷⁾. Je však potrebné poznamenať, že budúce štúdie s väčším počtom pacientov sú potrebné na ďalšie overenie týchto asociácií.

Záver

Problematika atypických parkinsonských syndrómov je veľmi rozsiahla a zložitá v dôsledku ich nízkej individuálnej frekvencie výskytu a fenotypovej rôznorodosti. V uvedenej práci sme sa zamerali predovšetkým na MSA, PSP a FTD ako na tri základné jednotky atypického parkinsonského syndrómu, pričom sme poukázali na širokú genotypovú rozmanitosť týchto ochorení. Genómové asociačné štúdie a testovanie génových panelov majú veľký potenciál vo výskume zriedkavých ochorení, a to aj v prípade neurodegeneratívnych zriedkavých ochorení. Rozvoj vedomostí, najmä v oblasti patofyziológie a genetiky parkinsonských syndrómov je kľúčom k pochopeniu genetického základu týchto ochorení a azda aj základným pilierom na vypracovanie efektívnejších terapeutických postupov.

Podakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekty ITMS:313011ATL7 (PanClinCov) a ITMS: 313011W988 (SANET II), spolufinancované zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

- Levin J, Kurz A, Arzberger T, et al. The Differential Diagnosis and Treatment of Atypical Parkinsonism. *Dtsch Arztebl Int.* 2016; 113: 61–69.
- Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol.* 2020; 19: 170–178.
- Coughlin DG, Litvan I. Progressive supranuclear palsy: Advances in diagnosis and management. *Parkinsonism Relat Disord.* 2020; 73: 105–116.
- Chen JA, Chen Z, Won H, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and

genetic correlation to neurodegenerative diseases. *Mol Neurodegener.* 2018; 13: 41.

5. Diez-Fairen M, Alvarez Jerez P, Berghausen J, et al. The Genetic Landscape of Parkinsonism-Related Dystonias and Atypical Parkinsonism-Related Syndromes. *Int J Mol Sci.* 2021; 22.

6. Vilas D, Sharp M, Gelpi E, et al. Clinical and neuropathological features of progressive supranuclear palsy in Leucine rich repeat kinase (LRRK2) G2019S mutation carriers. *Mov Disord.* 2018; 33: 335–338.

7. Sanchez-Contreras M, Heckman MG, Tacik P, et al. Study of LRRK2 variation in tauopathy: Progressive supranuclear palsy and corticobasal degeneration. *Mov Disord.* 2017; 32: 115–123.

8. Spanaki C, Latsoudis H, Plaitakis A. LRRK2 mutations on Crete: R1441H associated with PD evolving to PSP. *Neurology.* 2006; 67: 1518–1519.

9. Jabbari E, Woodside J, Tan MMX, et al. Variation at the TRIM11 locus modifies progressive supranuclear palsy phenotype. *Ann Neurol.* 2018; 84: 485–496.

10. Poewe W, Stankovic I, Halliday G, et al. Multiple system atrophy. *Nat Rev Dis Primers.* 2022; 8: 56.

11. Al-Chalabi A, Dürr A, Wood NW, et al. Genetic variants of the alpha-synuclein gene SNCA are associated with multiple system atrophy. *PLoS One.* 2009; 4: e7114.

12. Chen Y, Wei Q-Q, Ou R, et al. Genetic Variants of SNCA Are Associated with Susceptibility to Parkinson's Disease but Not Amyotrophic Lateral Sclerosis or Multiple System Atrophy in a Chinese Population. *PLOS ONE.* 2015. p. e0133776.

13. Gu X, Chen Y, Zhou Q, et al. Analysis of GWAS-linked variants in multiple system atrophy. *Parkinsonism & Related Disorders.* 2018. p. e30.

14. Hama Y, Katsu M, Takigawa I, et al. Genomic copy number variation analysis in multiple system atrophy. *Molecular Brain.* 2017.

15. Bang J, Spina S, Miller BL. Frontotemporal dementia. *Lancet.* 2015; 386: 1672–1682.

16. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GG-GGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011; 72: 245–256.

17. Yokoyama JS, Karch CM, Fan CC, et al. Shared genetic risk between corticobasal degeneration, progressive supranuclear palsy, and frontotemporal dementia. *Acta Neuropathol.* 2017; 133: 825–837.

Mgr. Natália Forgáčová

Vedecský park Univerzity Komenského v Bratislave

Ilkovičova 8, 841 04 Bratislava

e-mail: natalia.forgacova@geneton.sk

Neinvazívny skríning onkologických ochorení pomocou biomarkerov vo voľne cirkulujúcej DNA

Zuzana Holešová¹, Juraj Gazdarica^{1,2,4}, Jaroslav Budiš^{1,2,4}, Ondrej Pös^{1,2}, Tomáš Szemes^{1,2,3}

¹Geneton, s. r. o., Bratislava

²Vedecký park Univerzity Komenského, Bratislava

³Prírodovedecká fakulta, Univerzita Komenského v Bratislave

⁴Centrum vedecko-technických informácií, Bratislava

V súčasnosti sú účinné metódy skríningu dostupné len pre niektoré nádorové ochorenia a vo všeobecnosti sú relatívne komplikované. Preto je potrebné vyvinúť jednoduché a presné metódy na včasnú detekciu čo najväčšieho spektra nádorových ochorení. Takýto potenciál má detekcia genetických a epigenetických zmien v DNA voľne cirkulujúcej v plazme (cfDNA), vďaka ktorej by bolo možné veľmi skoro detegovať nádorové ochorenie a tak zlepšiť prežívanie pacientov. Molekulárne vlastnosti fragmentov cfDNA a ich distribúcia v genóme poskytujú informácie o tkanivách ich pôvodu. Organizácia nukleozómov a obsah nukleáz v pôvodnom tkanive ovplyvňujú veľkosť a motív zakončenia fragmentov cfDNA. Aj vďaka týmto vlastnostiam je tekutá biopsia jeden z najslubnejších nástrojov pre budúce klinické využitie.

Kľúčové slová: tekutá biopsia, cfDNA, ctDNA, biomarker, rakovina

Non-invasive screening of oncological diseases using biomarkers in freely circulating DNA

Effective screening methods are only available for some cancers and are generally relatively complicated. Therefore, it is necessary to develop simple and accurate methods for early detection of the most extensive possible spectrum of cancer. The detection of genetic and epigenetic changes in DNA freely circulating in plasma (cfDNA) has such potential, thanks to which it would be possible to detect cancer very early and thus improve the survival of patients. The molecular properties of cfDNA fragments and their distribution in the genome provide information about their tissue of origin. The organization of nucleosomes and the content of nucleases in the original tissue affect the size and termination motif of cfDNA fragments. These properties make liquid biopsy one of the most promising tools for future clinical use.

Keywords: liquid biopsy, cfDNA, ctDNA, biomarker, cancer

NewsLab, 2023; roč. 14 (1): 36 – 39

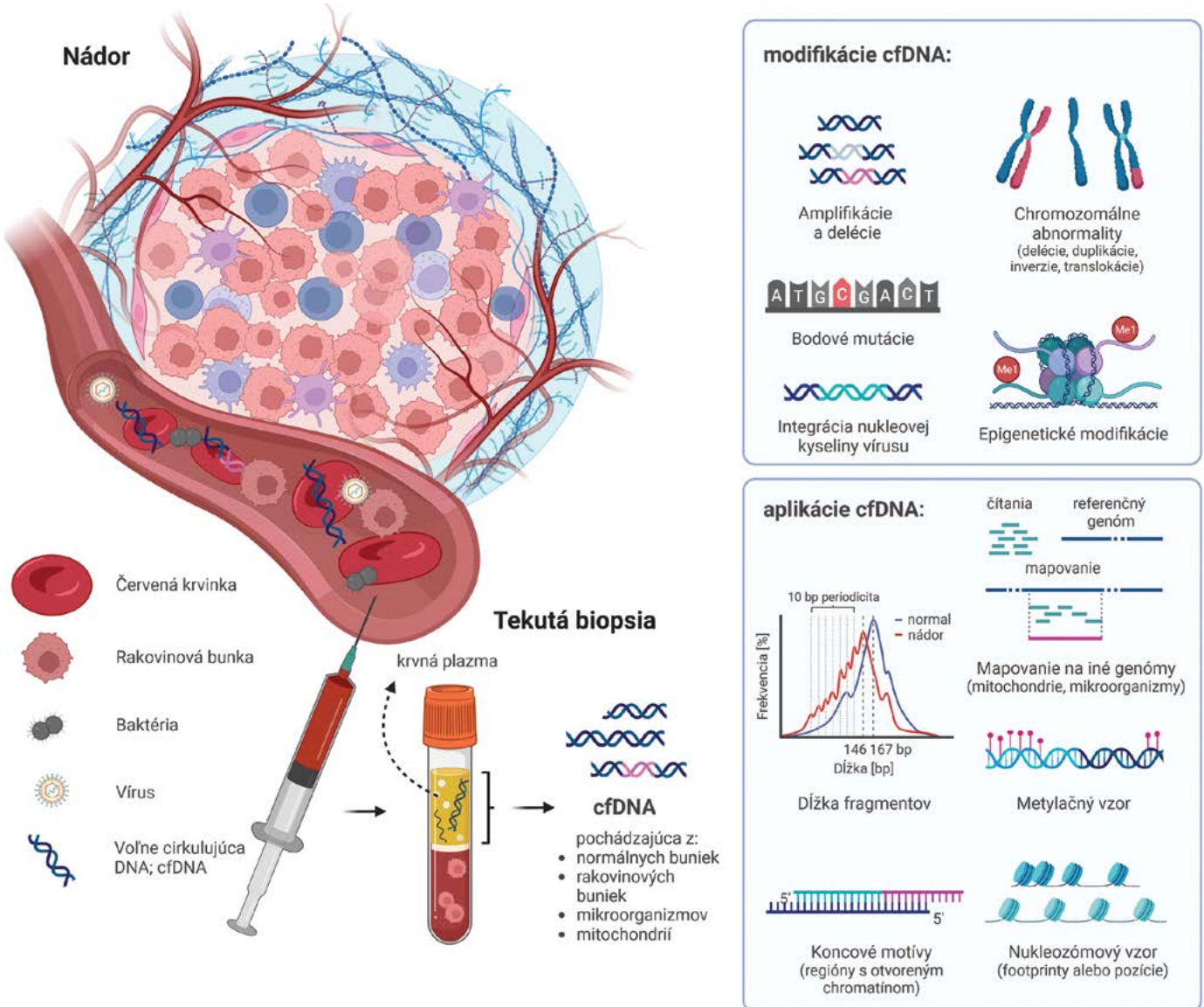
Úvod

Rakovina má významný vplyv na zdravie ľudí na celom svete. Podľa Svetovej zdravotníckej organizácie (WHO) je rakovina hlavnou príčinou úmrtí v mnohých štátoch a tento počet sa bude v dohľadnom čase zväčšovať⁽¹⁾. V roku 2020 zomrelo na rakovinu viac ako 10 miliónov ľudí najmä v dôsledku nedostatočného prístupu k vhodným diagnostickým metódam⁽²⁾. Príčinou vzniku nádorov je zmena v génoch riadiacich rast a delenie buniek, to spôsobí, že sa niektoré bunky začnú deliť bez zastavenia a šíria sa do okolitých tkanív. Rakovina tak môže vzniknúť takmer v akejkoľvek z miliárd buniek v ľudskom tele. Pri každom type nádoru dochádza k jedinečnej kombinácii genetických zmien, navyše, ako rastie, naďalej sa mení. Dokonca v rámci toho istého nádoru môžu bunky disponovať rôznymi genetickými zmenami. To je jeden z dôvodov, prečo je liečba rakoviny náročná. Preto je veľmi dôležité vyvinúť jednoduché, presné a čo najmenej invazívne metódy na jej včasné odhalenie, keďže štádium, v ktorom je diagnostikovaná, je jedným z najvýznamnejších prediktorov prežitia.

Jednou z takých metód je tzv. tekutá biopsia (LB). Jej prvé aplikácie boli založené na detekcii genetických markerov vo voľne cirkulujúcej DNA (cfDNA), ako sú napr. genetické polymorfizmy alebo mutácie. Štúdiom negenetických vlastností cfDNA – metylácie, fragmentácie a topológie – sa jej využitie rozšírilo, cfDNA si možno predstaviť ako genetický rezervoár, ktorý obsahuje genetickú informáciu zo všetkých buniek v tele⁽³⁾ vrátane zdravých, chorých i mikroorganizmov, ktoré sa v tele nachádzajú⁽⁴⁾. Ich pomer sa vzhľadom na aktuálny stav jedinca vrátane rakoviny môže meniť. Každý fragment cfDNA nesie molekulárne znaky bunky, z ktorej pochádza.

