

Sekvenování nové generace a možnosti jeho využití v onkologické praxi

Next Generation Sequencing – Application in Clinical Practice

Koubková L.¹, Vojtěšek B.¹, Vyzula R.^{1,2}

¹ Regionální centrum aplikované molekulární onkologie, Masarykův onkologický ústav, Brno

² Klinika komplexní onkologické péče, Masarykův onkologický ústav, Brno

Souhrn

Metoda masivně paralelního sekvenování umožnila rychlejší a ekonomičtější výzkum v oblasti genomiky. Tato technologie umožňuje osekvenovat kompletní lidský genom za zlomek ceny i času v porovnání s doposud používanou Sangerovou metodou. Zavedení této techniky do oblasti onkologického výzkumu významně přispělo k molekulární charakterizaci nádorů a hlubšímu porozumění jejich evoluce. Pomocí masivně paralelního sekvenování byly identifikovány nové kauzální mutace, které jsou podstatou nádorových dědičných syndromů, porovnáním sekvenování DNA nádorové a příslušné zdravé tkáně byly odhaleny nové mutace a strukturní aberace u více než 15 rozdílných nádorových onemocnění. V tomto přehledu jsou uvedeny technické charakteristiky nejrozšířenějších sekvenačních platform, krátce shrnuty jejich výhody a nevýhody a popsány možnosti uplatnění v klinické praxi.

Klíčová slova

technologie masivně paralelního sekvenování – genomika – mutace – onkologický výzkum – klinická aplikace – personalizovaná léčba

Summary

Development of new sequencing methods allowed faster and more economical genomic research. With these technologies, it is now possible to determine the complete sequence of human genome in a short time period and at a relatively low cost. Introduction of next generation sequencing methods to cancer research provided a comprehensive molecular characterization of cancers and enabled deeper insights into tumor complexity, heterogeneity and evolution. Next generation technologies have been applied to identify new causal mutations in genes in hereditary cancer syndromes. More than 15 various tumor types have been already sequenced and compared to that of normal cells allowing identification of new cancer driving mutations and genome structural rearrangements. In this review, we describe technical characteristics of main next generation sequencing platforms, briefly overview their pros and cons and clinical perspective.

Key words

high-throughput nucleotide sequencing – genomics – mutations – cancer research – clinical application – personalized treatment

Práce byla podpořena Evropským fondem pro regionální rozvoj a státním rozpočtem České republiky (OP VaVpl – RECAMO CZ.1.05/2.1.00/03.0101) a MZ ČR – RVO (MOÚ, 00209805).

This work was supported by the European Regional Development Fund and the State Budget of the Czech Republic (RECAMO, CZ.1.05/2.1.00/03.0101) and by MH CZ – DRO (MMCI, 00209805).

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 zaslané do biomedicínských časopisů.

The Editorial Board declares that the manuscript met the ICMJE "uniform requirements" for biomedical papers.



Mgr. Lucie Koubková
Regionální centrum aplikované
molekulární onkologie
Masarykův onkologický ústav
Žlutý kopec 7
656 53 Brno
e-mail: lucie.koubkova@mou.cz

Obdrženo/Submitted: 4. 2. 2014

Přijato/Accepted: 1. 4. 2014

Úvod

Deoxyribonukleová kyselina (DNA) jako genetický materiál byla popsána v roce 1944 Oswaldem Theodorem Averym [1]. Několik let poté (1953) objevili Watson a Crick [2] dvouvláknovou strukturu DNA a její čtyři základní stavební kameny. Genetická informace téměř každého druhu je uložena v primární struktuře nukleových kyselin a znalost její sekvence je zásadní informací pro výzkum buněčných struktur a jejich funkcí, s jejichž pomocí se moderní věda snaží odhalit „tajemství života“. Rozluštění DNA sekvence je nezbytné v podstatě pro všechna odvětví biologického výzkumu. Ještě na začátku 70. let 20. století bylo sekvenování DNA velmi obtížné a provádělo se nepřímo sekvenováním RNA molekul. Druhá polovina 80. let přinesla hned dvě metody DNA sekvenování. První metoda, kterou vynalezli Maxam a Gilbert [3], je založena na chemické modifikaci DNA a následném rozštěpení řetězce v místech modifikovaných nukleotidů. Druhá byla publikována Frederikem Sangerem et al v roce 1977 [4]. Je známa jako enzymatická metoda sekvenování (Sangerovo sekvenování) a využívá specifických vlastností DNA polymerázy při syntéze nového řetězce. Technologie enzymového sekvenování byla „první generací“ komerčně dostupného sekvenování a navzdory její pracnosti i časové náročnosti se Sangerovo sekvenování stalo velmi rozšířeným v laboratořích po celém světě. Významný pokrok přinesla automatizace této technologie. V roce 1987 uvedla firma Applied Biosystems na trh první automatický sekvenátor (AB370) využívající kapilární elektroforézu k separaci jednotlivých úseků DNA a umožňující osekvenovat 500 kilobází za den s délkou čteného fragmentu (tzv. read) okolo 600 bází. Současný model AB3730xl dokáže za den vyhodnotit až 2,88 megabází a délka čtení dosahuje 900 bází [5]. Automatické sekvenátory dominovaly v oblasti molekulární biologie téměř dvě desetiletí a vedly k významným objevům [6]. Za vynaložení velkého úsilí a peněz byl pomocí těchto sekvenátorů přečten první lidský genom. Projekt lidského genomu (human genome project) nicméně poukázal na limitace San-

gerova sekvenování a potřebu nových technologií s vyšší výkonností a přesností, nižšími náklady a menší pracností.

