

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra buněčné biologie a genetiky



**Optimalizace diagnostiky somatických
mutací vybraných solidních nádorů pomocí
masivně paralelního sekvenování**

Diplomová práce

Bc. Barbora Koblihová

Studijní program: Biologie

Studijní obor: Molekulární a buněčná biologie

Forma studia: Prezenční

Olomouc 2019

Vedoucí práce: Ing. Rastislav Slavkovský, Ph.D.

Čestně prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně pod vedením Ing. Rastislava Slavkovského, PhD. a za použití uvedených literárních zdrojů.

V Olomouci dne:

.....

SOUHRN

Cílem této diplomové práce je ověřit metodu panelového sekvenování genů pro detekci biomarkerů v diagnostických solidních nádorech. V teoretické části jsou diskutovány základní principy personalizované medicíny a role genomických biomarkerů v této léčebné strategii. Protože naše experimenty byly z velké míry prováděny na *BRCA*-mutovaných nádorech a chordomech, je zde popsána také cílená terapie pro tyto typy nádorů. Hlavní pozornost je však věnována srovnání různých přístupů sekvenační analýzy nádorového genomu, zejména s ohledem na uplatnění v klinické praxi.

Experimentální část je zaměřena na využitelnost panelu NEBNext Direct Cancer HotSpot Panel, který pokrývá specifické regiony 50 genů. Ačkoli tato metoda prokázala schopnost detekovat většinu hledaných variant, nezachytila jeden z důležitých prediktivních biomarkerů odpovědi na anti-*EGFR* terapii. Na základě toho byl genový panel posouzen pro odlišné aplikace, jako je analýza chordomů. Protože metoda umožňuje začlenění UMI barkódů do sekvenačních čtení, bylo dalším záměrem experimentů otestovat přínosy tohoto přístupu. Pro zpracování sekvenačních dat se začleněnými UMI byla s pomocí otevřených programů sestavena bioinformatická analýza a výsledky byly porovnány s výstupy programu MiSeq Reporter Software. Analýza byla posléze úspěšně ověřena i pro další panel NEBNext Direct BRCA1/BRCA2.

SUMMARY

The objective of this master's thesis is to evaluate a gene-panel sequencing method for detection of biomarkers in diagnostic solid tumours. In the theoretical part, the basic principles of personalized medicine and the role of genomic biomarkers in this treatment strategy are discussed. Targeted therapies for *BRCA*-mutated cancers and chordomas are also included, because the gene-panel testing in our experiments was performed mostly on these types of tumours. Nevertheless, the main attention is devoted to the comparison of different approaches for cancer genome sequencing analysis with respect to their utility in clinical practice.

The experimental part is focused on the applicability of NEBNext Direct Cancer HotSpot Panel covering specific regions of 50 genes. Even though the method proved to be able to detect most of the desired variants, it missed one of the important predictive biomarkers for response to anti-EGFR therapy. Accordingly, the gene panel was considered for different applications such as analysis of chordomas. Since the method enables incorporation of UMI adapters into the sequencing reads, the next goal of the thesis was to explore benefits of this approach. For processing UMI-containing sequencing data, the pipeline using open-source software was created and results were compared with outputs from MiSeq Reporter Software. The pipeline was then successfully evaluated also for another panel NEBNext Direct BRCA1/BRCA2.

Na tomto místě bych chtěla poděkovat svému školiteli Ing. Rastislavu Slavkovskému, Ph.D. za odborné vedení, nevyčerpatelnou trpělivost i všem věnovaný čas. Velmi si vážím předaných zkušeností a cenných rad. Dále děkuji Mgr. Lucii Kotkové za ochotnou pomoc v laboratoři a Ph.D. Janě Stránské za poskytnutá data k analýze. V neposlední řadě mé poděkování patří také kolektivu Ústavu molekulární a translační medicíny a CGB laboratoři Ostrava za spolupráci a poskytnutý materiál.

OBSAH

1	SEZNAM POUŽITÝCH ZKRATEK	6
2	ÚVOD	8
3	CÍLE PRÁCE	9
4	SUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY.....	10
4.1	Mutace jako základní charakteristika nádorů.....	10
4.2	Personalizovaná medicína a strategie cílených léčiv	11
4.3	Typy biomarkerů v klinice.....	13
4.4	Organizace genomických dat a identifikace biomarkerů v onkologii.....	14
4.5	Léčiva používaná na základě prediktivních biomarkerů v protinádorové terapii	15
4.5.1	Studie „basket trial“: nový pohled na význam rakovinných typů	16
4.6	Cílená terapie pro <i>BRCA</i> -mutované nádory vaječníků a prsu	17
4.7	Možnosti cílených léčiv u vzácných typů rakovin: chordomů	18
4.8	Rezistence nádorů k léčbě	19
4.9	Metody nádorové genomiky	20
4.9.1	Přehled současně nejběžnějších NGS technologií	21
4.9.2	Vliv kvality vzorků na výsledky NGS analýzy a metoda tekuté biopsie.....	26
4.10	Přístupy genomického profilování nádorů na základě rozsahu analýzy	28
4.10.1	Celogenomové sekvenování.....	28
4.10.2	Celoexomové sekvenování.....	28
4.10.3	Transkriptomové sekvenování (RNA-seq).....	29
4.10.4	Panelové sekvenování	29
4.10.5	Digitální PCR	34
4.11	Srovnání dostupných technologií pro sekvenování panelů genů.....	35
4.11.1	Cílené obohacení od Roche NimbleGen a Agilent Technologies	35
4.11.2	Panelové sekvenování Illumina	36
4.11.3	Panelové sekvenování Thermo Fisher Scientific	37
4.11.4	Genové panely Qiagen	38
4.11.5	Genové panely NEB	39
4.12	Použití molekulárních barkódů pro zvýšení přesnosti sekvenační analýzy.....	39
4.13	Zpracování dat ze sekvenátoru MiSeq.....	40
5	MATERÁL A METODY	42
5.1	Biologický materiál.....	42
5.1.1	Testované vzorky pomocí panelu NEBNext Direct Cancer HotSpot Panel	42
5.1.2	Testované vzorky pomocí panelu NEBNext Direct BRCA1/BRCA2.....	42
5.2	Seznam chemikálií, roztoků a kitů	42
5.3	Přístrojové a programové vybavení.....	43
5.4	Analýza panelu 50 genů referenčních a patientských nádorových vzorků.....	43
5.4.1	Příprava zásobních roztoků EGFR referenčních vzorků	43

5.4.2	Kontrola kvality a koncentrace DNA patientských vzorků.....	44
5.4.3	Fragmentace DNA vzorků s průměrnou velikostí molekul nad 200 bp	44
5.4.4	Příprava knihovny pomocí kitu NEBNext Direct Cancer HotSpot Panel.....	44
5.4.5	Analýza kvality eluované sekvenační knihovny.....	49
5.4.6	Pracovní postup sekvenování	49
6	VÝSLEDKY.....	51
6.1	Příprava DNA knihovny a primární výsledky sekvenování.....	51
6.2	Parametry kvality sekvenování	54
6.3	Sestavení bioinformatické analýzy s UMI barkódy pro data z panelů NEB	54
6.4	Ověření analýzy s UMI barkódy a porovnání výstupních dat z programu MSR.....	58
6.4.1	Stanovení sensitivity a specificity analýzy vzorku HD802 pro genotypizaci panelu 50 genů ..	58
6.4.2	Vliv pozičních filtrů na výsledná data z analýzy s UMI barkódy	63
6.4.3	Stanovení sensitivity a specificity analýzy s UMI pro genotypizaci <i>BRCA 1</i> a <i>2</i>	66
6.5	Detekce mutací s nízkou frekvencí NEB Cancer panelem za použití vzorků HD802 a HD709	67
6.6	Porovnání NEB Cancer panelu a panelu TruSight Tumor 170 na vzorku nádoru slinivky.....	69
6.7	Analýza mutací v chordomech pomocí NEB Cancer panelu	70
7	DISKUZE.....	72
8	ZÁVĚR	77
9	SEZNAM POUŽITÉ LITERATURY	79

1 SEZNAM POUŽITÝCH ZKRATEK

BW	Bead Wash Buffer
BWA	Burrows-Wheeler Aligner
CCD	charge-coupled device
cfDNA	volně cirkulující DNA (Cell-Free DNA)
CML	chronická myeloidní leukémie
CNV	varianta v počtu kopií (copy number variation)
UMI	molekulární barkód (unique molecular identifier)
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DEPC	dietyl pyrokarbonát
dNTP	deoxynukleotid
EMA	Evropská léková agentura (European Medicines Agency)
NHEJ	nehomologní spojování konců (non homologous end joining)
FDA	Úřad pro kontrolu potravin a léčiv (U.S. Food and Drug Administration)
FFPE	formalínem fixované v parafínu zalité biopsie (formalin-fixed paraffin-embedded)
gDNA	genomická DNA
HW	Hybridization Wash
HD	Horizon Discovery
HGVS	Human Genome Variation Society
LOD	limit detekce (limit of detection)
MSR	MiSeq Reporter Software
NGS	sekvenování nové generace
OS	celkové přežití (overall survival)
FISH	fluorescenční in situ hybridizace
PCR	polymerázová řetězová reakce
PE	párová koncová čtení (paired-end)
NEB	New England Biolabs
PFS	přežití bez progresu (progression-free survival)
pSTD	read position standard deviation (průměrná odchylka pozice ve čtení)

qPCR	kvantitativní polymerázová řetězová reakce
RECIST	hodnocení odpovědi nádoru na léčbu (Response Evaluation Criteria In Solid Tumors)
RKO	buněčná linie karcinomu tlustého střeva
SB	strand bias
SGS	sekvenování druhé generace (second generation sequencing)
SBS	sekvenování syntézou (sequencing by synthesis)
SNP	jednonukleotidový polymorfismus (single-nucleotide polymorphism)
SNV	jednonukleotidová varianta (single-nucleotide variant)
SV	strukturní varianta
TST170	TruSight Tumor 170
ÚMTM	Ústav molekulární a translační medicíny
ZMW	zero-mode waveguide

2 ÚVOD

Pokrok v onkologickém výzkumu a studiu rakovinných procesů na molekulární úrovni je úzce spjat s příchodem moderních vysokokapacitních metod, jako jsou technologie sekvenování nové generace. Velké množství získaných genomických poznatků souběžně podnítilo rozvoj personalizované medicíny, nového směru diagnostiky a léčby onemocnění založeném na individuální charakterizaci molekulárních změn každého sledovaného pacienta (Morash *et al.*, 2018; Collins *et al.*, 2017). V současnosti existuje již celá řada biomarkerů, které se využívají v cílené protinádorové léčbě s klinicky prokázanou účinností. Protože oproti konvenční chemoterapii cílená léčiva působí na nádorové abnormality specifickým způsobem, předpokládá se efektivnější výsledek léčby s nižší toxicitou (Joo *et al.*, 2013; Collins *et al.*, 2017).

Analýza genomických biomarkerů v nádorech dnes otevírá široké možnosti s potenciálním uplatněním pro mnoho odlišných aplikací. Jaké výhody a nevýhody jednotlivé přístupy přináší, představuje jeden z hlavních motivů této práce. Ačkoli v klinické praxi mají význam i další metody, největší pozornost je zde věnována technologiím sekvenování nové generace se zaměřením na panely genů. Citlivá a specifická detekce mutací v nádorové DNA je klíčový krokem cílené terapie, která ovlivňuje stanovení diagnózy i následná rozhodnutí o léčbě pacienta. Schopnost metody podávat nezkreslené výsledky i při nízkých vstupech testované genetické informace je nyní o to aktuálnější s ohledem na rostoucí význam tekuté biopsie a volně cirkulující nádorové DNA v krvi pacienta. Možné využití genového panelu NEBNext Direct Cancer HotSpot Panel s molekulárními barkódami právě pro tyto účely ověřuje experimentální část diplomové práce.

3 CÍLE PRÁCE

Teoretická část:

- Seznámit se s problematikou personalizované medicíny a prediktivních markerů pro solidní nádory.
- Sepsat rešerši o použití panelového multi-paralelního sekvenování (MPS) pro typizaci genů spojených s rozvojem nádorů.
- Seznámit se s tématem sekvenování na platformě Illumina MiSeq a možnostech zpracování dat s ultra vysokým sekvenačním pokrytím.
- Popsat nejnovější trendy a technologie v oblasti přípravy sekvenačních knihoven pro hluboké ampliconové sekvenování s důrazem pro diagnostiku solidních nádorů.
- Ověřit uplatnění molekulárních barkódů pro zlepšení kvality sekvenování, citlivosti detekce mutací a seznámit se s možnostmi jejich použití pro aplikaci genotypizace cirkulující nádorové DNA z tekuté biopsie.
- Popsat nejnovější trendy v oblasti predikce léčby solidních nádorů.

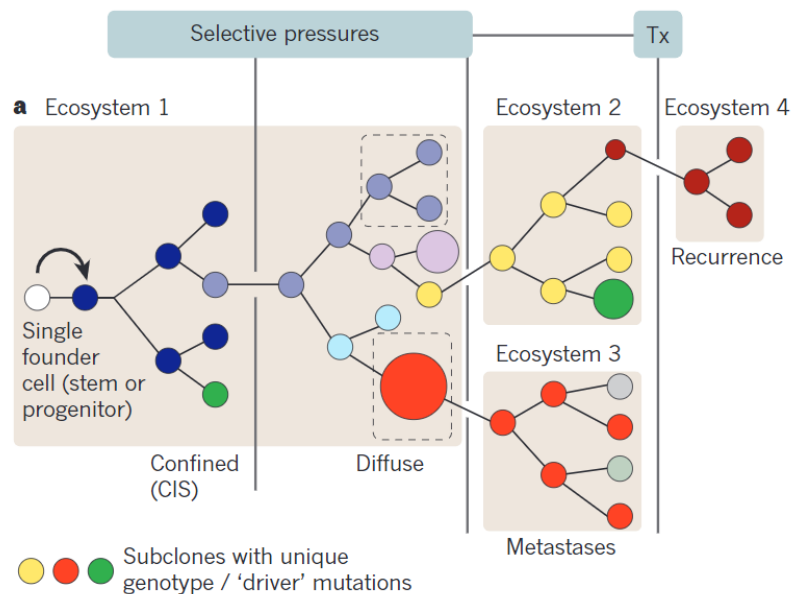
Experimentální část:

- Optimalizovat metodu panelového sekvenování min. 50 genů (např. NEBNext Direct Cancer HotSpot Panel).
- Zpracovat a sekvenovat referenční vzorky a vzorky DNA ze solidních nádorů.
- Ve spolupráci s dalšími pracovníky zpracovat hrubé sekvenační data a získat výsledky genotypizace.
- Porovnat výsledky genotypizace pro různé způsoby zpracování dat a různé parametry.
- Sumarizovat výsledky a vyhodnotit parametry nové metody (např. specificita, sensitivita).

4 SUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

4.1 Mutace jako základní charakteristika nádorů

Genetické faktory hrají klíčovou roli v predispozici, iniciaci i vývoji nádorového onemocnění (Burghel *et al.*, 2015). Ve své podstatě se rakovina řídí principy Darwinovské evoluce, kdy v adaptivním prostředí tkáňových ekosystémů dochází opakovaně ke klonální expanzi, genetické diverzifikaci a klonální selekci (Obrázek 1) (Greaves *et Maley*, 2012). Na základě dané teorie umožňuje získ výhodných mutací určitým buněčným skupinám (subklonům) rozšiřovat se, zatímco jiné zůstávají dormantní či vymizí úplně. Selekcí tlaky pohánějící adaptaci subklonů mohou pocházet z vlivu imunitního systému hostitele, peritumorálního mikroprostředí a/nebo protinádorové terapie. Proces kumulace odlišných genomických změn v jednotlivých subklonech podmiňuje tzv. nádorovou heterogenitu významně ovlivňující možnosti současné léčby. Některé rakoviny projevují zvýšenou genomickou nestabilitu, jako jsou melanomy či nádory s defektními opravami DNA, čímž jsou predisponovány k rozvoji nadměrných subklonálních aberací (Collins *et al.*, 2017).



Obrázek 1: Model klonální evoluce rakoviny. Selektivní tlaky (v obrázku svislé čáry) dovolují některým mutantním subklonům expandovat, zatímco jiné zcela vymizí nebo zůstávají dormantní. Ekosystémy 1–4 vyznačené v šedých polích představují různá tkáňová prostředí. Přerušovanou čarou jsou znázorněna lokálně odlišná mikroprostředí. Metastatické subklony se mohou odtrhnout od hlavních i minoritních klonů primárního nádoru (Převzato z: Greaves *et Maley*, 2012).

Během rozvoje rakoviny sice nastává v genomu mnoho somatických genetických i epigenetických změn, jen menšina z přítomných mutací má však na karcinogenezi zásadní vliv. Tyto změny, které se podílejí na zvýšené proliferaci a přežití buněk, jsou označovány

jako mutace řídící (driver mutations) a bývají doprovázeny četnými mutacemi neposkytující žádnou selektivní výhodu (passenger mutations) (Greaves *et al.*, 2012). Zároveň ne všechny z mutací řídících mohou sloužit k prognostickým, prediktivním nebo diagnostickým účelům; zavádí se proto další pojem tzv. zacílitelných mutací (actionable mutations) (Burghel *et al.*, 2015). Jak odlišit tyto terapeuticky významné mutace od nespočtu biologicky neutrálních v natolik různorodé a komplexní krajině, jako je genom nádorových buněk, zůstává jednou z hlavních otázek personalizované medicíny (Hyman *et al.*, 2017).

Počínaje identifikací prvních onkogenů a nádorových supresorových genů máme dnes k dispozici detailní genomické profily nejběžnějších rakovinných typů (Kandoth *et al.*, 2013; Hyman *et al.*, 2017). Zapojení rozsáhlých genomických projektů současně podnítilo vznik nové taxonomie nádorů založené na jejich podrobné molekulární charakterizaci, která nyní doplňuje tradiční klasifikaci rakovin vycházející z histologie (Hoadley *et al.*, 2014). K hodnotným poznatkům v tomto směru přispěla například důkladná integrační analýza provedená Hoadleyem *et al.* (2014) s použitím pěti genomických a jedné proteomické platformy na 3 527 vzorcích z 12 typů rakoviny. Výsledky této studie ukázaly, že klasifikace na základě genomických údajů pro pět typů rakovin odpovídá původnímu rozdělení dle tkáňového původu, avšak některé jiné typy spadají společně do nových skupin. Nyní je tak možné spojit spinocelulární karcinom plic, rakovinu hlavy a krku a několik dalších rakovin močového měchýře do jediného rakovinového podtypu, který je charakteristický změnami TP53, amplifikacemi TP63 a vysokou expresí genů imunitních a proliferačních drah. Kromě toho jsou mutace TP53 vůbec nejčastějšími změnami napříč různými typy nádorů s celkovou frekvencí až 41 %.

Ze spletnosti této problematiky lze snad jen shrnout, že rakovinné typy a podtypy sice charakterizuje společná řada řídících mutací, jejich specifická kombinace se však může významně lišit nejen mezi samotnými nádory, ale i v rámci buněčných subklonů jediného nádorového vzorku (Kandoth *et al.*, 2013; Collins *et al.*, 2017).

4.2 Personalizovaná medicína a strategie cílených léčiv

Odkrytí dějů, které se účastní procesu karcinogeneze, metastáze a rezistence, objev cílených léčiv a v neposlední řadě vznik technologií umožňujících molekulární analýzu nádorů, stimulovaly rozvoj personalizované medicíny (Tsimberidou, 2017). „Správný lék pro správného pacienta ve správný čas“, tak lze stručně vystihnout tento nový léčebný směr (Collins *et al.*, 2017).

Na každého pacienta může být nahlíženo jako na nositele jedinečného uskupení charakteristik, které ovlivňuje riziko i průběh jeho onemocnění. Jednotlivci se liší svým věkem, výživou, životním stylem, prostředím, etnickým původem apod. A stejně tak se mezi jednotlivci liší klinické projevy u každého typu a stupně rakoviny; dochází k rozdílům v chování a progresi nádoru, i v odpovědích na konkrétní léčebný režim. To je z velké části podmíněno právě jedinečným genomickým profilem (DNA, RNA a epigenetickým) jednotlivých nádorů (Uzilov *et al.*, 2016; Goetz *et Schork* 2018).

Personalizovaná medicína pak vychází z informací o nádorovém genomu a dalších individuálních vlastností každého pacienta, aby zajistila co možná nejvyšší účinnost léčby s co možná nejnižší toxicitou (Collins *et al.*, 2017). K terapii využívá cílených léčiv, která oproti klasickým chemoterapeutikům neinterferují jednoduše se všemi rychle se dělícími buňkami, ale blokují specifické dráhy související s karcinogenezí a růstem nádorů. V rakovinných buňkách dané látky mohou vyvolat apoptózu, inhibovat enzymy a receptory růstových faktorů podílejících se na proliferaci nebo modifikovat funkci proteinů, které regulují genovou expresi (Joo *et al.*, 2013).

Z chemického hlediska cílená léčiva spadají do dvou hlavních kategorií, a to malých molekul a monoklonálních protilátek. „Malé molekuly“ odkazují na organické sloučeniny s relativní molekulovou hmotností menší než 800 Da, které mohou pronikat skrze buněčnou membránu a interferují s molekulárními cíli nebo signálními cestami uvnitř buňky. Koncovka „-ib“ v názvu těchto látek vyjadřuje právě jejich přímé inhibiční účinky. Naopak monoklonální protilátky (koncovka „-mab“) obvykle přes buněčnou membránu nepronikají a vážou se na povrchové antigeny specifické pro rakovinné buňky (Joo *et al.*, 2013). Protilátky byly do strategie cílených léčiv zahrnuty teprve v nedávných letech, a zároveň s nimi byl také pozměněn původní koncept personalizované medicíny. Imunoterapie totiž zasahuje imunitní systém, a je tímto nepřímou formou léčby (Collins *et al.*, 2017).

Pro úplnost je třeba dodat, že personalizovaná medicína neřeší pouze otázky spojené s terapií, ale i diagnostikou a prevencí. Kromě genomiky čerpá také z mnoha dalších vědeckých směrů, jako jsou proteomika a metabolomika. V konečném důsledku proto personalizovaná medicína představuje složitý multidisciplinární vědní obor, který se týká nejenom nádorových, ale i mnoha dalších onemocnění (Abrahams *et Silver*, 2009; Goetz *et Schork*, 2018).

4.3 Typy biomarkerů v klinice

Podobně jako pojem personalizovaná medicína, ani biomarker nemá jednoznačnou definici a v literatuře bývá vymežován různě na základě daného kontextu. Biomarker obecně může představovat charakteristiku, která je objektivně měřitelná a vyhodnocována jako indikátor normálních biologických procesů, patogenních procesů nebo farmakologických odpovědí na terapeutický zásah (Biomarkers Definitions Working Group, 2001).

Biomarkery používané pro klinické účely lze vyčlenit na tři druhy: diagnostické, prognostické a prediktivní. Diagnostické biomarkery umožňují identifikaci typu nebo podtypu rakoviny, prognostické biomarkery pak slouží zejména při určování rizika relapsu nebo progresu onemocnění po léčbě. Pro volbu konkrétní terapie a odhad odpovědi pacienta na léčbu mají význam biomarkery prediktivní, přičemž právě tyto často identifikují molekulární cíle protinádorových léčiv (Roychowdhury *et al.*, 2016). Každý z daných typů biomarkerů může být s pomocí rozličných molekulárních metod testován pro detekci změn nádorového genomu, proteomu či fenotypových vlastností, jako je histopatologická klasifikace (Roychowdhury *et al.*, 2016).

Genomické biomarkery nemusí zahrnovat pouze charakteristiky spojené s molekulami DNA, ale také s RNA; reflektují tedy expresi, funkci i regulaci genů. V rámci změn vyskytujících se přímo na úrovni DNA se pak jedná o jednonukleotidové polymorfismy, inserce, delece, variabilitu v počtu kopií nebo mikrosatelity. Dále se do této skupiny řadí také epigenetické modifikace (např. metylace) a cytogenetické přestavby, jako jsou translokace, inverze nebo větší delece a inserce (Novelli *et al.*, 2008). Typické genomické biomarkery, prediktivní i prognostické, které jsou dnes běžnou součástí klinické praxe, ukazuje Tabulka 1.

Tabulka 1: Genomické biomarkery využívané v klinice na příkladu karcinomů plic (Minárik, 2016)

Biomarker	Běžné označení	Typické využití v klinické praxi	Vyšetřované geny
Bodová (jednonukleotidová) mutace	Mut SNV	Predikce léčebné odpovědi Odhad prognózy onemocnění	<i>EGFR</i> <i>KRAS</i> <i>BRAF</i>
Krátké inserce, delece a duplikace	Indel	Predikce léčebné odpovědi	<i>EGFR</i>
Změny v počtu kopií exonů/genů	Ampl	Predikce léčebné odpovědi	<i>ROS1</i> <i>cMET</i>
Strukturní varianty, aneuploidie, fúze	SV	Predikce léčebné odpovědi	<i>EML4-ALK</i>
Epigenetické poruchy	CpG, CIMP	Odhad prognózy onemocnění	<i>MGMT</i> <i>CDKN2A</i>

Personalizovaná medicína při posuzování léčebné odpovědi klade v zásadě dvojí otázku. Za prvé, zda je u pacienta přítomna mutace, pro kterou existuje cílené léčivo, a za druhé, zda u pacienta není přítomna mutace, která způsobuje na danou léčbu rezistenci. Dobře známé geny, které jsou společně vyšetřovány v této souvislosti, představují *EGFR* a *KRAS* (Hyman *et al.*, 2017).

Mutace receptoru epidermálního růstového faktoru (EGFR) hrají významnou roli v mnoha typech nádorů, včetně rakoviny plic, střev, prsu, mozku a dalších (Arteaga *et Engelman*, 2014). Aktivace této transmembránové kinázy stimuluje celou signalizační kaskádu prostřednictvím proteinů RAS, RAF, MAPK a ERK účastnících se procesů přežití, proliferace a metastáze nádorových buněk. V současné době jsou proti hyperaktivovanému EGFR k dispozici cílená léčiva, která ve výsledku blokují právě tyto přidružené dráhy (Brar *et al.*, 2017).

KRAS je zdaleka nejčastěji mutovaným onkogenem v rakovinách obecně a tyto mutace bývají zároveň spojovány se špatnou prognózou a rezistencí na terapii. Daný gen kóduje malou GTPázu z rodiny RAS, kterou doposud nelze zacílit žádným dostupným léčivem (Haigis, 2017). Protože je *KRAS* distálním (downstream) efektozem EGFR, nebude ani inhibice této nadřazené kinázy logicky účinná. Průkaz nemutovaného genu *KRAS* je tak nyní standardní podmínkou pro indikaci cílené anti-*EGFR* terapie (Brar *et al.*, 2017).

4.4 Organizace genomických dat a identifikace biomarkerů v onkologii

Moderní metody sekvenování sice umožnily identifikovat značný počet somatických mutací, většina z nich však postrádá biologickou či klinickou validaci. Je proto nezbytné genomické změny systematicky katalogizovat a charakterizovat jejich frekvence v rámci i mezi jednotlivými rakovinnými typy (Hyman *et al.*, 2017). Tento cíl si již vytyčila některá rozsáhlá konsorcia, jako je The Cancer Genome Atlas (TCGA), které dnešním dnem zmapovalo genomické změny v 33 rakovinných typech (National Cancer Institute, 2019). Projekt je provázán také s COSMIC (Catalogue of Somatic Mutations in Cancer), největším dostupným zdrojem popsaných somatických variant lidských rakovin (Forbes *et al.*, 2010).

Sestavení reprezentativních databází osekvenovaných nádorů je jen prvním krokem v procesu objevování a klinické validaci prediktivních biomarkerů. Jednotlivé genomické změny je potřeba dále klasifikovat na základě jejich abnormální funkce a odpovědi na léčiva. Funkční důsledky se mohou zásadně lišit i pro varianty vyskytujících se v jediném řídicím genu. V tomto směru se tudíž využívají jak výpočetní přístupy vycházející z principů

statistiky, tak experimentálně založené velkokapacitní analýzy, které umožňují rychlou biochemickou charakterizaci velkého počtu genomických variant současně. Své uplatnění zde nacházejí také tradiční funkčně genetické studie s použitím různých moderních metod molekulární biologie včetně CRISPR editování genů. Všechny tyto strategie jsou však primárně zaměřené na zkoumání jednotlivých alel izolovaně, což nepředstavuje typickou situaci v reálných nádorech. Jak terapeuticky postupovat u pacientů s přítomností více zacílitelných mutací proto patří mezi další současné slabiny personalizované medicíny (Hyman *et al.*, 2017).

Nejobtížnější z celého procesu je samotná validace biomarkerů a jejich přenesení do klinické praxe. Vyžaduje to řadu po sobě jdoucích studií a splnění náročných kritérií, díky kterým je úspěšně potvrzena souvislost s nemocí jen u naprosté menšiny potenciálních biomarkerů (Ziegler *et al.*, 2012) Pomalost a těžkopádnost tradičního postupu si již vynutila zavedení jistých změn, jako je umožnění klinické validace biomarkerů ještě před jejich úplnou biologickou a funkční charakterizací v laboratoři (Hyman *et al.*, 2017).

V komplexitě genomických dat by se jen stěží orientovalo bez databází, které se snaží uspořádat a standardizovat informace o biologické a klinické relevanci jednotlivých variant. Současně se proto objevují rozličné zdroje, které mají sloužit k vedení diagnózy, informovat o prognóze a usnadňovat rozhodnutí o léčebném postupu pacientů (Tabulka 2).

Tabulka 2: Dostupné databáze genomických variant pro klinické účely (Hyman *et al.*, 2017).

Název databáze	Oficiální webová stránka
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
OncKB	oncokb.org
MyCancerGenome	mycancergenome.org
CIViC	civic.genome.wustl.edu
Cancer Genome Interpreter (CGI)	cancergenomeinterpreter.org
CANDL	candl.osu.edu
Personalized Cancer Medicine Knowledge Base	https://pct.mdanderson.org/

4.5 Léčiva používaná na základě prediktivních biomarkerů v protinádorové terapii

Jako jeden z prvních úspěchů molekulárně cílené medicíny bývá klasicky uváděn imatinib pro léčbu chronické myeloidní leukémie (CML) (Joo *et al.*, 2013). CML je podmíněna translokací chromozomů 9 a 22, která způsobuje fúzi dvou genů *BRC* a *ABL*. Bcr-Abl onkoprotein získává svou transformační schopnost prostřednictvím zvýšené aktivity Abl tyrosinkinázy (Perazzona *et al.*, 2008). Ještě do roku 1993 byla léčba CML limitována

možnostmi klasické cytoreduktivní terapie, která zajišťovala pouze krátkodobou kontrolu nad nemocí s mediánem přežití mezi 45 až 55 měsíci. O něco pozdější se podařilo pomocí imunomodulačního léčiva IFN α medián přežití navýšit na 60 až 90 měsíců. Průlom v léčbě však nastal až s objevem imatinibu, prvního schváleného inhibitoru tyrosinkinázy, který blokuje vazebné ATP místo onkoproteinu Bcr-Abl. Rozsáhlá klinická studie IRIS s pěti sty pacienty zaznamenala přežití bez progresu u více jak 85 % pacientů po 5 letech testování. Imatinib je navíc pacienty relativně dobře snášen, a mnohé z vedlejších účinků lze potlačit pozdějším snížením podávané dávky (Fausel, 2007).