Prostredníctvom LB možno identifikovať rakovinové zmeny v cfDNA⁽⁵⁾, ktorá sa uvoľní z buniek po ich smrti⁽⁶⁾. Keďže väčšina nádorov je v kontakte s krvou, väčšinou sa pri LB odoberá vzorka krvi, je však možné analyzovať aj iné telesné tekutiny, napr. pleurálny výpotok, moč a mozgovomiechový mok⁽⁷⁾. Vďaka jednoduchosti odberu predstavuje analýza cfDNA minimálne invazívnu a nízkonákladovú alternatívu biopsie nádoru⁽⁸⁾. Okrem včasnej detekcie nádorov⁽⁹⁾

Obrázok 1. Schematické znázornenie tekutej biopsie. V periférnej krvi sa okrem krvných komponentov nachádza voľná cirkulujúca DNA (cfDNA), mikroorganizmy (baktérie, vírusy, huby), v prípade onkologického pacienta i cirkulujúce nádorové bunky a ďalšie komponenty (napr. RNA a vezikuly). Po oddelení krvnej plazmy centrifugáciou z nej možno izolovať cfDNA a následne ju analyzovať rôznymi prístupmi. Je možné sa zamerať na (epi)genetické zmeny a molekulárne vlastnosti fragmentov cfDNA, organizáciu nukleozómov, obsah nukleáz v súvislosti so zakončením fragmentov cfDNA a cfDNA pochádzajúcu z mitochondrií a mikrobiómu.



je možné pomocou LB rýchlejšie odhaliť recidívu nádorového ochorenia a na rozdiel od klasickej biopsie má vyššiu šancu zachytiť viaceré mutácie. Molekulárnou charakterizáciou cfDNA je zároveň možné získať informácie o pôvode a progresii nádoru, čo má veľký význam pre prevenciu a liečbu nádorového ochorenia.

Modifikácie a aplikácie cfDNA

Rozvoj vysokovýkonných sekvenčných technológií (NGS) umožňuje paralelné sekvenovanie miliónov fragmentov DNA rýchlym a ekonomickým spôsobom. Množstvo vygenerovaných údajov si vyžaduje okrem počítačového vybavenia bioinformatickú analýzu, ktorá vedie od spracovania surových údajov cez ich podrobnejšiu analýzu až k interpretácii variantov v klinickom kontexte⁽¹⁰⁾. Vďaka NGS možno odhaliť známe i nové mutácie, fúzie, abnormálny počet kópií génov

a variantov, zmeny v metylácii DNA i mutačnú záťaž alebo mikrosatelitovú nestabilitu (**obrázok 1**)⁽¹¹⁾. Veľkosť molekúl cfDNA, umiestnenie koncov fragmentov a koncové motívy sú ovplyvnené organizáciou nukleozómov, štruktúrou chromatinu, obsahom nukleáz a expresiou génov v tkanive, z ktorého pochádza.

Cirkulujúca DNA pochádzajúca z nádoru (ctDNA) tvorí u väčšiny onkologických pacientov nízky podiel k celkovej cfDNA, často < 1 %⁽¹²⁾, najmä v prípade raného štádia. Disponuje však rovnakými genetickými i epigenetickými zmenami, aké sú v samotnom nádore. Od cfDNA pochádzajúcej z normálnych buniek ju možno odlíšiť podľa prítomnosti bodových mutácií, aneuploidie, zmien počtu kópií génov, odlišnej metylácie alebo prítomnosti vírusových sekvencií DNA^(13,14). Vzhľadom na jej nízku koncentráciu je potrebné jej spoľahlivé odlíšenie od cfDNA zdravých buniek.

Metylácia DNA je nevyhnutná pre normálny vývoj a zohráva dôležitú úlohu v epigenetickej kontrole aktivity génov. Zmeny v metylácii DNA sa považujú za najčastejšie molekulárne zmeny v procese onkogenézy⁽¹⁵⁾. Je známe, že každé tkanivo má svoj charakteristický vzor metylácie, ktorý sa odlišuje medzi nádorovými a normálnymi bunkami⁽¹⁶⁾. Preto profilovanie metylácie celého genómu z cfDNA môže byť potenciálne účinným nástrojom na zisťovanie prítomnosti konkrétneho nádorového ochorenia.

Okrem zmien v metylácii je fragmentácia cfDNA a vzorce obsadenia nukleozómami ďalšími epigenetickými znakmi na sledovanie aktivity génov a pôvodu tkaniva. Počas bunkovej smrti dochádza k enzymatickému štiepeniu DNA nukleázami. DNA je počas toho chránená nukleozómami a tak majú fragmenty cfDNA charakteristickú veľkosť 167 bp, čo zodpovedá dĺžke otočenia okolo jedného nukleozómu⁽¹⁷⁾. Veľkosť fragmentov sa však môže u zdravých jedincov a onkologických pacientov líšiť. Fragmenty cfDNA pochádzajúce z nádorových buniek sú v porovnaní s fragmentmi zo zdravých buniek kratšie⁽¹⁸⁾. Sekvenovaním s rozlíšením na jeden bázy pár sa zistilo, že kratšie fragmenty cfDNA v plazme pacientov s hepatocelulárnym karcinómom majú viac zmien asociovaných s nádormi ako dlhšie fragmenty cfDNA⁽¹⁹⁾, hoci niektoré konkrétne zriedkavé mutácie sú v dlhších fragmentoch cfDNA⁽²⁰⁾. Zhutnenie nukleozómových štruktúr vytvára prekážku pre prístup transkripčných faktorov k regulačným elementom. Umiestnenie nukleozómov súvisí s aktiváciou a expresiou génov v závislosti od vývinu a konkrétneho tkaniva⁽²¹⁾. Preto môže skúmanie polohy nukleozómov v cfDNA a identifikácia pôvodného tkaniva odhaliť existenciu špecifického typu nádoru⁽²²⁾.

Okrem veľkosti fragmentov cfDNA je zaujímavé i zloženie koncových motívov. Fragmentácia molekúl cfDNA nevzniká náhodným procesom. Preferované konce fragmentov sa vzťahujú na špecifické koncové miesta v genóme, zvyčajne v oblastiach s otvoreným chromatinom (ktoré sú prístupnejšie)⁽²³⁾. K fragmentácii cfDNA dochádza aktivitou deoxyribonukleáz (DNáz), z ktorých každá rozpoznáva inú sekvenciu⁽²⁴⁾. Pri viacerých onkologických ochoreniach bol pozorovaný rôzny pomer koncových motívov. Napr. v prípade hepatocelulárneho karcinómu bol koncový motív CCCA výrazne menej zastúpený ako pri kontrole, pravdepodobne pre zníženú expresiu DN-ázy1L3. Navyše profil koncových motívov cfDNA pochádzajúcej z toho istého orgánu je vo všeobecnosti rovnaký⁽²⁵⁾. Dôvodom týchto odlišností môžu byť genetické alebo epigenetické zmeny, ktoré mohli spôsobiť poruchy v expresii endonukleáz DNA⁽²⁵⁾. Jiang a kol. skúmali prítomnosť prečnievajúcich koncov po štiepení nukleázami a zistili, že boli zastúpené vo väčšine molekúl cfDNA (~ 90 %)⁽²⁶⁾. V predchádzajúcich štúdiách prečnievajúce konce neboli odhalené, pretože sú v procese prípravy sekvenčných knižníc opravené na tupé. Prečnievanie sa líšilo podľa veľkosti fragmentov cfDNA a zdá sa, že súviselo s nukleozómovými vzormi. V prípade pacientov s hepatocelulárnym karcinómom vykazovala cfDNA vyššie zastúpenie prečnievajúcich koncov ako nenádorová cfDNA⁽²⁶⁾.

Mitochondrie svojím genómom takisto prispievajú k celkovej cfDNA. Mitochondriálna DNA (mtDNA) sa na rozdiel od jadrovej DNA nachádza v bunke v tisíckach kópií⁽²⁷⁾. Okrem toho má vysokú mutačnú rýchlosť a pri nádorových ochore-

niach dochádza u nej k zásadným modifikáciám^(28,29). Obe tieto vlastnosti ju predurčujú stať sa vhodným kandidátom na tekutú biopsiu. To potvrdzujú i niektoré štúdie, ktoré preukázali rozdiel v obsahu i fragmentácii voľne cirkulujúcej mtDNA (mt-cfDNA) pri onkologických ochoreniach^(30,31). Veľkosť fragmentov mt-cfDNA bola menšia v porovnaní s jadrovou cfDNA, dôvodom môže byť to, že na rozdiel od jadrovej cfDNA nie je chránená nukleozómami, preto pri nej nie je pozorovaný ani charakteristický veľkostný pík ako pri jadrovej cfDNA (167 bp)⁽¹⁹⁾. Čo sa týka množstva kópií, zrejme nejde o jednoznačný marker, keďže v prípade hepatocelulárneho karcinómu bol pozorovaný v plazme pacientov nižší⁽³¹⁾ i vyšší počet kópií mt-cfDNA⁽¹⁹⁾. Pokiaľ ide o dĺžku mt-cfDNA, bolo pozorované, že nepriamo súvisí s veľkosťou nádoru i s koncentráciou cirkulujúcej nádorovej DNA^(19,30).

Na 10-15 % prípadov vzniku a progresie rakoviny sa podieľajú vírusy a baktérie⁽³²⁾. V kontexte onkológie je však cirkulujúca mikrobiálna DNA málo preskúmaná, uplatňovala sa najmä pri infekčných chorobách a sepse, kde jej sekvenovanie zlepšilo detekciu ťažko kultivovateľných mikroorganizmov⁽³³⁾. Súvislosť medzi zmenami v črevnom mikrobióme, tzv. dysbiózou a mnohými neinfekčnými chorobami je relatívne dobre známa, rakovina nie je výnimkou^(34,35). Charakteristický „podpis“ cirkulujúcej bakteriálnej DNA (cbDNA) pre daný typ rakoviny bol rovnaký dokonca aj v I. a II. štádiu rakoviny a tiež pri rakovine, v ktorej nedošlo ku genomickým zmenám. Je to významné zistenie, keďže súčasná analýza ctDNA sa spolieha práve na zisťovanie genomických zmien, ktoré sú prítomné v primárnom nádore. Naznačuje to, že cbDNA by mohla byť citlivejším a širšie použiteľným biomarkerom rakoviny.

Medzi najčastejšie onkogénne vírusy patrí vírus hepatitídy B (HBV) a ľudský papilomavírus (HPV), ktoré sú príčinou približne polovice nádorov vyvolaných vírusovou infekciou⁽³⁶⁾. Ku karcinogéze prispievajú produkciou onkogénnych proteínov, moduláciou imunity alebo prostredníctvom integrácie vírusovej DNA do DNA hostiteľa (vh-DNA)⁽³⁷⁾. Práve integrované úseky vírusovej DNA môžu byť detegované v krvnom obehú ešte pred vznikom samotného nádoru. Nie je však možné pomocou nich identifikovať štádium nádoru alebo odlíšiť vírusovú DNA integrovanú do normálnych tkanív od nádorových tkanív. Stanovenie hraničnej hodnoty na odlíšenie nádorovo špecifickej vh-DNA od vh-DNA uvoľnenej z nenádorového tkaniva je rozhodujúce na dosiahnutie klinicky významnej citlivosti a špecifickosti⁽³⁷⁾.

Záver

Tekutá biopsia v dohľadnej budúcnosti nenahradí celkom biopsiu tkaniva, môže ju však doplniť. Vzhľadom na to, že väčšina súčasných techník tekutej biopsie nemá dostatočnú detekčnú schopnosť, je potrebné jej definíciu rozšíriť i o nenádorové informácie. Predpokladá sa, že po štandardizácii metodických postupov sa bude viac využívať v rutinných podmienkach a pomôže zachytiť nádor ešte pred vznikom symptómov. Pred jej úplným zavedením je však potrebné zodpovedať niektoré otázky. Napríklad, do akej miery sa presnosť testov líši medzi jednotlivými typmi nádorov a štádiami ochorenia. Poskytuje tekutá biopsia reprezentatívnu vzorku všetkých genetických klonov v nádore alebo dochádza ku skresleniu na špecifické subregióny? Na to je však potrebných ešte veľa klinických štúdií.

