

# PROTEOMICKÁ ANALÝZA VZORKŮ KONDENZÁTU VYDECHOVANÉHO VZDUCHU PROKÁZALA VYSOKOU REPRODUKOVATELNOST HMOTNOSTNĚ SPEKTROMETRICKÝCH MĚŘENÍ

JANA VÁCLAVKOVÁ, PAVLA KOUŘILOVÁ,  
JANA VRBKOVÁ, DUŠAN HOLUB, MARIÁN  
HAJDÚCH a PETR DŽUBÁK

Laboratoř experimentální medicíny, Ústav molekulární a translační medicíny, Lékařská fakulta, Univerzita Palackého v Olomouci, Hněvotinská 5, 779 00 Olomouc  
petr.dzubak@upol.cz

Došlo 27.5.20, přijato 1.6.20.

Klíčová slova: dechový kondenzát, hmotnostní spektrometrie, proteomika, biomarkery, onemocnění plic

## Úvod

Odběr kondenzátu vydechaného vzduchu (EBC, z angl. Exhaled Breath Condensate) je jednoduchá, neinvazivní, a tedy slibná metoda, pomocí které můžeme získat vzorek komplexního materiálu z dýchacích cest. Kondenzát vydechaného vzduchu je možné jednoduše odebrat a následně analyzovat i v případě pacientů z řad dětí předškolního věku<sup>1,2</sup>. Odběr EBC bývá prováděn s využitím různých odběrových přístrojů, a to buď vlastními, nebo komerčně dostupnými zařízeními. Z komerčně dostupných zařízení lze zmínit EcoScreen, TurboDECCS, RTube, a skleněný kondenzátor Anacon glass. V této studii bylo použito odběrové zařízení TurboDECCS, které se skládá z přenosné Turbo jednotky (Turbo je akronym anglického Transportable Unit for Research of Biomarkers Obtained) a odběrové soupravy DECCS (z angl. Disposable Exhaled Condensate Collection System). Odběrová souprava DECCS se skládá z náustku, jednocestného ventilu, odběrové zkumavky a sběrné cely, která je součástí chladicího systému Peltierova typu. Pokud dospělý člověk do přístroje klidně dýchá po dobu 10 min, je možné získat 1–3 ml EBC (cit.<sup>3</sup>). Kondenzát vydechaného vzduchu je tvořen z 99,9 % kondenzovanou vodní parou<sup>3</sup> a zbytek tvoří různé těkavé i netěkavé sloučeniny z dolního respiračního traktu<sup>4</sup>. Většina biomarkerů je zde přítomna ve velmi malém množství, které je blízké detekčnímu limitu většiny metod<sup>5</sup>.

V minulosti byly proteiny v EBC detegovány metodou vysokorozlišovací 2D elektroforézy v menším než nanogramovém množství<sup>5</sup>. U dětí s astmatem byly identifikovány cytokiny a růstové faktory s využitím enzymové imunoanalýzy<sup>6,7</sup> a multiplexní imunodetekce<sup>2</sup>. Prvním

hmotnostně spektrometrickým analýzám předcházelo rozdělení proteinů gelovou elektroforézou, po kterém následovala identifikace vyřezaných spotů s využitím hmotnostního spektrometru<sup>9–11</sup>. Tímto přístupem bylo v EBC identifikováno několik keratinů, a to jak při analýze EBC zdravých jedinců<sup>8</sup>, tak při analýze EBC mechanicky ventilovaných pacientů<sup>9</sup> a při analýze EBC u pacientů s plicním emfyzémem<sup>10</sup>. Autoři zmíněných studií se potýkali s problémem nízké koncentrace proteinů ve vzorcích EBC, a proto někteří z nich přistoupili ke spojení odebraných vzorků do jednoho vzorku<sup>8,10</sup>.

Nízká koncentrace proteinů ve vzorku je hlavním problémem při studiu proteomu EBC (cit.<sup>11</sup>). V minulosti byly testovány různé metody detekce proteinů bez předchozí separace gelovou elektroforézou a několik přístupů pro zakoncentrování vzorků. Byly otestovány a publikovány metody precipitace proteinů kyselinou trichloroctovou, precipitace pyrogalolovou červení, zachycení proteinů na porézním nosiči na bázi kopolymeru styrendivinybenzenu (POROS-20-R2-RP), odstranění lipidů a precipitace methanolem a chloroformem, precipitace deoxycholátem sodným a kyselinou trichloroctovou, ultramembránová centrifugace s využitím různých membrán, extrakce na pevné fázi s využitím kolon ZipTipC18, vysušování mrazem, vysušování pod vakuem a chromatografie na reverzní fázi<sup>11–13</sup>. Precipitace pyrogalolovou červení vykazovala vysokou návratnost proteinů, nicméně toto barvivo není kompatibilní s MS analýzou. Z výše zmíněných metod pro zakoncentrování proteinů bylo nejvýhodnější zachycení proteinů na nosiči POROS-20-R2-RP, současně bylo zjištěno, že vysušování mrazem je špatně reprodukovatelná metoda<sup>11</sup>. Jiná výzkumná skupina však publikovala, že lyofilizace je nejšetnější a nejrychlejší metodou s vysokým výtěžkem proteinů<sup>12</sup>. Také chromatografie na reverzní fázi byla publikována coby vhodná metoda pro zakoncentrování proteinů ve vzorku<sup>13</sup>. Implementace metody pro zakoncentrování vzorku a použití metody bez předchozí elektroforetické separace vedlo ke zlepšení citlivosti detekce. Důkazem toho je, že relativní kvantita jednoho z proteinů, kalgranulinu B, byla přibližně 20× nižší než v případě nejabundantnějšího proteinu, cytokeratinu 10 (cit.<sup>11</sup>).

Dalším krokem k charakterizaci kompletního proteomu EBC byla analýza vzorku vzniklého smícháním EBC odebraného od devíti zdravých jedinců. Proteiny v tomto vzorku byly identifikovány vysokorozlišovací hmotnostní spektrometrií (HRMS) s využitím technologie Orbitrap, ionizace nanoESI, ve spojení s předchozí elektroforetickou a HPLC separací. Tento přístup vedl k identifikaci 167 unikátních produktů genové exprese<sup>14</sup>. V další studii, v níž byla použita metoda elektroforetické separace ve spojení s detekcí proteinů pomocí nanoLC a hmotnostní spektrometrie s technologií Orbitrap, bylo identifikováno 229 unikátních proteinů. Zmíněných 229 proteinů bylo identifikováno ve dvou vzorcích, z nichž každý byl připraven smícháním 10 vzorků EBC od zdravých jedinců. Proteinů, které byly identifikovány v obou analyzovaných vzorcích, bylo 153 (cit.<sup>15</sup>). Pro analýzu jednotlivých vzorků od 24 intubovaných novorozenců napojených na me-

chanickou ventilaci byla použita metoda přímé přípravy vzorku v roztoku spojená s HPLC separací a detekcí na hmotnostním spektrometru založeném na iontové cyklotronové rezonanci s Fourierovou transformací (FT-ICR MS). S využitím semi-kvantitativní techniky bez značení proteinů bylo možno detegovat 119 proteinů ve všech vzorcích<sup>16</sup>.

Další zlepšení ve smyslu počtu identifikovaných proteinů přinesla analýza EBC ve studiích týkajících se plicních nádorů a chronické obstrukční plicní nemoci (CHOPN). Více než 300 proteinů bylo detegováno ve vzorcích, které byly zpracovány metodou bez předchozí elektroforetické separace a analyzovány metodou HPLC FT-ICR MS s vysokým rozlišením<sup>17</sup>. Ve studii zabývající se včasnou detekcí rakoviny plic bylo analyzováno ještě více vzorků, a tudíž vzrostl i počet identifikovaných proteinů. Kondenzáty vydechovaného vzduchu byly odebrány od 192 jedinců: 49 zdravých kontrol, 49 kuřáků, 46 pacientů s CHOPN a 48 pacientů s rakovinou plic. Analýza s využitím hmotnostního spektrometru nanoESI QqTOF vedla k identifikaci 348 různých proteinů<sup>18</sup>. Obě zmíněné studie také navrhly potenciální biomarkery pro včasnou diagnózu rakoviny plic<sup>17,18</sup>. Nicméně to, že v těchto studiích byl identifikován vyšší počet proteinů, může být ovlivněno také tím, že v EBC pacientů s rakovinou plic byla prokázána přítomnost více proteinů než v případě pacientů s CHOPN a zdravých kontrol<sup>18</sup>. Dalším přístupem bylo využití kvantitativní hmotnostní spektrometrie s vysokým rozlišením pomocí technologie Orbitrap, kdy bylo identifikováno 257 proteinů. Do této studie bylo zahrnuto 19 pacientů s CHOPN a 19 zdravých kontrol, přičemž byly analyzovány jednotlivé vzorky. Mezi zmíněnými 257 proteiny bylo objeveno 24 proteinů, které byly rozdílně exprimovány u pacientů s CHOPN ve srovnání s kontrolou<sup>19</sup>.

Studie založené na hmotnostní spektrometrii již byly schopny vyřešit problém s nízkou koncentrací proteinů ve vzorcích EBC. Nicméně publikované studie se zaměřují na kompletní charakterizaci proteomu EBC, nebo na identifikaci biomarkerů onemocnění. Ve většině studií, jež se zaměřily na kompletní charakterizaci proteomu, bylo více odebraných vzorků EBC spojeno do jednoho vzorku, což sice vedlo ke zvýšení počtu identifikovaných proteinů, ale ztratila se informace o interindividuální variabilitě vzorků od jednotlivých jedinců. Naše studie je první studií, která zavedla měření vzorků dechového kondenzátu v biologických a technických replikátech, porovnávala proteinový profil jednotlivých vzorků od zdravých jedinců a charakterizovala reprodukovatelnost měření tří technických replikátů jednoho vzorku.

## Experimentální část

### Vzorky ve studii

Vzorky byly odebrány od 3 zdravých členů jedné rodiny. Jednalo se o vzorky EBC dvou sester ve věku 8 a 11 let a jejich otce ve věku 45 let.

### Odběr EBC

Kondenzát vydechovaného vzduchu byl od účastníků studie odebrán pomocí přístroje Turbo 14 Turbo DECCS System (Medivac, Itálie). Účastníci studie dýchali do přístroje po dobu 10 min s volným nosem, přičemž vzorky byly odebrány ve třech časových bodech během dvou hodin. Po odběru byly vzorky EBC okamžitě zamrazeny a skladovány při  $-80\text{ }^{\circ}\text{C}$  do doby analýzy.

### Příprava vzorku

Kondenzáty vydechovaného vzduchu byly lyofilizovány v přístroji SpeedVac (Thermo Fisher Scientific). Lyofilizované proteiny byly rozpuštěny v TEZ pufru (10 mM Tris, pH 8,0; 1 mM EDTA; 0,002% Zwittergent 3-16). Vzorky byly stočeny v centrifuze při 14 000 g;  $20\text{ }^{\circ}\text{C}$ , po dobu 10 min a supernatanty byly přeneseny do 1,5 ml eppendorfek. Bylo provedeno teplem zprostředkované zpřístupnění antigenu při  $98\text{ }^{\circ}\text{C}$  po dobu 90 min a poté byly proteiny denaturovány sonikací ve vodní lázni po dobu 30 min při pokojové teplotě. Následně proběhla digestce trypsinem (Trypsin Gold, Mass Spectrometry Grade, Promega, výsledná koncentrace trypsinu ve vzorku 0,5  $\mu\text{g}$ ) při  $37\text{ }^{\circ}\text{C}$  přes noc. Proteiny byly redukovány 10  $\mu\text{M}$  dithiothreitem a získané peptidy byly purifikovány za použití kolonek Stage tips<sup>20</sup>. Purifikované vzorky byly zakonzentrovány ve vakuové odparce (Eppendorf) a rozpuštěny v 50  $\mu\text{l}$  roztoku 1% acetonitrilu a 0,05% kyseliny trifluoroctové (TFA).

### Hmotnostně spektrometrická analýza

Hmotnostně spektrometrická analýza každého vzorku byla provedena ve 3 technických replikátech na hmotnostním spektrometru Orbitrap Fusion<sup>TM</sup> s ionizačním zdrojem Proxeon EASY-Spray (Thermo Fisher Scientific) po jejich separaci na kapalinovém chromatografu s nanolitrovým průtokem Dionex UltiMate 3000 RSLC. Peptidy byly naneseny na předkolonu Acclaim PepMap<sup>TM</sup> 100 (nanoViper C18, 100  $\mu\text{m}$  vnitřní průměr  $\times$  2 cm, velikost částic 5  $\mu\text{m}$ , velikost pórů 100 Å; Thermo Fisher Scientific). Mobilní fázi pro nanesení vzorku na předkolonu byl 1% acetonitril s 0,05% TFA ve vodě. Bylo naneseno 10  $\mu\text{l}$  vzorku a pro nástřik byl zvolen typ nástřiku „partial-loop“. Předkolona byla přímo propojena s analytickou kolonou EASY-Spray column PepMap<sup>TM</sup> RSLC C18 (75  $\mu\text{m}$   $\times$  15 cm, velikost částic 3  $\mu\text{m}$ , velikost pórů 100 Å; Thermo Fisher Scientific), která byla vytemperována na  $35\text{ }^{\circ}\text{C}$ . Peptidy byly separovány po dobu 95 min a gradient organické mobilní fáze byl na počátku nastaven na 2% (0–5 min) a plynule rostl na 35% (5–65 min) a pak na 90% (65–73 min) a po 73. minutě gradient opět rychle klesl na 2% organické mobilní fáze. Vodná mobilní fáze obsahovala 0,1% kyseliny mravenčí ve vodě a organická mobilní fáze obsahovala acetonitril s 0,1% kyseliny mravenčí. Co se týče hmotnostní spektrometrie, statické napětí na iontovém spreji bylo 2000 V v pozitivním módu, teplota na nástřikové kapiláře byla nastavena na  $200\text{ }^{\circ}\text{C}$ , prekurzorový sken

proběhl v orbitrapu v pozitivním módu, rozlišení FTMS bylo nastaveno na 120 000 a rozpětí  $m/z$  pro skenování prekurzorových iontů bylo 400–1500. Byl zvolen cyklický, na datech závislý mód skenování (cycle time data dependent mode), čas mezi prekurzorovými skeny byl 3 s. Fragmentace iontů byla provedena kolizí indukovanou disociací (CID). Produktový sken MS2 byl proveden v iontové pasti, byl nastaven rychlý mód skenování (rapid scan rate), kolizní energie byla 30 % a čas aktivace 10 ms. Pro prekurzorový i produktový sken byl nastaven sběr dat typu profile.