Sekvenátory nové generace

Technika sekvenování DNA zaznamenala během posledních let velký posun kupředu díky novým sekvenátorům tzv. druhé generace. Sekvenátory první generace detekovaly DNA báze v řadě jednu po druhé, zatímco sekvenátory druhé generace umožňují masivní paralelní sekvenování až tisíců molekul DNA současně. Díky technologii masivního paralelního sekvenování se významně snížila doba potřebná k přečtení dlouhých DNA sekvencí. K určení sekvence prvního lidského genomu bylo potřeba 10 let. Nové technologie sekvenování umožňují určit sekvenci lidského genomu nebo jen jeho vybraných částí za zlomek času (několik týdnů) a za mnohem nižší cenu. Sangerovo sekvenování se v současnosti stále využívá a hodí se např. pro sekvenování PCR produktů nebo pro malé sekvenační projekty, ale pro náročnější aplikace je nahrazováno sekvenátory druhé generace. Trhu dnes dominují tři platformy: Roche 454 Genome Sequencer, Illumina Genome Analyzer a Life Technologies SOLiD System. Všechny tři platformy byly vyvinuty na konci 90. let 20. století a na trh uvedeny kolem roku 2005 [7]. Ačkoli všechny dostupné technologie sekvenování druhé generace (next generation sequencing – NGS) využívají rozdílnou chemii, sdílejí společně tyto kroky: a) příprava templátu neboli vytvoření knihovny ampikonů, b) sekvenování a detekce inkorporovaných nukleotidů a c) analýza dat. Unikátní kombinace přístupů k jednotlivým krokům celého sekvenačního běhu určuje rozdíly mezi jednotlivými technologiemi a typem a množstvím produkovaných dat. Rozdíly ve výstupních datech představují výzvu v interpretaci výsledků dosažených různými platformami při porovnávání stejné sekvence. Všichni výrobci poskytují odhad přesnosti a kvality jednotlivých čtení, ale neexistuje žádný konsenzus, že kvalita čtení z jedné platformy je ekvivalentní kvalitě čtení z platformy druhé [6]. NGS technologie mohou být rozděleny do dvou základních kategorií. První sku-

pinu představují technologie založené na PCR amplifikaci templátu (PCR-based technologies) a zahrnují tyto komerčně dostupné platformy: Roche 454 System, Illumina sequencers, AB SOLiD System and Ion Personal Genome Machine. Druhou skupinu tvoří technologie, které využívají tzv. single-molecule sequencing, tzn. že nedochází k amplifikačnímu kroku před vlastní sekvenací. Tato skupina zahrnuje např. technologie HelixScope a PacBio RS SMRT Ssystem [8] a Oxford Nanopore.

Roche 454

První komerčně dostupný NGS analyzátor 454 Genome FLX System byl uveden na trh v roce 2005. Pracuje na principu tzv. pyrosekvenování. To lze zjednodušeně popsat jako sled enzymatických reakcí, během kterých se začleňují jednotlivé báze do nově vznikajícího řetězce DNA, přičemž dochází k emisi viditelného světla. Množství uvolněného světla je úměrné počtu začleněných nukleotidů. Celý proces začíná přípravou DNA knihovny, kdy je DNA naštěpena na dvouvláknové fragmenty a k jednotlivým fragmentům jsou připojeny krátké koncově specifické adaptory (A a B) sloužící v dalších krocích k purifikaci a amplifikaci i k vlastní sekvenaci. B-adaptor je na svém 5'-konci označen biotinem, který slouží k imobilizaci fragmentů na streptavidinem pokryté magnetické kuličky. Dvouvláknové fragmenty navázané na magnetické kuličky jsou denaturovány, přičemž nebiotinylovaný komplementární řetězec je uvolněn [7]. Takto vzniklé jednořetězcové fragmenty (knihovna DNA) jsou připraveny k hybridizaci ke speciálním DNA kuličkám, které mají na svém povrchu komplementární sekvenci DNA sloužící jako primer pro následnou amplifikaci. Technologie 454 využívá klonální amplifikaci pomocí emulzní PCR (emPCR). Hybridizační podmínky jsou navrženy tak, aby se na jednu DNA kuličku navázal pouze jeden DNA fragment. Každá kulička je uzavřena ve svém vlastním „mikroreaktoru“ – olejové emulzi s vodou, kde probíhá nezávislá amplifikace jednoho fragmentu. Po skončení emPCR jsou kuličky uvolněny z emulze, přičemž každá nese v průměru 10 miliónů identických kopií původní

DNA. Následuje tzv. enrichment (obohacení) – krok, kdy se zachovávají pouze kuličky s amplifikovanou DNA, a zbytek se odmyje. Na obohacené kuličky, které nyní nesou jednořetězcové DNA fragmenty, se přichytí sekvenační primer a celá směs se nanese do jamky tzv. pikotitrační destičky (pico titer plate – PTP), která je navržena tak, aby do jedné jamky zapadla pouze jedna DNA kulička. PTP je rozdělena do čtyř vrstev, kam se centrifugací postupně vpraví další druhy kuliček důležité pro pyrosekvenační reakci. Klíčové komponenty reakce představují kuličky s enzymy sulfurylázou, luciferázou, apyrázou a DNA polymerázou a se substráty adenosin 5'-fosfosulfátem (APS) a luciferinem. Jednotlivé nukleotidy proudí nad PTP v pevném pořadí. Inkorporace komplementárního nukleotidu DNA polymerázou do nově vznikajícího řetězce je doprovázena uvolněním pyrofosfátu, který je pomocí enzymu ATP sulfurylázy za přítomnosti APS přiveden na ATP, jež umožní konverzi luciferinu na oxyluciferin katalyzované enzymem luciferázou. Při této reakci dochází k emisi viditelného světla, které je zaznamenáno CCD čipem. Nespotřebované nukleotidy a ATP jsou degradovány apyrázou. Proces se cyklicky opakuje a sekvence je odečítána z tzv. pyrogramu. Hrubá data jsou procesována 454 pyrosekvenačním softwarem a analyzována různými filtry kvality pro odstranění sekvencí nízké kvality nebo sekvencí, které mají více jak jeden iniciální DNA fragment. Pro analýzu dat jsou dostupné tyto bioinformatické nástroje: GS De Novo Assembler, GS Reference Mapper a GS Amplicon Variant Analyzer [7]. Současný sekvenátor GS FLX Titanium chemistry poskytuje v jednom běhu až jeden milion čtení dlouhých až 1 000 bází [9]. Na trhu je dostupný také stolní (tzv. benchtop) sekvenátor GS Junior s nižší kapacitou vhodný do rutinních laboratoří [8]. Jako všechny platformy má i technologie 454 své výhody a nevýhody. Její největší limitací jsou problémy v detekci homopolymerních úseků, jako např. AAA a GGG. V procesu inkorporace více nukleotidů stejného typu závisí počet začleněných nukleotidů pouze na množství emitovaného světla. Tato technologie je tedy náchylná k větším chybám v určení

počtu začleněných nukleotidů a z toho plynoucí inserce nebo delece v homopolymerních úsecích spíše než substituce nukleotidu [9]. Tento problém byl částečně vyřešen pomocí speciálních programů, které dokážou rozlišit a vyfiltrovat chybové sekvence [10]. Další nevýhoda je stále vysoká cena reagentů. Ve srovnání s ostatními platformami však 454 technologie poskytuje velmi dlouhá čtení, a výborně se tak hodí pro celogenomové *de novo* sekvenování, resekvenování (sekvence části nebo celého genomu se porovnává s již známou sekvencí za účelem nalézt rozdíly oproti referenční sekvenci) a pro studium bakteriální diverzity, kdy se analyzují mikrobiální genomy z různých vzorků (metagenomika) [9]. Jinou nespornou výhodou 454 technologie je její vysoká rychlost – pouhých 10 hod pro kompletní celého sekvenačního procesu [5].