Poďakovanie

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekty ITMS: 313011V578 (PreveLynch) a ITMS: 313021BUZ3 (USCCCORD) spolufinancované zo zdrojov Európskeho fondu regionálneho rozvoja. Práca bola tiež spolufinancovaná Agentúrou na podporu výskumu a vývoja v rámci grantov APVV-21-0296 (INCAM) a APVV-18-0319 (GenoMicrosat).

LITERATÚRA

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin*. 2021; 71(1): 7-33.
- Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: An overview. *Int J Cancer*. Published online April 5, 2021. doi: 10.1002/ijc.33588
- Qin Z, Ljubimov VA, Zhou C, et al. Cell-free circulating tumor DNA in cancer. *Chin J Cancer*. 2016; 35: 36.
- Kowarsky M, Camunas-Soler J, Kertesz M, et al. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. *Proc Natl Acad Sci U S A*. 2017; 114(36): 9623-9628.
- Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017; 17(4): 223-238.
- Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. *Eur J Hum Genet*. 2018; 26(7): 937-945.
- Pös Z, Pös O, Styk J, et al. Technical and Methodological Aspects of Cell-Free Nucleic Acids Analyzes. *Int J Mol Sci*. 2020; 21(22). doi: 10.3390/ijms21228634
- Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*. 2013; 10(8): 472-484.
- Zhu JW, Charkhchi P, Akbari MR. Potential clinical utility of liquid biopsies in ovarian cancer. *Mol Cancer*. 2022; 21(1): 114.
- Pereira R, Oliveira J, Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med Res*. 2020; 9(1). doi: 10.3390/jcm9010132
- Saghafinia S, Mina M, Riggi N, et al. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep*. 2018; 25(4): 1066-1080.e8.
- Pantel K, Alix-Panabières C. Liquid biopsy and minimal residual disease – latest advances and implications for cure. *Nat Rev Clin Oncol*. 2019; 16(7): 409-424.
- Chan KCA, Jiang P, Chan CWM, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci U S A*. 2013; 110(47): 18761-18768.
- McBride DJ, Orpana AK, Sotiriou C, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer*. 2010; 49(11): 1062-1069.
- Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet*. 2000; 16(4): 168-174.
- Chen Y, Breeze CE, Zhen S, et al. Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. *Epigenetics Chromatin*. 2016; 9: 10.
- Heitzer E, Auinger L, Speicher MR. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. *Trends Mol Med*. 2020; 26(5): 519-528.
- Mouliere F, Chandrananda D, Piskorz AM, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*. 2018; 10(466). doi: 10.1126/scitranslmed.aat4921
- Jiang P, Chan CWM, Chan KCA, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A*. 2015; 112(11): E1317-E1325.
- Liu X, Lang J, Li S, et al. Fragment Enrichment of Circulating Tumor DNA With Low-Frequency Mutations. *Front Genet*. 2020; 11: 147.
- Ulz P, Perakis S, Zhou Q, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nat Commun*. 2019; 10(1): 4666.
- Snyder MW, Kircher M, Hill AJ, et al. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016; 164(1-2): 57-68.
- Oberhofer A, Bronkhorst AJ, Uhlig C, et al. Tracing the Origin of Cell-Free DNA Molecules through Tissue-Specific Epigenetic Signatures. *Diagnostics (Basel)*. 2022; 12(8). doi: 10.3390/diagnostics12081834
- Han DSC, Ni M, Chan RWY, et al. The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. *Am J Hum Genet*. 2020; 106(2): 202-214.
- Jiang P, Sun K, Peng W, et al. Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov*. 2020; 10(5): 664-673.
- Jiang P, Xie T, Ding SC, et al. Detection and characterization of jagged ends of double-stranded DNA in plasma. *Genome Res*. 2020; 30(8): 1144-1153.
- Kelly RDW, Mahmud A, McKenzie M, Trounce IA, St John JC. Mitochondrial DNA copy number is regulated in a tissue specific manner by DNA methylation of the nuclear-encoded DNA polymerase gamma A. *Nucleic Acids Res*. 2012; 40(20): 10124-10138.
- Hertweck KL, Dasgupta S. The Landscape of mtDNA Modifications in Cancer: A Tale of Two Cities. *Front Oncol*. 2017; 7: 262.
- van Osch FHM, Voets AM, Schouten LJ, et al. Mitochondrial DNA copy number in colorectal cancer: between tissue comparisons, clinicopathological characteristics and survival. *Carcinogenesis*. 2015; 36(12): 1502-1510.
- An Q, Hu Y, Li Q, et al. The size of cell-free mitochondrial DNA in blood is inversely correlated with tumor burden in cancer patients. *Precis Clin Med*. 2019; 2(3): 131-139.
- Haupts A, Vogel A, Foersch S, et al. Comparative analysis of nuclear and mitochondrial DNA from tissue and liquid biopsies of colorectal cancer patients. *Sci Rep*. 2021; 11(1): 16745.
- Lunn RM, Jahnke GD, Rabkin CS. Tumour virus epidemiology. *Philos Trans R Soc Lond B Biol Sci*. 2017; 372(1732). doi: 10.1098/rstb.2016.0266
- Blauwkamp TA, Thair S, Rosen MJ, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol*. 2019; 4(4): 663-674.
- Vivarelli S, Salemi R, Candido S, et al. Gut Microbiota and Cancer: From Pathogenesis to Therapy. *Cancers*. 2019; 11(1). doi: 10.3390/cancers11010038
- Huang YF, Chen YJ, Fan TC, et al. Analysis of microbial sequences in plasma cell-free DNA for early-onset breast cancer patients and healthy females. *BMC Med Genomics*. 2018; 11(Suppl 1): 16.
- Plummer M, de Martel C, Vignat J, et al. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Health*. 2016; 4(9): e609-e616.
- Li CL, Yeh SH, Chen PJ. Circulating Virus-Host Chimera DNAs in the Clinical Monitoring of Virus-Related Cancers. *Cancers*. 2022; 14(10). doi: 10.3390/cancers14102531

Mgr. Zuzana Holešová, PhD.

Geneton, s. r. o.

Ilkovičova 8, 841 04 Bratislava

e-mail: zuzana.holesova@geneton.sk

Tandem repeat motif characterization for precision medicine: A brief overview of conventional methods and massive parallel sequencing approaches

Ingrid Lojova^{1,2,3}, Jaroslav Budis^{2,3,4,5}, Tomas Szemes^{2,3,4}, Monika Buchalova^{2,6}, Jan Radvanszky^{1,2,3}

¹Institute of Clinical and Translational Research, Biomedical Research Center of the Slovak Academy of Sciences, Bratislava, Slovakia

²Comenius University Science Park, Bratislava, Slovakia

³Department of Molecular Biology, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia

⁴Geneton Ltd., Bratislava, Slovakia

⁵Slovak Center of Scientific and Technical Information, Bratislava, Slovakia

⁶Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia

Repetitive motifs, also known as tandem repeats, are stretches of DNA that are repeated multiple times in a row within a genome. These motifs have important functional and clinical implications in human genomes, including their involvement in genetic disorders such as Huntington's disease, Fragile X syndrome or myotonic dystrophies. Reliable and accurate characterization of these motifs, especially on genome scales, has remained a challenge due to their structural complexity and variability. In this review, we explore the possibilities of various molecular methods for the reliable characterization of clinically-relevant repetitive motifs, including molecular biology techniques, high-throughput sequencing technologies and bioinformatics tools.

Keywords: tandem repeats, conventional TR genotyping, massively parallel sequencing

Charakterizácia tandemových repetitívnych motívov v precíznej medicíne: Stručný prehľad konvenčných metód a masívne paralelného sekvenovania

Repetitívne motívy, známe aj ako tandemové opakovania, sú úseky DNA, ktoré sa v genóme opakujú viackrát za sebou. Tieto motívy majú v ľudských genómoch dôležitý funkčný a klinický význam vrátane ich účasti na genetických poruchách, ako je napríklad Huntingtonova choroba, syndróm fragilného X alebo myotonické dystrofie. Spoľahlivá a presná charakterizácia týchto motívov, najmä na úrovni genómu, predstavuje vzhľadom na ich štruktúrnu zložitosť a variabilitu veľkú výzvu. V tomto prehľade sumarizujeme možnosti rôznych molekulárnych metód na spoľahlivú charakterizáciu klinicky relevantných repetitívnych motívov vrátane klasických metód molekulárnej biológie, vysokoparalelných sekvenčných technológií a bioinformatických nástrojov.

Kľúčové slová: tandemové opakovania, konvenčná TR genotypizácia, masívne paralelné sekvenovanie

NewsLab, 2023; roč. 14 (1): 40 – 44

Introduction

Tandem repeats (TRs) are stretches of DNA sequences, which contain repeated units several times. The sequence composition and length of the repeating units may be various in different TRs, and they can be found in various locations in the human genome, including in intergenic regions, introns, and exons of genes. Typically the number of repetitions is highly variable between individuals. Since they represent extremely variable genomic regions, TRs have many important biological roles in several physiological and pathophysiological processes. Their proper characterization may, therefore, explain observed phenotypes of traits in biomedical context, and thus markedly improve informativeness of clinical genomics assays. They also be-

came utilized in several non-biomedical applications, mainly in population genetics, forensic as well as genealogical applications⁽¹⁾.

It is generally known and well described, that the number of repeat units in certain TRs causes a pathogenic effect when it exceeds certain thresholds. Several studies have localized more than 50 TRs on the genome, where such pathogenic expansions cause severe neurological, neurodegenerative and neuromuscular disorders, conventionally known as repeat expansion disorders (REDs)⁽²⁾. These are, in general, monogenic diseases, although their inheritance is not always fully mendelistic and anticipation is commonly described in such families (e.g. Huntington's disease, Fragile X syndrome or Myotonic dystrophies). In addition, reports started to

appear which describe the potential role of TRs in the determination of complex phenotypes⁽³⁾, and very recently, the first genome-wide association studies of TRs were published, for example, for Parkinson's disease⁽⁴⁾. This clearly outlines the application potential of TR characterisation not only in the field of diagnostics and complex disease biology, but also in the possible preventive care against several common complex diseases, through potentially more effective identification of at-risk individuals.

Detailed and accurate characterization of TR motifs, in terms of complexity of either structures or biological effects, is an absolutely crucial aspect in assessing their clinical relevance, creating a major challenge for methods aimed at their genotyping. The main challenges of accurate TR characterization include the following factors: 1) The resistance of long TR motifs to amplification by PCR; 2) The complexity of some TR regions, in which interruptions of otherwise pure motifs may occur, or interruptions may be lost, or the clinically relevant motifs may be part of more complex repeat structures; 3) The stutter effect, caused by typical errors of DNA polymerases in TR regions, which makes it difficult to identify the correct fragment in a set of multiple non-specific fragments; and 4) The errors made by classical mapping tools used in sequencing analyses (designed to align sequencing reads containing single nucleotide variants or small insertion-deletions), which are not adapted to the high variability of TR sequences. Therefore, the success of TR analysis highly depends on several factors, including the length and complexity of the repeat sequence and the availability of specific laboratory technologies. Here we provide a basic overview of the most commonly used methods for this purpose. For this review, we decided to split them into two major categories, i.e. those offering more-or-less detailed characterisation of individual loci, with limited possibilities of multiplexing, and those based on sequencing, especially massively parallel sequencing (MPS), offering genome-wide high-throughput characterisation of TR motifs.

Conventional low-throughput TR genotyping methods

For decades, conventional molecular-biology techniques have been used to identify tandem repeat variations, which may be particularly helpful in more effective prediction and/or diagnosis of many human diseases, especially those having monogenic phenotype determination. From these, PCR modifications, and in certain cases also Southern blotting, are still the most commonly used methods, which are considered gold standard in DNA based laboratory diagnostics of TR associated diseases.

Southern blotting

The principle of the method is based on cleavage of DNA fragments by restriction enzymes, which are subsequently separated by gel electrophoresis. The next step is to transfer the DNA fragments to a carrier membrane (usually nylon or nitrocellulose) followed by detection of the target DNA fragment by hybridisation with a labeled probe specific for the TR region. The number of repeats can be determined based on the size of the fragment that hybridizes to the probe. Since it does not require PCR amplification, Southern blot allows to

determine the size of long, pathogenically expanded TR loci. On the other hand, it is not suitable for the differentiation of small alleles, especially those in the healthy and premutational range⁽⁵⁾. The limitations of this method lie mainly in its time-consuming and labor-intensive nature, technical difficulty, and it also requires large amounts of intact high-molecular-weight genomic DNA.