#### Zpracování dat

Zpracování hmotnostně spektrometrických dat bylo provedeno v beta verzi programu Proteome Discoverer™ 2.5 (Thermo Fisher Scientific). První vyhledávací krok (processing workflow) obsahoval kompletní lidskou databázi z UniProtKB, která obsahovala jak recenzované proteiny z databáze Swiss-Prot, tak nerecenzované výpočetně ověřené proteiny z databáze TrEMBL. Databáze byla stažena z <https://www.uniprot.org/uniprot/> v lednu 2020. Byl použit vyhledávací algoritmus Sequest HT, jako dynamické modifikace byly nastaveny oxidace, *N*-terminální acetylace, ztráta methioninu a ztráta methioninu v kombinaci s *N*-terminální acetylací. Pro validaci dat byl použit Fixed Value PSM Validator. Následný krok vyhledávání (consensus workflow) obsahoval prohledání 10 min okna retenčního času, zda neobsahuje píky již nalezené v jiném vzorku (retention time alignment), pro kvantifikaci byly použity pouze unikátní peptidy, byla použita normalizace na celkové množství peptidů (total peptide amount normalization), nebyla provedena změna měřítka (scaling), relativní kvantita proteinů byla vypočítána s použitím Top3 metody a pro výpočet poměru proteinů byla nastavena metoda založená na abundanci proteinů (protein abundan-

ce based method). Pro imputaci chybějících hodnot byl zvolen mód doplnění z hodnot nízkých abundancí (low abundance resampling) a byl zvolen test hypotéz založený na ANOVA modelu (ANOVA (individual proteins) hypothesis test). Byly vyfiltrovány výsledky pouze pro peptidy s vysokou spolehlivostí detekce.

#### Statistická analýza

Peptid popř. protein byl považován za detegovaný, pokud byl identifikován alespoň v jednom z technických replikátů vzorku. V každém ze vzorků byla abundance pro peptid popř. protein vypočtena jako  $\log_2(m)$ , kde  $m$  je medián z hodnot stanovených (pomocí softwaru Proteome Discoverer) ve třech technických replikátech vzorku. Statistická analýza byla provedena v programu R, ver. 3.5.2 (cit.<sup>21</sup>). Heat mapy byly vytvořeny funkcí Heatmap (balíček ComplexHeatmap).

#### Výsledky a diskuse

Byly analyzovány 3 vzorky od tří osob, jednoho dospělého jedince (otce, F) a dvou dětí (jeho dcer, D1 a D2), každý z nich byl změřen ve třech technických replikátech, tzn. celkem bylo analyzováno 27 vzorků. Vzorky byly analyzovány v softwaru Proteome Discoverer v jednom vyhledávacím kroku, přičemž analýza byla provedena nejen na úrovni proteinů, ale také na úrovni peptidů. Ve všech vzorcích bylo identifikováno 2797 proteinů a z nich bylo kvantifikováno 2263. Proteiny byly stanoveny na základě identifikace 5006 peptidů, z nichž 4189 bylo kvantifikováno. V naší studii byl identifikován vyšší počet proteinů než ve studiích doposud publikovaných. Přestože byl v předchozích studiích využíván přístup, kdy byly odebrané vzorky spojeny do jednoho vzorku a ten pak byl analyzován, v naměřených datech bylo identifikováno méně

#### Tabulka I

Počet proteinů detegovaných v jednotlivých vzorcích, resp. v technických replikátech daného vzorku.

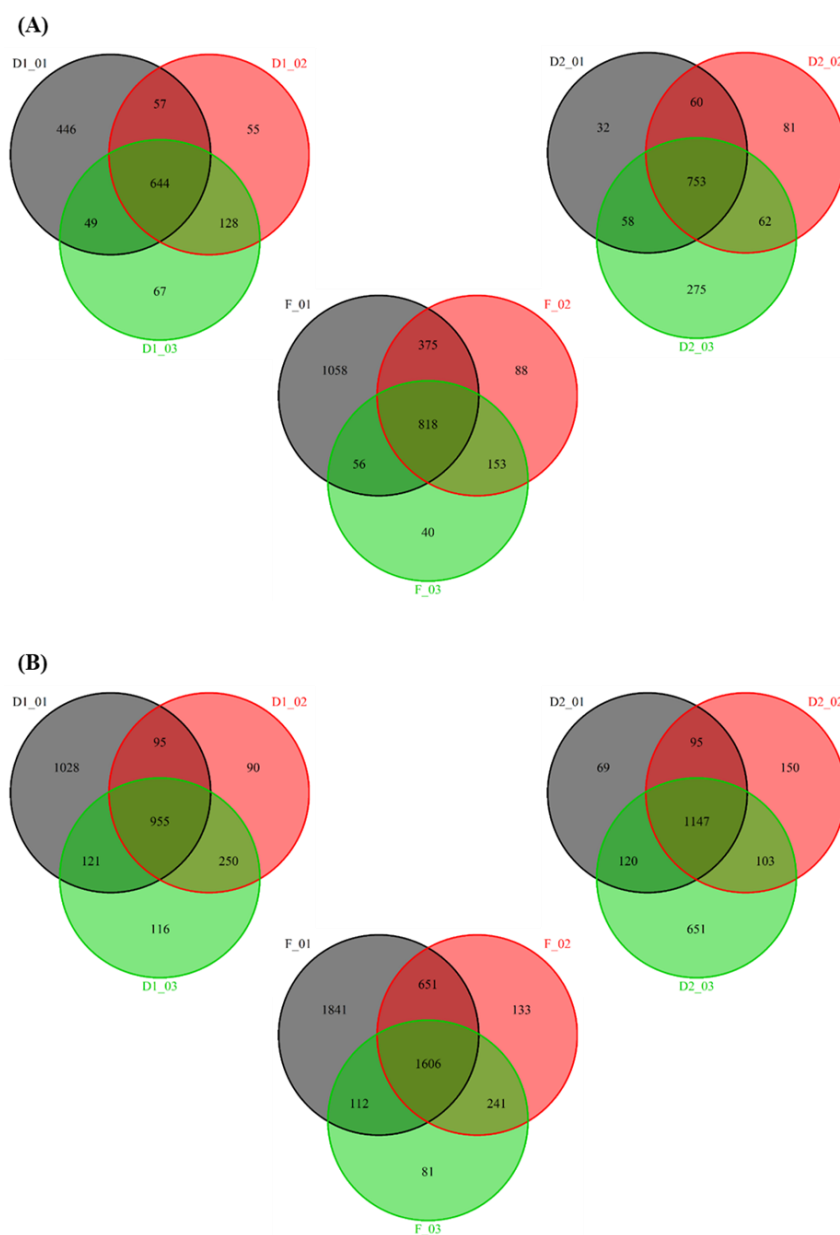
Vzorky byly odebrány třem osobám: otci (F) a jeho dvěma dcerám (D1 a D2). Každé osobě byly odebrány tři vzorky (s rozestupem mezi odběry cca jedna hodina), tj. celkem bylo analyzováno devět (biologických) vzorků. Každý ze vzorků byl vyhodnocen ve třech technických replikátech. Protein byl považován za detegovaný v daném vzorku, pokud byl detegovaný alespoň v jednom technickém replikátu tohoto vzorku

Vzorek	Počet detegovaných proteinů v jednotlivých vzorcích (biologických replikátech)	Počet detegovaných proteinů v jednotlivých technických replikátech (1.   2.   3.)
D1_01	1196	1052 1068 1069
D1_02	884	767  794  787
D1_03	888	753  778  803
D2_01	903	725  772  804
D2_02	956	793  859  764
D2_03	1148	986  972 1006
F_01	2307	1869 1871 1900
F_02	1434	1276 1204 1213
F_03	1067	993  949 930

proteinů. Ve vzorku, který vznikl spojením 9 vzorků zdravých jedinců, bylo identifikováno 167 proteinů<sup>14</sup>. Ve dvou vzorcích, kde každý z nich vznikl spojením 10 vzorků od zdravých jedinců, bylo identifikováno 229 proteinů<sup>15</sup>. A při analýze endogenních částic přítomných ve vydechaném vzduchu bylo ve dvou vzorcích, z nichž každý byl připraven smícháním vzorků od 6 zdravých jedinců, identifikováno 124 proteinů<sup>22</sup>. V následujících studiích již byly vzorky analyzovány samostatně a nebyly připravovány smícháním více vzorků. Při analýze vzorků od 24 intubovaných novorozenců, přičemž každý vzorek byl změřen ve 3 replikátech, bylo identifikováno

119 proteinů<sup>16</sup>. Doposud nejvyšší počet identifikovaných proteinů byl 348. Tohoto počtu bylo dosaženo při hledání biomarkerů rakoviny plic a CHOPN, přičemž studovaná skupina čítala 192 jedinců<sup>18</sup>.

V naší studii bylo detegováno průměrně  $1056 \pm 354$  proteinů v jednom technickém replikátu. Data ze tří technických replikátů byla spojena do jednoho vzorku. Protein byl považován za přítomný v daném vzorku, pokud byl detegován alespoň v jednom technickém replikátu tohoto vzorku. Překryv detegovaných proteinů mezi jednotlivými technickými replikáty byl vysoký, ale přesto došlo k navýšení počtu identifikovaných proteinů na  $1198 \pm 454$

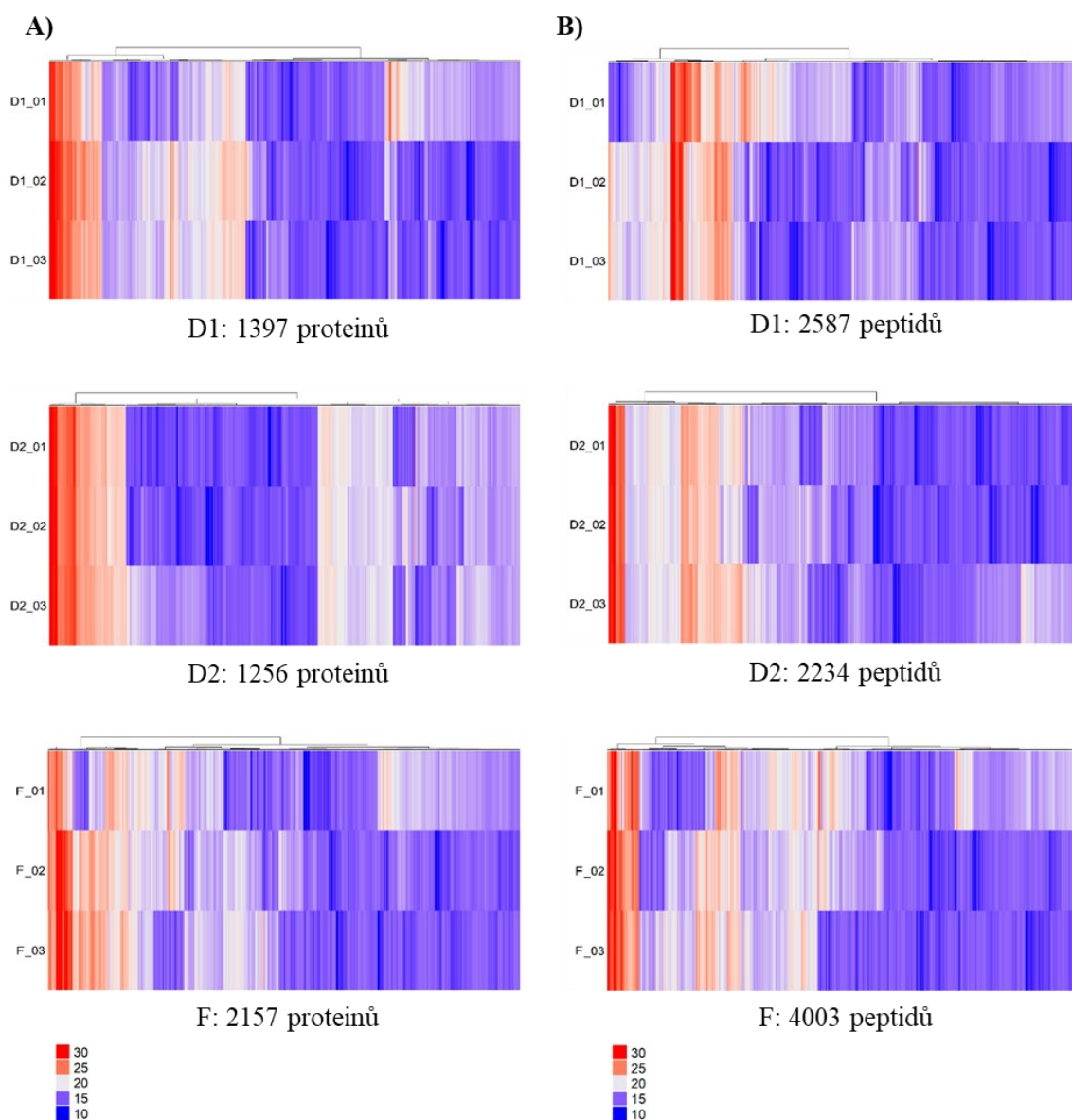


Obr. 1. Počty (A) proteinů a (B) peptidů detegovaných v jednotlivých vzorcích, vyjádřené pomocí Vennových diagramů

ve vzorku v porovnání s měřením jednoho technického replikátu. Počet proteinů detegovaných v jednotlivých technických replikátech a ve vzorcích je shrnut v tab. I. Ve vzorcích od otce byl detegován vyšší počet proteinů ( $1603 \pm 607$ ) než ve vzorcích od dcer ( $996 \pm 140$ ). V literatuře doposud nebyl popsán rozdíl mezi počtem identifikovaných proteinů u dětí a dospělých. Avšak porovnání vzorků od dětí a dospělého není cílem této studie a naše testovaná skupina je příliš malá na to, abychom mohli na základě zmíněného výsledku učinit závěr. Naším cílem bylo ukázat reprodukovatelnost mezi biologickými a technickými replikáty jednotlivých vzorků, ukázat výhody měření vzor-

ků ve více replikátech a charakterizovat vzorky kondenzátu vydechaného vzduchu odebraných od členů jedné rodiny.