Illumina/Solexa

První sekvenátor založený na principu sekvenace syntézou ve spojení s tzv. bridge amplifikací uvedla na trh společnost Illumina v roce 2007 [8]. U této technologie se templátová DNA hybridizuje na opticky transparentní pevný povrch reakční komůrky (flow cell) a využívá chemicky reverzibilně modifikovaných nukleotidů [9]. Příprava knihovny zahrnuje naštěpení DNA na fragmenty o velikosti < 800 bází. Konce fragmentů jsou zarovnané, fosforylovány a na 3'-konci adenylovány a k oběma koncům jsou přidány adaptory. Po denuraci jsou jednotlivé fragmenty hybridizovány k reakční komůrce, jejíž povrch je hustě pokryt komplementárními adaptory k adaptorům připojeným k DNA fragmentům. Každý fragment je tak svým jedním koncem imobilizován k povrchu reakční komůrky, kde následně probíhá amplifikace. Poté je přidána směs reagentů potřebných pro PCR, adaptory na povrchu reakční komůrky slouží jako primery pro syntézu dvouvláknové DNA. Ta je následně denaturována. Původní templát je odmyt a nově syntetizované vlákno DNA zůstává kovalentně navázáno k povrchu reakční komůrky. Svým volným koncem hybridizuje k adaptorům (primerům) na povrchu reakční komůrky, vlákno se ohne a dojde k pře-

mostění. V dalším PCR cyklu je adaptor (primer) opět prodloužen polymerázou a vytvoří se dvouvláknový most (bridge). Celý proces se cyklicky opakuje, je proto nazýván bridge amplification. Nakonec jsou dvouvláknové mosty denaturovány, reverzní řetězce odštěpeny a odmyty a zůstanou pouze tzv. klastry tvořené cca 1 000 identických kopií DNA fragmentů připravených k vlastní sekvenaci. Sekvenační primery jsou hybridizovány k adaptorovým sekvencím a do reakční komůrky s klastry je nalita směs polymerázy a čtyř rozdílně fluorescenčně značených nukleotidů s chemicky inaktivovanou 3'-OH skupinou. Je tak zaručeno, že v jednom cyklu je inkorporován pouze jeden nukleotid. Jakmile dojde k začlenění nukleotidu do řetězce DNA, pozice a typ nukleotidu jsou zaznamenány díky jeho fluorescenční značce pomocí CCD kamery. Terminační skupina na 3'-konci nukleotidu i fluorescenční barvička jsou odstraněny a cyklus je opakován. Sekvence každého klastru je generována speciálním algoritmem, který jednotlivým bázím přiděluje určitou hodnotu, na jejímž základě jsou vyřazeny sekvence nízké kvality. První sekvenátor společnosti Illumina, Illumina Genome Analyzer poskytoval přečtení sekvencí dlouhých 35 bází a generoval 1 gigabázi (Gb) dat za jeden sekvenační běh, který trval 2–3 dny. V současnosti je v nabídce společnosti Illumina několik přístrojů různé výkonnosti: (throughput) MiSeq, MiSeqDx, NextGen 500, HiSeq2500 a HiSeq X Ten. MiSeq a MiSeqDx jsou sekvenátory určené do malých laboratoří pro sekvenování malých genomů, ampikonové sekvenování a cílené sekvenování vybraných oblastí genů. MiSeqDx je prvním sekvenátorem schváleným americkou FDA (Food and Drug Administration) pro *in vitro* molekulární diagnostiku garantující spolehlivost výsledků. NextGen 500 představuje první vysokokapacitní stolní sekvenátor umožňující nejširší záběr DNA i RNA sekvenačních aplikací. HiSeq2500 je ultra vysokokapacitní sekvenátor umožňující opět nejširší záběr sekvenačních aplikací předurčený zejména pro velké sekvenační studie prováděné v specializovaných centrech. HiSeq X Ten je nejvýkonnější sekvenační platforma cílená

pro populační studie překonávající cenovou bariéru 1 000 \$ za osekvenování jednoho lidského genomu. Tato platforma se skládá z 10 ultra vysokokapacitních sekvenátorů, které jsou společně schopny osekvenovat více než 18 000 genomů za rok. HiSeq X Ten umožňuje cenově dostupnější a dosažitelnější sekvenování lidského genomu než cokoli jiného kdy předtím [11]. Sekvenátory Illumina jsou limitovány relativně krátkou délkou sekvenačních čtení v závislosti na typu přístroje, např. 2 × 125 (HiSeq 2500) 2 × 150 (NextSeq 500) nebo až 2 × 300 (MiSeq) bází. Tato limitace je dána zvýšenou nebo sníženou účinností inkorporace nukleotidu a selháním při odstraňování nebo přidávání terminační skupiny, což může způsobit nekompletní prodloužení vlákna. Proto je nejčastější chyba, která vzrůstá s délkou čtení, substituce nukleotidu. Další nevýhoda je nerovnoměrné pokrytí (tzv. coverage) v oblastech bohatých na AT a GC sekvenace. V porovnání se Sangerovým sekvenováním jsou Illumina sekvenátory schopné produkovat mnohem více dat za méně času i peněz, ale s vyšší chybovostí vedoucí k falešné pozitivě při identifikaci sekvenačních variací [6]. Ovšem díky ultra vysoké výkonosti, kapacitě a cenové hospodárnosti provozu sekvenátory Illumina v současnosti ovládají trh a využívají se k většině celogenomových i resekvenačních aplikací zahrnujících lidský genom i genomy modelových organismů [8].

