

Univerzita Karlova v Praze
Přírodovědecká fakulta

Biologie

Biologie obecná



Žaneta Chmelařová

*Sekvenování nové generace, jeho principy a aplikace pro
studium lidských nádorových onemocnění*

Principles of the next gen sequencing and its application for studies of human malignant diseases

Bakalářská práce

Školitel: MUDr. Petra Kleiblová, Ph.D.

Praha, 2015

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 12.05.2015

Žaneta Chmelařová

Ráda bych poděkovala vedoucí mé bakalářské práce MUDr. Petře Kleiblové, Ph.D. za čas, po který se mi trpělivě věnovala a za poskytnutí odborných konzultací a Doc. MUDr. Zdeňku Kleiblovi, Ph.D. za cenné připomínky a rady. Dále bych chtěla poděkovat mé rodině za důvěru a podporu, kterou mi projevovala po celou dobu mého studia.

Abstrakt

Moderní technologie, v čele se sekvenováním nové generace (NGS), mění řádově rychlost získávání genetických dat při odhalování genetických příčin vyvolávajících různá lidská onemocnění. Tím umožňují také změnu přístupu ke studiu a léčbě lidských nádorových onemocnění, která můžeme z řady důvodů chápat také jako onemocnění genetická.

Práce si klade za cíl zmapovat vývoj sekvenačních metod DNA se zaměřením na NGS. Popisuje základní poznatky vedoucí ke vzniku současných sekvenačních postupů a jejich chemické, fyzikální a biologické principy. Ve stručnosti uvádí přehled způsobu zpracování získaných sekvenačních dat, jejich analýzy a využití postupů, které slouží pro predikce významu nalezených sekvenčních variant. Na vybraných příkladech demonstruje hlavní aplikace NGS v onkogenetické problematice, mezi které patří charakterizace somatických genetických změn u sporadických nádorů a identifikace hereditárních patogenních variant v nádorových predispozičních genech u nádorů dědičných. Tím ukazuje, jak jsou poznatky a použití metod NGS zaváděny do klinického výzkumu a lékařské praxe.

Klíčová slova: DNA, PCR, sekvenování nové generace, třetí generace sekvenování, genové mapování, genová anotace, dědičné nádory, nádorová onemocnění

Abstract

Modern technologies, including the next-gen sequencing (NGS), increase the speed for genetic data acquisition in analyses of the genetic susceptibility to various human pathologies. This, for example, entirely changes the field of tumor biology research and treatment because malignant tumor could be also considered as a genetic disease.

This work aims to describe development of the DNA sequencing techniques with a focus on NGS. We summarize the basic scientific achievements stimulating the NGS development and its chemical, physical, and biological principles. Further, we overview the main steps in sequencing data acquisition and analysis, and principles used by the software tools that are required for the prediction of functional importance in identified genetic variants. The importance of NGS for current oncogenetics is illustrated using selected examples, including characterization of somatic genetic changes in sporadic tumors and identification of germinal pathogenic variants in case of hereditary cancer syndromes, that demonstrate the rapid implementation of NGS technologies into the clinical practice.

Key words: DNA, PCR, next-gen sequencing, 3rd generation sequencing, gene mapping, gene annotation, hereditary tumors, cancer

Obsah

Seznam zkratk	2
1 Genetické a genomické charakteristiky nádorových onemocnění	3
2 Metody sekvenování DNA	5
2.1 Historická poznámka	5
2.2 Sangerovo sekvenování.....	7
2.3 Sekvenační metody nové generace	9
2.3.1 Příprava sekvenačního templátu.....	9
2.3.1.1 Emulzní PCR.....	10
2.3.1.2 Můstková (bridge) PCR.....	10
2.3.2 Principy sekvenování nové generace a způsoby detekce pořadí nukleotidů	11
2.3.2.1 Pyrosequencing (pyrosekvenování)	11
2.3.2.2 Illumina Genome Analyzer	14
2.3.2.3 ABI SOLiD (Sequencing by Oligo Ligation and Detection)	15
2.3.2.4 Semiconductor sequencing technology	16
2.4 Třetí generace sekvenování	17
2.4.1 Helicos Biosciences	18
2.4.2 Technologie SMRT (Single - Molecule Real - Time).....	19
2.5 Technické charakteristiky a srovnání sekvenačních přístupů	19
3 Zpracování a analýza sekvenačních dat.....	21
3.1 Skládání sekvenačních dat - mapování (assembly)	22
3.2 Anotování sekvenačních dat	23
3.3 Predikce významu sekvenačních variant.....	24
4 Příklad využití technologií NGS pro studium lidských nádorů.....	25
4.1 NGS při analýze hereditárních mutací.....	25
4.2 NGS při analýze somatických mutací	26
5 Závěr.....	27
6 Seznam použité literatury	28
7 Elektronické zdroje:	31

Seznam zkratek

APS	-	3'-fosfoadenosin-5' fosfosulfát
ATP	-	adenosintrifosfát
BAM	-	binary alignment map
BRCA2	-	breast cancer gene 2
CCD	-	charge coupled device
CMOS	-	complementary metal-oxide semiconductor
ddNTP	-	dideoxyribonukleotid
DNA	-	deoxyribonukleová kyselina
DNA polymeráza	-	DNA dependentní DNA polymeráza
dNTP	-	deoxyribonukleotid
dsDNA	-	double strand (dvouvláknová) DNA
emPCR	-	emulzní polymerázová řetězová reakce
NGS	-	next generation sequencing (sekvenační metody nové generace)
PALB2	-	partner and localizer of BRCA2
PCR	-	polymerase chain reaction (polymerázová řetězová reakce)
PP _i	-	pyrofosfát
SAM	-	sequence alignment map
SMRT	-	single molecule real time sequencing
SOLiD	-	sequencing by oligo ligation and detection
ssDNA	-	single strand (jednovláknová) DNA
ZMV	-	zero mode waveguides

1 Genetické a genomické charakteristiky nádorových onemocnění

Nádorová onemocnění jsou významným zdravotnickým a celospolečenským problémem především v rozvinutých státech. V České republice jsou druhou nejčastější příčinou úmrtí. Více než tři čtvrtiny všech maligních tumorů u nás tvoří karcinomy prsu, tlustého střeva a konečníku, plic a prostaty. Ročně je v České republice přes 67 000 osob s nově diagnostikovaným zhoubným nádorovým onemocněním (www.svod.cz; www.uzis.cz).

Nádorová onemocnění se většinou vyskytují ve formě sporadické a vznikají na základě akumulace somatických mutací v průběhu života postižené osoby v určité populaci buněk. Somatické mutace genomové DNA vznikají z řady exogenních i endogenních příčin. Exogenní genotoxická poškození jsou důsledkem působení činitelů chemických (kancerogeny), fyzikálních (např. ionizující záření) či biologických (onkogenní viry). Mezi nejčastější endogenní příčiny vzniku mutací jsou řazeny poruchy biologických pochodů zodpovědných za syntézu DNA při její replikaci a opravách. Mutace, které mají významné pronádorové účinky, postihují především geny, které patří do rozsáhlých funkčních skupin protoonkogenů, tumor supresorových genů a genů kódujících proteiny zúčastněné v opravách genomové DNA. Rozlišení "příčinných" mutací (tzv. "driverů") přímo zodpovědných za vznik a rozvoj nádorového onemocnění od fenotypově "němých" mutací (tzv. "passengerů") je obtížným úkolem současné onkogenomiky (Stratton *et al.* 2008). Identifikace příčinných mutací v různých stádiích vývoje nádorového ložiska a jeho metastáz je zásadní nejen pro pochopení procesu tumorogeneze, ale je i významným faktorem pro určení prognózy onemocnění u konkrétního pacienta a důležitým předpokladem (prediktorem) pro odhad léčebné odpovědi.

Kromě většinou se vyskytujících nádorů sporadických existuje relativně malá, ale klinicky významná skupina nemocných, u kterých se na vzniku onkologického onemocnění významně podílí i genetická složka, tedy mutace v některém z více než 150 doposud charakterizovaných predispozičních genů (Rahman 2014). Identifikace dědičné nádorové predispozice má zásadní klinický význam pro pacienty a jejich rodinné příslušníky.

Charakterizace mutací u nádorových onemocnění má svoje významná úskalí. V případě studia somatických mutací je obtížným problémem nejen identifikace příčinných mutací (již zmíněná otázka drivers vs. passengers), ale také rovněž heterogenita nádorových vzorků. Ve vyšetřovaném okrsku tkáně se prolínají různé populace nádorových buněk (a tím i různé genomové DNA s různými variantami), buněk nenádorových a buněčných struktur nádorového stromatu (tvořeného především buňkami imunitního systému a mesenchymu). Při identifikaci dědičných mutací

v nádorových predispozičních genech je pak zásadní komplikací rozhodování o významu nalezených variant a jejich příčinnému vztahu k nádorovému onemocnění. Až na naprosté výjimky, se predispoziční mutace pro vznik nádorových onemocnění vyznačují neúplnou penetrancí¹.

Použití metod sekvenování nové generace přináší do studia změn genomů nádorových buněk a do analýzy nádorové predispozice zcela nový rozměr a významně přispívá k zásadnímu pokroku, který byl započat jejich zařazením do klinického výzkumu.

¹ Jako penetranci označujeme míru pravděpodobnosti, že u nosiče mutace dojde k očekávané změně fenotypu, tedy že onemocní sledovaným nádorovým onemocněním. Penetrance predispozičních variant pro vznik dědičných nádorových syndromů se pohybuje v širokém rozmezí (0 - <1). Čím vyšší je penetrance, tím klinicky významnějším nálezem je průkaz nosičství příslušné mutace.

2 Metody sekvenování DNA

2.1 Historická poznámka

Lidstvo po staletí pátralo po důvodu, proč se některé fenotypy (příp. onemocnění) přenášejí v rodinách do dalších generací. S revolučním vysvětlením příčiny dědičnosti některých fenotypových znaků přišel v 2. pol. 19. století Johann Gregor Mendel, když sledoval sedm znaků hrachu setého. Svým pozorováním a následným použitím matematických výpočtů zjistil, že se nedědí znaky, ale vlohy pro ně. Své poznatky shrnul do tří zákonů dědičnosti a tato pravidla (s danými výjimkami) platí pro nezávisle segregované geny dodnes (Nussbaum *et al.* 2004). Dědičnost, jako faktor vzniku nádorových onemocnění, je běžně uváděna v řadě literárních odkazů (včetně českých) již od poloviny 19. století. Klasicky je citována práce Pierra Brocy, ve které v roce 1866 popisuje rodinu s mnohočetným dědičným výskytem karcinomu prsu v několika generacích (Krush, 1979), přestože činitel zodpovědný za vznik těchto onemocnění nebyl v té době ani tušen.