Překryv mezi biologickými replikáty (jednotlivými vzorky) od každé osoby byl relativně velký. Pro jednotlivé vzorky byl počet překrývajících se proteinů 644, 753 a 818 (obr. 1A). Ze zmíněných proteinů, které se vyskytovaly ve všech biologických replikátech, bylo 358 proteinů společných pro všechny tři jedince. Analýza překryvu identifikovaných peptidů ukázala výsledky podobné analýze na úrovni proteinů. Překryv mezi biologickými replikáty byl opět relativně vysoký, konkrétně bylo identifikováno 955,

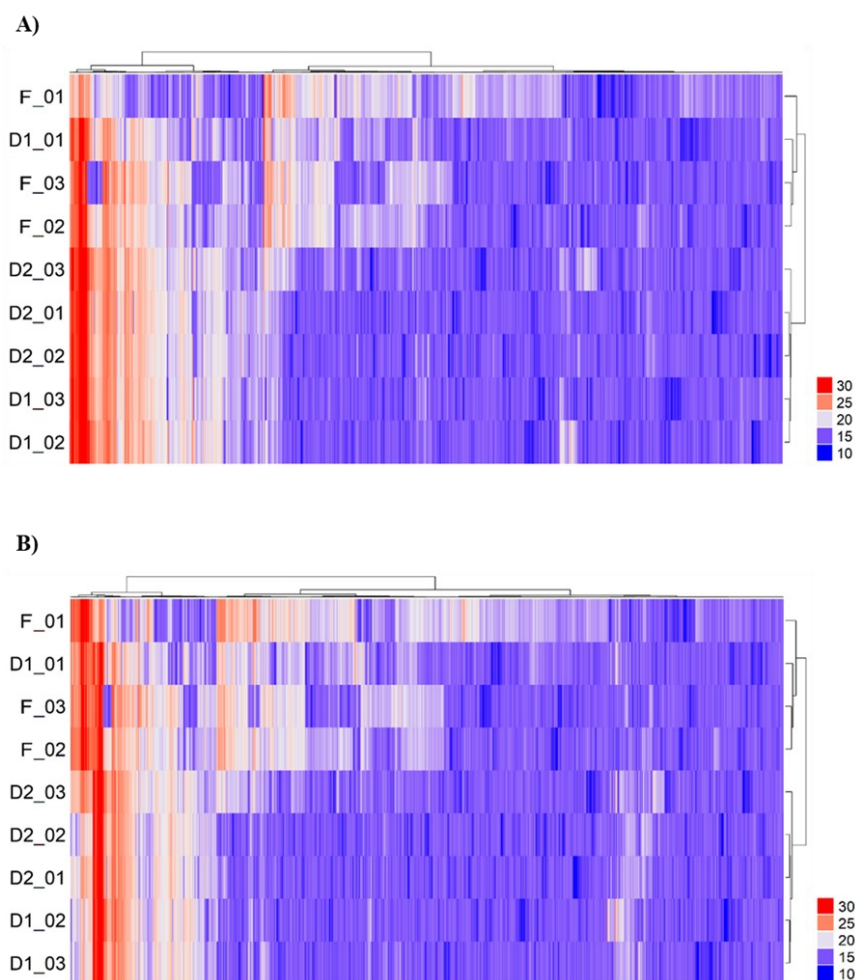


Obr. 2. Heat mapy zobrazující abundance (A) proteinů a (B) peptidů ve třech vzorcích u dané osoby. U každé osoby jsou do heat mapy zařazeny pouze ty proteiny, popř. peptidy, které byly detegovány alespoň v jednom jejím vzorku (a zároveň je u nich určena abundance)

1147 a 1606 peptidů, z nichž 532 bylo společných pro všechny vzorky od všech jedinců ve studii (obr. 1B).

Byla provedena analýza funkčních proteinových sítí pro proteiny společné pro všechny vzorky všech tří členů rodiny. Společných 358 proteinů bylo analyzováno programem STRING, který vyhodnocuje proteinové sítě na základě funkční asociace proteinů. Mezi biologické procesy, které byly vyhodnoceny s poměrným množstvím falešně pozitivních výsledků (FDR) < 0,001, patřily kornifikace, diferenciace keratinocytů, procesy spojené se stavbou kůže a epidermis, keratinizace, organizace cytoskeletu zprostředkovaná intermediárními filamenti, diferenciace epitelálních buněk, organizace intermediárních filament, vývoj tkání a epitelální vývoj. Avšak všech zmíněných procesů se účastní stejné klíčové proteiny, především keratiny. Keratiny byly historicky prvními proteiny, které byly

v EBC identifikovány<sup>8</sup> a bývají obvykle popisovány jako nejabundantnější proteiny v EBC vzorcích zdravých jedinců<sup>8,11,12,14,15,23</sup>. Do dalších biologických procesů byly již zapojeny různé skupiny identifikovaných proteinů. Mezi tyto procesy, které byly vyhodnoceny s parametrem FDR < 0,001, patří degranulace neutrofilů, aktivace myeloidních leukocytů, leukocyty zprostředkovaná imunita, buněčná aktivace, programovaná buněčná smrt, regulovaná exocytóza, procesy efektorů imunity, aktivace leukocytů, imunitní odpověď, buněčná sekrece a procesy imunitního systému. Na základě Gene Ontology analýzy bylo v minulosti ukázáno, že imunitní děje jsou hlavním biologickým procesem, do něhož jsou zapojeny proteiny obsažené v EBC intubovaných novorozenců<sup>16</sup>. Proteinů zapojených do imunitních procesů bylo u zdravých jedinců detegováno více v EBC než v plicním proteomu. Toto zjištění



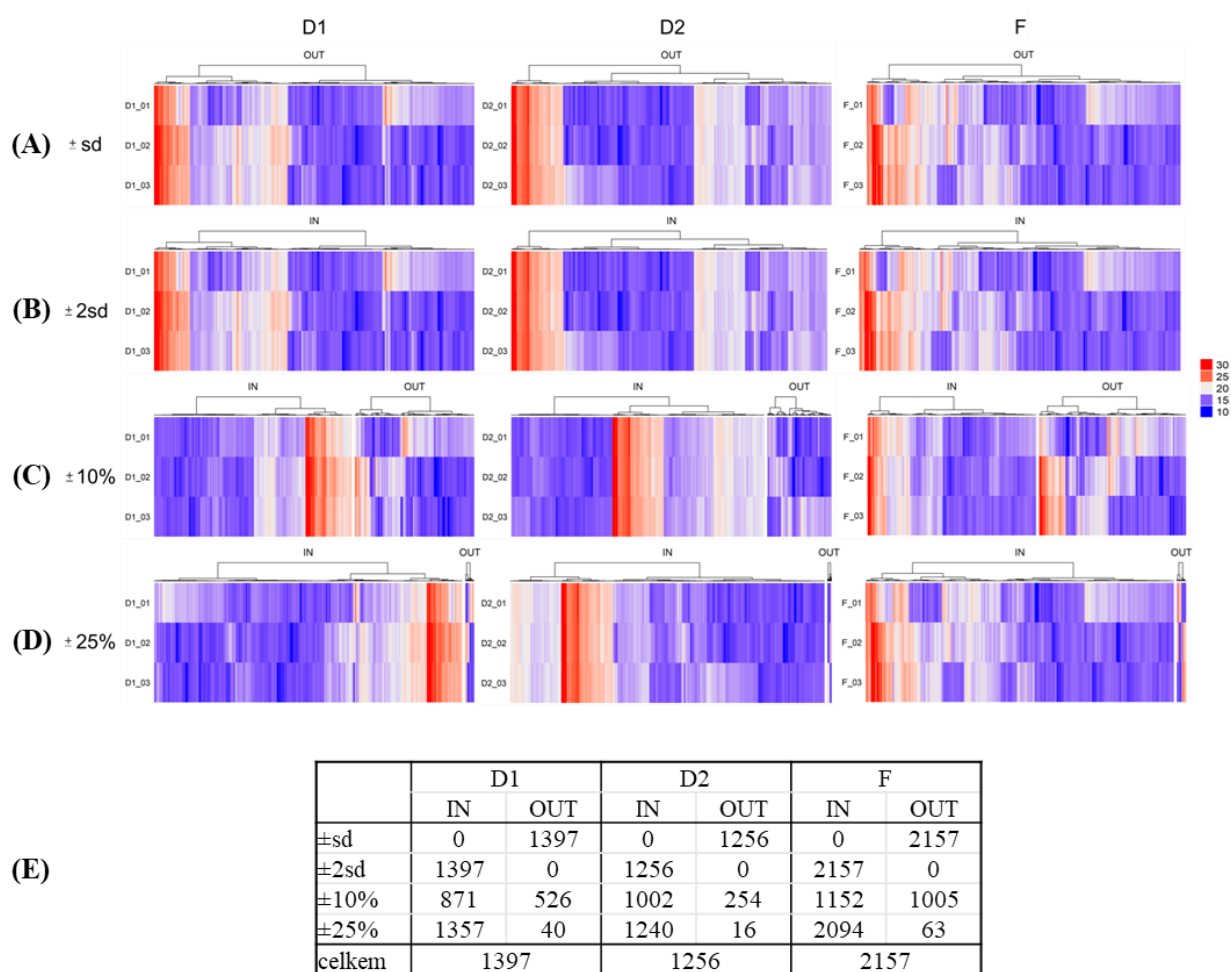
Obr. 3. Heat mapy zobrazující abundance všech proteinů, popř. peptidů, u všech devíti vzorků příslušníků zkoumané rodiny. (A) Heat mapa zahrnuje všech 2263 kvantifikovaných proteinů a (B) heat mapa pak všech 4189 kvantifikovaných peptidů. Sloupce v heat mapě reprezentují jednotlivé proteiny, popř. peptidy, řádky pak jednotlivé biologické vzorky. V obou případech je z dendrogramů na pravé straně heat map zřejmé, že vzorky jednotlivých osob se shlukují k sobě, s výjimkou jednoho vzorku od dcery D1 (s nejvyšším počtem proteinů, popř. peptidů), který vytváří spolu se vzorky od otce jeden shluk



je považováno za důkaz toho, že EBC je důležitou biologickou maticí pro studium funkce respiračního traktu, zejména mukózního sekretu, vrozených a adaptivních antimikrobiálních obranných mechanismů a odstranění inhalovaných částic<sup>15</sup>.