Life Technologies/SOLiD

V roce 2007 představila tehdy ještě společnost Applied Biosystems (dnes Life Technologies) platformu SOLiD (sequencing by oligonucleotide ligation and detection) pro sekvenování druhé generace. Na rozdíl od předchozích dvou je SOLiD technologie založena na sekvenování ligací. DNA knihovna se připravuje pomocí emPCR a k fragmentům DNA se připojí krátké adaptory komplementární k adaptorům imobilizovaným na povrchu magnetických kuliček. Po amplifikaci jsou kuličky kovalentně navázány na sklíčko se speciálně upraveným povrchem, které se vloží do kazety umožňující fluidní průtok. SOLiD systém využívá osm nukleotidů dlouhé sondy. Každá ze sond má známou

sekvenci prvních dvou bází, je označena jednou ze čtyř různých fluorescenčních barev, přičemž každá barva představuje 4 z 16 možných di-nukleotidových sekvencí. Solid technologie zaručuje přečtení každého nukleotidu 2krát, což velice zvyšuje přesnost, s jakou je určeno pořadí nukleotidů dané sekvence. Mezi výše uvedenými NGS platformami má tedy SOLiD systém nejnižší chybovost. Jeho nejběžnějším typem chyby je substituce nukleotidu. Využití nachází především v aplikacích, jako je celogenomové resekvenování, cílené resekvenování, transkriptomová analýza a epigenomika (chromatinová imunoprecipitace a metylace). V současnosti jsou v nabídce dvě varianty SOLiD sekvenátoru: 5500 System a 5500xl System. Z několika dnes dostupných platform je obtížné vybrat jednu optimální. Ze třech výše zmíněných NGS technologií Illumina HiSeq generuje nejvíce dat za nejnižší cenu, SOLiD System má nejvyšší přesnost a Roche 454 poskytuje největší délku čtení.

Life Technologies/Ion torrent

V roce 2010 představila firma Life Technologies přístroj Ion Personal Genome Machine (PGM) založený na nové technologii schopné přímo převádět chemický signál do digitální podoby. Místo optického způsobu zaznamenávání jednotlivých nukleotidů se zde využívá detekce vodíkových protonů uvolněných v průběhu syntézy nově vznikajícího řetězce katalyzovaného DNA polymerázou. Proces probíhá na polovodičovém čipu hustě pokrytém mikrojamkami, pod nimiž je umístěna na ionty citlivá vrstva. Inkorporace nukleotidu způsobí uvolnění H⁺, čímž dojde ke změně pH, kterou zaznamenává detektor. K přípravě knihovny se opět využívá emPCR. Kuličky s templátem jsou poté deponovány na čip tak, že v každé jamce je pouze jedna molekula DNA. Čip je postupně zapláván jednotlivými druhy nukleotidů a dochází k syntéze DNA. Pokud nukleotid není komplementární, detektor zaznamená nulový signál. Dojde-li k začlenění dvou nukleotidů, je signál dvojnásobný. Díky tomu, že technologie nepoužívá žádné fluorescenční detektory, kamery, skenery atp., trvá jeden sekvenační běh méně

než 2 hod [5]. Technologie je velice jednoduchá, rychlá a levná. Celkové množství produkovaných dat je určeno hustotou jamek na čipu. V roce 2012 byla představena nová generace polovodičového sekvenátoru – Ion Proton s kapacitou čipu slibující dostatečnou kapacitu i pro sekvenaci lidského genomu.

Technologie tzv. single-molecule sekvenování

Technologie single-molecule sekvenování představuje výraznou změnu ve vývoji nových sekvenačních metod. Tato technologie, někdy označovaná jako třetí generace sekvenování, přináší hned několik výhod. Nevyužívá žádný amplifikační krok před vlastním sekvenováním, což zkracuje dobu přípravy DNA, redukuje cenu, snižuje chybovost pramenící z amplifikace, umožňuje vyšší flexibilitu v délce čtení a přesnou kvantifikaci DNA molekul, protože signál je zaznamenáván v reálném čase [9].

První dostupnou technologií třetí generace byla platforma Helicos Genetic Analysis System využívající sekvenace syntézou. Vytvoří se jednovláknová DNA knihovna, která je neuspořádaně deponována na plochou destičku. Destička je v každém sekvenačním cyklu zaplavována DNA polymerázou a jedním ze čtyř fluorescenčně značených nukleotidů a signál je zaznamenán CCD kamerou. Po promytí se odstraní fluorescenční značky a prodloužování řetězce se opakuje [9]. V současné době není tato technologie dále vyvíjena.

Další z třetí generace sekvenačních technologií tzv. SMRT (single-molecule real-time) byla vyvinuta společností Pacific Bioscience. Tento systém využívá nanostruktury zvané Zero Mode Waveguide – destičky s deseti tisíci jamkami o průměru 10 nm. Během sekvenačního procesu se komplementární vlákno syntetizuje pomocí DNA polymerázy ukotvené na dně každé jamky. Fluorescenční značka je umístěna na fosfátové skupině nukleotidu, což má za následek uvolnění záblesku zároveň s jeho inkorporací. Proces inkorporace i uvolnění fluorescence trvá po určitou dobu, čehož se využívá pro určení identity báze. Tento přístup nevyžaduje promývací kroky při začleňování jednotlivých typů nuk-

leotidů, a zrychluje tak běh sekvenování. V současnosti poskytuje sekvenátor Pacific Bioscience model PacBio RS II největší délku čtení a extrémně vysokou přesnost. Využívá se zejména pro resekvenační projekty na malých prokaryotických i eukaryotických genomech díky schopnosti sekvenovat ultra dlouhé úseky (v průměru 4 000 bází). To umožňuje přesnější mapování na referenční sekvence.

Dalším příkladem využití nanotechnologie pro novou generaci sekvenování je tzv. nanopore sequencing. Koncept využití nanopóru jako biosenzoru byl navržen už v polovině 90. let, kdy začal výzkum nanopóru na půdě akademických institucí, jako je Oxford či Harvard. V roce 2005 vznikla společnost Oxford Nanopore s cílem převést akademický výzkum do komerční sféry [12]. Technologie je založena na biologických vlastnostech nanopóru. Nanopóry jsou součástí proteinových kanálků v membránách a dovolují výměnu iontů. Nanopórem protéká konstantní proud. Analyt, kterým je v případě využití pro sekvenování jednořetězcová molekula DNA, prochází nanopórem a dochází k detekci jednotlivých nukleotidů, přičemž pro každý typ nukleotidu je předem určena modulace proudu. Využití technologie nanopóru pro sekvenování DNA nabízí mnoho výhod oproti stávajícím platformám. Technologie má minimální požadavky na reagenty i přípravu vzorku, je levná, rychlá a nabízí analýzu DNA v reálném čase. Sekvenování stejné molekuly DNA může být neustále opakováno. Jednotlivá čtení mohou být dlouhá až desítky kilobází. Většina současných sekvenačních technik je ve srovnání s technologií nanopóru komplikovaná – vyžaduje kromě izolace DNA i její značení a namnožení, rozbití na malé fragmenty, které jsou pak mnohonásobně sekvenovány. I technologie nanopóru má však svá úskalí, která je nutno vyřešit, než na trh přijde první „nanotech“ sekvenátor. Vysoká rychlost, s jakou je DNA translokována skrz nanopór, totiž způsobuje problémy s rozlišením signálu nukleotidu od signálu pozadí. V roce 2013 bylo publikováno několik prací, které posunují řešení tohoto problému dále [13–15]. Zjistilo se

