

Kooperace genomických, transkriptomických a proteomických metod v detekci mutovaných proteinů

Cooperation of Genomic, Transcriptomics and Proteomic Methods in the Detection of Mutated Proteins

Zavadil Kokáš F., Faktor J., Vojtěšek B.

Regionální centrum aplikované molekulární onkologie, Masarykův onkologický ústav, Brno

Souhrn

Východiska: Současná protinádorová terapie se vyznačuje vysokou nespecifitou, a to z důvodu různorodé povahy nádorů, která významně snižuje účinnost léčby. Masivní rozvoj genomických, transkriptomických a proteomických metod v posledních desetiletích umožnil detailní charakterizaci nádorů na genomové, transkriptomové a proteomové úrovni a jejich vzájemná kombinace tak představuje potenciál, jak zvýšit efektivitu procesu detekce neoepitidů a následného navržení specifické terapie. Mezi v současné době široce používané genomické a transkriptomické metody patří zejména celogenomové, celotranskriptomové, příp. exomové sekvenování, která umožňují detekovat jednonukleotidové polymorfizmy. V případě proteomických metod, pokud je k dispozici peptidová knihovna, je možné detekovat mutované proteiny v biologickém vzorku. Nedílnou součástí kooperace těchto metod jsou softwary, které umožní interpretovat získané výsledky, jejich vizualizaci, příp. zprostředkují konverzi mezi datovými formáty často specifickými pro použítou metodu/přístroj. **Cíl:** Článek primárně popisuje bioinformatickou analýzu vzorků v rámci genomických a transkriptomických metod a jejich možné limitace a související problémy, které musí být zvaženy v průběhu analýzy, zejména týkající se kvality vstupních dat. V textu je rovněž věnována pozornost problémům vycházejícím ze zarovnání sekvencí na referenční genom. Součástí publikace je popis softwaru TransPEM, který byl vytvořen za účelem konverze výsledků analýzy jednonukleotidových polymorfizmů do podoby peptidové knihovny sekvencí využitelné k detekci neoepitidů pomocí proteomických metod. Nechybí ani stručný popis proteomických metod využívajících tuto knihovnu a představení jejich omezení.

Klíčová slova

genomika – transkriptomika – proteomika – bioinformatika – vývoj softwaru

Práce byla podpořena projektem MŠMT – NPU I – LO1413.

This work was supported by the project MEYS – NPS I – LO1413.

Autoři deklarují, že v souvislosti s předmětem studie nemají žádné komerční zájmy.

The authors declare they have no potential conflicts of interest concerning drugs, products, or services used in the study.

Redakční rada potvrzuje, že rukopis práce splnil ICMJE kritéria pro publikace zasílané do biomedicínských časopisů.

The Editorial Board declares that the manuscript met the ICMJE recommendation for biomedical papers.



Mgr. Bc. Filip Zavadil Kokáš, Ph.D.
Regionální centrum aplikované
molekulární onkologie
Masarykův onkologický ústav
Žlutý kopec 7
656 53 Brno
e-mail: filip.zavadil@mou.cz

Obdrženo/Submitted: 3. 6. 2019
Přijato/Accepted: 26. 8. 2019

doi: 10.14735/amko20193578

Summary

Background: Current anti-tumour therapy is characterised by high non-specificity due to the diverse nature of tumours, which can significantly reduce its efficiency. The massive development of genomic, transcriptomic, and proteomic methods has enabled the detailed characterisation of individual tumours at the genome, transcriptome and proteome levels. Whole-genome sequencing, whole-transcriptome sequencing and exome sequencing can be listed as examples of genomics and transcriptomics methods. Those methods are suitable for detecting single-nucleotide polymorphisms. In the case of proteomic methods, where a peptide library is available, it is possible to detect mutated proteins in a biological sample. Also important is software that interprets and visualises the results or facilitates conversion between data formats that are specific to the method. The combination of methods can in principle increase the likelihood of detecting new neoantigens and design-specific anti-tumour therapy. **Aim:** The article primarily describes the bioinformatics analysis of samples using the methods of genomics, transcriptomics and proteomics, and the possible problems which must be considered during the analysis. The article includes a description of TransPEM software designed to convert the results from the analysis of single nucleotide polymorphisms into a peptide library of sequences useful for the detection of neopeptides using proteomic methods. The publication is accompanied by a brief description of the proteomics methods using this peptide library and the summary of its limitations.

Key words

genomics – transcriptomics – proteomics – bioinformatics – software development

Úvod

Současná protinádorová terapie vychází zejména z histologického vyšetření nádoru, související analýzy symptomů pacienta a z jeho předchozí léčby. Terapie nádorů se vyznačují poměrně velkou nespecifitou a s tím související menší úspěšností a krátkou dobou přežití pacienta. Kvalitní charakteristika nádorů tedy představuje esenciální komponent pro zvolení, příp. vývoj terapie, která bude danému pacientovi tzv. ušita na míru.

Jednou z možností zkoumání nádorů představují tzv. omics metody [1], jejichž rozvoj v posledních desetiletích přinesl nové možnosti pro vývoj dokonalejší terapie. Z hlediska povahy zkoumané entity rozeznáváme metody genomické, epigenomické, transkriptomické, proteomické a metabolomické (schéma 1). Jejich vzájemná kooperace může být využita pro řadu výzkumných záležitostí, mezi něž bezesporu patří identifikace nových proteinových izoform [2] nebo detekce mutovaných proteinů. Prezentovaný článek je primárně věnován genomickým, transkriptomickým a proteomickým metodám a řeší jejich možnosti a limitace v souvislosti s bioinformatickou analýzou získaných dat.

