

Univerzita Karlova v Praze

Farmaceutická fakulta v Hradci Králové

Studijní program: Zdravotnická bioanalýtika

Katedra biochemických věd



**PŘÍNOS NEXT GENERATION SEQUENCING
PRO LABORATORNÍ DIAGNOSTIKU**

Diplomová práce

Vedoucí práce: Doc. PharmDr. Martin Beránek, Ph.D.

Hradec Králové 2015

Pavel Votýpka

Čestné prohlášení

„Prohlašuji, že tato práce je mým původním autorským dílem. Veškerá literatura a další zdroje, z nichž jsem při zpracování čerpal, jsou uvedeny v seznamu použité literatury a v práci řádně citovány. Práce nebyla využita k získání jiného nebo stejného titulu.“

V Hradci Králové, 14. května 2015

.....

Pavel Votýpka

Poděkování

Na tomto místě bych rád poděkoval především přednostovi ÚBLG 2. LF UK a FN Motol panu Prof. MUDr. Milanu Mackovi, DrSc. za umožnění studia Farmaceutické fakulty při práci a celému kolektivu ÚBLG za trpělivost při mém studiu a vypracování této diplomové práce. Dále bych rád poděkoval vedoucímu práce Doc. PharmDr. Martinu Beránkovi, Ph.D. a konzultantce Mgr. Nikole Ptákové za cenné rady, připomínky a předané zkušenosti při sepisování práce.

ABSTRAKT

Univerzita Karlova v Praze
Farmaceutická fakulta v Hradci Králové
Katedra biochemických věd

Kandidát: Bc. Pavel Votýpka

Školitel: Doc. PharmDr. Martin Beránek, Ph.D.

Konzultant: Mgr. Nikola Ptáková

Název diplomové práce: Přínos Next Generation Sequencing pro laboratorní diagnostiku

Snaha o přečtení lidského genomu ve svém důsledku vedla nejen k získání poznatků o lidské genetické informaci, ale také k vývoji nových sekvenačních metod a technologií. Pro udržení kroku s rozvojem v oblasti genetiky se na řadě klinických i výzkumných pracovišť do běžného provozu zavádějí přístroje druhé generace na principu masivního paralelního sekvenování. Na trhu dnes dominují čtyři platformy – Illumina, Solid, Ion Torrent a 454 Life Technologies. Proces celé sekvenační analýzy lze shrnout do tří hlavních kroků – příprava sekvenační knihovny, sekvenování a detekce nukleotidů a analýza dat. Každá část analýzy přináší v závislosti na použité platformě svá specifika, se kterými je nutné počítat a také úskalí, kterým je vhodné se vyhnout či minimalizovat jejich vliv na finální výsledek analýzy. V posledních několika letech jsou vyvíjeny i nové metody označované jako třetí generace sekvenování, které umožňují čtení jediné molekuly nukleové kyseliny bez její amplifikace. NGS technologie dnes umožňují, aby si každé pracoviště zvolilo svůj přístup k sekvenování, přičemž si laboratoř zvolí, zda ke svému využití potřebuje data získaná celogenomovým sekvenováním, sekvenováním exomu, či bude testovat určitou skupinu genů v tzv. panelu.

V praktické části práce bylo popsáno zavedení metody sekvenování nové generace na klinickém genetickém pracovišti a optimalizace potřebných kroků. Ve spolupráci s kardiologickou klinikou FN Motol a IKEM byl vybrán zkušební soubor 191 pacientů s diagnózou hereditární kardiomyopatie pro detekci zárodečných mutací. U 65 z nich byl použit panel TruSight Cardiomyopathy, který dokázal odhalit kauzální mutaci u 48 pacientů. U dalších 126 vzorků ze 42 rodin byl vyšetřen klinický exom kitem TruSight One, který odhalil kauzální mutace u 27 rodin. Účinnost NGS metod byla také otestována na skupině 8 pacientů z 3 rodin se závažnou genetickou zátěží, u nichž byl vyšetřen celý exom o délce 37 Mb.

ABSTRACT

Charles University in Prague
Faculty of Pharmacy in Hradec Králové
Department of Biochemical Sciences

Candidate: Bc. Pavel Votýpka

Supervisor: Doc. PharmDr. Martin Beránek, Ph.D.

Consultant: Mgr. Nikola Ptáková

Title of diploma thesis: Contribution of Next Generation Sequencing for Laboratory Diagnostics

The endeavor to sequence the whole human genome lead not only to the knowledge acquisition regarding the human genetic information but as well to the development of new sequencing methods and technologies. In order to keep up with progress in genetic field in many clinical and research laboratories the new massive parallel sequencing equipment is being utilized. On the market are currently established four leading platforms – Illumina, Solid, Ion Torrent and 454 Life Technologies. The process of sequencing analysis can be summarized into three main steps – the sequencing library preparation, sequencing itself, variant calling and data analysis. Each part of the sequencing analysis exhibits certain specifics, we need to count with and as well its pitfalls, we need to avoid or to minimize their impact on the analysis final result. Recently new methods termed sequencing of the 3rd generation are being developed, enabling sequence of a single DNA molecule to be determined without previous amplification. NGS technologies provide those days laboratories with ability to choose their own preferences in sequencing, where the laboratory can choose whether it needs data obtained by whole genome sequencing, exome sequencing or whether to test certain group of genes in so called panel.

In the practical part of this study is described the establishment of next generation sequencing in the diagnostic laboratory and optimization of required processes. The trial group of 191 patients with diagnosis of hereditary cardiomyopathy was selected in collaboration with cardiology clinic of IKEM and University Hospital Motol in order to detect germinal mutations. In 65 patients the TruSight Cardiomyopathy panel was utilized, and causal mutation was uncovered in 48 patients. In remaining 126 samples comprising 42 families the clinical exome was sequenced using TruSight One kit and revealing causal mutation in 27 more families. The usage of NGS methods was tested as well in group of 8 patients from 3 families with serious genetic condition, where expanded exome probes covering 37 Mb were employed.

Obsah

1. ÚVOD	8
2. ZADÁNÍ DIPLOMOVÉ PRÁCE – CÍL PRÁCE	9
3. HISTORIE SEKVENOVÁNÍ A LIDSKÝ GENOM	10
3.1 Cesta k lidskému genomu.....	10
3.2 Výsledky a přínos závodu o lidský genom.....	12
3.3 Variabilita lidského genomu.....	12
4. SEKVENOVÁNÍ NOVÉ GENERACE	14
4.1 Příprava sekvenační knihovny.....	14
4.1.1 Fragmentace genomické DNA	15
4.1.2 Úprava DNA fragmentů před amplifikací	16
4.1.3 Amplifikace DNA fragmentů	17
4.2 Sekvenování a detekce nukleotidů	18
4.2.1 Platforma Illumina (sekvenace syntézou).....	19
4.2.2 Platforma 454 Life Sciences (pyrosekvenování).....	20
4.2.3 Platforma Life Technologies Solid (sekvenace ligací).....	21
4.2.4 Platforma Life Technologies Ion Torrent (syntéza s chemickou detekcí).....	22
4.2.5 Ostatní platformy	24
4.3 Přístupy k sekvenování.....	26
4.3.1 Celogenomové sekvenování	26
4.3.2 Exomové sekvenování.....	26
4.3.3 Cílené sekvenování (targeted sequencing)	27
4.3.4 Transkriptomové sekvenování.....	27
4.4 Porovnání jednotlivých NGS platforem	28
4.4.1 Chybovost NGS platforem	28
4.4.2 Délka sekvenovaných fragmentů	29
5. ANALÝZA SEKVENAČNÍCH DAT	31
5.1 Primární analýza – převod signálu na data.....	31
5.2 Sekundární analýza – filtrování a mapování readů	33
5.3 Terciální analýza – spojení genotypu s fenotypem	36
6. APLIKACE NGS V KLINICKÉ PRAXI.....	38
6.1 Nádorová NGS diagnostika.....	39
6.1.1 Kolorektální karcinom (OMIM #114500).....	41
6.1.2 Nádory prsu a ovária (OMIM #114480, #167000).....	42
6.2 Hereditární kardiomyopatie.....	43

6.2.1 Hypertrofická kardiomyopatie (HCM, OMIM #192600).....	43
6.2.2 Dilatační kardiomyopatie (DCM, OMIM #115200)	44
6.2.3 Arytmogenní kardiomyopatie pravé komory (ARVD, OMIM #609040)	44
6.3 Preimplantační genetická diagnostika (PGD-NGS)	45
6.4 Cystická fibróza (CF, OMIM #217900).....	46
7. ETICKÉ OTÁZKY SPOJENÉ S NGS METODAMI.....	48
8. PRAKTICKÁ ČÁST	50
8.1 Materiál a pomůcky	50
8.2 Příprava experimentů	51
8.3. Izolace a ředění genomické DNA.....	52
8.3.1 Práce s přístrojem Qubit 2.0	53
8.4 Příprava DNA knihovny.....	54
8.4.1. Fragmentace genomické DNA a přečištění fragmentů.....	54
8.4.2 První PCR amplifikace a přečištění amplikonů.....	57
8.4.3 První hybridizace a obohacení.....	58
8.4.4 Druhá hybridizace a obohacení	60
8.4.5 Druhá PCR reakce	61
8.5 Kvantifikace, ředění a denaturace finální DNA knihovny	62
8.6 Sekvenace na přístroji HiSeq 1500.....	64
8.7 Sekvenace na přístroji MiSeq	65
9. VÝSLEDKY	66
9.1 Rozdělení analyzovaných vzorků.....	66
9.2 Optimalizace potřebných kroků	66
9.3 Kvalita sekvenačních dat.....	67
9.4 Výsledky sekvenování.....	68
9.5 Zpracování dat ze sekvenování panelu.....	70
9.6 Zpracování dat z exomového sekvenování.....	72
10. DISKUZE.....	75
11. ZÁVĚR.....	79
12. SEZNAM POUŽITÝCH ZKRATEK	80
13. POUŽITÁ LITERATURA A PRAMENY	82
14. PŘÍLOHY.....	88
Příloha 1: Nalezené předpokládané kauzální mutace na přístroji MiSeq	88

1. ÚVOD

Deoxyribonukleová kyselina byla poprvé izolována již v roce 1869 Friedrichem Miescherem, který již tehdy tušil, že by mohla hrát v dědičnosti důležitou roli (1). Na objevení její struktury si však lidstvo muselo počkat dlouhých 84 let. V roce 1953 publikovali James D. Watson a Francis Crick v časopise Nature dvoušroubovicový model molekuly DNA (2). Krátce po tomto objevu začaly první snahy o její čtení, které však byly úspěšné až v 70. letech 20. století. Nejprve se jednalo pouze o čtení v délce několika desítek bp a teprve v roce 1977 byly objeveny nezávisle na sobě dvě skutečné sekvenační metody. Jedna z nich – Sangerova metoda byla postupem času modernizována a je v řadě molekulárně genetických laboratoří používána dodnes (3). Snaha o přečtení lidského genomu však ukázala, že Sangerova metoda dosáhla svých kapacitních možností a je nutné začít vyvíjet metody nové. První lidský genom byl kompletně přečten v roce 2001, tedy po 11 letech od zahájení projektu Human Genom Project (HGP). K osekvenování lidského genomu bylo tehdy potřeba asi 300 sekvenačních přístrojů a celkové náklady překročily hranici 3 miliard amerických dolarů. Snaha rozluštit lidský genetický kód velmi pomohla vývoji nových metod a technologií. Aby totiž bylo možné využít získané znalosti v oblasti genetiky a implementovat je v diagnostice a léčbě pacientů, bylo nutné zaměřit se na vývoj levnějších a rychlejších metod sekvenace, než byly do té doby k dispozici. K tomuto účelu byla do dnešního dne otestována řada postupů a principů, z nichž některé byly použity při výrobě sekvenátorů nové generace. Ruku v ruce s vývojem nových a lepších sekvenačních přístrojů jde i vývoj výpočetní techniky a softwaru na zpracování získaných dat (4, 10).

Klasické Sangerovo sekvenování pomalu ale jistě nahrazuje sekvenování nové generace (Next Generation Sequencing – zkráceně NGS) a to především při tzv. cíleném sekvenování, kdy je sekvenován jeden gen či skupina vybraných genů. Neuvěřitelný vědecký a technologický pokrok však stále nepřináší očekávané výsledky v podobě přesných diagnóz a možností léčby zásahem do genetické výbavy člověka, jak se na počátku předpokládalo. Snižování ceny nových sekvenačních přístrojů však vede k jejich rozšiřování do většiny genetických laboratoří v Evropě i v celém světě. Velké množství získaných dat však klade velké nároky na kvalitu a rozvoj oboru bioinformatiky, jelikož pouze kvalitní zhodnocení získaných dat, která jsou posuzována společně klinickým genetikem i bioinformatikem, jsou podmínkou správného genetického vyšetření pomocí NGS (5).

2. ZADÁNÍ DIPLOMOVÉ PRÁCE – CÍL PRÁCE

1. Vytvoření přehledu dostupných NGS platforem, sekvenačních principů a přístupů k sekvenování nové generace, porovnání jednotlivých sekvenačních technologií
2. Zavedení a optimalizace metod sekvenování nové generace na klinickém genetickém pracovišti a ukázka účinnosti technologie na vybrané skupině vzorků

3. HISTORIE SEKVENOVÁNÍ A LIDSKÝ GENOM

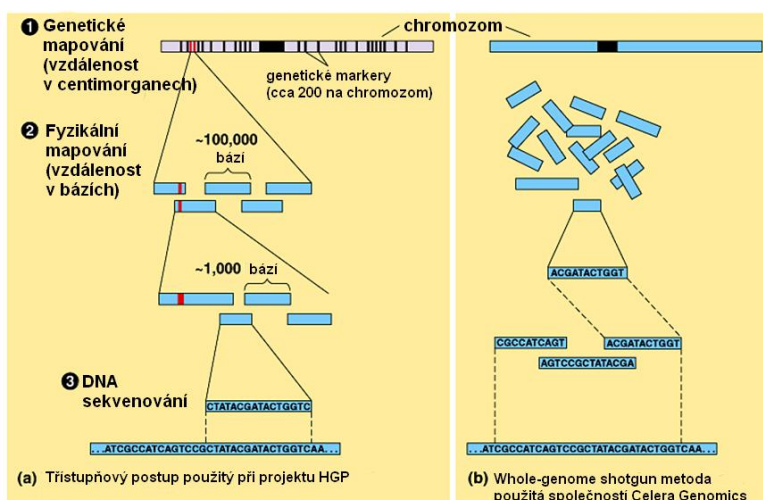
Snaha přečíst sekvenci molekuly DNA neboli pořadí nukleotidů v jejím v řetězci, trápila vědce už od roku 1953, kdy byla poprvé identifikována její struktura (2). Historie sekvenování se začala psát v roce 1955, kdy Frederick Sanger publikoval sekvenci inzulinu, za což získal v roce 1958 Nobelovu cenu za chemii (12). Průkopníkem v oblasti DNA sekvenování byl americký biochemik čínského původu Ray Wu z Cornellské univerzity, který k přečtení 12 nukleotidů z okrajových regionů DNA fága lambda využil modifikovanou metodu prodlužování primeru. Na jeho průlomovou práci navázali i další vědci a o několik let později byly představeny techniky schopné reprodukovatelného čtení delších úseků DNA. Nezávisle na sobě objevily v roce 1977 týmy vedené britským biochemikem Frederickem Sangerem a americkým biochemikem Walterem Gilbertem dvě metody, které identifikaci pořadí nukleotidů v nukleové kyselině umožnily. Oba tito vědci obdrželi za svůj výzkum v roce 1980 Nobelovu cenu v oboru chemie, u Sangera se tak jednalo již o druhou Nobelovu cenu. (6). Nové metody byly následně pojmenovány po svých objevitelích a začaly být rychle testovány a používány v řadě světových laboratoří. V 80. letech 20. století nastal velký rozvoj především Sangerovy metody. Metoda byla modernizována využitím kapilární elektroforézy, která umožnila nahradit polyakrylamidový gel mnohem citlivějším tekutým polymerem. Radioaktivní značení bylo nahrazeno bezpečnějšími fluorescenčními barvami, které mohly být detekovány laserem a vzniklý signál zaznamenáván počítačem (13).