Do dnešního dne je k dispozici celá řada cílených léků s prokázanou účinností, které jsou předepisovány na základě rutinního testování prediktivních biomarkerů. Mezi nejznámější můžeme jmenovat gefitinib působící v *EGFR*-pozitivních plicních karcinomech, crizotinib účinný v *ALK*-pozitivních plicních karcinomech či trastuzumab pro *HER2*-pozitivní rakovinu prsu (Collins *et al.* 2017; Joo *et al.*, 2013).

Jak uvádějí Sun *et al.* (2017), existuje již více než 80 FDA-schválených cílených léčiv efektivních v různých typech rakoviny, a tento počet strmě narůstá každým rokem. V případech, kdy pro specifickou mutaci nebo typ nádoru pacienta neexistuje zatím žádný lék schválený FDA či EMA, může lékař zvolit také možnost „off-label“. Takové léčivo cílí stejnou mutaci schválenou u jiného typu nádoru nebo cílí jinou mutaci, avšak s předpokládaným stejným funkčním účinkem (Johnson *et al.*, 2015; Lenk *et Duttge*, 2014).

4.5.1 Studie „basket trial“: nový pohled na význam rakovinných typů

Zaměření se právě na takové prediktivní biomarkery, které se vyskytují v širším spektru typů nádorů, může značně rozšířit možnosti cílené terapie na četnější populaci pacientů (Hyman *et al.*, 2017). S touto myšlenkou také vznikají nové třídy klinických studií, jež se snaží zefektivnit vývoj a validaci cílených léčiv. Tradičně se k jednotlivým rakovinným onemocněním přistupovalo zvláště na základě anatomické lokalizace primárního nádoru; tedy zda se jedná o rakovinu prsu, rakovinu plic apod. Z toho vycházel jak postup léčby, tak způsob testování nových léků. Studie třídy „basket trial“ tímto zavádí velmi odlišnou strategii, která si klade za cíl sledovat účinek léku na konkrétní mutaci, a to v různých typech nádorů současně. Přináší to značnou výhodu, pokud vezmeme v potaz typickou situaci, kdy jen malý podíl pacientů s určitým nádorovým onemocněním má specifickou zacílitelnou mutaci, avšak ta je přítomna u mnoha různých typů rakovin. „Basket trials“ pak podávají informaci nejen o účinku léčiva, ale i které typy nádorů jsou k němu citlivé

(Cunanan *et al.*, 2017). Úspěšnost tohoto nového přístupu dokazuje mimo jiné pembrolizumab, jenž se v minulém roce stal prvním FDA schváleným lékem čistě na základě přítomnosti specifického biomarkeru bez ohledu na typ nádoru a věk pacientů (U.S. Food and Drug Administration, 2019).

Význam rakovinných typů přesto nelze opomenout, stejně jako jednoznačně vyhodnotit efektivitu léčiva na základě přítomnosti jedné specifické mutace. Názornou ukázkou je lék vemurafenib účinně inhibující BRAF (V600E) onkoprotein v melanomech, avšak s velmi limitovanou odpovědí v rakovině tlustého střeva. Bylo zjištěno, že obě rakoviny se od sebe liší mírou exprese receptoru pro epidermálního růstový faktor (EGFR), který je zpětnovazebně aktivován inhibicí proteinu BRAF. V rakovině tlustého střeva inhibice proteinu BRAF vede k vyšší aktivaci EGFR, což podporuje buněčnou proliferaci i v přítomnosti vemurafenibu. Hladina EGFR je v melanomu nízká, a proto BRAF inhibitor účinný. Jako řešení se v tomto případě u rakoviny tlustého střeva nabízí použití dvou léčiv, které budou souběžně inhibovat BRAF i EGFR (Prahallad *et al.*, 2012).

4.6 Cílená terapie pro *BRCA*-mutované nádory vaječníků a prsu

Rakovina prsu a vaječníků je velmi často asociována se ztrátou funkce nádorových supresorových genů *BRCA1* a *BRCA2* s významným podílem mutací jak germinálního, tak somatického původu (Lord *et al.*, 2018; Faraoni *et al.*, 2018). Oba proteiny *BRCA1* i *BRCA2* jsou kritické pro opravu dvouvláknových zlomů DNA procesem homologní rekombinace, přičemž narušení této dráhy spouští alternativní opravné děje, jako je nehomologní spojování konců (non-homologous end joining; NHEJ). Přestože primární úlohou NHEJ je chránit buňku před chromozomálními aberacemi a genomovou nestabilitou, ve výsledku to bývají právě tyto mechanismy, které přispívají ke karcinogenezi (Mehta *et al.*, 2018; Sishc *et al.*, 2017).

Zacílení mutací nádorových supresorových genů sice není tak přímočaré, jako je tomu u běžných onkogenů, nicméně i pro *BRCA*-mutované rakoviny dnes existuje efektivní cílená léčba skrze tzv. syntetickou letalitu (Guo *et al.*, 2014). Zjednodušeně se jedná o strategii, kdy současná změna dvou genů způsobuje buněčnou smrt, zatímco při změně pouze jednoho z nich zůstává buňka životaschopná. Konkrétně v případě patogenních mutací *BRCA* lze syntetickou letalitu navodit inhibicí poly(ADP ribóza) polymerázy (PARP). Proteiny PARP fungují jako sensory poškození DNA, které zprostředkovávají navázání dalších efektorů podílejících se na reparačním procesu (Lord *et al.*, 2018). Inhibice PARP

v končeném důsledku vede k akumulaci dvouvláknových DNA zlomů, a tudíž ke smrti buněk s deficientními mechanismy homologní rekombinace (Nicolas *et al.*, 2018). Schváleno bylo již několik inhibitorů cílících tuto dráhu (olaparib, rucaparib, niraparib a talazoparib), které kromě *BRC*A-mutovaných nádorů vaječníku a prsu prokázaly účinky také například v rakovině prostaty (Faraoni *et Graziani*, 2018; De Felice *et al.*, 2017).

4.7 Možnosti cílených léčiv u vzácných typů rakovin: chordomů

Chordomy jsou relativně vzácné maligní nádory kostní tkáně s incidencí 0,08 na 100 000 případů. Nejvyšší výskyt se pohybuje mezi 50 až 60 leti s mediánem přežití přibližně šesti let. Primární ložiska se mohou objevovat ve spodině lebky, oblastech pohyblivé páteře nebo křížové kosti (Walcott *et al.*, 2012; McMaster *et al.*, 2001). Přestože se chordomy rozrůstají jen pomalu, vykazují silnou lokální agresivitu, kdy často tvoří nádory velkých rozměrů stlačující vitální nervy a krevní cévy (Casali *et al.*, 2007). Chordomy navíc obvykle neodpovídají na konvenční radioterapii a cytotoxickou chemoterapii. Primární terapeutickou volbou je tudíž chirurgický zákrok, který obvykle vede k lokálním a metastatickým recidivám. Bezvýhodná situace tohoto onemocnění tak zároveň otevírá diskuzi o cílených léčivech, jako potenciální možnosti terapie (Meng *et al.*, 2019; Bompas *et al.*, 2015).

Z genetického hlediska chordomy představují heterogenní nádory s častými abnormalitami velkých chromozomálních oblastí. Mezi klíčové aspekty patogeneze patří somatické duplikace genu notochordálního transkripčního faktoru brachyury a mutace se ztrátou funkce genů *PTEN*, *TSC*, *CDKN2*, *SMARCB1* a *PIK3CA*. Zásadní význam mají i mutované tyrosinkinázové receptory, které jsou současně hlavními molekulárními cíli tohoto onemocnění (Meng *et al.*, 2019). Jedná se zejména o geny *c-KIT*, *PDGFR* (α a β), *HER2/neu*, *EGFR*, *VEGFR* a *c-MET* (Bompas *et al.*, 2015; Weinberger *et al.*, 2015).

Cílená léčiva doposud nejsou léčbou první volby chordomů a mohou být doporučena pouze pro pokročilé nebo rekurentní nádory, které nereagují ani na radioterapii, ani na chirurgickou resekci. Odpověď léčby lze navíc vyjádřit více způsoby, jako je RECIST nebo Choiovo kritérium, hodnotící například velikost nebo densitu nádorů. Dostupné výsledky cílené terapie chordomů se proto poměrně obtížně vyhodnocují (Meng *et al.*, 2019).

Ze současně testovaných léčiv je jednoznačně nejpoužívanější imatinib, již dříve zmiňovaná látka v souvislosti s chronickou myeloidní leukémií. Tento specifický inhibitor tyrosinkinázy interferuje kromě produktu fúzního onkogenu *BCR-ABL* také s *PDGFR* a *c-*

KIT, které jsou silně exprimovány právě v chordomech (Meng *et al.*, 2019; Bompas *et al.*, 2015). Efektivitu imatinibu na základě těchto prediktivních biomarkerů hodnotili Stacchiotti *et al.* (2012) v jedné z doposud nejrozsáhlejších klinických studií chordomů s celkově 56 pacienty. Přežití bez progresu (PFS) a celkové přežití (OS) jsou důležitými ukazateli výsledků léčby, jejichž medián v tomto případě dosahoval 9 a 35 měsíců. Podle kritérií RECIST 70 % ze sledovaných pacientů vykazovalo stabilní stav onemocnění. Protože tyto výsledky byly považovány za úspěšné, brzy na analýzu účinků imatinibu v chordomech navázali také Hindi *et al.* (2015) v retrospektivní studii s 46 pacienty. Z tohoto počtu 74 % případů dosáhlo stabilního stavu a 26 % zůstalo s progresivním onemocněním. Pozorovaný medián PFS byl 9,9 měsíců a OS 30 měsíců.

Z významnějších analýz zaměřených na cílenou léčbu chordomů lze uvést také studii fáze II testující sorafenib s potenciálně inhibičními účinky kináz VEGFR provedenou Bompasem *et al.* (2015). Studie sledovala celkově 27 pacientů s progresivním onemocněním, z nichž 12 bylo již dříve léčeno chemoterapií a jinými molekulárně cílenými léčivy. Přítomnost specifických biomarkerů v tomto případě nebyla ověřována. Míra 9 měsíčního přežití bez progresu dosahovala 73 % a míra 12 měsíčního celkového přežití 86,5 %. Nicméně jak uvádějí Meng *et al.* (2019), použití sorafenibu je limitováno vyšším stupněm toxicity a s tím souvisejícími vážnými vedlejšími účinky.

Testována byla efektivita i jiných terapeutik, jako jsou erlotinib, cetuximab, sunitinib, afatinib a další, přičemž dané testy obvykle zahrnovaly jen malý počet pacientů (<20) nebo se jednalo pouze o případové studie. Z výsledků je však zřejmé, že použití těchto cílených léčiv na základě screeningu genetických mutací a imunohistochemie může být pro pacienty výhodnou léčebnou strategií (Meng *et al.*, 2019).

4.8 Rezistence nádorů k léčbě

Ačkoliv se cílená terapie může zdát jako slibné řešení v léčbě rakoviny, dosud nepřekonala jednu z nejzásadnějších překážek - rezistenci. Rezistence na léčiva se postupně vyvíjí u většiny pacientů s metastatickým onemocněním, kteří tak v konečném důsledku podléhají své nemoci. Významnou příčinou terapeutického selhání je intranádorová genetická heterogenita a je stále více zřejmé, že jednotlivé nádory mohou dosáhnout rezistence prostřednictvím hned několika cest současně (tzv. polyklonální rezistence) (Burrell *et al.*, 2014).

Na každou cílenou terapii bylo identifikováno mnoho mechanismů rezistence, běžně se však rozdělují do tří hlavních tříd vycházejících ze: (1) změny řídicích genů, (2) aktivace paralelní signální dráhy či podřazených komponent v kritické dráze, tj. aktivaci „downstream efektorů“ a (3) fenotypové transformace. Poslední třída rezistence zahrnuje histologickou proměnu z jedné buněčné linie, jako je epitelální, na jinou, jako je neuroendokrinní nebo mesenchymální. Daný proces byl zatím jen velmi málo vysvětlen (Neel *et Bivona*, 2017; Lin *et Shaw*, 2016).

Obecně lze všechny zmíněné typy rezistence považovat za projev vývoje nádorových buněk pod selekčním tlakem terapie. Mimo tuto tzv. získanou rezistenci existuje i rezistence původní (primární), která vzniká v důsledku přítomnosti některého rezistenčního mechanismu v nádorových buňkách ještě před zahájením samotné léčby (Lin *et Shaw*, 2016; Burrell *et Swanton*, 2014). Ať už je ale příčina rezistence v nádorových buňkách přítomna od počátku léčby, nebo se vyvíjí až později s narůstající heterogenitou nádoru, je pro úspěšnou cílenou terapii zcela zásadní daný mechanismus detekovat včas. V mnoha případech je již dnes pro pacienty k dispozici více léků, které mohou být různě kombinovány s cílem zastavit vznikající rezistenci ještě před jejím úplným propuknutím. S tím však souvisí také nutnost monitorovat změny nádorového genomu opakovaně, v průběhu celé léčby (Neel *et Bivona*, 2017). Pro tyto účely se nyní jako ideální metoda nabízí tekutá biopsie, jež umožňuje citlivě analyzovat cirkulující nádorovou DNA neinvazivním způsobem z krve pacienta (viz kapitola 4.9.2).

4.9 Metody nádorové genomiky

V diagnostických laboratořích se stále využívají základní metody detekce genetických změn, jako jsou imunohistochemie, fluorescenční *in situ* hybridizace (FISH), kvantitativní PCR a Sangerovo sekvenování, popřípadě pyrosekvenování (Burghel *et al.*, 2015; Garinet *et al.*, 2018). Jejich výhodou je přímočaré provedení, interpretace i validace. Typicky se jedná o různorodé testy, které sdílejí poměrně levné vybavení snadné na použití a údržbu (Kuo *et al.*, 2017). Nicméně tyto metody mají jen omezenou škálovatelnost a po vztažení na počet analyzovaných genů bývají pracné, pomalé a drahé. Často navíc vyžadují větší množství vzorku nebo DNA a vykazují vyšší míru selhání (Burghel *et al.*, 2015). NGS technologie se proto staly bezesporu převratným nástrojem v klinice, umožňující rozsáhlou analýzu genomických variant s vynikajícím rozlišením a přesností na základě efektivně generovaných sekvenčních dat (Xuan *et al.*, 2013).

4.9.1 Přehled současně nejběžnějších NGS technologií

Historie metod sekvenování nové generace (NGS), jinak známých jako masivně paralelní sekvenování, začíná teprve v nedávném roce 2004 zavedením prvního systému GS FLX od Roche 454. Tento nový typ sekvenátoru brzy následoval celý příliv NGS technologií, které do dnešního dne kontinuálně navyšují rychlost, a naopak snižují cenu sekvenační analýzy. Oproti Sangerově metodě je nyní možné sekvenovat nesrovnatelně více genomických regionů v jednom čase, a to bez požadavku předchozí znalosti sekvence (Ambardar *et al.*, 2016; Kamps *et al.*, 2017). Protože tento přístup zároveň umožňuje detekovat v zásadě všechny typy mutací, přináší NGS sekvenování téměř bezmezný zdroj dat pro genomiku včetně oblasti onkologie (Wang *et Xu*, 2017).

NGS metody lze rozdělit na dvě samostatné skupiny: sekvenování druhé a třetí generace (Ambardar *et al.*, 2016). Sekvenování druhé generace (SGS) se vyznačuje kratší délkou čtení (<500 bp) a nižší cenou na počet osekvenovaných bází. V potaz však musíme brát také doprovodné náklady na analýzu a uchovávání obrovského objemu dat, které jsou nezbytnou součástí těchto přístupů (Kamps *et al.*, 2017; Xuan *et al.*, 2013). Mezi přední zástupce SGS technologií patří Illumina, dále pak SOLiD, Ion Torrent, ustupující technologie Roche 454 a nově i Qiagen (Kamps *et al.*, 2017; Ambardar *et al.*, 2016).

4.9.1.1 Sekvenování metodou Illumina

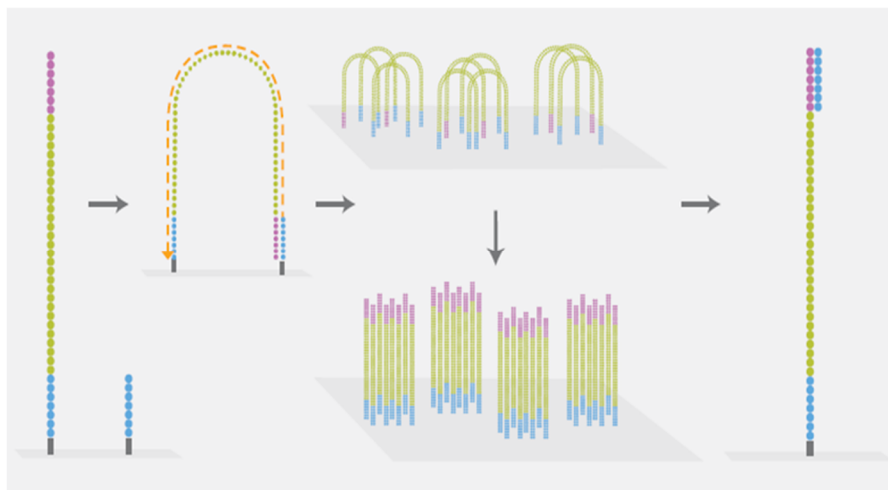
Technologie Illuminy se ve svém principu velmi podobá Sangerovu sekvenování založeném na syntéze DNA v přítomnosti polymerázy a fluorescenčně značených nukleotidů. V obou případech jsou nukleotidy na základě komplementarity začleněny do řetězce DNA a identifikovány excitací příslušného fluoroforu. Díky modifikované -OH skupině na 3' uhlíku je zabráněno další polymeraci, a to buď trvalým nebo reverzibilním způsobem. Hlavní rozdíl těchto přístupů však spočívá právě v počtu paralelně analyzovaných DNA fragmentů, který se u sekvenační reakce Illuminy pohybuje v řádech milionů až miliard (Chen *et al.*, 2013; Illumina, 2019h).

Typický postup sekvenování metodou Illuminy zahrnuje čtyři základní kroky:

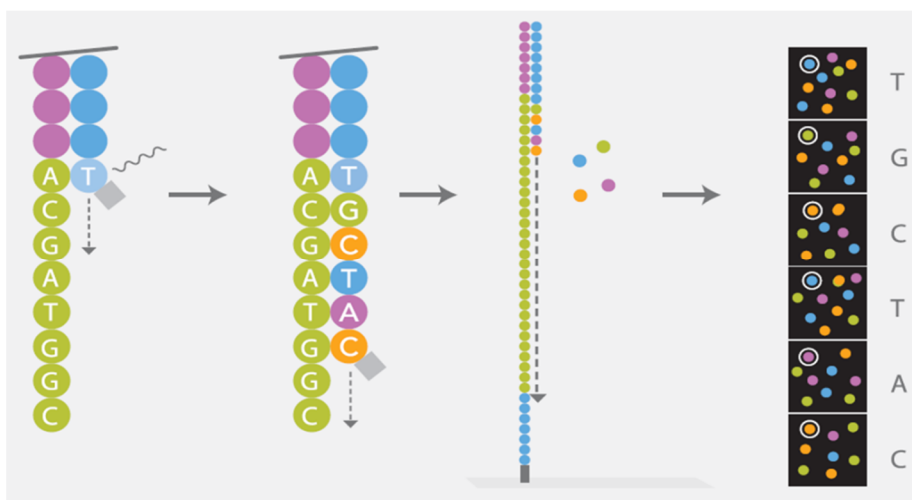
1. Příprava sekvenační knihovny: Celý proces obvykle začíná štěpením vzorku DNA nebo cDNA za vzniku náhodných fragmentů s velikostí do 800 bp, k nimž jsou následně naligovány 5' a 3' adaptory. Alternativně „tagmentace“ kombinuje fragmentační a ligační reakci v jediném kroku. Fragmenty s adaptory jsou poté

amplifikovány pomocí PCR a purifikovány např. metodami na bázi gelu (Koubková *et al.*, 2014; Illumina, 2019h).

2. Generování klastrů (Obrázek 2): Knihovna je vnesena do průtokové komůrky, kde jsou fragmenty zachyceny na povrchově vázaných oligonukleotidech komplementárních k adaptorům. Každý fragment pak generuje odlišný klonální shluk (klastr) pomocí můstkové amplifikace. Díky tomuto kroku je zesílen fluorescenční signál detekovaný v průběhu následného sekvenování (Ambardar *et al.*, 2016).
3. Sekvenační reakce (Obrázek 3): Technologie sekvenování syntézou (SBS) od Illuminy využívá princip reverzibilní terminace, kdy je v cyklech: 1) přidávána reakční směs polymerázy a značených dNTP do průtokové komůrky, 2) snímán signál pomocí excitace laserem, kdy je pro různé vlnové délky zaznamenána intenzita a pozice fluorescenčního signálu emitovaného při začlenění konkrétního nukleotidu do řetězce a 3) odštěpována terminační skupina na 3' uhlíku nukleotidu umožňující opakování celého procesu (Ambardar *et al.*, 2016; Illumina, 2019h).
4. Analýza dat: Sekvence každého klastru je generována specifickým algoritmem přidávající jednotlivým bázím určitou hodnotu, na jejímž základě jsou vyřazeny sekvence nízké kvality (Koubková *et al.*, 2014). Nově identifikovaná sekvenační čtení jsou poté bioinformatickým softwarem zarovnána (alignována) k referenčnímu genomu. Další fáze analýzy se může odvíjet značně odlišnými směry podle požadované aplikace NGS sekvenování (Illumina, 2019h).



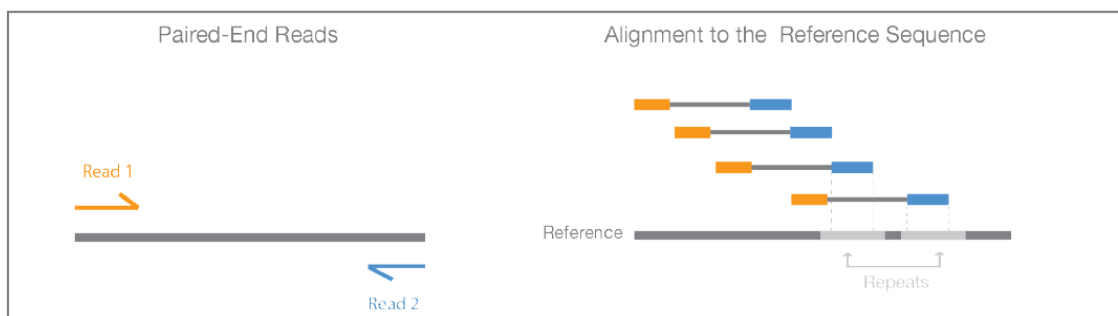
Obrázek 2: Generování klastrů na platformě Illumina. Fragменты DNA jsou přes adaptory přichyceny k průtokové komůrce a amplifikovány pomocí můstkové PCR. Amplifikační reakce představuje opakovaný proces hybridizace, elongace a denaturace, přičemž funkci primerů zde plní oligonukleotidy pokrývající průtokovou komůrku (Koubková *et al.*, 2014). Každý fragment DNA ve výsledku tvoří samostatný shluk identických molekul (Převzato z: CeGaT GmbH, 2019).



Obrázek 3: Princip sekvenační reakce Illuminy. Sekvenování vychází z procesu syntézy DNA v přítomnosti odlišně značených nukleotidů nesoucích reverzibilní terminační skupinu (znázorněna pomocí šedého čtverce). V jednotlivých cyklech je snímán fluorescenční signál z právě začleněného nukleotidu, a to paralelně na pozici každého klastru. Sekvence „TGCTAC“ byla z vyznačeného klastru v obrázku získána po šesti cyklech (Převzato z: CeGaT GmbH, 2019).

Sekvenační metoda využívá strategii párových koncových čtení, tedy sekvenování obou konců DNA fragmentů v knihovně, které zajišťuje přesnější výsledky a účinnější detekci indelů oproti „single-read“ přístupu (Obrázek 4). Analýza diferenčního rozestupu mezi páry čtení umožňuje také odstranění PCR duplikátů, jež jsou běžným artefaktem způsobeným amplifikací během přípravy knihovny (Illumina, 2019h). Obvyklá délka jednoho čtení je

u sekvenátorů Illumina 150 bp, případně až 300 bp s použitím systému MiSeq (Illumina, 2019i).



Obrázek 4: Metoda párových koncových čtení: Sekvenování je prováděno na obou koncích fragmentů DNA, přičemž vzdálenost mezi páry čtení je známa. Algoritmy aligmentu tak mohou dané informace využít k přesnějšímu mapování sekvencí, zejména v oblasti repetit (Převzato z: Illumina, 2019h).

Illumina se řadí k nejpřesnějším NGS metodám s dosahovanou chybovostí 0,4 % (Ambardar *et al.*, 2016). Nabízí navíc velmi flexibilní a škálovatelné platformy, od stolních sekvenátorů iSeq 100 s výkonem 1,2 Gb určených k panelovému sekvenování, až po systém NovaSeq 6000 pro rozsáhlé populační studie generující působivých 6 Tb a 20 miliard čtení v průběhu přibližně dvou dní (Tabulka 3) (Illumina, 2019h; Illumina, 2019i).

Tabulka 3: Současně dostupné sekvenační platformy Illumina (Illumina, 2019i).

Typ sekvenátoru	Výstupní kapacita	Počet čtení	Délka běhu
iSeq 100	1,2 Gb	4 miliony	9–17,5 hod
MiniSeq	7,5 Gb	25 milionů	4–24 hod
MiSeq	15 Gb	25 milionů	4–55 hod
NextSeq	120 Gb	400 milionů	12–30 hod
HiSeq 4000	1500 Gb	5 miliard	1–3,5 dní
NovaSeq 6000	6000 Gb	20 miliard	13–44 hod

4.9.1.2 Sekvenační systém Ion Torrent

Ion Torrent představuje ze současných SGS přístupů slibnou sekvenační metodu, která se již v mnohých studiích ukázala jako vhodná alternativa k Illumině (Hirsch *et al.*, 2018; Burghel *et al.*, 2015). Sekvenátory tohoto typu jsou založeny na polovodičovém čipu schopném detekovat změny pH (Koubková *et al.*, 2014). Samotné sekvenování předchází i v tomto případě příprava knihovny, kdy jsou fragmenty DNA ligovány k adaptorům a amplifikovány pomocí emulzní PCR na mikrokuličkách. Knihovna slouží během následné sekvenační reakce jako templát pro syntézu nového řetězce DNA (Grada *et Weinbrecht*, 2013). Mikrokuličky jsou po jedné umístěny do mikrojamek, kam je v cyklech přidávána reakční

směs obsahující vždy pouze jeden typ dNTP. Pokud je nukleotid inkorporován do narůstajícího řetězce, dojde k uvolnění vodíkového iontu a tento chemický signál je přímo převáděn do digitální podoby (Ambardar *et al.*, 2016).

Protože technologie Ion Torrent obchází nutnost použití fluorescenčně značených nukleotidů, jsou náklady spojené s analýzou sníženy (Ambardar *et al.*, 2016; Reuter *et al.*, 2015). Hlavní předností metody je rychlost, která se může pohybovat až k pouhým 2,5 hodinám na jeden běh (Mehrotra *et al.*, 2017). Nicméně výstupní kapacita zdaleka nedosahuje platform od Illuminy a sekvenování je zároveň méně přesné s celkovou chybovostí 1,78 % (Ambardar *et al.*, 2016). Metoda je náchylná zejména na detekci homopolymérů, tj. sekvenci identických nukleotidů, kdy při naměřeném signálu nad určitou mez dochází k chybnému stanovení počtu nukleotidů v dané repetici (Ivády *et al.*, 2018). Jak uvádějí Mehrotra *et al.* (2017), přesto lze tento sekvenátor s limitem detekce 5 % alelické frekvence při vysoké reprodukovatelnosti považovat za užitečný nástroj pro molekulární profilování nádorů v klinické praxi. Od roku 2018 Thermo Fisher Scientific navíc vypouští na trh vylepšenou sérii Ion GeneStudio S5, která slibuje výrazné navýšení kapacity a rychlosti analýzy (Thermo Fisher Scientific, 2019d). Nejvýkonnější platforma s čipem Ion 550 může generovat na dva sekvenační běhy v průběhu jednoho dne výstup až 50 Gb (Thermo Fisher Scientific, 2019e).

4.9.1.3 Sekvenační systém Qiagen

Srovnatelné výsledky s Ion Torrent přináší také jedna z nejnovějších sekvenačních technologií GeneReader od Qiagen (Heeke *et al.*, 2018). Metoda vychází ze sekvenování syntézou s cyklickou reverzibilní terminací; využívá tedy stejného detekčního mechanismu jako Illumina (Garrido-Cardenas *et al.*, 2017). Samotná příprava knihovny je založená na obohacení cílových oblastí pomocí PCR, přičemž sekvenační reakci je možné provést nejen na platformě GeneReader, ale i na jiných sekvenátorech (Kamps *et al.*, 2017). Nový systém byl navržen speciálně pro potřeby rutinní molekulární diagnostiky a jako takový přináší vysoce automatizovaný pracovní postup včetně integrované bioinformatické analýzy. Zajištění kompletní přípravy vzorků až po samotnou analýzu dat má zvýšit účinnost a reprodukovatelnost celého procesu. Za výhodu GeneReader může být považován i specifický způsob platby umožňující minimální počáteční náklady, které narůstají až za každý testovaný vzorek (Heeke *et al.*, 2018).

4.9.1.4 Technologie sekvenování třetí generace

Sekvenování třetí generace dosáhlo výrazného pokroku především v délce čtení. Kromě toho není vyžadována předchozí amplifikaci DNA, což zkracuje dobu přípravy vzorků a snižuje chybovost, která z tohoto kroku pramení (Xuan *et al.* 2013).