PCR-based methods

The most commonly used technology for amplification and characterisation of TR loci is polymerase chain reaction (PCR), followed by different options of evaluation of amplified fragments. Historically, several technologies have been used, such as evaluation based on restriction fragment length polymorphism (RFLP)⁽⁶⁾, on denaturing high-performance liquid chromatography (dHPLC)⁽⁷⁾, or even based on the denaturation characteristics and melting temperatures of the amplified fragments using high-resolution melting analysis (HRM)⁽⁸⁾. Historically, fragment size analysis by agarose or polyacrylamide gels (PAGE) are well known approaches, however, the most commonly used are automated capillary electrophoresis based separation techniques using genetic analyzers. Agarose electrophoresis allows efficient separation of PCR products only for alleles with a sufficiently large difference or in the case of large alleles. PAGE has a much higher resolution and can separate even fragments that differ by only a few base pairs. Capillary electrophoresis represents the most efficient platform to analyze length profiles of fluorescently labeled PCR products. The advantage of the methodology is that it is both affordable and not specifically labor-intensive, can be automated, and also multiplexed for certain extents^(9,10). Conventional PCR is mainly used to detect alleles belonging to normal ranges or premutation lengths, up to around 300 - 600 nucleotides. For longer alleles, i.e. larger repeats, this method is inaccurate and fails. On the other hand, long-range PCR protocols have been described too. Long-range PCR is used to characterize TRs that exceeded the detection capability of conventional PCR. This methodology uses high-fidelity polymerase and optimized reaction conditions that include extended annealing time and elevated annealing temperature. Such altered reaction conditions allow an extension of the effective range of allele detection, but still only up to a certain size. For visualization of larger alleles, hybridization-based detection may be required too⁽¹¹⁾. Another specific modification of conventional PCR is the so-called small-pool PCR that was described and used to resolve discrete bands in samples giving smeared or diffused bands by Southern blotting. This method is based on partial elimination of preferential amplification and amplification of too many fragments of different length. This is achieved by diluting and separating the template molecules into different individual amplifications. It also allows to increase the range of detected alleles, however, again only up to certain lengths limited by the PCR capabilities themselves⁽¹²⁾. Beyond the well known length limitations, conventional PCR based methods are limited also by the possibilities of getting false results due to: 1) „stutter effect“ caused by DNA polymerase errors; 2) migration changes due to sequence interruptions⁽¹³⁾; or even by 3) allelic dropout due to sequence changes under the primer binding sites.

A specific and highly effective method, called repeat-primed PCR (RP PCR), was developed to overcome the main limitation of conventional PCR (and also of the majority of its modifications which use primer pairs flanking the region of interest), i.e. the inability to amplify through large and complex motifs. It uses two (flanking primer and repeat-specific primer) or three primers (flanking primer, repeat-specific primer and tail primer) per reaction. Flanking primer is fluorescently labeled and it is localized upstream or downstream to the repetitive region. The second primer (repeat-specific) is designed directly in the region of the repetitive motif, generating multiple products. The largest of these products reflect the size of the entire allele, however, with certain limitations. The reactions may be made more specific if a third (tail) primer is used that has no complementarity to the human genome. However, this primer is present in limited amounts in the reaction, resulting in a third primer taking over after a few PCR cycles. This third primer targets the PCR products that were produced by the previous two primers and amplifies them all, generally in a way, in which the repeat specific primer is modified by a 5' tail that creates a template for the third primer. The evaluation is also combined with automated capillary electrophoresis. RP PCR is able to point out single locus expansions^(14,15), however, multiplex versions were also described^(9,16). The disadvantage is that it is unable to determine the specific length of the expanded allele and, for certain complex loci, it was determined to be not 100 % sensitive^(17,18). In addition to expansion detection, RP-PCR is able to identify the presence of sequence interruptions by other motifs within the base repetition. Sequence interruptions, on the other hand, may also disrupt RP-PCR signals to such an extent that they can result in false negative conclusions⁽¹³⁾. Similarly to conventional PCR, false negative results can also occur in a form of allelic dropout, due to sequence variants in the primer binding site. For the mentioned limitations, bidirectional RP-PCR, performed from both ends of the repeat motif, are generally offered to increase sensitivity^(9,13,17).

DNA sequencing methods

Following the introduction and spreading of commercial automatized genetic analyzers, direct DNA sequencing allowing the characterization of the primary structure of DNA, became established as a standard method for laboratory diagnostics and remained in this position for decades. However, in the field of TRs characterisation it had no special role, until second generation sequencing platforms began to appear. This, however, required also specific bioinformatic tools to be developed, which are dedicated to TRs genotyping from millions of short sequencing reads.

Sanger sequencing

Sanger sequencing represents the most widespread first generation sequencing method and was considered as a gold standard among molecular diagnostic techniques allowing screening and genotyping of several types of sequence variants. For genotyping of TR loci, however, it was not routinely used because of several reasons. The disadvantage of the methodology lies in the detection limit (up to approximately 1000 nucleotides). Another problem lies in the analysis of al-

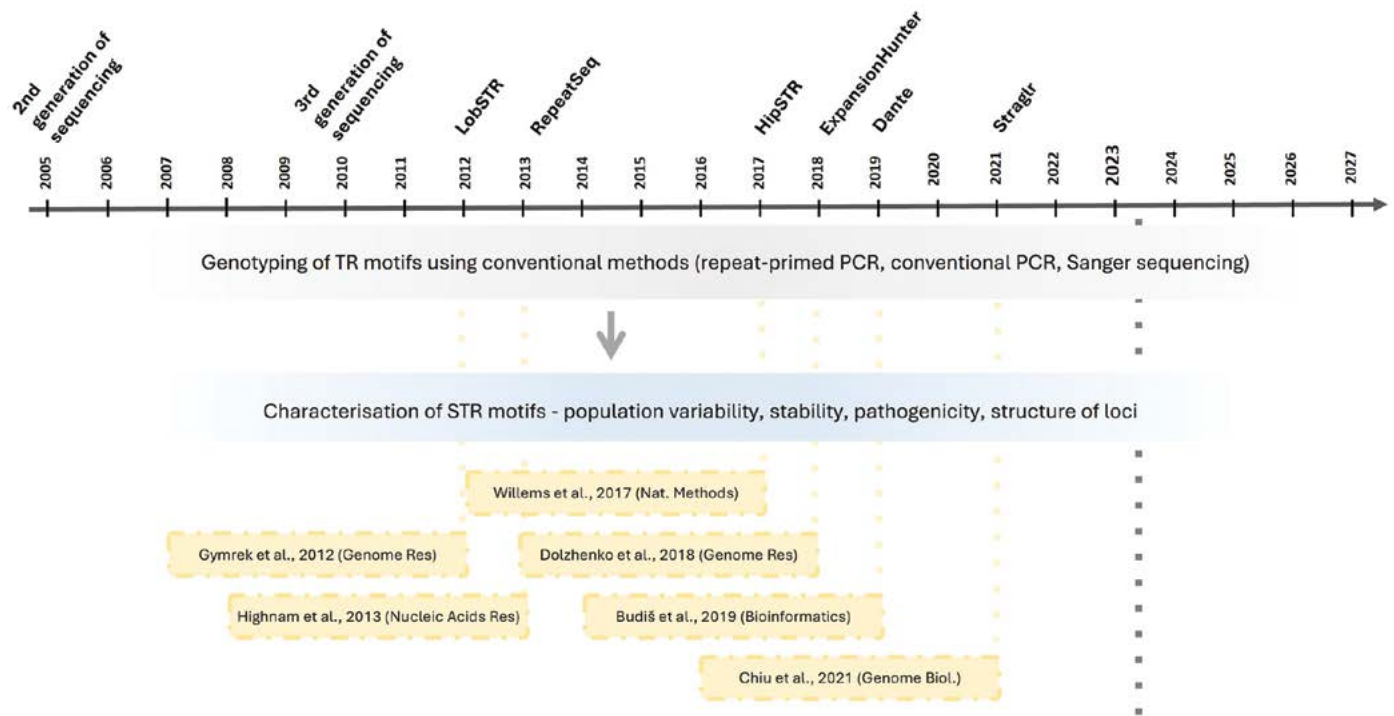
les with a heterozygous constitution, in which based on the shifted signal, a larger allele cannot be determined⁽¹⁹⁾. On the other hand, it can be used, for example, during standardization of other molecular methods, such as conventional and repeat-primed PCR for the fragment sizing. In addition, it was also described to be useful in more detailed characterisation of repeat structures, including sequence interruptions^(20,21).

Massively parallel sequencing

Within MPS applications, we distinguish at least two generations of sequencing methods including those having typically relatively short sequencing reads and requiring clonal amplification before sequencing (second generation) and those allowing longer reads and allowing single molecule sequencing, i.e. which does not require clonal amplification (third generation). From the technical point of view, when considering for example the target genomic region, sequencing can be aimed at the whole genome (WGS), but also can be narrowed to its certain parts, such as to smaller gene panels, larger panels, or even whole exomes (WES). Another relevant technical feature is, whether the preparation of the sequencing library is associated with PCR amplifications (may introduce errors) or are PCR-free (allows reduced error rate and is typically available for WGS). In addition to laboratory and technical aspects, bioinformatic processing of sequencing data may also be challenging and highly variable, while still under intensive development and diversification. In general, actually the most widespread, especially in a routine clinical setting, are platforms belonging to the second generation of sequencing supporting short-read sequencing (especially platforms of the Illumina company). Although still mainly in research settings, some third generation platforms (such as the PacBio or nanopore based sequencing) are appearing as potential tools for more effective TRs characterisation, especially when considering the possibilities of detection of large expansions, or the characterisation of highly repetitive regions. Possibilities and limitations of TR detection are largely connected to the chosen sequencing generation, platform, technical details of library preparation as well as on the applied bioinformatic pipelines.

Second generation platforms generally rely on sequencing following fragmentation of the DNA into small pieces (alternatively on amplicon sequencing), attaching adapters to the ends of the fragments, and then amplifying and sequencing the fragments using a next-generation sequencing instrument. These platforms are generally highly accurate and allow sequencing of short reads of DNA (up to 2 x 300 base pairs, yet). These approaches allow high-throughput characterisation of TRs throughout the genome, although with this regard the main limitations are clear if WES or smaller panels are used, i.e. the TR loci of interest need to be covered by reads. Another limitation is represented by the necessity of PCR amplification, each amplification step introduces challenges and possible errors to TR genotyping. Another limitation lies in the length of the TR regions, since to detailed characterisation they need to be covered by reads in their total length with certain unique flanking regions on their both ends^(22,23). The presence of repeat expansions, even the largest ones, may be indirectly identified from different read characteristics even if they exceed the sizing limit of the re-

Figure 1. Timeline of the evolution of TR locus characterization. Initially, it was possible to genotype TR motifs mainly by methods that represent the gold standard in molecular diagnostics (classical PCR, repeat-primed PCR, Sanger sequencing, Southern blotting,...), but they are still used as validation methods. After the spread of second-generation sequencing platforms into mainstream laboratories, several sophisticated tools have been developed over the last decade specifically for the specific nature of TR motifs (LobSTR, RepeatSeq, HipSTR) but also for the detection of expansions (Expansion Hunter, Dante, Straglr) and the number of these tools is still growing.



ads (i.e. even if they have no reads sequencing through the entire motif), such as the presence of partial reads (reads containing one of the flanking regions and part of the repeat region) or the position of read-pairs, if paired and sequencing was used^(24,25).

On the other hand, larger read lengths and overcoming amplification to a larger extent allows the use of third generation sequencing platforms for more effective TR characterisation, especially by allowing larger and more complex repeats to be directly sequenced by individual spanning reads, even of those belonging to larger expansions. This method makes it possible therefore to identify larger TRs that may be missed by shorter read lengths. These sequencing platforms have, however, some limitations too, including a relatively high error rate and a relatively high cost per base compared to other sequencing technologies. They are therefore most often used in combination with other sequencing technologies, such as those offered by second generation platforms, to generate high-quality, high-resolution genomic data.