Genomické a transkriptomické přístupy se vyznačují určitou podobností ve vlastním provedení, jelikož oba mohou být založeny na sekvenování nukleových kyselin. Široce používanými metodami jsou dnes techniky zprostřed-

kovávající genomové sekvenování, celotranskriptomové sekvenování a příp. exomové sekvenování [3]. Z pohledu přesnosti však existuje několik důležitých faktorů, které významně ovlivňují jejich výsledek a následnou interpretaci získaných dat.

Prvním důležitým aspektem všech metod je laboratorní příprava vzorků a následně experimentální provedení sekvenace. V současném výzkumu a diagnostice je nejrozšířenější sekvenáčnickou platformou Illumina [4], která se vyznačuje vysokou přesností, přijatelnou cenou a nenáročností na laboratorní vybavení. Její nevýhodou je ovšem relativně malá délka produkovaných readů ve srovnání s dříve využívanou Sangerovou metodou [5]. Tato skutečnost může výrazně komplikovat následnou bioinformatickou analýzu, zejména pokud se jedná o celotranskriptomové sekvenování [6]. Bioinformatická analýza získaných sekvenáčnických dat může být rozdělena do několika kroků, jejichž specifikace je závislá na povaze experimentu. Také volba vhodného softwaru v závislosti na vstupních datech a cílech experimentu představuje zásadní faktor pro korektní bioinformatickou analýzu. Ta je rozdělena na kontrolu kvality vstupních dat, mapování readů na referenci, vlastní analýzu v závislosti na požadavcích experimentu a konečně interpretaci výsledků [7].

V případě proteomiky je v současnosti dostupná široká škála hmotnostně spektrometrických metod, které je možné

rozdělit z hlediska přípravy vzorku na „top-down“ a „bottom-up“ přístupy. Metody „top-down“ zkoumají intaktní proteiny, zatímco metody „bottom-up“ analyzují peptidy získané proteomickým štěpením proteinů. „Bottom-up“ metody představují nejvýhodnější přístup pro kvalitativní i kvantitativní detekci proteinů ve vysoce komplexních biologických vzorcích, jako je nádorová tkáň, plazma nebo buněčná kultura, a proto se jim budeme dále věnovat. Tandemovou hmotnostní spektrometrií (liquid chromatography tandem-mass spectrometry – LC-MS/MS) je možné využít pro kvalitativní analýzu, tj. identifikaci proteinů, a lze ji úspěšně kombinovat s genomickými i transkriptomickými metodami. Analýzu LC-MS/MS dat však komplikuje dostupnost tzv. prohledávacích knihoven, které slouží jako reference pro správnou identifikaci zkoumaného proteinu. A právě genomická a transkriptomická analýza je schopna nabídnout nejkvalitnější prohledávací knihovny reprezentující zkoumaný vzorek.

Metody kvantitativní „bottom-up“ analýzy, jako např. monitorování vybraných reakcí [8], je možné kombinovat s proteogenomikou zejména za účelem validace již identifikovaných mutovaných proteinů/peptidů. Platforma kombinující všechny popsané metody má potenciál detailně charakterizovat nádorovou tkáň a popsat proteiny/peptidy, které doposud unikaly pozornosti (schéma 2).