Technologie SMRT (single molecule real-time) vyvinutá společností Pacific Biosciences generuje délku čtení bažně až 40 kb či dokonce delší (Pacific Biosciences, 2019). Během sekvenačního procesu touto metodou je v přítomnosti fluorescenčně značených nukleotidů syntetizováno komplementární DNA vlákno polymerázou, která je ukotvena na dně struktury zvané „zero mode waveguide“ (ZMW). ZMW struktury si lze představit jako jamky v hliníkové folii s průměrem přibližně 70 nm. Vzhledem k miniaturním rozměrům jamek je možno zachytit fluorescenční signál uvolněný jen z právě inkorporovaného nukleotidu (Ambardar *et al.*, 2016). Celková chybovost sekvenování 11-15 % je však poměrně vysoká a výstupní kapacita 0,5-1 Gb na jeden čip (SMRT cell) naopak nízká (Rhoads *et al.*, 2015).

Velmi dlouhých čtení nad 150 kb bylo dosaženo pomocí MinION, přenosného sekvenátoru vážícím sotva 100 g. Tento systém od Oxford Nanopore Technologies využívá proteinových nanopórů, jimiž procházejí jednořetězcové molekuly DNA. Nanopór funguje jako biosenzor, který měří změny iontového proudu způsobenými rozdílnou strukturou jednotlivých bází. Metoda tímto způsobem umožňuje detekovat také epigenetické modifikace (Jain *et al.*, 2016). Uváděná přesnost sekvenátoru se však pohybuje pouze kolem 85 % (Rang *et al.*, 2018).

Kvůli vysoké chybovosti technologií, jako jsou SMRT nebo MinION, bývá prováděno vícenásobné čtení a vytvoření konsenzuální sekvence s dosažením přesnosti až nad 99 % (Rang *et al.*, 2018; Rhoads *et al.*, 2015). Tento přístup je ale také nákladnější, vzhledem k čemuž je sekvenování s krátkým čtením stále používanější metodou. SGS se navíc hodí i pro analýzu méně kvalitní, fragmentované DNA. Sekvenování s dlouhým čtením naopak nabízí výhodu zejména pro detekci strukturních variant, chromosomálních přestaveb a translokací. Díky jednodušší bioinformatice pak může být také velmi usnadněno *de novo* skládání genomů (Kamps *et al.*, 2017).

4.9.2 Vliv kvality vzorků na výsledky NGS analýzy a metoda tekuté biopsie

Navzdory velkému potenciálu NGS analýzy mohou být výsledná data kvalitní jen do takové míry, jako samotná zdrojová tkáň (Collins *et al.*, 2017). V tomto ohledu proto představují

velkou výzvu tzv. FFPE vzorky, nejdostupnější materiál pro genetické testování pacientů s rakovinou. Jedná se o formalínem fixované nádorové biopsie zalité v parafínu, jejichž primárním záměrem je zachovat buněčnou strukturu nádoru pro histologické vyšetření a diagnostiku. Fixace formalínu však způsobuje zesíťování jednotlivých molekul DNA včetně proteinů, což dělá tento proces značně nešetrný k následné genomické analýze. Dochází nejenom k poškození DNA z důvodů její fragmentace, ale i vzniku sekvenačních artefaktů, jako jsou změny bází C>T a G>A (Wong *et al.*, 2014). Kvůli heterogenní povaze nádorů navíc biopsie z malých oblastí vůbec nemusí reprezentovat celkovou buněčnou architekturu a biomarkery nádorové tkáně. Kromě toho pak izolované vzorky nejsou jen směsí maligní genetické informace, ale i DNA a RNA z nezhoubné části tkáně (Joo *et al.*, 2013; Gerlinger *et al.*, 2012).

Jako nejpřímější řešení k výše uvedeným problémům se přirozeně nabízí použití mnohočetných čerstvých biopsií z různých částí primárního nádoru i sekundárních metastáz. Takový invazivní postup by však znamenal významné riziko pro pacienta, nehledě na doprovodné náklady a reálnou proveditelnost (Collins *et al.*, 2017). V poslední době proto přilákala velkou pozornost tzv. tekutá biopsie, jako možnost detekce somatických mutací solidních nádorů z volně cirkulující DNA (circulating cell-free DNA, cfDNA) v krvi a dalších tělesných tekutinách pacienta. Předpokládá se, že cfDNA jsou uvolňovány z nádorových buněk během apoptózy a/nebo nekrózy, případně z nádorů odvozených exosomů (Han *et al.*, 2017). cfDNA tímto mohou poskytovat cenné informace o mutačním profilu mnoha buněčných populací, které se vzájemně podílí na průběhu nádorového onemocnění (Finotti *et al.*, 2018). Reálnou vypovídající hodnotu cfDNA ověřili ve své studii mimo jiné Janku *et al.* (2015) sledováním mutačního statusu genů *BRAF*, *EGFR*, *KRAS* a *PIK3CA* s pomocí tekuté i klasické biopsie. Obě metody dosahovaly pro jednotlivé geny shodu 91 %, 99 %, 83 % a 91 %.

Nicméně i s tekutou biopsií souvisí několik praktických překážek, zejména omezené množství cfDNA, která má být analyzována. Molekuly cfDNA jsou v tělesných tekutinách naředěny nenádorovou DNA, a tvoří tak u pacientů s rakovinou variabilní frakci, která může sahát až pod hranici 0,01 %. K citlivé detekci cílových molekul proto klasické metody, jako je Sangerovo sekvenování, zdaleka nepostačují (Diehl *et al.*, 2008; Finotti *et al.*, 2018). Na základě toho se dnes rozvíjí celá řada novějších technik často založených na alelově specifické PCR, digitální PCR nebo NGS s molekulárními barkódy. Jednotlivé přístupy se liší nejen svou citlivostí, ale i rozsahem a finanční náročností (Han *et al.*, 2017).

4.10 Přístupy genomického profilování nádorů na základě rozsahu analýzy

Vývoj moderních vysokokapacitních metod, jako je NGS, podnítil rychlý přechod z testů jednotlivých genů ke komplexnějšímu profilování nádorových genomů (Burghel *et al.*, 2015). Sekvenování celých genomů a transkriptomů je však zatím otázkou zejména klinického výzkumu, přičemž hlavní uplatnění v klinické diagnostice dnes nacházejí panely genů o rozdílných velikostech (Jennings *et al.*, 2017).

4.10.1 Celogenomové sekvenování

Analýza celého genomu dříve představovala rozsáhlý multicentrický projekt, který stál miliardy a trval několik let. V současnosti je její provedení možné v jediné laboratoři během několika dní, a to v ceně přibližně 1000 dolarů. Protože není vyžadován krok obohacení, je pracovní postup dokonce jednodušší než u cíleného sekvenování. Analýza však generuje velké množství dat; zásadní omezení tohoto přístupu tedy netkví v části chemické, ale výpočetní. Hlavní výhodou celogenomového sekvenování je zahrnutí „všeho“, přičemž fúze a změny v počtu kopií jsou komplexněji hodnoceny než u menších panelů a detekovatelné jsou i varianty na úrovni jednotlivých nukleotidů. Zvláště slibnou aplikací je „low-pass“ analýza celého genomu umožňující stanovení počtu kopií a detekci strukturních variant z fragmentované DNA, jenž je srovnatelná s cytogenetickou analýzou. Hlavní nevýhodou je však také zahrnutí „všeho“. Miliony variant, které jsou jedinečné pro každého člověka, je téměř nemožné vyhodnotit, nehledě na nízkou hloubku pokrytí spojenou se ziskem tolika sekvenčních dat na každého pacienta. Omezena je také možnost analyzovat subklony a heterogenní vzorky (Kuo *et al.*, 2017).

4.10.2 Celoxomové sekvenování

Sekvenování exomu nádorů se týká speciálních forem panelů genů založených na hybridizaci, které zahrnují všechny (nebo téměř všechny, protože některé geny nemohou být spolehlivě zachyceny či rozlišeny od homologních oblastí) známé lidské geny. Postup je technicky totožný s ostatními genovými panely s tím rozdílem, že množství generovaných dat je více než řádově vyšší (Kuo *et al.*, 2017). Pro dostatečně specifickou a sensitivní detekci mutací je vyžadováno pokrytí větší než 20x (Kamps *et al.*, 2017). Tento přístup proto také zvyšuje výpočetní požadavky i náklady na ukládání dat, navíc mnohé ze získaných genetických údajů zatím přesahují naši schopnost je vůbec nějak interpretovat. Navzdory významnému nárůstu informací pak není ani možné detekovat většinu

strukturních variant s intronovými signálními sekvencemi sestřihu, protože běžné komerčně dostupné kity introny nezachycují (Kuo *et al.*, 2017).

4.10.3 Transkriptomové sekvenování (RNA-seq)

RNA-Seq aplikuje technologie NGS pro kvalitativní a kvantitativní analýzu kompletní sady transkriptů (tj. transkriptomu), včetně mRNA, malých RNA a dalších nekódujících RNA. Profilování transkriptomu zachycuje informaci o genové expresi a regulačních prvcích v buňce, tkáni nebo organismu za různých fyziologických stavů, což je důležité pro pochopení biologických procesů ve vývoji i nemoci. Ačkoli transkriptom tvoří pouze malou frakci lidského genomu (<5 %), je obsahově velmi rozmanitý, přičemž zahrnuje také produkty odvozené z alternativního sestřihu a editace RNA či „antisense“ transkripty (Xuan *et al.*, 2013). Metoda RNA-Seq navíc přináší účinné řešení pro identifikaci genových fúzí, které hrají významnou roli v rakovinných procesech (Parker *et Zhang*, 2013).

4.10.4 Panelové sekvenování

Panelové sekvenování je typem cíleného re-sekvenování, které je zaměřeno pouze na určitý počet specifických genů či zájmových oblastí (Kamps *et al.*, 2017). V současné době panely genů představují vůbec nejpoužívanější NGS metodu pro diagnostické testování somatických mutací solidních nádorů a hematologických malignit (Jennings *et al.*, 2017). Již mnohé studie prokázaly cenovou efektivitu těchto metod i jejich schopnost identifikovat genetické změny, které jsou relevantní pro onkogenezi či rezistenci na léčiva, a to jak v celé populaci rakovinných buněk, tak v buněčných subklonech (Kamps *et al.*, 2017).

Omezení celkové velikosti sekvence jen na cílové oblasti umožňuje využít vysoký výkon NGS k analýze i více pacientů v jednom běhu. Jak rozsah sekvenování ovlivňuje množství současně testovaných vzorků, znázorňuje Tabulka 4 na příkladu genových panelů od Qiagen (zde rozsah analýzy vychází z počtu primerů použitých v amplifikačním obohacení). Panely genů lze kombinovat s různými sekvenačními platformami, z nichž nejvyužívanější jsou Illumina a Iont Torrent. Tyto se ověřily mimo jiné jako vhodná volba pro analýzu DNA o nižší kvalitě, včetně FFPE vzorků (Kamps *et al.*, 2017).

Tabulka 4: Množství vzorků, které lze sekvenovat současně v závislosti na rozsahu analýzy a kapacitě sekvenátoru pro genové panely od Qiagen při sekvenačním pokrytí 500x (Qiagen, 2019b).

Sekvenátor	Verze	Kapacita ^{1 2} (počet čtení)	Počet vzorků/jeden běh		
			1 000 primerů	2 500 primerů	12 000 primerů
MiSeq	v2	30 milionů	60	24	5
HiSeq 2500 rapid run	Dual Flowcell v2	1200 milionů	2400	960	200
Ion S5	Ion 530 Chip	15 milionů	21	8	1
Ion Proton	I Chip	80 milionů	112	44	9

¹ Kapacita je vyjádřena v počtu párových koncových čtení pro Illuminu a jednoduchých čtení pro Ion Torrent

² Počítá se s délkou čtení 2 x 150 bp (Illumina) a 200 bp (Ion Torrent)

Detekční limit analýzy závisí na hloubce pokrytí (Garinet *et al.*, 2018). Pro testování germinálních mutací, které jsou přítomny v homo- nebo heterozygotním stavu, dostačuje pokrytí 80x. Oproti tomu detekce somatických změn vyžaduje pokrytí několikanásobně vyšší (>500x) z důvodů obvyklého výskytu těchto mutací na subklonální hladině (Kamps *et al.*, 2017). Průměrný detekční limit panelů genů se pohybuje okolo 2 % alelické frekvence, což umožňuje analýzu mutací s nízkou frekvencí i přes malé množství vstupní nádorové DNA. U velkých panelů dosahujících až k velikosti celého exomu je mez detekce okolo 10 %, přičemž se nedoporučuje používat vzorky s nízkou kvalitou nebo vzorky s méně než 50 % obsahem nádorových buněk (Garinet *et al.*, 2018).

Detekce somatických mutací o velmi nízké frekvenci (<5 %) při limitním množství vstupního materiálu (<30 ng) je často spojená s metodou tekuté biopsie a cirkulující nádorovou DNA. Také pro analýzu za těchto podmínek byly přizpůsobeny některé panely genů, a to v kombinaci se sekvenováním s ultra vysokým pokrytím (>2000x) (Vollbrecht *et al.*, 2018; Bennett *et al.*, 2016; Page *et al.*, 2017).

Od hloubky sekvenování se kromě detekčního limitu odvíjí také sensitivita a specifita analýzy. Sensitivita je definována jako schopnost správně identifikovat variantu, která je skutečně přítomná (true positive). Specifita je naopak schopnost neidentifikovat variantu, která přítomna ve skutečnosti není (true negative). S rostoucí sekvenační hloubkou významně roste sensitivita analýzy, zároveň však klesá její specifita. Vyšší počet čtení totiž navýší nejenom pravděpodobnost identifikace správných, ale také falešných variant. Vliv na sensitivitu a specifitu má v obdobné korelaci také množství vstupní DNA (Illumina, 2019g). Tyto závislosti znázorňuje Tabulka 5.

Tabulka 5: Závislost sensitivity a specificity na množství vstupní DNA, sekvenačním pokrytí a detekčním limitu (Illumina, 2019g).

Vstupní DNA	Surové sekvenační pokrytí	Specificita	Sensitivita při limitu detekce (LOD)			
			LOD = 0,20 %	LOD = 0,40 %	LOD = 0,60 %	LOD = 0,80 %
10 ng	10,000x	100%	7.26%	30.06%	54.86%	73.74%
	20,000x	100%	13.72%	46.97%	73.65%	88.48%
	30,000x	100%	15.63%	51.06%	77.34%	90.82%
	40,000x	100%	17.27%	54.32%	80.07%	92.42%
30 ng	10,000x	99.99%	54.77%	92.60%	99.17%	99.92%
	20,000x	99.98%	73.62%	98.22%	99.92%	100%
	30,000x	99.98%	77.28%	98.79%	99.96%	100%
	40,000x	99.98%	80.02%	99.12%	99.98%	100%
50 ng	10,000x	99.93%	85.64%	99.61%	99.99%	100%
	20,000x	99.90%	95.33%	99.97%	100%	100%
	30,000x	99.92%	96.54%	99.99%	100%	100%
	40,000x	99.93%	97.32%	99.99%	100%	100%
70 ng	10,000x	99.83%	96.31%	99.98%	100%	100%
	20,000x	99.74%	99.35%	100%	100%	100%
	30,000x	99.79%	99.59%	100%	100%	100%
	40,000x	99.82%	99.72%	100%	100%	100%

Počet genů/cílových regionů tímto představuje důležitý faktor panelů udávající dosažitelnou hloubku sekvenování a s ní spojené další parametry. S větším počtem genů opět vyvstává i více otázek při následném zpracování dat, a to zejména jak se při klinickém rozhodování vypořádat s variantami neznámé významnosti. Také z technického hlediska se stává obtížnější kontrola kvality a validace větších panelů (Garinet *et al.*, 2018). V neposlední řadě pak stoupá cena a doba celé analýzy.

Nehledě na požadovaný rozsah testu ovlivní výběr vhodné platformy například i to, zda je cílem analyzovat solidní nádory nebo hematologické malignity. Na trhu jsou k dispozici nejen panely určené pro konkrétní typy nádorů, ale také panely méně specifikované obsahující sadu nejčastěji mutovaných genů asociovaných s rakovinou (Jennings *et al.*, 2017; Illumina, 2019a). Další dostupnou možností je vlastní návrh cílových oblastí (Keibl, 2016).

Panelové sekvenování umožňuje detekci jednonukleotidových variant (SNV), malých inzercí a delecí (indelů), změnu počtu kopií (CNA) i genových fúzí. V rámci jednoho panelu mohou být navrženy cílové sekvence zahrnující oblasti „hotspotů“ daného genu (např. exony 18 až 21 *EGFR*), případně pokrývající celé kódující i nekódující sekvence relevantní pro cílový gen (např. *KRAS*, *NRAS* nebo *TP53*). Dané uspořádání má zásadní význam, protože rozhoduje o potenciální schopnosti panelu detekovat jednotlivé typy mutací. Stanovení změn počtu kopií vyžaduje specifický přístup vycházející z počtu sond nebo ampliconů pokrývajících určitý gen. Odhadnutí počtu kopií z jedné „hotspotové“ oblasti v genu není tak přesné jako ze zprůměrovaného množství sond či ampliconů pokrývajících

všechny exonické regiony. Mez detekce (zejména pro genové ztráty) je silně závislá na frakci nádorových buněk přítomných v testovaném vzorku. Stanovení jiného typu mutací, genových fúzí, vyžaduje navržení sond zahrnující intronové sekvence, a to z důvodů nejčastějšího výskytu zlomů právě v těchto oblastech. Jako alternativní možnost se pak nabízí panelové sekvenování RNA (cDNA) (Jennings *et al.*, 2017).

Základem každého panelového sekvenování je obohacení cílových sekvencí, které může vycházet ze dvou hlavních strategií: PCR a hybridizace (Kuo *et al.*, 2017).

4.10.4.1 Obohacení založené na PCR

Metody obohacení založené na PCR využívají sadu primerů specificky navržených na požadované genomové sekvence, které v reakci dávají vzniknout mnoha překrývajícím se amplikonům (Xuan *et al.*, 2013). K cílovým sekvencím jsou připojeny indexy specifické pro jednotlivé vzorky a adaptory umožňující následnou NGS analýzu. Tento přístup je velmi univerzální, avšak náchylný na problémy spojené s chemií PCR primerů (Jennings *et al.*, 2017). Existence variant v místech hybridizace může snížit účinnost nasedání a způsobit alelické zkreslení (bias) nebo úplnou ztrátu alel (Xuan *et al.*, 2013). Příprava knihovny na bázi amplifikace není příliš efektivní ani pro geny s vysokým obsahem guanin-cytosinu (např. *CEBPA8*) nebo oblasti s vysoce repetitivními sekvencemi (Jennings *et al.*, 2017).

Zabránění nespecifickému vázání primerů, jejich křížových reakcí či tvorbě dimerů může vyžadovat rozsáhlou výpočetní analýzu. Z těchto důvodů mají také přístupy založené na PCR jen omezený rozsah, a to do několika megabáz (Xuan *et al.*, 2013). V porovnání s hybridizačním obohacením jsou však obecně rychlejší, méně nákladné a mohou poskytovat lepší citlivost pro subklonální populace nebo heterogenní vzorky (Kuo *et al.*, 2017).

4.10.4.2 Hybridizační obohacení

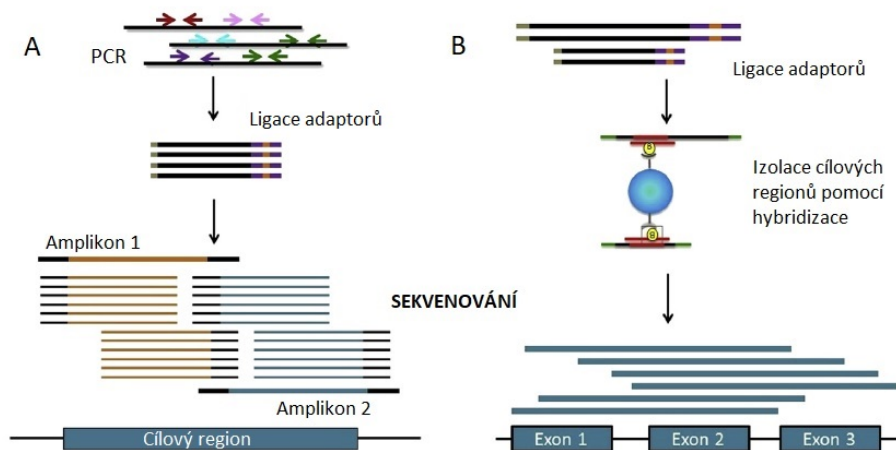
Metody obohacení na bázi hybridizace přímo vychytávají cílové sekvence pomocí komplementárních oligonukleotidů s typickou délkou 60-120 bp (Gray *et al.*, 2015; Kuo *et al.*, 2017). Může se tak dít na čipu (array), nebo častěji v roztoku za použití biotinylovaných jednovláknových RNA či DNA sond. Hlavní výhodou těchto metod je schopnost pokrýt velké genomové oblasti, a to až do celé šíře lidského exomu (~30 Mb) (Xuan *et al.*, 2013). Sondy jsou významně delší než PCR primery, mohou proto tolerovat přítomnost i několika neshod ve vazebném místě oligonukleotidu bez narušení specifické hybridizace, a zabránit tím ztrátě alel. Hybridizační metodou však často dochází nejenom k izolaci cílových, ale i

obklopujících sekvencí. To sice může přinášet nové informace do následné analýzy, ale zároveň snižovat pokrytí oblastí zájmu (Jennings *et al.*, 2017).

Obecné kroky zpracování vzorků pro hybridizační obohacení zahrnují počáteční fragmentaci DNA, několik enzymatických reakcí zajišťujících opravu konců, adenylaci a ligaci adaptorů a následnou PCR amplifikaci. Po přečištění je NGS knihovna připravena na hybridizaci s biotinylovanými sondami (Jennings *et al.*, 2017).

Modifikovaný postup nabízí systém HaloPlex od Agilent Technologies. Tato metoda vychází ze štěpení genomické DNA restrikcí enzymem, kdy jsou v navazujícím kroku obohacení použity biotinylované DNA sondy s navrženou homologií pouze na 5' a 3' konce oblastí zájmu. Tím je po hybridizaci sondy podpořena cirkularizace cílových sekvencí a současně zvýšena specifita této metody. Hybridní molekuly jsou poté vycity pomocí kuliček potažených streptavidinem; genomický materiál je promyt, cirkulární cílové sekvence zligovány a nakonec amplifikovány za použití primerů specifických pro sekvenační adaptory (Coonrod *et al.*, 2014).

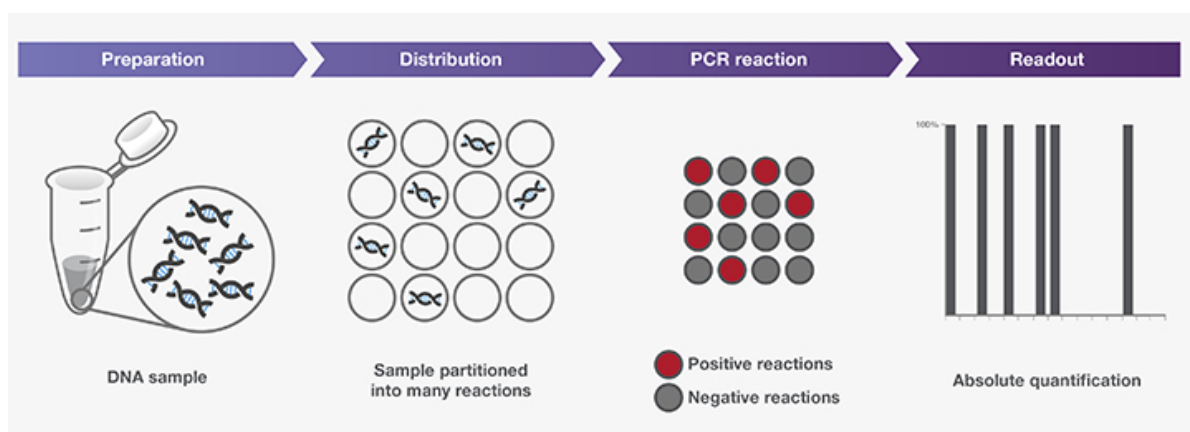
Nevýhodou hybridizačního obohacení je jeho náchylnost na bázevé složení vzorku. Může docházet jak ke ztrátě sekvencí bohatých na adenin-thymin z důvodů slabého navázání, tak oblastí s vysokým obsahem guanin-cytosinu kvůli tvorbě sekundárních struktur (Mamanova *et al.*, 2010). Porovnání základních rozdílů hybridizačního a amplifikačního obohacení shrnuje Obrázek 5.



Obrázek 5: Porovnání obohacení pomocí PCR a hybridizace: A) Amplifikační obohacení cílových oblastí dává vzniknout mnoha překrývajícím se amplikonům. Generovaná sekvenační čtení mají stejné počáteční a koncové souřadnice určené návrhem primerů. B) Hybridizační obohacení vychází ze zachycení cílových molekul pomocí biotinylovaných sond za použití magnetických kuliček. Sondy obvykle zachycují větší oblasti DNA, což vede k sekvenování nejen specifických cílů, ale i obklopujících regionů (Převzato a upraveno z: Jennings *et al.*, 2017).

4.10.5 Digitální PCR

Digitální polymerázová řetězová reakce je ke dnešnímu dni nejčastějším způsobem detekce cirkulující nádorové DNA (Zhang *et al.*, 2019). Vysokou přesnost a citlivost této amplifikační metody zajišťuje strategie rozdělení vzorku do mnoha identických miniaturních kompartmentů/kapének, díky nimž je možné provádět jednotlivé PCR nezávisle a současně v jedné reakci (Obrázek 6) (Liao *et Huang*, 2017). Metoda se tímto hodí zvláště pro detekci vzácných mutací, které jsou klasickou qPCR často jen obtížně zachytitelné z důvodů převládnutí „wild-type“ alel v reakční směsi. S využitím Poissonovy statistiky lze tímto přístupem navíc určit absolutní počet templátových molekul ve vzorku, a to bez kalibračních křivek, které jsou nutnou součástí postupu stanovení koncentrace DNA pomocí qPCR (Huggett *et Whale*, 2013; Liao *et Huang*, 2017).



Obrázek 6: Princip digitální PCR. Vysoce zředěný vzorek DNA je rozdělen do mnoha kompartmentů, které neobsahují více než jednu kopii cílové molekuly. Jednotlivé reakce následně poskytují pozitivní nebo negativní výsledek (vycházející z 1 nebo 0 templátových kopií; odtud „digitální“ PCR). Na základě Poissonova rozdělení může být ze získaných hodnot provedena absolutní kvantifikace (Převzato z: Thermo Fisher Scientific, 2019a).

V porovnání s NGS cíleným sekvenováním jsou však možnosti digitální PCR značně limitovány, jak ukázali například Zhang *et al.* (2018) v analýze nádorové cfDNA u pacientů s kolorektálním karcinomem. Přes vysokou sensitivitu a specificitu digitální PCR bylo možné oproti NGS metodě souběžně genotypovat pouze malý počet genů (*RAS*, *EGFR*, *NRAS*, *MEK1* a *BRAF*), a z těchto navíc jen vybrané „hotspotové“ mutace. NGS sekvenování podle této studie může poskytovat podobně kvalitní výsledky s výrazně širším rozsahem analýzy.

4.11 Srovnání dostupných technologií pro sekvenování panelů genů

Technologii panelového sekvenování vyvíjí celá řada společností, z nichž na současném trhu vedou zejména Illumina, Thermo Fisher Scientific, Qiagen, Agilent Technologies a Roche NimbleGen (Jennings *et al.*, 2017; Samorodnitsky *et al.*, 2015; Kamps *et al.*, 2017). K dispozici je proto nespočet typů panelů genů s rozmanitými výhodami i nevýhodami pro odlišné aplikace.

4.11.1 Cílené obohacení od Roche NimbleGen a Agilent Technologies

Roche NimbleGen a Agilent Technologies jsou hlavními představiteli technik obohacení na zakázku, které poskytují velmi variabilní testy přizpůsobené specifickým požadavkům každého uživatele. S širokými možnostmi panelového sekvenování však narůstá také potřeba objektivních studií, které by dostupné platformy systematicky ověřily a vzájemně porovnaly (Samorodnitsky *et al.*, 2015).

Na základě toho si Bodi *et al.* (2013) pro svou studii vyžádali od NimbleGen a Agilent návrh toho nejlepšího řešení k zachycení cílové oblasti o 3,5 Mb zahrnující 31 genů a 2Mb kontinuální interval, přičemž k tvorbě panelů poskytly oběma společnostem stejný vzorek genomické DNA.

Vybrány byly panely SureSelect Target Enrichment System (Agilent) a SeqCap EZ Choice (NimbleGen) založené na hybridizačním obohacení v roztoku. Následné sekvenování proběhlo na sekvenátorech Analyzer IIX od Illuminy. Sekvenční data byla hodnocena podle citlivosti, specificity a pokrytí cílových regionů, kdy obě platformy pro svou sadu sond podaly dobré výsledky. Z variant byly hodnoceny jednonukleotidové polymorfismy (SNP), které u všech vzorků vykazovaly ~98% shodu s databází (dbSNP). Agilent však prokázal jen omezené možnosti v designu panelu, což zásadně ovlivnilo úspěšnost dosaženého obohacení. Test tímto pokryl pouze necelou polovinu 3,5 Mb požadované cílové oblasti (Bodi *et al.*, 2013). Podobné srovnání SureSelect a SeqCap později provedli i Samorodnitsky *et al.* (2015) na panelech s 257 geny souvisejících s rakovinou, kteří do své studie zahrnuli také informace o množství vstupní DNA, ceně a době analýzy (Tabulka 6).

Tabulka 6: Vlastnosti panelového sekvenování Agilent Technologies a Roche NimbleGen (Samorodnitsky *et al.*, 2015)

	SureSelect Custom Target Enrichment	SeqCap EZ Choice
Délka sond (bp)	120	50-105
Typ sond	RNA	DNA
Počet sond (v dané studii)	14 368	2,1 x 10 ⁶
Doporučený rozsah obohacení	1 kb - 24 Mb	1 kb - 50 Mb
Vstupní množství DNA	2 µg	2 µg
Doba přípravy knihovny	72 hod	72-96 hod
Cena na jednu knihovnu	568 USD	759 USD

Roche NimbleGen a Agilent byly dále ověřeny pro možnost použití FFPE vzorků karcinomu plic v rámci celoexomového sekvenování pomocí kitů SeqCap EZ Human Exome v.3.0 a SureSelect XT Human All Exon v.5. Obě technologie bez výrazných rozdílů poskytly vysoce kvalitní knihovny a sekvenační data z relativně nízkého vstupu značně fragmentované genomické DNA (300 ng). Přestože u detekovaných inzercí/delecí byla shoda nižší, procento sdílených SNV dosahovalo nad 90 %. Na základě dané studie mohou být obě metody celoexomového sekvenování doporučeny pro translační výzkum i klinické účely (Bonfiglio *et al.*, 2016).