Bioinformatic tools

TR typing at the genomic level has become possible with some delay and took longer to catch on, mainly due to certain technical limitations of the conventionally used bioinformatics tools processing MPS data, and high variability in individual TR loci. The majority of genotyping tools are not adapted to TR sequence variability, and they tend to „lose“ some reads in certain specific cases⁽²⁶⁾. In such genotypes, mapping often fails, which can lead to the loss of reads and

incorrect clinical interpretation of predicted genotypes. Several more sophisticated tools have been lately designed specifically for the specific nature of the TR motifs, or even for expansion detection. These, such as STRetch, ExpansionHunter, Dante^(24,27,28), use complex statistical modeling to estimate the underlying genotypes of TR motifs influenced by the stutter effect (Figure 1).

Conclusion

Tandem repeat genotyping is a challenging task due to several inherent characteristics of these DNA sequences. These include the repetitive nature itself, creating challenging sequences for amplification, different and highly variable lengths of repeating motifs, the possible presence or loss of sequence interruptions, but also the complexity of motif structures. Currently, TR motifs are most commonly characterized, in routine practice, by conventional, low-throughput methods. Within these, conventional PCR and repeat-primed PCR are likely the most commonly used. Other PCR modifications, such as long-range PCR or small-pool PCR, or even Southern blotting, are still used, however, they have some specificities and limitations. On the other hand, progressing genomic technologies, such as MPS applications, started to revolutionize genomics and disease diagnostics. The main focus in current development is, therefore, on the possibilities of characterization of TR loci by MPS applications and on improving the accuracy and reliability of TRs genotyping by more effective integration of molecular biology and bioinformatics techniques.

Acknowledgment

This publication was created with the support of the following resources: Operational Program Integrated Infrastructure for the projects: Center for Biomedical Research – BIOMEDIRES – II. Phase, ITMS: 313011W428; PanClinCov, ITMS: 313011ATL7, co-financed by the European Regional Development Fund; Slovak Research and Development Agency under the project ID APVV-18-0319; and Scientific Grant Agency under the project ID VEGA_2/0146/23.

Podakovanie

Táto publikácia vznikla s podporou týchto zdrojov: Operačný program Integrovaná infraštruktúra pre projekty: Centrum pre biomedicínsky výskum - BIOMEDIRES – Fáza II., kód ITMS: 313011W428; PanClinCov, kód ITMS: 313011ATL7, spolufinancované zo zdrojov Európskeho fondu regionálneho rozvoja; Agentúra na podporu výskumu a vývoja SR v rámci projektu ID APVV-18-0319 a Vedecká grantová agentúra v rámci projektu ID VEGA_2/0146/23.

REFERENCES

- Balzano E, Pelliccia F, Giunta S. Genome (in) stability at tandem repeats. *Seminars in Cell & Developmental Biology* 2021; Vol. 113: 97-112.
- Malik I, Kelley CP, Wang ET, et al. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews Molecular Cell Biology* 2021; 22(9): 589-607.
- Gymrek M, Goren A. Missing heritability may be hiding in repeats. *Science* 2021; 373(6562): 1440-1441.
- Bustos BI, Billingsley K, Blauwendraat C, et al. Genome-wide contribution of common short-tandem repeats to Parkinson's disease genetic risk. *Brain* 2023; 146(1): 65-74.
- Buxton J, Shelbourne P, Davies J, et al. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 1992; 355: 547-548.
- Takahashi S, Miyamoto A, Oki J, et al. CTG trinucleotide repeat length and clinical expression in a family with myotonic dystrophy. *Brain and Development* 1996; 18(2): 127-130.
- Oefner PJ, et Underhill PA. DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). *Current protocols in human genetics* 1998; 19(1): 7-10.
- Nicklas JA, Noreault-Conti T, Buel E. Development of a fast, simple profiling method for sample screening using high resolution melting (HRM) of STRs. *Journal of forensic sciences* 2012; 57(2): 478-488.
- Radvansky J, Ficek A, Kadasi L. Upgrading molecular diagnostics of myotonic dystrophies: multiplexing for simultaneous characterization of the DMPK and ZNF9 repeat motifs. *Molecular and Cellular Probes* 2011; 25(4): 182-185.
- Bauer PO, Kotliarova SE, Matoska V, et al. Fluorescent multiplex PCR: fast method for autosomal dominant spinocerebellar ataxias screening. *Russian Journal of Genetics* 2005; 41: 675-682.
- Cheng S, Barcelo JM, Korneluk RG. Characterization of large CTG repeat expansions in myotonic dystrophy alleles using PCR. *Hum Mutat* 1996; 7: 304-310.
- Wong LJ, Ashizawa T, Monckton DG, et al. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet* 1995; 56: 114-122.
- Radvansky J, Ficek A, Minarik G, et al. Effect of unexpected sequence interruptions to conventional PCR and repeat primed PCR in myotonic dystrophy type 1 testing. *Diagnostic Molecular Pathology* 2011; 20(1): 48-51.
- Warner JP, Barron LH, Goudie D, et al. A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J Med Genet* 1996; 33: 1022-1026.
- Falk M, Vojtiskova M, Lukas Z, et al. Simple procedure for automatic detection of unstable alleles in the myotonic dystrophy and Huntington's disease loci. *Genet Test* 2006; 10: 85-97.
- Lian M, Limwongse C, Yoon CS, et al. Single-tube screen for rapid detection of repeat expansions in seven common spinocerebellar ataxias. *Clinical Chemistry* 2022; 68(6): 794-802.
- Radvansky J, Ficek A, Kadasi L. Repeat-primed polymerase chain reaction in myotonic dystrophy type 2 testing. *Genetic Testing and Molecular Biomarkers* 2011; 15(3): 133-136.
- Catalli C, Morgante A, Iraci R, et al. Validation of sensitivity and specificity of tetraplet-primed PCR (TP-PCR) in the molecular diagnosis of myotonic dystrophy type 2 (DM2). *J Mol Diagn* 2010; 12: 601-606.
- Gettings KB, Kiesler KM, Faith SA, et al. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Science International: Genetics* 2016; 21: 15-21.
- Musova Z, Mazanec R, Krepelova, A, et al. Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *American journal of medical genetics Part A* 2009; 149(7): 1365-1374.
- Braida C, Stefanatos RK, Adam B, et al. Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Human molecular genetics* 2010; 19(8): 1399-1412.
- Gymrek M, Golan D, Rosset S, et al. lobSTR: a short tandem repeat profiler for personal genomes. *Genome research* 2012; 22(6): 1154-1162.
- Highnam G, Franck C, Martin A, et al. Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles. *Nucleic acids research* 2013; 41(1): e32-e32.
- Budiš J, Kucharik M, Ďuriš F, et al. Dante: genotyping of known complex and expanded short tandem repeats. *Bioinformatics* 2019; 35(8): 1310-1317.
- Dolzhenko E, Van Vugt JJ, Shaw RJ, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome research* 2017; 27(11): 1895-1903.
- Willems T, Zielinski D, Yuan J, et al. Genome-wide profiling of heritable and de novo STR variations. *Nature methods* 2017; 14(6): 590-592.
- Dashnow H, Lek M, Phipson B, et al. STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome biology* 2018; 19(1): 1-13.
- Tankard RM, Bennett MF, Degorski P, et al. Detecting expansions of tandem repeats in cohorts sequenced with short-read sequencing data. *The American Journal of Human Genetics* 2018; 103(6): 858-873.

Mgr. Ingrid Lojová

Institute of Clinical and Translational Research,
Biomedical Research Center of the Slovak Academy of Sciences
Dúbravská cesta 9, 845 05 Bratislava, Slovakia
e-mail:ingrid.lojova@gmail.com

Zmena expresie génov spojených s imunitnou odpoveďou u pacientov s COVID-19

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

COVID-19 je multiorgánové ochorenie, ktoré je od vypuknutia pandémie subjektom mnohých štúdií. Klinická manifestácia má široké spektrum príznakov od asymptomatického priebehu až po ťažký priebeh, s hlavnými rizikovými faktormi ako vyšší vek a komorbidita. Pochopenie interakcií medzi hosťiteľom a vírusom môže poskytnúť pohľad na nové potenciálne terapeutické ciele v prevencii a liečbe týchto infekcií. Transkriptomová analýza je účinný nástroj na identifikáciu imunitnej odpovede hosťiteľa po napadnutí vírusom SARS-CoV-2. Umožnila identifikovať zmeny v expresii génov a dysregulované dráhy zapojené do aktivácie imunitných buniek. Niekoľko štúdií identifikovalo miRNA a zmeny v ich expresii ako faktor ovplyvňujúci priebeh COVID-19. Rovnako aj defenzíny ako kľúčové komponenty vrodenej imunity majú zmenenú expresiu počas infekcie SARS-CoV-2. Niektoré zmeny v expresii genov sú však vekovo a pohlavne špecifické.

Kľúčové slová: SARS-CoV-2, expresia, cytokíny, miRNA, defenzíny

Altered expression of genes associated with the immune response in patients with COVID-19

COVID-19 is a multi-organ disease that has been the subject of many studies since its outbreak. The clinical manifestation has a broad spectrum from asymptomatic to severe, with older age and comorbidity being the main risk factors. Understanding host-virus interactions may provide insight into new potential therapeutic targets in preventing and treating these infections. Transcriptome analysis is a powerful tool to identify the host's immune response after SARS-CoV-2 infection. Studies identified changes in gene expression and dysregulated pathways involved in immune cell activation. Several studies have identified miRNAs and changes in their expression as a factor affecting the course of COVID-19. Likewise, defensins, as crucial components of innate immunity, have differential expression during SARS-CoV-2 infection. However, some changes in gene expression during COVID-19 are age-specific and sex-specific.

Keywords: SARS-CoV-2, expression, cytokines, miRNA, defensins

NewsLab, 2023; roč. 14 (1): 45 – 48

Úvod

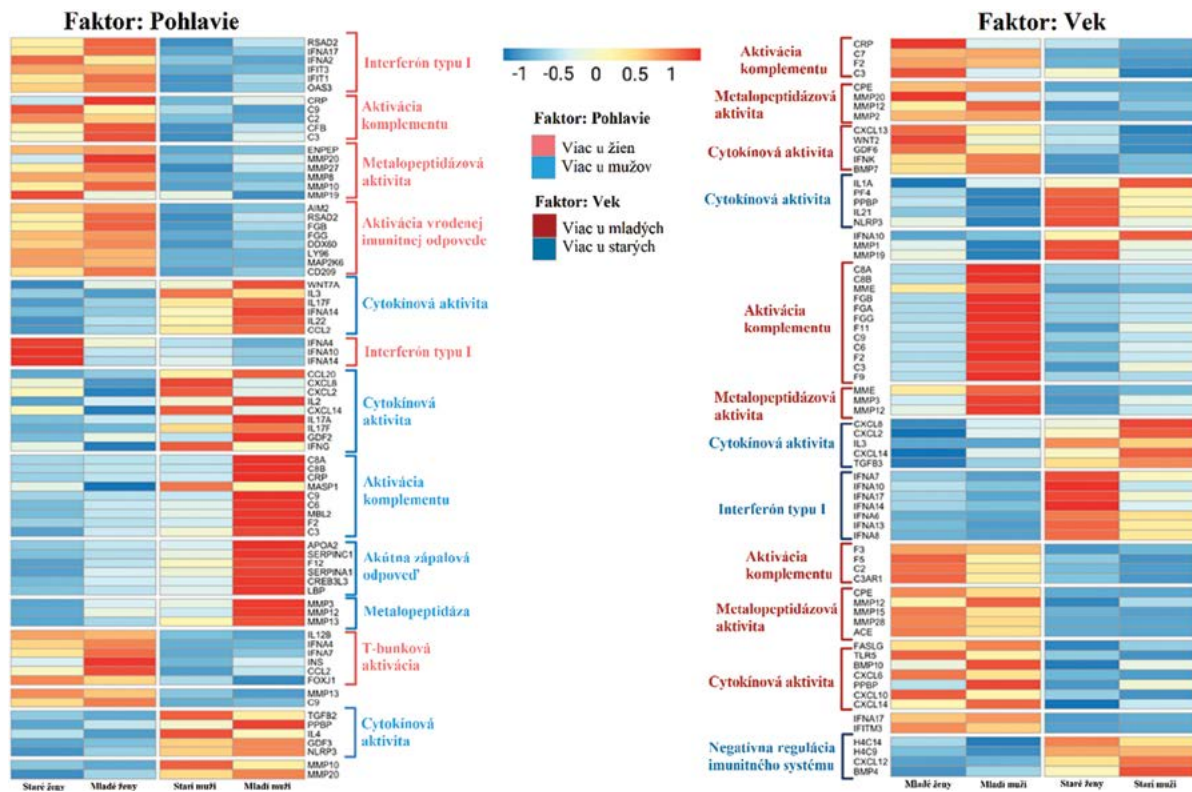
Pandémia COVID-19 spôsobená vírusom SARS-CoV-2 spôsobila veľké preťaženie zdravotného systému po celom svete. Vírus SARS-CoV-2 je jednovláknový RNA vírus s pozitívnou polaritou RNA. Jeho genóm obsahuje minimálne 9 otvorených čítacích rámcov a 4 štruktúrne proteíny. Počiatkové kroky infekcie zahŕňajú väzbu povrchového proteínu S na bunkový receptor ACE-2 (angiotenzín konvertujúci enzým 2). Na samotné naviazanie S-proteínu na ACE-2 receptor je potrebná aktivácia S-proteínu hosťiteľskou bunkovou serín proteázou TMPRSS2. Uzavretá forma S proteínu sa stáva otvorenou a prístupnou pre ACE-2 receptor⁽¹⁾. ACE-2 je lokalizovaný v mnohých bunkách a tkanivách, ako je srdce, obličky, črevný trakt, žľazník, semenníky a, čo je najdôležitejšie, v epiteli dýchacích ciest v nose, ústach a pľúcach. V pľúcach je silne exprimovaný v pneumocytoch v alveolách⁽²⁾. Infekcia SARS-CoV-2 má u pacientov rôzne druhy prejavu, od asymptomatického až po ťažký prejav spojený s hospitalizáciou. Symptómy sa líšia od pacienta k pacientovi a závisia od rôznych faktorov, ako je vek, pohlavie, krvná skupina a ďalšie komorbidity. Niekoľko znakov infekcie SARS-CoV-2 skomplikovalo efektívnu klinickú liečbu, napr. zmena dĺžky trvania inkubačnej lehoty zo 4-5 dní dokonca aj na 8 – 27 dní, vírusová