Teprve koncem 19. století Švýcar Friedrich Miescher objevil při pokusu izolování proteinů z bílých krvinek nebílkovinnou substanci bohatou na fosfor, a nazval ji „nuclein“ (později DNA). Od té doby se genetikou začalo zabývat více vědců, kteří tak svými objevy značně rozšířili poznatky o podstatě dědičnosti. Mezi lety 1900 - 1920 objasnil Phoebus Levene, že základními stavebními kameny DNA jsou cukr-fosfátová kostra a čtyři dusíkaté báze a zavedl termín nukleotid (Dahm, 2004). Mezi další významné objevy patří například práce Thomase Hunta Morgana o lineárním uspořádání genů na chromozomech a významná zjištění o genové vazbě (Morgan 1914). Dalším mezníkem v oboru genetiky můžeme nazvat experiment Avery – MacLeod- McCarty (publikovaný roku 1944), který dokazoval, že změna virulence bakterie *Streptococcus pneumoniae* je způsobena přejímáním informace mezi různě virulentními kmeny pneumokoků (tzv. transformace), a že za tento přenos je zodpovědná DNA, nikoli protein, jak se doposud předpokládalo (Avery *et al.* 1943). V tomto období se o DNA začal zajímat také americký vědec Erwin Chargaff. Z různých organismů izoloval DNA a smísením s kyselinou rozrušil fosfodiesterové vazby mezi jednotlivými nukleotidy. Množství jednotlivých nukleotidů následně kvantifikoval pomocí papírové chromatografie. Tímto experimentem prokázal a roku 1952 publikoval, že v DNA je relativní zastoupení adeninových a thyminových bází velmi podobné, ne-li stejné, a zároveň relativní zastoupení cytosinových bází je

velmi podobné zastoupení guaninových bází. Toto zjištění se později do povědomí veřejnosti dostalo jako tzv. první Chargaffovo pravidlo.²

Sekundární strukturu DNA popsali v dubnu roku 1953 Francis Crick a James Watson. Ve své práci podpořené experimentem měření difrakce DNA metodou rentgenové krystalografie ukázali, že DNA je dvoušroubovice, která se skládá ze dvou antiparalelních řetězců obtočených kolem jedné osy. Cukr-fosfátová kostra polynukleotidu je složena ze zbytků kyseliny fosforečné a 2'-deoxy-D-ribózy, na kterou jsou navázány dusíkaté báze pomocí $\beta - N -$ glykosidové vazby. Nukleotidy v jednom řetězci jsou mezi sebou spojeny kovalentní fosfodiesterovou vazbou, která se vytváří mezi volnou 3' alkoholovou skupinou ribózy a 5' fosfátovou skupinou dNTP. Antiparalelně uspořádané řetězce jsou ve dvoušroubovici stabilizovány pomocí vodíkových můstků vytvářených mezi dusíkatými bázemi. Přitom ve Watson-Crickově modelu platí pravidlo, že purinová báze páruje pouze s pyrimidinovou a naopak (Watson *et al.* 1953). K objevu přispěla svými poznatky i Rosalind Franklinová a Maurice Willkins. Za objasnění struktury DNA obdržela trojice Watson, Crick, Willkins v roce 1962 Nobelovu cenu. Díky těmto vědcům dnes daleko lépe rozumíme molekulární genetice a jsem schopni dělat velké pokroky ve zkoumání a chápání lidského genomu.

Po zjištění struktury DNA v roce 1953 bylo velké úsilí věnováno nalezení postupu pro odhalení přesného pořadí nukleotidů v řetězci. První úspěšný pokus určení primární struktury DNA provedl biolog Ray Wu v roce 1970, když se mu podařilo na Cornellské univerzitě osekvenovat krátký úsek genomu bakteriofága lambda (Wu 1970). Jako první použil strategii nasednutí specifického primeru na jednovláknovou DNA a následné syntézy komplementárního vlákna s využitím DNA-dependentní DNA polymerázy a specificky značených nukleotidů. Tím nastínil možnou cestu analyzování genomu.

U procesu sekvenování se využívá vlastnosti replikace vybraného úseku DNA, katalyzované specifickými enzymy. Klíčovou úlohu má v procesu replikace enzym DNA-dependentní DNA polymeráza (DNA polymeráza). Ke správné činnosti DNA polymerázy je potřeba dodat nebo pomocí enzymu primázy nasyntetizovat krátkou oligonukleotidovou sekvenci komplementární k templátu, která slouží jako počáteční místo replikace, tzv. primer. Tento oligonukleotid poskytne volný 3'-OH konec, který je nezbytný pro navázání dalšího nukleotidu fosfodiesterovou vazbou, již pomocí

² Chargaffovo druhé pravidlo říká, že procentuální zastoupení poměru A -T , C -G bází platí pro obě vlákna a liší se mezi druhy.

enzymu DNA polymerázy. Sangerovo sekvenování i metody sekvenování nové generace (s výjimkou ligačního sekvenování) DNA polymerázu vyžadují k replikaci informace o sekvenci z templátu. Takový přístup sekvenování je nazýván "sekvenování pomocí syntézy"(Chen 2014). Některé typy DNA polymeráz (např. DNA polymeráza I) jsou schopny rozkládat řetězec DNA - mají exonukleázovou aktivitu a jsou schopny degradovat vlákno ve směru 3'→5' (proofreading). Tato aktivita byla popsána také u proteinu zvaného Klenowův fragment, který byl původně získán proteolytickým štěpením DNA polymerázy I *E. coli* (Jacobsen *et al.* 1974) a který byl využit v Sangerově metodě, protože byl vhodný pro inkorporování 2', 3'- dideoxynukleotidů (Sanger *et al.* 1977). S rozvojem sekvenování se v technikách NGS stal kvůli velkým modifikacím terminálních nukleotidů neúčinným, a byl proto nahrazen jinými typy polymeráz (Chen 2014).

Ještě v 70. letech 20. století byly vyvinuty dvě sekvenační metody (společně oceněné Nobelovou cenou v roce 1980). Metoda Sangerova, kterou můžeme označit jako metodu biochemickou a metoda Maxam-Gilbertova, která je metodou spíše chemickou.³ Sangerova metoda se ve své automatizované podobě užívá v molekulárně-biologických laboratořích dodnes.

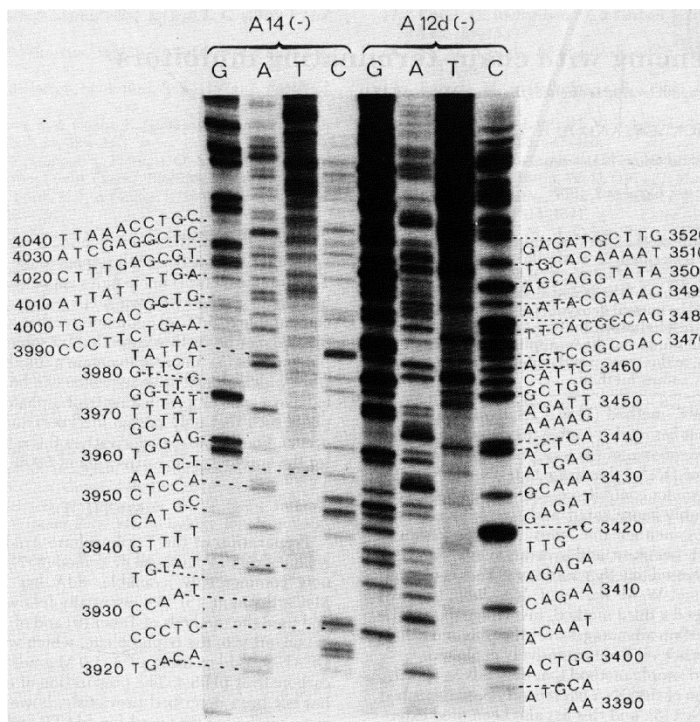
Velký průlom v sekvenování DNA nastal v roce 1983, kdy Kary Mullis objevil způsob, jakým lze efektivně namnožit úsek vybraný DNA *in vitro*. Principem je použití termostabilní polymerázy izolované z bakterie *T. aquaticus* (Taq polymeráza), která si udržuje polymerázovou aktivitu i přes opakovanou inkubaci za vysoké teploty, jež slouží pro denaturaci po jednotlivých cyklech *in vitro* replikace DNA (Saiki *et al.* 1988). Tuto metodu nazval polymerázová řetězová reakce (PCR; Polymerase Chain Reaction) a o deset let později byl za ni oceněn Nobelovou cenou.

2.2 Sangerovo sekvenování

Frederick Sanger vyvinul metodou založenou na principu enzymové replikace DNA *in vitro*. Reakční směs se skládá z templátu DNA, DNA-dependentní DNA polymerázy, primerů značených ³²P, 2'-

³ Principem Maxam-Gilbertova sekvenování je degradace předem určené nukleotidové báze pomocí specifického chemického štěpení (purinové báze jsou štěpeny dimethyl-sulfátem, pyrimidinové hydrazinem). Oligonukleotidy označené radioaktivním ³²P se po štěpení rozdělí na různě dlouhé fragmenty, přenesou se na polyakrylamidový gel a proběhne elektroforéza. Následně jsou pomocí autoradiografie vidět na elektroforeogramu radioaktivní proužky. Tyto proužky představují jednotlivé úseky DNA lišící se o jeden nukleotid (Maxam, Gilbert 1976). Vzestupným čtením těchto proužků můžeme složit maximálně 100 bází dlouhou sekvenci. Tato metoda se však kvůli svým omezením z hlediska velikosti čtení bází, závislosti na velikosti elektroforetického gelu, práci s radioaktivitou a nízkém uživatelském komfortu stala slepou větví sekvenování.

deoxyribonukleotidů (dNTPs) a asi 10 % 2',3' - dideoxyribonukleotidů (ddNTPs; Sanger *et al.* 1977). Po denaturaci dsDNA dochází k hybridizaci sekvenačního primeru na komplementární vlákno ssDNA. Tento krátký úsek dvouřetězcové DNA rozpoznává DNA-dependentní DNA polymeráza, která následně syntetizuje komplementární vlákno DNA od 3'-OH konce primeru. Při zabudování ddNTP přítomného v reakční směsi (na místo příslušného dNTP) dochází k ukončení syntézy polynukleotidu, protože molekula ddNTP neobsahuje volnou 3'-OH skupinu, potřebnou pro navázání dalšího deoxyribonukleotidu. V původní podobě reakce probíhá ve čtyřech zkumavkách, kdy každá obsahuje jiný typ ddNTP (Sanger *et al.* 1977). Po denaturaci sekvenační reakce byly vzniklé ssDNA zakončené ddNTPs rozděleny elektroforeticky na akrylamidovém gelu. (Sanger a Coulson 1978). Radioaktivně označené fragmenty ssDNA elektroforeticky rozdělené na základě mobility nepřímo úměrné logaritmu délky (v počtu bází) byly zobrazeny pomocí autoradiografie. (Sanger *et al.* 1977) Pomocí Sangerova sekvenování s radioaktivně značenými primery (později značenou reakcí pomocí ^{32}P -dATM nebo ^{33}P -dATP) bylo možné přečíst sekvenci



Obr. 1. Autoradiogram sekvenační reakce z původní práce Sanger *et al.* 1977. Autoradiogram na 12% akrylamidovém gelu pro stanovení sekvence při použití restričních fragmentů A12d a A14 jako primerů komplementárního vlákna ΦX174 . Fragmenty jsou zakončené inhibitory ddGTP, ddATP, ddTTP a araCTP. Kratší fragmenty putují gelem rychleji, tudíž doputují dále. Intenzita barvy proužku je úměrná množství kopií daného fragmentu ve vzorku. DNA sekvence je čtena po proužcích směrem vzhůru. (Převzato z Sanger *et al.* 1977)

o délce do 350bp (Obr. 1). Sangerova metoda byla po dlouhá léta díky své spolehlivosti a jednoduchosti nejpoužívanějším sekvenačním přístupem. Ovšem i přes postupné odstraňování nevýhod (radioaktivní značení nahrazeno fluorescenčním; Smith *et al.* 1986) a modernizaci

(zavedení kapilární elektroforézy), nedokázala uspokojit poptávku po stále se zvyšující sekvenační kapacitě.