např., že působením světla o určité vlnové délce lze docílit zpomalení průtoku DNA skrz nanopór. Technologie nanopóru prošla rozličnými technickými zlepšeními, která umožnila lepší rozlišení i stabilitu nanopóru. V roce 2012 ohlásila firma Oxford Nanopore příchod prvního sekvenátoru – GridION System, který slibuje přečtení celého lidského genomu v 15 minutách. Nedávno byl také představen miniaturizovaný sekvenátor, jenž se vejde do kapsy – MinION od Oxford Nanopore schopný přečíst DNA fragmenty o velikosti do 10 kilobází. V současnosti jsou oba systémy ještě komerčně nedostupné.

Aplikace NGS technologií

Schopnost NGS technologií produkovat obrovské množství sekvenačních dat v poměrně krátké době za relativně nízkou cenu je dělá užitečnými nástroji pro řadu aplikací, např. *de novo* genomové sekvenování, celogenomové nebo cílené resekvenování pro objevování nových mutací a polymorfizmů, analýza transkriptomu a DNA metylovaných oblastí, mapování DNA-proteinových interakcí pomocí chromatinové imunoprecipitace (ChIP-Seq) atd. Nástup NGS technologií umožnil charakterizovat molekulární podstatu nejrozličnějších onemocnění. NGS technologie nacházejí široké uplatnění v molekulární diagnostice dědičných chorob, infekčních onemocněních, prenatální diagnostice, farmakogenomice, v molekulární diagnostice nádorů i jejich prognóze [16]. Nádorové onemocnění je způsobeno nahromaděním mutací v genetickém materiálu. Tyto mutace, které mohou být germinální (zděděné od rodičů) i somatické (získané v průběhu života), mají zásadní efekt na onkogeny, nádorové supresory nebo geny mající na starosti reparaci DNA, umožňují uniknout kontrolním buněčným mechanismům vedoucím v konečném důsledku ke vzniku nádoru. Studium mutačních profilů jednotlivých typů nádorů tedy významně pomáhá objasnit mechanismy kancerogeneze. V posledních několika letech bylo pomocí NGS technologií provedeno velké množství studií, které poskytly komplexnější molekulární charakterizaci nádorových onemocnění

a vedly k objevení nových genů asociovaných zejména s rakovinou mléčné žlázy, vaječnicků, tlustého střeva a konečníku, plic, jater, ledvin, rakovinou hlavy a krku, melanomem a akutní myeloidní leukemií [17] a k identifikaci velkého počtu rekurentních mutací (identické mutace detekované ve velkém počtu vzorků) [18], které byly standardními cytogenetickými technikami nezachytitelné. NGS technologie se ukázaly být velice slibným diagnostickým nástrojem pro identifikaci germinálních mutací, jež jsou podkladem familiárních syndromů. V případě familiárního melanomu jsou známy pouze dvě mutace odpovědné za rozvoj onemocnění. Mutace v *CDKN2A* (*cyclin-dependent kinase inhibitor*) se podílí na 40 % familiárních případů melanomu, kdežto mutace v *CDK4* (*cyclin-dependent kinase 4*) byla nalezena jen u velmi malého počtu probandů. Sekvenování nové generace probandů z několika rodin s výskytem familiárního melanomu odhalilo pro melanom novou rizikovou variantu c.G1075A genu *MITF* (*microphthalmia-associated transcription factor*) [18].

Celogenomové sekvenování

Celogenomové sekvenování zahrnuje *de novo* sekvenování a tzv. resekvenování. Při resekvenování se sekvenují stejné genomy jako u *de novo* přístupu, ale s tím rozdílem, že je k dispozici referenční sekvence, na kterou se nová sekvence zpětně mapuje. Výhoda celogenomového sekvenování je sekvenace celé chromozomální DNA, pokrytí promotorových a regulačních sekvencí a tím i poskytnutí informace o úplném mutačním profilu daného genomu zahrnujícím bodové mutace, malé inserce a delece, chromozomální změny a strukturní variace typu CNVs (copy number variations). Celogenomové sekvenování se proto nejvíce využívá k identifikaci nových a vzácně se vyskytujících mutací [16]. První celogenomový sekvenační projekt byl realizován na nádorovém a normálním vzorku téhož pacienta s cytogeneticky normální akutní myeloidní leukemií (AML) v roce 2008 [19] a vedl k identifikaci osmi nových somatických mutací zúčastněných v patogenezi AML a několika somatických insercí

a delecí. Pomocí celogenomového sekvenování lze také úspěšně detekovat chromozomální přestavby. Znalost chromozomálních translokací je velmi důležitá z biologického i klinického pohledu. Například Gleevec® byl vyvinut pro inhibici fúzní tyrozinkinázy Bcr/Abl – chromozomální přestavby typické pro chronickou myeloidní leukémii (CML). CML je nádorové onemocnění se specifickou chromozomální aberací tzv. Philadelphským chromozomem, kde vzniká fúzní gen *BCR-ABL*. Produktem toho genu je tyrozinkináza fosforylující celou řadu proteinů klíčových pro proliferaci a diferenciaci hematopoetických buněk. Gleevec je příklad léku, který byl vyvinut na základě molekulární podstaty CML. Podobné chromozomální přestavby u solidních tumorů byly zatím charakterizovány jen chabě z důvodu nedostatku systematických přístupů, avšak s nástupem NGS technologií je detekce chromozomálních aberací i u solidních nádorů realizovatelná a cenově dostupná. Nedávná studie objevila translokace genů *EML4-ALK* (*echinoderm microtubule associated protein like 4 – EML 4, anaplastic lymphoma kinase – ALK*) u nemalobuněčného karcinomu plic a také *TMPRSS2-ERG* (*transmembrane protease serine 2 – TMPRSS2, ETS-related gene – ERG*) u karcinomu prostaty [20]. Celogenomové sekvenování je velmi výhodným přístupem pro nalezení nových potenciálně terapeutických cílů.

