

Technologie sekvenování nové generace: celogenomové, celoexomové a cílené – hotspot – sekvenování

Základní principy sekvenování nové generace

Sekvenování nové generace (NGS – *next generation sequencing*) vykazuje oproti tradičním metodám používaným pro detekci genomových variant řadu výhod. Tradiční technologie jako Sangerovo sekvenování, sekvenčně specifická PCR nebo fluorescenční *in situ* hybridizace (FISH) jsou vždy cílené pouze na konkrétní gen nebo variantu. V současné době, kdy je v onkologii často nezbytné paralelní vyšetřování většího množství genů/variant, kladou tyto konvenční metody vyšší nároky na množství biologického materiálu, často opakované biopsie a také delší čas do poskytnutí výsledků, protože jednotlivá vyšetření musejí být prováděna vždy nezávisle na sobě. Oproti tomu NGS umožňuje paralelní testování většího množství genů, případně celého exomu či genomu, a to i od většího množství pacientů v jednom experimentu. Tento přístup je spojen s nižšími nároky na množství vstupního biologického materiálu, a především s redukcí času do poskytnutí výsledků vyšetření, což má v onkologii zásadní význam.

NGS je založeno na masivním paralelním sekvenování mnoha fragmentů DNA. Tato technologie někdy bývá označována také jako vysokokapacitní sekvenování, hluboké sekvenování nebo sekvenování druhé generace. V současnosti již existuje řada technologií (nejčastěji používané *Illumina* a *Ion Torrent*) umožňujících diagnostickou aplikaci NGS. Napříč všemi platformami jsou zachovány základní principy NGS. **Technologie NGS umožňuje sekvenovat DNA**, a i když 3 základní typy genomových variant – jednonukleotidové záměny (SNV), variabilita v počtu kopií genů (CNV) a krátké inserce

a delece (*indels*) – lze identifikovat na úrovni nádorové DNA, neméně důležité **fúzní geny** detekujeme pomocí **analýzy nádorové RNA**. Na jednu kopii fúzního genu v nádorovém genomu totiž připadají řádově stovky až tisíce kopií jeho RNA transkriptu. Detekci fúzních genů na úrovni RNA proto ve srovnání s DNA dosahujeme mnohonásobně vyšší citlivosti. Před vlastní sekvenační reakcí je ovšem nezbytné provést **přepis (reverzní transkripci) nádorové RNA do cDNA**.

Prvním krokem NGS je příprava vzorku DNA pro účely sekvenace neboli příprava tzv. sekvenační knihovny. V tomto kroku je vždy zahrnuta izolace DNA ze vzorku nádoru (případně přepis nádorové RNA do cDNA), fragmentace DNA, navázání tzv. adaptorových sekvencí na konce fragmentů DNA, což umožní jejich globální namnožení (amplifikaci), a dále jejich fixaci na solidní povrch.

Následně, **ve druhém kroku**, dochází k vlastní **sekvenaci syntézou**. K imobilizovaným fragmentům DNA jsou komplementárně připojovány jednotlivé nukleotidy (ACTG), což je spojeno s emisí vždy jiného fyzikálního signálu zachycovaného detektorem.

Takto získané sekvence jednotlivých fragmentů jsou **ve třetím kroku bioinformaticky analyzovány**. V první fázi jsou získané sekvence **přiřazovány (*alignment*) k referenčnímu genomu** či sekvenci považované za sekvenci referenční. Následně probíhá analýza, jejímž cílem je **nalézt varianty v sekvenovaném vzorku DNA (*variant calling*)**, poté **anotace nalezených variant** a dále nejsložitější krok bioinformatické analýzy, kterým je

filtrování nalezených variant. Zde jsou postupně eliminovány varianty podle různých kritérií, v první fázi automaticky (například dle schopnosti varianty měnit strukturu kódovaného proteinu, tzv. nesynonymní varianty) a následně manuálně (například filtrování podle klinické evidence). Cílem je zúžit seznam nalezených variant na takové, které jsou biologicky relevantní v kontextu daného nádorového onemocnění (*drive* mutace) a optimálně také klinicky relevantní (*actionable* mutace) s jasným dopadem pro terapeutické plánování. *Workflow* technologie sekvenování nové generace je shrnuto na *obr. 1*.