Pro vyjádření abundancí identifikovaných proteinů v biologických replikátech jednotlivých členů rodiny a zobrazení proteinových shluků byly vykresleny heat mapy (obr. 2A). Heat mapy byly vykresleny také pro stejné analýzy na úrovni peptidů (obr. 2B). Do heat map jsou zařazeny pouze ty proteiny, které byly kvantifikovány. Proteiny, které byly detegovány, ale nebyly kvantifiková-

ny ani v jednom ze vzorků, byly z heat map vyloučeny. Vyloučení bylo provedeno, protože k vykreslení heat mapy je potřebná hodnota abundance a nemá smysl zahrnovat do heat map ty hodnoty, které jsou u všech proteinů imputovány, přestože imputované hodnoty jsou velmi nízké. Bylo studováno, zda se vzorky jednotlivých osob shlukují k sobě, proto byla vytvořena heat mapa pro všech devět vzorků (obr. 3A). Vzorky jednotlivých osob se shlukovaly k sobě s výjimkou jednoho biologického replikátu od dceři D1 (D1\_01). Jednalo se o vzorek, který obsahoval větší počet identifikovaných proteinů než všechny další vzorky od dcer a tvořil shluk spolu se vzorky od otce. Vzhledem

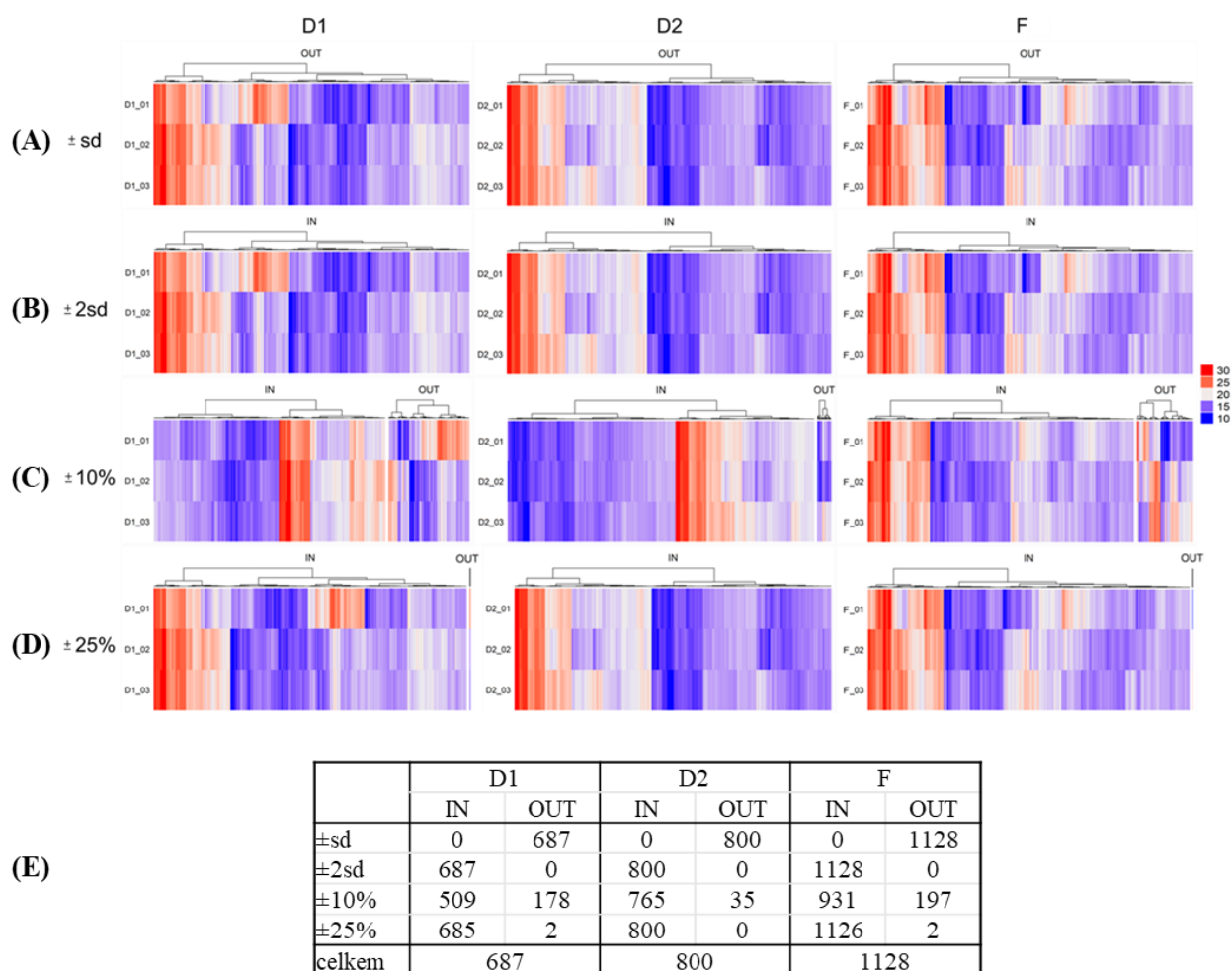


Obr. 4. U jednotlivých proteinů byla k vyjádření variability abundancí, naměřených ve třech vzorcích u každé z osob, použita vzdálenost těchto abundancí od jejich průměru. Do heat map pro danou osobu byly zahrnuty pouze proteiny, které byly kvantifikovány alespoň v jednom ze vzorků této osoby. Z tohoto důvodu je celkový počet proteinů nižší, než bylo uvedeno výše v textu. Pro charakterizaci vzdálenosti od průměru byly uvažovány čtyři různé intervaly. Pro každý typ intervalu a danou osobu pak byly proteiny rozděleny do dvou disjunkčních skupin, označených jako IN a OUT. Pokud příslušný interval obsahoval všechny tři abundancí naměřené ve vzorcích dané osoby, byl zařazen do skupiny IN. V opačném případě, tj. pokud u proteinu ležela hodnota aspoň jedné abundance mimo tento interval, byl protein zařazen do skupiny OUT. Heat mapy (A)–(D) zobrazují abundancí u těchto dvou skupin proteinů samostatně. Typy intervalů byly zvoleny takto: (A)  $\pm sd = (\text{průměr} - sd; \text{průměr} + sd)$ , (B)  $\pm 2sd = (\text{průměr} - 2 \times sd; \text{průměr} + 2 \times sd)$ , (C)  $\pm 10\% = (0,9 \times \text{průměr}; 1,1 \times \text{průměr})$ , (D)  $\pm 25\% = (0,75 \times \text{průměr}; 1,25 \times \text{průměr})$ , kde průměrem rozumíme průměr z abundancí proteinu ve třech vzorcích u dané osoby a jako sd označujeme směrodatnou odchylku tohoto průměru. V tabulce (E) jsou shrnuty počty proteinů patřících do skupin IN a OUT v jednotlivých případech

k tomu, že vzorky byly odebrány od členů jedné rodiny, lze předpokládat, že vzorky budou vzájemně vykazovat velkou podobnost. Vzorky od otce obecně obsahovaly více proteinů než vzorky od dcer. To může být zapříčiněno tím, že otec je dospělá osoba s větším objemem a aktivním povrchem plic, což může mít za následek i vyšší počet proteinů ve vzorku EBC. Počet proteinů může mít při analýze velmi podobných vzorků velký vliv na shlukovou analýzu. Stejná situace byla pozorována i v případě analýzy na úrovni peptidů (obr. 3B).

Byla analyzována vzdálenost abundancí jednotlivých proteinů naměřených ve třech vzorcích u každé z osob od

jejich průměru. Pro každý protein samostatně byla vypočítána průměrná hodnota abundance a byl vypočítán rozdíl tří naměřených hodnot abundance od tohoto průměru. Tato analýza byla provedena pro všechny tři vzorky (biologické replikáty) a pro všechny tři osoby. Pro charakterizaci vzdálenosti od průměru byly uvažovány následující čtyři intervaly:  $\pm sd = (\text{průměr} - sd; \text{průměr} + sd)$ ,  $\pm 2 \times sd = (\text{průměr} - 2 \times sd; \text{průměr} + 2 \times sd)$ ,  $\pm 10\% = (0,9 \times \text{průměr}; 1,1 \times \text{průměr})$  a  $\pm 25\% = (0,75 \times \text{průměr}; 1,25 \times \text{průměr})$ . Bylo zkoumáno, zda naměřené abundance proteinů leží v příslušném intervalu, nebo mimo něj (obr. 4). Opět byla stejná analýza provedena i na úrovni peptidů. Pro analýzu



Obr. 5. U jednotlivých peptidů byla k vyjádření variability abundancí, naměřených ve třech vzorcích u každé z osob, použita vzdálenost těchto abundancí od jejich průměru. Pro každou z osob byly do heat map zahrnuty pouze peptidy, které byly u této osoby kvantifikovány ve všech třech vzorcích ve všech třech replikátech, tj. pouze peptidy kvantifikované ve všech devíti replikátech (technických a biologických), a tedy žádná z hodnot abundancí nebyla imputována softwarem Proteome Discoverer. Pro charakterizaci vzdálenosti od průměru byly uvažovány čtyři různé intervaly. Pro každý typ intervalu a danou osobu pak byly peptidy rozděleny do dvou disjunktních skupin, označených jako IN a OUT. Pokud příslušný interval obsahoval všechny tři abundance naměřené ve vzorcích dané osoby, byl zařazen do skupiny IN. V opačném případě, tj. pokud u peptidu ležela hodnota aspoň jedné abundance mimo tento interval, byl peptid zařazen do skupiny OUT. Heat mapy (A)–(D) zobrazují abundance u těchto dvou skupin peptidů samostatně. Typy intervalů byly zvoleny takto: (A)  $\pm sd = (\text{průměr} - sd; \text{průměr} + sd)$ , (B)  $\pm 2sd = (\text{průměr} - 2 \times sd; \text{průměr} + 2 \times sd)$ , (C)  $\pm 10\% = (0,9 \times \text{průměr}; 1,1 \times \text{průměr})$ , (D)  $\pm 25\% = (0,75 \times \text{průměr}; 1,25 \times \text{průměr})$ , kde průměrem rozumíme průměr z abundancí peptidů ve třech vzorcích u dané osoby a jako sd označujeme směrodatnou odchylku tohoto průměru. V tabulce (E) jsou shrnuty počty peptidů patřících do skupin IN a OUT v jednotlivých případech



peptidů byly do individuálních heat map zahrnuty pouze ty peptidy, které byly kvantifikovány ve všech technických a biologických replikátech vzorků od dané osoby. Počet peptidů kvantifikovaných ve všech vzorcích byl v případě dcer 687 a 800 a v případě otce 1128 (obr. 5). V analýze proteinů i peptidů se vzdálenost jedné směrodatné odchylky od průměru ukázala jako příliš striktní kritérium, neboť u každého z peptidů i proteinů ležela vždy alespoň jedna jeho hodnota abundance mimo interval  $\pm$  sd, a tedy (vzhledem k intervalu  $\pm$  sd) byly všechny peptidy i proteiny zařazeny do skupiny označené OUT (OUT = vyskytující se mimo nastavený interval). V případě použití dvojnásobku směrodatné odchylky byly naopak všechny peptidy i proteiny zařazeny (vzhledem k intervalu  $\pm$  2 sd) do skupiny označené IN (IN = vyskytující se v nastaveném intervalu). Pokud byla odchylka abundance od průměru vyjádřena jako procento z průměru, pak požadavek odchylky nejvýše 10 % byl opět dosti striktní. I přesto však byla (vzhledem k intervalu  $\pm$  10 %) více než polovina peptidů i proteinů zařazena do skupiny IN, tj. vzdálenost všech tří abundance od průměru byla menší než 10 % tohoto průměru. Pokud byl požadavek odchylky nejvýše 25 %, pak mimo tento interval leželo pouze pár desítek proteinů, popř. 2 peptidy, které byly kvantifikovány ve všech vzorcích.

Následně bylo zkoumáno, zda tyto zmíněné odchylky abundance jsou důsledkem odchylky měření pomocí hmotnostní spektrometrie, nebo zda jsou způsobeny intraindividuální biologickou variabilitou vzorků. Analyzovali jsme maximální vzdálenost abundance od průměru pro jednotlivé technické replikáty měření komerčního standardu hovězího sérového albuminu (Pierce™ BSA Protein Digest Standard, LC/MS Grade, Thermo Fisher Scientific), který je v naší laboratoři rutinně používán pro kontrolu kvality hmotnostně spektrometrických měření. Hovězí sérový albumin (BSA) je dodáván lyofilizovaný, naštěpený na peptidy a pro každé měření na hmotnostním spektrometru bylo nastříknuto množství 100 fmol BSA. Pro analýzu technických replikátů BSA byla vypočítána průměrná log<sub>2</sub> abundance ze 3 technických replikátů pro každý peptid samostatně. Byly vypočítány vzdálenosti jednotlivých replikátů od zmíněné průměrné abundance. Byla vybrána největší vzdálenost v absolutní hodnotě. Dále byl vypočten podíl této odchylky na průměru a vyjádřen v %. Tento výpočet byl proveden pro všechny peptidy, které byly kvantifikovány ve všech změřených replikátech. V případě BSA byly všechny identifikované peptidy také ve všech replikátech kvantifikovány. Analýza ukázala, že u všech peptidů byla maximální vzdálenost abundance od průměru menší než 8,11 %. Z těchto dat je jasné, že mezi výsledky měření technických replikátů standardu je odchylka a že je menší než odchylka mezi výsledky měření biologických replikátů. Ovšem standard BSA je komerční směs peptidů o přesně definované koncentraci peptidů a obsahuje výhradně peptidy proteinu BSA, proto lze očekávat, že odchylka měření v případě standardu bude nižší než v případě komplexních biologických vzorků. Abychom tuto hypotézu ověřili, byla vypočítána maximální vzdálenost abundance od průměru pro technické replikáty vzorků. Výpočet byl proveden stejně jako v případě analýzy tech-

Tabulka II

Maximální vzdálenost abundance, naměřených v jednotlivých technických replikátech, od jejich průměru v jednotlivých vzorcích

Vzorek	Max. vzdálenost abundance [%] od průměru
D1_01	12,66
D1_02	19,69
D1_03	17,61
D2_01	13,13
D2_02	10,33
D2_03	8,31
F_01	9,04
F_02	13,39
F_03	13,44

nických replikátů BSA. Opět byly do analýzy zahrnuty pouze peptidy, které byly u této osoby kvantifikovány ve všech jejich technických a biologických replikátech. Výpočet byl proveden pro každou osobu, každý vzorek a každý peptid a vypočítané hodnoty byly dále porovnávány. V tab. II jsou pro jednotlivé vzorky uvedeny maximální vzdálenosti abundance, které byly naměřeny v jednotlivých technických replikátech. Všechny tyto odchylky jsou menší než 19,69 %. Tyto výsledky ukázaly, že variabilita mezi biologickými replikáty vzorků od jedné osoby je větší než v případě technických replikátů, ale rozdíl ve variabilitě mezi biologickými a technickými replikáty není výrazný. Závěrem je možné konstatovat, že rozdíl v abundance proteinů a peptidů, jejichž vzdálenost od průměru je do 10 %, je způsoben více odchylkou měření pomocí hmotnostní spektrometrie než biologickou variabilitou vzorků. Naše výsledky také ukazují, že změna v abundance proteinu pod 20 % průměru je způsobena variabilitou metody a neměla by být považována za významnou změnu v expresi proteinu, pokud používáme label-free techniky hmotnostní spektrometrie.