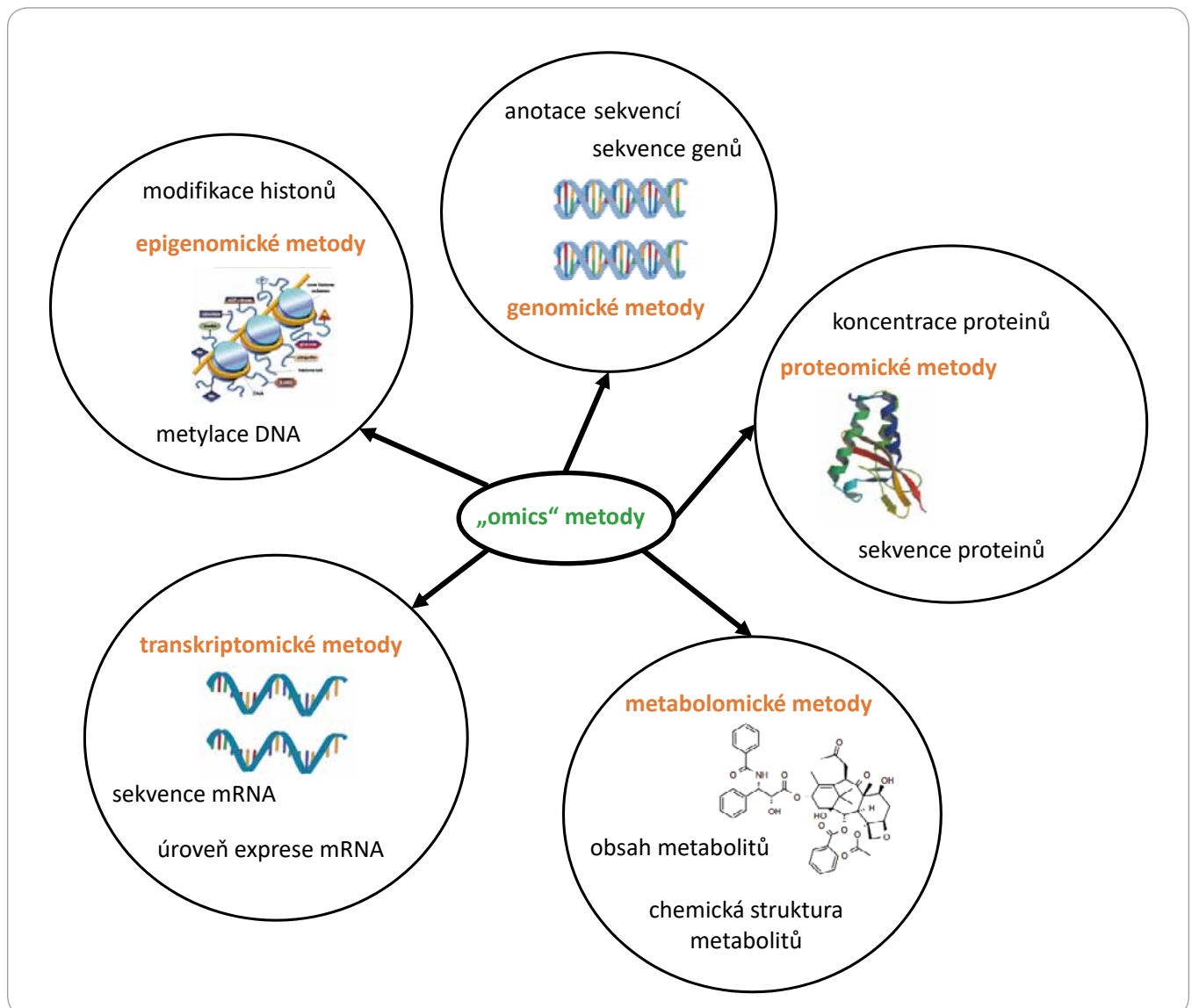


Schéma 1. Diagram zobrazující rozdělení „omics“ metod (zapsáno oranžovou barvou) a výsledků, které lze obdržet jejich prostřednictvím.

mRNA – mediátorová RNA

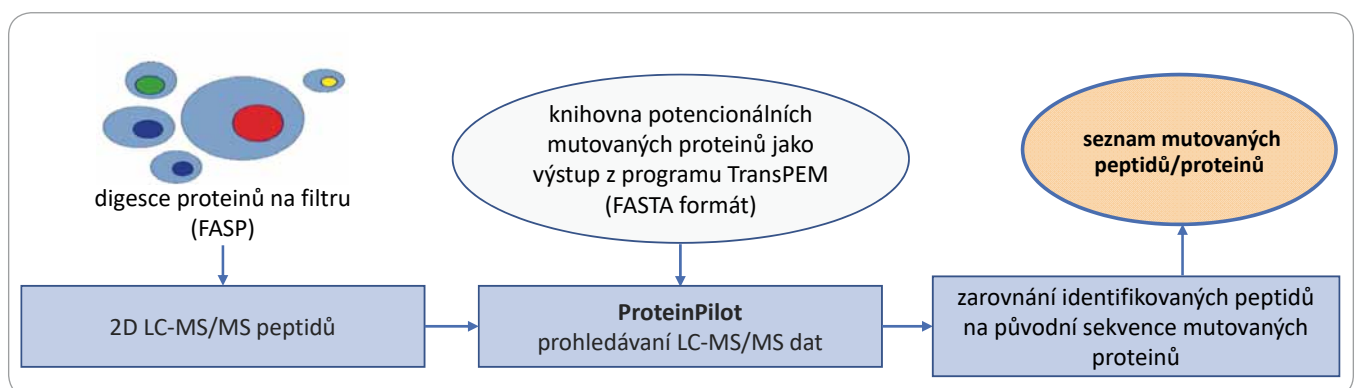


Schéma 2. Diagram zobrazující kombinaci softwaru TransPEM navazujícím na bioinformatickou analýzu jednonukleotidových polymorfizmů s proteomickou analýzou za účelem detekce mutovaných peptidů/proteinů.

LC-MS/MS – tandemová hmotnostní spektrometrie

Kvalita vstupních dat je základ

Prvním a zásadním krokem bioinformatické analýzy u sekvenace nukleových kyselin je kontrola kvality vstupních dat. Data jsou nejčastěji obdržena ve formě krátkých sekvencí, tzv. readů, přičemž kvalita každé báze v sekvenci může být odlišná [9]. U sekvenační platformy Illumina se například často setkáváme se sníženou kvalitou osekvenovaných bází na koncích readů. Za účelem posouzení kvality readů byl vyvinutý značný počet programů. Příkladem je FastQC [10], který je specifický pro platformu Illumina, nebo NGSQC (next-generation sequencing quality control) [11]. Mezi kritéria hodnocení kvality náleží zejména detekce obsahu sekvenačních adaptérů, zastoupení nukleotidů GC ve zkoumaných readech nebo sledování skóre kvality získaného ze vstupního souboru. Pokud se kvalita readů na základě těchto parametrů ukáže jako nevyhovující, je nezbytné provést opatření pro zlepšení jejich spolehlivosti, zejména ořezání readů, příp. odstranění vybraných readů. Operace ořezávání readů je široce používaná za účelem eliminace nekvalitních konců readů. K odstranění readů se přistupuje při výskytu kontaminace sekvenačními adaptéry. Obě operace mohou být samozřejmě kombinovány a v důsledku zlepšují kvalitu readů pro následné zarovnání na referenční sekvenci [2,12].