3.1 Cesta k lidskému genomu

Od chvíle, kdy vědecký svět získal nástroj, kterým šlo poměrně snadno přečíst genetickou informaci, začal v této oblasti prudký výzkum. Již v roce 1977 publikoval F. Sanger se svým týmem sekvenci genomu bakteriofága ϕ X174 o délce 5 375 bp v časopise Nature (6). Úspěchy sekvenování pokračovaly i v 80. letech 20. století, kdy byl osekvenován první lidský mitochondriální genom o délce 16 569 bp (7). Teprve v roce 1995 byl však poprvé osekvenován kompletní genom bakterie, tedy prvního organismu schopného samostatného života. Byla jí *Haemophilus influenzae* s genomem o délce 1,8 Mb a o rok později byl přečten genom prvního eukaryontního organismu – kvasinky *Saccharomyces cerevisiae*, u níž bylo identifikováno téměř 6 tisíc genů (8, 9). V roce 1988 začal být připravován a v roce 1990 formálně zahájen americký Projekt lidského genomu (the Human

Genome Project), který byl naplánován na dobu 15 let a byl řízen americkým Úřadem pro energii (Department of Energy) a Národními ústavami zdraví (the National Institutes of Health). Společnost International Human Genome Sequencing Consortium si jako hlavní postup pro sekvenování vybrala metodu „clone-by-clone“ využívající klonování sekvencí o délce asi 150 tisíc bp, které jsou vloženy ve formě umělého plazmidu do bakterie a následně jsou tyto klony osekvenovány (19, 20).

V květnu 1998 byla Craigem Venterem založena soukromá společnost Celera Genomics, Inc., která vyhlásila svůj záměr osekvenovat celý lidský genom do 3 let od svého založení za celkovou částku 300 milionů dolarů (4). Společnost si jako hlavní metodu sekvenování zvolila „Whole Genome Shotgun Assembly“, která byla poprvé použita v roce 1979, a která byla v roce 1995 také využita pro sekvenování genomu bakterie *Haemophilus Influenzae* (8, 16). Při této metodě se nevyužívá genetické a fyzikální mapování, ale náhodně velké naštěpené fragmenty DNA v rozmezí asi 2 - 50 tisíc párů bází, které jsou vloženy do plasmidového vektoru a namnoženy pomocí bakterie *E.coli* k získání milionů kopií každého fragmentu. Tyto fragmenty byly osekvenovány a výsledná data pak seřazena pomocí špičkových počítačů společnosti Celera Computers. Samotné sekvenování firma spustila 8. září 1999 a bylo při něm použito 300 sekvenátorů ABI PRISM 3700 s 96 paralelně pracujícími kapilárami. Rozdílnost obou postupů při sekvenování prvních lidských genomů je znázorněna na obrázku 1. Mezi oběma společnostmi začal závod o to, která z nich jako první opublikuje první osekvenovaný lidský genom. Společnost Celera přečetla přibližně $1,48 \times 10^{10}$ párů bází z více než 27 milionů vygenerovaných sekvenčních readů, což reprezentuje 5,11 x pokrytí celého lidského genomu (14).



Obr. 1: Rozdíl mezi sekvenačními postupy při sekvenování prvního lidského genomu (17, převzato a upraveno)

3.2 Výsledky a přínos závodu o lidský genom

Závod o první lidský genom skončil roku 2001, kdy 15. února publikovala společnost International Human Genome Sequencing Consortium draft (zhruba 90 %) lidského genomu v časopisu Nature a následně o den později 16. února společnost Celera Genomics, Inc. publikovala svou sekvenci lidského genomu v časopisu Science. Mezinárodní tým vědců však oficiálně oznámil dokončení plné identifikace lidského genomu až 14. dubna 2003 (k 50. výročí objevu dvoušroubovice DNA Jamesem Watsonem a Francisem Crickem) (21, 22).

Politika otevřeného přístupu HGP projektu měla pomoci výzkumu v oblasti genomiky, diagnostiky a léčby dědičných onemocnění všem vědcům, zatímco Celera genomics byla příznivcem více komerčně orientovaného přístupu, v rámci něhož se dokonce snažila patentovat některé nalezené geny. Z těchto zdrojů chtěla platit své investory a navíc měla v plánu zavedení poplatků za poskytnutí dat ze sekvenace. Vzhledem k tomu, že více než polovina dat společnosti Celera byla převzata z HGP a vzhledem k volné dostupnosti dat ze sekvenování lidského genomu, však musela společnost Celera změnit svůj plán a stát se společností diagnostickou (23). Závod o přečtení prvního lidského genomu znamenal velký přínos pro rozvoj nových sekvenačních technologií. Zatímco v roce 1985 bylo v jedné laboratoři osekvenováno 500 bp/den, v roce 2000 bylo ve společnosti Celera osekvenováno cca 175 000 bp/den a ve společnosti HGP dokonce 1000 bp každou vteřinu. I když už v té době byla Sangerova metoda značně sofistikována (polyakrylamidový gel byl nahrazen tekutým polymerem, radioaktivní značení bylo nahrazeno bezpečnějším fluorescenčním, celý proces byl automatizován a vyhodnocení obstarával počítač), přesto bylo jasné, že současná technika dosáhla svých kapacitních možností a nové poznatky a rozvoj v oblasti genomiky budou umožněny pouze vývojem technologií nových (24).

3.3 Variabilita lidského genomu

V rámci projektu 1000 Genomes, který si vzal za cíl popsat všechny genetické varianty v lidském genomu s frekvencí alespoň 1 %, se ukázalo, že variabilita zárodečného lidského genomu je mnohem větší, než se původně předpokládalo. Překvapivé množství zjištěných variant se týkalo nejen jednonukleotidových polymorfismů (SNP, z angl. Single Nucleotide Polymorphism), ale také ztrát (delecí) nebo zisků (insercí) úseků DNA. V neposlední řadě však byla oproti očekávání zjištěna velká proměnlivost v počtu kopií (CNV

z angl. Copy Number Variant), při němž se jednotlivé alely vyskytují v různém počtu kopií. Cíl projektu byl tak rozšířen z původního plánu 1 000 osekvenovaných lidských genomů na 2 500 genomů (15). Získaná data z projektu slouží jako zdroj k porovnání sekvenačních dat v rámci populace. Umožňují tak filtraci genetických variant, které se v populaci vyskytují často, od méně se vyskytujících variant, které mohou být kauzální příčinou dědičného onemocnění.

Naprostá většina kauzálních mutací, které v lidském genetickém kódu detekujeme, jsou tzv. stabilní mutace. Jedná se o mutace, které se při přenosu na novou generaci nemění a mají nejčastěji rozsah jednoho či několika málo nukleotidů. Nejčastěji se setkáváme se záměnami nukleotidů, které se podle vyplývající funkční změny na úrovni aminokyseliny dělí na mutace typu nonsense, missense či sestřihové mutace. U nonsense mutace dochází ke vzniku stop kodonu a tak k předčasnému ukončení translace. Při missense mutaci dochází k záměně nukleotidu za jiný, což může, ale také nemusí vést ke vzniku jiné aminokyseliny. Při sestřihové mutaci dochází k zániku sestřihového místa, či vzniku místa nového. Další časté mutace jsou inserce či delece, které vedou k tzv. frame-shiftu neboli posunu čtecího rámce. Ten má za následek začlenění jiných aminokyselin či opět vznik stopkodonu, který translaci předčasně ukončí. Druhou velkou skupinou jsou dynamické mutace, u nichž dochází k expanzi nestabilního, většinou tripletového elementu, který se s další generací dále amplifikuje uvnitř genu nebo mimo něj a počet opakujících se jednotek tak extrémně narůstá. Příkladem dědičného onemocnění způsobeného dynamickou mutací je Huntingtonova choroba, myotonická dystrofie či syndrom fragilního X. Podle četnosti výskytu mutací můžeme ještě rozlišit mutace prevalentní, které jsou v populaci časté a mutace privátní, které jsou pro daného pacienta a rodinu specifické. Kompletní přehled mutací, které byly v lidském genomu nalezeny je k dispozici na stránkách The Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk/ac/index.php>) (50).

4. SEKVENOVÁNÍ NOVÉ GENERACE

Objevení techniky tzv. masivního paralelního sekvenování (MPSS - z anglického „Massively Parallel Signature Sequencing“) umožnilo rozvoj sekvenačních metod označovaných jako NGS (Next Generation Sequencing). MPSS byla poprvé použita v komerčně dostupném zařízení v roce 2005. Jedná se o univerzální techniku pro charakterizaci a kvantifikaci nukleových kyselin. Zatímco při klasické Sangerově metodě sekvenace je sekvenovaný úsek amplifikován a následně jsou čteny báze jedna po druhé, v případě MPSS je nukleová kyselina rozdělena na fragmenty o určité délce, z nichž je každý amplifikován samostatně a mnoho takto vzniklých kopií každého fragmentu je poté čteno různými technologiemi podle použité platformy. Zavedení metody masivního paralelního sekvenování umožnilo v oblasti genomiky nebývalý rozvoj, jelikož bylo náhle možné osekvenovat celý lidský genom za zlomek ceny a času v porovnání s do té doby používanými metodami (25). V lednu 2014 dokonce společnost Illumina představila svůj nový sekvenační systém HiSeq X Ten, který je schopný osekvenovat kompletní lidský genom za částku 1 tisíc amerických dolarů (26).

Na trhu dnes dominují čtyři platformy sekvenátorů druhé generace – Illumina genome Analyzer, 454 Life Sciences, Solid system a Ion Torrent od společnosti Life Technologies. I přesto, že všechny dnes běžně dostupné NGS platformy využívají rozdílnou chemii a různé přístupy k sekvenování, samotné sekvenační schéma je u všech platforem společné a skládá se ze tří hlavních kroků:

- a) Příprava sekvenační knihovny**
- b) Sekvenování a detekce nukleotidů**
- c) Softwarová analýza získaných dat**

4.1 Příprava sekvenační knihovny

Příprava sekvenační knihovny je často zdlouhavý a náročný proces, během něhož je kladen důraz na přesnost provádějícího pracovníka, stejně jako na přístroje a pomůcky, které ke své práci používá. Nevhodná příprava vzorku pro sekvenování se může stát výrazným zdrojem sekvenačních chyb, které snižují či v některých případech znemožňují správné vyhodnocení získaných dat. Velký důraz je přitom kladen především na přesnost

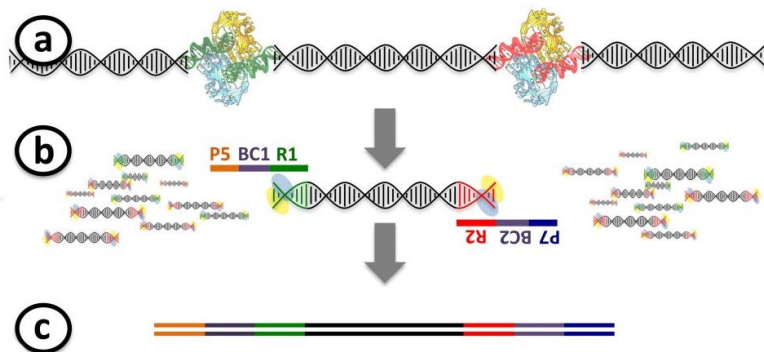
laboratorních pipet a jiných měřidel a přístrojů, které jsou při přípravě používány. Pokud je připravováno více vzorků současně, je nutné zajistit reprodukovatelnost jednotlivých úkonů tak, aby všechny vzorky byly zpracovávány pokud možno za stejných podmínek. Důležitým faktorem při přípravě je i kvalita a kvantita vstupní DNA. Během přípravy knihovny je DNA naštěpena a na její upravené konce jsou ligací připojeny pomocné adaptéry, které umožňují PCR amplifikaci. Během této polymerázové řetězové reakce je každý fragment mnohonásobně zkopírován a je tak připraven k samotnému sekvenování (27).

4.1.1 Fragmentace genomické DNA

Prvním krokem přípravy sekvenační knihovny je rozdělení nukleové kyseliny na potřebně dlouhé úseky (fragменты), jejichž délka se liší podle použité technologie. Například nejnovější přístroje firmy Roche dokáží pracovat i s fragmenty dlouhými až 1 000 bp, v praxi se však u nich nejčastěji využívá délka kolem 700 bp (28). Jinak je tomu v případě platformy Illumina, kdy je optimální délka fragmentů zhruba 150-300 bp (29). K fragmentaci lze použít fyzikální či enzymatickou metodu a protokol musí být vždy optimalizován tak, aby vyhovoval technologii, která je použita k samotnému sekvenování. Z fyzikálních metod se využívá nebulizace dusíkem či sonikace pomocí ultrazvukových vln. Nevýhodou mechanického štěpení je, že se jedná o štěpení nespecifické a tedy nevíme přesně, v jakých místech byla DNA rozštěpena. Další nevýhodou je nutnost vyhlazení ostrých konců pomocí enzymu. Nebulizace je na druhou stranu reprodukovatelná technika, která je zároveň rychlá a levná (30).

Velký pokrok v přípravě knihoven přinesla enzymatická metoda využívající enzymu Tn5 transposázy, kterou jako první použila společnost Epicentre Biotechnologies (Madison, WI) pod označením Nextera. Ta zprostředkovává fragmentaci dvouřetězové DNA a následnou ligaci syntetických oligonukleotidů na obou koncích během 5-10 minutové reakce. DNA je při této reakci fragmentována pomocí transpozomu (komplex transposázy a transposonu) a následně je na oba konce vzniklého fragmentu přidána invertovaná repetice o délce 19 bp (31). Postup tagmentace je znázorněn na obrázku 2. Po fragmentaci je vždy bezpodmínečně nutné zkontrolovat velikost vzniklých fragmentů pomocí elektroforézy. Starší gelovou elektroforézou dnes v NGS laboratořích většinou nahrazuje modernější čipová

elektroforéza, která umožňuje kvalitativní i kvantitativní vyhodnocení vzniklých fragmentů. K měření postačí 1 μ l fragmentované DNA.