latencia v hosťiteľovi a taktiež prítomnosť SARS-CoV-2 u klinicky zotavených pacientov⁽³⁾. SARS-CoV-2 inhibíciou kľúčových hosťiteľských proteínov modifikuje bunkové procesy, ktoré majú za následok aktiváciu antivírusových dráh, ako aj zlepšenie transkripcie a translácie vlastných proteínov⁽⁴⁾. Pre lepšie pochopenie interakcií medzi hosťiteľom a vírusom počas infekcie SARS-CoV-2 hrá dôležitú úlohu zmena v transkripcii génov u hosťiteľa.

Cytokíny a chemokíny

Cytokíny majú dôležitú úlohu v reakcii na infekciu koronavírusom. Pacienti s COVID-19 vykazovali vyššiu signalizáciu interferónu γ (IFN- γ) v porovnaní s pacientmi s pneumóniou bez COVID-19, čo poukazuje na IFN- γ ako jeden z počiatkových signálov pri infekcii SARS-CoV-2⁽⁵⁾. Gény kódujúce prozápalové cytokíny (*IL12B*, *IL15*, *IL6*, *IL12A* a *IL1B*) a chemokíny (*CXCL9*, *CXCL11* a *CXCL10*) mali zníženú expresiu u COVID-19 pacientov. Nadmerná aktivácia prozápalových cytokínov a znížená expresia určitých génov stimulujujúcich interferón (*IFNA1*, *APOBEC3G*, *FADD*) môže byť asociovaná s ťažkým priebehom COVID-19⁽⁶⁾. Počas infekcie SARS-CoV-2 bola zaznamenaná zvýšená hladina prozápalových cytokínov, nazývaná aj „cytokínová búrka“. Cytokínová

Obrázok 1. Signifikantné rozdiely v expresii podľa pohlavia (vľavo) a veku (vpravo) pre vybrané gény a dráhy COVID-19. Upravené podľa Liu a kol., 2020.



búrka je závažný jav pozorovaný u pacientov na jednotke intenzívnej starostlivosti, ktorý bol identifikovaný už v súvislosti s ochoreniami SARS a MERS. Spôsobuje hyperaktiváciu a hromadenie zápalových buniek v pľúcach, čo má za následok poškodenie pľúc⁽³⁾. Je jednou z možných príčin ťažkých foriem COVID-19 a jeho úmrtnosti⁽⁷⁾. Okrem toho je u pacientov so SARS-CoV-2 rozdielne exprimovaných 14 génov vrátane *MX1*, *OAS1* a *OAS3*, z ktorých všetky sa podieľajú na antivírusových odpovediach vrátane cytokínovej signalizácie. Gén pre cytotoxickú a regulačnú T-bunkovú molekulu (*CRTAM*) je zodpovedný za reguláciu aktivácie a diferenciácie niekoľkých podskupín T-buniek, ako sú NK bunky. *CRTAM* mal zníženú expresiu u pacientov s COVID-19, ale tiež u pacientov s už existujúcimi pľúcnymi ochoreniami, čo naznačuje jeho možnú úlohu v prognóze COVID-19⁽⁸⁾. Dôležité signálne dráhy, ktoré sú zodpovedné za imunitné reakcie, ako apoptóza, tvorba fagozómov, cytotoxicita sprostredkovaná NK bunkami a signalizácia toll-like receptorov, mali u pacientov s COVID-19 zníženú expresiu⁽³⁾. Gény súvisiace s aktiváciou neutrofilov vrátane *S100A8*, *S100A9* a *S100A12* boli u pacientov s ťažkým priebehom COVID-19 exprimované v monocytoch na vyššej úrovni ako u pacientov s miernym priebehom. Gény súvisiace s cytokínmi a pre interferón boli downregulované⁽⁹⁾.

MikroRNA

MikroRNA (miRNA) sú malé nekódujúce RNA, patria medzi postranskripčné regulátory. Viazu sa na komplementárne molekuly mRNA, čím zabraňujú ich translácii. Zmena v expresii miRNA môže mať rozličné vplyvy na hostiteľský organizmus. MiRNA, ktoré majú za úlohu inhibíciu expície S proteínu a replikácie SARS-CoV-2, miR-7-5p, miR-24-

3p, miR-145-5p, miR-223-3p, sú znížené u starších pacientov s COVID-19⁽¹⁰⁾. U hospitalizovaných pacientov korelovala znížená expresia miRNA, hsa-miR-146a a hsa-miR-126-3p so závažnosťou priebehu COVID-19⁽¹¹⁾. Hladina miR-155, ktorá zohráva úlohu v T-bunkovej a B-bunkovej diferenciácii, bola zvýšená u pacientov s miernym priebehom v porovnaní s pacientmi s ťažkým priebehom COVID-19⁽⁶⁾.

Defenzíny

Defenzíny sú antimikrobiálne proteíny, ktoré sa nachádzajú v živočíchoch aj v rastlinách. Tieto prvky vrodenej imunity slúžia na ochranu hostiteľa pred baktériami a vírusmi. Nachádzajú sa v bunkách imunitného systému, napr. neutrofilných granulocytoch, a pomáhajú zabíjať fagocytované bunky tým, že sa viažu na membránu mikrobiálnych buniek a vytvárajú v nej malé otvory, cez ktoré unikajú životne dôležité ióny. Ich antivírusová aktivita proti iným vírusom naznačuje, že majú úlohu nielen v obrane proti baktériám. Skutočnosť, že sú exprimované v rôznych bunkách a tkanivách, poukazuje na možnosť ich pôsobenia v prvej obrannej línii proti rôznym druhom vírusov. Na základe vzoru disulfidových väzieb klasifikujeme cicavčie defenzíny do α , β a θ podrodiny. α defenzíny sú produkované prevažne neutrofilmi a pri ich aktivácii sú uvoľňované vo veľkých množstvách, β defenzíny sú exprimované najmä v dýchacích cestách⁽¹²⁾, a θ defenzíny sú zriedkavé a vyskytujú sa len v leukocytoch makaka rhesusa⁽¹³⁾. Gény pre β defenzíny *DEFB4A*, *DEFB107B*, *DEFB106B*, *DEFB4B*, *DEFB103A* a 1 α defenzín gén *DEFA1B* boli významne downregulované u pacientov s COVID-19 v porovnaní so zdravou kontrolnou skupinou⁽¹²⁾. Downregulácia týchto génov spôsobená infekciou SARS-CoV-2 poukazuje na zníženú odpoveď vrodenej imunity a taktiež môže

mať za následok zvýšenú kolonizáciu dýchacích ciest baktériami vedúcu k zápalu pľúc.

Rozdiely v transkripčnom profile u mužov a u žien

Epidemiologické štúdie z rôznych častí sveta naznačujú, že pravdepodobnosť nakazenia SARS-CoV-2 je takmer rovnaká pre mužov aj ženy, avšak poukazujú na vyššiu chorobnosť a úmrtnosť mužov⁽¹⁴⁾. Muži majú vyššiu expresiu *ACE2* a *TM-PRSS2*, čo pomáha zvýšenej náchylnosti na infekciu SARS-CoV-2. Orgánom s najvyššou expresiou *ACE2* a *TM-PRSS2* sú obličky⁽¹⁵⁾. Ženy vykazujú vyššiu expresiu INF génov a génov komplementu C2, C3, C9 a CRP (obrázok 1), na rozdiel od mužov, ktorí majú zvýšenú expresiu génov pre prozápalový cytokín Th17 a vyššiu bazálnu expresiu chemokínov CXCL8, CXCL14, CXCL2 a cytokínov IL2, IL3 a IL22. Tieto výsledky korelujú aj so zistením, že pacienti s primárnou interferónovou odpoveďou majú lepšiu prognózu COVID-19 ako pacienti s masívnou aktiváciou cytokínov⁽¹⁶⁾.

Rozdiely v transkripčnom profile založené na veku

Analýza vzoriek pacientov s COVID-19 mladších ako 40 rokov a starších ako 60 rokov poukázala na vekovo špecifické zmeny v expresii niektorých génov. Gény stimulované interferónom, *IFIT1*, *IFIT2* a *IFIT3* zodpovedné za inhibíciu virálnych procesov ako replikácia a translácia, boli upregulované vo vzorkách mladých pacientov s COVID-19. U starších pacientov s nízkou vírusovou záťažou bola expresia génov *IFIT1* a *IFIT2* znížená. Taktiež gén *ADAP2*, čo je gén stimulovaný interferónom, a proteín kódovaný týmto génom môže byť schopný blokovat vstup niektorých RNA vírusov; *TRIM5*, ktorý kóduje retrovírusový restričný faktor; a *TRIM22* a *TRIM38*, kódujúce proteíny zapojené vo vrodenej imunite

proti DNA a RNA vírusom, boli významne downregulované vo vzorkách starších pacientov v porovnaní so zdravou kontrolnou skupinou rovnakého veku. Chemokíny, ktoré hrajú dôležitú úlohu vo funkcii imunitného systému, CXCL11, CCL4 a CCL4 a receptor CCR5 exprimovaný T-bunkami a makrofágmi boli takisto významne downregulované vo vzorkách starších pacientov v porovnaní so zdravou kontrolnou skupinou rovnakého veku. Znížená expresia týchto génov môže mať za následok nerozpoznanie vírusu hosťiteľskými bunkami, čo môže viesť k závažnejšiemu priebehu COVID-19 u starších pacientov. Naopak, u mladých pacientov boli tieto chemokíny a receptor upregulované⁽¹⁷⁾.

Záver

Transkriptómová analýza sa stala jedným z kľúčových nástrojov na pochopenie interakcie hosťiteľa a vírusu SARS-CoV-2 a molekulárnych mechanizmov, ktoré stoja za patogenézou COVID-19. Počas ochorenia COVID-19 sú gény v bunkách rozdielne exprimované v porovnaní so zdravým stavom. Najčastejšie sú rozdielne exprimované gény imunitného systému a dráhy spojené s vrodenu imunitnou odpoveďou. Rozdielne exprimované gény sú aj medzi pohlaviami a rôznymi vekovými kategóriami. Tieto zmeny v expresii môžu byť jedným z vysvetlení, prečo majú ženy a mladší ľudia lepšiu prognózu pri ochorení COVID-19.