Volba softwaru pro mapování

Zarovnání readů na referenci (nejčastěji referenční genom) tvoří druhý krok v rámci bioinformatické analýzy sekvenačních dat. Volba programu určeného pro provedení typu sekvenování (exomové nebo celotranskriptomové) je dalším důležitým aspektem, který musí být zvažován zejména z důvodu odlišnosti mezi povahou dat poskytnutých sekvenováním exomů v porovnání se sekvenováním transkriptomu. Exomová sekvenování vycházejí z DNA jako vstupního materiálu, a tedy v průběhu mapování není třeba brát do úvahy alternativní sestřih primárního transkriptu RNA (pre-mRNA). Ve srovnání s celotranskriptomovým sekvenováním, v jehož případě ready obsahují sekvence transkriptu, které mohou být modifikované

alternativním sestřihem, je z hlediska výpočetní náročnosti vyhodnocení exomového sekvenování snadnější [13,14].

Mapování readů na referenci je rovněž ovlivněno několika negativními faktory, které musí být při vyhodnocování brány do úvahy. Nejčastěji uváděným negativním faktorem je tzv. problém multi-mapování, při němž je jeden read přiřazen na několik míst v použité referenci a který nastává zejména v oblastech genomu s vysokým obsahem repetitivních sekvencí. Tento problém může být do jisté míry odstraněn použitím technologie „paired-end“ sekvenování [15,16], při níž ready, které jsou v páru, poskytují přesnější zarovnání na referenci díky v součtu delší sekvenci. Velmi často se též objevuje problém krátkých readů, které jsou obecně náchylné k falešně pozitivnímu zarovnání ve srovnání s delšími ready. Řešení tohoto problému spočívá v použití jiné technologie sekvenování produkující delší ready, což ovšem výrazně zvyšuje cenu sekvenování [14,17].

Na základě povahy dat a zohlednění alternativního sestřihu se programy pro mapování readů na referenci dělí na dvě skupiny: „unspliced read aligners“ a „spliced read aligners“. První skupina sdružuje programy, které v rámci mapování readů nezahrnují alternativní sestřih, např. BWA [18] nebo Bowtie2 [19] a jsou primárně určeny pro zpracování dat z exomového nebo genomového sekvenování. Druhá skupina programů bere do úvahy alternativní sestřih transkriptů a je vhodná pro sekvenování transkriptomu (RNA-seq). Typickými a široce používanými softwary jsou Tophat2 [20], GSNAP [21] nebo QPALMA [22].

Detekce jednonukleotidových polymorfizmů

Analýza jednonukleotidových polymorfizmů (single nucleotide polymorphism – SNP) je nezbytná pro získání peptidové prohlédávací knihovny, která je nutná pro detekci potenciálních mutovaných peptidů proteomickými metodami. Hlavním problémem při detekci pravých SNP je jejich separace od artefaktů vzniklých v průběhu sekvenování, příp. v průběhu mapování readů na referenci [23]. Existuje několik volně do-

stupných softwarových balíčků široce používaných pro analýzu SNP, jako např. GATK [24], VarScan2 [25] nebo SAMtools [26], vzájemně se však ve schopnosti detekovat SNP liší [27]. Jejich výstupem je seznam predikovaných SNP. Na základě skutečnosti, jestli má SNP vliv na výsledný protein, rozlišujeme mutace nesynonymní (s vlivem na výslednou proteinovou sekvenci) a synonymní (bez vlivu na sekvenci aminokyselin, a to v důsledku degenerace genetického kódu). Pro případnou proteomickou analýzu jsou tedy důležité pouze mutace nesynonymní. Proteomické analýzy ovšem vyžadují SNP ve formátu peptidových sekvencí obsahujících tyto mutace, které tvoří referenční databázi nutnou pro následné prohlédávání MS dat, tzv. prohlédávací knihovny. Vyhledání specifické sekvence a změna příslušné aminokyseliny mohou být u dané sekvence provedeny manuálně na základě informací získaných z analýzy SNP, nicméně s narůstajícím počtem peptidových sekvencí a za předpokladu, že jedna mutace může vést ke vzniku několika výsledných proteinů (alternativní sestřih pre-mRNA), podstatným způsobem narůstá časová náročnost těchto operací a rovněž pravděpodobnost chyby způsobené lidským faktorem.

Za účelem automatizace popsaného procesu zejména z důvodu úspory času a minimalizace pravděpodobnosti vytvoření chybné sekvence byl vytvořen software TransPEM (software for Translation sequence into PEptide based on Mutation information) [28]. Představuje nový bioinformatický nástroj pro rychlou a robustní extrakci kódujících sekvencí z genomové sekvence a jejich translace do proteinů zohledňující výsledky analýzy SNP.

Vytvořené soubory obsahující všechny takto vytvořené proteinové sekvence mohou následně posloužit proteomické analýze jako prohlédávací knihovny. Pro svou činnost program vyžaduje celkem tři vstupní soubory. První ve formátu FASTA obsahující genomovou sekvenci, která musí být shodná s referencí, na kterou bylo provedeno mapování readů. Druhým je GTF soubor, který obsahuje informace o poloze genů a kódujících sekvencích v rámci genomu. Poslední

A. Neopeptid IIP^TVLMTEDIK (mutace A na P) se třemi produktovými ionty (y2, y3, y3 2+), potvrzujícími přítomnost mutované aminokyseliny ve zkoumané sekvenci.