K panelovému sekvenování od Agilent se řadí také novější systém HaloPlex, který představuje rychlejší řešení oproti SureSelect a zároveň vyžaduje menší množství vstupní DNA. Samorodnitsky *et al.* (2015) však ukázali, že příprava knihovny založená na enzymatické fragmentaci způsobuje nízkou uniformitu pokrytí a méně spolehlivou detekci SNV.

V současnosti Agilent ke svým panelům nabízí reagenční kity SureSelect^{XT HS}, které díky molekulárním barkódům umožňují dosažení limitu detekce pod 1 % alelické frekvence. Tyto testy jsou optimalizované pro FFPE vzorky a nízké množství vstupní DNA, jako je 10 ng (Agilent Technologies, 2019).

4.11.2 Panelové sekvenování Illumina

K analýze somatických mutací o větším rozsahu lze kromě celoexomového sekvenování využít panel zahrnující exonické regiony 409 genů asociovaných s rakovinou (The AmpliSeq for Illumina Comprehensive Cancer Panel). Stejně velký panel poskytuje také Thermo Fisher Scientific; podobnosti a rozdíly těchto dvou přístupů proto uvádí Tabulka 7 (viz kapitola 4.11.3).

O něco menší genový panel představuje TruSight Tumor 170, který pokrývá 170 genů spojených s běžnými typy solidních nádorů. Panel umožňuje analýzu genových fúzí, sestřihových variant, inzercí/delecí, SNV i amplifikací. Za použití 40 ng vstupního materiálu (DNA a/nebo RNA) včetně FFPE vzorků je možné detekovat varianty s frekvencí výskytu ~ 5 %. Příprava knihovny je založená na hybridizačním obohacení s délkou pracovního postupu dva dny (Illumina, 2019f). V kombinaci s reagenčním kitem TruSight Oncology UMI Reagents lze pak pomocí molekulárních barkódů snížit chybovost sekvenanční analýzy až na <0,007 % a detekovat varianty s frekvencí výskytu ~0,4 %. Takto modifikovaný test je určen pro cfDNA bez možnosti použití FFPE vzorků, nicméně 30 ng doporučeného množství vstupního materiálu je stále poměrně vysoké (Illumina, 2019c; Illumina, 2019d).

Illumina poskytuje také panely založené na amplifikačním obohacení určené pro limitní výchozí materiál v množství 1 ng (AmpliSeq for Illumina Cancer Hotspot Panel v2). Za těchto podmínek se však doporučuje použití pouze vysoce kvalitní genomické DNA, pro FFPE vzorky je analýza vhodná až nad 10 ng. Jedná se o „hotspotový“ panel zahrnující ~2800 známých somatických mutací (SNP a inzercí/delece) v 50 genech souvisejících s rakovinou. Detekční limit testu je uváděn kolem 5 %, příprava knihovny trvá 5 hodin a oproti výše zmíněným panelům je sekvenování prováděno na menších sekvenátorech, jako je MiSeq (Illumina, 2019e).

4.11.3 Panelové sekvenování Thermo Fisher Scientific

Thermo Fisher Scientific se mimo jiné zaměřuje na cílené re-sekvenování v kombinaci se sekvenátory Ion Torrent. Z největších platforem je nabízen Ion AmpliSeq Comprehensive Cancer Panel pokrývající všechny exony 409 genů asociovaných s rakovinou (Tabulka 7). Singh *et al.* (2014) genový panel otestovali v klinické diagnostické laboratoři pro detekci somatických mutací v solidních nádorech a svými výsledky potvrdili vysokou reprodukovatelnost i citlivost tohoto testu.

Panel s 52 geny (Oncomine Pan-Cancer Cell-Free) přináší možnost analýzy volně cirkulujících nukleových kyselin (DNA a RNA) získaných z jediné 10ml zkumavky krve, kdy je při vstupním množství cfDNA 5 ng slibován detekční limit až 0,1 % (SNV). Test pokrývá „hotspotové“ oblasti vybraných genů zahrnující SNV, inzercí/delece, CNV a genové fúze. Metoda vychází z amplifikačního obohacení a zahrnuje použití molekulárních barkódů pro snížení falešných pozitiv. Příprava knihovny trvá pouhé 4 hodiny a je vhodná i

pro FFPE vzorky, k analýze je však vyžadován nový typ sekvenátoru Ion GeneStudio S5 (Thermo Fisher Scientific, 2019b; Li *et al.*, 2018).

Tabulka 7: Srovnání panelů od Illuminy a Thermo Fisher Scientific pokrývajících 409 genů (Illumina, 2019b; Thermo Fisher Scientific, 2019c; Singh *et al.*, 2014)

	The AmpliSeq for Illumina Comprehensive Cancer Panel	Ion AmpliSeq Comprehensive Cancer Panel
Metoda obohacení	Multiplexní PCR	Multiplexní PCR
Počet ampliconů	15 992	16 000
Doba přípravy knihovny	6 hod	16 hod
Typ vzorku	DNA včetně FFPE vzorků	DNA včetně FFPE vzorků
Množství vstupního materiálu	1 ng vysoce kvalitní gDNA nebo 10–100 ng FFPE	40 ng
Typy detekovaných variant	CNV, inzerce/delece, SNP	CNV, inzerce/delece, SNP
Detekční limit	~5 %	~5 %
Sekvenátor	NextSeq	Ion PGM System

4.11.4 Genové panely Qiagen

Další řadu genových panelů přizpůsobených pro analýzu somatických variant z FFPE vzorků a cfDNA představují platformy QIAseq od Qiagen, které lze navíc výhodně kombinovat i se sekvenátory typu Illumina či Ion Torrent. Příprava knihovny vychází z amplifikačního obohacení za použití dvou druhů primerů, cílově specifických a univerzálních, díky čemuž je snížen celkový počet vyžadovaných oligonukleotidů na reakci. Největší rozsah z platform QIAseq zaměřených na rakovinu umožňuje Human Tumor Mutational Burden Panel, existuje k němu však jen omezené množství informací z důvodů jeho teprve nedávného uvedení na trh. Panely QIAseq jsou přizpůsobeny pro detekci SNV, malých inzercí/delecí a CNV a součástí analýzy jsou také molekulární barkódy výrazně zvyšující přesnost panelového sekvenování. Limit detekce SNP při 90% sensitivitě je s ohledem na výchozí materiál a hloubku pokrytí vyjádřen v Tabulce 8 (Qiagen, 2019a; Qiagen, 2019b).

Tabulka 8: Detekční limit panelů QIAseq při 90% sensitivitě v závislosti na množství vstupní DNA a sekvenačním pokrytí (Qiagen, 2019b).

Frekvence variant (SNP)	Vstupní množství čerstvé DNA	Průměrné pokrytí
5 %	10 ng	7200
1 %	40 ng	25 600

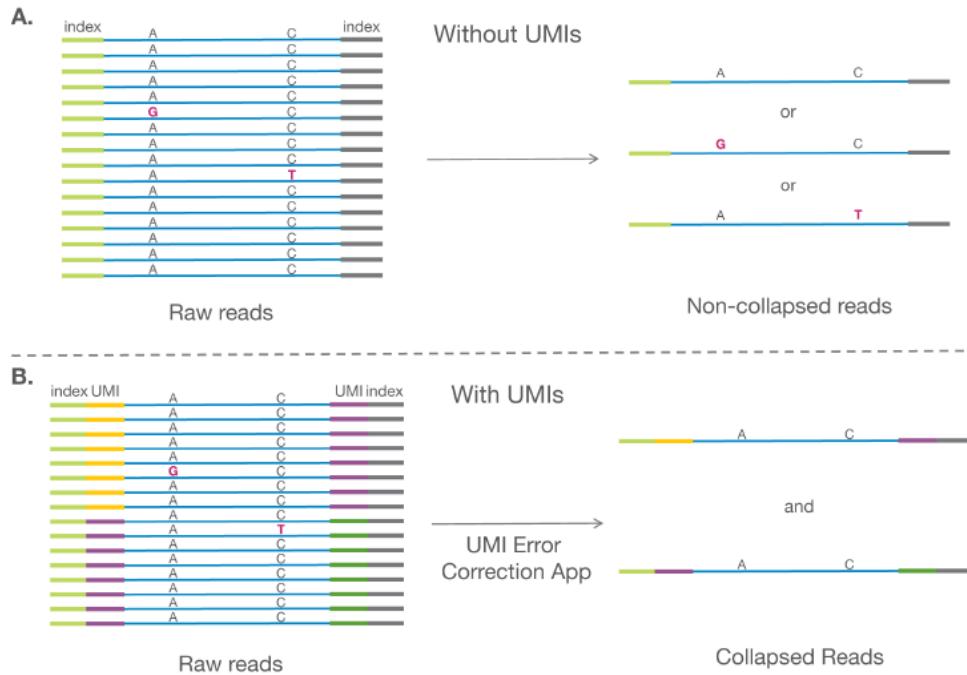
4.11.5 Genové panely NEB

Panel NEBNext Direct HotSpot Cancer Panel pokrývá 190 „hotspotových“ oblastí s celkovou velikostí 40 kb napříč 50 nejběžnějších genů spojených s rakovinou. Strategie obohacení je založená na hybridizaci pomocí biotinylovaných oligonukleotidových sond, které určují 3' konec oblastí zájmu. Hybridní molekuly tvořené cílovou DNA a sondou jsou v reakci zachyceny streptavidinovými kuličkami s následným enzymatickým odstraněním 3' přesahujícího konce DNA. Jak výrobce uvádí, tento krok v kombinaci s krátkou dobou hybridizace má ve srovnání s běžnými metodami hybridizačního obohacení zajistit vyšší účinnost sekvenční analýzy (New England Biolabs, 2019a). Protože jsou zároveň součástí analýzy molekulární barkódy, je metoda určená i pro malé množství vstupního materiálu s nízkou kvalitou včetně FFPE vzorků a cfDNA (10 ng – 1 µg). Příprava knihovny je proveditelná během jediného dne (New England Biolabs, 2019b).

V zásadě identické parametry platí také pro panel NEBNext Direct BRCA1/BRCA2, jenž je navržený k obohacení kompletních exomových oblastí genů *BRCA1* a *BRCA2*. Panel je tak vhodný nejen pro vysoce citlivou analýzu germinálních, ale i somatických mutací (New England Biolabs, 2019c).

4.12 Použití molekulárních barkódů pro zvýšení přesnosti sekvenční analýzy

Molekulární barkódy nebo také UMI barkódy (unique molecular identifiers) jsou obvykle navrženy jako řetězec náhodných nukleotidů, které jsou připojeny k templátovým molekulám DNA v průběhu přípravy knihovny, a to ještě před provedením PCR. UMI vytváří odlišnou identitu pro každou vstupní molekulu, čehož lze využít k identifikaci a korekci zkreslení způsobeným krokem amplifikace. V principu se předpokládá, že po PCR jsou molekuly sdílející UMI odvozeny ze stejného výchozího templátu (Obrázek 7) (Sena *et al.*, 2018). Během zpracování sekvenčních dat může být na základě toho proveden „kolaps“ čtení s identickými UMI, díky kterému jsou odfiltrovány falešné pozitivita a současně zpřesněna identifikace skutečných variant vyskytujících se při nízkých frekvencích (Illumina, 2019g).



Obrázek 7: Korekce chyb pomocí UMI. A) Na obrázku je celkem 16 čtení se dvěma variantami, které mohou představovat buď skutečné varianty, nebo artefakty. Bez UMI není možné správnou a falešnou pozitivitu rozlišit. B) Integrace UMI umožňuje seskupit více čtení ze stejné výchozí molekuly do jedné společné „rodiny“. Každá rodina čtení zde obsahuje jednu chybu, po korekci tak zůstává pouze jediná správná sekvence.

UMI nelze zaměňovat s barkódy specifickými pro jednotlivé vzorky (sample barcodes), které jsou sice také součástí adaptorů sekvenační knihovny, avšak pro zásadně odlišné účely. Tyto barkódy jsou obvykle začleňovány do knihovny přímo v průběhu PCR za použití modifikovaných primerů a jsou identické pro všechny molekuly daného vzorku. Na základě toho je pak možné provést smíchání (pooling) více vzorků do společné sekvenační reakce, a posléze výstupní informaci opět separovat pro každý vzorek zvlášť v procesu označovaném jako „demultiplexing“. Možnost sekvenovat větší počet vzorků paralelně v jednom běhu výrazně zvyšuje efektivitu a dostupnost současných NGS metod (Smith *et al.*, 2010; Illumina, 2019j).

4.13 Zpracování dat ze sekvenátoru MiSeq

Přístroj MiSeq od Illuminy se řadí k menším stolním sekvenátorům umožňující přizpůsobení mnoha parametrů analýzy. Výstupní kapacita dosahuje 15 Gb s maximálním počtem 25 milionů jednoduchých čtení (v případě párových koncových čtení 50 milionů). Délka čtení se pohybuje v rozmezí od 1×36 až po 2×300 párů bází a doba jednoho běhu může trvat 4 až 55 hodin (Illumina, 2019k).

Primární analýzu dat vykonává sekvenátor automaticky pomocí programu Real-Time Analysis (RTA) již v průběhu jednotlivých cyklů sekvenční reakce. Dochází ke zpracování obrazu, identifikaci bází (base calling) a stanovení příslušných hodnot kvality, jako je Phred skóre. Základním výstupem tohoto procesu jsou soubory BCL (base call) (Illumina, 2019k, Illumina, 2019j).

Sekundární analýza dat může být provedena buď opět přímo na sekvenátoru základním programem MiSeq Reporter Software (MSR), nebo za využití specifických bioinformatických nástrojů s přídatnými funkcemi. Mezi první kroky této fáze patří „demultiplexování“ (viz kapitola 4.12) a konverze dat do formátu FASTQ - textových souborů pro uložení sekvence a příslušných skóre kvality. FASTQ soubory jsou výchozím vstupem pro zarovnání (alignment) čtení podle referenčního genomu pomocí algoritmů, jako jsou BWA či Bowtie. V datech jsou následně vyhledány varianty (variant calling) typu SNV, inzercí/delece, případně dalších strukturní změn a získané informace jsou zapsány do souborů VCF (Variant Call Format). Illumina k tomto účelu vyvinula Somatic Variant Caller speciálně přizpůsobený k identifikaci variant při nízkých frekvencích (Illumina, 2019j).

V rámci poslední fáze zpracování sekvenačních dat, tj. terciální analýzy, jsou detekované varianty interpretovány, zejména s ohledem na jejich biologický význam a možnosti léčby daného onemocnění. K dispozici jsou například programy VariantStudio a VarAFT nebo databáze, jako je Cancer Genome Interpreter (Illumina, 2019l; Illumina, 2019m; VarAFT, 2019; Cancer Genome Interpreter, 2019).

5 MATERIÁL A METODY

5.1 Biologický materiál

5.1.1 Testované vzorky pomocí panelu NEBNext Direct Cancer HotSpot Panel

- Referenční standard HD802 EGFR Gene-Specific Multiplex Reference Standard gDNA: 18-0042-DG-2282
- Referenční standard HD709 EGFR Wild Type Reference Standard: 17-9064-DG-1840
- FFPE patientský vzorek nádoru slinivky: 060-3588(1)
- FFPE patientské vzorky chordomů: CH22042/07B, CH7612/09B, CH17069/09-2, CH16165/13B. Vzorky a kit k jejich zpracování poskytla CGB laboratoř a.s. Ostrava (vzorky byly odebrány ve Fakultní nemocnici Ostrava).

5.1.2 Testované vzorky pomocí panelu NEBNext Direct BRCA1/BRCA2

Všechny níže uvedené vzorky zanalyzovala Mgr. Jana Stránská, Ph.D. ve spolupráci s laborantkami DNA laboratoře ÚMTM.

- Referenční standard HD802 EGFR Gene-Specific Multiplex Reference Standard gDNA: 18-0042-DG-2282
- Patientské vzorky nádoru vaječníků: 18-0708-DG-3652 (č. 1), 8-0708-DG-3688 (č. 2), 18-0708-DG-3689 (č. 3), 18-0708-DG-3690 (č. 4), 18-0708-DG-3691 (č. 5), 18-0708-DG-3692 (č. 6), 18-0708-DG-3693 (č. 7), 18-0708-DG-3694 (č. 8), 18-0708-DG-3695 (č. 9), 18-0708-DG-3696 (č. 10)

5.2 Seznam chemikálií, roztoků a kitů

- DEPC voda (Sigma-Aldrich, kat. č. W4502)
- Ethanol absolutní (Penta, kat. č. 71250-11000)
- Hydroxid sodný 5 mol/l (Sigma-Aldrich, kat. č. S8263)
- MiSeq Reagent Kit v2 (300 cyklů) (Illumina, kat. č. MS-102-2002)
- MiSeq Reagent Kit v3 (150 cyklů) (Illumina, kat. č. MS-102-3001)
- MiSeq Reagent Nano Kit v2 (300 cyklů) (Illumina, kat. č. MS-103-1001)
- NEBNext Direct BRCA1/BRCA2 Panel (New England Biolabs, kat. č. E6627S)
- NEBNext Direct Cancer HotSpot Panel (New England Biolabs, kat. č. E7000S)
- QIAxcel DNA High Resolution Kit (1200) (QIAGEN, kat. č. 929002)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, kat. č. Q32851)
- Tris-EDTA pH = 8 (TE) (Sigma-Aldrich, kat. č. 93283)
- TruSight Tumor 170 Kit (24 vzorků) (Illumina, kat. č. OP-101-1004)

- TWEEN 20 (Sigma-Aldrich, kat. č. P9416)

5.3 Přístrojové a programové vybavení

- BioRuptor Plus (Diagenode)
- Magnetický stojánek (vyvinuto na UPOL/ÚMTM)
- Minicentrifuga MiniSpin (Eppendorf)
- Minicentrifuga/vortex Combi-Spin FVL-2400N (Biosan)
- MiSeq System (Illumina)
- QIAxcel Advanced System (QIAGEN)
- Qubit 2.0 Fluorometer (Invitrogen)
- Termoblok ThermoStat plus (Eppendorf)
- Termocykler PTC-200 (MJ Research)
- Vortex Mixer PV-1 (Grant-bio)
- Server Ubuntu 14.04 (výpočetní cluster ÚMTM, LF)

5.4 Analýza panelu 50 genů referenčních a patientských nádorových vzorků

5.4.1 Příprava zásobních roztoků EGFR referenčních vzorků

- EGFR-MUT3 (c = 1,02 ng/μl; frekvence *EGFR* mutací 3,1 %) byl připraven smísením:

0,45 μl	HD802 EGFR Gene-Specific Multiplex Reference Standard gDNA (c = 50 ng/μl)
1,35 μl	HD709 EGFR Wild Type Reference Standard (c = 50 ng/μl)
88,2 μl	1X TE
- EGFR-MUT12 (c = 1,88 ng/μl; frekvence *EGFR* mutací 12,5 %) byl připraven smísením:

1,5 μl	HD802 EGFR Gene-Specific Multiplex Reference Standard gDNA (c = 50 ng/μl)
78,5 μl	1X TE
- EGFR-WT (c = 1,88 ng/μl; frekvence *EGFR* mutací 0 %) byl připraven smísením:

1,5 μl	HD709 EGFR Wild Type Reference Standard (c = 50 ng/μl)
78,5 μl	1X TE

5.4.2 Kontrola kvality a koncentrace DNA patientských vzorků

5.4.2.1 Fluorometrické stanovení koncentrace DNA

1. Pro každé měření bylo připraveno 199 μ l reakční směsi smísením 198 μ l Qubit dsDNA HS Buffer a 1 μ l Qubit dsDNA HS Reagent (Qubit dsDNA HS Assay Kit).
2. K reakční směsi byl přidán 1 μ l DNA vzorku. Koncentrace DNA byla změřena na přístroji Qubit 2.0 Fluorometer.

5.4.2.2 Elektroforetická analýza velikostí DNA molekul

Míra fragmenatce DNA ve FFPE vzorcích byla stanovena kapilární elektroforézou QIAxcel Advanced a pomocí kitu QIAxcel DNA High Resolution Kit:

1. K analýze bylo od každého patientského vzorku odebráno 0,5 μ l DNA a rozpuštěno do 9,5 μ l Dilution Buffer.
2. Následná elektroforéza (program 0L500) byla provedena s použitím Size markerů a Aligment markerů o velikostech 100 bp – 2,5 kb a 50 bp – 5 kb.

5.4.3 Fragmentace DNA vzorků s průměrnou velikostí molekul nad 200 bp

Vzorky s fragmenty DNA >200 bp byly naštěpeny v sonikátoru BioRuptor Plus:

1. Z patientských vzorků bylo nutné štěpit pouze vzorek 060-3588(1). Z tohoto bylo 10 μ l rozpuštěno v 90 μ l 1X TE.
2. Referenční EGFR vzorky byly štěpeny v plném objemu připravených zásobních roztoků.
3. K sonikaci byl zvolen program LOW; 30/90 sekund (ON/OFF); 10 cyklů.
4. Velikost fragmentů byla ověřena pomocí přístroje QIAxcel Advanced stejným způsobem jako v kroku 5.4.2 za použití 3 μ l roztoku DNA a 7 μ l Dilution Buffer.

5.4.4 Příprava knihovny pomocí kitu NEBNext Direct Cancer HotSpot Panel

Knihovnu pro sekvenančí analýzu pomocí NEBNext Direct Cancer HotSpot Panel lze podle výrobce připravit z DNA v množství 10 ng – 1 μ g, přičemž nejspolehlivějších výsledků má být dosaženo nad hodnotou 100 ng. Množství DNA použité v případě jednotlivých vzorků včetně koncentrace výchozích roztoků uvádí Tabulka 9.

Tabulka 9: Koncentrace a vstupní množství DNA na přípravu sekvenanční knihovny

Název vzorku	Koncentrace DNA	Použité množství DNA na 1 knihovnu
EGFR-MUT3	1,00 ng/μl	42 ng
EGFR-MUT12	0,94 ng/μl	39,4 ng
EGFR-WT	0,94 ng/μl	39,4 ng
060-3588(1)	1,00 ng/μl	42 ng
CH22042/07B	6,78 ng/μl	100 ng
CH7612/09B	3,70 ng/μl	68,5 ng
CH17069/09-2	9,12 ng/μl	91,2 ng
CH16165/13B	4,92 ng/μl	91 ng

5.4.4.1 Pracovní postup NEBNext Direct Cancer HotSpot Panel

a) Fosforylace FFPE DNA

Tabulka 10: Příprava reakční směsi pro FFPE fosforylaci

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
FFPE Phosphorylation Buffer	5 μl	5,5 μl
FFPE Phosphorylation Enzyme	1 μl	1,1 μl
Celkem	6 μl	6,6 μl

1. K 6 μl připravené reakční směsi (Tabulka 10) přidat 42 μl fragmentované DNA.
2. Vzorčky inkubovat v termocykleru při 37 °C 15 min s vyhřátým víkem na 45 °C.

b) Hybridizace DNA s biotinylovanými sondami

Tabulka 11: Příprava hybridizační reakční směsi

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
Hybridization Buffer	47 μl	51,7 μl
Hybridization Additive	20 μl	22 μl
Cancer HotSpot Baits	5 μl	5,5 μl
Celkem	72 μl	79,2 μl

1. K připravené reakční směsi (Tabulka 11) přidat 48 μl fragmentované DNA z předchozí fosforylační reakce na finální objem 120 μl.
2. Spustit následující program s vyhřátým víkem na 105 °C (vzorčky umístit do termocykleru až teplota bloku dosáhne 95 °C):

10 min	95 °C
90 min	60 °C
∞	60 °C

c) Příprava streptavidinových kuliček

1. Na každou reakci je vyžadováno 75 μl kuliček.
2. Mikrozkumavku s požadovaným objemem kuliček umístit na magnet a po vyčerení roztoku odebrat supernatant.

3. Na každou reakci přidat 150 μl Hybridization Wash (HW).
4. Mikrozkušavku opět umístit na magnet a odebrat supernatant.
5. Kroky 5 a 6 opakovat dvakrát (celkově 3 promytí).
6. Kuličky resuspendovat v 30 μl /reakce Bead Prep Buffer.

d) Navázání DNA na kuličky přes biotinylované sondy

1. Ke každému vzorku přidat 30 μl resuspendovaných kuliček.
2. Reakce inkubovat v termobloku při 48 $^{\circ}\text{C}$ po dobu 10 min.
3. Vzorky umístit na magnet a odstranit supernatant.
4. Přidat 150 μl HW do každého vzorku a inkubovat v termobloku při 62 $^{\circ}\text{C}$ 5 min.
5. Umístit vzorky na magnet a odstranit supernatant.
6. Kroky 5 a 6 opakovat, tj. celkově dvě promyty při 62 $^{\circ}\text{C}$.
7. Přidat 150 μl Bead Wash Buffer 2 (BW2) do každého vzorku.

e) Zatupení 3' konců DNA

Tabulka 12: Příprava reakční směsi pro zatupení 3' konců DNA

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
3' Blunting Buffer	97 μl	106,7 μl
3' Blunting Enzyme Mix	3 μl	3,3 μl
Celkem	100 μl	110 μl

1. Kuličky s navázanou DNA umístit na magnet a odstranit supernatant.
2. Do každého vzorku přidat 100 μl připravené reakční směsi (Tabulka 12) a inkubovat při 37 $^{\circ}\text{C}$ po dobu 10 min v termobloku.
3. Post-reakční promytí
 - Vzorky umístit na magnet a odstranit supernatant.
 - Do každého vzorku přidat 150 μl Bead Wash Buffer 1 (BW1). Reakce šetrně promíchat, umístit na magnet a odstranit supernatant.
 - Do každého vzorku přidat 150 μl Bead Wash Buffer 2 (BW2). Reakce opět šetrně promíchat, umístit na magnet a odstranit supernatant.

f) Polyadenylace DNA

Tabulka 13: Příprava reakční směsi pro polyadenylace DNA

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
dA-Tailing Buffer	97 μl	106,7 μl
dA-Tailing Enzyme	3 μl	3,3 μl
Celkem	100 μl	110 μl

1. Kuličky s navázanou DNA umístit na magnet a odstranit supernatant.

2. Do každého vzorku přidat 100 μl připravené reakční směsi (Tabulka 13) a inkubovat při 37 °C po dobu 10 min v termobloku.
3. Provést post-reakční promytí stejným způsobem jako v sekci e) zatupení 3' konců DNA.

g) Ligace 3' adaptoru

Tabulka 14: Příprava reakční směsi pro ligaci 3' adaptoru

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
Adaptor Ligation Buffer	80 μl	88 μl
3' Adaptor	10 μl	11 μl
Ligáza	10 μl	11 μl
Celkem	100 μl	110 μl

1. Kuličky s navázanou DNA umístit na magnet a odstranit supernatant.
2. Do každého vzorku přidat 100 μl připravené reakční směsi (Tabulka 14) a inkubovat při 20 °C po dobu 15 min v termobloku.
3. Provést post-ligační promytí. Postup je identický jako v případě post-reakčního promytí, s tím rozdílem, že jsou vzorky v BW1 pufru promyty dvakrát po sobě.

f) Zatupení 5' konců DNA

Tabulka 15: Příprava reakční směsi pro zatupení 5' konců DNA

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
5' Blunting Buffer	97 μl	106,7 μl
5' Blunting Enzyme Mix	3 μl	3,3 μl
Celkem	100 μl	110 μl

1. Kuličky s navázanou DNA umístit na magnet a odstranit supernatant.
2. Do každého vzorku přidat 100 μl připravené reakční směsi (Tabulka 15) a inkubovat při 20 °C po dobu 10 min v termobloku.
3. Provést post-reakční promytí.

g) Ligace 5' adaptoru

Tabulka 16: Příprava reakční směsi pro ligaci 5' adaptoru

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
Adaptor Ligation Buffer	80 μl	88 μl
5' UMI Adaptor	10 μl	11 μl
Ligáza	10 μl	11 μl
Celkem	100 μl	110 μl

1. Kuličky s navázanou DNA umístit na magnet a odstranit supernatant.
2. Do každého vzorku přidat 100 μl připravené reakční směsi (Tabulka 16) a inkubovat při 20 °C po dobu 20 min v termobloku.
3. Provést post-ligační promytí.

h) Štěpení adaptoru

Tabulka 17: Příprava reakční směsi pro štěpení adaptoru

Reagencie	Objem pro 1 reakci	Objem s 10% rezervou
Cleaving Buffer	95 μ l	104,5 μ l
Cleaving Enzyme Mix	5 μ l	5,5 μ l
Celkem	100 μ l	110 μ l

1. Kuličky s navázanou DNA umístit na magnet a odstranit supernatant.
2. Do každého vzorku přidat 100 μ l připravené reakční směsi (Tabulka 17) a inkubovat při 37 °C po dobu 15 min v termobloku.
3. Provést post-reakční promytí

i) Amplifikace sekvenační knihovny

1. Vzorky umístit na magnet a odstranit supernatant.
2. Do každé reakce přidat 45 μ l DEPC vody a připravit PCR reakční směs podle Tabulky 18.

Tabulka 18: Složení reakční směsi pro PCR

Reagencie	Objem pro 1 reakci
Q5 Master Mix	50 μ l
Index Primer Mix	5 μ l
Resuspendované kuličky	45 μ l
Celkem	100 μ l

3. Spustit program podle Tabulky 19 s vyhřátým víkem na 105 °C (vzorky umístit do termocykleru až teplota dosáhne 98 °C).

Tabulka 19: Podmínky PCR pro amplifikaci DNA knihovny

Fáze cyklu	Teplota	Čas	Počet cyklů
Počáteční denaturace	98 °C	30 s	1
Denaturace	98 °C	10 s	25
Hybridizace	62 °C	15 s	
Elongace	72 °C	20 s	
Konečná elongace	72 °C	5 min	1

j) Přečištění amplifikovaných fragmentů

2. Přidat 85 μ l Sample Purification Beads do každé PCR reakce.
3. Otevřené vzorky inkubovat 10 min při pokojové teplotě.
4. Umístit vzorky na magnet a po ~2 min odebrat supernatant.
5. Do každé reakce přidat 200 μ l čerstvě připraveného 80% ethanolu, zatímco jsou mikrozkušavky na magnetu. Inkubovat při pokojové teplotě 30 sekund, a poté odstranit supernatant.
6. Zopakovat krok 4.

7. Inkubovat otevřené vzorky při 37 °C po dobu 5 min na termobloku, aby došlo k vysušení kuliček.
8. Vzorky resuspendovat ve 102 µl vody a inkubovat 2 min při pokojové teplotě.
9. Vzorky umístit na magnet a po ~2 min převést 100 µl eluované knihovny do nové mikrozkušavky.
10. Do reakcí přidat 85 µl Sample Purification Beads a inkubovat 10 min při pokojové teplotě.
11. Vzorky umístit na magnet a po ~2 min odebrat supernatant.
12. Do mikrozkušavek umístěných na magnetu přidat 200 µl čerstvě připraveného 80% ethanolu. Inkubovat při pokojové teplotě 30 sekund, a poté odstranit supernatant.
13. Zopakovat krok 11.
14. Vzorky inkubovat otevřené při 37 °C po dobu 2 min na termobloku.
15. Vysušené kuličky resuspendovat ve 30 µl 1X TE a inkubovat 2 min při pokojové teplotě.
16. Vzorky umístit na magnet a po ~2 min přemístit 28 µl eluované knihovny do nové mikrozkušavky.