## Závěr

V této práci jsme popsali metodu pro analýzu proteinů v kondenzátu vydechovaného vzduchu založenou na hmotnostní spektrometrii. Současně jsme prokázali, že se jedná o citlivou a reprodukovatelnou analýzu, kdy variabilita měření je přijatelná pro další interpretaci výsledků a je schopná identifikovat stovky unikátních proteinů. Metoda je tedy vhodným nástrojem pro biologicky zaměřené vědecko-výzkumné práce, a to zejména pro hledání biomarkerů plicních onemocnění.

*Tato práce byla podpořena z grantu Ministerstva zdravotnictví ČR s reg. č. 16-32302A. Veškerá práva podle předpisů na ochranu duševního vlastnictví jsou vyhrazena.*

## LITERATURA

- Horváth I. a 33 spoluautorů: *Eur. Respir. J.* 26, 523 (2005).
- Rosias P. P., Robroeks C. M., van de Kant K. D., Rijkers G. T., Zimmermann L. J., van Schayck C. P., Heynens J. W., Jöbsis Q., Dompeling E: *Pediatr Allergy Immunol.* 21, e235 (2010).
- Konstantinidi E. M., Lappas A. S., Tzortzi A. S., Behrakis P. K.: *Sci. World J.* 2015, 435160.
- Mutlu G. M., Garey K. W., Robbins R. A., Danziger L. H., Rubinstein I.: *Am. J. Respir. Crit. Care Med.* 164, 731 (2001).
- Griese M., Noss J., von Bredow C.: *Proteomics* 2, 690 (2002).
- Shahid S. K., Kharitonov S. A., Wilson N. M., Bush A., Barnes P. J.: *Am. J. Respir. Crit. Care Med.* 165, 1290 (2002).
- Leung T. F., Wong G. W., Ko F. W., Li C. Y., Yung E., Lam C. W., Fok T. F.: *Int. Arch. Allergy Immunol.* 137, 66 (2005).
- Gianazza E., Allegra L., Bucchioni E., Eberini I., Puglisi L., Blasi F., Terzano C., Wait R., Sirtori C. R.: *Am. J. Med.* 117, 51 (2004).
- Gessner C., Dihazi H., Brettschneider S., Hammerschmidt S., Kuhn H., Eschrich K., Keller T., Engelmann L., Sack U., Wirtz H.: *Respir. Med.* 102, 299 (2008).
- Fumagalli M., Dolcini L., Sala A., Stolk J., Fregonese L., Ferrari F., Viglio S., Luisetti M., Iadarola P.: *J. Proteomics* 71, 211 (2008).
- Bloemen K., Hooyberghs J., Desager K., Witters E., Schoeters G.: *Proteomics Clin. Appl.* 3, 498 (2009).
- Kurova V. S., Anaev E. C., Kononikhin A. S., Fedorchenko K. Y., Popov I. A., Kalupov T. L., Bratanov D. O., Nikolaev E. N., Varfolomeev S. D.: *Clin. Chem. Lab. Med.* 47, 706 (2009).
- Núñez-Naveira L., Mariñas-Pardo L. A., Montero-Martínez C.: *Lung* 197, 523 (2019).
- Muccilli V., Saletti R., Cunsolo V., Ho J., Gili E., Conte E., Sichili S., Vancheri C., Foti S.: *J. Pharm. Biomed. Anal.* 105, 134 (2015).
- Lacombe M., Marie-Desvergne C., Combes F., Kraut A., Bruley C., Vandenbrouck Y., Chamel Mossuz V., Couté Y., Brun V.: *J. Breath Res.* 12, 021001 (2018).
- Kononikhin A. S. a 13 spoluautorů: *J. Chromatogr. B* 1047, 97 (2017).
- Fedorchenko K. U. a 11 spoluautorů: *Mosc. Univ. Chem. Bull.* 71, 134 (2016).
- López-Sánchez L. M., Jurado-Gámez B., Feu-Collado N., Valverde A., Cañas A., Fernández-Rueda J. L., Aranda E., Rodríguez-Ariza A.: *Am. J. Physiol. Lung Cell. Mol. Physiol.* 313, L664 (2017).
- Sun C., Zhou T., Xie G., Fu S., Gao L., Liao J., Wu Y., Wang G.: *J. Proteomics* 206, 103392 (2019).
- Rappsilber J., Ishihama Y., Mann M.: *Anal. Chem.* 75, 663 (2003).
- <https://www.R-project.org/>, staženo 20. 12. 2018.
- Bredberg A., Gobom J., Almstrand A. C., Larsson P., Blennow K., Olin A. C., Mirgorodskaya E.: *Clin. Chem.* 58, 431 (2012).
- Fumagalli M. a 11 spoluautorů: *Int. J. Mol. Sci.* 13, 13894 (2012).

**J. Václavková, P. Kouřilová, J. Vrbková, D. Holub, M. Hajdúch, and P. Džubák** (*Laboratory of Experimental Medicine, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc*): **Proteomic Analysis of Exhaled Breath Condensate Samples Showed High Reproducibility of Mass Spectrometric Measurements**

Exhaled breath condensate is considered a rich source of biomarkers that can provide valuable information for early diagnosis of respiratory and systemic diseases. However, exhaled breath condensate (EBC) consists mainly of water vapor and the concentration of analytes in the exhaled breath condensate sample is very low, close to the detection limit of sensitive methods such as mass spectrometry. Here, we have developed a mass spectrometry-based approach to characterize the EBC proteome. We compared biological and technical replicates of the measured samples, compared the protein profile between individual samples of healthy individuals and described the reproducibility between the three technical replicates of the sample taken. We have found that repeated measurements of samples are reproducible and can increase the number of proteins identified. Analysis led to the detection of 2797 unique proteins in all samples. Based on statistical analysis, it was shown that the 10% difference from the mean is due more to the deviation of the mass spectrometry measurements than to the biological variability of the samples. And second, the change in protein amount below 20% of the average in the case of EBC is due to method variability and should not be considered a significant change in protein expression using label-free mass spectrometry techniques.

Full text English translation available in the on-line version.

**Keywords:** Exhaled breath condensate, mass spectrometry, proteomics, biomarkers, lung disease

#### Acknowledgements

*This work was supported by Ministry of Health of the Czech Republic, grant nr. 16-32302A. All rights reserved.*

# PROTEOMIC ANALYSIS OF EXHALED BREATH CONDENSATE SAMPLES: HIGH REPRODUCIBILITY OF MASS SPECTROMETRIC MEASUREMENTS

JANA VÁCLAVKOVÁ, PAVLA KOUŘILOVÁ,  
JANA VRBKOVÁ, DUŠAN HOLUB, MARIÁN  
HAJDÚCH, and PETR DŽUBÁK

*Laboratory of Experimental Medicine, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc, Czech Republic.  
petr.dzubak@upol.cz*

Keywords: Exhaled breath condensate, mass spectrometry, proteomics, biomarkers, lung disease

## Introduction

Exhaled breath condensate (EBC) sampling is a promising, easy and non-invasive method to obtain samples from human lungs. EBC was successfully taken and analyzed even in case of preschool children<sup>1,2</sup>. EBC collection could be performed by various devices including some lab-made or commercially available ones such as EcoScreen, TurboDECCS, RTube, and Anacon glass condenser. In this study, the TurboDECCS device was used. It consists of a portable Turbo Unit (Turbo is an acronym for Transportable Unit for Research of Biomarkers Obtained) and a Disposable Exhaled Condensate Collection System (DECCS). DECCS is equipped with a mouthpiece, a non-return valve, a tube, and a collecting cell inserted in an electrical cooling system of the Peltier type. After 10 minutes of quiet breathing, 1–3 mL of EBC could be collected from adults<sup>3</sup>. Approx. 99.9% of the EBC sample is constituted by condensed water vapor<sup>3</sup>, while the rest contains various volatile and nonvolatile molecules from the lower respiratory tract<sup>4</sup>. Most biomarkers are highly diluted, close to the lower assay sensitivity limits of the methods<sup>3</sup>.

Proteins in exhaled breath condensates were previously detected in subnanogram amounts by high-resolution 2D electrophoresis<sup>5</sup>. Cytokines and growth factors were identified in children with asthma using enzymatic immunoassays<sup>6–7</sup> or multiplex immunoassays<sup>2</sup>. The first mass spectrometry-based analyses included gel-based electrophoretic separation followed by mass spectrometric identification of excised spots<sup>8–10</sup>. This approach led to a few keratin identifications in healthy controls using an HPLC-MS / Q-TOF mass spectrometer<sup>8</sup>, as well as in mechanically ventilated patients using MALDI-TOF MS<sup>9</sup> and in patients with pulmonary emphysema using the HPLC-MS / ESI-ION TRAP model LCQ-Advantage<sup>10</sup>. The authors of these studies faced a problem of a low protein concentra-

tion in EBC samples and therefore some of them used pooled samples<sup>8,10</sup>.

As the low protein concentration represents the main problem in the study of the EBC proteome<sup>11</sup>, the following methods to concentrate the proteins were studied: protein precipitation by trichloroacetic acid or pyrogallol red, precipitation on POROS-20-R2-RP [poly(styrene-*co*-divinylbenzene)] beads, EBC delipidation and precipitation with methanol and chloroform, protein precipitation with sodium deoxycholate and trichloroacetic acid, ultramembrane centrifugation on various membranes, solid phase extraction with ZipTipC18, freeze-drying, vacuum concentration and reversed-phase chromatography<sup>11–13</sup>. The precipitation with pyrogallol red showed high protein recovery but this dye is not compatible with MS protein analysis. Protein capture on POROS-20-R2-RP beads was selected by some authors as the best method for EBC samples protein concentration, while lyophilization was found to be highly irreproducible<sup>11</sup>. However, another group found the lyophilization as the finest and most rapid method of protein concentration with high protein recovery<sup>12</sup>. The reversed-phase chromatography concentration was also chosen as a suitable method<sup>13</sup>. In a study by Bloemen's group, the relative quantity of one of the proteins, calgranulin B, was found to be approximately 20-fold lower, as compared to the most abundant protein, cytokeratin 10. This was considered as an evidence that using a sample concentration method and gel-free approach improve the sensitivity<sup>11</sup>.

The next step to complete the characterization of the EBC proteome was the analysis of pooled EBC samples from nine healthy subjects using the high-resolution Orbitrap-based mass spectrometry (HRMS), nanoESI ionization, combined with electrophoretic separation for protein identification. This approach led to the identification of 167 unique gene products<sup>14</sup>. Another study using electrophoretic separation and high resolution Orbitrap-based mass spectrometry has led to identification of 229 unique proteins in two pooled samples (each consisted of 10 individual EBC samples) and 153 proteins common to both pools<sup>15</sup>. A direct in-solution sample preparation and HPLC analysis in conjunction with Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry (FT-ICR MS) of EBC samples collected from 24 mechanically ventilated newborns were able to detect 119 proteins using label-free semi-quantitative approach<sup>16</sup>.

Another increase of the number of protein identifications came with lung cancer and COPD studies. More than 300 proteins were identified in the samples that were processed by the gel-free method and analyzed by high-resolution HPLC FT-ICR MS<sup>17</sup>. In another study on early diagnosis of lung cancer, more samples were used, thereby increasing the coverage of the EBC proteome. EBCs were collected from 192 individuals: 49 controls, 49 smokers, 46 COPD, and 48 lung cancer patients; here, the nanoESI QqTOF mass spectrometry analysis revealed 348 different proteins<sup>18</sup>. Both studies suggested potential biomarkers of early-stage lung cancer<sup>17,18</sup>. However, higher number of proteins in these studies should be influenced by the fact that more proteins were identified in EBC from lung can-

cer patients, as compared to healthy controls and COPD patients<sup>18</sup>. In another approach<sup>19</sup>, a total of 257 proteins were identified in the high resolution and quantitative Orbitrap-based mass spectrometry study. This study was performed with individual samples (not sample pools), comprising 19 COPD patients and 19 healthy controls. It revealed 24 proteins that were differentially expressed in COPD patients<sup>19</sup>.

Thus, assays based on mass spectrometry have already solved successfully the problem of low protein concentrations in EBC samples. However, these studies are aimed at complete characterization of the EBC proteome or disease biomarkers discovery. Most studies focused on complete proteome analysis used pooled samples, which increased the number of identifications but lost information about the interindividual variability of the samples. In the present study, the measurements of the EBC samples in biological and technical replicates are introduced for the first time, comparing the protein profile between individual samples of healthy subjects and describing the reproducibility between three technical replicates of a single sample.

## Materials and methods

### Study cohort

The samples were taken from 3 individuals from one family. These individuals included two sisters of age 8 and 11 and their father (45).

### EBC collection

The exhaled breath condensates were collected from the study participants using the Turbo 14 Turbo DECCS System (Medivac, Italy). The EBC was being collected for a period of 10 min without wearing a nose clip. Samples were taken at three time points in the interval of two hours. After collection, samples were immediately frozen and stored at  $-80\text{ }^{\circ}\text{C}$  until analysis.