EIF4G1_IIP^TVLMTEDIK A na P

Zbytek	b	b+2	y	y+2
I	114.0913	57.5493	1372.7756	686.8914
I	227.1754	1140913	1259.6916	630.3494
P	324.2282	162.6177	1146.6075	573.8074
T	425.2758	213.1416	1049.5547	525.2810
V	524.3443	262.6758	948.5070	474.7572
L	637.4283	319.2178	849.4386	425.2230
M	768.4688	384.7380	736.3546	368.6809
T	869.5165	435.2619	605.3141	303.1607
E	998.5591	499.7832	504.2664	252.6368
D	1113.5360	557.2967	375.2238	188.1155
I	1226.6701	613.8387	260.1969	130.6021
K	1354.7551	677.8862	147.1128	74.0600

B. Neopeptid VSGSPEQAVEENLSSYFLDR (mutace S na F) s osmi produktovými ionty (y4-1, y9 2+), potvrzujícími přítomnost mutované aminokyseliny.

TACC3_VSGSPEQAVEENLSSYFLDR S na F

Zbytek	b	b+2	y	y+2
V	100.0757	50.5415	2227.0513	1114.0293
S	187.1077	94.0575	2127.9829	1064.4951
G	244.1292	122.5682	2040.9509	1020.9791
S	331.1612	166.0842	1983.9294	992.4684
P	428.2140	214.6106	1896.9874	948.9523
E	557.2566	279.1319	1799.8446	900.4260
Q	685.3151	343.1612	1670.8020	835.9047
A	756.3523	378.6798	1542.7435	771.8754
V	855.4207	428.2140	1471.7064	736.3568
E	984.4633	492.7353	1372.6379	686.8226
E	1113.5059	557.2566	1234.5953	622.3013
N	1227.5438	614.2780	1114.5527	557.7800
L	1340.6329	670.8201	1000.5098	500.7585
S	1427.6649	714.3361	887.4258	444.2165
S	1514.6969	757.8521	800.3937	400.7005
Y	1677.7602	839.3838	713.3617	357.1845
F	1824.8286	912.9180	550.2984	275.6528
L	1937.9127	969.4600	403.2300	202.1186
D	2025.9397	1026.9735	290.1459	145.5766
R	2209.0408	1105.0240	175.1190	88.0631

Obr. 1. Seznam produktových iontů reprezentujících dva vybrané neopeptidy identifikované prohlédávacím algoritmem ProteinPilot 4.5.0.0 s vysokou peptidovou konfídencí (peptidová konfidence > 99 %). Zelenou barvou jsou vyznačeny produktové ionty spolehlivě identifikované v LC-MS/MS fragmentačním spektru. Červeně jsou podtrženy produktové ionty, které potvrzují aminokyselinovou záměnu a zahrnují ve své aminokyselinové sekvenci mutovanou aminokyselinu. Z obrázku je patrné, že peptidová konfidence nereflektuje pravděpodobnost přítomnosti mutace v sekvenci neopeptidů a že fragmentační spektra neopeptidů je nutné po identifikaci zkontrolovat.

soubor obsahuje výsledky z analýzy SNP poskytnuté např. programem VarScan2 [25] v tabulkovém formátu obsahující pro každé SNP chromozom, pozici v rámci chromozomu, referenční a mutantní alelu. Software TransPEM [28] poskytuje mutované a původní sekvence všech možných transkriptů daného genu. Ve srovnání s manuálním přístupem přináší významnou časovou úsporu a rovněž minimalizuje možné chyby prostřednictvím kontroly vstupních souborů a systému hodnocení a hlášení chyb vzniklých v průběhu získávání mutovaných sekvencí proteinů. Jelikož jsou generovány všechny možné peptidové

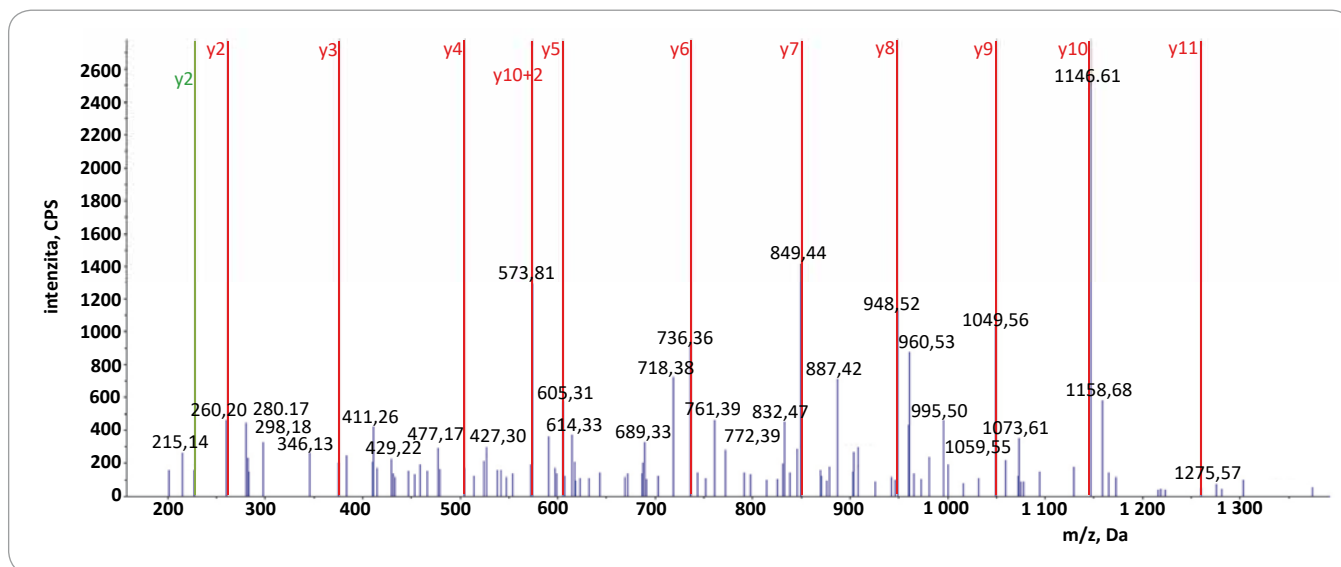
sekvence pro daný gen, je rovněž navýšena pravděpodobnost detekce neopeptidů, jelikož program tímto postihuje všechny možné podoby výsledného genového produktu.