5.4.5 Analýza kvality eluované sekvenační knihovny

5.4.5.1 Stanovení koncentrace DNA knihovny

1. Pro každé měření bylo připraveno 195 µl reakční směsi smísením 194 µl Qubit dsDNA HS Buffer a 1 µl Qubit dsDNA HS Reagent (Qubit dsDNA HS Assay Kit).
2. K reakční směsi bylo přidáno 5 µl DNA vzorku. Koncentrace DNA byla změřena na přístroji Qubit 2.0 Fluorometer.

5.4.5.2 Kontrola velikosti fragmentů DNA knihovny pomocí přístroje QIAxcel Advanced a reagensů kitu QIAxcel DNA High Resolution Kit

1. K analýze bylo odebráno 5 µl DNA knihovny referenčního vzorku (EGFR-MUT3) a smícháno s 5 µl Dilution Buffer.
2. Z patientských vzorků chordomů byl z DNA knihovny o nejvyšší koncentraci (CH17069/09-2) odebrán 1 µl a smíchán s 9 µl Dilution Buffer.
3. Vzorky byly elektroforeticky separovány (program 0L500) za použití Size markerů s velikostí 50-800 bp a Alignment markerů 15 bp – 1 kb.

5.4.6 Pracovní postup sekvenování

Vzorky pro sekvenování připravil Ing. Rastislav Slavkovský, PhD. za asistence laborantek.

1. Před samotným sekvenováním byla rozmrazena kazeta a sekvenátor promyt pomocí programu Post-RUN WASH anebo Maintenance WASH (za použití 0,5 % roztoku Tween 20).
2. Z naměřených koncentrací DNA knihovny (viz kapitola 5.4.5.1) a průměrných velikostí DNA molekul po indexaci byla vypočítána molární koncentrace jednotlivých vzorků dle vzorce:

$$c [\text{nmol} \cdot \mu\text{l}^{-1}] = \frac{\rho_i [\text{ng} \cdot \mu\text{l}^{-1}] \cdot 10^6}{660 \text{ g} \cdot \text{mol}^{-1} \cdot 260 \text{ bp}}$$

kde ρ_i je hmotnostní koncentrace DNA knihovny; $660 \text{ g} \cdot \text{mol}^{-1}$ je průměrná molární hmotnost jedné báze a 260 bp je orientační průměrná velikost molekuly DNA po indexaci

3. Vzorky byly naředěny DEPC vodou na $2 \text{ nmol}/\mu\text{l}$.
4. Byl připraven roztok $0,2 \text{ mol/l}$ NaOH smísením $480 \mu\text{l}$ H_2O s $20 \mu\text{l}$ 5 mol/l NaOH.
5. Bylo smícháno $20 \mu\text{l}$ naředěného vzorku s $20 \mu\text{l}$ čerstvě připraveného $0,2 \text{ mol/l}$ NaOH (1:1) a směs byla inkubována 5 min při pokojové teplotě.
6. K $10 \mu\text{l}$ denaturované DNA bylo přidáno $990 \mu\text{l}$ vychlazeného HT1 z lednice (1:100).
7. Do sekvenační kazety bylo pipetováno po $600 \mu\text{l}$ vzorků.
8. Byl připraven Sample Sheet a spuštěn sekvenační program (Illumina MiSeq software).
9. Po ukončení sekvenačního programu byl přístroj promyt 0,5 % roztokem Tween 20 pomocí programu Post-RUN WASH za použití promývací kazety a promývací láhve.
10. Data byla na sekvenátoru umístěna ve složce:
D:\Illumina\MiSeqOutput\datum+identifikace běhu\Data\ Intensities\BaseCalls
11. K následné analýze byly použity soubory formátu fastq.gz, pro každý vzorek celkem 3 soubory (read1, read2 a index2) a soubory VCF.
12. Data byla poté zpracovávána na serveru Ubuntu 14.04. za použití dalších bioinformatických nástrojů (viz kapitola 6.3)

6 VÝSLEDKY

Jedním z cílů experimentální části diplomové práce bylo otestovat panel NEBNext Direct Cancer HotSpot Panel (dále uváděno jako „NEB Cancer panel“) s 50 geny pro jeho potenciální diagnosticko-terapeutické využití u pacientů s karcinomem plic. Předpokládanou výhodou analýzy pomocí tohoto panelu je možnost začlenění UMI barkódů do sekvenačních dat, které mají snižovat počet přítomných artefaktů, a zajišťovat tak přesnější zachycení variant s nízkou frekvencí.

Aby však mohla být informace z molekulárních barkódů v analýze uplatněna, je potřeba data generovaná sekvenátorem dále zpracovat specifickými nástroji. Právě pro tyto účely byla sestavena bioinformatická analýza založená na otevřených (open-source) programech, která za využití UMI barkódů vytváří z původních FASTQ souborů výstupní soubory VCF. Přínos tohoto přístupu byl zhodnocen porovnáním výsledných dat s výstupy z programu MiSeq Reporter Software (MSR), který s informací UMI barkódů nepracuje. Nově sestavená analýza byla posléze ověřena i pro další panel NEBNext Direct BRCA1/BRCA2.

Detekční limit stejně jako požadovaný vstupní materiál jsou důležitými kritérii, které zásadně ovlivňují možnosti aplikace dané metody. Protože v současném molekulárním profilování nádorů plic stále roste význam metody tekuté biopsie, byla snaha NEB Cancer panel ověřit pro schopnost detekce mutací při nízkém vstupu DNA pod 100 ng. Ačkoli panel vykazoval obecně vysokou citlivost, neumožnil za daných podmínek detekci delece genu *EGFR*, důležitého biomarkeru v terapii plicních karcinomů. Analýza byla z těchto důvodů přeměřována k alternativnímu využití panelu pro jiné typy nádorů, zejména chordomy.

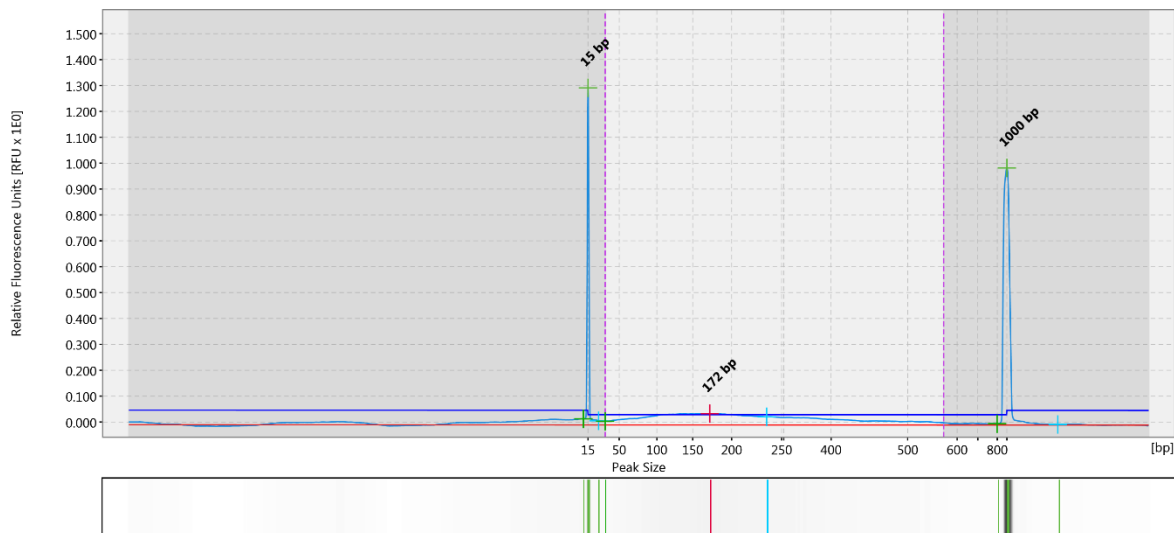
Další diskutovanou otázkou v této práci představoval rozsah metody a jaké výhody analýza větších genomických regionů přináší. Výstupy NEB Cancer panelu byly proto experimentálně srovnány s větším panelem TruSight Tumor 170 (Illumina) pokrývající celé exony 170 genů v testu patientského vzorku nádoru slinivky. Detekované mutace z obou panelů byly zhodnoceny s ohledem na možnosti cílené terapie.

6.1 Příprava DNA knihovny a primární výsledky sekvenování

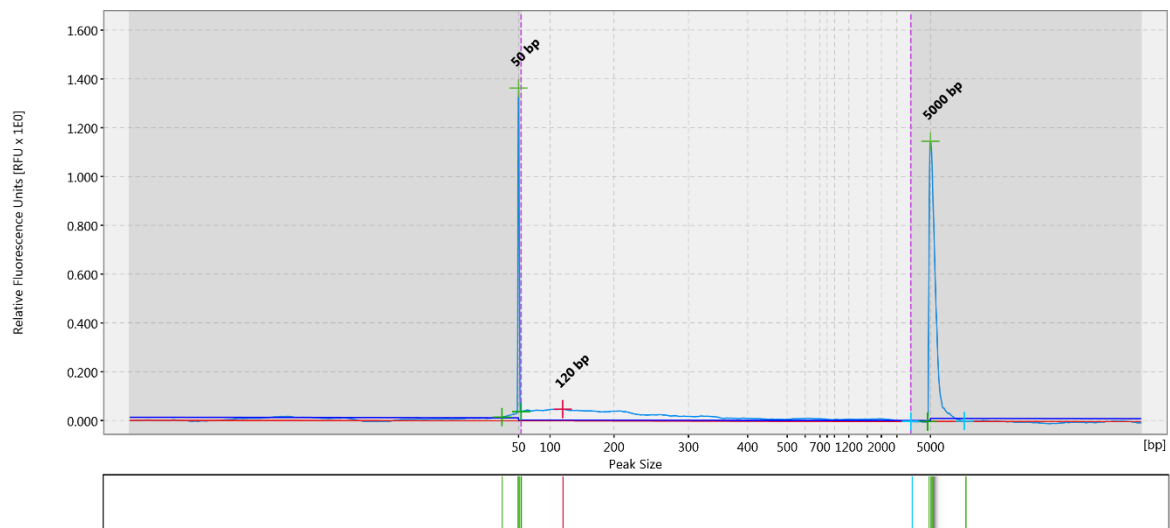
Pro sekvenační analýzu za použití NEB Cancer panelu bylo zpracováno celkem 8 vzorků; 3 referenční a 5 patientských. Mezi patientskými vzorky byly 4 chordomy a 1 nádor slinivky.

Příprava knihovny pomocí tohoto panelu vyžaduje vstupní DNA s délkou fragmentů přibližně 200 bp. Genomická DNA referenčních vzorků a DNA patientského vzorku nádoru

slinivky byla proto před přípravou knihovny fragmentována sonikací a zkontrolována elektroforézou na přístroji QIAxcel Advanced (Obrázek 8). Jak zobrazuje Obrázek 9, DNA fragmenty ve FFPE vzorcích chordomů odpovídaly požadovaným velikostem i bez štěpení v sonikátoru. Koncentrace DNA knihovny po eluci pro většinu vzorků dosahovala poměrně nízkých hodnot, pravděpodobně z důvodů malého množství vstupního materiálu (<100 ng) a/nebo vysokého stupně poškození DNA (Tabulka 20, Tabulka 21). Velikosti DNA fragmentů eluované knihovny na příkladu jednoho ze vzorků chordomů znázorňuje Obrázek 10. V porovnání s Obrázkem 9 lze pozorovat posun velikostí molekul do rozmezí 150-400 bp s maximem v oblasti 260 bp, které odpovídá selektovaným fragmentům s ligovanými sekvenačními adaptory na 3' a 5' koncích. Slabé intenzity signálu ze všech uvedených elektroforetických separací jsou způsobené nízkou koncentrací DNA, která byla použita z důvodů zachování vzorků pro další kroky analýzy.



Obrázek 8: Velikosti fragmentů DNA referenčního vzorku po sonikaci stanovené elektroforézou na přístroji QIAxcel Advanced (před přípravou sekvenační knihovny)



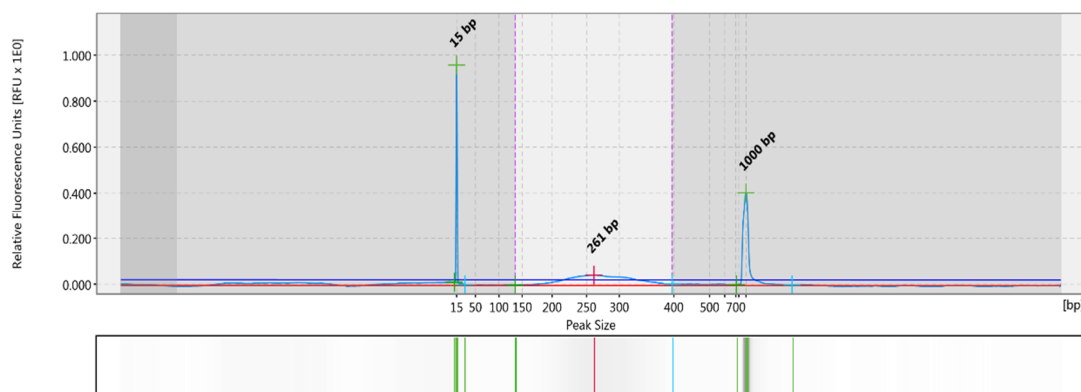
Obrázek 9: Elektroforetické stanovení velikosti fragmentů DNA vzorku chordomu (CH17069/09-2) na přístroji QIAxcel Advanced před přípravou sekvenační knihovny

Tabulka 20: Koncentrace DNA ve vzorcích a vstupní množství DNA pro přípravu knihovny

Název vzorku	Koncentrace DNA [ng/μl]	Použité množství DNA na 1 knihovnu [ng]
EGFR-MUT3	1,00	42
EGFR-MUT12	0,94	39,4
EGFR-WT	0,94	39,4
060-3588(1)	1,00	42
CH22042/07B	6,78	100
CH7612/09B	3,70	68,5
CH17069/09-2	9,12	91,2
CH16165/13B	4,92	91

Tabulka 21: Koncentrace a molarita eluované knihovny po přečištění

Název vzorku	Koncentrace DNA knihovny [ng/μl]	Molarita DNA [nmol/μl]
EGFR-REF 1	2,2	12,8
EGFR-REF 2	0,93	5,4
EGFR-REF-WT	0,68	3,9
060-3588(1)	3,48	20,3
CH22042/07B	0,12	0,7
CH7612/09B	1,03	6,0
CH17069/09-2	4,41	25,7
CH16165/13B	0,31	1,8



Obrázek 10: Elektroforetické stanovení velikosti fragmentů DNA vzorku chordomu (CH17069/09-2) na přístroji QIAxcel Advanced po eluci sekvenační knihovny

6.2 Parametry kvality sekvenování

Pomocí programu bedtools coverage byly pro každý osekvenovaný vzorek zjištěny hodnoty dosaženého průměrného sekvenačního pokrytí a množství jednotlivých bází splňující kryterium nejnižšího pokrytí 200x, 500x a 1000x (Tabulka 22). Použitý soubor BED pro vypočtení výsledků celkově zahrnoval 55712 bází (BED soubor poskytuje výrobce NEB Cancer panelu). Sekvenační výstupy zpracoval Ing. Rastislav Slavkovský, Ph.D.

Tabulka 22: Průměrné pokrytí a bázové pokrytí osekvenovaných referenčních a patientských vzorků.

Název vzorku	Průměrné pokrytí	Počet bází > 200x	% bází > 200x	Počet bází > 500x	% bází > 500x	Počet bází > 1000x	% bází > 1000x
EGFR-MUT3	5593	55420	99	54473	97	51932	93
EGFR-MUT12	4827	47894	85	46851	84	45059	80
EGFR-WT	3503	49886	89	47898	85	46044	82
060-3588(1)	3314	54938	98	52817	94	48902	87
CH22042/07B	1471	50685	90	44767	80	33291	59
CH7612/09B	5790	54503	97	52799	94	50388	90
CH17069/09-2	8525	55289	99	54453	97	52883	94
CH16165/13B	4764	53742	96	51282	92	48185	86

Pozn.: Celkový počet bází v souboru BED = 55712

6.3 Sestavení bioinformatické analýzy s UMI barkódy pro data z panelů NEB

Ke zpracování sekvenačních dat získaných pomocí panelů NEB byl vytvořen vlastní shellový skript využívající otevřené (open-source) programy. Série příkazů byla seskládána na základě již existujícího softwaru napsaném ve skriptovacím jazyce Scala (GitHub, 2019a). Oproti generování VCF souborů přes program MiSeq reporter přímo na sekvenátoru jsou tímto přístupem začleněny do analýzy UMI molekulární barkódy.

V analýze jsou použity následující programy:

1. BWA – provádí alignment sekvenačních čtení na referenční genom (Burrows-Wheeler Aligner, 2010).
2. Picard – využívá se zejména pro konverze formátů souborů (GitHub, 2019b)
3. SAMtools – komprimuje soubory SAM (Sequence Alignment Map) na BAM (Binary Alignment Map) a indexuje referenční genom (SAMtools, 2012)
4. fgbio – začleňuje UMI barkódy do sekvenačních dat, generuje konsenzuální čtení a filtruje nespolehlivé sekvence (GitHub, 2019c)
5. VarDictJava – provádí vysoce citlivé vyhledání variant (variant calling) (GitHub, 2019d)

Ke spuštění skriptu jsou kromě těchto programů vyžadovány:

1. Linux nebo Mac operační systém
2. Java verze 1.8
3. Nainstalované programovací jazyky R a Perl
4. Referenční sekvence lidského genomu včetně přidružených souborů (viz Příprava reference)
5. Soubor BED cílových oblastí
6. Tři FASTQ soubory pro každý vzorek (čtení 1, čtení 2, index 2 = UMI) pojmenované podle vzoru SampleName_S1_L001_R1_001.fastq.gz

Příprava reference zahrnuje příkazy:

1. `bwa index -a bwtsv reference.fa`
#Indexování lidského genomu. Vytvoří se kolekce souborů, které BWA později používá k alignmentu.
2. `samtools faidx reference.fa`
#Vytvoření indexu FASTA souboru za vzniku reference.fa.fai. Pro každý z kontigů v referenčním souboru FASTA tímto vzniká záznam, který se skládá z názvu kontigu, velikosti, umístění, "basePerLine" a "bytesPerLine".
3. `java -jar picard.jar CreateSequenceDictionary REFERENCE=reference.fa OUTPUT=reference.dict`
#Generuje soubor s názvem reference.dict formátovaný jako hlavička SAM, popisující obsah referenčního souboru FASTA.

Kompletní obsah skriptu:

```
#Skript musí být spuštěn ze složky se všemi požadovanými FASTQ soubory.
```

```
FILES=.*R1_001.fastq.gz
```

```
for f in $FILES
```

```
do
```

```
#Jedná se o jednoduchý cyklus, který pro jednotlivé soubory v proměnné FILES provede následující příkazy v celém skriptu.
```

```

soubor=${f::-15}
sample_name=${f::-21}

soubor1=${soubor}R1_001.fastq.gz"
soubor2=${soubor}R2_001.fastq.gz"
soubor3=${soubor}I2_001.fastq.gz"
#Vytváří proměnné soubor1, soubor2, soubor3 obsahující příslušné FASTQ soubory, které jsou vstupem pro
následující programy.

echo "Processing files "$soubor1" and "$soubor2" and using indexes in "$soubor3
#Vypíše na obrazovku kontrolní hlášku o spuštění požadovaného procesu.

soubor_out=${soubor1%.*} #bez gz
soubor_out=${soubor_out%.*} #bez fastq.gz
soubor_out=${soubor_out::-12} #bez _L001_R1_001
soubor_out=${soubor_out}_1.bam #s _1.bam
#Příkazy upravují název výstupního souboru z programu FastqToSam. V tomto případě vzniká soubor
SampleName_1.bam. Obdobná úprava výstupních souborů je použita pro každý program v celém skriptu.

java -jar '/path/to/picard.jar' FastqToSam F1=${soubor1} F2=${soubor2} O=${soubor_out} SM=${sample_name}
#Generuje nemapovaný BAM ze vstupních souborů FASTQ 1 a 2.

soubor_in=${soubor_out}
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-2}
soubor_out_unmap=${soubor_out}_2.bam
java -Xmx24g -jar '/path/to/fgbio-0.6.1.jar' AnnotateBamWithUmis -i $soubor_in -f $soubor3 -o
$soubor_out_unmap
#V nemapovaném BAM souboru jsou k jednotlivým čtením přiřazeny UMI barkódy.

soubor_in1=${soubor1}
soubor_in2=${soubor2}
soubor_out=${soubor1%.*}
soubor_out=${soubor_out%.*}
soubor_out=${soubor_out::-12}
soubor_out=${soubor_out}_3.sam
bwa mem -t 12 '/path/to/hg19/genome.fa' $soubor_in1 $soubor_in2 > $soubor_out
# Proveďte alignment sekvenačních čtení na referenční genom za vzniku souboru SAM. Soubor nezahrnuje
informace o UMI.

soubor_in=${soubor_out}
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-2}
soubor_out=${soubor_out}_4.bam
samtools view -bS -@ 14 $soubor_in > $soubor_out
#Zkomprimuje SAM na mapovaný BAM soubor.

soubor_in1=${soubor_out}
soubor_in2=${soubor_out_unmap}
soubor_out=${soubor_in1%.*}
soubor_out=${soubor_out::-2}
soubor_out=${soubor_out}_5.bam
java -Xmx24g -jar '/path/to/picard.jar' MergeBamAlignment ALIGNED=${soubor_in1}
UNMAPPED=${soubor_in2} O=${soubor_out} R=/path/to/hg19/genome.fa
#Do mapovaného BAM jsou doplněny UMI za použití již dříve anotovaného nemapovaného souboru BAM.

soubor_in=${soubor_out}
soubor_out_MD=${soubor_in%.*}
soubor_out_MD2=${soubor_out_MD::-2}
soubor_out_MD3=${soubor_out_MD2}_6.bam

```

```

metrics=${soubor_out_MD2}_MD.txt
java -Xmx24g -jar '/path/to/picard.jar' MarkDuplicates I=$soubor_in O=$soubor_out_MD3
METRICS_FILE=$metrics BARCODE_TAG=RX
#V souboru BAM jsou na základě UMI vyhledány a označeny duplicitní čtení. Duplikáty jsou definovány jako
čtení pocházející z jediného fragmentu DNA.

```

```

soubor_in=$soubor_out
soubor_out=${soubor_in%.*}
soubor_out2=${soubor_out::-2}
soubor_out3=${soubor_out2}_7.bam
famSizes=${soubor_out2}_FS.txt
java -Xmx24g -jar '/path/to/fgbio-0.6.1.jar' GroupReadsByUmi -i $soubor_in -o $soubor_out3 --raw-tag=RX --
family-size-histogram=$famSizes --min-map-q=10 --strategy=Adjacency --edits=1
#Seskupuje dohromady čtení, která pocházejí ze stejné původní molekuly.

```

```

soubor_in=$soubor_out3
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-2}
soubor_out_unmap=${soubor_out}_8.bam
java -Xmx24g -jar '/path/to/fgbio-0.6.1.jar' CallMolecularConsensusReads -i $soubor_in -o
$soubor_out_unmap --min-reads=1 --m 30 --2 35
#Čtení se stejným UMI barkódem jsou analyzována báze po bázi a podle toho je vyhodnocována
pravděpodobnost každé báze ve zdrojové molekule. Vytváří se konsenzuální sekvence pro všechna čtení se
stejným UMI barkódem. Na základě pravděpodobnostního modelu současně dochází k modifikaci původní
hodnoty bazové kvality (bazová kvalita představuje pravděpodobnost chyby sekvenování).

```

```

soubor_in=$soubor_out_unmap
soubor_in=$soubor_out_unmap
soubor_out1=${soubor_in%.*}
soubor_out1=${soubor_out1::-2}
soubor_out1=${soubor_out1}_9_R1.fastq
soubor_out2=${soubor_in%.*}
soubor_out2=${soubor_out2::-2}
soubor_out2=${soubor_out2}_9_R2.fastq
java -jar '/path/to/picard.jar' SamToFastq I=$soubor_in F1=$soubor_out1 F2=$soubor_out2
#Soubor BAM je převeden opět na FASTQ 1 a 2, které se však liší od výchozích souborů. Formát FASTQ
vyžaduje následující program bwa mem, dochází však ke ztrátě informací o UMI.

```

```

soubor_in1=$soubor_out1
soubor_in2=$soubor_out2
soubor_out=${soubor_in1%.*}
soubor_out=${soubor_out::-5}
soubor_out=${soubor_out}_10.sam
bwa mem -t 12 '/path/to/hg19/genome.fa' $soubor_in1 $soubor_in2 > $soubor_out
# Provede alignment sekvenačních čtení na referenční genom za vzniku souboru SAM.

```

```

soubor_in=$soubor_out
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-3}
soubor_out=${soubor_out}_11.bam
samtools view -bS -@ 14 $soubor_in > $soubor_out
#Zkomprimuje SAM na mapovaný BAM soubor.

```

```

soubor_in1=$soubor_out
soubor_in2=$soubor_out_unmap
soubor_out=${soubor_in1%.*}
soubor_out=${soubor_out::-3}
soubor_out=${soubor_out}_12.bam
java -Xmx24g -jar '/path/to/picard.jar' MergeBamAlignment ALIGNED=$soubor_in1
UNMAPPED=$soubor_in2 O=$soubor_out R=/path/to/hg19/genome.fa

```

#Do mapovaného BAM jsou doplněny UMI za použití již dříve anotovaného nemapovaného souboru BAM.

```
soubor_in=$soubor_out
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-3}
soubor_out=${soubor_out}_13.bam
java -Xmx24g -jar '/path/to/fgbio-0.6.1.jar' FilterConsensusReads -i $soubor_in -o $soubor_out -r
/path/to/hg19/genome.fa --min-reads=1 --max-base-error-rate=0.1 --max-read-error-rate=0.05 --min-base-
quality=40 --max-no-call-fraction=0.1
#Podle zvolených parametrů, jako je minimální bázová kvalita 40, jsou ze souboru BAM odstraněna, případně
skryta nespolehlivá čtení.
```

```
soubor_in=$soubor_out
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-3}
soubor_out=${soubor_out}_14.bam
java -Xmx24g -jar '/path/to/fgbio-0.6.1.jar' ClipBam -i $soubor_in -o $soubor_out -r /path/to/hg19/genome.fa -
-clipping-mode=Hard --clip-overlapping-reads=true
#Eliminuje překryvy mezi čteními pocházejícími z jednoho templátu. Díky tomu nedochází k nadhodnocení
vypovídající hodnoty o výskytu varianty v místech překryvu těchto čtení.
```

```
soubor_in=$soubor_out
soubor_out=${soubor_in%.*}
soubor_out=${soubor_out::-3}
soubor_out=${soubor_out}_p0.vcf
'/path/to/VarDict-1.5.8/bin/vardict' -G /path/to/hg19/genome.fa -f 0.01 -N $sample_name -b $soubor_in -z -c 1
-S 2 -E 3 -g 4 -P 5 /path/to/E7000baitedregionsGRCh37hg191.bed | '/path/to/VarDict-
1.5.8/bin/teststrandbias.R' | '/path/to/VarDict-1.5.8/bin/var2vcf_valid.pl' -N $sample_name -E -f 0.01 >
$soubor_out
#Z BAM souborů jsou vyhledány SNV a jednoduché inserce/delece s frekvencí nad 1 %. Současně dochází k
odfiltrování variant s neodpovídající kvalitou (viz parametry dostupné na stránkách GitHub/VarDictJava). K
variantám jsou přiřazeny značky (flags), popisující např. „strand bias“ (Bias), průměrou pozici varianty ve
četních apod. Značky mohou být v následné analýze použity k dalšímu odfiltrování nespolehlivých variant.
Výstupem programu je textový soubor ve formátu VCF (Variant Call Format).
```

```
done
#Skript pro každý vzorek generuje celkem 15 souborů (BAM, SAM nebo FASTQ) očíslovaných na základě
pořadí jednotlivých programů, ve kterých byly vytvořeny + 2 textové soubory metrik + výsledný VCF.
```

6.4 Ověření analýzy s UMI barkódy a porovnání výstupních dat z programu MSR

6.4.1 Stanovení sensitivity a specificity analýzy vzorku HD802 pro genotypizaci panelu 50 genů

Výstupy sestavené bioinformatické analýzy s UMI barkódy byly porovnány s daty z programu MiSeq Reporter Software (MSR) využívající Somatic Variant Caller s detekčním limitem alelické frekvence 1 %. K analýze byly použity FASTQ soubory referenčních vzorků EGFR-MUT12 (HD802 EGFR Gene-Specific Multiplex Reference Standard gDNA) a EGFR-WT (HD709 EGFR Wild Type Reference Standard) získané panelovým sekvenováním na platformě MiSeq s délkou čtení 75 bp (párová koncová čtení). DNA knihovna byla připravena pomocí kitu NEBNext Direct Cancer HotSpot Panel. Výsledky z obou bioinformatických analýz byly vyhodnoceny podle exomových dat

referenčního standardu HD802 od výrobce Horizon Discovery (HD) a daty pro buněčnou linii RKO, které poskytuje AccuRef Diagnostics (Horizon Discovery, 2019; AccuRef Diagnostics, 2019).

S cílem nalézt společné varianty ve vytvořených VCF souborech vzorku EGFR-MUT12 s exomovými daty HD byly využity příkazy BEDTools (BEDTools, 2019):

1. `bedtools intersect -a EGFR_Multiplex_DNA.bed -b NEBbaitedregions.bed > NEB_intersect_standard.bed`
#Vymezení genomických regionů standardu pouze na oblasti, které zahrnuje NEB Cancer panel (Soubor EGFR_Multiplex_DNA.bed musel být ručně vytvořen z formátu .xlsx dostupném na webových stránkách HD).
2. `bedtools intersect -a NEB_intersect_standard.bed -b EGFR-MUT12.vcf -wb > in_both.txt`
#Vyhledání společných variant jednotlivých VCF souborů a HD dat standardu omezených na požadované genomické regiony.
3. `bedtools intersect -a NEB_intersect_standard.bed -b EGFR-MUT12.vcf -wb -v >not_in_vcf.txt`
#Nalezení variant, které nebyly a měly být přítomny ve VCF souborech podle HD dat standardu.