Exomové sekvenování

Exomové sekvenování je přístup tzv. cíleného sekvenování, kdy jsou sekvenovány pouze kódující oblasti – exomy. Lidský exom představuje asi 1 % velikosti celého lidského genomu, exomové sekvenování je tudíž snazší, cenově méně náročné a umožňuje vyšší pokrytí (coverage). Pomocí celoexomového sekvenování (whole exome sequencing – WES) byly např. úspěšně identifikovány mutace odpovědné za familiární nádory pankreatu [21], dědičného feochromocytomu [22] či familiárního melanomu [23]. Studium mutačního profilu adenokarcinomu plic odhalilo nové mutace v řadě protoonkogenů (*ERBB4, KDR, FGFR4* a *NTRK*) a nádorových supresorů (*NF1, RB1, ATM* a *APC*), jejichž znalost na-

bízí nové způsoby léčby. Podobně byly u glioblastomu detekovány aberace *RTK, TP53, RB* signální dráhy a mutace v *NF1*, které mohou v budoucnu vést k vývoji inhibitorů drah MEK, nebo RAF a k zefektivnění léčby [20]. Velký počet WES studií nových genetických aberací je shrnut v práci Tran et al [24].

Sekvenování transkriptomu

Transkriptom představuje soubor všech molekul RNA (mRNA, rRNA, tRNA a další nekódující RNA molekuly). Transkriptomová analýza je jedním z nejdůležitějších přístupů pro komplexní molekulární charakterizaci nádorů. Sekvenování transkriptomu (RNA-seq) velmi usnadnilo detekci mezigenových fúzí, somatických mutací a alternativních sestřihových variant. Na rozdíl od čipových technologií není technologie RNA-seq limitována předchozí znalostí genomu, dynamickým rozlišením nebo zkříženou (cross) hybridizací. Fúze genů jsou v genomech nádorů (převážně u hematologických malignit) velmi časté. Pomocí RNA-seq byly objeveny nové fúze genů např. u rakoviny mléčné žlázy, prostaty, lymfomu a melanomu [9]. Výskyt určitých fúzí genů je spojován s procesem kancerogeneze specifických tkání nebo orgánů a tyto změny mohou být využity jako diagnostické markery onemocnění. Některé fúze byly naopak nalezeny napříč několika rozdílnými druhy nádorů. Takovým příkladem je fúze genů z RAF signální dráhy, která byla identifikována u karcinomu prostaty, žaludku i melanomu [9], mající potenciál terapeutického cíle pro všechny tři malignity. RNA-seq umožňuje studium i nekódujících ncRNA molekul, jež se účastní širokého spektra biologických procesů zahrnujících regulaci proliferace, diferenciace a apoptózy. Typickým příkladem jsou mikroRNA (miRNA) jako klíčové regulátory genové exprese.

Cílené sekvenování vybraných oblastí

Cílené sekvenování (targeted sequencing) je inovativní technika, která umožňuje sekvenovat pouze vybrané geny nebo definované oblasti genomu, což šetří čas, peníze a vyžaduje i méně pro-

storu pro skladování dat. Typicky se tato technika využívá pro sekvenování velkého počtu vzorků při screeningu nebo validaci genetických variant v populaci. NGS zde umožňuje identifikovat vzácnější varianty, které nejsou Sangerovým sekvenováním zachytitelné, anebo je jejich identifikace příliš drahá. Využije-li se pro cílené sekvenování vysokého pokrytí (coverage), mohou být tímto způsobem charakterizovány neobvyklé sekvenovací varianty v populaci s frekvencí o četnosti pod 1 %. Amplikonové sekvenování je vhodné do klinického prostředí pro odhalování somatických mutací v komplexních biologických vzorcích, jako je např. smíšený vzorek nádorové a normální DNA.

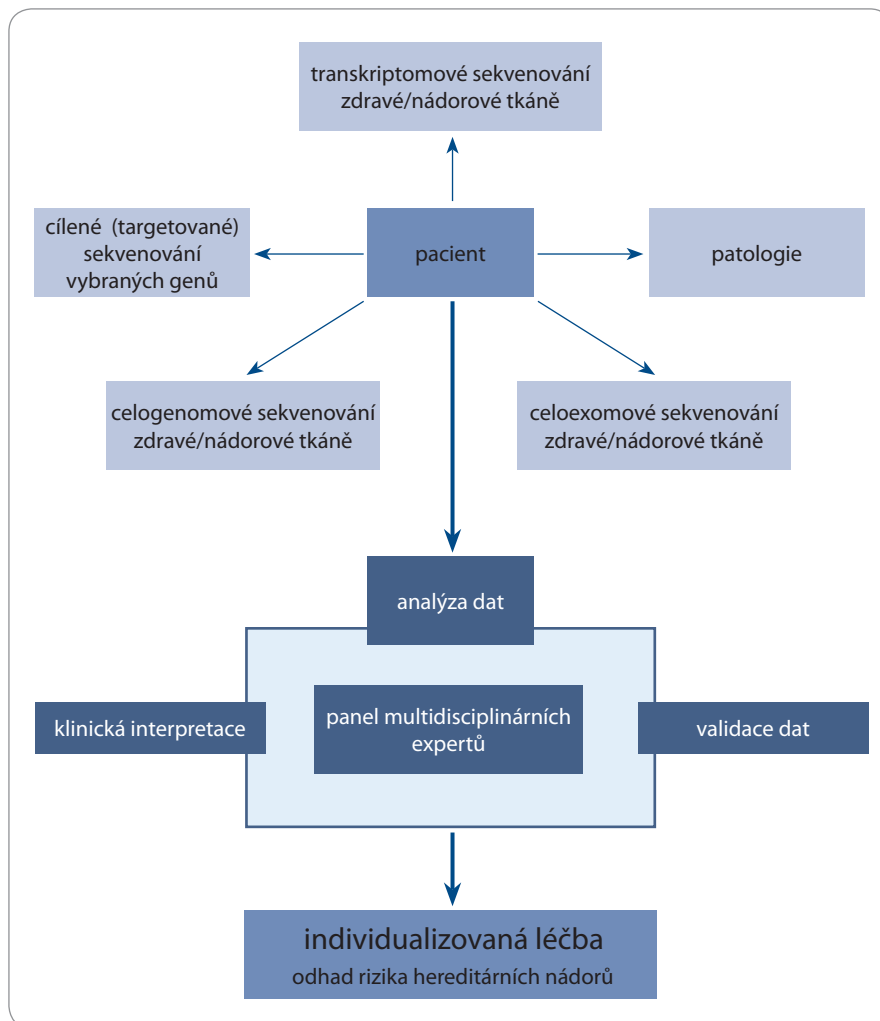
U targetovaného sekvenování existují dnes tři rozdílné způsoby přípravy knihovny – multiplex PCR, single-plex PCR a tzv. targeted capture (cílený „zachyt“) s následnou multiplex PCR. Technika multiplex PCR je reprezentována AmpliSeq technologií, kterou nabízí Life Technologies. Je možné využít až 6 144 párů primerů, jež se smíchají v jedné PCR zkumavce a umožní selektivní amplifikaci oblasti našeho zájmu. AmpliSeq technologie vyžaduje pouze 10 ng vstupní DNA z různých typů nádorových vzorků včetně FFPE. Nedávno byla tato technologie upravena i pro amplifikaci celého lidského exomu. K tomuto účelu byl navržen Ion AmpliSeq Exome kit s 294 000 páry primerů [25]. Obohacení vzorku založené na single-plex PCR představuje technika zvaná Microdroplet PCR od RainDance Technologies. Umožňuje vysoce efektivní simultánní amplifikaci až 4 000 vybraných sekvencí na jeden vzorek. Další technikou přípravy amplikonové knihovny je Access Array System od společnosti Fluidigm. Tato technologie využívá metodu, která umožňuje amplifikovat 48 různých vzorků s použitím až 48 rozdílných párů primerů. Poslední technika přípravy amplikonové knihovny využívá specifických sond pro zachyt oblastí našeho zájmu a následného obohacení vzorku o tyto oblasti pomocí PCR. Tyto hybridní techniky jsou reprezentovány technologiemi TrueSeq Amplicon od společnosti Illumina, HaloPlex od Agilent Technologies a SeqCap EZ technology od společnosti