Obr. 2: Štěpení DNA pomocí Tn5 transpozomu (Nextera) (71)

(a) Transpozom se integruje do genomické DNA. (b) Tagmentační reakce produkuje amplifikovatelné a neamplifikovatelné molekuly knihovny, dokud nedojde k vyčerpání transpozomu (c) Knihovna je přečištěna a je odstraněn Tn5 protein navázaný na koncích fragmentu, poté jsou pomocí PCR reakce přidány adaptéry pro uchycení k průtokové komůrce a indexy k rozlišení jednotlivých vzorků.

4.1.2 Úprava DNA fragmentů před amplifikací

Po fragmentaci jsou na oba konce získaných fragmentů ligací připojeny adaptérové molekuly, které slouží jako templát v klasické PCR reakci i pro nasedání sekvenačních primerů v samotné sekvenační reakci. V případě, že budeme provádět emulzní PCR, jsou na oba konce připojeny adaptéry A a B, z nichž B adaptér obsahuje biotinovou značku. Pomocí této značky se fragment váže na magnetické kuličky obalené streptavidinem, na nichž dochází k samotné amplifikaci. Jsou použity pouze fragmenty, které obsahují oba adaptéry. Fragmenty, které oba adaptéry neobsahují, jsou v promývacím kroku ze vzorku odstraněny. Po enzymatické fragmentaci metodou Nextera je fragmentovaná DNA amplifikována krátkou PCR reakcí za použití primerů, která obsahují sekvence cílené k transposonu. Tyto primery přidávají na fragmenty konkrétní adaptéry a sekvenační kódy, které umožňují identifikovat jednotlivé vzorky (32).

4.1.3 Amplifikace DNA fragmentů

I když už dnes existují technologie, které umožňují sekvenovat DNA bez její amplifikace, u většiny postupů je stále potřebné po naštěpení genomické DNA a úpravě vzniklých fragmentů tyto fragmenty namnožit pomocí polymerázové řetězové reakce. K tomu slouží především dvě amplifikační metody – můstková amplifikace, jež je hlavní metodou v přístrojích společnosti Illumina a emulzní PCR, kterou využívají ve svých přístrojích společnosti 454 Life Sciences i Life Technologies (Solid a Ion Torrent). Amplifikace za pomoci polymerázy však činí z tohoto kroku jeden z možných zdrojů chyb ve výsledných datech.

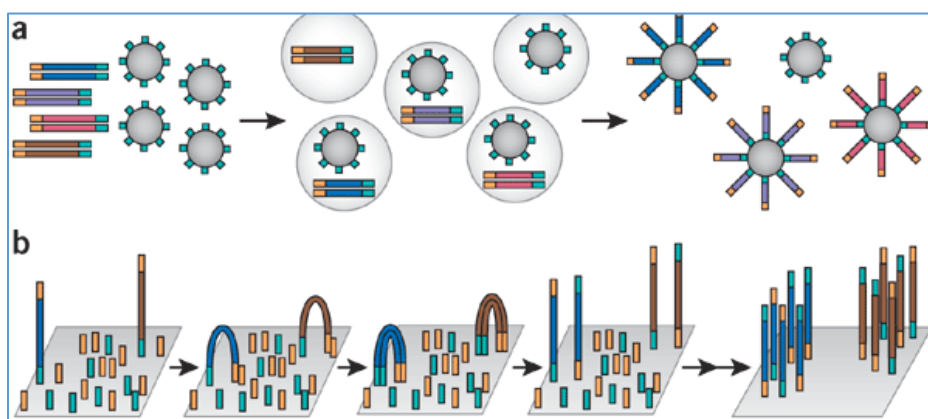
4.1.3.1 Emulzní PCR (emPCR)

Při emulzní PCR reakci jsou fragmenty pomocí B-adaptérů obsahujících na 5'-konci biotinovou značku připojeny na magnetické kuličky pokryté streptavidinem (obrázek 3a). Takto navázaná dsDNA je denaturována na jednotlivé řetězce, přičemž je uvolněn nebiotinizovaný komplementární řetězec a na kuličce tak zůstávají pouze jednovláknové fragmenty DNA. Takto získané řetězce jsou hybridizovány ke kuličkám, které mají na svém povrchu komplementární sekvence sloužící jako primery v následné amplifikační reakci. Celý systém je navržen tak, aby došlo k navázání pouze jednoho DNA fragmentu na jednu kuličku. Amplifikace fragmentů probíhá pomocí tzv. emulzní PCR reakce, která pro každou kuličku probíhá v jejím vlastním „reaktoru“. Ten je tvořen emulzí oleje a vody. Výsledkem reakce je vznik asi 10 milionů identických kopií daného fragmentu na jedné kuličce. Po emulzní PCR reakci následuje obohacení (enrichment), jehož cílem je selekce a odstranění těch kuliček, na nichž nedošlo k navázání fragmentu a následné amplifikaci. Provádí se pomocí sekvenačního primeru, který se přichytí na kuličky s amplikony a společně se nanesou do tzv. pikotitrační destičky s tak malými otvory, že se do každého vejde pouze jediná kulička s amplikony (33,34).

4.1.3.2 Můstková polymerázová řetězová reakce (bridge-PCR)

Při můstkové amplifikaci se templátová DNA hybridizuje na transparentní povrch komůrky (flow cell, obrázek 3b). K tomu se využívá chemicky modifikovaných nukleotidů. Po denaturaci dvouvláknové DNA knihovny jsou fragmenty pomocí adaptérů hybridizovány k reakční komůrce. Každý fragment je tak jedním svým koncem připojen k povrchu komůrky, na níž probíhá následně amplifikace. Po přidání PCR reagensů slouží adaptéry na komůrce

jako primery pro syntézu dsDNA. Ta je následně denaturována, původní templát je odmyt, nově vzniklé vlákno však zůstává pevně spojeno s komůrkou. Dojde k ohnutí vlákna a přemostění (odtud označení můstková PCR) a vlákno svým volným zakončením hybridizuje k adaptérům na povrchu komůrky. Oblast P5 adaptérů se štěpí a k reakční komůrce tak zůstává vlákno připojené pouze pomocí P7 adaptéru, což zajišťuje, že bude každé vzniklé vlákno čteno ve stejném směru. V dalším PCR cyklu sekvenační primer nasedá na P5 oblast a opět dochází ke vzniku komplementárního vlákna a celý proces se cyklicky opakuje. Ve finální fázi jsou denaturovány vzniklé dvouřetězce a reverzní řetězce jsou odmyty. Výsledkem je asi 1 000 identických kopií DNA fragmentů k vlastní sekvenaci (klastry) (33,34, 45).



Obr. 3: Schématické znázornění emulzní PCR (a) a můstkové PCR (b) amplifikace (72)

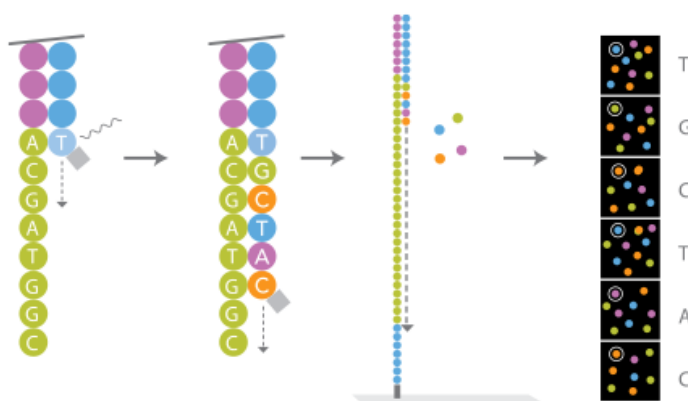
4.2 Sekvenování a detekce nukleotidů

V tomto posledním kroku laboratorní části jde v podstatě už jen o přečtení sekvence nukleotidů ve fragmentech, které jsme připravili a namnožili v předchozích krocích. Rozlišujeme tři nejčastější způsoby sekvenace – pyrosekvenování, sekvenování syntézou a sekvenování ligací. Zaznamenáním sekvence jednoho fragmentu vzniká tzv. read, neboli čtení. Pro kvalitu sekvenování je důležitá tzv. hloubka čtení. Jedná se o počet, kolikrát byl daný nukleotid během sekvenačního běhu přečten. V klinické praxi existují 2 různé přístupy k NGS sekvenování - ultraširoké sekvenování a ultrahluboké sekvenování. Pokud pouze hledáme zárodečné mutace, které jsou v homozygotním či heterozygotním stavu, potom je vhodné zvolit ultraširoké sekvenování. To nám umožňuje sekvenovat velké množství vzorků v jednom runu. V některých případech, jakými jsou např. genetické mozaiky či somatické mutace, však nastává potřeba ultrahlubokého sekvenování, kdy mnohokrát sekvenujeme jeden

úsek určitého vzorku. Tento postup nám zajistí vysokou citlivost, která může výrazně snížit chybovost. Mnohonásobné prosekvenování jednoho úseku DNA - vysoká „coverage“ tak kompenzuje vyšší chybovost sekvenačních metod nové generace (33, 34, 45).

4.2.1 Platforma Illumina (sekvenace syntézou)

Sekvenační technologie společnosti Illumina využívá sekvenaci syntézou, kdy jsou po vytvoření klastrů k adaptérům na koncích fragmentů hybridizovány sekvenační primery a do průtokové komůrky je přivedena směs polymerázy a nukleotidů, které jsou označeny fluorescenční značkou. Tyto nukleotidy mají deaktivovanou 3'-OH skupinu, aby bylo zajištěno, že se v každém cyklu začlení pouze jeden nukleotid. Po úspěšném začlenění báze do vlákna nukleové kyseliny jsou její pozice a typ zaznamenány CCD kamerou, která fotografuje průtokovou komůrku. Kamera obsahuje 2 lasery, z nichž každý zaznamenává 2 typy bází. Červený laser zaznamenává báze A a C a zelený laser zaznamenává G a T. Detailnější rozlišení je poté zajištěno pomocí filtrů. Na konci každého cyklu je poslednímu nukleotidu odstraněna terminační skupina na 3'-OH konci i fluorescenční barva, aby mohl být cyklus opakován. Postup čtení nukleotidů na platformě Illumina je znázorněn na obrázku 4. Na základě speciálního algoritmu je přidělována jednotlivým bázím hodnota, na základě níž jsou vyřazeny sekvence o nízké kvalitě. Některé sekvenátory společnosti Illumina umožňují pro různé typy aplikací pracovat ve 2 režimech – „Rychlý“ režim (Rapid Run) a „Vysoce výkonný“ režim (High Output Run). Při Rapid Run lze výsledky sekvenování získat během jednoho dne. High output mod je vhodný pro rozsáhlé studie s velkým množstvím vzorků, kde je požadováno vyšší pokrytí čtených úseků (33, 34, 45).



Obr. 4: Čtení nukleotidů při sekvenování syntézou na platformě Illumina (73)

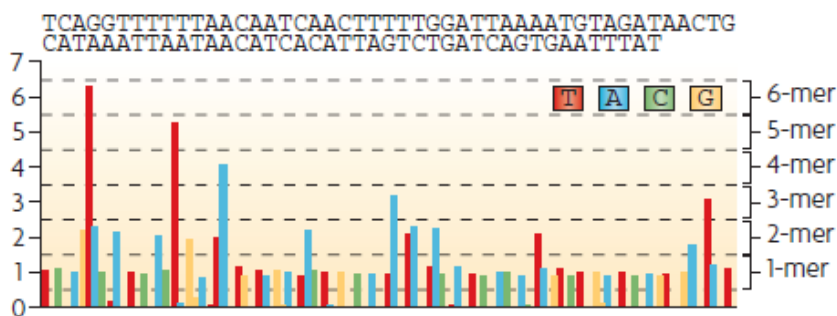
Společnost Illumina nabízí dvě řady sekvenátorů podle výkonu a nároků pracoviště. Pro menší provozy jsou nabízeny přístroje MiSeq, případně MiSeqDx, které se hodí jako náhrada klasické kapilární sekvenace pro cílené sekvenování jednotlivých genů, panelů více genů, exomové sekvenování či resekvenování malých genomů. Pro velká vědecká pracoviště, která mají za svůj cíl mimo jiné *de novo* sekvenaci genomů, Illumina nabízí přístroje NextSeq 500, HiSeq2500 či HiSeq X Ten. NextSeq je dostupný za cenu kolem 250 tisíc amerických dolarů a zahrnuje přípravu (generování klastrů) a následnou sekvenaci v jednom přístroji. Během 12-ti hodinového běhu dokáže například osekvenovat celý lidský genom, 16 exomů, 20 vzorků neinvazivního prenatalního testování, 20 transkriptomů, 48 vzorků genové exprese nebo až 96 vzorků pomocí cílených panelů. HiSeq X Ten způsobil průlom v NGS sekvenování tím, že poprvé prolomil magickou hranici 1 000 dolarů za lidský genom. Jeho cena se pohybuje kolem 10 milionů dolarů a průchodnost přístroje je $1,8 \times 10^{12}$ bází za třídní run (600 Gb/den). HiSeq X Ten je ve skutečnosti vhodný spíše pro velké společnosti, které se sekvenováním zabývají, nikoliv pro klinické provozy či malá výzkumná pracoviště. Aby bylo totiž reálně dostat se na částku 1 tisíc dolarů za genom, bylo by nutné ročně osekvenovat 18-20 tisíc lidských genomů (36).

4.2.2 Platforma 454 Life Sciences (pyrosekvenování)

Sekvenátory 454 Life Sciences pracují na principu pyrosekvenování. Jedná se o enzymatickou reakci, během níž se do nově vznikajícího řetězce začleňují nové báze, při čemž dochází k emisi viditelného světla. Množství vznikajícího světla je pak úměrné počtu začleněných nukleotidů. Metodu monitorování DNA polymerázy pomocí bioluminiscence navrhl již v roce 1987 P. Nyrén (41). Samotná pyrosekvenace však byla uvedena do praxe až v roce 1998. V roce 2000 vznikla společnost 454 Life Science, která se v roce 2005 spojila s firmou Roche Diagnostics. Ta měla zajistit výrobu sekvenátorů a komerčních kitů. Ještě v témže roce byl představen první sekvenátor na principu pyrosekvenace (33,34).