### Sample preparation

The exhaled breath condensates were freeze-dried using a SpeedVac (Thermo Fisher Scientific). Proteins were solubilized in the TEZ buffer (10 mM Tris, pH 8.0; 1 mM EDTA; 0.002% Zwittergent 3-16), cleared at 14,000 g, at  $20\text{ }^{\circ}\text{C}$  for 10 min and the supernatants were transferred to 1.5 mL Eppendorf tubes. Heat-mediated antigen retrieval was performed at  $98\text{ }^{\circ}\text{C}$  for 90 min. Proteins were denatured by sonication in a water bath for 30 minutes at room temperature and digested with trypsin at  $37\text{ }^{\circ}\text{C}$  overnight. Trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega) was added at a final concentration of  $0.5\text{ }\mu\text{g}$  per sample. Proteins were reduced with dithiothreitol ( $10\text{ }\mu\text{M}$  final concentration) and peptides were purified using Stage tips<sup>20</sup>. The purified samples were concentrated in vacuum concentrator (Eppendorf) and dis-

solved in  $50\text{ }\mu\text{L}$  solution of 1% acetonitrile, 0.05% trifluoroacetic acid (TFA).

### Mass spectrometric analysis

Mass spectrometric analysis was performed in 3 technical replicates using the Orbitrap Fusion<sup>TM</sup> mass spectrometer with the Proxeon EASY-Spray ionization source (Thermo Fisher Scientific) coupled to the Dionex UltiMate 3000 RSLC Nanoliquid chromatograph. The peptides were loaded on the Acclaim PepMap<sup>TM</sup> 100 nano trap column (nanoViper C18  $100\text{ }\mu\text{m}$  i.d.  $\times$  2 cm,  $5\text{ }\mu\text{m}$  particle size,  $100\text{ }\text{\AA}$  pore size; Thermo Fisher Scientific), loading solvent was 1% acetonitrile with 0.05% TFA in water and sample loading volume was  $10\text{ }\mu\text{L}$  using a partial-loop injection mode. The trap column was directly connected to analytical EASY-Spray column PepMap<sup>TM</sup> RSLC C18 ( $75\text{ }\mu\text{m} \times 15\text{ cm}$ ,  $3\text{ }\mu\text{m}$  particle size,  $100\text{ }\text{\AA}$  pore size; Thermo Fisher Scientific) heated to  $35\text{ }^{\circ}\text{C}$ . The peptides were separated for 95 minutes; the starting organic phase gradient was set to 2% (0 to 5 min), then it increased gradually to 35% (5 to 65 min), then to 90% (65 to 73 min), and finally dropped quickly back to 2%. As the aqueous and organic mobile phases, water and acetonitrile were used, respectively, both with 0.1% formic acid. Concerning MS analysis, the static positive ion spray voltage was set to 2000 V, ion transfer tube temperature was  $200\text{ }^{\circ}\text{C}$ , master scan was performed in Orbitrap, positive mode, FTMS resolution was set to 120,000 and the  $m/z$  mass range for scanning of precursor ions was 400–1500. Cycle time data dependent mode was chosen, time between master scans was set to 3 s. Ion fragmentation was performed by the collision-induced dissociation (CID). The MS2 scan was performed in an ion trap, the rapid scan rate was set, the collision energy was 30% and the activation time was 10 ms. The profile data type was recorded in both master scan and MS2.

### Data processing

Protein search was performed in beta version of Proteome Discoverer<sup>TM</sup> 2.5 Software (Thermo Fisher Scientific). The processing workflow includes a search against the complete human UniProtKB database, which included both Swiss-Prot reviewed and TrEMBL computationally analyzed but unreviewed proteins (<https://www.uniprot.org/uniprot/>, downloaded in January 2020). Sequest HT search engine was used where oxidation, *N*-terminal acetylation, methionine loss and, finally, methionine loss in combination with *N*-terminal acetylation were set as dynamic modifications. The Fixed Value PSM Validator was used. The consensus workflow included retention time alignment, only unique peptides were used for quantification, total peptide amount normalization was performed, no scaling was done, protein relative quantities were calculated using the Top 3 average method, protein ratio calculation was performed using protein abundance based method, low abundance resampling mode was chosen for missing values imputation and ANOVA

(individual proteins) hypothesis test was used. The results were filtered only for high confidence peptides.

### Statistical analysis

A peptide/protein was considered to be present in a sample, if it had been detected at least in one of the technical replicates of this sample. For each sample and each peptide/protein, abundance was calculated as the log<sub>2</sub> value of the median of the three values (from Proteome Discoverer) from technical replicates of this sample. Statistical analysis was performed using the R program, ver. 3.5.2 (cit.<sup>21</sup>). Heatmaps were created with the Heatmap function (from the ComplexHeatmap library).

## Results and discussion

We have evaluated 3 samples from 3 persons, namely, two children (two daughters, D1 and D2) and an adult (father, F); each sample was measured in three technical replicates, i.e. 27 samples altogether. These samples were evaluated by Proteome Discoverer software in one search. In our study, in addition to protein analysis, analyses were also performed at the peptide level. We have identified 2797 proteins across all samples and 2263 of them were quantified. The proteins were determined on the basis of the identification of 5006 peptides, of which 4189 were quantified. The number of proteins identified in our study was higher than that in previously published studies. Even if the pooled samples were used, 167 proteins were identified in a pool of nine healthy subjects<sup>14</sup>, 229 proteins in two pooled samples from 10 individuals<sup>15</sup>, and 124 proteins in two pooled samples from 6 individuals where endogenous particles were collected in exhaled air<sup>22</sup>. The number of 119 proteins were identified using individual

samples in a cohort of 24 intubated newborns, where each sample was measured in triplicate<sup>16</sup>. Up to date, the highest protein count was 348 different proteins identified in the search for biomarkers of lung cancer and COPD in the entire group of 192 individuals<sup>18</sup>.

We have detected 1056 ± 354 proteins per technical replicate. Three technical replicates were combined into one sample. A protein was considered to be detected in a sample, if it was detected in at least one of the technical replicates of the sample. The overlap of technical replicates was high; in spite of this, however, the number of detected proteins increased to 1198 ± 454 per sample, as compared to a single measurement (the number of proteins detected in individual replicates and samples are shown in Table I). The number of proteins in the sample from the adult (1603 ± 607) was higher than that from the daughters (996 ± 140). In the literature, there is still no evidence of differences in protein counts between children and adults. However, a comparison between children and adult samples is not the aim of this paper, and the group under study is too small to make any conclusion based on this difference. Instead, our intention was to show the reproducibility of biological and technical replicates of the samples, to present the advantages of introducing these replicates and to characterize samples of exhaled breath condensate taken in one family.

The overlap between biological replicates (individual samples) of each subject was relatively high and contained 644, 753 and 818 proteins for particular samples (Figure 1A). Of those proteins, which overlapped between biological replicates, 358 proteins were common to all the three individuals. Peptide analysis of peptide overlap showed results similar to protein analysis. However, as expected, all the numbers increased. The overlap between biological replications also remained relatively high, in particular 955, 1147, and 1606 peptides, 532 of which were common

Table I

Numbers of proteins detected in samples and technical replicates of samples. Samples were taken from three persons: father (F) and two daughters (D1 and D2). Three samples were taken from each person (the time delay between samples was about one hour), i.e. there were nine samples in the analysis. Each of the nine samples was measured in three technical replicates. A protein was considered to be detected in a sample if it was detected in at least one of the technical replicates of the sample

Sample	Number of detected proteins in the sample (biological replicate)	Number of detected proteins in the technical replicate (1 <sup>st</sup>   2 <sup>nd</sup>   3 <sup>rd</sup> )
D1_01	1196	1052 1068 1069
D1_02	884	767  794  787
D1_03	888	753  778  803
D2_01	903	725  772  804
D2_02	956	793  859  764
D2_03	1148	986  972 1006
F_01	2307	1869 1871 1900
F_02	1434	1276 1204 1213
F_03	1067	993  949 930

to all the three individuals (Figure 1B).

We performed a functional protein enrichment analysis common to all samples of all the three family members. Common proteins included 358 proteins that were submitted to STRING functional protein association networks analysis. Cornification, keratinocyte differentiation, skin development, keratinization, epidermis development, intermediate filament cytoskeleton organization, epithelial cell differentiation, intermediate filament organization, tissue development and epithelium development were among biological processes with false discovery rate (FDR) below 0.001. However, the same key proteins, especially

keratins, take part in all these processes. Historically, the keratins were the first proteins identified in EBC samples<sup>8</sup> and were usually referred to as the most abundant proteins in EBC samples of healthy controls<sup>8,11,12,14,15,23</sup>. Other processes with FDR < 0.001, such as neutrophil degranulation, myeloid leukocyte activation, leukocyte-mediated immunity, cell activation, programmed cell death, regulated exocytosis, immune effector process, leukocyte activation, immune response, secretion by cell and immune system process, involved various proteins from interaction network. Using Gene Ontology, immune system processes have been reported to be the main biological processes of

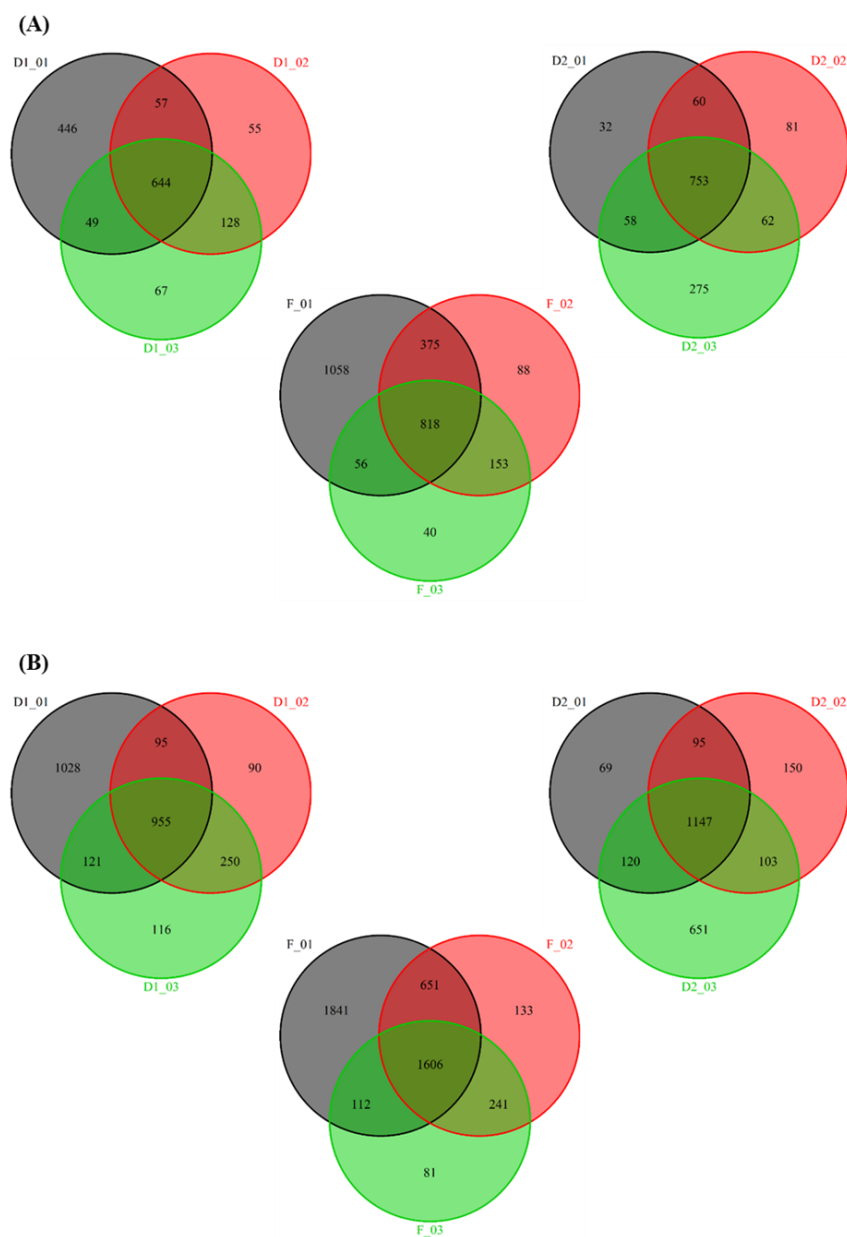


Fig. 1. The number of (A) proteins and (B) peptides detected in each sample and the overlap between biological replicates of each person expressed by Venn diagrams



proteins in EBC of intubated newborns<sup>16</sup> and were also overrepresented in EBC compared to lung proteome in healthy controls. This finding was considered as an evidence that EBC was a relevant matrix to study the major physiological functions of the respiratory tract, particularly mucosal secretion, innate and adaptive antimicrobial defense mechanisms and clearance of inhaled particles<sup>15</sup>.