Představíme-li si konkrétní pozici v lidském genomu, resp. genu, např. *BRAF* V600E (v sekvenaci genu pro *BRAF* to odpovídá pozici a záměně c. 1799T > A), počet fragmentů DNA získaných z nádorového vzorku topologicky pokrývajících tuto pozici určuje, kolikrát bude daná pozice genu v sekvenační reakci tzv. přečtena, tedy jaká bude v dané pozici **hloubka čtení** (*sequencing depth*), resp. jaké zde bude pokrytí neboli **coverage**.

Představme si, že máme biologický vzorek tvořený tisícem buněk, ze kterých je 50 % nádorových. Z těchto 500 nádorových buněk 400 buněk vykazuje heterozygotní stav pro aktivační mutaci *BRAF* V600E – to je 40 % všech buněk obsahuje mutaci v jedné alele, tzn. 20 % DNA ve vzorku nese mutaci. Pokud danou pozici při sekvenaci přečteme s hloubkou čtení 20X, teoreticky 4 z 20 čtení zachytí mutovanou alelu A, 16 z 20 čtení potom nemutovanou alelu T.

Základní aplikace sekvenování nové generace

Za dvě základní aplikace NGS lze považovat **sekvenaci genomové DNA nebo jejich částí** s cílem identifikovat SNV, CNV a *indels* a dále **sekvenaci transkriptomu, tzv. RNAseq**; technicky se jedná o sekvenaci DNA získané přepisem nádorové RNA s cílem identifikovat

přítomnost fúzních genů. Z hlediska genomové DNA lze potom sekvenovat zárodečnou DNA (germinální varianty), získanou obvykle z periferních leukocytů či buňkami buňkami stěru, zpravidla pro účely klinické genetiky, případně DNA nádorovou (somatické varianty), získanou izolací ze vzorku nádorové tkáně pro účely terapeutického plánování.

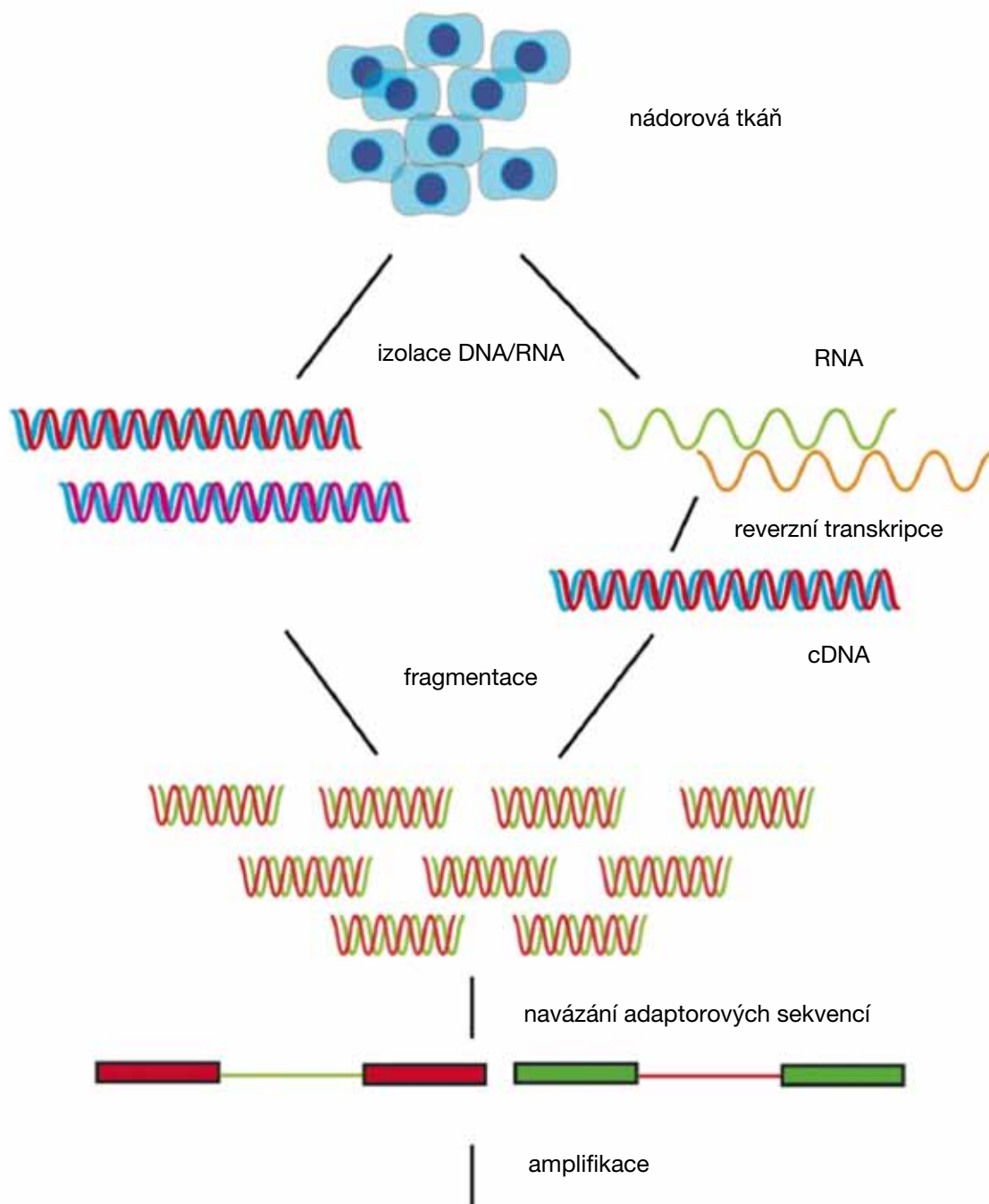
Z hlediska tzv. **šíře pokrytí** (*breadth of coverage*) neboli procenta genomu pokrytého sekvenační reakcí dělíme základní aplikace NGS na **celogenomové sekvenování (WGS – whole genome sequencing)**, **celoexomové sekvenování (WES – whole exome sequencing)** a **cílené (hotspot) sekvenování, založené na použití tzv. sekvenačních panelů (srovnání v tabulce)**.