Po emulzní PCR jsou do jamiček v pikotitrační destičce postupně přiváděny i další kuličky s enzymy důležitými pro pyrosekvenační reakci a jednotlivé nukleotidy, které jsou do jamek přiváděny v pevně určeném pořadí. Pro zamezení enzymatické aktivity luciferázy vůči ATP přidávanému do reakce jsou použity α -thiosubstituované nukleotidy (deoxyadenosin-5'-thiotrifosfát), které s luciferázou nereagují, současně jsou však substrátem DNA polymerázy. V případě komplementarity a inkorporace daného nukleotidu do nově vznikajícího

komplementárního řetězce dochází k uvolnění pyrofosfátu. Ten je enzymem ATP-sulfurylázou převeden na ATP. Vzniklý adenosintrifosfát umožní reakci katalyzovanou enzymem luciferázou, při níž se luciferin přeměňuje na oxyluciferin za emise viditelného světla, které je snímáno CCD kamerou. Proces přidávání nukleotidů se cyklicky opakuje, až dojde k vytvoření celého vlákna, které je komplementární s původním templátem. Nezařazené nukleotidy jsou podobně jako zbylé ATP degradovány pomocí enzymu apyrázy. Výsledkem sekvenace je tzv. flowgram, tedy graf, ze kterého lze na základě velikosti píku odečíst pořadí a počet jednotlivých bází (obrázek 5) (33,34).



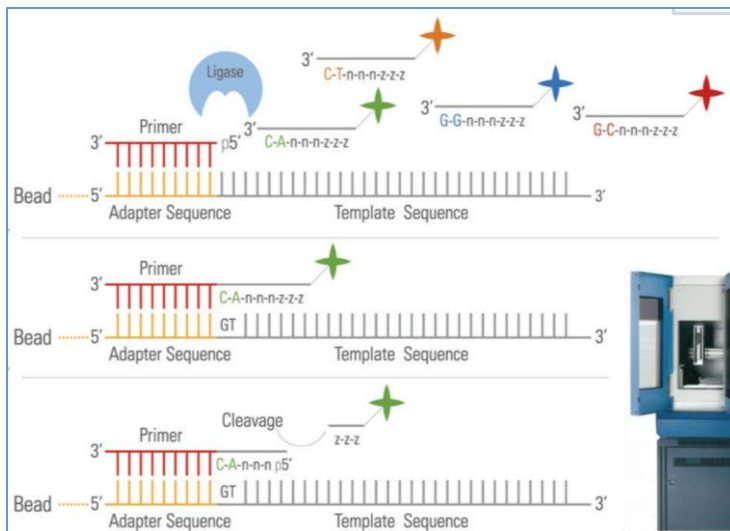
Obr. 5: Flowgram s výsledky pyrosekvenování (74)

V říjnu 2013 společnost 454 Life Sciences oznámila, že má v plánu do poloviny roku 2016 ukončit výrobu svých sekvenátorů GS Junior a GS FLX+, které jsou založeny na pyrosekvenování. Hlavním důvodem ukončení výroby je podle slov společnosti obtížná modernizace současných přístrojů. Z důvodu hledání nových technologií navázala v roce 2013 společnost 454 Life Sciences spolupráci s do té doby konkurenční společností Pacific Biosciences, s níž má v plánu pracovat na nových přístrojích založených na technologii Single molecule real time sequencing (SMRT) (36).

4.2.3 Platforma Life Technologies Solid (sekvenace ligací)

Po emulzní PCR amplifikaci jsou při tomto typu sekvenování kuličky s DNA fragmenty připojeny na povrch sklíčka. Takto připravené sklíčko se vloží do speciální kazety, která umožňuje fluidní průtok. Systém využívá sondy o délce 8 nukleotidů, u nichž známe vždy sekvenci prvních dvou. Sondy jsou označeny jednou ze čtyř fluorescenčních barev, celkem tedy každá barva představuje 4 z 16 možných nukleotidových sekvencí. K primeru je pomocí ligázy připojena sonda, jejíž první dvě báze jsou komplementární s prvními bázemi

čteného fragmentu. Po připojení následuje štěpení sondy nukleázou za pátou bázi, čímž dojde k uvolnění fluorescenční barvy. Tento proces se opakuje, dokud nedojde k pokrytí celého fragmentu. Postup prvního kola ligace je znázorněn na obrázku 6. Následuje druhé kolo, při němž je použit primer o jednu bázi kratší a opět dochází k ligaci sond a jejich štěpení. Po pěti kolech ligace a štěpení můžeme přesně určit, jaký nukleotid se v daném místě nachází (54).

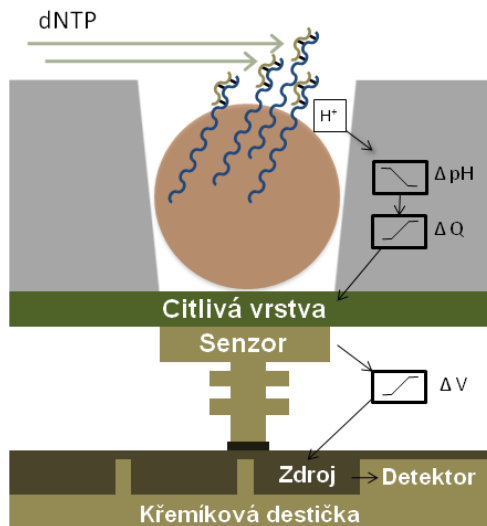


Obr. 6: Postup při sekvenaci ligací na platformě Solid společnosti Life Technologies (54)

4.2.4 Platforma Life Technologies Ion Torrent (syntéza s chemickou detekcí)

Také Ion Torrent stejně jako Illumina využívá při sekvenování syntézu komplementárního vlákna. Hlavním rozdílem této technologie od těch předchozích je však skutečnost, že Ion Torrent umožňuje převádět chemický signál přímo do digitální podoby. Sekvenátory vyráběné na této platformě nevyužívají k zaznamenávání dat žádné optické detektory, což velmi zrychluje sekvenční proces. Nákladné a často komplikované hodnocení signálů fluorescence či chemiluminiscence je zde nahrazeno měřením vodíkových iontů, které jsou uvolňovány v průběhu polymerázové řetězové reakce. Po amplifikaci jsou kuličky s namnoženým teplátem umístěny do jamek v polovodičovém čipu technologie CMOS (z anglického “Complementary Metal–Oxide–Semiconductor“), který obsahuje deset miliónů submiskropických cell, v nichž v reálném čase probíhá analýza změn pH (75, 76) (obrázek 7).

Do jamek jsou postupně přiváděny nukleotidy a dochází k syntéze komplementárního vlákna DNA. Pokud dojde k inkorporaci nukleotidu, uvolní se vodíkový kation, čímž dojde ke snížení pH. Tato změna je zaznamenána detektorem. Jediným limitujícím faktorem této technologie je kapacita čipu. Lze tedy předpokládat, že v budoucnu ve svých přístrojích firma nabídne vyšší kapacitu, což umožní ještě rychlejší a levnější sekvenování (75, 76). Společnost Life Technologies, která byla v únoru 2014 zakoupena firmou Thermo Fisher Scientific, v současnosti staví své NGS přístroje na dvou platformách – Solid a Ion Torrent. Většina pozornosti v rámci vylepšení se soustředí však pouze na druhou jmenovanou platformu. Podle uváděných specifikací dokáže systém Ion PGM™ během čtyřhodinového runu vyprodukovat až 6 milionů readů o délce 400 bp. Tento přístroj využívá čip Ion 318. Oproti tomu přístroj používající čip Ion PI, který dostal typové označení Ion Proton, produkuje až 80 milionů readů o délce 200 bp ve stejném čase, nebo 10-14 Gb za jeden sekvenační run. Nejnovější Ion PII čip vhodný pro transkriptomové sekvenování by měl vyprodukovat až 300 milionů readů o délce 100 bází. K významnému zpřesnění analýz by měla také přispět nově vyvíjená polymeráza nazvaná Hi-Q, u níž je výrobcem uváděna až o 90 % menší chybovost. (36).

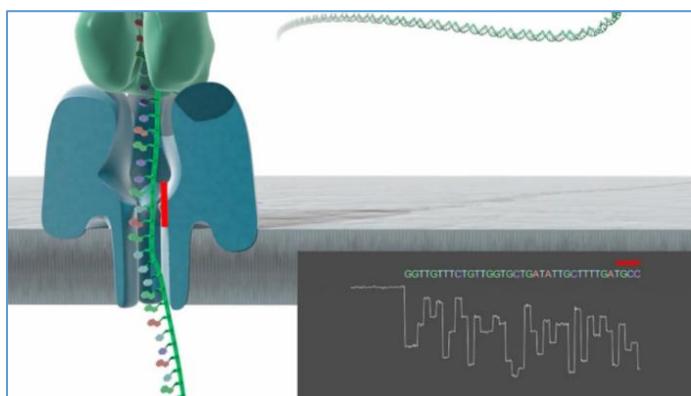


Obr. 7: Schématické znázornění detektoru přístroje Ion Torrent (76, převzato a upraveno)

4.2.5 Ostatní platformy

Mimo zde popsané platformy však dochází k vývoji platform nových, které však zatím nenašly masivní uplatnění v molekulárně genetických provozech. Mezi jejich společný znak patří skutečnost, že nevyžadují amplifikační krok a k sekvenování jim tak stačí jediná molekula DNA. Pokud budeme skupinu výše popsaných sekvenačních technologií nazývat druhou generací sekvenování, lze bezpochyby bezamplifikační sekvenační metody označit za generaci třetí. Jak již bylo zmíněno, tyto metody nevyžívají před vlastním sekvenováním žádný amplifikační krok, což zkracuje dobu přípravy DNA, redukuje cenu, snižuje chybovost pramenící z amplifikace, umožňuje vyšší flexibilitu v délce čtení a přesnou kvantifikaci DNA molekul, protože signál je zaznamenáván v reálném čase. Mezi tyto platformy patří technologie vyvíjené společnostmi Oxford Nanopore a Pacific Biosciences (34-36).

Společnost Oxford Nanopore Technologies se zaměřuje na co nejsnadnější přípravu vzorků a celého sekvenačního postupu. K analýze DNA využívá protažení její molekuly přes systém biologických nanopórů tvořených proteiny (obrázek 8). Tyto nanopóry jsou spolu se systémem nádržek a polymerních membrán součástí analyzačního čipu. Ve chvíli, kdy projde určitým nanopórem část DNA se specifickým nukleotidem, dojde ke změně vodivosti celého obvodu. Tato změna je změřena a je specificky vyhodnocena pro každý typ nukleotidu. Zřejmě nejzajímavější technologií této společnosti je sekvenátor MinION o velikosti USB klíčenky, který na principu nanopórového sekvenování umožňuje jednomolekulovou DNA analýzu. Přístroj obsahuje sensorový čip a rozvětvený analyzátor na bázi mikrofluidiky, který využívá znalostí o proudění kapalin v zařízeních s miniaturními rozměry. Zařízení je přes USB rozhraní připojitelné k běžnému stolnímu počítači či notebooku, na němž je nainstalován příslušný software (34-36).



Obr. 8 : Průchod DNA nanopórem při sekvenování platformou Oxford Nanopore (76)

Společnost Pacific Biosciences využívá ve svých sekvenátorech technologii sekvenování jednotlivých DNA molekul v reálném čase (Single Molecule Real-Time sequencing, SMRT) na principu ZMW (Zero Mode Waveguide), při něm jsou jednotlivá vlákna DNA uchycena v dutinách ZMW čipu o průměru 70 nm. Na dně každé dutiny je tak umístěn jednořetězcový templát spolu s DNA polymerázou. Metoda využívá fluorescenčně značené deoxyribonukleotidy. Na fosfátovou skupinu každého nukleotidu je navázána jedna ze čtyř specifických fluorescenčních barev. Při začlenění komplementárního nukleotidu je specifický světelný signál zaznamenán detektorem a následně dojde k odstranění fluoroforu z fosfátové skupiny. Tím může dojít k začlenění dalšího nukleotidu (obrázek 9). Tímto postupem je možné přečíst až 10 nukleotidů za sekundu, přičemž délka čtených fragmentů může být i více než 10 kb (35,36).



Obr. 9: Metoda sekvenování jednotlivých DNA molekul v reálném čase - SMRT (52)

Firma při použití DNA polymerázy označené P6 a chemie označené C4 garantuje získat v průměru 10 kb dlouhé ready, ale i ready přesahující 40 kb. Nová sekvenační chemie vytváří jakési „ochranné lešení“, které zabraňuje fyzickému kontaktu mezi enzymem a fluoroforem a tím možnému poškození enzymu. Podle společnosti byla fotodestrukce při sekvenování omezena do té míry, že jediným limitujícím faktorem pro délku čtení je kvalita samotné DNA. Při čtení stále delších sekvencí je totiž stále obtížnější získat templáty, které neobsahují žádné poškození, jež by vedlo ve svém důsledku k předčasnému ukončení čtení. V současnosti používaný přístroj PacBio RSII využívající technologii SMRT dokáže vyprodukovat okolo 50 tisíc readů či asi 400 milionů bází za jeden sekvenační běh. Společností Pacific Bioscience byl také vyvinut systém HGAP (Hierarchical Genome Assembly Process), který umožňuje kompletní a přesné sestavení sekvence z takto dlouhých readů (36,37).

4.3 Přístupy k sekvenování

Každé pracoviště, které chce zavést některou z NGS metod, musí pečlivě uvážit všechny její aspekty, především přístup k sekvenování, který bude využívat a porovnat výhody a nevýhody dostupných platform. Pochopitelně je nutné zvolit jinou platformu pro vědecká pracoviště a jinou pro rutinní provozy. Často je samozřejmě při koupi nového přístroje zohledňována kupní cena, ale je velmi důležité také to, zda bude mít kapacita přístroje na daném oddělení dostatečné uplatnění. Je nutné kalkulovat také cenu reagensů a v neposlední řadě je nutné počítat s dostatečným množstvím pracovníků, kteří budou sekvenační experimenty připravovat, provádět a vyhodnocovat. Je nutné také počítat s jejich edukací v této oblasti. S nárůstem množství získaných sekvenačních dat jsou navíc kladeny vyšší nároky nejen na bioinformatiky, ale i na dostatečné počítačové vybavení pro všechny, kdo budou se sekvenačními daty pracovat. Při správném výběru a využití jsou NGS technologie výborným nástrojem, který umožňuje velmi rozšířit spektrum analyzovaných genů, což je jednoznačně výzvou k zajištění dostatečných preventivních a léčebných postupů pro rizikové pacienty.

4.3.1 Celogenomové sekvenování

Cena sekvenování lidského genomu je dnes mnohonásobně nižší, než tomu bylo před několika lety. I přesto je tento přístup pro svou časovou i finanční náročnost stále předurčen spíše pro výzkumná pracoviště, než k rutinnímu klinickému použití a jedná se stále o minoritní aplikační oblast. Nezanedbatelné je také ohromné množství získaných dat, které by bylo nutné zpracovat a která by často zůstala bez klinického využití. Výhodou celogenomového sekvenování je hlavně skutečnost, že je přečtena celá chromozomální DNA a to včetně promotorových a regulačních sekvencí. Největší význam tak má tato metoda v identifikaci nových a vzácně se vyskytujících mutací (45).