Heatmaps were used to express abundancies and clustering of proteins (Figure 2A) and peptides (Figure 2B) in biological replicates of individual family members. Only quantified proteins were used to construct the heatmap. Proteins that were detected but not quantified across all

samples were excluded because we needed an abundance to draw a heatmap, and it was useless to include values that were all imputed, despite the fact that the imputed values were very low. We assessed whether or not the samples of the individuals were clustered together, so the heatmap was constructed for all the nine samples (Figure 3A). Daughters' samples were clustered together, except for the first biological replicate from D1 (D1\_01). It was a sample with a higher number of proteins identified than that of the other samples of daughters and was clustered together with father's ones. Since the samples were taken from the members of the same family, it could be assumed

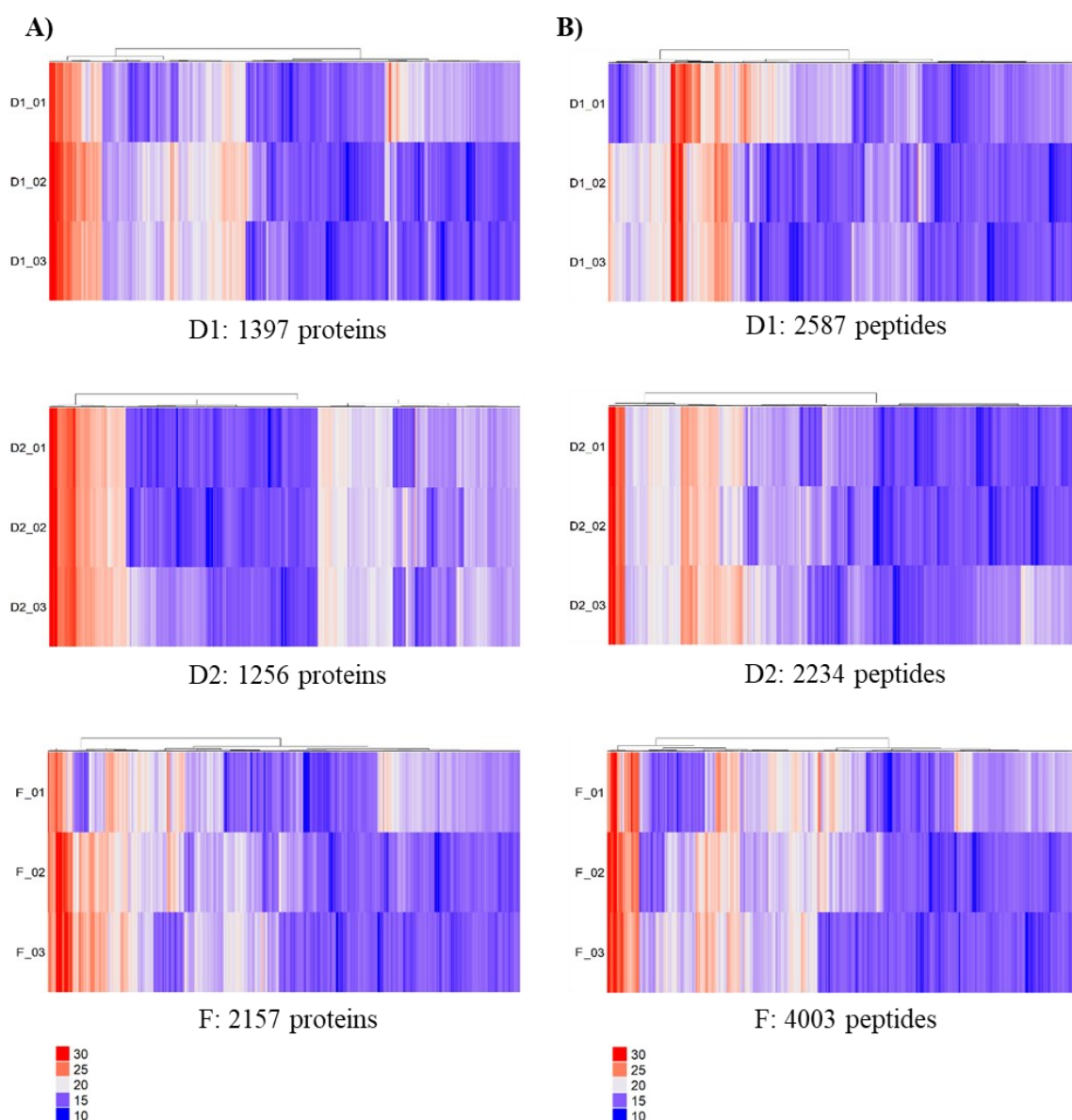
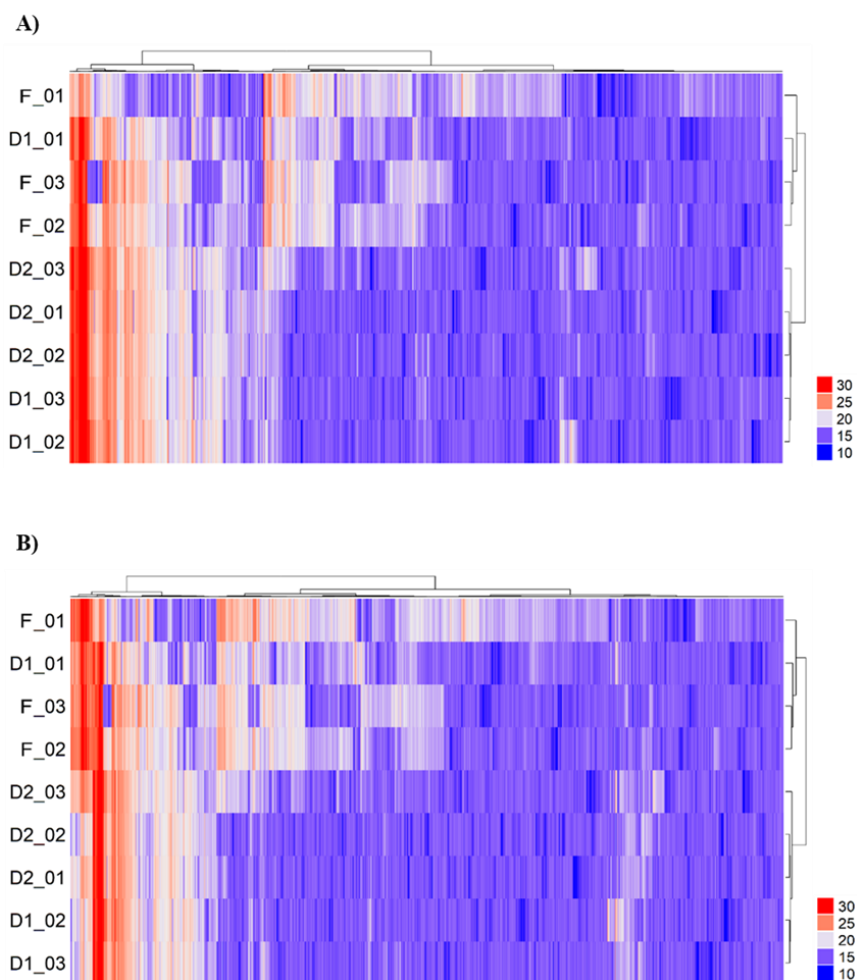


Fig. 2. Heatmaps of abundance. (A) proteins and (B) peptides in three samples of each person. In each person-heatmap there are proteins or peptides detected and quantified in at least one sample of that person

that a high similarity will be observed between the samples. Father's samples generally showed higher protein numbers than daughters' ones. This may be caused by the fact that the father is an adult with a higher lung volume and active lung surface; this, as a result, may lead to an increase in protein concentration at the same time. The number of proteins can have a big impact on such a situation. The same situation was observed in the heatmap at the peptide level (Figure 3B).

We have evaluated the distance of the abundances of the three samples for each person and each protein. The mean value of protein abundancies and the difference of the three values of abundancies from the mean were calculated for each protein separately. This was done for each sample (biological replicate) of all three individuals. The following four intervals were set:  $\pm$  sd (mean - sd; mean +

sd),  $\pm 2 \times$  sd (mean -  $2 \times$  sd; mean +  $2 \times$  sd),  $\pm 10\%$  ( $0.9 \times$  mean;  $1.1 \times$  mean), and  $\pm 25\%$  ( $0.75 \times$  mean;  $1.25 \times$  mean) and it was assessed whether the values for individual proteins were in or out of the interval (Figure 4). This analysis was also performed on peptide level. For peptide analysis, we used only fully quantified peptides, which were quantified in all technical and all biological samples (biological replicates) from the given individual. The number of this fully quantified peptides was 687 and 800 for daughters' samples (D1 and D2) and 1128 for father's samples (Figure 5). In both protein and peptide analyses, the difference of one standard deviation from the mean appeared to be too strict criterion because, for each of the proteins and peptides, at least one abundance value lay outside of the  $\pm$  sd interval (denoted as OUT). In the case of double the standard deviation ( $\pm 2 \times$  sd), all the proteins

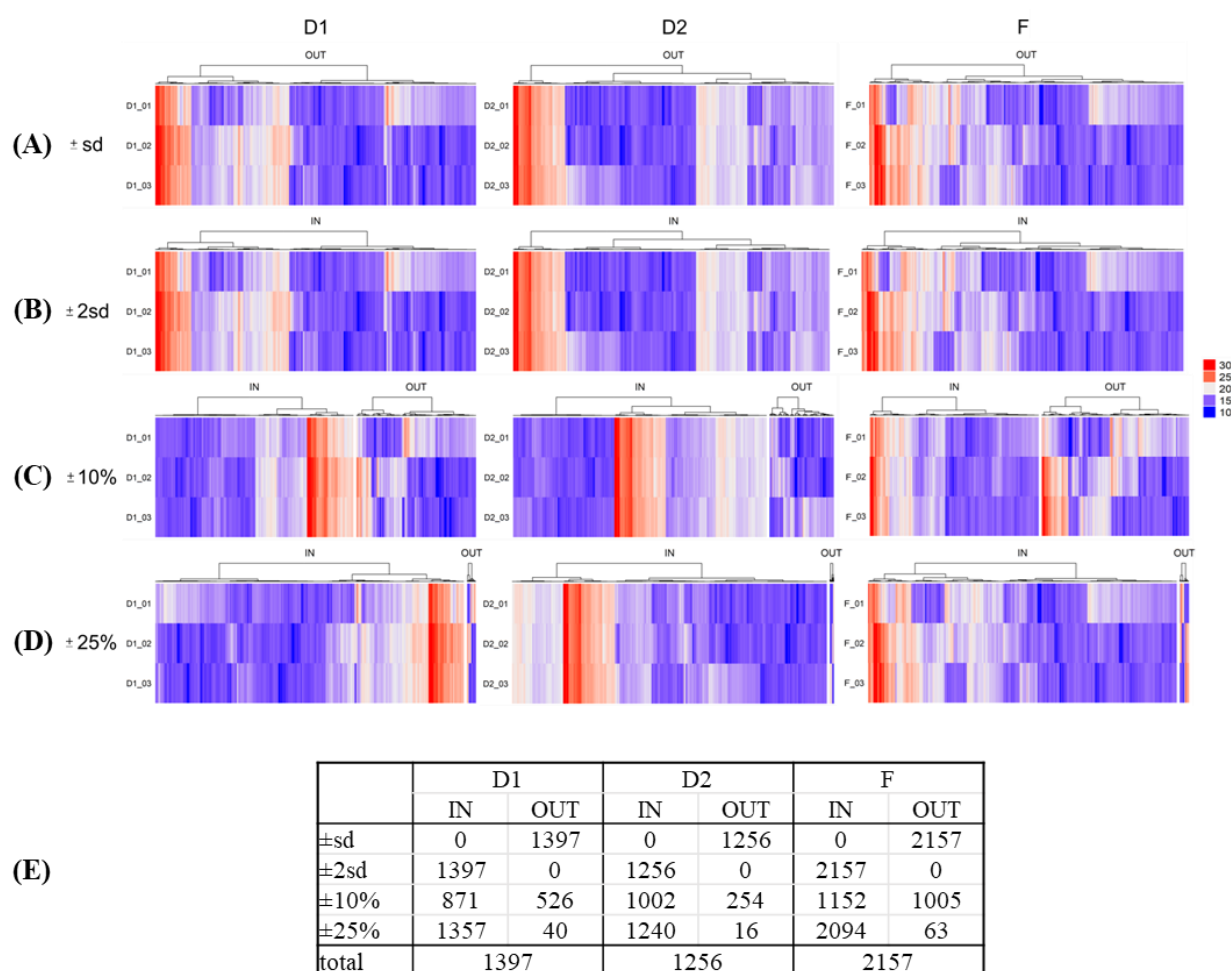


**Fig. 3. Heatmap expressing the abundance of all proteins and peptides across nine samples from individual family members.** (A) Heatmap with all 2263 quantified proteins and (B) heatmap with all 4189 quantified peptides. The presented proteins and peptides are those detected in at least one of the 9 samples. The columns represent individual proteins (or peptides), the rows represent the clustering of EBC samples of family members. Daughters' samples are clustered together except for the first replicate of D1, which contained the highest number of proteins and formed a cluster with father's samples

and peptides lay in this interval (denoted as IN). Analysis using the percent distance from the mean showed that 10% was also too strict criterion. However, more than half of the proteins and peptides belonged to this interval (IN), i.e. the distance of all the three abundances from the mean was less than 10%. When we analyzed proteins the distance of which from the mean was more than 25%, there were only a few tens of proteins or only 2 peptides lying outside this interval.