S nástupem automatizovaného sekvenování bylo potřeba zjistit kvalitu přístrojové identifikace inkorporovaného nukleotidu (pravděpodobnost, s jakou byl začleněný nukleotid správně pojmenován). Pro tento účel bylo zavedeno tzv. Phred skóre (Q). Phred skóre je definováno jako $Q = -10\log_{10}P$, kde P je pravděpodobnost špatné identifikace přítomné báze. Například pokud máme Phred skóre $Q=10$, máme 90% jistotu, že je zařazena báze správná. Jestliže $Q=20$, pravděpodobnost správného začlenění báze bude již 99%. Vyšší skóre tedy znamená vyšší přesnost identifikace a zároveň nižší pravděpodobnost, že došlo k nesprávnému určení nukleotidu. Maximální Phred skóre nabývá hodnoty 60. Toto skóre je využíváno v různých modifikacích ke kontrole kvality všemi metodami sekvenování dalších generací.

Z hlediska diagnostiky nádorových onemocnění je Sangerovo sekvenování po dlouhá desetiletí "zlatým standardem". Jeho zásadní důležitost byla obzvláště patrná v diagnostice mutací v nádorových predispozičních genech, kde Sangerovo sekvenování sehrává zásadní úlohu pro definitivní potvrzení přítomnosti patogenní varianty. Zásadní je však jeho omezení při hledání somatických mutací ve směsných vzorcích buněk nádorových tkání, kdy klasickým sekvenování obvykle nelze zachytit mutace, které se objevují s alelickou frakcí nižší než 10% (tj. obtížně jsou zachytitelné varianty, pokud je v analyzovaném vzorku tkáně méně než 20% nádorových buněk).

2.3 Sekvenační metody nové generace

Pro řadu aplikací vyžadujících získání zevrubné informace o složení komplexních genetických struktur (celých genů, exomů, transkriptomů či genomů) či skladbě somatických variant v různých často velmi málo zastoupených populacích buněk bylo v nedávné době Sangerovo sekvenování nahrazeno metodami nové generace (NGS; Next-Gen Sequencing). NGS, též nazýváno sekvenováním 2. generace, umožňuje průběh mnoha sekvenačních reakcí současně – tzv. masivní paralelní sekvenování (massive parallel sequencing; Tucker *et al.* 2009).

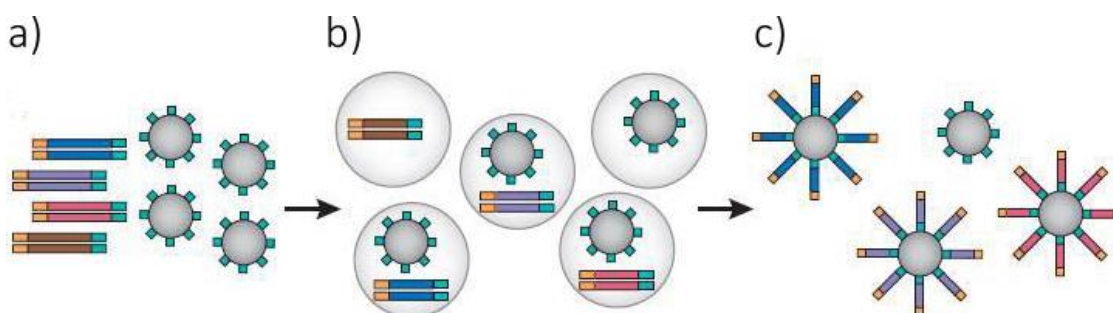
2.3.1 Příprava sekvenačního templátu

Všechny v současnosti zavedené NGS metody umožňují sekvenování sice mnoha, avšak délkově omezených sekvencí DNA. Proto je pro přípravu vzorku k sekvenování vysokomolekulární genomové DNA ve všech systémech nejprve nutno tuto DNA naštěpit (fragmentovat) na kratší

úseky o délce několika desítek páru bází až několika kilobází. Vlákna DNA lze fragmentovat enzymaticky pomocí endonukleáz (například směsí restričních endonukleáz či transponáz) nebo fyzikálními postupy (sonikace, nebulizace). Nebulizací rozumíme rozdělení DNA na fragmenty pomocí působení stlačeného vzduchu v přístroji nazývaném nebulizér. Nejpoužívanější a patrně nejvhodnější metodou pro získání sekvenačních templátů o kontrolované délce je sonikace – mechanická fragmentace dlouhých vláken ultrazvukem.⁴

2.3.1.1 Emulzní PCR

Emulzní amplifikace je příprava klonálně namnoženého sekvenačního templátu pomocí PCR, která probíhá v emulzi voda - olej (Shao *et al.* 2011; Obr. 2). Emulzní PCR využívají pro přípravu sekvenačního templátu platformy 454, SOLiD a Ion torrent. Pro jednotlivé technologie se postup emulzního PCR liší v technických detailech.



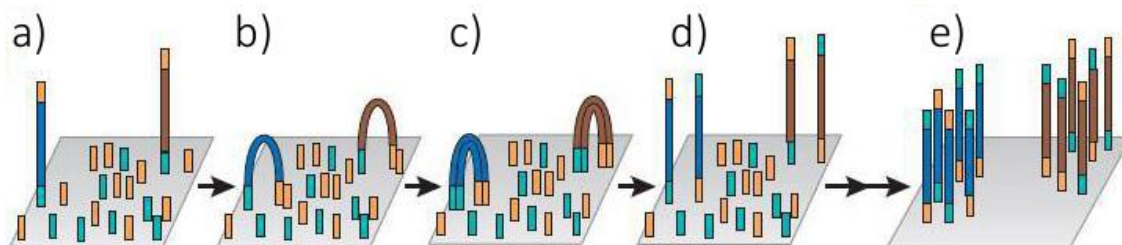
Obr. 2. Schéma emulzní PCR. a) Adaptory opatřené fragmenty DNA jsou ve vhodném poměru přidány do emulze obsahující také mikročástičky s navázaným PCR primerem. **b)** Za ideálních podmínek obsahují vodní kapénky v emulzi pouze jednu mikročástičku s jediným templátem DNA. Následně fragmenty přes adaptory hybridizují k mikročástičkám. **c)** Při PCR tak dochází ke klonální amplifikaci ve „vodních mikroreaktorech“. Po amplifikaci dochází k rozrušení emulze a kuličky nesou kopie fragmentu (Převzato z Shendure *et al.* 2008).

2.3.1.2 Můstková (bridge) PCR

Amplifikace sekvenačního templátu pro platformu Solexa/Illumina nese název můstková PCR nebo také "shluková (cluster) PCR" a slouží ke klonálnímu namnožení templátu na pevné podložce

⁴ Použití specifických (restričních endonukleázy) či nespecifických (např. shearáza) endonukleáz vede u řady sekvencí DNA v důsledku složení její primární struktury k preferenčnímu štěpení určitých úseků.

v ohraničeném prostoru za vzniku tzv. shluků (clusterů). PCR amplifikace využívá univerzálních primerů, které jsou kovalentně vázány na povrchu skleněné destičky (Obr. 3).



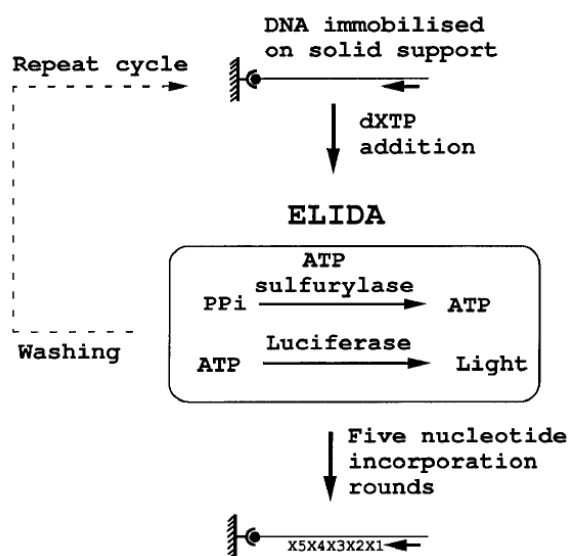
Obr. 3. Schéma můstkové PCR (bridge PCR). **a)** Na oba konce fragmentů jsou ligačně připojeny univerzální adaptory specifické pro 5' a 3' konec PCR templátu.. **b)** Tyto adaptory jsou sekvenčně komplementární (a po denaturaci mohou hybridizovat) k jednovláknovým fragmentům kovalentně připojeným k pevnému povrchu destičky, které v následné PCR slouží jako primery pro dosyntetizování komplementárního vlákna. Následně je templátová DNA odmyta. Kovalentně vázané řetězce DNA jsou ohnuty tak, aby hybridizovaly s přilehlými komplementárními primery. **c)** Ve třetím kroku dochází k amplifikaci druhého vlákna. **d)** Po amplifikaci jsou řetězce znovu denaturovány. **e)** Opakováním tohoto způsobu amplifikace vznikají ohraničené tzv. shluky (clusters) PCR produktů, které jsou sekvenčně identické s původním jedním templátem. V těchto shlucích jsou přítomna forwardová i reverzní vlákna. Reverzní vlákna jsou na počátku následného vlastního sekvenování odštěpena a odmyta. (Převzato z Shendure *et al.* 2008)

2.3.2 Principy sekvenování nové generace a způsoby detekce pořadí nukleotidů

V současné době používané metody NGS jsou založeny na dvou různých principech – sekvenování pomocí syntézy ("sequencing by synthesis") a sekvenování založené na ligaci ("ligation based sequencing").