4.3.2 Exomové sekvenování

Exomové sekvenování lze označit za určitý typ cíleného sekvenování, které spojuje výhody sekvenování celogenomového a cíleného. Na rozdíl od celogenomového sekvenování jsou při něm však sekvenovány pouze kódující oblasti DNA, tedy exony kódující proteiny. V praxi rozlišujeme dva typy exomů – kompletní exom, kdy sekvenujeme všech cca 20 tisíc genů či tzv. klinický exom (mendeliom), který zahrnuje vyšetření 2-4 tisíc genů známých

v souvislosti s některým z dědičných onemocnění. Jelikož lidský exom představuje asi 1 % celého genomu, jeho sekvenováním dochází k výraznému snížení nákladů i potřebného času v porovnání s celogenomovým sekvenováním. Nezanedbatelné je také vyšší pokrytí (coverage) sekvenovaných úseků. Tyto výhody předurčují exomové sekvenování k použití na mnohem větším množství pracovišť, než je tomu u celogenomového sekvenování (45).

4.3.3 Cílené sekvenování (targeted sequencing)

Metoda ideální pro sekvenování velkého počtu vzorků umožňuje sekvenaci pouze vybraných genů či oblastí jediného genu spojeného s dědičným onemocněním (např. nádorová onemocnění, kardiomyopatie, cystická fibróza apod.). Pro tuto metodu sekvenace lze zakoupit komerční kit s již pevně zvoleným panelem vyšetřovaných genů, nebo si laboratoř může sama zvolit, které geny bude v rámci panelu vyšetřovat. Pro cílené sekvenování se hodí jednoduché stolní sekvenátory typu MiSeq či 454 GS Junior. Tento přístup je nejjednodušší cestou, jak na klinickém pracovišti nahradit klasické kapilární sekvenování NGS technologií (5, 45).

4.3.4 Transkriptomové sekvenování

Analýza transkriptomu je jednou z nejdůležitějších metod pro kompletní molekulární charakteristiku zejména u nádorových onemocnění. Transkriptom je kompletním souhrnem všech molekul ribonukleové kyseliny (tRNA, rRNA, mRNA i dalších nekódujících RNA molekul) v buněčné populaci či konkrétní buňce. Zatímco genom se v každé buňce nachází v podstatě pouze jeden, transkriptom oproti tomu odráží aktuální stav dané buňky. Jeho sekvenování je tedy velmi využitelné pro detekci genových fúzí, alternativních sestřihových variant či somatických mutací. Hluboké transkriptomové sekvenování dnes umožňuje identifikovat i dodnes neznámé transkripty a získat také informaci o transkripci jednotlivých genů mnohem přesněji, než pomocí dodnes používaných metod. Před samotným sekvenováním je vyizolovaná RNA převedena pomocí reverzní transkriptázy na cDNA, která je následně sekvenována (45).

4.4 Porovnání jednotlivých NGS platforem

V posledních několika letech jsme v oblasti sekvenování svědky ohromného technologického rozvoje. Jednotlivé firmy mezi sebou soutěží o místo na trhu se sekvenačními přístroji i sekvenační chemií. Vznikají nové přístroje, které si kladou za cíl eliminovat nevýhody přístrojů předchozích. Především se jedná o snížení chybovosti čtení, zlevnění, zrychlení a zjednodušení celého sekvenačního postupu. Právě jednoduchost obsluhy a relativně nízká cena stolních sekvenátorů je jedním z důvodů, proč si dnes nacházejí uplatnění na velkém množství klinických pracovišť a není daleko chvíle, kdy zcela jistě vytěsní klasické metody sekvenování. Nové technologie však stále mají několik nevýhod, mezi něž patří především jejich chybovost ve srovnání s klasickou metodou sekvenace a často i krátká délka sekvenovaných fragmentů, která činí složitější následné sestavení původní sekvence.

4.4.1 Chybovost NGS platforem

Jedním z největších problémů, se kterými se v praxi při sekvenování nové generace setkáváme, je jeho chybovost. První skupina chyb vychází z chybovosti samotné DNA polymerázy při syntéze komplementárního vlákna a je společná pro všechny platformy, které využívají PCR amplifikaci. Druhá skupina chyb je specifická pro každou platformu, respektive pro technologii, která je při sekvenování použita. Většina dnes dostupných platforem dokonce vykazuje vyšší chybovost ve srovnání se Sangerovou metodou. Zřejmě nejvíce problémů způsobuje sekvenátorům čtení v homopolymerech, dále pak úseky, v nichž došlo k delecii či inzerci části DNA řetězce. V tabulce č. 1 jsou uvedeny primární zdroje chyb, které jsou specifické pro každou platformu a vycházejí z použité technologie. Například technologie ION Torrent je nastavená tak, že při inkorporaci dvou stejných po sobě jdoucích nukleotidů do syntetizovaného vlákna, signál u druhého nukleotidu vzroste a senzor tak zaznamená dva stejné nukleotidy (70).

Právě tento na první pohled sofistikovaný mechanismus je příčinou chybovosti čtení v delších homopolymerech. Signál je totiž možné zvyšovat jen do určité míry a od 4. stejného nukleotidu již nedochází k jeho nárůstu, což vede k chybě. Chybovost při čtení v sekvenátorech nové generace je kompenzována mnohonásobným pokrytím čtených úseků (coverage). Zjednodušeně řečeno je každý fragment přečten opakovaně, aby bylo možné ve

výsledku vyfiltrovat menší procento těch, které se liší. Coverage by v případě sekvenace DNA z diploidního organismu měla být vyšší, než 10x, abychom odlišili sekvenační chybu od heterozygota. Konečně třetí a poslední skupina možných chyb vychází z použitého vstupního materiálu. Pokud zpracováváme normální tkáň, lze předpokládat, že genom všech buněk bude identický. To však nemusí platit např. pro mitochondriální DNA či tkáň nádorovou, v níž je skupina buněk postižena somatickou mutací (15, 70).

Tab. 1: Porovnání chybovosti dostupných NGS platform (70)

Sekvenátor:	Primární zdroj chyb:	Chybovost při jednom průchodu (%)	Celková chybovost (%)
3730xl (kapilární sekvenování)	substituce	0,1-1	0,1-1
454 všechny modely	indel	1	1
Illumina všechny modely	substituce	~0,1	~0,1
Ion Torrent – všechny čipy	indel	~1	~1
SOLiD – 5500xl	A-T bias	~5	≤0,1
Oxford Nanopore	delece	≥4*	4*
PacBio RS	indel	~13	≤1

4.4.2 Délka sekvenovaných fragmentů

Další nevýhodou nových platform je často krátká délka čtených úseků. Nejen pro *de novo* sekvenování je naopak velmi výhodné, když máme délky sekvenovaných fragmentů co největší z důvodu snazšího sestavení původní sekvence. Tato starost nám částečně odpadá v případě, že chceme především provádět resekvenování, kdy známe referenční sekvenci a můžeme se tak zaměřit na další parametry přístroje, jakými jsou cena, délka sekvenačního běhu, propustnost apod. Na základě porovnání dalších údajů, které jsou shrnuty v tabulce 2, můžeme konstatovat, že pro resekvenování a cílené sekvenování se v celkovém porovnání jeví jako nejlepší platforma Illumina. I přes poměrně krátkou délku čtených fragmentů se Illumina stává postupně jedničkou na trhu s NGS přístroji pro rutinní diagnostiku, což je dáno především jejich vysokou propustností, tedy množstvím přečtených bází za jeden sekvenační běh a v neposlední řadě také nízkou provozní cenou v porovnání s ostatními přístroji (33, 38).

Společnost Illumina také nabízí v rámci základního softwaru přístup ke cloudovému prostředí Illumina BaseSpace (<https://basespace.illumina.com/home/sequence>), který umožňuje přímé online ukládání dat ze sekvenátoru, jejich sdílení v rámci pracovní skupiny i jednoduché nástroje pro kontrolu a základní zpracování získaných dat. Pro *de novo* sekvenování, kde bychom byli částečně omezeni délkou fragmentů, lze naopak s výhodou

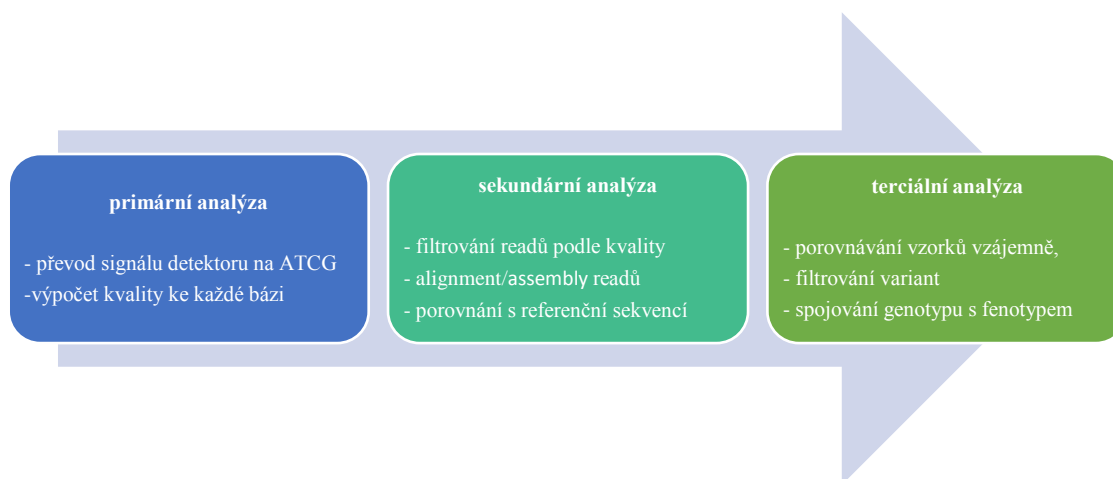
využít například platformu 454 Life technologies, která poskytuje průměrnou délku čtení přibližně 700 bází, případně některou z metod třetí generace sekvenování. Největší přesnosti lze dosáhnout při využití platformy Solid. Na některých pracovištích je tak vhodné kombinovat několik NGS platform. (33, 38).

Tab. 2: Porovnání parametrů dostupných NGS platform (70)

Sekvenátor	Princip sekvenování:	Délka čtení (bp)	Počet čtení:	Délka sekvenačního běhu:	Propustnost/běh:	Uváděná přesnost (%):	Přibližná pořizovací cena (\$):
3730xl	Sangerova metoda	400-900	1	20 min – 3 hod	1,9-84 Kb	99,999	95 000
HiSeq 2500	Sekvenace syntézou	36/50/100	3×10^9	2-11 dní	600 Gb	>99	740 000
MiSeq	Sekvenace syntézou	25/36/100/150/250	$1,7 \times 10^7$	4-27 hodin	8,5 Gb	>99	125 000
454 GS FLX+	Pyrosekvenování	700	1×10^6	23 hodin	0,7 Gb	99,997	450 000
454 GS Junior	Pyrosekvenování	400	1×10^6	10 hodin	0,035 Gb	>99	108 000
5500 Solid	Ligace	75+35	$1,4 \times 10^9$	7 dní	90 Gb	99,99	350 000
5500xl Solid	Ligace	75+35	$2,8 \times 10^9$	7 dní	180 Gb	99,99	595 000
Ion Personal Genome Machine (PGM)	Sekvenace syntézou/ Protonová detekce	35/200/400	$1,2 \times 10^7$	2 hodiny	2 Gb	>99	80 000
Ion Proton Chip I/II	Sekvenace syntézou/ Protonová detekce	>200	$6-8 \times 10^7$	2 hodiny	10 Gb/100 Gb	>99	243 000
PacBio RS	Jednomolekulové sekvenování v reálném čase (SMRT)	~ 3000	5×10^4	2 hodiny	13 Gb	84-85	750 000
gridION	Nanoporové exonukleasové sekvenování	Desítky kB	0,4- 1×10^7	Podle experimentu	Desítky Gb	96	-

5. ANALÝZA SEKVENAČNÍCH DAT

Jednou z nejsložitější částí NGS procesů, ne-li tou vůbec nejsložitější, je vyhodnocení získaných sekvenačních dat. Jejich objem je dnes bohužel často tak velký, že převyšuje kapacitní možnosti laboratorních provozů a část dat tak někdy zůstane bez dostatečného vyhodnocení, natož interpretace. Množství dat klade velké nároky na bioinformatiky, kteří musí v rámci analýzy sekvenačních dat úzce spolupracovat s laboratorními i klinickými pracovníky molekulárně genetických pracovišť. Analýzu dat získaných sekvenováním nové generace můžeme rozdělit na 3 fáze, které jsou znázorněny na obrázku 10. Jelikož data pro praktickou část této diplomové práce byla získána na sekvenátorech společnosti Illumina, bude v této části popsána především analýza dat na této platformě při znalosti referenční sekvence.

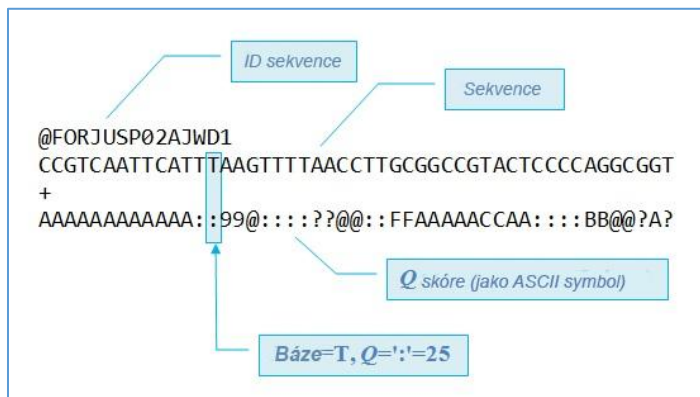


Obr. 10 : Postup při analýze dat získaných sekvenováním nové generace

5.1 Primární analýza – převod signálu na data

Primární analýzu dat zajišťuje samotný sekvenátor, v němž dochází k převedení světelného či chemického signálu detektoru na čtyřpísmenný kód a k výpočtu kvality pro získaná data. Tato data mohou být ukládána lokálně na pevný disk, připojený server s diskovým polem či jsou odesílána na síťový server a jsou tak přístupná prostřednictvím webového rozhraní. Výstupem ze sekvenátoru Illumina HiSeq jsou fotografie průtokové komůrky ve vysokém rozlišení, které jsou na připojeném serveru převáděny na soubory ve formátu FASTQ. Výstupem ze sekvenátoru Illumina MiSeq je již přímo soubor ve FASTQ

formátu. Soubor ve formátu FASTQ má standardní čtyřřádkovou strukturu, která je znázorněna na obrázku 11.