Then we tested whether these distances were due to variations in mass spectrometric measurements or rather to biological variability. We evaluated the maximum distance

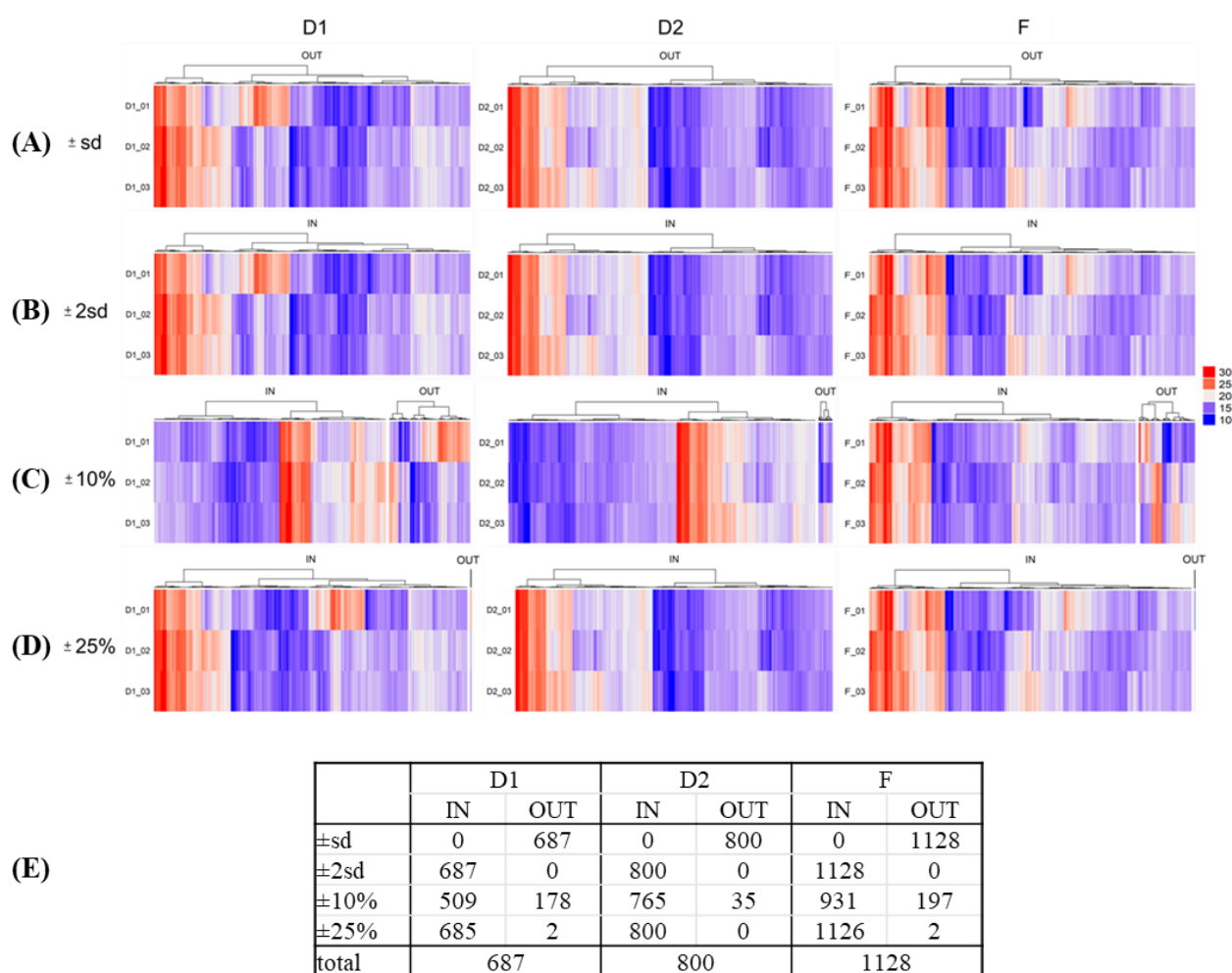
from the mean for technical replicates of the commercial bovine serum albumin standard (Pierce™ BSA Protein Digest Standard, LC/MS Grade, Thermo Fisher Scientific) which is routinely used to check the quality of mass spectrometry measurements by our group. The 500 pmol of lyophilized digested bovine serum albumin (BSA) was dissolved with 500  $\mu\text{L}$  of 0.1% formic acid in water, and then 15  $\mu\text{L}$  of this solution was further diluted with 285  $\mu\text{L}$  of 0.1% formic acid in water to obtain 50 fmol/ $\mu\text{L}$  BSA. An amount of 100 fmol BSA was injected for each MS measurement. For an analysis of the technical replicates, we calculated the mean of log<sub>2</sub> abundance of 3 technical



**Fig. 4. The distance of the abundances of the three samples analyzed for each person and each protein to evaluate the variability of the measured values.** Only proteins that have been quantified in at least one sample of individual person are included in the heatmaps. Thus, the total number of proteins is lower than reported before. Four types of intervals were determined and the proteins were divided into two groups: IN = the abundance of all three samples lies in the respective interval; OUT = the abundance of at least one of the three samples is outside this interval. Heatmaps (A – D) expressing proteins that lie in or out of intervals. Selected intervals were: **(A)**  $\pm$  sd (mean – sd; mean + sd), **(B)**  $\pm 2 \times$  sd (mean –  $2 \times$  sd; mean +  $2 \times$  sd), **(C)**  $\pm 10\%$  ( $0.9 \times$  mean;  $1.1 \times$  mean), **(D)**  $\pm 25\%$  ( $0.75 \times$  mean;  $1.25 \times$  mean). The table **(E)** lists the number of proteins which lie in and out of the respective intervals. Mean = mean of the abundances of the protein of the three samples of each person; sd = standard deviation of the mean

replicates for each peptide separately. The difference from the mean was calculated for all three values. The absolute value of the largest difference was chosen and expressed as a percentage of the mean value. This was performed for all fully quantified peptides of all replicates. For the BSA standard, all identified peptides were fully quantified. The analysis has shown that all maximum distances from the mean were less than 8.11%. Based on these results, it is clear that there is some deviation between the measurements of the technical replicates of the standard which is significantly lower than in the case of biological replicates.

However, the BSA standard is a commercial mixture of peptides with defined concentrations of the components, which includes only the peptides of this single protein, so that the deviation can be expected to be lower for the standard than for complex biological samples. To test this hypothesis, we examined the maximum distance from the mean in the case of technical replicates of a single person sample. This was done using the same calculations as for the BSA. Calculations were performed for all fully quantified peptides of all samples of all subjects and the values thus obtained were further compared. In the case of tech-



**Fig. 5. The distance of the abundancies of the three samples analyzed for each person and each peptide.** For each individual, peptides quantified in all three technical replicates and in all samples of the person were included, i.e. only peptides that were quantified in 9 replicates (technical and biological) of the individuals, no values imputed by Proteome Discoverer software were included. Four types of intervals were set and the peptides were divided into two groups: IN = the abundance of all three samples lies in the respective interval; OUT = the abundance of at least one of the three samples is outside this interval. Heatmaps (A – D) expressing peptides that lies in or out of intervals. Selected intervals were: (A)  $\pm sd$  (mean –  $sd$ ; mean +  $sd$ ), (B)  $\pm 2 \times sd$  (mean –  $2 \times sd$ ; mean +  $2 \times sd$ ), (C)  $\pm 10\%$  ( $0.9 \times$  mean;  $1.1 \times$  mean), (D)  $\pm 25\%$  ( $0.75 \times$  mean;  $1.25 \times$  mean). Table (E) lists the number of peptides that lie within and out the respective intervals. Mean = mean of the abundancies of the peptide in the three samples of each person;  $sd$  = standard deviation of the mean

Table II

The maximum distances from the mean abundance value of the three technical replicates (expressed in %) for individual samples

Sample	Max. distance (%) from the mean
D1_01	12.66
D1_02	19.69
D1_03	17.61
D2_01	13.13
D2_02	10.33
D2_03	8.31
F_01	9.04
F_02	13.39
F_03	13.44

nical replicates of human samples, the maximum distances from the mean received for each sample are given in Table II. All maximum distances from the mean were smaller than 19.69%. These results have shown that there is a greater variability between biological replicates of a given person than in the case of technical replicates of a single EBC sample. Thus, the difference in the protein or peptide abundancies which are distant less than 10% from the mean are much more due to the deviation of the mass spectrometry measurements than to the biological variability of the samples. These results also showed that, when mass spectrometry data are evaluated by label-free quantification, the change in protein abundancies below 20% of the mean value is due to the method variability and should not be considered as a change in protein expression.

## Conclusion

In this work, a method of protein analysis in exhaled breath condensate based on high resolution mass spectrometry is described. At the same time, it is shown that this analysis is sensitive and reproducible, and the variability of the measurements is acceptable for further interpretation of the results and able to identify hundreds of unique proteins. This method therefore represents a suitable tool for other biologically oriented research work, especially for the search for biomarkers of lung diseases.

*This work was supported by Ministry of Health of the Czech Republic, grant nr. 16-32302A. All rights reserved.*

## REFERENCES

- Horváth I. a 33 coauthors: *Eur. Respir. J.* 26, 523 (2005).
- Rosias P. P., Robroeks C. M., van de Kant K. D., Rijkers G. T., Zimmermann L. J., van Schayck C. P., Heynens J. W., Jöbsis Q., Dompeling E: *Pediatr Allergy Immunol.* 21, e235 (2010).
- Konstantinidi E. M., Lappas A. S., Tzortzi A. S., Behrakis P. K.: *Sci. World J.* 2015, 435160.
- Mutlu G. M., Garey K. W., Robbins R. A., Danziger L. H., Rubinstein I.: *Am. J. Respir. Crit. Care Med.* 164, 731 (2001).
- Griese M., Noss J., von Bredow C.: *Proteomics* 2, 690 (2002).
- Shahid S. K., Kharitonov S. A., Wilson N. M., Bush A., Barnes P. J.: *Am. J. Respir. Crit. Care Med.* 165, 1290 (2002).
- Leung T. F., Wong G. W., Ko F. W., Li C. Y., Yung E., Lam C. W., Fok T. F.: *Int. Arch. Allergy Immunol.* 137, 66 (2005).
- Gianazza E., Allegra L., Bucchioni E., Eberini I., Puglisi L., Blasi F., Terzano C., Wait R., Sirtori C. R.: *Am. J. Med.* 117, 51 (2004).
- Gessner C., Dihazi H., Brettschneider S., Hammerschmidt S., Kuhn H., Eschrich K., Keller T., Engelmann L., Sack U., Wirtz H.: *Respir. Med.* 102, 299 (2008).
- Fumagalli M., Dolcini L., Sala A., Stolk J., Fregonese L., Ferrari F., Viglio S., Luisetti M., Iadarola P.: *J. Proteomics* 71, 211 (2008).
- Bloemen K., Hooyberghs J., Desager K., Witters E., Schoeters G.: *Proteomics Clin. Appl.* 3, 498 (2009).
- Kurova V. S., Anaev E. C., Kononikhin A. S., Fedorchenko K. Y., Popov I. A., Kalupov T. L., Bratanov D. O., Nikolaev E. N., Varfolomeev S. D.: *Clin. Chem. Lab. Med.* 47, 706 (2009).
- Núñez-Naveira L., Mariñas-Pardo L. A., Montero-Martínez C.: *Lung* 197, 523 (2019).
- Muccilli V., Saletti R., Cunsolo V., Ho J., Gili E., Conte E., Sichili S., Vancheri C., Foti S.: *J. Pharm. Biomed. Anal.* 105, 134 (2015).
- Lacombe M., Marie-Desvergne C., Combes F., Kraut A., Bruley C., Vandenbrouck Y., Chamel Mossuz V., Couté Y., Brun V.: *J. Breath Res.* 12, 021001 (2018).
- Kononikhin A. S. a 13 spoluautorů: *J. Chromatogr. B* 1047, 97 (2017).
- Fedorchenko K. U. a 11 coauthors: *Mosc. Univ. Chem. Bull.* 71, 134 (2016).
- López-Sánchez L. M., Jurado-Gámez B., Feu-Collado N., Valverde A., Cañas A., Fernández-Rueda J. L., Aranda E., Rodríguez-Ariza A.: *Am. J. Physiol. Lung Cell. Mol. Physiol.* 313, L664 (2017).
- Sun C., Zhou T., Xie G., Fu S., Gao L., Liao J., Wu Y., Wang G.: *J. Proteomics* 206, 103392 (2019).
- Rappsilber J., Ishihama Y., Mann M.: *Anal. Chem.* 75, 663 (2003).
- <https://www.R-project.org/>, staženo 20. 12. 2018.
- Bredberg A., Gobom J., Almstrand A. C., Larsson P., Blennow K., Olin A. C., Mirgorodskaya E.: *Clin. Chem.* 58, 431 (2012).
- Fumagalli M. a 11 coauthors: *Int. J. Mol. Sci.* 13, 13894 (2012).

## Abstract

Exhaled breath condensate (EBC) is considered to be a rich source of biomarkers that can provide valuable information for early diagnosis of respiratory and systemic diseases. However, EBC consists mainly of water vapor and the concentration of analytes in its condensed sample is very low, close to the detection limit of sensitive methods such as mass spectrometry. Here, we have developed a mass spectrometry-based approach to characterize the EBC proteome. We compared biological and technical replicates of the measured samples, compared the protein profile between individual samples of healthy individuals and described the reproducibility between the three technical replicates of the sample taken. We have found that repeated measurements of samples are reproducible and can increase the number of proteins identified. The analyses led to the detection of 2797 unique proteins in all samples. Based on statistical analysis, it was shown that (i) the 10% difference from the mean is due more to the deviations of the mass spectrometry measurements than to the biological variability of the samples and (ii) the change in protein amount of 20% of the average in the case of EBC is due to the method variability and should not be considered as a significant change in protein expression using label-free mass spectrometry techniques.