Roche NimbleGen. Jak Microdroplet PCR, tak Access Array System jsou vhodné zejména pro sekvenování několika genů ve velkém množství vzorků. V opačném případě (velké množství genů v malém počtu vzorků) jsou tyto techniky neekonomické. Obě technologie vyžadují také přídatné specializované vybavení, což představuje jejich další nevýhodu. Klinickému využití nejlépe vyhovují techniky AmpliSeq, TruSeq Amplicon, HaloPlex a SeqCap EZ System, jelikož jsou vhodné pro sekvenování stovek genů v malém, ale i větším počtu vzorků. Příprava knihovny zabere pouze jeden den. Senzitivita a specifita je srovnatelná.

Lze si vybrat ze standardních kitů navržených např. i pro záchyt nejběžnějších mutací u onkologických pacientů – tzv. nádorové panely – nebo si navrhnout projekt dle oblastí vlastního zájmu. Společnost Ambry Genetics vyšla vstříc klinickým potřebám a vyvinula několik diagnostických kitů pro záchyt hereditárních nádorových onemocnění za využití NGS technologií. Společnost GeneDx se rovněž specializuje na testování vzácných hereditárních poruch a nabízí komplexní panel 29 genů významně asociovaných s vrozenými nádorovými onemocněními nebo panely zaměřené na jednotlivé nádorové typy.

Využití NGS v onkologické praxi

Pomocí sekvenování nové generace byly identifikovány nové genetické změny přispívající k onkogenezi, nádorové progresi i metastazování. Významných pokroků v identifikaci nových s nádory asociovaných genů bylo dosaženo u nádorů mléčné žlázy, ovaria, tlustého střeva a konečníku, plic, jater, renálního karcinomu, u nádorů hlavy a krku, melanomu a akutní myeloidní leukémie [17]. NGS objevy nových s nádory asociovaných genů jsou shromažďovány na celosvětově dostupných portálech. Například National Cancer Institute (NCI) podpořil úspěšný projekt vytvoření atlasu nádorových genů – The Cancer Genom Atlas (TCGA) za účelem zlepšit nádorovou prevenci, včasnou detekci a léčbu. Další projekt z oblasti nádorové genomiky podporovaný International Cancer Genome Consortium (ICGC) je Cancer Genome Projekt. Ten si klade za cíl komplexně



Obr. 1. Shrnutí workflow integrace omických dat.

U daného pacienta je osekvenován nádorový a zdravý genom. Genetická informace je analyzována a interpretována multidisciplinárními experty. Pacientovi je navržena léčba „šitá na míru“. Z analýzy profituje také pacientova rodina, protože se zjistí i hereditární riziko vzniku nádorového onemocnění a mohou být provedena příslušná opatření [16].

charakterizovat genomické, transkriptomické a epigenomické změny u 50 různých nádorů nebo jejich subtypů, které mají klinickou i společenskou důležitost.

WGS poprvé ukázalo svůj diagnostický potenciál u pacienta s nejednoznačnou diagnózou akutní myeloidní leukemie neznámého typu. Sekvenováním byla zjištěna fúze *PML-RARA* genů potvrzující diagnózu akutní promyeloidní leukemie vedoucí ke změně terapie a zlepšení prognózy [16]. Walsh et al [26] pomocí cíleného přístupu detekovali 21 nových mutací asociovaných s hereditárními nádory mléčné žlázy a vaječníku. Testování *BRCA* genů se standardně provádí pomocí PCR a následného Sangerova sek-

venování produktů. Větší exonové delece a duplikace se dodatečně testují pomocí MLPA (multiplex ligation-dependent probe amplification). Dodatečné testování je avšak omezeno pouze na známé variace. Díky tomu je výsledek testu u mnoha pacientů s hereditární dispozicí negativní. NGS studie provedená mezi 300 rodinami s hereditárním výskytem rakoviny prsu odhalila předtím nedetekované mutace u 52 probandů [16]. Na základě výsledků této studie byla vyvinuta metoda kombinující amplifikaci dlouhých úseků pomocí PCR a jejich NGS sekvenaci pro detekci mutací v genech *BRCA1* a *BRCA2* [27]. Technologie sekvenování nové generace je díky vývoji

malých stolních sekvenátorů dostupná i pro klinické aplikace. Je vhodná především pro sekvenování vybraných panelů genů. Díky tomu je analýza a interpretace dat méně náročná. Kombinace všech NGS technik poskytuje globální pohled na nádorový genom. Onkologická praxe stojí na pokraji doby, kdy budou mít lékaři k dispozici osekvenovaný nádorový/zdravý genom pacienta dříve, než rozhodnou o způsobu léčby. Na obr. 1 je shrnuto workflow takové integrace genomických, transkriptomických a epigenomických dat.