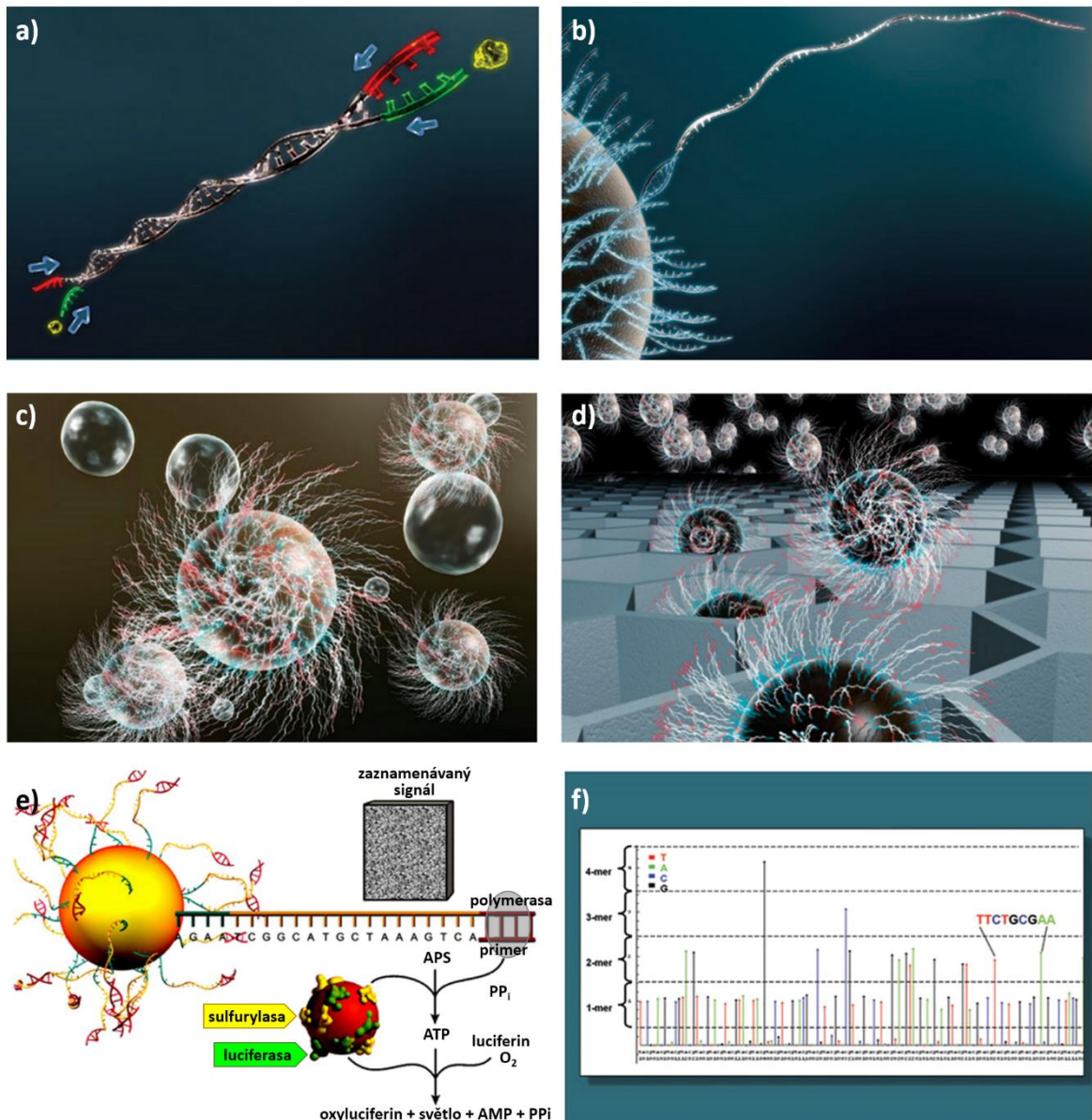
2.3.2.1 Pyrosequencing (pyrosekvenování)

Princip pyrosekvenační detekce vychází z prací Nyréna *et al.* (Nyrén *et al.* 1985; Obr. 4). Sekvenační reakce je založena na detekci pyrofosfátu (PPi), který je uvolňován z trifosfátu při vzniku fosfodiesterové vazby při syntéze polynukleotidu. Množství vznikajícího pyrofosfátu je detekováno spřaženou chemickou reakcí vedoucí ke vzniku luciferinu, který vytváří světelný signál zaznamenávaný citlivou CCD kamerou (Ronaghi *et al.* 1996).



Obr. 4. Schéma principu pyrosequenování v reálném čase. Při zařazení správného nukleotidu k templátu připojeného na mikrokuličku se uvolní úměrné množství PP_i. Ten dále reaguje s 3'-fosfoadenosin-5'-fosfosulfátem (APS) za vzniku ATP v reakci katalyzované ATP sulfurylázou. Vzniklý ATP je využit luciferázou na oxidaci luciferinu na oxyluciferin, a zároveň dochází ke generování světelného signálu. (V případě, že nedojde k inkorporování nukleotidu, jsou nukleotidy degradovány enzymem apyrázou.) Tato reakce se cyklicky opakuje, dokud není přečtena poslední báze templátu. (Převzato z Ronaghi *et al.* 1996)

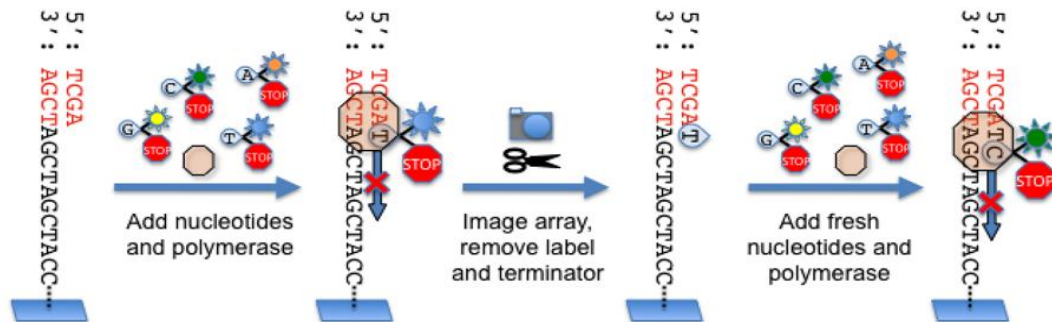
Pyrosequenování je základem první rutinně používané metody NGS uvedené do komerční praxe firmou 454 (Roche; Obr. 5). Při 454 sekvenování dochází k sekvenování DNA templátu ukotveného na mikrokuličce v tzv. pikotitrační destičce, což je plochá destička s mnoha jamkami opatřenými optickým vláknem (picotiter plate; Margulies *et al.* 2005). Jamky jsou používány jako velmi malé zkumavky, ve kterých DNA polymeráza inkorporuje postupně do reakce přidávané dNTPs. Pokud je daný dNTP komplementární k sekvenci templátové DNA, je zařazen do syntetizovaného vlákna a uvolní se patřičný počet molekul PP_i, čímž je hodnocen druh a počet nukleotidy začleněného v konkrétním kroku sekvenační reakce.



Obr. 5. Schéma průběhu pyrosekvenování, které využívá systém 454 (Roche). Části obrázku a - c), které schematicky znázorňují ligaci adaptorů na templátovou DNA, připojení k mikrokuličce a průběh klonální amplifikace templátu pro pyrosekvenování jsou v detailu vysvětleny na Obr. 2. d) Sekvenování. Kuličky se v roztoku nanesou rovnoměrně na destičku. Povrch destičky je pokryt jamkami ve tvaru plástve, které svou velikostí umožňují, aby do každé jamky zapadla právě jedna kulička. Destička je poté vložena do sekvenačního přístroje, kde je celý její povrch postupně promýván jednotlivými nukleotidy. Každé inkorporování komplementárního nukleotidu vede ke vzniku světelného signálu zaznamenaného vysoce citlivou CCD kamerou. e) Chemický průběh pyrosekvenování je podrobně rozepsán v Obr. 4. f) Zpracování dat a analýza. Program využívá intenzitu světelného signálu po každé inkorporaci v každé jamce ke stanovení sekvence všech čtení najednou. Intenzita signálu je úměrná počtu inkorporovaných bází (Ronaghi *et al.* 1996). V případě začlenění více bází najednou (homopolymerů), se objevují problémy s přesností detekovaného signálu, které tento přístup omezují – správně lze hodnotit inkorporaci do max. pěti stejných nukleotidů (Shendure *et al.* 2008; převzato a upraveno z <http://454.com/products/technology.asp>).

2.3.2.2 Illumina Genome Analyzer

Sekvenování firmy Illumina má počátky v pracích Turcattiho *et al.*, ve kterých je popsán způsob připojení DNA ke skleněné destičce s navázanými oligonukleotidy a vytváření shluků amplifikovaného DNA fragmentu (Turcatti *et al.* 2006). Stejně jako pyrosekvenování využívá postupu sekvenování pomocí syntézy. Liší se ovšem v postupu detekce inkorporovaných nukleotidů. Namnožené a ukotvené shluky ssDNA fragmentů, sloužících jako sekvenační templát, jsou po navázání sekvenačního primeru postupně omývány fluorescenčně značenými nukleotidy s modifikovaným 3'-OH koncem (Obr. 6). Tato modifikace zajišťuje, že v každém kroku sekvenace se zařadí pouze jeden nukleotid (Bentley *et al.* 2008). Po začlenění značeného nukleotidu je jím po excitaci emitováno charakteristické fluorescenční záření, které je detekováno systémem CCD kamer (www.illumina.com).

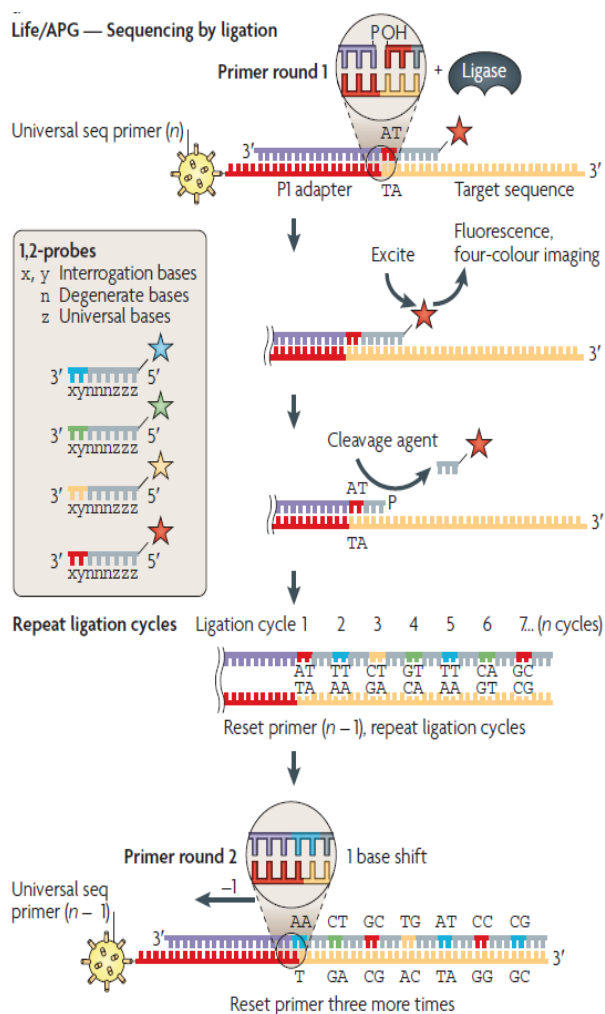


Obr. 6. Schéma průběhu sekvenace na platformě Solexa/Illumina. Na forwardové vlákno templátového clusteru, připevněné k pevné podložce hybridizuje primer a je přidána DNA polymeráza a čtyři typy fluorescenčně značených nukleotidů s modifikovanými 3'-OH konci, které zajišťují v jednom cyklu sekvenace zařazení pouze jednoho nukleotidu (Bentley *et al.* 2008). DNA polymeráza zprostředkuje navázání jednoho nukleotidu. Po jeho inkorporování se po excitaci uvolní pro něj typické fluorescenční záření, které je detekováno. Terminační část modifikující 3'-OH konec a fluorescenční značka jsou odštěpeny a odstraněny, což umožňuje navázání dalšího nukleotidu v dalším kroku syntézy. Tento proces se cyklicky opakuje. Takový postup sekvenace se opakuje také pro reverzní vlákno templátu. Délka čtení je omezena délkou templátu a typem použitého sekvenačního kitu (Převzato a upraveno z Anderson *et al.* 2010).