Celogenomové sekvenování umožňuje detekovat všechny typy variant napříč celým genomem, v kódujících (tzv. exom) i nekódujících oblastech, a tak umožňuje identifikovat i nové, dříve nepopsané nádorové varianty. Srovnání zárodečného genomu s genomem nádorovým poskytne skutečně komplexní pohled na specifické změny přítomné v nádorovém genomu. Nicméně množství dat získané pomocí WGS klade vysoké nároky na jejich interpretaci, a to především v případě náhodných potenciálně rizikových, ale těžce interpretovatelných nálezu.

Celoexomové sekvenování svojí šíří pokrývá pouze kódující oblasti genomu, tedy přibližně 1% z jeho celkové sekvence. Nicméně přibližně 85% patogenních variant se nachází právě v kódujících oblastech, což činí výstupy WES oproti WGS jednodušší z hlediska interpretace. Nižší šíře pokrytí WES umožňuje oproti WGS větší hloubku čtení, a tak i vyšší senzitivitu pro záchyt variant v heterogenním vzorku nádorové tkáně. Z hlediska klinicky významného informačního obsahu WES představuje velice dobrou alternativu k WGS.

V současnosti je pro účely precizní onkologie nejvíce používáno cílené (*hotspot*) sekvenování

1. příprava sekvenanční knihovny



2. sekvenování

3. bioinformatická analýza

(přiřazování referenčního genomu, hledání a filtrování variant)

Aplikace NGS	Popis	Výhody	Nevýhody
Celogenomové sekvenování (WGS)	stanovení kompletní sekvence genomové DNA	<ul style="list-style-type: none"> • Poskytuje sekvence kódujících i nekódujících oblastí genomu • Představuje nejkomplexnější způsob charakterizace genomové DNA • Velký význam z hlediska translačního onkologického výzkumu • Umožňuje výpočet mutační nálože (TMB) 	<ul style="list-style-type: none"> • Velké množství získaných dat klade vysoké nároky na jejich interpretaci, a to především v případě náhodných potenciálně rizikových nálezů • Finanční a časová náročnost
Celoexomové sekvenování (WES)	stanovení sekvence přibližně 1% genomové DNA představující kódující oblasti	<ul style="list-style-type: none"> • Většina patogenních variant (85%) leží v kódujících částech genomu • Funkční dopady těchto variant jsou lépe interpretačně uchopitelné • Oproti WGS umožňuje větší hloubku čtení • Umožňuje výpočet mutační nálože (TMB) 	<ul style="list-style-type: none"> • Neumožňuje identifikaci variant v nekódujících oblastech genomu • Problém s náhodnými potenciálně rizikovými nálezy
Cílené (<i>hotspot</i>) sekvenování (sekvenační panely)	stanovení sekvence vybraných genů, případně vybraných genomových oblastí	<ul style="list-style-type: none"> • Méně finančně a časově náročné • Cílení na konkrétní geny a varianty (<i>hotspoty</i>) umožňuje jejich snadnou klinickou interpretaci • Malé riziko náhodných nálezů • Vysoká senzitivita • Metodika umožňuje snadnější optimalizaci a standardizaci ve srovnání s WGS a WES • Umožňuje výpočet mutační nálože TMB (při počtu genů na panelu > 200) 	<ul style="list-style-type: none"> • Největší limitací je nízká šíře pokrytí, determinovaná počtem sekvenovaných genů, která může vést k nezachycení potenciálně klinicky významné varianty (s rostoucím počtem genů pokrytých panelem tato limitace úměrně ztrácí na významu)