Obr. 11: Schéma struktury FASTQ souboru s využitím kódování Illumina v1.8 (40, převzato a upraveno)

Řádek 1: Začíná symbolem @ a obsahuje přesné informace o sekvenci, včetně použitého přístroje a umístění klastru na sekvenační komůrce.

První řádek tak může mít například tvar:

@EAS141:123:FB716VJ:2:5:1000:12850 1:Y:18:ATCACG

EAS141	Unikátní identifikace přístroje
123	Číslo běhu přístroje
FB716VJ	Identifikace průtokové komůrky
2	Číslo dráhy na komůrce
5	Číslo dlaždice v dráze na komůrce
1000	x-souřadnice klastru
12850	y-souřadnice klastru
1	Pouze při párovém čtení, označuje číslo readu 1 či 2
Y	Pokud byl read filtrován Y=yes (ano), N=no (ne)
18	Počet kontrol, 0 znamená žádná kontrola
ATCACG	Sekvence indexu

Řádek 2: Samotná sekvence ve formátu IUPAC nomenklatury

(„...ACTGCTGAAATTCGATCCA...“)

Řádek 3: Začíná symbolem + a může obsahovat další informace o sekvenci.

Řádek 4: Obsahuje informaci o kvalitě přečtení dané báze - **Phred Quality Scores (Q)**.

Tato hodnota se vypočítá podle vzorce $Q = -10 \log_{10}(P)$ a většinou dosahuje hodnot 1 až 41. Hodnota P značí pravděpodobnost, se kterou byla daná báze přečtena chybně. Pokud je tedy například hodnota Q rovna 3, potom je 50%

šance, že byla báze přečtena špatně. Pokud $Q = 0$, potom $P = 1$, což znamená, že báze byla s jistotou špatně přečtena. Tento případ je však raritní záležitostí a v praxi se nejčastěji setkáme s nejnižší hodnotou $Q = 2$ ($P = 0,63$). Aby bylo možné pro zápis kvality použít pouze jednu řádku, je zvoleno nahrazení číslic ASCII kodem (39). Od zavedení hodnocení Illumina verze 1.8 je použito kódování Phred+33. To znamená, že symboly pro označení Q jsou od klasického ASCII kódu posunuty o 33 pozic. U starších verzí bylo použito kódování Phred+64 s posunutím o 64 pozic. Rozdíl v kódování oběma způsoby je znázorněn na obrázku 12.

Sanger, Illumina v1.3 to 1.7 (ASCII_BASE=64)											
Q	ASCII	P	Q	ASCII	P	Q	ASCII	P	Q	ASCII	P
1	A	0.79433	12	L	0.06310	23	W	0.00501	34	b	0.00040
2	B	0.63096	13	M	0.05012	24	X	0.00398	35	c	0.00032
3	C	0.50119	14	N	0.03981	25	Y	0.00316	36	d	0.00025
4	D	0.39811	15	O	0.03162	26	Z	0.00251	37	e	0.00020
5	E	0.31623	16	P	0.02512	27	[0.00200	38	f	0.00016
6	F	0.25119	17	Q	0.01995	28	\	0.00158	39	g	0.00013
7	G	0.19953	18	R	0.01585	29]	0.00126	40	h	0.00010
8	H	0.15849	19	S	0.01259	30	^	0.00100			
9	I	0.12589	20	T	0.01000	31	~	0.00079			
10	J	0.10000	21	U	0.00794	32	`	0.00063			
11	K	0.07943	22	V	0.00631	33	a	0.00050			

Illumina v1.8 and later (ASCII_BASE=33)											
Q	ASCII	P	Q	ASCII	P	Q	ASCII	P	Q	ASCII	P
1	"	0.79433	12	-	0.06310	23	8	0.00501	34	C	0.00040
2	#	0.63096	13	.	0.05012	24	9	0.00398	35	D	0.00032
3	\$	0.50119	14	/	0.03981	25	:	0.00316	36	E	0.00025
4	%	0.39811	15	0	0.03162	26	;	0.00251	37	F	0.00020
5	&	0.31623	16	1	0.02512	27	<	0.00200	38	G	0.00016
6	'	0.25119	17	2	0.01995	28	=	0.00158	39	H	0.00013
7	(0.19953	18	3	0.01585	29	>	0.00126	40	I	0.00010
8)	0.15849	19	4	0.01259	30	?	0.00100	41	J	0.00008
9	*	0.12589	20	5	0.01000	31	@	0.00079			
10	+	0.10000	21	6	0.00794	32	A	0.00063			
11	,	0.07943	22	7	0.00631	33	B	0.00050			

Obr. 12: Převodní tabulka pro ASCII kódování využívané pro hodnocení kvality Illumina v1.3 a v1.8 (39)

5.2 Sekundární analýza – filtrování a mapování readů

Data získaná NGS sekvenováním obsahují chyby, které mohou být způsobeny buď přípravou knihoven či samotným sekvenováním. Kvalitu získaných dat je možné a dokonce vhodné zkontrolovat některým ze zdarma dostupných nástrojů. Tím nejpoužívanějším je zřejmě nástroj FastQC, který po importování FASTQ souboru umožňuje kromě základních statistik i grafické znázornění několika kvalitativních parametrů sekvenovaného úseku. Mezi ty nejdůležitější patří Phred kvalita bází na každé pozici readu, přítomnost adaptérů, obsah GC či přítomnost duplikovaných sekvencí. Na obrázku 13 vidíme typický příklad rozložení kvality čtení bází v jednotlivých readech v programu FastQC.

Tím, že před sestavováním sekvence ze získaných readů vyfiltrujeme ready o špatné kvalitě, snížíme možnost zkreslení výsledků. Z vybraných readů, jejichž kvalita neklesne pod určitou stanovenou mez, můžeme pomocí softwarových nástrojů rekonstruovat původní sekvenci. Alignment neboli mapování readů je součástí analýzy dat získaných tzv. resekvenováním. Tedy sekvenování, při němž známe referenční sekvenci, s níž naši sekvenci porovnááme. Sekvenování více jedinců stejného druhu navíc umožňuje mimo porovnání s referenční sekvencí také porovnání jedinců mezi sebou. Jakmile se nám podaří ready poskládat do kompletní sekvence, můžeme hledat variační rozdíly, jakými jsou například SNPs.

Na rozdíl od sestavování původní sekvence po *de novo* sekvenování, se při resekvenování jedná o poměrně jednoduchý proces, který má však řadu úskalí. Tím prvním je, že délka referenční sekvence v případě sekvenování lidské DNA může být např. i 3×10^9 bází, zatímco délka našich readů se pohybuje podle použité platformy průměrně asi kolem 300 bází. Je tedy velká šance, že podobných úseků, kam by náš read mohl patřit, najdeme na referenční sekvenci hned několik. Tato pravděpodobnost se ještě zvyšuje v případě repetitivních sekvencí. Navíc, abychom odhalili variační rozdíly oproti referenční sekvenci, je nutné povolit určitou nepřesnost mezi oběma sekvencemi. Stejně tomu je i u chybovosti sekvenování, kdy musíme tolerovat určitou chybovost čtení a pokusit se odlišit ji od skutečné varianty genu. Přiřazování readů k referenční sekvenci je znázorněno na obrázku 15.

CACGTTGTGTCGCCTCACTACGGTGGTGAGGTGACGCTTA	Referenční sekvence
ACGTTGTGCCGCTCACTACGGTGGTG	Read č. 1
CGTTGTG T CGCCTCACTACGGTGGTGA	Read č. 2
GTTGTGCCGCTCACTACGGTGGTGAG	Read č. 3
CGTTGTGCCGCTCACTACGGTGGTGAG	Read č. 4
GTTGTG T CGCCTCACTACGGTGGTGAG	Read č. 5
GTTGTGCCGCTCACTACGGTGGTGAG	Read č. 6
CACGTTGTGCCGCTCACTACGGTGGTGA	Read č. 7

Obr. 15: Přiřazování (alignment) readů k referenční sekvenci

Sedm krátkých readů bylo přiřazeno k referenční sekvenci. I přesto, že 2 ready se kompletně shodují s referenční sekvencí, ostatních 5 se liší od referenční sekvence v jednom nukleotidu. Lze tak předpokládat, že se v tomto případě jedná o SNV.

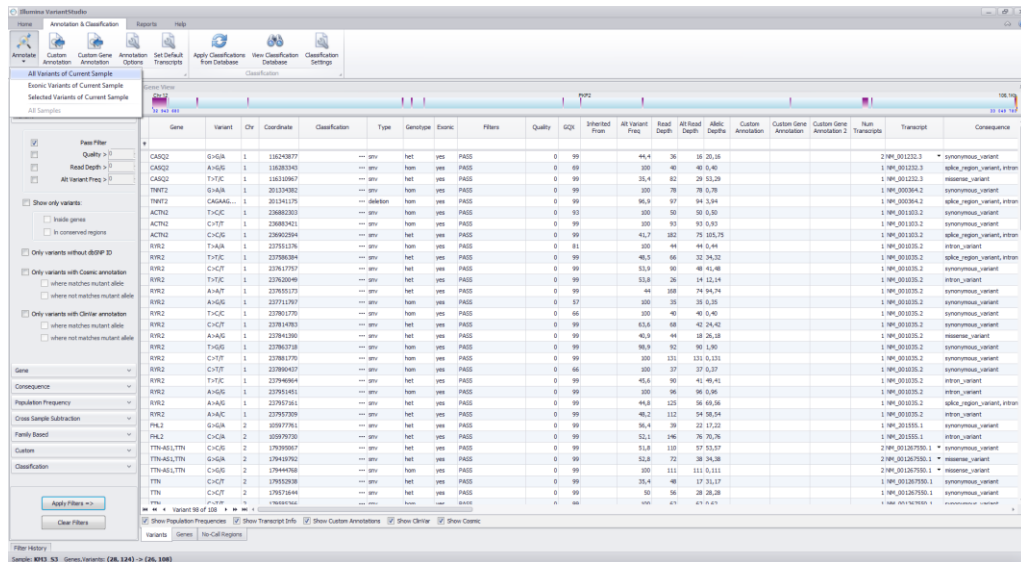
Poté, co se nám podaří sestavit původní sekvenci vyšetřovaného vzorku, je dalším krokem sekundární analýzy detekce variant, tedy odchylek od sekvence referenční. Tyto odchylky mohou být na úrovni jednotlivých bází (inzerce, delece, substituce) nebo se může jednat o větší změny (translokace, inverze či CNV). Typickým výstupem sekundární

analýzy je soubor typu VCF. VCF formát byl vytvořen v rámci projektu 1000 Genomes (<http://www.1000genomes.org/>) jako způsob uchování informací o variantách a je od té doby považován za standardní způsob ukládání těchto dat. VCF soubory obsahují text ve formátu ASCII, který je rozdělen na dvě hlavní části. První částí souboru je hlavička, která obsahuje informaci o verzi VCF, definované filtry a další technické údaje. Za ní následuje samotná tabulka s nalezenými variantami.

5.3 Terciální analýza – spojení genotypu s fenotypem

Díky této poslední fázi analýzy dat lze propojit data popisující pacientův genotyp s jeho fenotypem pomocí analytických a genetických nástrojů. Terciální analýza již není tolik náročná, jako analýzy předchozí a lze ji tak provádět na výkonnějších osobních počítačích. V této poslední fázi analýzy dat by na základě předchozích kroků mělo dojít k interpretaci zanalyzovaných dat, tedy činit závěry vyplývající z daného experimentu. V podstatě se jedná o porovnání více sekvenovaných vzorků a hlubší analýzu získaných sekvencí. Jelikož se proces liší podle aplikace, nemůže být nijak standardizovaný. Jinak budeme postupovat při *de novo* sekvenacích a jinak při resekvenování, při němž můžeme porovnávat získané sekvence se sekvencí referenční. Jedním z programů, který pracuje se soubory typu VCF je Illumina Variant Studio (Illumina, USA; obrázek 16).

V tomto programu jsou po vložení souboru s daty určitého pacienta v tabulce vypsány nalezené varianty, jejich typ (insece, delece či snv), genotyp (heterozygot-het, homozygot-hom, případně somatická mutace-som), označení, pozice na genu, frekvence alely v různých populacích a další důležité informace. Po naimportování následuje provedení filtrace nalezených variant. V záložce „Variants“ lze podle uvážení nastavit filtr podle nedostatečné kvality či hloubky čtení. Případně použít automatickou filtraci zvolením „Pass filter“, která provede filtraci podle výrobcem přednastavených parametrů kvality jako je frekvence minor alely, hloubka čtení a kvalita sekvenace. Po filtraci dat je možné přistoupit k „annotaci“. Tato funkce vyhledá a přiřadí známé informace ke každé z nalezených variant včetně odkazu na databázi variant (<http://www.ncbi.nlm.nih.gov/projects/SNP>).



Obr. 16: Prostředí aplikace Illumina Variant Studio 2.2.1

Bioanalytici, kteří zpracovávají obrovské množství dat, ať už z celogenomového, exomového či cíleného sekvenování panelu genů, si musí být dostatečně vědomi řady úskalí, které s sebou masivní paralelní sekvenování přináší. Ty nejdůležitější můžeme shrnout do čtyř následujících bodů (43):

- 1) Počítat s možností nekompletního pokrytí sekvenovaného úseku, které může způsobit falešně negativní výsledek analýzy
- 2) Znat možné zdroje falešně pozitivních výsledků a pokusit se o jejich omezení
- 3) Identifikovat možné kauzální mutace mezi stovkami až tisíci nalezených variant
- 4) Využití rodokmenů a genotypových informací při analýze příbuzných

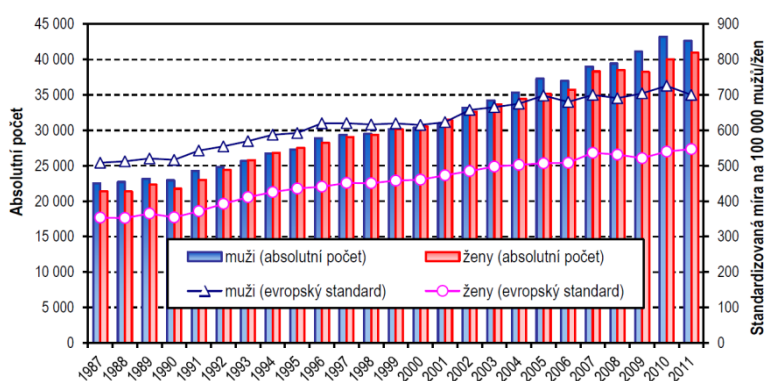
6. APLIKACE NGS V KLINICKÉ PRAXI

Přístroje na principu masivního paralelního sekvenování za 10 let své existence našly uplatnění na řadě klinických pracovišť a usnadnily molekulárně genetickou diagnostiku nejen postnatální, ale také prenatální a dokonce preimplantační. Mimo jiné jsou využívány v oblasti farmakogenetiky, výzkumu a stanovování prognózy u nádorových onemocnění, detekci chromozomálních aberací či k vyšetření monogenních chorob. Klasická molekulárně genetická diagnostika se v případě prevalentních mutací opírá především o vyšetření specifických úseků nukleové kyseliny metodami, jako je restriční analýza či alelově specifická PCR. V případě, že je ve vyšetřovaném genu pravděpodobnost výskytu privátní mutace, je vhodné zvolit sekvenační analýzu celého vyšetřovaného úseku. Pro detekci rozsáhlejších delecí či přestaveb je vhodné ještě doplnění o MLPA metodu (Multiplex Ligation Probe Amplification). Ideálním postupem při vyšetřování mutací je však analýza DNA spolu s cDNA, která vzniká přepisem RNA reverzní transkriptázou (44).