Proteomická analýza mutovaných proteinů

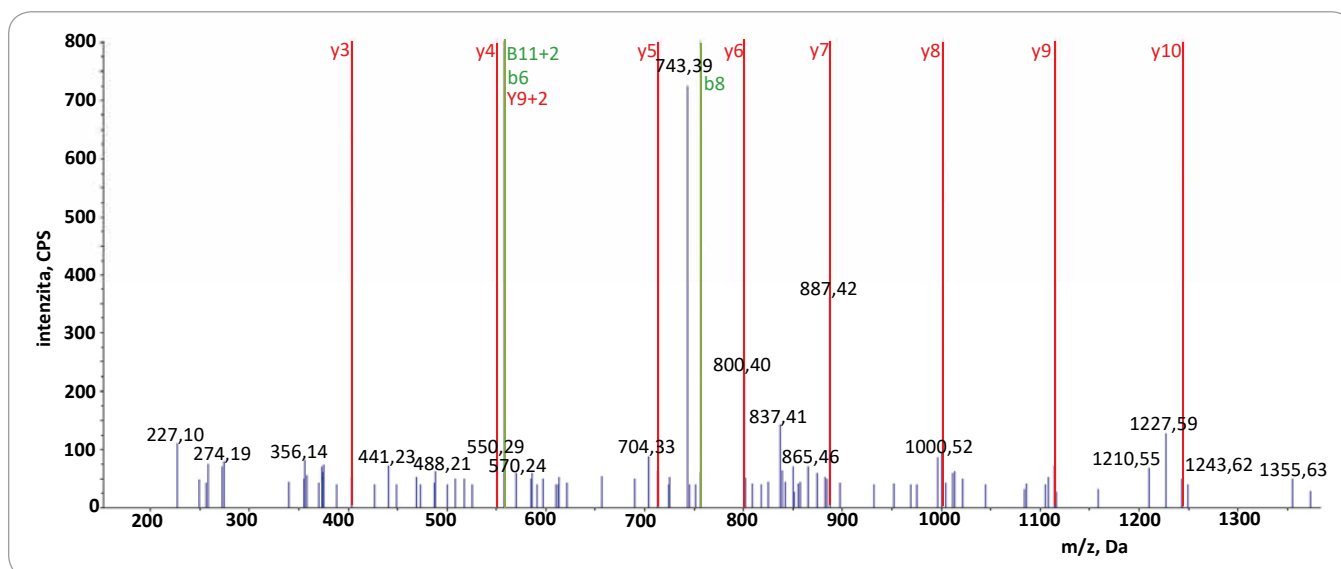
Tandemová hmotnostní spektrometrie nabízí informace o genové expresi na úrovni proteinů. Získané výsledky poskytují vyšší výpovědní hodnotu v porovnání s genomickými metodami, jelikož přímo popisují efektorové molekuly. Mezi největší úskalí LC-MS/MS patří různorodá ionizovatelnost peptidů, vysoký

dynamický koncentrační rozsah proteinů ve vzorku a nutnost mít k dispozici prohlédávací knihovnu.

Výsledky LC-MS/MS analýzy jsou obvykle zatíženy nedostatečným pokrytím sekvence a nízkou efektivitou přiřazení změřených spekter k teoretickým spektrům obsaženým v prohlédávací knihovně. Ukázalo se, že až 10 % vysoce kvalitních a 65 % MS/MS spekter s nižší kvalitou není vůbec přiřazeno k sekvenci proteinu/peptidu, ze kterého pocházejí [29]. Genomické/transkriptomické metody umožňují získat prohlédávací knihovnu šitou na míru pro danou LC-MS/MS analýzu. Přípravou těchto pro-



Obr. 2. Fragmentační spektrum produktových iontů reprezentující peptid IPTVLMTEDIK z proteinu EIF4G1



Obr. 3. Fragmentační spektrum produktových iontů reprezentující peptid VSGSPEQAVEENLSSYFLDR z proteinu TACC3.

hledávacích knihoven (jak bylo popsáno výše) je možné zvýšit efektivitu identifikace proteinů pomocí LC-MS/MS, a dokonce je možné identifikovat i peptidy/proteiny s mutovanou sekvencí.