V současnosti nejvíce užívané platformy byly uvedeny na trh firmou Illumina v roce 2006 (www.illumina.com). Od té doby byly uvedeny na trh ještě další modely lišící se svou sekvenační kapacitou (Mi-Seq, Next-Seq a Hi-Seq).

2.3.2.3 ABI SOLiD (Sequencing by Oligo Ligation and Detection)

Sekvenování využívané platformou SOLiD není založeno na principu syntézy sekvenovaného vlákna pomocí DNA polymerázy, ale na principu ligace (Obr. 7). Při sekvenování jsou na jednořetězcové templátové vlákno ukotvené na mikrokuličkách hybridizovány oligomerní hybridizační sondy obsahující na 5' konci známou sekvenci dimerů označenou specifickými flouorofory. Po sekvenčně specifické hybridizaci těchto sond je DNA ligáza kovalentně spojuje s prodlužujícím se sekvenovaným vláknem (Shendure *et al.* 2005).



Obr. 7. Schéma sekvenování pomocí ligace.

K fragmentované DNA, namnožené pomocí emulzní PCR hybridizuje po denuraci primer komplementární s adaptorovou sekvencí fragmentu.

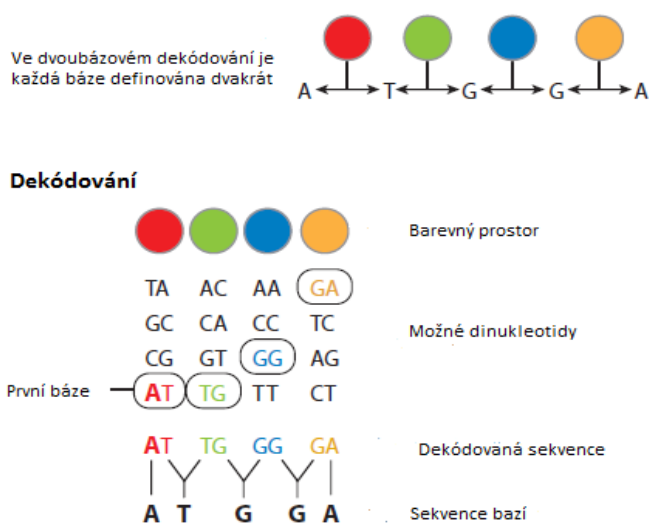
V dalším kroku je pomocí enzymu ligázy připojen komplementární oktamer složený ze dvou bází komplementárních s templátovým řetězcem (interrogation bases; zde AT; xy v levém boxu), tří degenerovaných bází (nnn v boxu vlevo) a tří univerzálních bází (zzz v boxu vlevo), nesoucích fluorescenční značku, která je po začlenění detekována citlivou kamerou.

Univerzální báze jsou poté chemicky odštěpeny (přesný postup štěpení je předmětem patentové ochrany) a ligační proces se opakuje v cyklech až do konce čtení templátu DNA. V druhém kole čtení stejného templátu dochází k odštěpení nově vzniklého vlákna i původního primeru. K templátovému vláknu se hybridizuje primer, který je komplementární k sekvenci o jeden nukleotid posunutý. Tak je zajištěno dibázické čtení sekvence posunutý o jeden nukleotid. Tento cyklus zkrácení primeru, přiligování oktamerů a odštěpení univerzálních nukleotidů se opakuje ještě třikrát. (Převzato a upraveno z Metzker, 2010, review)

V krocích následujících po identifikaci fluoroforu ligačně připojeného oligomeru je fluorofor s několika bázemi chemicky odštěpen tak, aby volná fosfátová skupina na 5' konci zkráceného oligomeru byla připravena na navázání následujícího značeného oligomeru. Tímto způsobem jsou

diskontinuálně čteny vždy dvě báze. Výsledná sekvence je vypočítána na základě sekvenování templátu kotveného na mikrokuličce v několika různých krocích posunutých o jeden nukleotid, čímž dojde k pročení celé sekvence templátu.

Dvoubázové kódování (two base encoding; Obr. 8) je typickou vlastností SOLiD systému. V tomto postupu je díky postupnému přemístování primerů každý cílový nukleotid kódován dvakrát, postupně ve dvou kolech ligace. Tato vlastnost ligačního sekvenování zajišťuje velkou sekvenační přesnost (McKernan *et al.* 2006).



Obr. 8. Schéma dvoubázového barevného kódování sekvenované DNA v systému SOLiD.

Pro dekódování je nutné znát první nukleotid skenovaného úseku (zde A).

Každý následující nukleotid je kódován dvěma fluorofovy.

V případě diskordantního barevného kódování jednoho fluoroforu lze takto detekovat sekvenační chybu a tuto chybu odlišit od sekvenační varianty.

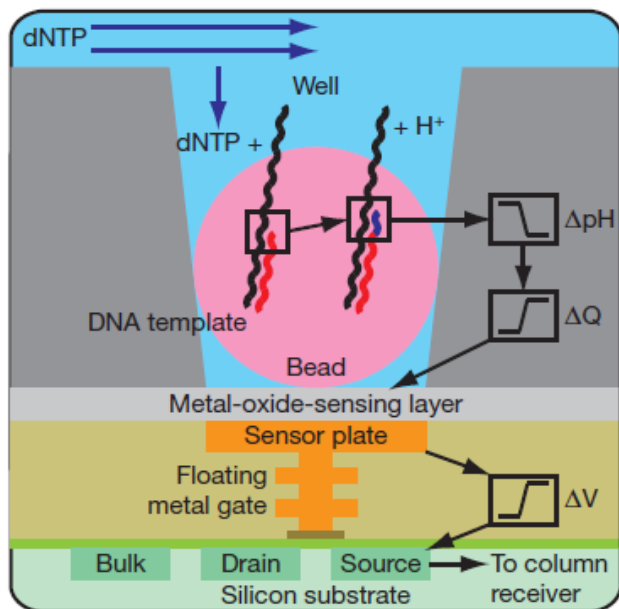
(Převzato a upraveno z Mardis Elaine R., 2008).

Systém SOLiD byl poprvé představen firmou Applied Biosystems (www.appliedbiosystems.com) v roce 2006, avšak v současné době vývoj této technologie již dále neprobíhá.

2.3.2.4 Semiconductor sequencing technology

Sekvenování pomocí polovodičových technologií využívá toho, že začlenění jednoho nukleotidu do polymeru DNA vede k hydrolýze jeho fosfátové skupiny, a tím dojde k uvolnění jednoho protonu (H^+). Polovodičový čip technologie CMOS (Complementary Metal - Oxide Semiconductor) se skládá z vrstvy jamek, navržených tak, aby do každé jamky zapadla přesně jedna 2 μm velká kulička s amplikonem, ion - citlivé vrstvy a vlastního senzoru. Uvolnění protonů má za následek změnu pH roztoku, úměrnou počtu inkorporovaných bází (0,2 jednotek pH za jednu bázi), která je detekována

ion senzitivní vrstvou. Tyto změny pH jsou následně převedeny na napětí a softwarem jsou převedeny do digitální informace. Není zde tedy použito žádného fluorescenčního značení ani modifikovaných terminátorů (Rothberg *et al.* 2011), což zvyšuje rychlost sekvenace.



Obr. 9. Zobrazení jedné jamky polovodičového čipu a schéma průběhu sekvenování. Po průběhu emPCR (Obr. 2) se k mikrokuličkám s navázanými několika miliónovými kopiemi DNA templátu přidají sekvenační primery a DNA polymeráza. Mikrokuličky s navázanými kopiemi DNA templátu plují čipem a pomocí centrifugy je rozmístěna jedna mikrokulička do jedné jamky. Následně jsou sekvenčně promývány neznačenými nukleotidy. Po začlenění nukleotidu dochází k uvolnění protonu (H^+), v důsledku čehož se v jamce změni pH (ΔpH). Pod snímací vrstvou se nachází ion senzitivní tranzistor detekující změny v napětí (ΔV) a ihned převádí chemický signál na digitální. V případě inkorporování více než jedné báze se napětí násobí. (Převzato z Rothberg *et al.* 2011)

Přístroj na sekvenaci pomocí polovodičových čipů uvedla na trh v roce 2010 firma Applied Biosystems.

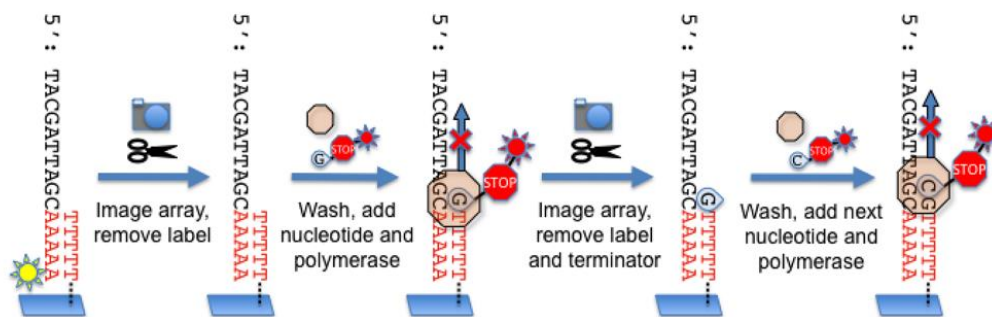
2.4 Třetí generace sekvenování

Přestože technologie sekvenování první a druhé generace zaznamenaly mnoho průkopnických objevů v oblasti genomiky i mimo ni a tím umožnily lepší a úplnější pochopení lidského genomu, jsou zde stále aspekty, které jsou mimo dosah současných technologií. Hlavním nedostatkem sekvenování druhé generace je omezená délka čtení a časová náročnost přípravy biologických vzorků a sekvenační reakce. Zatímco u metod druhé generace se průměrná délka čtení pohybuje v rozmezí 75 - 500 párů bází, některé technologie třetí generace slibují více než 1000 párů bází (Podrobný přehled na Tab. 1, str. 20). Sekvenování třetí generace je také označováno jako technologie "single-molecule" sekvenování. Výhodou je také to, že zde není potřeba amplifikace

templátu. Tím je zachováno i kvantitativní hledisko a (v případě možnosti dlouhého čtení) možnost detekce velkých přestaveb i chybění či amplifikací velkých úseků genomové DNA. Současně se samotné sekvenování výrazně zkracuje a zlevňuje (Schadt *et al.* 2010).