VCF soubor vzorku EGFR-MUT12 vytvořený nově sestavenou analýzou s UMI barkódy obsahoval celkem 131 variant, VCF soubor z MSR celkem 359 variant. V exomových datech HD bylo zahrnuto celkem 22 variant, které by měly být detekované NEB testem. Z obou těchto souborů bylo 20 z 22 (~90 %) variant nalezeno také v datech HD (Tabulka 23). Naopak dvě z 22 variant nebyly detekovány, ačkoli měly být podle dat výrobce v referenčním vzorku přítomny; jedna z nich představovala validovanou delecii v genu *EGFR* (c.2235_2249delGGAATTAAGAGAAGC), druhá se nacházela v genu *FGFR3* (c.797T>G).

Tabulka 23: Varianty potvrzené v exomových datech Horizon Discovery, které byly detekovány ve vzorku EGFR-MUT12 dvěma různými bioinformatickými metodami: programem MiSeq Reporter Software (MSR) a sestavenou analýzou s UMI barkódy.

Gen	Zápis varianty na úrovni cDNA	EGFR-MUT12-REF-MSR			EGFR-MUT12-REF-UMI		
		Frekvence [%]	Pokrytí	Filtr	Frekvence [%]	Pokrytí	Filtr
<i>EGFR</i>	c.2582T>A	12,8	12010	PASS	13,3	105	p8
<i>EGFR</i>	c.2573T>G	11,2	12577	PASS	14,3	105	PASS
<i>EGFR</i>	c.2369C>T	7,1	12577	PASS	2,3	302	PASS
<i>PIK3CA</i>	c.1173A>G	41,2	6624	PASS	36,1	72	PASS
<i>PIK3CA</i>	c.3140A>G	52,0	4891	PASS	43,4	53	PASS
<i>FGFR3</i>	c.882T>C(p.=)	41,7	4203	PASS	25,7	74	PASS
<i>FGFR3</i>	c.1959G>A(p.=)	100	1534	PASS	100	17	PASS
<i>PDGFRA</i>	c.1701A>G(p.=)	99,0	6136	PASS	97,4	77	PASS
<i>PDGFRA</i>	c.2472C>T(p.=)	99,7	7124	PASS	97,4	76	PASS
<i>KIT</i>	c.252G>T(p.=)	46,9	12399	PASS	40,0	135	PASS
<i>KIT</i>	c.2586G>C(p.=)	31,2	5522	PASS	34,2	85	PASS
<i>KDR</i>	c.889G>A	37,1	4936	PASS	27,6	58	PASS
<i>SMO</i>	c.1164G>C(p.=)	100	3666	PASS	100	26	PASS
<i>BRAF</i>	c.1799T>A	75,6	12811	PASS	67,8	152	PASS
<i>NOTCH1</i>	c.6555C>T(p.=)	67,6	7551	PASS	53,3	75	PASS
<i>NOTCH1</i>	c.5094C>T(p.=)	74,4	2390	PASS	64,7	17	PASS
<i>RET</i>	c.2307G>T(p.=)	71,3	19065	PASS	63,8	160	p8
<i>HRAS</i>	c.81T>C(p.=)	22,0	2536	SB	20,8	24	PASS
<i>FLT3</i>	c.1310-3T>C	47,8	5853	PASS	43,5	69	PASS
<i>STK11</i>	c.816C>T(p.=)	38,6	4964	PASS	48,8	41	PASS

Pozn.: Označení (p.=) v zápisu na úrovni cDNA představuje synonymní variantu

Z výsledků je patrné, že obě bioinformatické metody zachytily velké množství variant, které jsou pravděpodobně pouze artefakty. Data byla proto dále zpracována pomocí programu VariantStudio, kde byly z VCF souboru s UMI barkódy vyřazeny varianty s frekvencí pod 10 % a se statusem pSTD nebo Bias (= strand bias). Celkový počet variant se tímto snížil na 36, kdy z tohoto počtu 19 variant představovalo varianty potvrzené v HD datech a 17 variant zůstalo neobjasněných. Všechny z těchto 17 variant se nacházely také ve VCF souboru z MSR, ačkoli v některých případech pod limitní frekvencí 10 % (v Tabulce 24 vyznačeny tučně). Téměř všechny zachycené varianty pak byly detekovány i v datech EGFR-WT.

Jak uvádí Horizon Discovery, genomická DNA referenčního standardu byla získána z buněčné linie RKO. Protože v naší analýze byl stále neobjasněn velký počet variant, které se navíc často vyskytovaly ve frekvencích na úrovni germinálních mutací v hetero- nebo dokonce homozygotním stavu, byl k porovnání použit třetí zdroj sekvenčních dat pro tuto buněčnou linii od AccuRef Diagnostics (AccuRef Diagnostics, 2019). V daném VCF souboru bylo následně potvrzeno dalších 6 variant z naší analýzy, a to v genech *TP53*, *KDR*, *ATM*, *IDH1* a *APC*, přestože nebyly přítomny v datech HD (Tabulka 24).

Tabulka 24: Varianty detekované analýzou s UMI barkódy ve vzorku EGFR-MUT12 při limitní frekvenci výskytu 10 % a aplikaci filtrů (pSTD a Bias), které nebyly potvrzeny v datech Horizon Discovery.

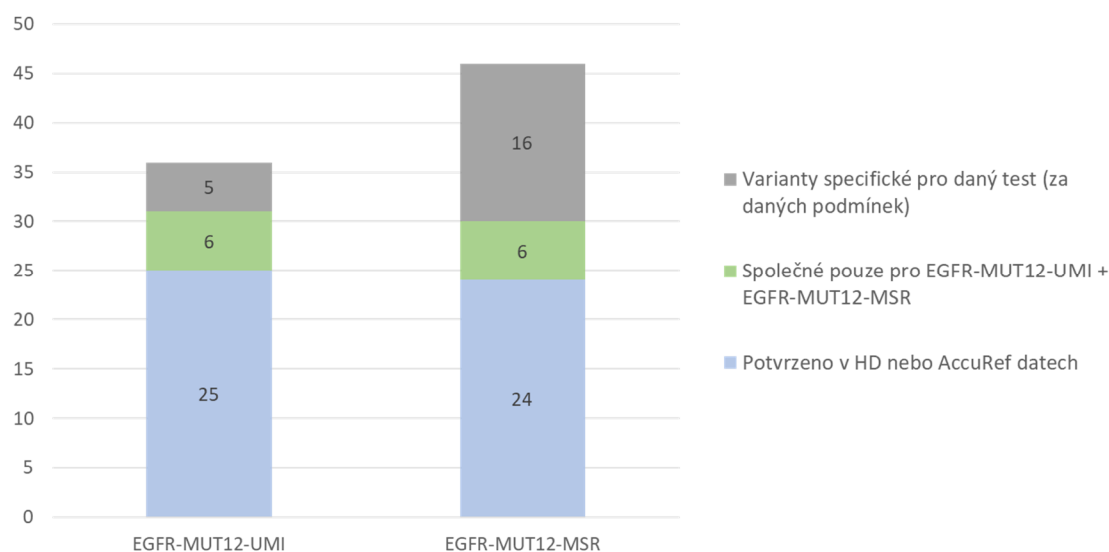
Gen	Zápis varianty na úrovni cDNA	Frekv. [%]	Pokrytí	Filtr	Výskyt v EGFR-MUT12-REF-MSR [%]	Výskyt v EGFR-WT-REF-UMI [%]	Výskyt v datech AccuRef
<i>FGFR3</i>	c.849_850delCCinsAA	37,3	115	p8	14,2	41,0	-
<i>FGFR3</i>	c.1090G>T	14,8	27	PASS	3,5	-	-
<i>JAK2</i>	c.1847_1848delGTinsAA	20,0	39	PASS	14,8	13,3	-
<i>TP53</i>	c.215C>G	52,5	59	PASS	54,9	44,7	ano
<i>RBI</i>	c.1366delG	28,6	12	p8	26,6	-	-
<i>RBI</i>	c.1366G>A	50,0	12	PASS	52,3	14,9	-
<i>KDR</i>	c.1416A>T	45,2	37	PASS	51,5	34,8	ano
<i>PTEN</i>	c.802-11T>C	32,7	53	PASS	31,6	46,2	-
<i>ATM</i>	c.2614C>T	61,5	13	PASS	57,8	57,7	ano
<i>APC</i>	c.4360A>G	25,0	32	PASS	26,2	44,2	ano
<i>APC</i>	c.4479G>A(p.=)	100	38	PASS	100	91,7	ano
<i>MET</i>	c.3688C>T(p.=)	17,1	38	PASS	4,4	20,0	-
<i>NOTCH1</i>	c.5014delC	36,7	31	p8	17,3	32,6	-
<i>NOTCH1</i>	c.4958_4959delTCinsAA	23,3	50	p8	9,5	30,2	-
<i>NOTCH1</i>	c.4957_4958delCTinsAA	12,9	48	p8	9,6	12,8	-
<i>IDHI</i>	c.315C>T(p.=)	38,9	54	PASS	46,2	39,5	ano
<i>ERBB4</i>	c.884-9_884-7delTTT	55,1	106	PASS	36,5 (R8 filtr)	50,0	-

Pozn.: Pro srovnání je uveden výskyt variant ve VCF souboru z programu MiSeq Reporter Software (MSR) a ve VCF souboru kontrolního vzorku EGFR-WT-REF (zpracován s UMI barkódy); v těchto případech bez aplikace veškerých filtrů. Pro potvrzení správnosti výsledků byla použita data s variantami buněčné linie RKO od AccuRef Diagnostics. Označení (p.=) v zápisu na úrovni cDNA představuje synonymní variantu.

Podobně byl ve VariantStudios zpracován VCF soubor z programu MSR s vyřazením variant vyskytujících se pod 10 % nebo bez statusu PASS (tj. filtr R8 = indely delší než 8 bází a filtr SB = strand bias). Poziční filtry (pSTD) v tomto případě aplikovat nelze. Z původního počtu 359 variant tímto zůstalo 46; 24 představovalo varianty potvrzené v HD nebo AccuRef datech a 22 přebývalo navíc. Jak znázorňuje Obrázek 11, z těchto 22 neověřených variant se 6 shodovalo s výsledky analýzy s UMI barkódy. Varianty zachycené pouze touto bioinformatickou analýzou předkládá Tabulka 25.

Tabulka 25: Varianty zachycené pouze programem MiSeq Reporter Software při požadované nejnižší alelické frekvenci 10 % a filtru PASS, které nebyly potvrzeny v žádných kontrolních datech.

Gen	Zápis varianty na úrovni cDNA	Frekv. [%]	Pokrytí	Filtr	Výskyt v datech HD	Výskyt v datech AccuRef	Výskyt v EGFR-MUT12-REF-UMI
FGFR3	c.831G>T	17.63	4424	PASS	-	-	-
SMAD4	c.435A>T	19.46	1079	PASS	-	-	-
RB1	c.2107A>T	20.91	5643	PASS	-	-	-
KIT	c.2362-1delG	23.99	8757	PASS	-	-	-
KIT	c.2362-1G>T	34.68	6442	PASS	-	-	-
ATM	c.960delG	20.48	11422	PASS	-	-	-
ATM	c.962delA	16.77	11321	PASS	-	-	-
ATM,C11ORF65	c.8647delG	10.66	3134	PASS	-	-	-
ABL1	c.879+6CT>AA	17.2	872	PASS	-	-	-
NOTCH1	c.7488C>A	15.05	4000	PASS	-	-	-
FGFR3	c.831G>T	17.63	4424	PASS	-	-	-
RB1	c.2107delA	10.36	6284	PASS	-	-	-
EGFR,EGFR-AS1	c.2366T>A	12.88	11789	PASS	-	-	-
EGFR,EGFR-AS1	c.2367C>A	19.93	12512	PASS	-	-	-
KDR	c.1414C>T	10.56	2406	PASS	-	-	-
FBXW7	c.1832G>T	10.32	4255	PASS	-	-	-



Obrázek 11: Zastoupení společně detekovaných variant vzorku EGFR-MUT12 jednotlivými testy, tj. sestavenou analýzou s UMI barkódou (EGFR-MUT12-UMI) nebo programem MiSeq Reporter Software (EGFR-MUT12-MSR) při nejnižší alelické frekvenci 10 % a nastavených filtrech pSTD + Bias nebo PASS. Varianty byly porovnány s daty standardu Horizon Discovery (HD) a daty AccuRef pro buněčnou linii RKO.

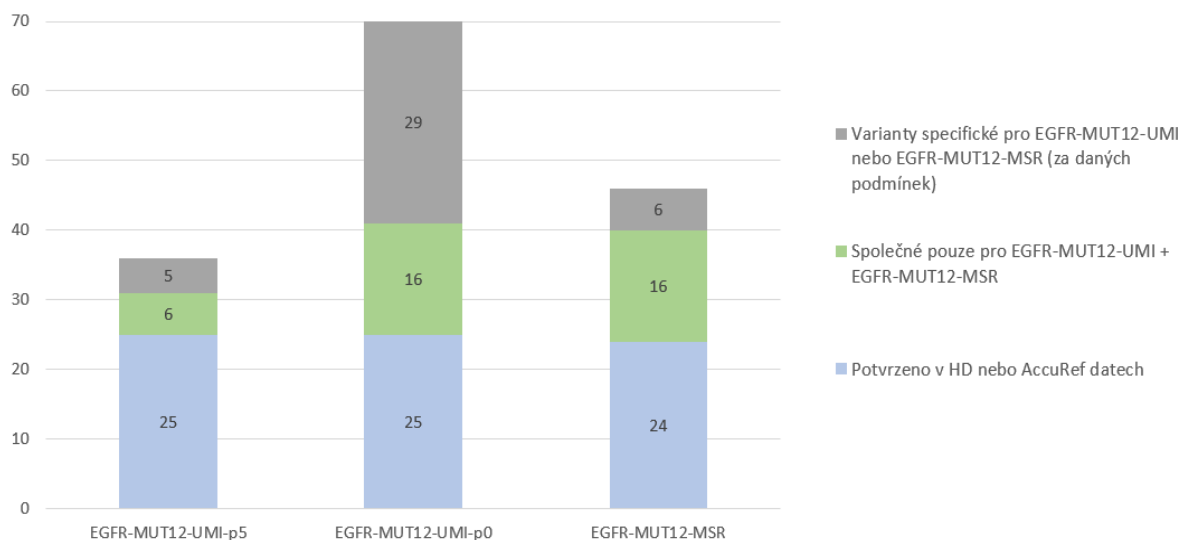
Pro velký počet zachycených variant byla při zpracování dat nastavena poměrně vysoká hodnota minimální alelické frekvence (10 %), kvůli čemuž byly v obou analýzách (s UMI barkódou a MSR) detekovány pouze 2 ze 4 validovaných mutací v *EGFR* genu. Nevýhodou

analýzy s UMI barkódy v tomto testu bylo nápadné snížení frekvence právě jedné ze 4 validovaných mutací oproti očekávání. Většina potvrzených variant daty HD nebo AccuRef (celkem 25 z 26 variant) však zůstala zachována i při nastavení přísných podmínek. Analýzou s UMI barkódy byl tímto detekován stejný počet ověřených variant jako programem MSR, zároveň se ale znatelně snížil počet variant, které se ve vzorku pravděpodobně nacházet neměly. Obě bioinformatické analýzy lze tak považovat za podobně sensitivní, přičemž UMI barkódy a poziční filtry zvyšují zejména specificitu metody. Současně bylo správně potvrzeno, že se žádné z těchto validovaných mutací nevyskytovaly v referenčním vzorku EGFR-WT, a to ani při frekvencích pod 10 %.

6.4.2 Vliv pozičních filtrů na výsledná data z analýzy s UMI barkódy

V přechozí kapitole (6.3.1) byla ověřována data generovaná pomocí shellového skriptu, jehož součástí je „variant caller“ VarDictJava. Daný program při výchozích (default) parametrech v příkazu způsobí odfiltrování variant, jejichž výskyt se nachází v koncových oblastech čtení, konkrétně v pozici pátého nukleotidu od konce nebo méně (po zprůměrování pozic ze všech čtení s touto variantou). Program vychází z předpokladu, že na koncích čtení výrazně narůstá chybovost sekvenační analýzy. Současně však může dojít ke ztrátě skutečných variant, které se vyskytují u krajů cílových regionů s obvykle nižším pokrytím. Na základě toho bylo vyzkoušeno, jaké výstupy generuje analýza bez odfiltrování těchto variant (tj. v příkazu vardict byl nastaven parametr $-P = 0$). Program VarDictJava automaticky přiřazuje status p8 všem variantám, jejichž průměrný výskyt se nachází v pozici osmého nukleotidu od konce nebo méně; varianty nejsou odstraněny z výsledného VCF souboru, ale díky statusu p8 mohou být odlišeny od ostatních variant.

Pomocí programu VariantStudio byly z VCF souboru získaného pozměněnou analýzou s UMI barkódy opět vyřazeny varianty s výskytem pod 10 %, se statusem pSTD nebo Bias. Celkový počet detekovaných variant se tímto rovnal 70; detekováno bylo tedy téměř o polovinu více variant než v předchozím testu s UMI barkódy, kde byl vardict parametr $-P$ nastaven na hodnotu 5. Ve výsledcích však nepřebýla žádná z variant potvrzených daty HD ani AccuRef, pouze se navýšil počet variant společných pro VCF soubory z MSR (Obrázek 12).



Obrázek 12: Zastoupení společně detekovaných variant vzorku EGFR-MUT12 jednotlivými testy, tj. sestavenou analýzou s UMI barkódy s pozičním filtrem (p5) nebo bez pozičního filtru (p0) a programem MiSeq Reporter Software (MSR) při nejnižší alelické frekvenci 10 % a nastavených filtrech pSTD + Bias nebo PASS. Varianty byly porovnány s daty standardu Horizon Discovery (HD) a daty AccuRef pro buněčnou linii RKO.

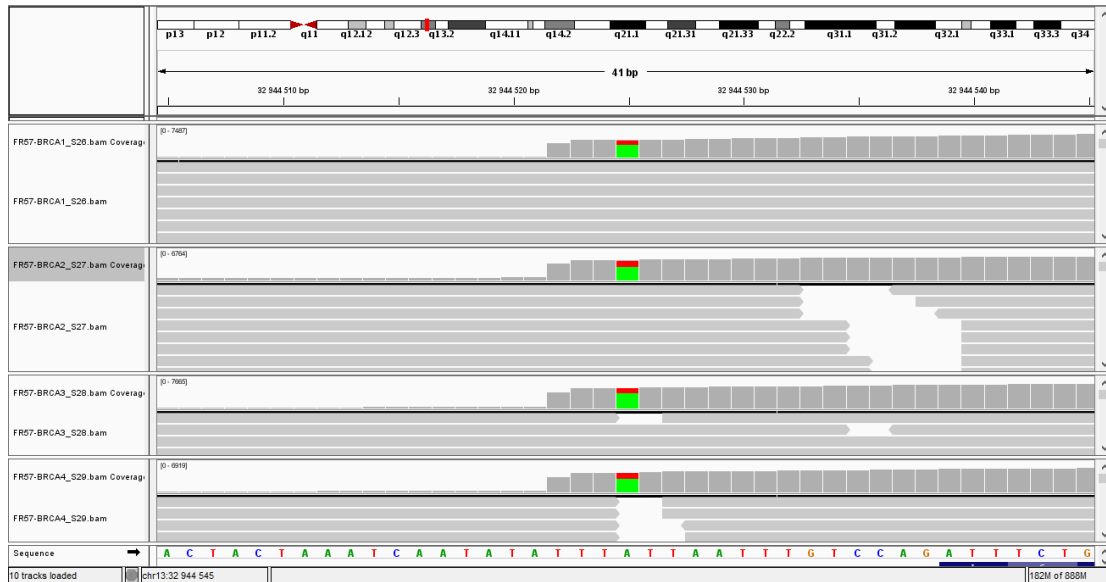
Vliv pozičních filtrů na výsledky sestavené analýzy s UMI barkódy byl sledován dále u patientských vzorků nádoru vaječníku testovaných na panelu NEBNext Direct BRCA1/BRCA2 (data pro bioinformatickou analýzu poskytl Mgr. Jana Stránská, Ph.D.). Počet detekovaných variant byl o více než 50 % nižší v analýze s UMI barkódy oproti datům z MSR, přičemž aplikace pozičního filtru (-P = 5) výsledný počet ještě více snížila. Z testu však nelze jednoznačně určit, zda kromě falešných pozitiv nebyly odfiltrovány také pozitivní správné (Tabulka 26).

Tabulka 26: Porovnání počtu detekovaných variant ve vzorcích patientských nádorů vaječníků pomocí programu MiSeq Reporter Software a analýzy s UMI barkódy bez nebo s aplikací jednotlivých filtrů.

Číslo vzorku	1	2	3	4	5	6	7	8	9	10	Celkem
MSR.vcf (celkem)	116	86	80	89	92	70	103	108	102	119	965
MSR.vcf (>5 % + PASS)	36	36	33	32	38	27	41	45	35	38	366
sUMI-p0.vcf (celkem)	91	65	54	61	64	42	70	82	74	94	697
sUMI-p0.vcf (>5 % bez STD nebo Bias)	16	19	15	11	20	11	20	21	13	15	161
sUMI-p5.vcf (>5 % bez pSTD nebo Bias)	13	16	13	10	13	8	16	17	9	12	127

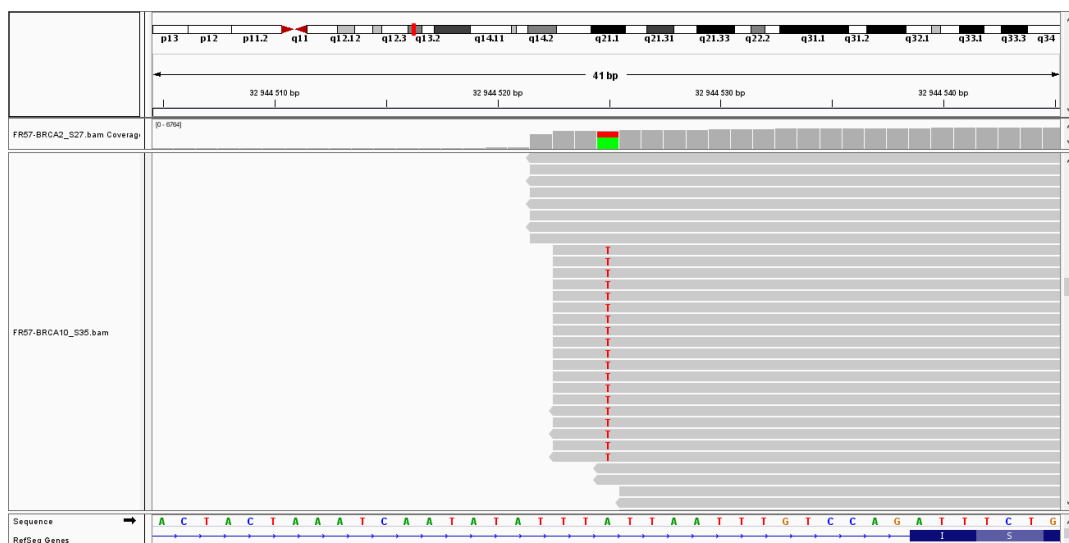
Pozn.: Soubor sUMI-p0.vcf označuje vypnutí pozičního filtru v programu VarDietJava, soubor sUMI-p5.vcf označuje naopak ponechání pozičního filtru. Kromě toho je uveden také celkový počet variant z obou analýz bez aplikace veškerých filtrů včetně odstranění limitní alelické frekvence 5 % (označeno „celkem“).

Obrázek 13 vizualizuje v programu IGV typický příklad varianty, která je odfiltrována z výsledných VCF na základě pozičního filtru ($-P < 5$). Jak je z obrázku patrné, ve čtyřech různých patientských vzorcích nádorů vaječníku byla zachycena identická varianta, a je tudíž vysoce nepravděpodobné, aby se jednalo o správnou pozitivitu.



Obrázek 13: Výskyt pozičních artefaktů ($-P < 5$) v patientských vzorcích nádorů vaječníku č. 1-4 vizualizovaný v programu IGV.

Další poziční filtr pSTD (read position standard deviation) umožňuje označit varianty, jejichž průměrná odchylka pozice ve čtení odpovídá 0. Znamená to, že se varianta vyskytuje ve většině případů ve stejné vzdálenosti od konce čtení (Obrázek 14). Výskyt varianty za těchto podmínek opět napovídá pouze vzniku artefaktu.



Obrázek 14: Výskyt pozičního artefaktu ($-P < 5$ a současně pSTD) v patientském vzorku nádorů vaječníku č. 10 vizualizovaný pomocí programu IGV.

6.4.3 Stanovení sensitivity a specificity analýzy s UMI pro genotypizaci *BRCA 1* a *2*

Sestavená bioinformatická analýza s UMI barkódy byla dále na EGFR referenčním vzorku (HD802) ověřena pro panel NEBNext Direct *BRCA1/BRCA2*. Sekvenování bylo provedeno na sekvenátoru MiSeq s délkou čtení 75 bp (párová koncová čtení). Knihovnu pro tento experiment připravily laborantky DNA laboratoře ÚMTM a veškerá data z analýzy poskytla Mgr. Jana Stránská, Ph.D.

Z VCF souborů generovaných analýzou s UMI (-P = 5) byly vyřazeny varianty s frekvencí pod 5 % a se statusem pSTD nebo Bias. Podle exomových dat pro EGFR referenční standard, jež jsou uvedena na stránkách výrobce Horizon Discovery, mělo být celkem zachyceno 23 variant. Všechny byly úspěšně potvrzeny, ačkoli 7 variant chybělo ve výsledném souboru VCF. Tyto varianty však byly detekovány v datech ve formátu BAM a k jejich závěrečnému odfiltrování došlo z důvodů výskytu těchto variant v oblastech mimo regiony zahrnutých v BED souboru používaném pro „variant calling“ (soubor poskytuje výrobce genového panelu NEBNext Direct). V analýze byly naopak detekovány některé varianty navíc, které Horizon Discovery neuvádí, přičemž všechny tyto se nacházely buď v intronech, nebo v 3' nepřekládané oblasti (3' UTR). V datech HD jsou tyto varianty řazeny do kategorie Impact MODIFIER, která byla z výsledků odfiltrovaná. Nelze proto určit, zda byly analýzou zachyceny pouze artefakty, nebo varianty skutečně přítomné (Horizon Discovery, 2019) (Tabulka 27).

Tabulka 27: Hledané varianty pro zjištění sensitivity analýzy s UMI barkódy.

Správně detekované varianty			Nedetekované varianty ve VCF			
Gen	Zápis varianty na úrovni cDNA	Konsekvence	Gen	Zápis varianty na úrovni cDNA	Důvod neshody	
<i>BRCA1</i>	c.4900A>G	substituce AK	<i>BRCA1</i>	c.5531-121delAAA	Regiony mimo BED soubor (NEBNext Direct)	
	c.3548A>G	substituce AK		c.5470-151G>T		
	c.2612C>T	substituce AK		c.5470-193A>G		
	c.2458A>G	substituce AK		c.5469+68T>C		
	c.2082C>T	synonymní var.		c.5215+66G>A		
	c.1303G>T	substituce AK		c.5137+65G>A		
<i>BRCA2</i>	c.865A>C	substituce AK		c.5531-121delAAA		Možný důvod neshody
	c.1365A>G	synonymní var.		Detekované varianty navíc oproti datům Horizon Discovery		
	c.2229T>C	synonymní var.	Gen	Zápis varianty na úrovni cDNA		
	c.2971A>G	substituce AK	<i>BRCA1</i>	c.*36C>G	Kategorie Impact MODIFIER	
	c.3516G>A	synonymní var.		c.212+23T>A		
	c.3807T>C	synonymní var.	<i>BRCA2</i>	c.1909+21_1909+22delTT		
	c.4563A>G	synonymní var.		c.*105A>C		
	c.5351delA	posun čtecího rámce				
	c.6513G>C	synonymní var.				
c.7397T>C	substituce AK					

Ke stanovení specifity byly z publikace Spectrum and characterisation of BRCA1 and BRCA2 deleterious mutations in high-risk Czech patients with breast and/or ovarian cancer (Machackova *et al.*, 2008) vybrány nejčastější varianty genů BRCA (SNP, delece, duplikace, sestřihové varianty) v České republice (vyskytující se ve více než 1 rodině) a byly porovnány jak s exomovými daty HD, tak s daty získaných z námi ověřované analýzy, kde by se vyskytovat neměly. Celkem bylo hledáno 29 variant a žádná z nich se v testovaných datech správně nevyskytla (Tabulka 28).

Tabulka 28: Hledané varianty pro zjištění specifity analýzy s UMI barkódy

Gen	Exon	Zápis varianty na úrovni cDNA	Genotyp v datech:		Gen	Exon	Zápis varianty na úrovni cDNA	Genotyp v datech:	
			HD	s UMI				HD	s UMI
BRCA1	2	c.68_69del2	wt	wt	BRCA2	IVS5	c.476-2A>G*	wt	wt
	3	c.115T>C	wt	wt		10	c.1389_1390del2	wt	wt
	5	c.181T>G	wt	wt		10	c.1813dupA	wt	wt
	IVS6	c.302-3C>G*	wt	wt		11A	c.2808_2811del4	wt	wt
	11A	c.1600C>T	wt	wt		11B	c.3847_3848del2	wt	wt
	11A	c.1687C>T	wt	wt		11D	c.5645C>A	wt	wt
	11B	c.2157dupA	wt	wt		11E	c.6449_6450del2	wt	wt
	11B	c.2263G>T	wt	wt		11E	c.6638delC	wt	wt
	11B	c.2411_2412del2	wt	wt		17	c.7913_7917del5	wt	wt
	11B	c.2488_2497dup10	wt	wt		18	c.8042_8043del2	wt	wt
	11B	c.2762delA	wt	wt		20	c.8537_8538del2	wt	wt
	11C	c.3642_3643del2	wt	wt		IV523	c.9117+2T>A*	wt	wt
	11C	c.3700_3704del5	wt	wt		25	c.9403delC	wt	wt
	11C	c.3756_3759del4	wt	wt					
	12	c.4185G>A*	wt	wt					
20	c.5266dupC	wt	wt						

Analytická sensitivita a specifita byly vypočteny na základě následujících vzorců:

$$\text{Sensitivita} = \frac{\text{správná pozitivita}}{\text{správná pozitivita} + \text{falešná negativita}} = \frac{16}{16+0} = 100 \%$$

$$\text{Specifita} = \frac{\text{správná negativita}}{\text{správná negativita} + \text{falešná pozitivita}} = \frac{29}{29+0} = 100 \%$$

6.5 Detekce mutací s nízkou frekvencí NEB Cancer panelem za použití vzorků HD802 a HD709

Genový panel NEBNext Direct Cancer HotSpot Panel byl s pomocí EGFR referenčních vzorků (HD802 EGFR Gene-Specific Multiplex Reference Standard gDNA) ověřen pro schopnost detekce somatických mutací. V testech byly sledovány dva hlavní aspekty: 1) zda

jsou spolehlivě detekovány všechny validované referenční mutace při limitním množství vstupní DNA (<100 ng) a 2) jaká je nejnižší detekovatelná frekvence těchto mutací.