Významně z toho benefitují LC-MS/MS přístupy zaměřené na identifikaci tzv. neoepitidů, které pokrývají mutovanou část sekvence proteinu. Software na prohledávání LC-MS/MS dat (např. ProteinPilot, Proteome Discoverer) v principu použije prohledávací mutantní knihovnu založenou na genomických/transkriptomických datech, kterou *in silico* naštěpí proteázou a vzniklé pep-

tidy posttranslačně modifikuje na základě pravidel definovaných uživatelem. Získaná data LC-MS/MS analýzy jsou následně importována a zpracována. Po dobu jejich zpracování dochází k přiřazení změřených spekter ke spektrům teoretickým, a pokud je mezi spektry nalezena dostatečná shoda, je peptid považován za identifikovaný. Identifikované peptidy jsou následně přiřazeny k příslušným sekvencím proteinů. Výsledkem analýzy dat LC-MS/MS je seznam identifikovaných proteinů a peptidů. V seznamu identifikací je potřeba vyhledat neoepitidy. V následujícím kroku je dále nutné

zjistit, jestli jsou neoepitidy unikátní, a tedy jestli jejich sekvence není sdílena s jinými proteiny. K tomuto účelu je využívána funkce peptid/protein BLAST. Za klíčovou součást nově vzniklé proteogenomické platformy považujeme zavedení verifikačního kroku, kterým ověřujeme, zda prohledávací algoritmus skutečně identifikoval neoepitid s aminokyselinovou záměnou. Problémem však zůstává, že současné prohledávací algoritmy neposkytují funkci zohledňující množství a kvalitu produktových iontů, které přímo obsahují mutované/ou aminokyselinu/ou v jejich sekvenci. Pro lepší pochopení pro-

blému uvádíme obr. 1, která reprezentuje dva neopeptidy identifikované se stejnou peptidovou konfidencí s pomocí prohledávacího algoritmu ProteinPilot 4.5.0.0 (peptidová konfidence = 99 %). V obr. 1 jsou zeleně vyznačeny m/z produktových iontů, které byly spolehlivě identifikovány v MS/MS spektrech obou neopeptidů (obr. 2 a 3). Z hlediska potvrzení přítomnosti mutované aminokyseliny v sekvenci ovšem neopeptidy nejsou rovnocenné a jejich neopeptidová konfidence není identická. Je zřejmé, že mutace (A na P) v neopeptidu ILPTVLMTEEDIK potvrzuje jen produktové ionty z y a y 2+ série (y₂, y₃, y₃ 2+, podtrženy červenou barvou; obr. 1A), které jsou navíc velmi nespecifické vzhledem k nízkému m/z. K opačné situaci dochází u neopeptidu VSGSPEQAVEENLSSYFLDR (S na F; obr. 1B), u něhož vidíme až osm produktových iontů ze série y a y 2+ (y-1, y 2+, podtrženy červenou barvou) potvrzující mutaci. Z obr. 1A a B je zřejmé, že pravděpodobnost správné identifikace mutací je odlišná u obou neopeptidů a nedá se ztotožňovat s peptidovou konfidencí. Z tohoto důvodu doporučujeme při identifikaci neopeptidů důkladně kontrolovat fragmentační spektra a zároveň spatřujeme potenciál v tvorbě softwaru, který by určil neopeptidovou konfidenci a zautomatizoval proces jejího určení.

Závěr

Detailní popis nádorů zprostředkovaný vzájemnou kooperací genomických, transkriptomických a proteomických metod vede k účinnější detekci neopeptidů a potenciálně k výběru účinnější protinádorové terapie. Software TransPEM [28] zprostředkovává konverzi výsledků získaných prostřednictvím SNP analýzy do tvaru prohledávací peptidové knihovny, kterou lze následně využít proteomickými metodami k detekci neopeptidů. Software je ovládán prostřednictvím příkazové řádky a v současnosti neobsahuje grafické uživatelské rozhraní. Podrobnější popis softwaru je uveden v manuálu, který je ke stažení společně se zdrojovým kódem softwaru na webu Recamo.cz [28]. Popisováním kooperací genomických, transkriptomických a proteomických metod lze identifikovat imunogenní neoantigeny

dále využitelné v rámci přípravy protinádorových vakcín. Je ovšem nutné poukázat na skutečnost, že biologická funkce identifikovaných neoantigenů musí být dále objasněna jinými molekulárně biologickými metodami. Identifikovaný neoantigen s potenciálem pro tvorbu neoantigenových protinádorových vakcín by měl být rozpoznáván CD8 T lymfocyty. Rovněž navození efektivní imunitní odpovědi cílené na tumor specifické neoantigeny je komplikované a vyžaduje zapojení multioborového týmu vědců. Dalším důležitým aspektem je, že k navození tumor supresivního účinku neoantigenové vakcinace často dochází až po kombinaci s jinou formou imunoterapie, jako např. anti-PD1 anti-CTLA4 biologická léčba [30,31]. Úspěšná proteogenomická identifikace mutovaného peptidu, resp. neoantigenu, tedy představuje pouze velmi malý krok k vývoji specifické terapie.