2.4.1 Helicos Biosciences

Helicos Genetic Analysis Platform vychází z teoretické práce Braslavskyho *et al.* 2003, ve které je popsána sekvenace jedné molekuly DNA ukotvené na povrchu krystalu pomocí biotinylovaného primeru.



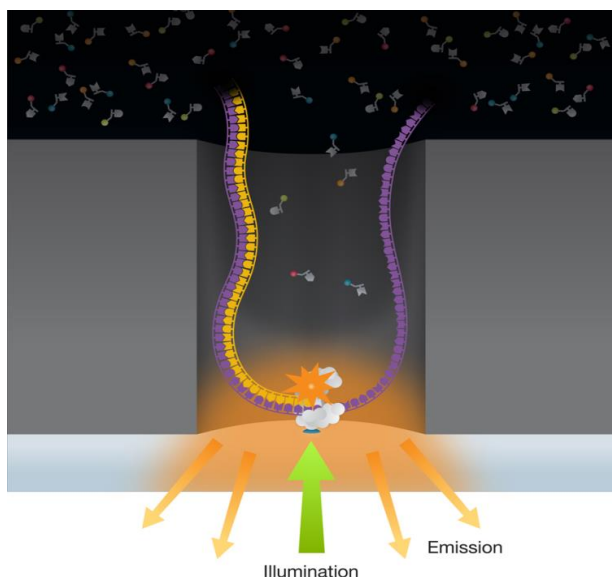
Obr. 10. Průběh sekvenace na sekvenátoru HeliScope. Jednovláknové fragmenty DNA o délce 100 - 200 bp jsou nejprve opatřeny poly - A adaptory, které zajišťují uchycení vláken k poly - T sekvenci primeru, který je ukotven na povrchu destičky, na které probíhá sekvenace. Následně jsou neuspořádaně rozmístěny na destičku, která je cyklicky zaplavována DNA polymerázou a jedním ze čtyř typů fluorescenčně značených nukleotidů. Podobně jako u platformy Illumina, mají jednotlivé nukleotidy modifikované 3'-OH konce, které zabraňují inkorporaci více nukleotidů v jednom cyklu. Každá inkorporace jednotlivých nukleotidů k DNA fragmentům je snímána velice citlivou CCD kamerou. Po zaznamenání signálu jsou fluorescenční značky odštěpeny a odmyty a cyklus se opakuje. (Převzato a upraveno z Anderson *et al.* 2010)

Sekvenace probíhá za přítomnosti DNA polymerázy s využitím fluorescenční mikroskopie (Braslavsky *et al.* 2003). Unikátní vlastností této metody je, že k sekvenování není potřeba žádné klonální amplifikace a ke zjištění původní sekvence dochází použitím velmi citlivého fluorescenčního FRET systému při sekvenaci syntézou jedné DNA molekuly (Shendure *et al.* 2008; Obr. 10).

Firma Helicos Biosciences sekvenátor HeliScope uvedla na trh v roce 2008. V současné době ovšem již nedochází k vývoji této technologie.

2.4.2 Technologie SMRT (Single - Molecule Real - Time)

Technologie SMRT provádí sekvenování jednotlivých molekul v reálném čase. Tento systém využívá nanostrukturu zvanou Zero Mode Waveguide (ZMW), což jsou jamky o průměru asi 70 nm, umístěné v čipu. Každý SMRT čip obsahuje desetitisíce těchto jamek a v každé jamce probíhá sekvenace jedné molekuly DNA (Levene *et al.* 2003). Během sekvenačního procesu je ukotvené komplementární vlákno syntetizováno za účasti enzymu DNA polymerázy, která je také ukotvena na dně každé jamky (Korlach *et al.* 2008; Obr. 11). Jamky (ZMW) mají velmi malý objem, což zabezpečuje efektivní detekci začleněných značených nukleotidů. Tento způsob sekvenace je velice rychlý (asi 10 bází za sekundu) a má velmi dlouhou délku čtení (až 10 000 bp). Sekvenaci lze provádět v několika tisících jamek (ZMW) paralelně. Přístroj využívající technologii SMRT vytvořila firma Pacific Biosciences.



Obr. 11. Inkorporace nukleotidu v ZMW.

DNA polymeráza a DNA templát jsou uchyceny na spodní straně jamky, která je zespoda osvětlena laserem. Do jamky jsou přidávány nukleotidy, které mají na fosfátové skupině navázané rozdílné fluorofory. Po začlenění báze se uvolní světelný puls, který je zachycen detektorem. Poté dochází k odštěpení fluorescenční značky. Převzato z opfocus.org

2.5 Technické charakteristiky a srovnání sekvenačních přístupů

Všechny sekvenační systémy nové generace umožňují výrazně vyšší sekvenační výkon než automatické kapilární sekvenování. Zároveň je možno stejné množství genetické informace získat za výrazně kratší čas. Technické parametry dostupných sekvenačních přístrojů uvádí Tab. 1.

Metody NGS jsou tedy vhodné pro použití v laboratořích, kde je třeba rychle získat velké množství genetické informace. Pro menší projekty (rozsahem kilobáze - megabáze) je (a pravděpodobně stále bude) Sangerovo sekvenování vhodnou metodou (Shendure *et al.* 2008).

Tab. 1. Přehled v současnosti používaných platform sekvenujících DNA. (Převzato a upraveno z Li et al. 2011)

Platforma	Technologie	Model	Délka čtení	Propustnost (za den)	Výrobce
Automatizované Sangerovo sekvenování	Kapilární elektroforéza, ddNTP metoda	3730 xl	<900 bp	<3 Mb	Applied Biosystems
454	Pyrosekvenování	GS FLX ⁺	500 - 700 bp	700 Mb	454 Life Sciences (Roche)
SOLiD™	Sekvence ligací	5500 xl	75 bp	30 Gb	Life Technologies (ABI)
Illumina	Reverzní terminátory	HiSeq2000	50 - 150 bp	<55 Gb	Illumina, Inc.
Ion Torrent	Detekce vodíkového iontu	Ion 316 Chip	100 bp	100 Mb	Life Technologies (ABI)
HeliScope™	Zobrazení inkorporovaného NTP	Sekvenátor jednotlivých molekul	35 bp	1 Gb	Helicos
PacBio	SMRT™ technologie	PacBio RS	>1000 bp	500 Mb	Pacific Biosciences

Metoda pyrosekvenování má v porovnání s ostatními systémy NGS vysokou rychlost a větší délku čtení. Jeho hlavním nedostatkem je relativně velká chybovost při začlenění homopolymerů, jeho finanční náročnost (v porovnání s ostatními systémy NGS), a také neschopnost získat větší sekvenační kapacitu (množství sekvenovaných templátů je omezeno velikostí pikotitrační destičky).

Charakteristickým znakem systému SOLiD je díky dvoubázovému dekódování velká přesnost sekvence (99,99%), ale i přes snahu postupného zlepšení tento systém naráží na problém, kterým je jeho krátká délka čtení.

Pro svůj relativně dobrý uživatelský komfort (mimo jiné není nutná emPCR) je v současnosti nejhojněji užívaný systém Illumina Genome Analyzer. Proti pyrosekvenování umožňuje sice sekvenování kratších úseků, není však zatížen tak významnou chybou při čtení homopolymerních oblastí a umožňuje též aplikaci čtení z obou stran templátu při oboustranném čtení (pair-end sekvenování).

Platforma Ion Torrent se od ostatních liší tím, že v ní není potřeba žádného fluorescenčního značení či reverzibilních terminátorů. Tím je výrazně snížena finanční nákladnost a zvýšena rychlost sekvenování. Stejně jako u pyrosekvenování se zde ovšem objevují chyby v identifikaci homopolymerních úseků.

3 Zpracování a analýza sekvenačních dat

Zatímco u Sangerova sekvenování je možno v rámci jedné sekvenační analýzy identifikovat sekvenci do maximálně 900 párů bází a sekvenování unikátní cílové sekvence o velikosti např. 1 milión bp je tedy velice časově náročné, u technologií nové generace jsou z důvodu možnosti masivního paralelního sekvenování analyzovány sice kratší úseky (fragmenty DNA), nicméně sekvenování probíhá paralelně, čímž se doba potřebná k vyšetření cílové sekvence stejné velikosti výrazně zkracuje. V současné době se výzkum v oblasti genomiky velmi rychle rozvíjí a pomocí používaných nových technologií se velmi zvyšuje získané množství sekvenačních dat.

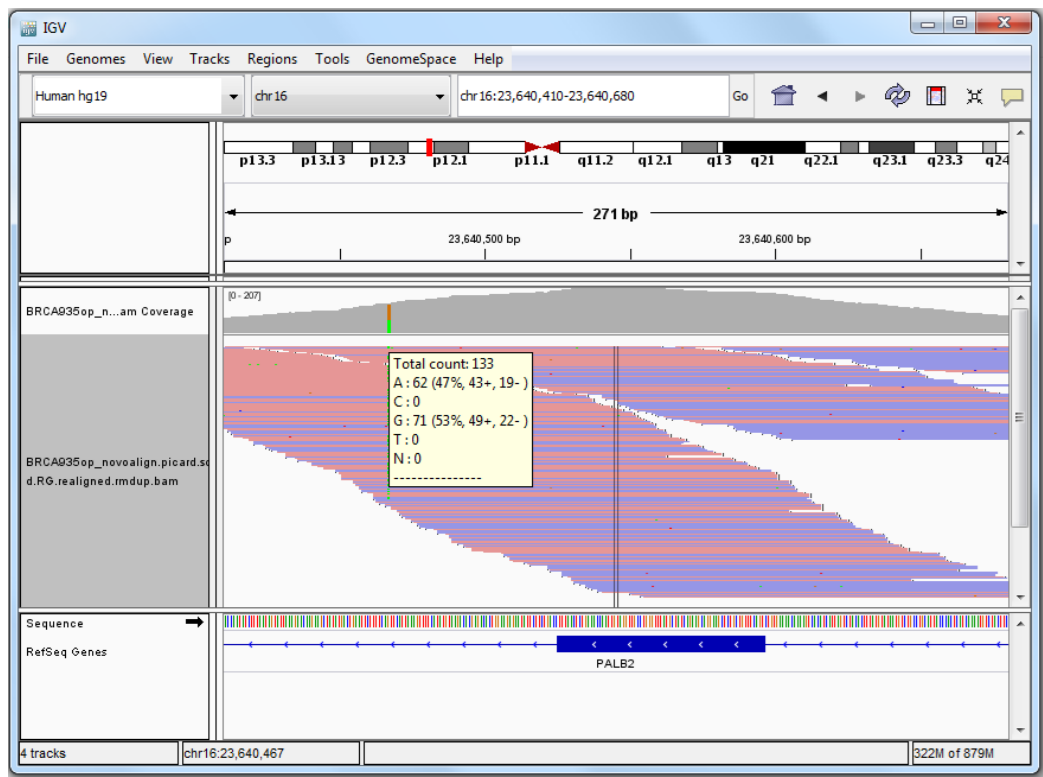
Výstupem vlastního sekvenování pomocí současně používaných sekvenačních technologií druhé generace je získání tzv. readů, neboli jednotlivých sekvenačních čtení amplifikovaných templátů. Jejich délka odpovídá použité sekvenační technologii a pohybuje se od desítek bází (u SOLiDu) až po cca 1000 bází (u FLX, Roche, přehled v Tab. 1). Těchto sekvenačních čtení je v průběhu sekvenování zaznamenáno 10^5 - 10^7 . Kolekce takto zaznamenané informace reprezentuje hrubá data. Tato data je nezbytné zpracovat následnými bioinformatickými postupy, při nichž nejprve dochází k odstranění sekvenačních artefaktů a následnému mapování, které představuje přiřazení jednotlivých překrývajících se čtení k sobě navzájem a jejich porovnání (namapování) na referenční genomovou sekvenci.