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. Benton DJ, Wrobel AG, Xu P, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 2020; 588(7837), 327-330. <https://doi.org/10.1038/s41586-020-2772-0>
2. Wang Y, Wang Y, Luo W, et al. A comprehensive investigation of the mrna and protein level of ace2, the putative receptor of sars-cov-2, in human tissues and blood cells. *International Journal of Medical Sciences* 2020; 17(11), 1522-1531. <https://doi.org/10.7150/ijms.46695>
3. Islam ABMMK, Khan MAAK, Ahmed R, et al. Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2. *Journal of Translational Medicine* 2021; 19(1). <https://doi.org/10.1186/s12967-020-02695-0>
4. Cavalcante LTde F, da Fonseca GC, Amado Leon L A, et al. Buffy Coat Transcriptomic Analysis Reveals Alterations in Host Cell Protein Synthesis and Cell Cycle in Severe COVID-19 Patients. *International Journal of Molecular Sciences* 2022; 23(21). <https://doi.org/10.3390/ijms232113588>
5. Zhang H, Ai JW, Yang W, et al. Metatranscriptomic Characterization of Coronavirus Disease 2019 Identified a Host Transcriptional Classifier Associated with Immune Signaling. *Clinical Infectious Diseases* 2021; 73(3), 376-385. <https://doi.org/10.1093/cid/ciaa663>
6. Li S, Duan X, Li Y, et al. (2021). Differentially expressed immune response genes in COVID-19 patients based on disease severity *Equal contribution 2021; 13(7). www.aging-us.com
7. Montazersaheb S, Hosseiniyan Khatibi SM, Hejazi MS, et al. COVID-19 infection: an overview on cytokine storm and related interventions. In *Virology Journal* 2022; 19(1). BioMed Central Ltd. <https://doi.org/10.1186/s12985-022-01814-1>

8. Alqutami F, Senok A, & Hachim M. (2021). COVID-19 Transcriptomic Atlas: A Comprehensive Analysis of COVID-19 Related Transcriptomics Datasets. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.755222>
9. Xu G, Qi F, Li H, et al. The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. *Cell Discovery* 2020; 6(1). <https://doi.org/10.1038/s41421-020-00225-2>
10. Wang Y, Zhu X, Jiang X-M, et al. (2021). ARTICLE Decreased inhibition of exosomal miRNAs on SARS-CoV-2 replication underlies poor outcomes in elderly people and diabetic patients. *Signal Transduction and Targeted Therapy*. <https://doi.org/10.1038/s41392-021-00716-y>
11. Mitchell MI, Ben-Dov IZ, Liu C, et al. (2021). Extracellular Vesicle Capture by AnTibody of Choice and Enzymatic Release (EV-CATCHER): A customizable purification assay designed for small-RNA biomarker identification and evaluation of circulating small-EVs Masood Kamali-Moghaddam. <https://doi.org/10.1002/jev2.12110>
12. Idris MM, Banu S, Siva AB, & Nagaraj R. (n.d.). Downregulation of Defensin genes in SARS-CoV-2 infection. <https://doi.org/10.1101/2020.09.21.20195537>
13. Tran D, Tran P, Roberts K, et al. Microbicidal Properties and Cytocidal Selectivity of Rhesus Macaque Theta Defensins. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY* 2008; 52(3), 944-953. <https://doi.org/10.1128/AAC.01090-07>
14. Bwire GM. Coronavirus: Why Men are More Vulnerable to Covid-19 Than Women? *SN Comprehensive Clinical Medicine* 2020; 2, 874-876. <https://doi.org/10.1007/s42399-020-00341-w/Published>
15. Sun M, Shankar R, Ko M, et al. Sex differences in viral entry protein expression and host transcript responses to SARS-CoV-2 Meng-

ying Sun Michigan State University 2020. <https://doi.org/10.21203/rs.3.rs-100914/v2>

16. Liu T, Balzano-Nogueira L, Lleo A, & Conesa A. Transcriptional differences for covid-19 disease map genes between males and females indicate a different basal immunophenotype relevant to the disease. *Genes* 2020; 11(12): 1-14. <https://doi.org/10.3390/genes11121447>

17. Singh K, Chen Y C, Hassanzadeh S, et al. Network Analysis and Transcriptome Profiling Identify Autophagic and Mitochondrial Dysfunctions in SARS-CoV-2 Infection. *Frontiers in Genetics* 2021; 12. <https://doi.org/10.3389/fgene.2021.599261>

Mgr. Klaudia Babišová

MEDIREX GROUP ACADEMY. n. o.

Novozámocká 1/67, 949 05 Nitra

e-mail: klaudia.babisova@medirexgroupacademy.sk

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,

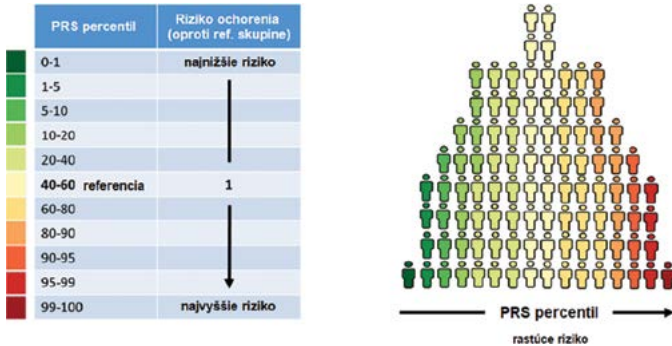
prípadne v dôsledku pôsobenia faktorov vonkajšieho prostredia. Väčšina variantov má minimálny, resp. žiadny fenotypový prejav, a teda neovplyvňujú zdravotný stav ich nositeľa, avšak viaceré z nich sa môžu prejaviť aj na úrovni fenotypu nevynímajúc ani zmeny patologického charakteru manifestovateľné až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

Urothelial carcinoma of the bladder with adamantinoid (ameloblastoma-like) features. Case report

Michal Zámečník, MD^{1,2}, Peter Kohútek, MD, MPH³

¹Medicyt s. r. o., Laboratory of Surgical Pathology, Trenčín, Slovak Republic

²Agel, a. s., Laboratory of Surgical Pathology, Nový Jičín, Czech Republic

³Department of Urology, Faculty Hospital, Trenčín, Slovak Republic

An unusual case of bladder urothelial carcinoma that showed areas of adamantinoid (ameloblastoma-like) morphology is described. The tumor occurred in 69-ys-old patient. It was a recurrence of non-invasive low-grade urothelial carcinoma which had been resected 3 years ago. Currently, two transurethral resections were performed (to ensure the tumor removal was complete). The tissue from both resections contained urothelial carcinoma without invasion in lamina propria. Histologically, besides of conventional morphology of low- and high-grade papillary urothelial carcinoma, numerous foci of the tumor showed inverted growth with ameloblastoma-like features such as reversed polarization of basal cells and loose network-like arrangement of the cells in the lobules. At present, the patient is well one month after the resection, and he will receive intravesical BCG treatment. We suppose that adamantinoid (ameloblastoma-like) urothelial carcinoma can be regarded as a variant of inverted urothelial carcinoma.

Keywords: urothelial carcinoma, urinary bladder, transurethral resection, ameloblastoma, adamantinoid

Urotelový karcinóm močového mechúra s adamantinoidnou morfológiou. Kazuistika

Prezentujeme prípad urotelového papilokarcinómu s neobvyklou morfológiou, ktorá je podobná ameloblastómu (adamantinómu). Išlo o rekurenciu low-grade noninvazívneho karcinómu močového mechúra u 69-ročného muža. Tumor bol odstránený transuretrálnou resekciou v dvoch etapách. Histologicky šlo o urotelový karcinóm, high-grade, bez invázie do strómy. Okrem tradičnej papilárnej a invertovanej morfológie obsahoval tumor invertované lobuly s polarizáciou bazálnych buniek a dehiscenciou, resp. retikuláciou epitelu vnútri lobulov. Stav pacienta je mesiac po resekcii primeraný a je u neho plánovaná BCG terapia. Predpokladáme, že urotelový karcinóm s adamantinoidnou morfológiou možno považovať za podtyp urotelového karcinómu s invertovaným rastom.

Kľúčové slová: urotelový karcinóm, močový mechúr, transuretrálna resekcia, ameloblastóm, adamantinoidný

NewsLab, 2023; roč. 14 (1): 52 – 54

Introduction

Common urothelial carcinomas (UC) consist of stratified urothelial epithelium with nuclear atypia and mitoses⁽¹⁾. Histological diagnosis of these cases is usually straightforward and their biological behavior is well-known. However, there exist variants of UC which show unusual histological and immunohistochemical features. These tumor can cause difficulties, regarding both histological diagnosis and subsequent treatment. Spectrum of variant types includes squamous, glandular, micropapillary, nested, microcystic, inverted, villous-like, basaloid, sarcomatoid and lymphoepithelioma-like⁽¹⁻³⁾. Recently, we have seen in our practice a case of UC with papillary and inverted pattern⁽⁴⁾ which, moreover, showed ameloblastoma-like (adamantinoid) morphology with reversed polarization of basal cells and loose, network-like arrangement of cell islands⁽⁵⁻⁷⁾. We would like to briefly present this unusual case here.

Case report

A 69-ys-old male was admitted for recurrence of non-invasive low-grade UC that had been resected 3 years ago, without subsequent chemotherapy or BCG therapy. Otherwise, his

medical history includes stage 1 hypertension, type 2 diabetes mellitus, obesity, metabolic syndrome, combined hyperlipoproteinemia, and hyperuricemia. Recent control cystoscopy revealed 4 polypoid tumors in the posterior wall of the bladder, measuring 5 cm x 4 cm, 3 cm x 2 cm, 2 cm x 2 cm, and 1 cm x 1 cm, respectively. The lesions were removed by two subsequent transurethral resections (TUR). The tissue fragments were examined histologically and immunohistochemically. In both TUR specimens, we have found common papillary UC with areas of both high- and low-grade and with focal inverted growth pattern. Invasion into the lamina propria was not present, and numerous pieces of muscularis propria did not contain tumor infiltration as well. In addition to common histological features of UC, we have seen areas with adamantinoid (ameloblastoma-like) morphology (**figures 1-4**). They were found in approximately one third of the tumor tissue. These areas included islands and cords of the basaloid appearing epithelium, with reversed polarization of the nuclei of the basal cells (**figures 3 and 4**). The cells inside the islands and cords showed loss of cohesion, resembling reticulated change seen typically in ameloblastoma⁽⁵⁻⁷⁾. In addition, numerous lobules contained round hyalin-

Figure 1. Low-power shows adamantinoid lobules (right) and conventional UC (left).

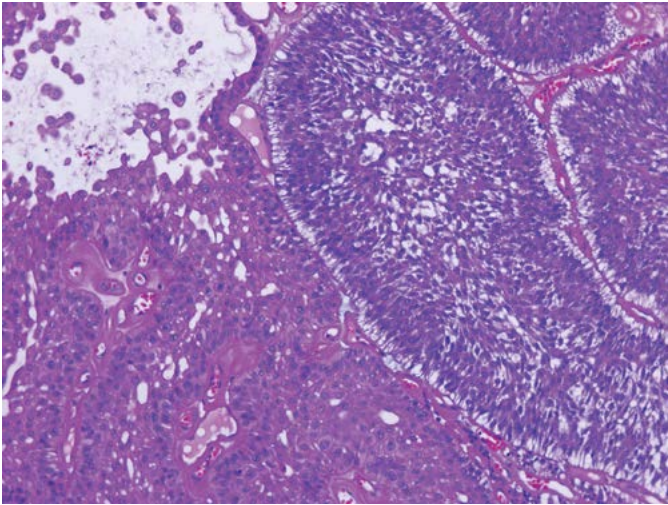


Figure 2. Superficial papillae show reversed polarity of the basal cells.

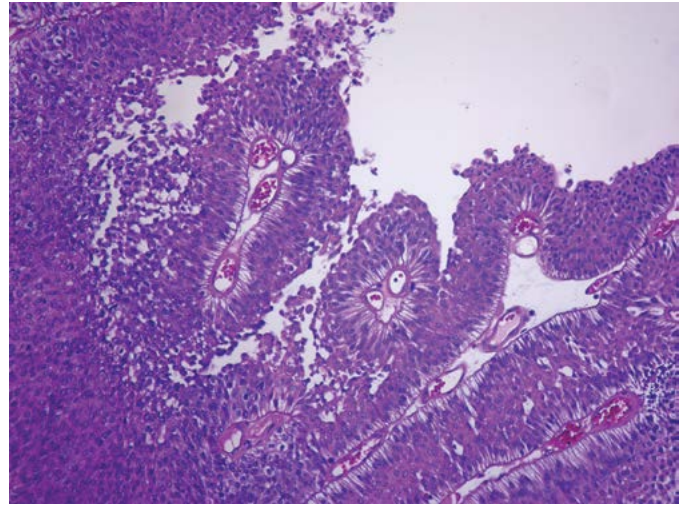


Figure 3. Some of the lobules contain gland-like spaces (right and bottom).