Literatura

- Vucic EA, Thu KL, Robinson K et al. Translating cancer "omics" to improved outcomes. *Genome Res* 2012; 22(2): 188–195. doi: 10.1101/gr.124354.111.
- Conesa A, Madrigal P, Tarazona S et al. A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016; 17: 13. doi: 10.1186/s13059-016-0881-8.
- Morey M, Fernández-Marmiesse A, Castiñeiras D et al. A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* 2013; 110(1–2): 3–24. doi: 10.1016/j.jmgme.2013.04.024.
- Liu L, Li Y, Li S et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012; 2012: 251364. doi: 10.1155/2012/251364.
- Sanger F, Air GM, Barrell BG et al. Nucleotide sequence of bacteriophage φX174 DNA. *Nature* 1977; 265(5596): 687–695. doi: 10.1038/265687a0.
- Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics* 2016; 107(1): 1–8. doi: 10.1016/j.jygeno.2015.11.003.
- Wolf JBW. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol Ecol Resour* 2013; 13(4): 559–572. doi: 10.1111/1755-0998.12109.
- Faktor J, Michalova E, Bouchal P. p-SRM, SWATH a HRM – cílené proteomické přístupy na hmotnostním spektrometru TripleTOF 5600+ a jejich aplikace v onkologickém výzkumu. *Klin Onkol* 2014; 27 (Suppl 1): 110–115. doi: 10.14735/amko201415110.
- Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; 8(3): 186–194.
- Babraham Bioinformatics. FastQC: a quality control tool for high throughput sequence data. [online]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Dai M, Thompson RC, Maher C et al. NGSQC: cross-platform quality analysis pipeline for deep sequencing data. *BMC Genomics* 2010; 11 (Suppl 4): S7. doi: 10.1186/1471-2164-11-S4-S7.
- Yang IS, Kim S. Analysis of whole transcriptome sequencing data: workflow and software. *Genomics Inform* 2015; 13(4): 119–125. doi: 10.5808/GI.2015.13.4.119.

- Garber M, Grabherr MG, Guttman M et al. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods* 2011; 8(6): 469–477. doi: 10.1038/nmeth.1613.
- Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet* 2011; 12(10): 671–682. doi: 10.1038/nrg3068.
- Campbell PJ, Stephens PJ, Pleasance ED et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet* 2008; 40(6): 722–729. doi: 10.1038/ng.128.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; 10(1): 57–63. doi: 10.1038/nrg2484.
- Góngora-Castillo E, Buell CR. Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Nat Prod Rep* 2013; 30(4): 490–500. doi: 10.1039/c3np20099j.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25(14): 1754–1760. doi: 10.1093/bioinformatics/btp324.
- Langmead B, Trapnell C, Pop M et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; 10(3): R25. doi: 10.1186/gb-2009-10-3-r25.
- Kim D, Pertea G, Trapnell C et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013; 14(4): R36. doi: 10.1186/gb-2013-14-4-r36.
- Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 2010; 26(7): 873–881. doi: 10.1093/bioinformatics/btq057.
- De Bona F, Ossowski S, Schneeberger K et al. Optimal spliced alignments of short sequence reads. *Bioinformatics* 2008; 24(16): 174–180. doi: 10.1093/bioinformatics/btn300.
- De Wit P, Pespeni MH, Ladner JT et al. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol Ecol Resour* 2012; 12(6): 1058–1067. doi: 10.1111/1755-0998.12003.
- McKenna A, Hanna M, Banks E et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20(9): 1297–1303. doi: 10.1101/gr.107524.110.
- Koboldt D, Zhang Q, Larson D et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; 22(3): 568–576. doi: 10.1101/gr.129684.111.
- Li H, Handsaker B, Wysoker A et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; 25(16): 2078–2079. doi: 10.1142/S0219720015500250.
- Yu X, Sun S. Comparing a few SNP calling algorithms using low-coverage sequencing data. *BMC Bioinformatics* 2013; 14: 274. doi: 10.1186/1471-2105-14-274.
- Recamo.cz. TransPEM (Software for Translation sequence into peptide based on mutation information). [online]. Dostupné z: <https://www.recamo.cz/cz/software/transpem/>.
- Ning K, Fermin D, Nesvizhskii A. Computational analysis of unassigned high-quality MS/MS spectra in proteomic data sets. *Proteomics* 2010; 10(14): 2712–2718. doi: 10.1002/pmic.200900473.
- Sahin U, Derhovanessian E, Miller M et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 2017; 547(7662): 222–226. doi: 10.1038/nature23003.
- Ott PA, Hu Z, Keskin DB et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017; 547(7662): 217–221. doi: 10.1038/nature22991.

Redakce časopisu **Klinická onkologie** vypisuje

SOUTĚŽ O NEJLEPŠÍ PRÁCI

v kategoriích:

Původní práce
Přehled
Kazuistika

Podmínky soutěže:

1. Do soutěže budou automaticky zařazeny všechny práce publikované v řádných číslech v roce 2019.
2. Práce budou hodnoceny členy redakční rady.
3. Hlavními kritérii hodnocení budou odborná úroveň, originalita a přínos zveřejněných údajů.
4. Výsledky soutěže budou vyhlášeny v časopise Klinická onkologie 1/2020.

Nejlepší práce v každé kategorii bude oceněna částkou 10 000 Kč.

Instrukce pro autory naleznete na internetových stránkách České onkologické společnosti ČLS JEP www.linkos.cz nebo na stránkách www.klinickaonkologie.cz. Dotazy můžete zasílat na adresu klinickaonkologie@mou.cz a své příspěvky vkládat do redakčního systému časopisu Klinická onkologie <http://redakce.ambitmedia.cz/ko>.

Proč publikovat v časopise Klinická onkologie?

Vaše práce budou dohledatelné ve 4 renomovaných světových bibliografických databázích MEDLINE/PubMed, EMBASE/Excerpta Medica, SCOPUS, Index Copernicus a tuzemské databázi Bibliographia medica chechoslovaca.

Vaše práce budou uznávány při hodnocení grantů, pro obhajoby doktorského studia a pro habilitační a profesorské řízení.

Vaše práce budou čteny. Časopis Klinická onkologie patří k nejčtenějším onkologickým časopisům!