V prvním testu byl analyzován referenční vzorek EGFR-MUT3 při vstupním množství DNA 42 ng a frekvenci referenčních mutací 3,1 %. Sekvenování proběhlo na platformě MiSeq s délkou čtení 75 bp (párová koncová čtení), přičemž k hodnocení výsledků byla použita data generovaná programem MSR. Tři z celkových čtyř mutací byly správně detekovány s téměř odpovídajícími hodnotami frekvence výskytu (Tabulka 29). Test však nezachytil očekávanou delecí $\Delta E746-A750$.

Tabulka 29: Analýza EGFR-MUT3: množství DNA na přípravu knihovny 42 ng; vstupní frekvence *EGFR* mutací 3,1 %. Uvedená data pochází z VCF generovaným programem MSR.

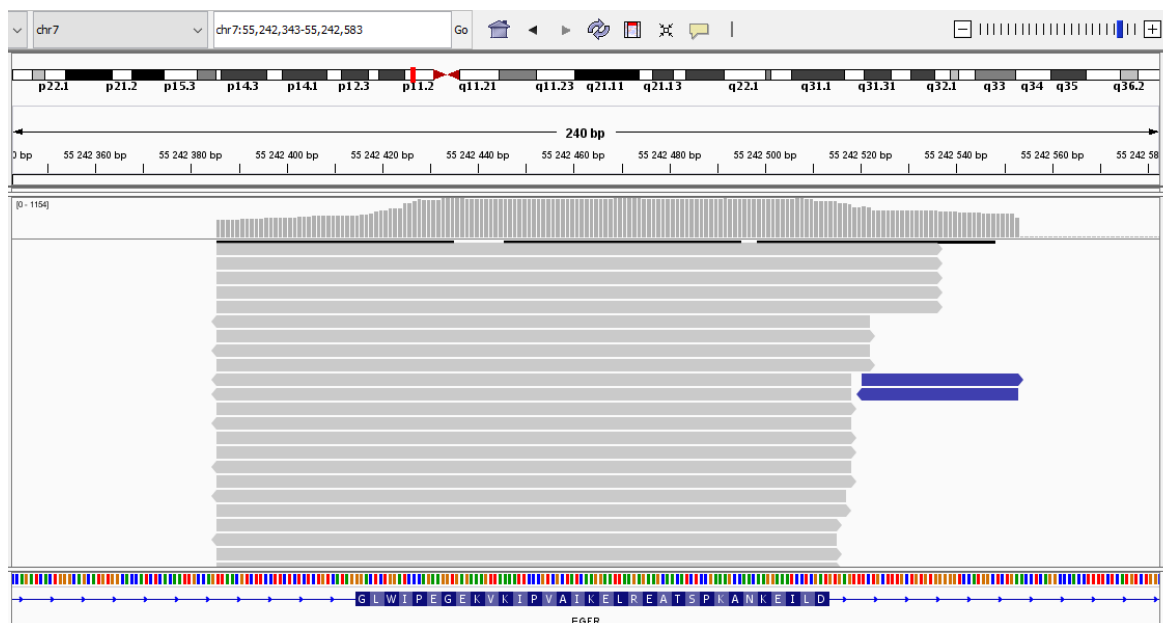
Gen	Varianta	Frekvence	Pokrytí
<i>EGFR</i>	T790M	2,12 %	5321
	L861Q	3,48 %	13398
	L858R	3,05 %	7282
	$\Delta E746-A750$	-	-

V druhém testu byl analyzován referenční vzorek EGFR-MUT12 při podobném vstupním množství DNA (39,4 ng), ale výrazně vyšší frekvenci referenčních mutací 12,5 %. Sekvenování proběhlo na platformě MiSeq s délkou čtení 75 bp (párová koncová čtení). K hodnocení výsledků byly použity VCF soubory generované programem MSR i nově sestavenou analýzou s UMI barkódy. Informace o referenčních mutacích byly ve VCF souborech vyhledány pomocí programu VariantStudio bez aplikace jakéhokoliv filtru (Tabulka 30). Protože očekávaná delecce nebyla opět detekována, byl vzorek osekvenován ještě jednou s délkou čtení 2 x 150 bp. Nicméně jak ukazuje Obrázek 15, ani přes prodlouženou délku čtení se mutaci zachytit nepodařilo.

Tabulka 30: Analýza vzoru EGFR-MUT12: množství DNA na přípravu knihovny 39,4 ng; vstupní frekvence *EGFR* mutací 12,5 %. Uvedená data pochází z VCF generovaných MiSeq reporterem při dvou délkách čtení 2 x 75 bp a 2 x 150 bp.

Gen	Varianta	PE 75		PE 150	
		Frekvence	Pokrytí	Frekvence	Pokrytí
<i>EGFR</i>	T790M	7,08 %	12577	8,35 %	2755
	L861Q	11,23 %	9930	11,5 %	2486
	L858R	12,76 %	12010	7,95 %	2500
	$\Delta E746-A750$	-	-	-	-

Pozn.: PE označuje párová koncová čtení (paired-end)



Obrázek 15: Vizualizace genomického regionu s očekávanou delecí $\Delta E746$ – $A750$ genu *EGFR* (koordináta: chromozom 7; 55242463–55242478) pomocí programu IGV; vzorek EGFR-MUT12 (PE 150).

6.6 Porovnání NEB Cancer panelu a panelu TruSight Tumor 170 na vzorku nádoru slinivky

Výstupy „hotspotového“ panelu NEB o 50 genech byly srovnány s panelem TruSight Tumor 170 (knihovnu připravila Mgr. Lucie Kotková) pokrývající celé exony 170 genů. K oběma testům byl použit stejný patientský vzorek nádoru slinivky. Sekvenování proběhlo na platformě MiSeq s délkou čtení 75 bp (párová koncová čtení) bez zahrnutí UMI barkódů do analýzy. VCF soubory generované přímo na sekvenátoru byly následně zpracovány v programu VariantStudio.

Z výsledků (Tabulka 31, Tabulka 32) byly vyřazeny varianty: 1) s frekvencí pod 5 %, 2) bez statusu PASS, 3) s pokrytím méně než 100x, 4) bez vlivu na kódující potenciál transkriptu (aplikace filtru konsekvence) a 5) s výskytem v populaci nad 5 %.

Jak zobrazuje Tabulka 31, mutace zachycené oběma panely se vyskytovaly v genech *KRAS*, *SMAD4* a *TP53* (vyznačené tučně). Absence téměř všech mutací identifikovaných pouze s pomocí TST170 je v testu NEB zcela očekávaná, protože sondy tohoto panelu příslušné regiony nepokrývají. Výjimkou je mutace genu *GNAS*, která měla být zachycena, a přesto se ve VCF souboru z NEB testu nevyskytuje. Je však zahrnuta v příslušném souboru BAM a k jejímu odfiltrování došlo až programem Somatic Variant Caller. Co se týče mutací

identifikovaných pouze s NEB Cancer panelem, jedná se vždy o varianty s velmi nízkou frekvencí (<10 %), což v tomto případě naznačuje spíše vznik artefaktů (všechny dané regiony TST170 pokrývá).

Tabulka 31: Detekované mutace v nádoru slinivky panelem TruSight Tumor 170

Gen	HGVS zápis varianty	Typ	Frekv. [%]	Pokrytí	Významnost dle ClinVar	PolyPhen ¹
<i>ARID1A</i>	NP_006006.3:p.Arg2236ProfsTer42	inzerce	25,15	3790	-	-
<i>MSH2</i>	NP_000242.1:p.Gly322Asp	snv	50,44	2522	pravděp. benigní	benigní
<i>NOTCH1</i>	NP_060087.3:p.Pro1377Ser	snv	30,51	590	benigní	benigní
<i>KRAS</i>	NP_203524.1:p.Gly12Val	snv	39,11	2002	patogenní	může poškodovat
<i>CDH1</i>	NP_004351.1:p.Asp805Asn	snv	46,65	1582	pravděp. patogenní	může poškodovat
<i>TP53</i>	NP_000537.3:p.Arg273His	snv	37,25	1949	pravděp. patogenní	může poškodovat
<i>BRCA1</i>	NP_009231.2:p.Asp1152Asn	snv	44,39	2354	-	benigní
<i>BRCA1</i>	NP_009231.2:p.Ser1040Asn	snv	39,02	2458	benigní	benigní
<i>SMAD4</i>	NP_005350.1:p.Leu535ThrfsTer15	delece	24,92	1553	-	-
<i>GNAS</i>	NP_536350.2:p.Asp837Asn	snv	33,63	1567	-	pravděp. poškozuující
<i>CHEK2</i>	NP_001005735.1:p.Lys416Glu	snv	19,4	1206	-	pravděp. poškozuující

¹PolyPhen podává *in silico* predikci účinku AK substituce na protein

Tabulka 32: Detekované mutace v nádoru slinivky kitem NEBNext Direct Cancer HotSpot Panel

Gen	HGVS zápis varianty	Typ	Frekv. [%]	Pokrytí	Významnost dle ClinVar	PolyPhen ¹
<i>ERBB4</i>	NP_005226.1:p.Leu932Phe	snv	7.45	3935	-	možná poškozuující
<i>PIK3CA</i>	NP_006209.2:p.Ile330Lys	snv	5.46	2693	-	benigní
<i>EGFR-ASI</i>	NP_005219.2:p.Ile789Asn	snv	8.53	2651	-	pravděp. poškozuující
<i>KRAS</i>	NP_203524.1:p.Gly12Val	snv	42.81	5622	patogenní	možná poškozuující
<i>RBI</i>	NP_000312.2:p.Ile703Phe	snv	9.2	1457	-	pravděp. poškozuující
<i>TP53</i>	NP_000537.3:p.Arg273His	snv	43.31	5301	pravd. patogenní	možná poškozuující
<i>SMAD4</i>	NP_005350.1:p.Leu535ThrfsTer15	delece	13.18	4446	-	-

¹PolyPhen podává *in silico* predikci účinku AK substituce na protein

6.7 Analýza mutací v chordomech pomocí NEB Cancer panelu

Panelové sekvenování za použití NEB Cancer panelu se v rámci základního kitu neprokázalo jako vhodná metoda diagnostiky plicních karcinomů. Panel byl proto využit k analýze vzorků chordomů, jakožto málo prostudovaných nádorů, pro které jsou možnosti cílené terapie ověřeny jen velmi okrajově. Následné experimenty si tímto kladly za cíl zjistit,

jaké mutace budou panelem v chordomech detekovány, a zda na základě těchto dat mohou být nalezena potenciálně efektivní léčiva pro testované pacienty.

Analyzovány byly celkem čtyři vzorky chordomů: CH22042/07B, CH7612/09B, CH17069/09-2 a CH16165/13B. DNA knihovny byly připraveny za použití kitu NEBNext Direct Cancer HotSpot Panel pokrývající 50 genů, a poté osekvenovány na platformě MiSeq s délkou čtení 75 bp (párová koncová čtení). Výsledky byly zpracovány sestavenou bioinformatickou analýzou s UMI barkódou, přičemž poziční filtr -P pro „variant calling“ byl nastaven na výchozí hodnotu dle programu VarDictJava (tj. -P=5).

Výsledné VCF soubory byly vyhodnocovány v programu VariantStudio, kde byly z analýzy vyřazeny varianty: 1) s frekvencí pod 5 %, 2) bez vlivu na kódující potenciál transkriptu (aplikace filtru konsekvence), 3) s výskytem v populaci nad 5 %, 4) se statusem pSTD nebo Bias (= strand bias) a 5) varianty mimo kódující oblasti. Všechny detekované varianty za těchto podmínek uvádí (Tabulka 33).

Tabulka 33: Detekované mutace ve vzorcích chordomů

Vzorek	Gen	Typ	Zápis varianty na úrovni proteinu, příp. cDNA	Frekv. [%]	Filtr	Významnost dle ClinVar	PolyPhen ¹
CH17069/09-2	<i>NOTCH1</i>	mnp	p.Asp1670Val	5,4	p8	-	možná poškozující
CH22042/07B	<i>MET</i>	snv	p.Arg988Cys	60,8	PASS	spíše benigní	benigní
	<i>JAK2</i>	snv	c.1777-2A>T	6,1	PASS	-	-
CH7612/09B	<i>TP53</i>	snv	p.Arg110Cys	6,8	PASS	spíše patogenní	pravděp. poškozující
CH16165/13B	<i>MLH1</i>	snv	p.Leu447Ser	9,8	p8	-	benigní
	<i>PIK3CA</i>	snv	p.Ile112Asn	5,3	p8	patogenní	možná poškozující
	<i>PIK3CA</i>	snv	p.Arg916Cys	5,1	p8	-	možná poškozující
	<i>FGFR3</i>	mnp	p.His284Asn	29,6	p8	-	-
	<i>KDR</i>	mnp	p.Trp485Lys	7,2	p8	-	pravděp. poškozující
	<i>KDR</i>	snv	p.Cys482Arg	34,6	PASS	-	pravděp. poškozující
	<i>NOTCH1</i>	mnp	p.GlnSer2315 HisCys	17,9	p8	-	
<i>RBI</i>	snv	p.Tyr155Phe	5,6	p8	-	pravděp. poškozující	

¹PolyPhen podává *in silico* predikci účinku AK substituce na protein

7 DISKUZE

Schopnost metody citlivě a specificky detekovat somatické mutace o nízkých frekvencích v širším spektru genů a při limitním množství vstupní DNA je dnes jedním z hlavních požadavků současné personalizované medicíny (Dong *et al.*, 2015; Vollbrecht *et al.*, 2018). V experimentální části práce byl na základě tohoto předpokladu ověřován genový panel NEBNext Direct Cancer HotSpot Panel, zejména s ohledem na jeho možné využití pro testování pacientů s karcinomem plic. Jak uvádějí Vollbrecht *et al.* (2018), panelové sekvenování s UMI barkódy může nabízet výhodnou diagnostickou strategii včetně analýzy cfDNA.

Plicní karcinomy jsou velmi často podmíněny mutacemi onkogenu *EGFR*, které se osvědčily jako účinné molekulární cíle v protinádorové terapii. Již několik klinických studií fáze III potvrdilo vyšší efektivitu monoterapie inhibitory EGFR oproti chemoterapii, díky čemuž jsou nyní cílená léčiva v daných typech nádorů standardní léčbou první volby (Bulbul *et Husain*, 2018). Experimenty v této práci zaměřené na možné diagnostické uplatnění NEB Cancer panelu byly proto prováděny na referenčním standardu se čtyřmi *EGFR* validovanými mutacemi. Vzhledem k původnímu záměru analýzy cfDNA byla snaha pro panel stanovit detekční limit, při němž jsou spolehlivě detekovány mutace *EGFR* i s velmi nízkou frekvencí výskytu.

Testy ke stanovení LOD byly provedeny na dvou referenčních vzorcích s odlišnou alelickou frekvencí (3 % a 12,5 %), a ačkoliv panel všechny regiony příslušných mutací pokrývá, ani jeden z těchto testů nebyl úspěšný v detekci delece $\Delta E746-A750$. Protože *EGFR* delece v exonu 19 včetně této specifické mutace jsou důležitými ukazateli odpovědi na cílenou léčbu, chyba v jejich detekci znamenala závažný nedostatek analýzy (Lynch *et al.*, 2004). Jak bylo následně zjištěno, pozice mutace se nacházela právě v koncové oblasti sond panelu. Na základě komunikace s technickou podporou společnosti New England Biolabs (NEB) bylo vyzkoušeno zopakování sekvenování s prodlouženou délkou čtení na 150 bp. Problém však tímto nebyl vyřešen, na základě čehož nám společnost NEB nabídla alternativní možnost navržení odlišného designu sond pro vybraný exon.

Zakoupení nové verze kitu v našem případě ale nepředstavovalo výhodné řešení a od dalších experimentů zaměřených na možnosti detekce mutací s nízkou frekvencí bylo z těchto důvodů upuštěno. Pokud bychom však skutečně chtěli stanovit detekční limit panelu, bylo by potřeba provést analýzu i minimálně při alelické frekvenci ≤ 1 %, a také na větším počtu

vzorků či opakování. Nicméně samotné uspořádání experimentů pro testování detekčního limitu se může značně lišit. Například Vollbrecht *et al.* (2018) stanovili LOD metody na základě analýzy referenčního vzorku s 8 validovanými mutacemi při třech různých alelických frekvencích (0,1 %, 1 % a 5 %), přičemž za dosažený limit detekce považovali nejnižší alelickou frekvenci, kdy byly detekovány všechny sledované mutace. Podle Jennings *et al.* (2017) však může být nejnižší detekční limit definován také jako minoritní alelická frakce, při které bude 95 % vzorků spolehlivě detekováno. V té nejjednodušší podobně to na příkladu testu 20 vzorků znamená, že se spolehlivostí $\geq 95\%$ při daném LOD bude z celkového počtu vzorků dosažen správný výsledek v 19 případech. Ideální počet analyzovaných vzorků v souladu se statistickými principy pak může být stanoven podle vzorce $r^n = \alpha$, kde r je spolehlivost, n počet vzorků a α hladina významnosti. Při běžně používané hladině významnosti 5 % a spolehlivosti 95 % je tímto pro validaci metody doporučeno testovat až 59 vzorků, což je evidentně daleko nad rámec této diplomové práce (Jennings *et al.*, 2017).

Metoda detekce mutací nádorové DNA pomocí NEB Cancer panelu kromě přípravy knihovny a sekvenační analýzy zahrnuje také zpracování a interpretaci výstupních dat. Pro tyto účely byla sestavena bioinformatická analýza využívající informaci z molekulárních barkódů a výsledky byly porovnány s výstupy, které generuje program MiSeq Reporter Software. Data byla zpracovávána na EGFR referenčních vzorcích a k vyhodnocení správnosti obou bioinformatických přístupů byla použita exomová data EGFR referenčního standardu poskytovaná výrobcem Horizon Discovery a VCF soubor s variantami RKO buněčné linie od AccuRef Diagnostics. Obě bioinformatické analýzy prokázaly vysokou sensitivitu s téměř 100% shodou v detekci variant potvrzených daty HD či AccuRef. V případě analýzy s UMI barkódy nebyly v porovnání s daty HD za daných podmínek detekovány celkem varianty tři, v případě MSR varianty čtyři. Pouze dvě varianty (společně pro analýzu s UMI i MSR) pak nebyly ve výstupních souborech vůbec zachyceny ani po odstranění veškerých filtrů. Jedna z nich představovala právě validovanou delecí *EGFR*, druhá se měla nacházet v genu *FGFR3* (varianta typu SNP). Je možné, že jednonukleotidová varianta byla jen falešnou pozitivitou v exomových datech HD, protože není uvedena ani ve VCF souboru AccuRef Diagnostics. Nicméně k jednoznačnému určení správného závěru by bylo vhodné provést další test na referenčním standardu, například osekvenovat daný region pomocí Sangerovy metody.

Výsledky současně ukázaly, že metoda panelového sekvenování zachytila velké množství variant bez objasněného původu. I po nastavení přísných podmínek pro odfiltrování potenciálních artefaktů bioinformatická byla analýza s UMI barkódy detekováno 36 variant, z nichž jen 25 potvrdila data HD nebo AccuRef. Přesto tento přístup prokázal výhodnější řešení oproti zpracování sekvenačních dat pomocí základního programu MSR, který sice provádí „variant calling“, ale nepracuje ani s UMI barkódy, ani s pozičními filtry. Sestavenou bioinformatickou analýzou s UMI barkódy tak bylo zachyceno o celých 50 % méně variant, které se ve vzorku vyskytoval pravděpodobně neměly.

Jedním z vysvětlení vzniku velkého počtu artefaktů ve vzorku EGFR-MUT12 může být nápadně nižší dosažené bázové pokrytí v porovnání s ostatními vzorky testovanými v této práci. Zda všechny z těchto nepotvrzených variant kontrolními daty skutečně ve vzorku přítomny nebyly, přesto nemusí být ale jednoznačné. Celoxomová analýza bývá prováděna s pokrytím několikanásobně nižším oproti panelovému sekvenování (HD uvádí průměrné pokrytí 60x), metoda proto není zcela spolehlivá v zachytu variant s nižší frekvencí výskytu. Koneckonců i neshoda dat Horizon Discovery a AccuRef Diagnostics získaných z totožné buněčné linie RKO ukazuje na možné odchylky výsledků.

Bioinformatická analýza s UMI barkódy byla na stejném EGFR referenčním standardu dále ověřena pro panel NEBNext Direct BRCA1/BRCA2 pokrývající celé exony dvou genů *BRCA1* a *BRCA2*. V tomto případě byla stanovena 100% analytická sensitivita i specificita, ačkoli podobně jako je výše diskutováno u výpočtu detekčního limitu, k získání statisticky hodnotných výsledků by bylo nutné analýzu otestovat na větším počtu vzorků.

Výhodnost metod pokrývajících širší genomické regiony byla sledována v experimentu porovnávající NEB Cancer panel s panelem TruSight Tumor 170 pokrývajícím celé exony 170 genů. Analýza pomocí TST170 vyžaduje použití 40 ng DNA, množství tedy není vyšší oproti doporučenému vstupu pro menší panel. Příprava knihovny o délce dva dny je však přibližně jednou tak delší oproti pracovnímu postupu NEB Cancer panelu (Illumina, 2019c; New England Biolabs, 2019b). Samotný panel TST170 sice nezahrnuje UMI barkódy, lze je ale přikoupit v rámci dalšího kitu (Illumina, 2019c).

Panely byly testovány na stejném patientském vzorku nádoru slinivky, přičemž z potenciálně patogenních mutovaných genů na základě databází ClinVar a PolyPhen byly oběma metodami detekovány *TP53* a *KRAS*. Mutace *KRAS* jsou obecně asociovány se špatnou prognózou a rezistencí na léčiva, role *TP53* v predikci léčebné odpovědi není

jednoznačná (Hyman *et al.*, 2017; Cicenas *et al.*, 2017). Jako nejlépe zacílitelná mutace se tedy jeví mutace *CHEK2* zachycená pouze panelem TST170. *CHEK2* se podílí na opravných mechanismech DNA skrze homologní rekombinaci, jejichž defekty způsobují tzv. fenotyp „BRCAness“ (Moschetta *et al.*, 2016). V těchto případech mohou být efektivní léčiva typu olaparib účinkující na principu syntetické letality, jak také dokazují Golan *et al.* (2018) ve studii fáze II pro léčbu nádoru slinivky.

I když je použití panelu TST170 finančně i časově náročnější, výsledky experimentu dokazují, jak rozsah analýzy může zásadně ovlivnit možnosti pacientovi léčby. Kromě toho, že jsou větší panely pokrývající celé kódující regiony genů schopné zachytit více klinicky významných biomarkerů, jsou také vhodné pro detekci změn v počtu kopií a strukturních variant (Hyman *et al.*, 2017). Panely nad ~300 genů lze pak využít i pro stanovení mutační zátěže (tumor mutational burden; TMB), která hraje důležitou roli v imunoterapii pomocí „checkpoint“ inhibitorů (Meléndez *et al.*, 2018).

Přesto i malé panely s 50 nebo ještě méně geny nalézají své klinické uplatnění a ověřily se již v celé řadě studií (Kamps *et al.*, 2017). Také v experimentech s chordomy NEB Cancer panel přinesl v této práci některé užitečné informace. V jednom ze čtyř testovaných vzorků (CH22042/07B) byla zachycena bodová mutace genu *MET* o velmi vysoké frekvenci (61 %), díky které je velmi nepravděpodobné, aby se jednalo pouze o falešnou pozitivitu. c-Met je tyrozinkinázový receptor pro hepatocytární růstový faktor, jehož abnormální funkce bývá v chordomech často popisována (Weinberger *et al.*, 2005; Walcott *et al.*, 2012). Podle databází PolyPhen a ClinVar je mutace R988C sice pravděpodobně benigní, bylo však prokázáno, že tlumí ubikvitinaci a degradaci receptoru za podpoření jeho signalizace (Tovar *et Graveel*, 2017). Inhibitory c-Met včetně jejich účinku na mutaci R988C jsou aktivně studovány a v některých dalších typech rakovin přinesly velmi pozitivní účinky (Ou *et al.*, 2011; Ricciuti *et al.*, 2018). V chordomech byl pro mutovaný *MET* klinicky ověřen crizotinib, ačkoliv pouze v jedné případové studii a bez uvedení konkrétní mutace tohoto genu. Léčený pacient dosáhl 17 měsíčního přežití bez progresu (Liang *et al.*, 2015). Další z detekovaných mutací vzorku CH22042/07B se nacházela v genu *JAK2*, vyskytovala se ale jen při nízké frekvenci (<10 %) a pro chordomy není charakteristická (Meng *et al.*, 2019; Walcott *et al.*, 2012).

Z potenciálně patogenních genů detekovaných ve vzorku CH16165/13B jsou s chordomy významně asociovány *KDR* a *PIK3CA* (Meng *et al.*, 2019; Fischer *et al.*, 2015). Proti mutovanému *KDR*, označovaném také jako *VEGFR-2*, bylo již v chordomech testováno

několik cílených léčiv. Klinický efekt prokázaly zejména tyrozinkinázové inhibitory sorafenib, pazopanib a sunitinib. Šanci na příznivou léčebnou odpověď v tomto případě snižuje právě přítomnost mutací *PIK3CA*, které jsou v chordomech spojovány s rezistencí na terapii (Meng *et al.*, 2019). Pro potvrzení těchto variant by bylo z důvodů jejich nízké frekvence a statusu p8 vhodné provést další ověřující test.

Ve vzorcích CH17069/09-2 a CH7612/09B nebyly nalezeny žádné klinicky významné mutace v souvislosti s chordomy. Změny v Notch signální dráze jsou sice u tohoto onemocnění pozorovány, výskyt somatických mutací v „hotspotových“ oblastech genu *NOTCH1* není ale typický (Heery, 2016; Fischer *et al.*, 2015). Vzhledem k nízké frekvenci této varianty (<10 %) a statusu p8 se může s velkou pravděpodobností jednat pouze o artefakt. Tumor supresorový gen *TP53* zde byl již zmiňován v souvislosti s nádorem slinivky a ani v případě chordomů nepatří tyto mezi významné pro současné možnosti terapie (Meng *et al.*, 2019).

8 ZÁVĚR

Analýza panelu genů je díky výhodné kombinaci cenové efektivity, citlivosti a rozsahu současně nejpoužívanější metodou sekvenování nové generace pro diagnostické testování somatických mutací v nádorové DNA. Omezení testu pouze na určité oblasti zájmu umožňuje dosáhnout vyššího sekvenačního pokrytí, které je vyžadováno pro detekci variant vyskytujících se i při nízkých frekvencích. Zároveň jsou větší panely schopny pokrýt většinu, ne-li všechny klinicky relevantní biomarkery, pro které dnes existují dostupná cílená léčiva. S ohledem na uplatnění v klinice proto další rozšíření analýzy na úroveň celoxomového nebo dokonce celogenomového sekvenování zatím spíše jen komplikuje zpracování a interpretaci výstupních dat. Rostoucí význam panelů zahrnující celé exony mnoha genů oproti testům zaměřených pouze na „hotspotová“ místa je však bezpochybný, a to nejen z důvodů zachycení většího množství biomarkerů, ale také možnosti stanovení mutační zátěže. Přínos analýzy s širším rozsahem ukázal i jeden z experimentů této práce srovnávající „hotspotový“ panel NEBNext Direct Cancer HotSpot Panel o 50 genech s panelem TruSight Tumor 170 pokrývající všechny exonické regiony 170 genů. Oba panely byly ověřeny na patientském vzorku nádoru slinivky, ve kterém byla detekována jediná klinicky významná mutace, a to v genu *CHEK2*. Protože mutace byla zachycena jen pomocí panelu TruSight Tumor 170, pacient testovaný pouze s menším panelem by tak v tomto případě přišel o potenciálně účinnou léčbu.

Panelové sekvenování genů s molekulárními barkódy se dnes mimo jiné prokázalo jako užitečný nástroj detekce cirkulující nádorové DNA, kdy je vyžadován detekční limit pod 5 % i při použití nízkého množství vstupní DNA (~10 ng). S cílem ověřit využití panelu NEBNext Direct Cancer HotSpot Panel pro analýzu karcinomů plic za těchto podmínek byly v experimentální části práce prováděny testy na referenčním standardu se čtyřmi validovanými mutacemi v genu *EGFR*. Přestože metoda celkově vykazovala vysokou analytickou sensitivitu, nebyla úspěšná při detekci specifické delece, která je významným biomarkerem v cílené terapii tohoto typu nádorů.

Panel NEBNext Direct Cancer HotSpot Panel byl z těchto důvodů uplatněn k jiným účelům, konkrétně pro analýzu chordomů patřící mezi poměrně málo prostudované nádory. Nové poznatky o molekulárních charakteristikách tohoto onemocnění, stejně jako o možnostech léčby a efektivních diagnostických metodách mohou být v tomto směru proto cenným přínosem. V rámci provedených experimentů v této práci byly zajímavé výsledky získány

především v případě vzorku s mutovaným genem *MET*, proti kterému bylo již u chordomu klinicky ověřeno cílené léčivo crizotinib.

Významnou část experimentální práce tvořilo také sestavení a otestování bioinformatické analýzy umožňující začlenění molekulárních barkódů do sekvenačních dat. Analýza ve srovnání s programem MiSeq Reporter Software, který informaci molekulárních barkódů nevyužívá, prokázala prakticky stejně vysokou citlivost, a současně znatelně snížila počet zachycených variant, jež se ve vzorku správně vyskytovat neměly.

9 SEZNAM POUŽITÉ LITERATURY

Abrahams E. *et* Silver M. (2009): The case for personalized medicine. *Journal of Diabetes Science and Technology*: 3(4): 680-4.

AccuRef Diagnostics (2019): Frequently Asked Questions. [online] [navštíveno 2.5. 2019]. Dostupné z: <https://www.accuref.com/faq>

Agilent Technologies (2019): SureSelectXT HS Reagent Kits. [online] [navštíveno 1.3. 2019] Dostupné z: [https://www.agilent.com/en/product/hybridization-based-next-generation-sequencing-\(ngs\)/dna-seq-reagents-kits-library-preparation-kits/sureselectxt-hs-reagent-kits-232857#productdetails](https://www.agilent.com/en/product/hybridization-based-next-generation-sequencing-(ngs)/dna-seq-reagents-kits-library-preparation-kits/sureselectxt-hs-reagent-kits-232857#productdetails)

Ambardar S., Gupta R., Trakroo D., Lal R., Vakhlu J. (2016): High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Medical Microbiology*: 56(4): 394-404.