Závěr

Zavedení sekvenačních metod nové generace do klinické praxe je stále limitováno několika faktory. První z nich je cena celogenomového sekvenování. Cíl, za kterým se ubírá vývoj NGS technologií, je dosažení ceny 1 000 \$ za jeden kompletní sekvenační běh s průměrným pokrytím 30krát. Zdá se, že nový sekvenátor společnosti Illumina nedávno uvedený na trh bude schopen dosáhnout cenového limitu 1 000 \$. Za jak dlouho bude využitelný pro rutinní klinickou praxi, zůstává otázkou. Druhý limitující faktor je obrovské množství generovaných dat, pro která je nutné nalézt způsob uchování pro pozdější analýzu a interpretaci výsledků. Náklady plynoucí z uchování a interpretace tak velkého množství dat jsou mnohdy vyšší než náklady na vlastní sekvenování, což si malé diagnostické laboratoře nemohou dovolit. Další komplikací je absence standardu pro určení kvality sekvenačních dat. Je známo, že každá z platform NGS má specifický typ chyb. Správně rozlišit mezi genetickou variantou a chybou sekvenační platformy je nezbytné pro následné rozhodování o typu léčby. Rozdílné bioinformatické strategie se také mohou odrazit na výsledku analýzy

dat. Kromě jiného vyvstává také etická otázka nakládání s NGS daty a jejich zpětného předávání pacientům. Realita NGS dat je zatím taková, že klinicky a biologicky důležitá informace zůstává často pohřbena v obrovském množství falešně pozitivních či negativních výsledků. Zavedení NGS technologie do klinické praxe má tedy před sebou ještě četné překážky. K uvedení do klinické praxe je nyní nejbližší genetické testování hereditárních syndromů. Navzdory všem výše popsaným problémům poskytuje NGS nebyvalé příležitosti pro studium molekulární podstaty nádorových onemocnění.

Literatura

1. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* 1944; 79(2): 137–158.
2. Watson JD, Crick FH. The structure of DNA. *Cold Spring Harb Symp Quant Biol* 1953; 18: 123–131.
3. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977; 74(2): 560–564.
4. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74(12): 5463–5467.
5. 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.
6. Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* 2010; 11(1): 31–46. doi: 10.1038/nrg2626.
7. Zhou X, Ren L, Meng Q et al. The next-generation sequencing technology and application. *Protein Cell* 2010; 1(6): 520–536. doi: 10.1007/s13238-010-0065-3.
8. Shokralla S, Spall JL, Gibson JF et al. Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* 2012; 21(8): 1794–1805. doi: 10.1111/j.1365-294X.2012.05538.x.
9. Xuan J, Yu Y, Qing T et al. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 2013; 340(2): 284–295. doi: 10.1016/j.canlet.2012.11.025.
10. Quince C, Lanzen A, Curtis TP et al. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* 2009; 6(9): 639–641. doi: 10.1038/nmeth.1361.
11. Illumina.com [homepage on the Internet]. Illumina, Inc.; c2014 [updated 2014; cited 2014 Mar 19]. Available from: <http://www.illumina.com/systems.ilmn>.
12. Nanoporetech.com [homepage on the Internet]. Oxford Nanopore Technologies, Oxford, UK;

c2008–2014 [updated 2014; cited 2014 Mar 17]. Available from: <https://www.nanoporetech.com/technology/introduction-to-nanopore-sensing/introduction-to-nanopore-sensing>.

13. Di Fiori N, Squires A, Bar D et al. Optoelectronic control of surface charge and translocation dynamics in solid-state nanopores. *Nat Nanotechnol* 2013; 8(12): 946–951. doi: 10.1038/nnano.2013.221.
14. Squires AH, Hersey JS, Grinstaff MW et al. A nanopore-nanofiber mesh biosensor to control DNA translocation. *J Am Chem Soc* 2013; 135(44): 16304–16307.
15. Anderson BN, Muthukumar M, Meller A. pH tuning of DNA translocation time through organically functionalized nanopores. *ACS Nano* 2013; 7(2): 1408–1414. doi: 10.1021/n3051677.
16. Guan YF, Li GR, Wang RJ et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer* 2012; 31(10): 463–470. doi: 10.5732/cjc.012.10216.
17. Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. *Biol Proced Online* 2013; 15(1): 4. doi: 10.1186/1480-9222-15-4.
18. Ku CS, Cooper DN, Ziogas DE et al. Research and clinical applications of cancer genome sequencing. *Curr Opin Obstet Gynecol* 2013; 25(1): 3–10. doi: 10.1097/GCO.0b013e32835af17c.
19. Ley TJ, Mardis ER, Ding L et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008; 456(7218): 66–72. doi: 10.1038/nature07485.
20. Dong H, Wang S. Exploring the cancer genome in the era of next-generation sequencing. *Front Med* 2012; 6(1): 48–55. doi: 10.1007/s11684-012-0182-x.
21. Jones S, Hruban RH, Kamiyama M et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 2009; 324(5924): 217. doi: 10.1126/science.1171202.
22. Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F et al. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet* 2011; 43(7): 663–667. doi: 10.1038/ng.861.
23. Yokoyama S, Woods SL, Boyle GM et al. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature* 2011; 480(7375): 99–103. doi: 10.1038/nature10630.
24. Tran B, Dancy JE, Kamel-Reid S et al. Cancer genomics: technology, discovery, and translation. *J Clin Oncol* 2012; 30(6): 647–660. doi: 10.1200/JCO.2011.39.2316.
25. Chang F, Li MM. Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genet* 2013; 206(12): 413–419. doi: 10.1016/j.cancergen.2013.10.003.
26. Walsh T, Lee MK, Casadei S et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci USA* 2010; 107(28): 12629–12633. doi: 10.1073/pnas.1007983107.
27. Ozelik H, Shi X, Chang MC et al. Long-range PCR and next-generation sequencing of BRCA1 and BRCA2 in breast cancer. *J Mol Diagn* 2012; 14(5): 467–475. doi: 10.1016/j.jmoldx.2012.03.006.