Výsledkem je získání namapovaných oblastí, kde jednotlivé úseky sekvenované genetické informace (např. genomu) jsou zmapovány (pokryty) několika různými ready. Počet readů (čtení), které zaznamenávají jednotlivé úseky referenční sekvence, se nazývá sekvenačním pokrytím. Míra pokrytí, na kterou je konkrétní sekvenování navrženo se liší podle účelu daného sekvenování (např. odlišné pokrytí u vyšetřování zárodečných vs. somatických mutací).

Namapované sekvence je možno převést do grafické podoby pomocí různých programových nástrojů (např. Integrative Genomic Viewer; IGV; Obr. 12), které ukazují sekvenační pokrytí zobrazeného úseku referenční sekvence, identifikovanou sekvenci i výskyt přítomných sekvenačních variant. Pro celkové hodnocení sekvenování, obzvláště v případech, kdy se jedná o studium sekvenačních variant v rozsáhlých úsecích kódujících oblastí genomu (exomu) je toto zobrazení nevhodné.

Další úroveň bioinformatického zpracování má za úkol identifikovat nalezené sekvenační změny a tyto změny popsat v kontextu změn na úrovni DNA resp. cDNA a proteinu. Tento postup se nazývá (funkční) anotace. Následně je nezbytné ze získaných dat extrahovat genetické změny, u nichž lze

předpokládat, že mají funkční význam. K tomuto slouží řada filtračních kroků a nástrojů umožňujících prioritizaci především jednonukleotidových missense variant.



Obr 12. Mapování části genu *PALB2* (umístěného na chromosomu 16p) pomocí programu IGV. Jednotlivé ready (modré a růžové úsečky) pokrývají sekvenci oblasti exonu 5 a přilehlých intron exonových přechodů (271 bp) genu *PALB2*. V místě výskytu varianty (hnědo-zelený sloupec v šedém grafu celkového pokrytí) je v obdélníku vyznačeno celkové sekvenační pokrytí 133x, nukleotid A byl osekvenován 62x, zatímco nukleotid G 71x.

3.1 Skládání sekvenačních dat - mapování (assembly)

Skládání sekvenačních dat je prvním krokem k úspěšnému dokončení analýzy. Získané osekvenované fragmenty DNA (ready) se musí zarovnat a sloučit do řetězce na sebe navazujících fragmentů DNA, čímž se vytvoří spojitá sekvence, která je následně použita k získání sekvence původní. Takto sloučené fragmenty DNA se nazývají kontigy. Proces skládání sekvencí může být proveden dvojím způsobem. V případě dostatečné hloubky pokrytí lze využít překryvu jednotlivých

sekvencí a sestavit kontig *de novo*⁵. Druhým způsobem je sestavení původní sekvence pomocí již známé, referenční sekvence (mapování), která pochází ze stejného nebo blízkého příbuzného druhu. Referenční sekvence jsou uloženy v online databázích (GenBank v Severní Americe, Ensembl databáze v Evropě) a jsou veřejně přístupné (Yandell *et al.* 2012). Při mapování genomu program z tzv. PCR duplikátů vybere nejkvalitnější osekvenovaná čtení, vytvoří se prvotní SAM (Sequence Alignment Map) textový formát a následně BAM (Binary Alignment Map) formát, který textovou informaci převede na binární, čímž se objem dat zmenší a zrychlí se vyhledávání a přiřazení určitých regionů (Li *et al.* 2009). Poté program provede namapování na referenční sekvenci (Obr 13).

3.2 Anotování sekvenačních dat

Získaná sekvenační data musí být anotována, aby sekvenování genomů, exomů, či genů mělo význam a získané informace mohly být nadále využity. Genovou anotací rozumíme identifikování nalezených sekvenačních variant v kontextu struktury genů (kódujících/nekódujících sekvencí) a jimi kódovaných proteinů (vliv nalezených alterací na změnu genetického kódu). Anotace stovek tisíc variant není možné provádět manuálně. Proto k anotování slouží řada anotačních softwarových nástrojů (např. Annovar). Výsledkem anotace je uživatelem definovaný souhrn nalezených variant. U variant v některém ze známých genů jsou určeny její genomové koordináty, vztah ke konkrétnímu genu, respektive k jeho kódujícím oblastem (exonům) či oblastem nekódujícím (intronům a regulačním sekvencím). U variant kódujících oblastí je identifikována případná změna čtecího rámce (v případě krátkých insercí/ delecí) nebo dopad nukleotidových záměn na výslednou aminokyselinovou sekvenci vznikajícího proteinu (silent, missense nebo nonsense varianty). K těmto informacím lze doplnit řadu dalších údajů, jako je frekvence výskytu popisovaných variant ve veřejně přístupných databázích (dbSNP;

⁵ Sekvenováním *de novo* nazýváme sekvenování, při kterém zjišťujeme sekvenci genomu druhu, který dosud nebyl osekvenován. Jednotlivá čtení při tomto způsobu sekvenování musíme tedy sestavit na základě identických sekvencí v různých readech. Po složení kontigu je nutné identifikovat a odstranit repetitivní sekvence, které značně anotaci komplikují. Toto se děje pomocí postupu tzv. maskování, při kterém se všechny repetitivní nukleotidy přemění na "N", tedy signál, že regiony jsou opakovány. Identifikaci genu ze seskládaných DNA sekvencí zajišťují *ab initio* genové prediktory. Ty k určení exon - intronové struktury genu používají matematické modely a nepotřebují tedy k zjištění struktury genu žádné referenční sekvence. Takto dokáží identifikovat pouze kódující sekvenci, nevyhledají nepřekládané regiony nebo alternativně sestříhané transkripty (Yandell *et al.* 2012)

<http://www.ncbi.nlm.nih.gov/SNP>) shromažďujících údaje o sekvenaci exomů (ESP 6500; evs.gs.washington.edu) či genomů (1000 genomes project; www.1000genomes.org).

3.3 Predikce významu sekvenačních variant

Množství identifikovaných variant DNA získané metodami sekvenování druhé generace je proti výstupu ze sekvenování generace první enormní. Nejhojněji zastoupenou skupinou genetických změn, které se nacházejí při sekvenování, jsou jednonukleotidové záměny. Jejich frekvence se odhaduje zhruba na 1/1000 bp. Zatímco hodnocení významu nonsense variant vedoucích k zařazení stop kodonu s následnou syntézou zkrácené a tedy obvykle funkčně defektní formy proteinu je poměrně málo komplikované, významně obtížnější je funkční popis významu zaznamenaných missense variant. K predikci jejich funkčního významu byla vyvinuta celá řada predikčních modelů. Tyto modely hodnotí význam missense variant s ohledem na i) charakter aminokyselinové záměny z hlediska jejích chemických vlastností, ii) konzervativnost dotčeného aminokyselinového zbytku v průřezu fylogenetického vývoje, iii) strukturní význam aminokyselinového zbytku pro lokální konformaci proteinu, nebo iv) selekční genetický tlak genetické záměny v průběhu evoluce (CADD score). Pro vytvoření predikčního algoritmu CADD byly identifikovány všechny možné jednonukleotidové varianty (8,6 miliard). Na základě definovaných kritérií jim bylo přiděleno skóre, které můžeme rozdělit na raw a scaled. Raw skóre je přiděleno každé variantě na základě matematického modelu. Varianty jsou poté seřazeny podle tohoto čísla od největšího po nejmenší, kde vyšší číslo reprezentuje větší závažnost mutace. Scaled (C-) skóre je následně generováno podle pořadí variant v raw skóre a nabývá hodnot 0 -50. Výpočet je založen na principu Phred skóre původně zavedeného pro hodnocení pravděpodobnosti identifikace správného nukleotidu při automatickém sekvenování (kapitola 2.2 na konci). Je-li tedy například C-skóre = 20, daná varianta se nachází ve skupině 1 % nejvíce patogenních variant, čili 99% ostatních je zařazeno pod umístěním dané varianty – predikováno jako méně patogenní. Tento algoritmus objektivně sjednocuje obsah mnoha predikčních nástrojů (např. SIFT, PolyPhen, Grantham) v jednu hodnotu, čímž proces vybírání prioritizovaných variant zjednodušuje, avšak k získání co nejvěrohodnější predikce je zapotřebí různé predikční algoritmy kombinovat. (Kircher *et al.* 2014). V současné době se jedná o nejnovější variantu v rámci všech predikčních algoritmů. Predikce důsledku dané varianty ovšem ve skutečnosti umožňuje spíše prioritizaci variant k dalším analýzám, než konečnou identifikaci patogenních mutací. K definitivnímu potvrzení charakteru vybrané varianty je nutné ji dále zpracovat např. funkční nebo segregační analýzou.

4 Příklad využití technologií NGS pro studium lidských nádorů

Ze stručného úvodu vyplývá, že základní charakteristikou všech nádorových onemocnění jsou poruchy genové DNA, v podobě somatických mutací v genomu nádorových buněk nebo hereditárních predispozičních variant v genomu pacienta/probanda. Z tohoto pohledu lze nádorová onemocnění zařadit mezi nejčastější geneticky podmíněné choroby. Charakterizace příčinných mutací na úrovni primární struktury DNA má proto v onkogenetice, resp. onkogenomice nezastupitelnou úlohu. Genetická testování v onkologii lze rozdělit na analýzy somatických mutací (vznikajících v průběhu života pacienta) v nádorově transformované tkáni a analýzy dědičných variant.

Do doby zavedení sekvenačních metod NGS byla mutační analýza využívána v diagnostice nádorových onemocnění omezena na testování hlavních známých genů (resp. mutací v nich) z důvodů kapacitních i finančních.

Principy sekvenování somatických a hereditárních variant se neliší; zásadní rozdíl v přístupu a provedení vlastního sekvenování je v návrhu pokrytí cílové unikátní sekvence (hloubky čtení; viz kapitola 3). Význam sekvenačního pokrytí je zcela zásadní pro analýzu komplexních vzorků, kterými jsou vzorky genetického materiálu ze směsných populací, například vzorky nádorové tkáně. Při jejich analýze lze předpokládat, že nádorové buňky, tvoří pouze malou část všech přítomných buněk (často jednotky procent). Při dostatečném sekvenačním pokrytí (10^3 - 10^5 readů, tzv. "deep sequencing") lze však pomocí NGS zachytit somatické mutace v nádorově transformovaných buňkách s vysokou spolehlivostí.