založené na tzv. **sekvenačních panelech**. V rámci tohoto přístupu je sekvenován předem definovaný panel řádově desítek až stovek genů významných v nádorové biologii, případně pro terapeutické plánování. Tento přístup má oproti WGS a WES řadu výhod především z praktického hlediska. Ještě menší šíře pokrytí, než má WES, umožňuje větší hloubku čtení, a tedy nejvyšší senzitivitu, což může mít svůj význam především u nádorových biopsií s nízkým zastoupením nádorových buněk. Sekvenační panely jsou méně finančně nákladné, jsou jednodušší z hlediska interpretace a nevedou k náhodným nálezům. Z hlediska technologického umožňují panely lepší optimalizaci sekvenačního protokolu a vyšší míru jeho standardizace.

Mezi komerčně nabízenými sekvenačními panely se setkáme s panely obsahujícími jak desítky, tak i stovky genů, přičemž s narůstajícím počtem genů zařazených do panelu logicky narůstá také pravděpodobnost záchytu klinicky relevantní mutace a rovněž přesnost výpočtu tzv. mutační nálože nádoru – viz níže. Na druhou stranu cílené (*hotspot*) sekvenování nikdy nepokrývá všechny potenciálně klinicky relevantní genomové varianty, a pravděpodobnost jejich záchytu pomocí WGS a WES bude proto vždy vyšší. Informace o individuálních nádorových exomech či genomech má rovněž významnou hodnotu pro translační výzkum v onkologii.

Ať již použijeme kteroukoliv z výše uvedených aplikací genomového sekvenování, je možné její výstupy použít pro **výpočet tzv. mutační nálože nádoru** neboli TMB (*tumor mutational burden*). V případě sekvenačních panelů lze výpočet TMB provést pouze v případě, že panel pokrývá > 200 genů. Teprve od tohoto limitního počtu genů začíná hodnota TMB získaná ze sekvenačního panelu korelovat s TMB získanou pomocí WGS a WES. TMB je definována jako počet nesynonymních (protein měnících) mutací na 1 Mb (na milion bází) DNA a jeho hodnota může sloužit jako citlivý ukazatel odpovědi na protinádorovou imunoterapii.

Role sekvenování nové generace ve fungování molekulárně onkologické indikační komise

Použitím jakékoliv z výše uvedených aplikací NGS získáme více či méně komplexní informace o individuálním nádorovém genomu. Tyto informace jsou použity jako podklad pro přípravu detailních výsledkových protokolů obsahujících přehled klinicky relevantních variant a stupeň klinické evidence jejich vztahu ke konkrétnímu preparátu a léčebné odpovědi. Takto připravené protokoly jsou následně projednávány v kontextu celkového klinického stavu pacienta na **multioborové molekulárně onkologické indikační komisi (MTB – molecular tumor board)** s cílem vytvoření individuálního terapeutického plánu založeného na biologických vlastnostech daného tumoru.

prof. RNDr. Ondřej Slabý, Ph.D.
CEITEC, Masarykova Univerzita Brno
Masarykův onkologický ústav, Brno

Zdroje:

1. Moorcraft S. Y., Gonzalez D., Walker B. A. *Understanding next generation sequencing in oncology: a guide for oncologists. Crit Rev Oncol Hematol* 2015; 96 (3): 463–474.
2. Strom S. P. *Current practices and guidelines for clinical next-generation sequencing oncology testing. Cancer Biol Med* 2016; 13 (1): 3–11.
3. Carter T. C., He M. M. *Challenges of identifying clinically actionable genetic variants for precision medicine. J Healthc Eng* 2016; 3617572, doi: 10.1155/2016/3617572.