Příčinou toho, že cDNA není ve většině případů vyšetřována je nejen finanční i manuální náročnost vyšetření, ale také prostá skutečnost, že RNA se v případě některých genů vyskytuje pouze v určitých tkáních a není tak snadné a často ani možné cDNA získat. Jak ukázaly mnohé studie, současné NGS technologie dokáží nahradit nejen současnou sekvenační technologii, ale i většinu dnes používaných technik, včetně MLPA za kratší čas a při nižších nákladech. Pomocí nového sekvenování již byla navíc objasněna řada diagnóz, u nichž nebylo možné do té doby dostupnými metodami odhalit kauzální mutace. Byla také prokázána možnost NGS metod detekovat chromozomální mozaiky, nezachytitelné metodou a-CGH či QFPCR ve vzorku tkáňové DNA (46-49). V následující části jsou popsány příklady jen malé skupiny onemocnění, pro jejichž výzkum a diagnostiku jsou již dnes významným přínosem.

6.1 Nádorová NGS diagnostika

Nádorové onemocnění vzniká na podkladě nahromadění germinálních i somatických mutací v genetickém materiálu. Tyto mutace vedou ke vzniku nádoru svým vlivem na geny ovlivňující buněčné dělení či reparaci DNA. Oblast onkologie a hematookologie nabízí zřejmě největší klinickou oblast pro využití NGS metod. Podle statistik Ústavu zdravotnických informací a statistik (ÚZIS) se u nás zvýšil počet nově hlášených zhoubných novotvarů od roku 1987 do roku 2011 ze 43 913 na 83 581 případů (51). Vývoj incidence na našem území je znázorněn na obrázku 17.

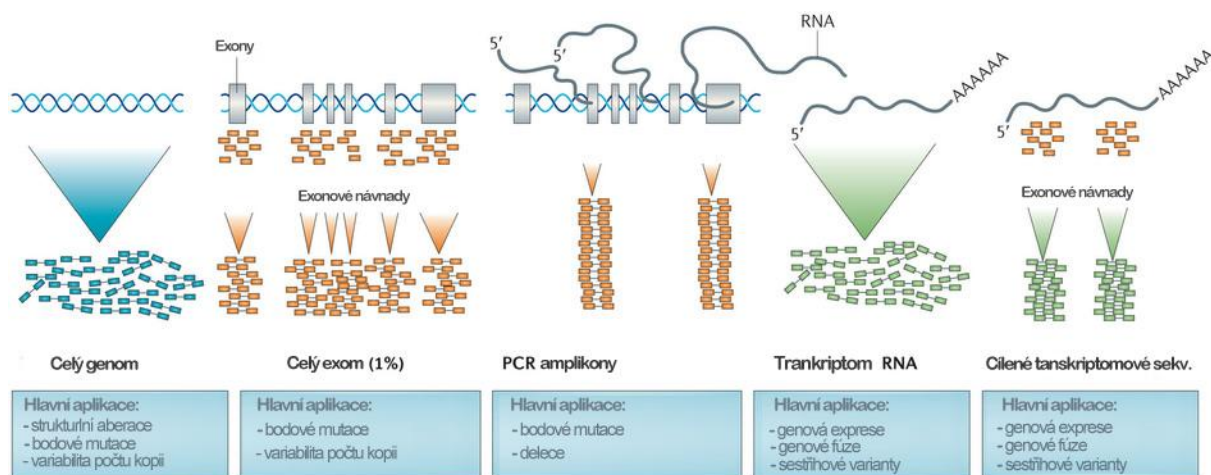


Obr. 17: Vývoj incidence zhoubných novotvarů na území ČR mezi lety 1987-2011(51)

Svou roli na vzestupném trendu má samozřejmě dostupnost nových vyšetřovacích metod a technologických postupů, které jsou schopny zachytit novotvary v časnějším stádiu. Spolu se vzestupem nově diagnostikovaných případů zhoubných novotvarů se zvyšuje i poptávka po dostatečně účinných laboratorních metodách časně diagnostiky, které jsou schopny zachytit onkologické onemocnění ještě v latentním stádiu či v případě molekulárně genetické diagnostiky identifikovat jedince s dědičnou predispozicí k nádoru. Pro tyto ohrožené jedince pak zvolit cílenou prevenci a včasnou léčbu (51, 53).

Pomocí klinických studií využívajících nové metody sekvenování byla objevena řada nových kauzálních genů, nové techniky pomáhají také v objasnění mechanismů kancerogeneze a umožňují komplexnější charakterizaci nádorových onemocnění na molekulární úrovni. V rámci NGS diagnostiky nádorových onemocnění je využívána řada postupů, které jsou znázorněny na obrázku 18. Jednou z nejdůležitějších vyšetřovacích technik při komplexní diagnostice nádorového onemocnění je transkriptomové sekvenování. Tento přístup k NGS sekvenování již umožnil identifikovat řadu genových fúzí, které jsou

spojovány s procesem kancerogeneze ve tkáni, a to např. u nádorů tlustého střeva, zhoubného melanomu, lymfomu či nádoru prostaty. Při detekci somatických mutací hraje zásadní roli citlivost metody, jelikož vzorek často obsahuje směs buněk postižených mutací spolu s ostatními buňkami, které touto změnou postiženy nebyly a obsahují zárodečný genom v nezměněné podobě (15, 45).



Obr. 18: Přehled NGS metod užívaných v onkogenetice (49, převzato a upraveno)

TruSight Cancer kit je jedním z nabízených komerčních kitů pro diagnostiku dědičně podmíněných nádorových onemocnění platformou Illumina. Byl vyvinut ve spolupráci s londýnským Institutem pro výzkum rakoviny a testuje 94 genů a 284 jednonukleotidových polymorfismů, u nichž byla předchozími studiemi prokázána souvislost s nádorovým onemocněním. Seznam testovaných genů tímto panelem ukazuje tabulka 3.

Tab. 3: Seznam genů vyšetřovaných pomocí kitu TruSight Cancer (Illumina, USA)

<i>AIP</i>	<i>CDH1</i>	<i>EPCAM</i>	<i>FANCE</i>	<i>KIT</i>	<i>PALB2</i>	<i>RET</i>	<i>SUFU</i>
<i>ALK</i>	<i>CDK4</i>	<i>ERCC2</i>	<i>FANCF</i>	<i>MAX</i>	<i>PHOX2B</i>	<i>RHBDF2</i>	<i>TMEM127</i>
<i>APC</i>	<i>CDKN1C</i>	<i>ERCC3</i>	<i>FANCG</i>	<i>MEN1</i>	<i>PMS1</i>	<i>RUNX1</i>	<i>TP53</i>
<i>ATM</i>	<i>CDKN2A</i>	<i>ERCC4</i>	<i>FANCI</i>	<i>MET</i>	<i>PMS2</i>	<i>SBDS</i>	<i>TSC1</i>
<i>BAP1</i>	<i>CEBPA</i>	<i>ERCC5</i>	<i>FANCL</i>	<i>MLH1</i>	<i>PRF1</i>	<i>SDHAF2</i>	<i>TSC2</i>
<i>BLM</i>	<i>CEP57</i>	<i>EXT1</i>	<i>FANCM</i>	<i>MSH2</i>	<i>PRKAR1A</i>	<i>SDHB</i>	<i>VHL</i>
<i>BMPR1A</i>	<i>CHEK2</i>	<i>EXT2</i>	<i>FH</i>	<i>MSH6</i>	<i>PTCH1</i>	<i>SDHC</i>	<i>WRN</i>
<i>BRCA1</i>	<i>CYLD</i>	<i>EZH2</i>	<i>FLCN</i>	<i>MUTYH</i>	<i>PTEN</i>	<i>SDHD</i>	<i>WT1</i>
<i>BRCA2</i>	<i>DDB2</i>	<i>FANCA</i>	<i>GATA2</i>	<i>NBN</i>	<i>RAD51C</i>	<i>SLX4</i>	<i>XPA</i>
<i>BRIP1</i>	<i>DICER1</i>	<i>FANCB</i>	<i>GPC3</i>	<i>NF1</i>	<i>RAD51D</i>	<i>SMAD4</i>	<i>XPC</i>
<i>BUB1B</i>	<i>DIS3L2</i>	<i>FANCC</i>	<i>HNFI1A</i>	<i>NF2</i>	<i>RB1</i>	<i>SMARCB1</i>	
<i>CDC73</i>	<i>EGFR</i>	<i>FANCD2</i>	<i>HRAS</i>	<i>NSD1</i>	<i>RECQL4</i>	<i>STK11</i>	

6.1.1 Kolorektální karcinom (OMIM #114500)

Nádory tlustého střeva a konečníku jsou třetím nejčastějším nádorovým onemocněním na světě a jsou také nejčastější příčinou úmrtí na nádorová onemocnění. Česká republika pravidelně obsazuje přední místa statistik v rámci četnosti tohoto onemocnění. Ročně je u nás diagnostikováno 7 800–8 100 pacientů s kolorektálním karcinomem a 3 800–4 200 pacientů tomuto onemocnění podlehnou. Včasné odhalení premalignit, jakými jsou například střevní polypy či provádění screeningových vyšetření typu testu na okultní krvácení, pochopitelně snižují riziko rozvinutí tohoto nádorového onemocnění, stále však existují případy, kdy je vznik nádoru nezjistitelný a pokročilá stadia onemocnění stále nejsou bohužel léčitelná. Rozvoj kolorektálního karcinomu je multifaktoriální proces, na němž se mimo genetických vlivů podílí také vliv životního stylu, stravy, prostředí a další (51,53). Díky NGS technologiím došlo k výraznému pokroku v charakterizaci genetických a epigenetických změn, které vypovídají o velkém významu genetické predispozice pro toto onemocnění.

Na základě rozsáhlých celogenomových a exomových studií bylo prokázáno, že zhruba u 16 % případů kolorektálního karcinomu je přítomna somatická hypermutace. Jedná se o stav, kdy je počet genetických mutací enormně vysoký z důvodu narušení opravných mechanismů DNA. Asi u $\frac{3}{4}$ z nich byla prokázána mikrosatelitní nestabilita (MSI), při níž dochází k mutaci v genech odpovídajících za opakující se úseky DNA, následkem čehož dochází ke zkrácení či prodloužení těchto úseků. U zbylé jedné čtvrtiny případů jsou nalézány mutace v mismatch-repair genech a v *POLE* genu pro polymerázu ϵ . Transkriptomovým sekvenováním byla u pacientů identifikována řada genových fúzí. Mimo jiné rekurentní genová fúze genů *RSPO2* a *RSPO3*, která může potencovat Wnt signalizaci a mít tak zásadní úlohu v kancerogenezi či fúze genů *NAV2-TCF7L1*. Výrazné množství somatických mutací (přes 36 tisíc) bylo nalezeno ve 24 genech. Kromě očekávaných genů *APC*, *TP53*, *SMAD4*, *PIK3CA* a *KRAS*, jsou nejčastěji nalézány také mutace v genech *ARID1A*, *SOX9* či *FAM123B* (55, 57).

NGS analýza také dokáže v nádorové tkáni detekovat přítomnost cizorodých sekvencí DNA pocházejících z bakteriálních či virových patogenů. I když byla v minulosti prokázána souvislost mezi patogenem a vznikem určitého nádorového onemocnění, u kolorektálního nádoru začala být tato možnost zkoumána teprve nedávno. Především je zkoumána onkogenní role anaerobních bakterií *Fusobacterium nucleatum*, které jsou ve velkém množství

nacházeny ve tkáni tlustého střeva pacientů s kolorektálním karcinomem. Předpokládá se, že tato bakterie má vliv na prvotní neoplastické změny ve střevní tkáni. Do budoucna se očekává, že NGS technologie umožní objasnění také v této oblasti (55, 56).

6.1.2 Nádory prsu a ovária (OMIM #114480, #167000)

Karcinom prsu je nejčastějším diagnostikovaným nádorovým onemocněním u žen v rozvinutých zemích (51). Zavedení včasné diagnostiky nádoru pomocí mamografického vyšetření sice vedlo k výraznému snížení úmrtnosti na toto onemocnění, stále je však úmrtnost na tyto nádory značně vysoká a velké naděje jsou vkládány do zlepšeného genetického screeningu. Pomocí genetického screeningu pacientek s nádorem prsu dnes bývá nalezena mutace pouze u 5-10 % případů s hereditárním výskytem. U řady postižených rodin tedy nebývá kauzální genetická příčina odhalena. Za posledních 20 let se molekulární diagnostika pacientek s nádory prsu a vaječníku soustředila převážně na vyšetření nejčastěji zasažených genů *BRCA1* a *BRCA2*. Bylo prokázáno, že ztrátové mutace v těchto genech vedou ke zvýšení rizika bilaterálního nádoru a často přidruženého nádoru vaječníků, přičemž kumulativní riziko pro vznik nádoru prsu je odhadováno na 84 % do věku 70 let. U karcinomu vaječníků je riziko odhadováno na 63 % při výskytu mutace v *BRCA1* genu a 27 % při výskytu mutace v *BRCA2* genu (59, 60).

Nové sekvenční metody umožňující rozsáhlé genetické studie však u pacientů s nádory prsu a vaječníků kromě mutací v genech *BRCA* našly také časté mutace v dalších genech jako např. *PALB2*, *ATM*, *CHEK2*, *NBT*, *PTEN* či *MSH2*, které ukazují nejen genetickou heterogenitu, ale i na skutečnost, že až 10 % nádorů, které jsou dnes klasifikovány jako sporadické, může mít velmi výrazný genetický podklad. Velký význam má testování nových genů u pacientek, které mají v rodinné anamnéze nádory prsu či ovárií a nebyla u nich nalezena mutace v *BRCA1/2* genech. Například ve studii, při níž bylo testována 141 *BRCA1/2* negativních pacientek pomocí panelu 42 genů byly u 15 z nich (11 %) nalezeny zárodečné mutace v genech *ATM*, *BLM*, *CDH1*, *CDKN2A*, *MUTYH*, *MLH1*, *NBN*, *PRSS1* a *SLX4* (61). Řada nových genů, které jsou nalézány u pacientek s tímto typem rakoviny, by se tak mohla do budoucna stát součástí panelů pro screeningové genetické testování žen v riziku a některé společnosti již nabízejí testování 10-20 kauzálních genů. V současné době však nemáme k dispozici dostatek podkladů, které by u řady z těchto genů potvrdily jejich přímý vliv na

vznik nádoru prsu či ovária. Výsledky studií však umožňují komplexnější pohled na tato onemocnění a na možnou cílenou terapii (62).