Annala M., Vandekerkhove G., Khalaf D., Taavitsainen S., Beja K., Warner E. W., Sunderland K., Kollmannsberger C., Eigl B. J., Finch D. (2018): Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. *Cancer Discovery*: 8(4): 444-457.

Arteaga C. L. *et* Engelman J. A. (2014): ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell*: 25(3): 282-303.

BEDTools (2019): Intersect. [online] [navštíveno 19.4. 2019]. Dostupné z: <https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html>

Bennett C. W., Berchem G., Kim Y. J., El-Khoury V. (2016): Cell-free DNA and next-generation sequencing in the service of personalized medicine for lung cancer. *Oncotarget*: 7(43): 71013-71035.

Biomarkers Definitions Working Group (2001): Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*: 69(3): 89-95. In: Ziegler A., Koch A., Krockenberger K., Grosshennig A. (2012): Personalized medicine using DNA biomarkers: a review. *Human Genetics*: 131(10): 1627-38.

Bodi K., Perera A. G., Adams P. S., Bintzler D., Dewar K., Grove D. S., Kieleczawa J., Lyons R. H., Neubert T. A., Noll A. C. (2013): Comparison of Commercially Available Target Enrichment Methods for Next-Generation Sequencing. *Journal of biomolecular techniques*: 24(2): 73–86.

Bompas E., Le Cesne A., Tresch-Bruneel E., Lebellec L., Laurence V., Collard O., Saada-Bouzid E., Isambert N., Blay J. Y., Amela E. Y. *et al.* (2015): Sorafenib in patients with

locally advanced and metastatic chordomas: a phase II trial of the French Sarcoma Group (GSF/GETO). *Annals of Oncology*: 26(10): 2168-73.

Bonfiglio S., Vanni I., Rossella V., Truini A., Lazarevic D., Dal Bello M. G., Alama A., Mora M., Rijavec E., Genova C. *et al.* (2016): Performance comparison of two commercial human whole-exome capture systems on formalin-fixed paraffin-embedded lung adenocarcinoma samples. *BMC Cancer*: 16: 692.

Brar G., Marshall J. L., Pishvaian M. J. (2017): Targeted Therapy and the Use of Molecular Profiling in Metastatic Colorectal Cancer. *AJHO*: 13(9): 4-12.

Bulbul A. *et* Husain H. (2018): First-Line Treatment in EGFR Mutant Non-Small Cell Lung Cancer: Is There a Best Option? *Frontiers in Oncology*: 8:94.

Burghel G. J., Hurst C. D., Watson C. M., Chambers P. A., Dickinson H., Roberts P., Knowles M. A. (2015): Towards a Next-Generation Sequencing Diagnostic Service for Tumour Genotyping: A Comparison of Panels and Platforms. *BioMed Research International*: 2015: 478017.

Burrell R. A. *et* Swanton C. (2014): Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular Oncology*: 8(6): 1095-111.

Burrows-Wheeler Aligner (2010). [online] [navštíveno 2.4. 2019]. Dostupné z: <http://bio-bwa.sourceforge.net/>

Cancer Genome Interpreter (2019). [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.cancergenomeinterpreter.org/>

Casali P. G., Stacchiotti S., Sangalli C., Olmi P., Gronchi A. (2007): Chordoma. *Current Opinion in Oncology*: 19: 367–70. In: Meng T., Jin J., Jiang C., Huang R., Yin H., Song D., Cheng L. (2019): Molecular Targeted Therapy in the Treatment of Chordoma: A Systematic Review. *Frontiers in Oncology*: 9: 30.

CeGaT GmbH (2019): Next-Generation Sequencing. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.cegat.de/en/services/next-generation-sequencing/>

Cicenas J., Kvederaviciute K., Meskinyte I., Meskinyte-Kausiliene E., Skeberdyte A., Cicenas J. (2017): KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 Mutations in Pancreatic Cancer. *Cancers (Basel)*: 9(5). pii: E42.

Collins D. C., Sundar R., Lim J. S. J., Yap T. A. (2017): Towards Precision Medicine in the Clinic: From Biomarker Discovery to Novel Therapeutics. *Trends in Pharmacological Sciences*: 38(1): 25-40.

Coonrod E. M., Durtschi J. D., VanSant Webb C., Voelkerding K. V., Kumánovics A. (2014): Next-generation sequencing of custom amplicons to improve coverage of HaloPlex multigene panels. *Biotechniques*: 57(4): 204-7.

Cunanan K. M., Iasonos A., Shen R., Begg C. B., Gönen M. (2017): An efficient basket trial design. *Statistics in medical journals*: 36(10): 1568-1579.

De Felice F., Tombolini V., Marampon F., Musella A., Marchetti C. (2017): Defective DNA repair mechanisms in prostate cancer: impact of olaparib. *Drug Design, Development and Therapy*: 11: 547-552.

Diehl F., Schmidt K., Choti M. A., Romans K., Goodman S., Li M., Thornton K., Agrawal N., Sokoll L., Szabo S. A. *et al.* (2008): Circulating mutant DNA to assess tumor dynamics. *Nature Medicine*: 14(9): 985-90.

Dong L., Wang W., Li A., Kansal R., Chen Y., Chen H., Li X. (2015): Clinical Next Generation Sequencing for Precision Medicine in Cancer. *Current Genomics*: 16(4): 253-63.

Faraoni I. *et* Graziani G. (2018): Role of BRCA Mutations in Cancer Treatment with Poly(ADP-ribose) Polymerase (PARP) Inhibitors. *Cancers (Basel)*: 10(12). pii: E487.

Fausel C. (2007): Targeted chronic myeloid leukemia therapy: Seeking a cure. *American Journal of Health-System Pharmacy*: 64(24 Suppl 15): S9-15.

Finotti A., Allegretti M., Gasparello J., Giacomini P., Spandidos D. A., Spoto G., Gambari R. (2018): Liquid biopsy and PCR-free ultrasensitive detection systems in oncology (Review). *International Journal of Oncology*: 53(4): 1395-1434.

Fischer C., Scheipl S., Zopf A., Niklas N., Deutsch A, Jorgensen M., Lohberger B., Froehlich E. V., Leithner A. *et al.* (2015): Mutation Analysis of Nine Chordoma Specimens by Targeted Next-Generation Cancer Panel Sequencing. *Journal of Cancer*: 6(10):984-9.

Forbes S. A., Tang G., Bindal N., Bamford S., Dawson E., Cole C., Kok C. Y., Jia M., Ewing R., Menzies A., *et al.* (2010): COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Research*: 38(Database issue): D652-7.

Garinet S., Laurent-Puig P., Blons H., Oudart J. B. (2018): Current and Future Molecular Testing in NSCLC, What Can We Expect from New Sequencing Technologies? *Clinical Medicine*: 7(6). pii: E144.

Garrido-Cardenas J. A., Garcia-Maroto F., Alvarez-Bermejo J. A., Manzano-Agugliaro F. (2017): DNA Sequencing Sensors: An Overview. *Sensors (Basel)*: 17(3). pii: E588.

Gerlinger M., Rowan A. J., Horswell S., Math M., Larkin J., Endesfelder D., Gronroos E., Martinez P., Matthews N., Stewart A., *et al.* (2012): Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine*: 366(10): 883-892.

GitHub (2019a): SingleStrandUmiSomaticVariantCalling.scala. [online] [navštíveno 2.4. 2019]. Dostupné z: <https://github.com/nh13/bfx->

examples/blob/master/pipelines/src/main/scala/com/nilshomer/bfx/examples/pipelines/SingleStrandUmiSomaticVariantCalling.scala#L118

GitHub (2019b): Picard. [online] [navštíveno 2.4. 2019]. Dostupné z: <https://broadinstitute.github.io/picard/>

GitHub (2019c): fgbio [online] [navštíveno 2.4. 2019]. Dostupné z: <http://fulcrumgenomics.github.io/fgbio/>

GitHub (2019d): VarDictJava. [online] [navštíveno 2.4. 2019]. Dostupné z: <https://github.com/AstraZeneca-NGS/VarDictJava>

Goetz L. H. *et* Schork N. J. (2018): Personalized medicine: motivation, challenges, and progress. *Fertility and Sterility*: 109(6): 952-963.

Golan T., Varadhachary G. R., Sela T., Fogelman D. R., Halperin N., Shroff R. T., Halparin S., Xiao L., Dan Aderka D., Anirban Maitra A. *et al.* (2018): Phase II study of olaparib for BRCAness phenotype in pancreatic cancer. *Journal of Clinical Oncology*: 36: 4_suppl, 297-297

Grada A. *et* Weinbrecht K. (2013): Next-generation sequencing: methodology and application. *Journal of Investigative Dermatology*: 133(8): e11.

Gray P. N., Dunlop C. L., Elliott A. M. (2015): Not All Next Generation Sequencing Diagnostics are Created Equal: Understanding the Nuances of Solid Tumor Assay Design for Somatic Mutation Detection. *Cancers (Basel)*: 7(3): 1313-32.

Greaves M. *et* Maley C. C. (2012): Clonal evolution in cancer. *Nature*: 481(7381): 306-13.

Guo X. E., Ngo B., Modrek A. S., Lee W. H. (2014): Targeting tumor suppressor networks for cancer therapeutics. *Current Drug Targets*: 15(1): 2-16.

Haigis K. V. (2017): KRAS Alleles: The Devil Is In The Detail. *Trends in Cancer*: 3(10): 686-697.

Han X., Wang J., Sun Y. (2017): Circulating Tumor DNA as Biomarkers for Cancer Detection. *Genomics Proteomics Bioinformatics*: 15(2): 59-72.

Heeke S., Hofman V., Long-Mira E., Lespinet V., Lalvée S., Bordone O., Ribeyre C., Tanga V., Benzaquen J., Leroy S. *et al.* (2018): Use of the Ion PGM and the GeneReader NGS Systems in Daily Routine Practice for Advanced Lung Adenocarcinoma Patients: A Practical Point of View Reporting a Comparative Study and Assessment of 90 Patients. *Cancers (Basel)*: 10(4). pii: E88.

Heery C. R. (2016): Chordoma: The Quest for Better Treatment Options. *Oncology and Therapy*: 4(1):35-51.

Hindi N., Casali P. G., Morosi C., Messina A., Palassini E., Pilotti S., Tamborini E., Radaelli S., Gronchi A., Stacchiotti S. (2015): Imatinib in advanced chordoma: A retrospective case series analysis. *Eur J Cancer*: 51(17): 2609-14.

Hirsch B., Endris V., Lassmann S., Weichert W., Pfarr N., Schirmacher P., Kovaleva V., Werner M., Bonzheim I., Fend F. *et al.* (2018): Multicenter validation of cancer gene panel-based next-generation sequencing for translational research and molecular diagnostics. *Virchows Archiv.*: 472(4): 557-565.

Hoadley K. A., Yau C., Wolf D. M., Cherniack A. D., Tamborero D., Ng S., Leiserson M. D. M., Niu B., McLellan M. D., Uzunangelov V. *et al.* (2014): Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*: 158(4): 929-944.

Horizon Discovery (2019): EGFR Multiplex DNA - Exome Data. [online] [navštíveno 10. 4. 2019]. Dostupné z: https://www.horizondiscovery.com/resources/support-materials/data?d_cat=reference-standards&p=2#pager

Huggett J. F. *et* Whale A. (2013): Digital PCR as a novel technology and its potential implications for molecular diagnostics. *Clinical Chemistry*: 59(12): 1691-3.

Hyman D. M., Taylor B. S., Baselga J. (2017): Implementing Genome-Driven Oncology. *Cell*: 68(4): 584–599.

Chen F., Dong M., Ge M., Zhu L., Ren L., Liu G., Mu R. (2013): The history and advances of reversible terminators used in new generations of sequencing technology. *Genomics Proteomics Bioinformatics*: 11(1):34-40.

Illumina (2019): Pan-cancer research with molecular profiling. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/areas-of-interest/cancer/clinical-cancer-research/somatic-mutations/pan-cancer.html?langsel=/cz/>

Illumina (2019b): AmpliSeq for Illumina Comprehensive Cancer Panel. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/ampliseq-comprehensive-cancer-panel.html>

Illumina (2019c): TruSight Oncology UMI Reagents. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/products/by-type/clinical-research-products/trusight-oncology-umi.html>

Illumina (2019d): Using TruSight Oncology UMI Reagents with TruSight Tumor 170 DNA Content for Detection of Low-Frequency Variants in Cell-Free DNA. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/content/dam/illumina-marketing/documents/products/technotes/trusight-umi-trusight-tumor-170-app-note-1000000050427.pdf>

Illumina (2019e): AmpliSeq for Illumina Cancer Hotspot Panel v2. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/ampliseq-cancer-hotspot-panel.html>

Illumina (2019f): TruSight Tumor 170. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/products/by-type/clinical-research-products/trusight-tumor-170.html>

Illumina (2019g): Error Rate Reduction with the TruSight™ Oncology UMI Reagents for Variant Detection. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/trusight-umi-rare-variant-technote-1000000050426.pdf>

Illumina (2019h): An introduction to Next-Generation Sequencing Technology. [online] [navštíveno 1.3. 2019] Dostupné z: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

Illumina (2019i): Illumina sequencing platforms. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/systems/sequencing-platforms.html>

Illumina (2019j): MiSeq Reporter Software Guide. [online] [navštíveno 1.3. 2019] Dostupné z: https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/miseqreporter/miseq-reporter-software-guide-15042295-05.pdf

Illumina (2019k): Specifications for the MiSeq System. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/systems/sequencing-platforms/miseq/specifications.html?langsel=/cz/>

Illumina (2019l): Sequencing Data Analysis Solutions. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/informatics/sequencing-data-analysis.html>

Illumina (2019m): VariantStudio Software Highlights. [online] [navštíveno 1.3. 2019] Dostupné z: <https://emea.illumina.com/products/by-type/informatics-products/variantstudio.html?langsel=/cz/>

Ivány G., Madar L., Dzsudzsák E., Koczok K., Kappelmayer J., Krulisova V., Macek M. Jr., Horváth A., Balogh I. (2018): Analytical parameters and validation of homopolymer detection in a pyrosequencing-based next generation sequencing system. *BMC Genomics*: 19(1): 158.

Jain M., Olsen H. E., Paten B., Akeson M. (2016): The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*: 17(1): 239.

Janku F., Angenendt P., Tsimberidou A. M., Fu S., Naing A., Falchook G. S., Hong D. S., Holley V. R., Cabrilo G., Wheler J. J. *et al.* (2015): Actionable mutations in plasma cell-free DNA in patients with advanced cancers referred for experimental targeted therapies. *Oncotarget* 6(14):12809-21.

Jennings L. J., Arcila M. E., Corless C., Kamel-Reid S., Lubin I. M., Pfeifer J., Temple-Smolkin R. L., Voelkerding K. V., Nikiforova M. N. (2017): Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics*: 19(3): 341-365.

Johnson A., Zeng J., Bailey A. M., Holla V., Litzenburger B., Lara-Guerra H., Mills G. B., Mendelsohn J., Shaw K. R., Meric-Bernstam F. (2015): The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. *Drug Discov Today*: 20(12): 1433-8.

Joo W. D., Visintin I., Mor G. (2013): Targeted cancer therapy--are the days of systemic chemotherapy numbered? *Maturitas*: 76(4): 308-14.

Kamps R., Brandão R. D., Bosch B. J., Paulussen A. D., Xanthoulea S., Blok M. J., Romano A. (2017): Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *International Journal of Molecular Sciences*: 18(2). pii: E308.

Kandoth C., McLellan M. D., Vandin F., Ye K., Niu B., Lu C., Xie M., Zhang Q., McMichael J. F., Wyczalkowski M. A. *et al.* (2013): Mutational landscape and significance across 12 major cancer types. *Nature*: 502: 333–339.

Kleibl Z. (2016): Sekvenování nové generace - nová generace diagnostiky dědičných nádorových syndromů. *Labor Aktuell*: 2016(2): 21-24.

Koubková L., Vojtěšek B., Vyzula R. (2014): Sekvenování nové generace a možnosti jeho využití v onkologické praxi. *Klinická onkologie*: 27 (Suppl 1): S61–S68.

Kuo F. C., Mar B. G., Lindsley R. C., Lindeman N. I. (2017): The relative utilities of genome-wide, gene panel, and individual gene sequencing in clinical practice. *Blood*: 130(4): 433-439.

Lenk C. *et* Duttge G. (2014): Ethical and legal framework and regulation for off-label use: European perspective. *Therapeutics and Clinical Risk Management*: 10: 537-46.

Li Y., Lea K., Kshatriya P., Cao R., Gu J., Schageman J., Bagai V., Hanif K., Bramlett K. S. (2018): PO-086 An efficient ion torrent™ next generation sequencing workflow for liquid biopsy research to assess cell-free total nucleic acid. *ESMO Open* 3(Suppl 2): A1–A463.

Liang W. S., Millis S. Z., Gatalica Z., Reddy S. K., Little A., Van Tine B. A. (2015): Identification of actionable targets in chordomas using a multiplatform molecular analysis,

and response with targeted therapy. *Journal of Clinical Oncology*: 33 (suppl; abstract e22215).

Liao P. *et* Huang Y. (2017): Digital PCR: Endless Frontier of 'Divide and Conquer'. *Micromachines* (Basel): 8(8). pii: E231.

Lin J. J. *et* Shaw A. T. (2016): Resisting Resistance: Targeted Therapies in Lung Cancer. *Trends in Cancer*: 2(7): 350-364.

Lord C. J. *et* Ashworth A. (2017): PARP inhibitors: Synthetic lethality in the clinic. *Science*: 355(6330): 1152-1158.

Lynch T. J., Bell D. W., Sordella R., Gurubhagavatula S., Okimoto R. A., Brannigan B. W., Harris P. L., Haserlat S. M., Supko J. G., Haluska F. G. *et al.* (2004): Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *The New England Journal of Medicine*: 350(21): 2129-39.

Machackova E., Foretova L., Lukesova M., Vasickova P., Navratilova M., Coene I., Pavlu H., Kosinova V., Kuklova J., Claes K. (2008): Spectrum and characterisation of BRCA1 and BRCA2 deleterious mutations in high-risk Czech patients with breast and/or ovarian cancer. *BMC Cancer*: 8: 140.

Mamanova L., Coffey A. J., Scott C. E., Kozarewa I., Turner E. H., Kumar A., Howard E, Shendure J., Turner D. J. (2010): Target-enrichment strategies for next-generation sequencing. *Nature Methods*: 7(2): 111-8.

Matheson C. J., Backos D. S., Reigan P. (2016): Targeting WEE1 Kinase in Cancer. *Trends in Pharmacological Sciences*: 37(10): 872-881.

McMaster M. L., Goldstein A. M., Bromley C. M., Ishibe N., Parry D. M. (2001): Chordoma: incidence and survival patterns in the United States, 1973-1995. *Cancer Causes & Control*: 12(1): 1-11.

Mehrotra M., Duose D. Y., Singh R. R., Barkoh B. A., Manekia J., Harmon M. A., Patel K. P., Routbort M. J., Medeiros L. J., Wistuba II *et al.* (2017): Versatile ion S5XL sequencer for targeted next generation sequencing of solid tumors in a clinical laboratory. *PLoS One*: 12(8): e0181968.

Mehta A., Vasudevan S., Sharma S. K., Kumar D., Panigrahi M., Suryavanshi M., Gupta G. (2018): Germline BRCA1 and BRCA2 deleterious mutations and variants of unknown clinical significance associated with breast/ovarian cancer: a report from North India. *Cancer Management and Research*: 10: 6505-6516.

Meléndez B., Van Campenhout C., Rorive S., Rummelink M., Salmon I., D'Haene N. (2018): Methods of measurement for tumor mutational burden in tumor tissue. *Translational Lung Cancer Research*: 7(6):661-667.

Meng T., Jin J., Jiang C., Huang R., Yin H., Song D., Cheng L. (2019): Molecular Targeted Therapy in the Treatment of Chordoma: A Systematic Review. *Frontiers in Oncology*: 9:30.

Minárik M. (2016): Molekulární profilování nádorů plic. *Onkologie*: 10(1): 7-10.

Morash M., Mitchell H., Beltran H., Elemento O., Pathak J. (2018): The Role of Next-Generation Sequencing in Precision Medicine: A Review of Outcomes in Oncology. *Journal of Personalized Medicine*: 8(3). pii: E30.

Moschetta M., George A., Kaye S. B., Banerjee S. (2016): BRCA somatic mutations and epigenetic BRCA modifications in serous ovarian cancer. *Annals of Oncology*: 27(8): 1449-55

National Cancer Institute (2019): The Cancer Genome Atlas Program. [online] [navštíveno 2.4. 2019]. Dostupné z: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>

Neel D. S. *et* Bivona T. G. (2017): Resistance is futile: overcoming resistance to targeted therapies in lung adenocarcinoma. *NPJ Precision Oncology*: 1: 3.

New England Biolabs (2019a): NEBNext Direct Cancer HotSpot Panel Instruction Manual. [online] [navštíveno 2.4. 2019]. <https://www.neb.com/-/media/catalog/datacards-or-manuals/manuale7000.pdf?rev=dca60d68fb0548bfb0de161e23eed4f6>

New England Biolabs (2019b): NEBNext Direct Cancer HotSpot Panel. [online] [navštíveno 2.4. 2019]. Dostupné z: <https://www.neb.com/products/e7000-nebnext-direct-cancer-hotspot-panel#Product%20Information>

New England Biolabs (2019c): NEBNext Direct BRCA1/BRCA2 Panel. [online] [navštíveno 2.4. 2019]. <https://www.neb.com/products/e6627-nebnext-direct-brca1-brca2-panel#Product%20Information>

Nicolas E., Bertucci F., Sabatier R., Gonçalves A. (2018): Targeting BRCA Deficiency in Breast Cancer: What are the Clinical Evidences and the Next Perspectives? *Cancers (Basel)*: 10(12). pii: E506.

Novelli G., Ciccacci C., Borgiani P., Papaluca Amati M., Abadie E. (2008): Genetic tests and genomic biomarkers: regulation, qualification and validation. *Clinical Cases in mineral and bone metabolism*: 5(2): 149-54.

Ou S. H., Kwak E. L., Siwak-Tapp C., Dy J., Bergethon K., Clark J. W., Camidge D. R., Solomon B. J., Maki R. G., Bang Y. J. *et al.* (2011): Activity of crizotinib (PF02341066), a dual mesenchymal-epithelial transition (MET) and anaplastic lymphoma kinase (ALK) inhibitor, in a non-small cell lung cancer patient with de novo MET amplification. *Journal of Thoracic Oncology*: (5):942-6.

Pacific Biosciences (2019): High-quality, long reads produce exceptional results. [online] [navštíveno 2.5. 2019]. Dostupné z: <https://www.pacb.com/smrt-science/smrt-sequencing/read-lengths/>

Page K., Guttery D. S., Fernandez-Garcia D., Hills A., Hastings R. K., Luo J., Goddard K., Shahin V., Woodley-Barker L., Rosales B. M. *et al.* (2017): Next Generation Sequencing of Circulating Cell-Free DNA for Evaluating Mutations and Gene Amplification in Metastatic Breast Cancer. *Clinical Chemistry*: 63(2):532-541.

Parker B. C. *et* Zhang W. (2013): Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chinese Journal of Cancer*: 32(11): 594-603.

Perazzona B., Lin H., Sun S., Wang Y., Arlinghaus R. (2008): Kinase domain mutants of Bcr enhance Bcr-Abl oncogenic effects. *Oncogene*: 27(15): 2208-2214.

Prahallad A., Sun C., Huang S., Di Nicolantonio F., Salazar R., Zecchin D., Beijersbergen R. L., Bardelli A., Bernards R. (2012): Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature*: 483(7387): 100-3.

Qiagen (2019a): QIAseq Targeted DNA Panels. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/qiaseq-targeted-dna-panels/?catno=DHS-6600Z#orderinginformation>

Qiagen (2019b): QIAseq™ Targeted DNA Panel Handbook. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.qiagen.com/us/resources/download.aspx?id=8907edbe-a462-4883-ae1b-2759657e7fd0&lang=en>

Rang F. J., Kloosterman W. P. , de Ridder J. (2018): From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*: 19(1): 90

Reuter J. A., Spacek D. V., Snyder M. P. (2015): High-throughput sequencing technologies. *Molecular Cell*: 58(4): 586-97.

Rhoads A. *et* Au K. F. (2015): PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*: 13(5): 278-89.

Roychowdhury S. *et* Chinnaiyan A. M. (2016): Translating cancer genomes and transcriptomes for precision oncology. *CA: A Cancer Journal for Clinicians*: 66(1): 75-88.

Samorodnitsky E., Datta J., Jewell B. M., Hagopian R., Miya J., Wing M. R., Damodaran S., Lippus J. M., Reeser J. W., Bhatt D., Timmers C. D., Roychowdhury S. (2015): Comparison of custom capture for targeted next-generation DNA sequencing. *The Journal of Molecular Diagnostics*: 17(1): 64-75.

SAMtools (2012). [online] [navštíveno 2.4. 2019]. Dostupné z: <http://samtools.sourceforge.net/>

Sena J. A., Galotto G., Devitt N. P., Connick M. C., Jacobi J. L., Umale P. E., Vidali L., Bell C. J. (2018): Unique Molecular Identifiers reveal a novel sequencing artefact with implications for RNA-Seq based gene expression analysis. *Scientific Reports*: 8(1): 13121.

Sena J. A., Galotto G., Devitt N. P., Connick M. C., Jacobi J. L., Umale P. E., Vidali L., Bell C. J. (2018): Unique Molecular Identifiers reveal a novel sequencing artefact with implications for RNA-Seq based gene expression analysis. *Scientific Reports*: 8(1): 13121.

Singh R. R., Patel K. P., Routbort M. J., Aldape K., Lu X., Manekia J., Abraham R., Reddy N. G., Barkoh B. A., Veliyathu J. *et al.* (2014): Clinical massively parallel next-generation sequencing analysis of 409 cancer-related genes for mutations and copy number variations in solid tumours. *British Journal of Cancer*: 111(10): 2014-23.

Sishe B. J. *et* Davis A. J. (2017): The Role of the Core Non-Homologous End Joining Factors in Carcinogenesis and Cancer. *Cancers (Basel)*: 9(7). pii: E81.

Smith A. M., Heisler L. E., St Onge R. P., Farias-Hesson E., Wallace I. M., Bodeau J., Harris A. N., Perry K. M., Giaever G., Pourmand N. *et al.* (2010): Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Research*: 38(13): e142.

Stacchiotti S., Longhi A., Ferraresi V., Grignani G., Comandone A., Stupp R, Bertuzzi A., Tamborini E., Pilotti S., Messina A. *et al.* (2012): Phase II study of imatinib in advanced chordoma. *Journal of Clinical Oncology*: 30(9): 914-20.

Sun J., Wei Q., Zhou Y., Wang J., Liu Q., Xu H. (2017): A systematic analysis of FDA-approved anticancer drugs. *BMC Systems Biology*: 11(Suppl 5): 87.

Thermo Fisher Scientific (2019a): PCR Methods—Top Ten Strategies. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.thermofisher.com/cz/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-methods.html>

Thermo Fisher Scientific (2019b): OncoPrint™ Pan-Cancer Cell-Free Assay. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.thermofisher.com/order/catalog/product/A37664>

Thermo Fisher Scientific (2019c): Ion AmpliSeq™ Comprehensive Cancer Panel. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.thermofisher.com/order/catalog/product/4477685>

Thermo Fisher Scientific (2019d): Meet the New Ion GeneStudio S5 Systems. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.thermofisher.com/blog/behindthebench/meet-the-new-ion-genestudio-s5-systems/>

Thermo Fisher Scientific (2019e): Ion GeneStudio S5 Next-Generation Sequencing Series Specifications. [online] [navštíveno 1.3. 2019] Dostupné z: <https://www.thermofisher.com/cz/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-s5-ngs-targeted-sequencing/ion-s5-specifications.html>

Tovar E. A. *et* Graveel C. R. (2017): MET in human cancer: germline and somatic mutations. *Annals of Translational Medicine*: 5(10):205.

Tsimberidou A. M. (2017): Initiative for Molecular Profiling and Advanced Cancer Therapy and challenges in the implementation of precision medicine. *Current Problems in Cancer*: 41(3): 176-181.

U.S. Food and Drug Administration (2019). FDA Grants Accelerated Approval to Pembrolizumab for First Tissue/Site Agnostic Indication. [online] [navštíveno 20.2. 2019] Dostupné z: <https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm560040.htm>

Uzilov A. V., Ding W., Fink M. Y., Antipin Y., Brohl A. S., Davis C., Lau C. Y., Pandya C., Shah H., Kasai Y. *et al.* (2016): Development and clinical application of an integrative genomic approach to personalized cancer therapy. *Genome Medicine*: 8(1): 62.

VarAFT (2019): Variant Annotation and Filter Tool. [online] [navštíveno 1.3. 2019] Dostupné z: <https://varaft.eu/>

Vollbrecht C., Lehmann A., Lenze D., Hummel M. (2018): Validation and comparison of two NGS assays for the detection of EGFR T790M resistance mutation in liquid biopsies of NSCLC patients. *Oncotarget*: 9(26): 18529-18539.

Walcott B. P., Nahed B. V., Mohyeldin A., Coumans J. V., Kahle K. T., Ferreira M. J. (2012): Chordoma: current concepts, management, and future directions. *The Lancet Oncology*: 13(2): e69-76.

Wang K. *et* Xu C. (2017): Applications of Next-Generation Sequencing in Cancer Research and Molecular Diagnosis. *Journal of Clinical and Medical Genomics*: 5: 1.

Weinberger P. M., Yu Z., Kowalski D., Joe J., Manger P., Psyrri A., Sasaki C. T. (2005): Differential expression of epidermal growth factor receptor, c-Met, and HER2/neu in chordoma compared with 17 other malignancies. *Archives of Otolaryngology–Head & Neck Surgery*: 131(8): 707-11.

Wong S. Q., Li J., Tan A. Y., Vedururu R., Pang J. M., Do H., Ellul J., Doig K., Bell A., MacArthur G. A. *et al.* (2014): Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Medical Genomics*: 7:23.

Xuan J., Yu Y., Qing T., Guo L., Shi L. (2013): Next-generation sequencing in the clinic: promises and challenges. *Cancer Letters*: 340(2): 284-95.

Zhang H., Liu R., Yan C., Liu L., Tong Z., Jiang W., Yao M., Fang W., Chen Z. (2019): Advantage of Next-Generation Sequencing in Dynamic Monitoring of Circulating Tumor DNA over Droplet Digital PCR in Cetuximab Treated Colorectal Cancer Patients. *Translational Oncology*: 12(3): 426–431.

Ziegler A., Koch A., Krockenberger K., Grosshennig A. (2012): Personalized medicine using DNA biomarkers: a review. *Human Genetics*: 131(10): 1627-38.