4.1 NGS při analýze hereditárních mutací

Lidský genom obsahuje asi 25 000 kódujících genů, z toho prozatím bylo identifikováno asi 150 genů, které mají ve svých mutantních variantách známou spojitost se vznikem nádorového onemocnění (Rahman 2014). Podrobným studiem těchto predispozičních genů se zabývá onkogenetika. Protože sekvenování celého genomu při podezření na přítomnost již popsané mutace v konkrétním genu by bylo časově i finančně velmi neekonomické, provádí se vyšetření pomocí analýzy jednotlivých predispozičních genů. Nejprve dochází k analýze těch genů, ve kterých je nejvíce očekáván výskyt patogenní mutace a postupně se se snižující pravděpodobností provádí analýzy genů následujících. V současné době se již více využívá tzv. panelové sekvenování (Targeted cancer sequencing), tedy vyšetření pouze vybraných skupin genů (regionů), které jsou asociovány

s určitým nádorovým syndromem pomocí NGS K osekvenování žádaných genů slouží technologie nazývaná vychytávání DNA (Sequence Capture) která v jednom kroku nabohacuje zvolené (například kódující oblasti) z celé DNA. K nabohacení dochází hybridizací s jednovláknovými oligonukleotidovými sondami navázanými na biotinylované kuličky. Ty jsou po hybridizaci vychytávány z roztoku magnetickými kuličkami s navázaným streptavidinem (systém SeqCap EZ System; vytvořen firmou Nimblegen). Vzhledem k tomu, že se sekvenují pouze žádané geny, sekvenování umožňuje hlubší pokrytí a tím se zvyšuje citlivost ke zjištění mutantní varianty (www.nimblegen.com). Toto vyšetření více genů v jednom kroku s sebou přináší výhody, jako jsou zkrácení doby analýzy a snížení materiálních nákladů na analýzy.

Přestože exom reprezentuje méně než dvě procenta lidského genomu, právě v exonech se nachází cca 85% všech známých mutantních variant způsobujících nádorová onemocnění (www.sciencemag.org). U sekvenování exomu je cílem nabohacení všech exonů v genomu (obvykle je to kolem 30 - 100 Mb, v závislosti na tom, zda je požadováno vychytávání i regulačních oblastí genů či analýza genů kódujících také heterogenní regulační RNA).

Nejkomplexnější informace o genomu dostaneme při tzv. celogenomovém sekvenování (Whole genome sequencing). To nám udává informace o nukleotidových variantách v kódujících i nekódujících oblastech, delecích, a také o chromozomálních přestavbách. Jeho nevýhodou je ovšem vysoká cena a náročnost na zpracování získaných dat.

4.2 NGS při analýze somatických mutací

Analýzu mutací v nádorových ložiscích komplikuje fakt, že nádor není homogenní strukturou a mutantní varianty se tedy nevyskytují v celé populaci jeho buněk. Nadto je při analýze důležité rozlišit tzv. "driver" a "passenger" mutace (viz kapitola 1).

Rozlišení mutantních variant v nádorové tkáni se dosud provádí Sangerovou metodou, kde je ovšem k průkazné analýze potřeba mít v analyzované tkáni alespoň 15-20 % nádorových buněk, případně komerčně dostupnými kity. Jejich princip se může lišit, citlivost jednotlivých přístupů se pohybuje mezi pěti a jedním procentem přítomné mutantní varianty. Toto je využíváno zejména v molekulárně-patologické diagnostice při identifikaci vybraných známých mutací v genech *KRAS*, *NRAS* a *BRAF*.

Pro studium těchto heterogenních nádorových ložisek jsou vhodné metody NGS, protože jejich použitím lze významně rozšířit spektrum vyšetřovaných variant. Toto cílené sekvenování je levnější, umožňuje hlubší pokrytí a jeho pomocí bylo odhaleno několik nových mutací v protoonkogenech (*ERBB4*, *KDR*, *FGFR4*; Koubková *et al.* 2014).

Současně je možno také s přizpůsobením různé hloubky sekvenačního pokrytí identifikovat i minoritně se vyskytující subklony nádorové populace a podle výsledků vyšetření navrhnout pacientům specifickou léčbu a opětovným sekvenováním v časových rozestupech lze včas detekovat příznaky možného relapsu.(Shyr a Liu 2013).

5 Závěr

Sekvenování nové generace se koncepčně odlišuje od klasických sekvenačních postupů paralelním sekvenováním tisíců sekvenačních templátů. Tato strategie tak umožňuje sekvenování velmi komplexních souborů DNA. NGS sekvenování přináší do studia a diagnostiky nádorových onemocnění zcela nové možnosti, které významně posunují naše znalosti onkogeneze i klinickou aplikovatelnost těchto poznatků. Umožňuje vyšetřovat nejen geny, u kterých byly popsány jasné patogenní mutace, ale i velké kolekce (panely) predispozičních genů. Při vyšetření somatických mutací v nádorech pak umožňuje odhalit patogenní varianty DNA v raritních nádorových populacích, které by byly pouze obtížně zachytitelné jinými vyšetřovacími přístupy.

6 Seznam použité literatury

Pozn. Přehledné práce jsou vyznačeny modře.

Anderson MW, Schrijver I. Next generation DNA sequencing and the future of genomic medicine. *Genes (Basel)*. 2010;1(1):38-69.

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):137-58.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, AnastasiC, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling NgB, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Racz C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53-9

Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci U S A*. 2003;100(7):3960-4.

Dahm R. Friedrich Miescher and the discovery of DNA. *Dev Biol*. 2005 Feb 15;278(2):274-88.

Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*. 2006;34(3):e22.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol*. 2007;8(7):R143.

Chen CY. DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front Microbiol*. 2014;5:305.

- Jacobsen H, Klenow H, Overgaard-Hansen K. The N-terminal amino-acid sequences of DNA polymerase I from *Escherichia coli* and of the large and the small fragments obtained by a limited proteolysis. *Eur J Biochem.* 1974;45(2):623-7.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-5.
- Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc Natl Acad Sci U S A.* 2008;105(4):1176-81.
- Koubková L, Vojtešek B, Vyzula R. Sekvenování nové generace a možnosti jeho využití v onkologické praxi. *Klinická onkologie Journal.* 2014;27(1):S61-S68.
- Krush A J. Contributions of Pierre Paul Broca to cancer genetic. *Transactions of the Nebraska Academy of Sciences and Affiliated Societies.* 1979;Paper 316.
- Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science.* 2003;299(5607):682-6.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
- Li L, Goedegebuure P, Mardis ER, Ellis MJ, Zhang X, Herndon JM, Fleming TP, Carreno BM, Hansen TH, Gillanders WE. Cancer genome sequencing and its implications for personalized cancer vaccines. *Cancers (Basel).* 2011;3(4):4191-211.
- Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9:387-402.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, PlantR, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005;437(7057):376-80.
- Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A.* 1977;74(2):560-4.
- McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC, Zhang Z, Ranade SS, Dimalanta ET, Hyland FC, Sokolsky TD, Zhang L, Sheridan A, Fu H, Hendrickson CL, Li B, Kotler L, Stuart JR, Malek JA, Manning JM, Antipova AA, Perez DS, Moore MP, Hayashibara KC, Lyons MR, Beaudoin RE, Coleman BE, Laptewicz MW, Sannicandro AE, Rhodes MD, Gottimukkala RK, Yang S, Bafna V, Bashir A, MacBride A, Alkan C, Kidd JM, Eichler EE, Reese MG, De La Vega FM, Blanchard AP. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* 2009;19(9):1527-41.
- Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31-46.
- Morgan TH. Has the white man more chromosomes than the negro? *Science.* 1914;39(1014):827-8.
- Nussbaum RL, McInnes RR, Willard HF, Thompson J, Thompson MW. *Klinická genetika: Thompson.* 2004. Vyd. 6. Praha: Triton, 426, lix s. ISBN 80-725-4475-6.
- Nyrén P, Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal Biochem.* 1985;151(2):504-9.

- Rahman N. Realizing the promise of cancer predisposition genes. *Nature*. 2014 Jan 16;505(7483):302-8.
- Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*. 1996;242(1):84-9.
- Ronaghi M, Nygren M, Lundeberg J, Nyrén P. Analyses of secondary structures in DNA by pyrosequencing. *Anal Biochem*. 1999;267(1):65-71.
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011;475(7356):348-52.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 1988;239(4839):487-91.
- Sanger F, Coulson AR. The use of thin acrylamide gels for DNA sequencing. *FEBS Lett*. 1978;87(1):107-10.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-7.
- Shao K, Ding W, Wang F, Li H, Ma D, Wang H. Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection. *PLoS One*. 2011;6(9):e24910.
- Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet*. 2010;19(R2):R227-40.
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26(10):1135-45.
- Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. 2005;309(5741):1728-32.
- Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. *Biol Proced Online*. 2013 Feb 13;15(1):4.
- Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986;321(6071):674-9.
- Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-24.
- Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet*. 2009;85(2):142-54.
- Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171(4356):737-8.
- Wu R. Nucleotide sequence analysis of DNA. I. Partial sequence of the cohesive ends of bacteriophage lambda and 186 DNA. *J Mol Biol*. 1970;51(3):501-21.
- Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet*. 2012;13(5):329-42.

7 Elektronické zdroje

Illumina, Inc. [online]. 2015 [cit. 2015-04-29]. Dostupné z: <http://www.illumina.com>

Applied Biosystems by Life Technologies [online]. 2013 [cit. 2015-04-28]. Dostupné z: <http://www.appliedbiosystems.com/absite/us/en/home.html>

454 Life Sciences, a Roche Company [online]. 1996-2015 [cit. 2015-04-30]. Dostupné z: <http://www.454.com/index.asp>

DUŠEK Ladislav, MUŽÍK Jan, KUBÁSEK Miroslav, KOPTÍKOVÁ Jana, ŽALOUĐÍK Jan, VYZULA Rostislav. Epidemiologie zhoubných nádorů v České republice [online]. Masarykova univerzita, [2005], [cit. 2015-5-12]. Dostupné z: <http://www.svod.cz>

1000 Genomes. A Deep Catalogue of Human Genetics Variations [online databáze] 2008-2012 [cit 2015-05-10] Dostupné z: <http://www.1000genomes.org>

NHLBI Exome Sequencing Project (ESP). Exome Variant Server [online databáze] 2015 [cit. 2015-05-10] Dostupné z: <http://evs.gs.washington.edu/EVS/>

Roche NimbleGen, Inc. [online]. 2002-2015 [cit. 2015-05-07] Dostupné z: <http://www.nimblegen.com>

Ústav zdravotnických informací a statistiky ČR [online]. 2010-2015 [cit. 2015-05-10] Dostupné z: <http://www.uzis.cz>

Optics & Photonics Focus. Decoding the Code of Life [online] 2009 [cit. 2015-05-08] Dostupné z: <http://opfocust.org/index.php?topic=about>

NCBI dbSNP Short Genetic Variations [online databáze] 2009 [cit. 2015-05-08] Dostupné z: <http://www.ncbi.nlm.nih.gov/projects/SNP/>