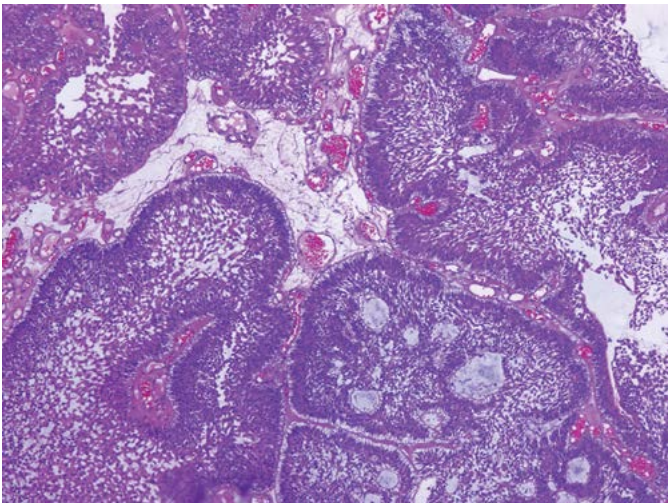


Figure 4. Gland-like spaces are clearly developed due to inclusion of stroma into the lobule (creating "solid papillary" appearance). Arrow indicates direct transition between perilobular and hyalinized intralobular stroma.

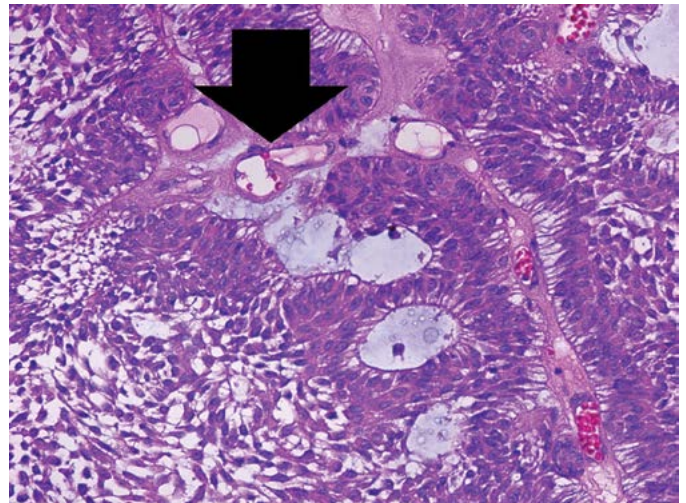
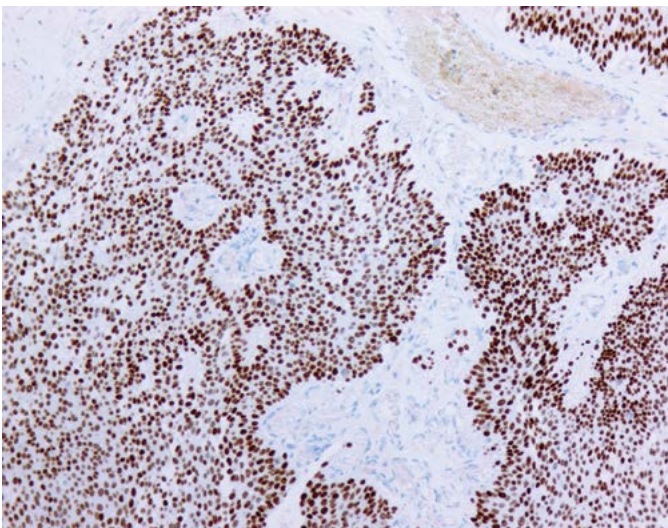


Figure 5. Expression of GATA3 in the tumor.



ized and edematous cores of papillae that mimicked glands. These pseudoglands showed, however, polarized basal cells indicating their stromal (non-glandular) nature (**figure 4**), and they were negative for mucicarmine. In addition, we reviewed slides from the tumor removed three years ago. We have found only focal reversed polarity of the basal cells in some papillae (as depicted in **figure 2**), without apparent adamantinoid lobules.

For immunohistochemistry, following antibodies were used: CK20 (clone Ks20.8), GATA3 (L50-823), p53 (DO-7), CK HMW (34BE12), p63 (DAK-P63), p40 (poly), CK5/6 (D5/16B4), CK19 (RCK108), HER2 (4B5), glypican 3 (GC33), alpha-fetoprotein (poly), CDX2 (DAK-CDX2), beta-catenin (14) calretinin (DAK-Calret 1), alpha-smooth muscle actin (1A4), and CD56 (123C3). The tumor cells were positive for GATA3 (**figure 5**), p63, p40, CK19, CK-HMW, CK20 and CK5/6. P53 was positive in 30% of tumor cells. Neoplastic cells were negative for CDX2, alpha-fetoprotein, beta-catenin, calretinin, glypican 3,

CD56, alpha-smooth muscle actin, and HER2. Currently, the patient is well one month after the tumor resection, and he will receive intravesical BCG treatment.

Discussion

The present tumor shows focal unusual features that resembled ameloblastoma (adamantinoma), such as reversed polarization of basal cells and loose and network-like arrangement of the cells in the lobules⁽⁵⁻⁷⁾. To our knowledge, this morphology was not explicitly described in UC before. Regarding reverse cell polarity in urothelial lesions, we have found in the literature one case of inverted glandular papilloma⁽⁸⁾. Sundaram et al. shows in their Fig. 1 the reversed polarity of the glandular epithelium. In common UC, one can sometimes see subtle reversed polarity with subnuclear clearing of the cytoplasm, as depicted for example in Figure 2 in study by Comperat et al.⁽⁹⁾. However, these features are so mild that they do not resemble ameloblastoma. So-called adamantinoid morphology, typical of ameloblastoma, was rarely reported in non-odontogenic tumors. Well known are adamantinoid basal cell carcinoma⁽¹⁰⁾, adamantomatous craniopharyngioma⁽⁷⁾, and adamantinoid trichoblastoma⁽¹¹⁾. Differential diagnosis in our case included UC with glandular differentiation and rare yolk sac tumor with glandular and

reticular pattern^(1-3,12,13). As mentioned previously, true glands in UC contain epithelial mucin positive secretion, and this contrasts with stromal features in gland-like spaces seen in our case⁽¹⁻³⁾. Yolk sac tumor of the bladder or UC with yolk sac tumor differentiation can contain primitive endodermal cells with subnuclear vacuoles and with reticular pattern^(12,13). However, our tumor was negative for markers of yolk sac tumor, such as alpha-fetoprotein, glypican and CDX2.

Estimation of biological aggressiveness of presented UC is difficult, and observation of additional cases is needed. Currently, it seems that among UCs with variant histology, only micropapillary, sarcomatoid, squamous cell and basaloid types indicate worse prognosis, although knowledge is still limited⁽²⁾. In our case, the tumor showed features of high-grade UC with both exo- and endophytic inverted growth and without stromal invasion. Therefore, we suppose that behavior of the tumor will be similar to common high-grade non-invasive UC and that management of the patient should be in accord with current guidelines for such tumors.

In sum, we described unusual case of UC with adamantinoid (ameloblastoma-like) morphological features seen in substantial part of the lesion. The tumor showed both high- and low-grade areas, and it was non-invasive. Additional similar cases are needed for better knowledge of these tumors.

REFERENCES

1. Raspollini MR, Comperat EM, Lopez-Beltran A, et al. News in the classification of WHO 2022 bladder tumors. *Pathologica* 2022; 115(1): 32–40.
2. Lopez-Beltran A, Blanca A, Cimadamore A, et al. T1 bladder carcinoma with variant histology: pathological features and clinical significance. *Virchows Arch* 2022; 480(5): 989–998.
3. Lopez-Beltran A, Cheng L (2006) Histologic variants of urothelial carcinoma: differential diagnosis and clinical implications. *Hum Pathol* 2006; 37(11): 1371–1388.
4. Bang H, Park H, Park S, et al. Clinicopathologic study of 60 cases of urothelial neoplasms with inverted growth patterns: Reclassification by international consultation on urologic disease (ICUD) recommendations. *Ann Diagn Pathol* 2020; 44: 151433.
5. Ide F, Ito Y, Miyazaki Y, et al. New Look at the history of peripheral ameloblastoma. *Head Neck Pathol* 2020; 14(4): 1052–1057.
6. Ludvíková M, Michal M, Zámečník M, et al. Desmoplastic ameloblastoma. *Cesk Patol* 1998; 34(3): 94–98.
7. da Silva LAM, Filho SRC, Saraiva MJD, et al. Clinical, radiographic and histopathological analysis of craniopharyngiomas and ameloblastomas. a systematic review. *Head Neck Pathol* 2022; 16(4): 1195–1222.
8. Sundaram A, Gupta K, Singla S, et al. Glandular-type inverted papilloma of urinary bladder: report of a unique case with immunohistochemistry. *Indian J Pathol Microbiol* 2020; 63(1): 96–97.
9. Compérat EM, Burger M, Gontero P, et al. Grading of urothelial carcinoma and the new “World Health Organisation Classification of Tumours of the Urinary System and Male Genital Organs 2016”. *European Urology Focus* 2019, 5: 457–466.
10. Berk DR, Ball Arefiev KL, et al. Adamantinoid basal cell carcinoma: a predictor of more-aggressive clinical behavior. *Dermatol Surg* 2012; 38(8):1346–1350.
11. Betti R, Alessi E. Nodular trichoblastoma with adamantinoid features. *Am J Dermatopathol* 1996; 18(2): 192–195.
12. Magers MJ, Kao CS, Cole CD, et al. “Somatic-type” malignancies arising from testicular germ cell tumors: a clinicopathologic study of 124 cases with emphasis on glandular tumors supporting frequent yolk sac tumor origin. *Am J Surg Pathol* 2014; 38(10): 1396–1409.
13. Zamecnik M. Urothelial carcinoma of the bladder with foci of yolk sac tumor. *J Urol Pathol* 1999; 11: 161–170.

Michal Zamecnik, MD

Medicyt, s.r.o., lab. Trencin
Legionarska 28, 911 71 Trencin, Slovak Republic
e-mail: zamecnikm@seznam.cz

MEDIREX GROUP ACADEMY n. o.,
Slovenská parazitologická spoločnosť pri SAV
a Trnavská univerzita, Fakulta zdravotníctva
a sociálnej práce, Trnava



člen
MEDIREX GROUP
všetko pre vaše zdravie



POZVÁNKA

na vedeckú konferenciu

II. AKTUÁLNE PROBLÉMY HUMÁNNEJ PARAZITOLÓGIE

20. septembra 2023 | od 12.30 hod.

v konferenčnej miestnosti **MEDIREX GROUP ACADEMY** n. o.,
laboratórneho komplexu na Galvaniho ul. 17/C, Bratislava,
5. poschodie

PROGRAMOVÝ VÝBOR

prof. RNDr. František Ondriska, PhD.
RNDr. Vojtech Boldiš, PhD.
doc. RNDr. Michal Stanko, DrSc.
doc. RNDr. Martina Horváthová, PhD.

INFORMÁCIA O KREDITOCH

Vedecká konferencia je zaradená do kontinuálneho medicínskeho
vzdelávania Slovenskej lekárskej komory a bude hodnotená kreditmi.

KOORDINÁTORKA

Mgr. Lenka Palšová | +421 917 533 743, e-mail: lenka.palsova@medirexgroup.sk

ODBORNÝ GARANT PODUJATIA

prof. RNDr. František Ondriska, PhD., Medirex, a.s., člen skupiny **MEDIREX GROUP**

REGISTRÁCIA

Svoju aktívnu alebo pasívnu účasť na konferencii potvrdíte, prosím, prostredníctvom
webovej stránky www.medirexgroupacademy.sk v sekcii **Vzdelávanie – Odborné
podujatia** alebo zaslaním e-mailu na adresu lenka.palsova@medirexgroup.sk.

TERMÍNY REGISTRÁCIE

Aktívna účasť a poslanie rukopisu **do 30. 5. 2023**, pasívna účasť **do 8. 9. 2023**.

**Príspevky z konferencie budú
publikované v recenzovanom
časopise newslab v plnom znení**
(kategória publikačnej činnosti AED).

www.newslab.sk
www.medirexgroupacademy.sk