6.2 Hereditární kardiomyopatie

Tato skupina chorob postihuje srdeční svalovinu, přičemž dochází k postupnému narušení funkce celého srdce. Jde o skupinu život ohrožujících onemocnění, jejich kauzální diagnostika je nesnadná, jelikož prvním příznakem může být srdeční selhání či náhlá srdeční smrt (SCD, Sudden Cardiac Death). V řadě případů je tak diagnostika onemocnění prováděna až post mortem. Také proto se dostala tato skupina onemocnění do zájmu vědců i klinických pracovníků a jsou vyvíjeny nové postupy laboratorní diagnostiky, které by identifikovaly co největší množství rodin s dědičnou formou kardiomyopatie. V rámci kardiomyopatií již bylo také díky NGS identifikováno více než 80 různých asociovaných genů a tyto nové technologie tak konečně přinášejí výkonný nástroj ke zlepšení prevence a léčby těchto onemocnění u dětí i dospělých. Pro kvalitní diagnostiku je však nutná spolupráce kardiologických klinik s genetickými pracovišti při výběru pacientů pro záchyt kauzálních mutací a v případě pozitivního nálezu také možnost genetického poradenství a testování dalších členů rodiny (58). Právě rodiny, u nichž se předpokládá genetická predispozice pro některou z forem kardiomyopatie, byly také vybrány pro praktickou část této diplomové práce. K nejčastějším typům hereditárních kardiomyopatií patří hypertrofická kardiomyopatie (HCM), dilatační kardiomyopatie (DCM) či arytmogenní dysplázie pravé srdeční komory (ARVD). Výhodou při testování cíleného panelu genů u této skupiny onemocnění je skutečnost, že mutace u jednotlivých typů kardiomyopatií se často nacházejí ve stejných genech.

6.2.1 Hypertrofická kardiomyopatie (HCM, OMIM #192600)

Toto onemocnění se projevuje hypertrofií srdce s abnormálním uspořádáním svaloviny, při němž dochází ke ztluštění komorového septa, případně dalších srdečních stěn a ke zvětšení cípů u mitrální chlopně. To bývá doplněno fibrózou a vznikem abnormálních ložisek v myokardu. Až 70 % pacientů bývá asymptomatických a prvním příznakem onemocnění u nich může být náhlá srdeční smrt, která ohrožuje až 20 % pacientů. U pacientů s příznaky jsou nejčastěji popisovány pocity dušnosti, palpitace, angina pectoris či drobné poruchy vědomí. V současné době je prevalence tohoto onemocnění v České republice 0,2 %

(tedy cca 20 000 nemocných s HCM. U 50-60 % pacientů jsou dnes identifikovány příčinné mutace. Dodnes bylo v souvislosti s tímto onemocněním popsáno více než 1 400 mutací v nejméně 12 genech. Nejčastěji se jedná o mutace v genech kódujících těžký myosinový řetězec, vazebný myosinový protein C a T₆ Troponin (63). Frekvenci výskytu mutací v jednotlivých genech ukazuje tabulka 4.

Tab. 4: Frekvence výskytu mutací v genech pro proteiny při HCM (63)

Mutovaný protein:	Frekvence výskytu mutace:
těžký řetězec β -myozinu	20-25 %
vazebný protein C	15-20 %
troponin T	3-5 %
troponin I	1-2 %
α -tropomyosin	1-2 %
reg. podjednotka lehkého řetězce myozinu	< 1 %
ostatní	< 1 %

6.2.2 Dilatační kardiomyopatie (DCM, OMIM #115200)

Pro tuto chorobu je typická dilatace srdečních komor, čímž dochází k narušení jejich fyziologické funkce. Jedná se o poměrně časté onemocnění s prevalencí 1: 2 500. V současné době převažuje názor, že onemocnění není přímo geneticky podmíněno a roli v jeho rozvoji mají i jiné faktory, jakými jsou například viry. Přesto je predispozice k tomuto onemocnění odhadována na 20-35 % případů. Bylo vytipováno velké množství (>20) genů podezřelých z účasti na vzniku tohoto onemocnění, z nichž nejvíce pozornost vědců upoutaly geny ze skupiny cytoskeletálních a sarkomerických proteinů. K nim patří například titin, který je kódován genem *TTN* na chromozomu 2, lamin A/C kódovaný genem *LMNA* na chromozomu 1 či dystrofin kódovaný genem *DMD* na chromozomu X. Bylo také prokázáno, že mutace v genu pro lamin A/C výrazně zvyšuje riziko SCD (58, 64).

6.2.3 Arytmogenní kardiomyopatie pravé komory (ARVD, OMIM #609040)

Toto onemocnění primárně postihuje pravou srdeční komoru, až u 70 % pacientů však dochází k postižení také levé komory. To vede k silné kardiomegalii, častějším arytmiickým příhodám a vyššímu výskytu srdečního selhání. ARVD je geneticky podmíněné onemocnění s nejčastěji autozomálně dominantní formou dědičnosti, popsány jsou však i případy dědičnosti autozomálně recesivního typu. Tento typ přenosu je spojen s etnickými variacemi a byl pozorován např. na řeckém ostrově Naxos, kde je ARVC spojena s palmoplantární

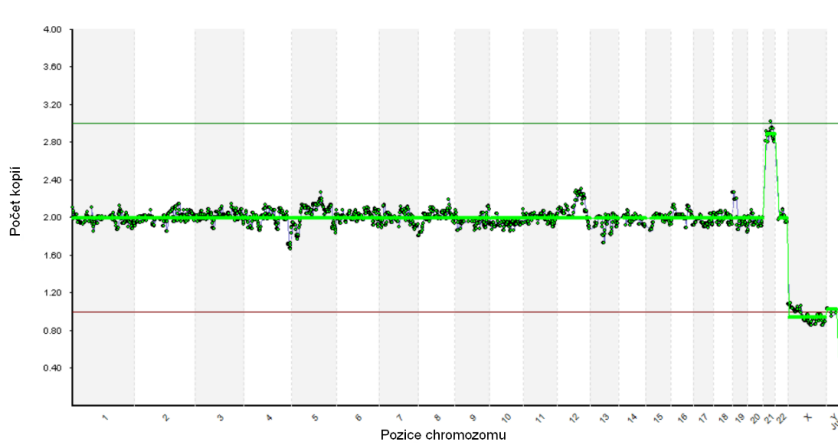
keratózou a autozomálně recesivní dědičností (Naxon disease). Jedná se o delecii v *JUP* genu pro kódování plakoglobinu, který je umístěn na 17. chromozomu. Plakoglobin je protein, který je součástí buněčných spojů desmozomů, které mají význam pro buněčnou adhezi v různých orgánech. V případě srdeční svaloviny tak dochází k narušení spojení mezi jednotlivými kardiomyocyty. Mezi další geny, které jsou dávány do spojitosti s ARVD se řadí *PKP2*, *DSG2*, *DSP* či *DSC2* (65).

6.3 Preimplantační genetická diagnostika (PGD-NGS)

Velké ambice má NGS diagnostika v metodách asistované reprodukce, zejména preimplantační genetické diagnostice (PGD). Ta má sloužit párům, u nichž reálně hrozí přenos vrozené chromozomální aberace na potomka, ve výběru vhodného embrya, jehož genetický kód není touto aberací poznamenán. Preimplantační diagnostika se od té prenatalní liší tím, že zatímco u prenatalní diagnostiky vyšetřujeme plod až v průběhu těhotenství, v případě PGD je embryo vyšetřeno před samotnou implantací do dělohy a lze tak selektovat ta embrya, u nichž je genetická vada potvrzena. Odpadá tak stres a zdravotní komplikace spojené s případným umělým přerušением těhotenství. K samotnému vyšetření se využívá 1 buňka (blastomera) či několik buněk, které jsou získány biopsií trofoektodermu (66).

NGS již několik let v rámci klinických studií ukazuje, že je vhodným nástupcem klasických metod pro vyšetření monogenních onemocnění i chromozomálních odchylek v preimplantační diagnostice. Například studie provedená v roce 2013 na 38 darovaných blastocytách ukázala vysokou přesnost NGS technologií. Tato studie byla zaměřena na detekci aneuploidií a nevyvážených chromozomových přestaveb pomocí sekvenátoru Illumina HiSeq 2000. NGS platforma byla porovnávána s mikročipovou analýzou. Obě tyto metody dokázaly shodně identifikovat všech 12 embryí s chromozomovými aberacemi. Navíc však NGS metoda dokázala určit i částečné aneuploidie u dvou embryí, které se mikročipovou analýzou odhalit nepodařilo. V porovnání s čipovou analýzou bylo dosaženo dvojnásobné četnosti vyšetřovaných úseků a jako jediná nevýhoda se ukázala délka celého vyšetření. Doba přípravy knihovny a sekvenování totiž trvá několik dní. Po tuto dobu je nutné embryo kryokonzervovat. Nové sekvenační techniky však již umožňují sekvenaci zkrátit až na několik málo hodin a nic tak nebrání rychlému transferu čerstvých embryí bez nutnosti jejich zamražení (66).

VeriSeq PGS kit (Illumina, USA) je zástupcem PGD-NGS kitů, který umožňuje snadnou diagnostiku chromozomálních aberací v rámci všech chromozomů včetně submikroskopických delecí a duplikací na úrovni jedné buňky. K reakci stačí pouze 1 ng DNA. Ta je získána celogenomovou amplifikací pomocí SurePlex DNA Amplification System, který umožňuje získat až 5 µg amplifikované DNA za 2 hodiny. Do amplifikační reakce je přidána 1 buňka (blastomera) či několik buněk, které jsou získány biopsií trofoektodermu (67). Jelikož lze kitem vyšetřovat jedinou buňku, lze teoreticky vyšetřit i samotnou pohlavní buňku, tedy spermii či oocyt, což již bylo v minulosti vyzkoušeno. Praktické využití testování jednotlivých gamet je však zatím mizivé, jelikož při vyšetření je tato buňka zničena a nelze ji tak použít k umělému oplodnění. Výhodou tohoto testu je i jednoduchost vyhodnocení. Jak je vidět na obrázku 19, výstupem analýzy je na rozdíl od jiných NGS metod jednoduchý graf znázorňující počet kopií jednotlivých chromozomů.



Obr. 19: Výsledek vyšetření DNA kitem VeriSeq u pacienta s trizomií chromozomu 21 (67)

6.4 Cystická fibróza (CF, OMIM #217900)

Cystická fibróza je závažné a poměrně časté autozomálně recesivní onemocnění u kavkazské populace. Její incidence se na našem území dlouhodobě uvádí 1: 2 700 - 1: 3300 narozených dětí (77). Na základě nových dat získaných novorozeneckým screeningem však byla nově incidence stanovena na 1: 4 500 dětí (78). Samotné onemocnění cystickou fibrózou je podloženo dvěma mutacemi v genu *CFTR* (z angl. Cystic Fibrosis Transmembrane Conductance Regulator) v pozici trans. Tento gen se skládá z 27 exonů, rozprostírá se přibližně v oblasti 250 kb a tvoří mRNA dlouhou 6,5 kb (68). Do října 2014 bylo v *CFTR* genu popsáno 1 983 mutací a sekvenčních variant. Hlavní aplikací NGS v diagnostice

monogenních onemocnění jakým je právě cystická fibróza, jsou případy, při nichž pacient vykazuje klinické projevy onemocnění, avšak rutinně používanými postupy včetně Sangerova sekvenování nebylo možné nalézt očekávané mutace. Pro tyto případy jsou již na trhu validované kity pro rutinní klinickou diagnostiku (CE-IVD), které umožňují cílenou sekvenaci jediného genu.

Jedním ze zástupců těchto kitů je i CFTR MASTR Dx (Multiplicon, Belgie), který umožňuje vyšetření všech kódujících exonů s asi 30 přilehlými intronovými sekvencemi, důležitých hlubokých intronových sekvencí a část promotorů *CFTR* genu. K reakci je potřeba pouze 20-50 ng DNA. Kit je validován pro rutinní diagnostiku na stolních sekvenátorech typu MiSeq (Illumina, USA), 454 GS Junior a 454 FLX+ (454 Life Sciences). Tento kit zatím nezískal validaci na analýzu variability počtu kopií segmentů DNA (CNV). Jelikož samotné snížení či absence sekvence při čtení konkrétního ampliconu je samo o sobě nedostatečně průkazné, musí být případná delece potvrzena alternativními metodami (69). Tento kit je v současnosti testován i v Národní referenční laboratoři pro cystickou fibrózu ve FN Motol v Praze s možností rutinního prenatálního i postnatálního testování cystické fibrózy na přístroji MiSeq.

7. ETICKÉ OTÁZKY SPOJENÉ S NGS METODAMI

Žijeme nyní v době, kdy se NGS technologie stávají rutinními vyšetřovacími metodami a zcela jistě není daleko doba, kdy se ještě díky dalšímu vylepšení a zlevnění současných platforem bude rutinně sekvenovat i celý lidský genom. Již od samotných příprav Projektu lidského genomu však panují obavy ze zneužití získaných informací a sekvenování lidského genomu vyvolává i po řadě let mnoho eticko-právních a sociálních otázek. Ta první se týká zacházení se získanými daty. Nové sekvenační technologie získávají a uchovávají obrovské množství dat, která jsou často sdílena v rámci národních i mezinárodních vědeckých projektů a je tak na místě obava z jejich možného zneužití. Možnost, že na základě genetické identifikace dojde k propojení zdravotních a sociálních informací, které by měly negativní dopad na testovaného jedince v osobním životě, vyvolává jednu z hlavních otázek. Tedy tu, kdo by měl mít k získaným datům přístup. Prvním stupněm, kterým by měla být zajištěna ochrana osobních údajů pacienta je bezpochyby informovaný souhlas s vyšetřením. Tím pacient dává nejen souhlas s vlastním vyšetřením, ale vymezuje také pravidla, na základě nichž je možno nakládat s informacemi získanými během vyšetřování jeho genetické informace (79).

Pro Českou republiku vydává doporučení týkající se informovaného souhlasu Společnost lékařské genetiky ČLS JEP. V současnosti jsou platná doporučení týkající se informovaného souhlasu pro genetická laboratorní vyšetření aktualizovaná dne 22. 5. 2013 (79). Mimo ochrany osobních údajů však také pacient může v rámci informovaného souhlasu poskytnout svůj genetický materiál k výzkumným účelům. Každý účastník takového výzkumu má pak právo získávat také informace z jeho výsledků a zde narážíme především u exomových a celogenomových studií na jeden podstatný problém. Na začátku výzkumu lidského genomu si řada lidí představovala, že rozluštěním lidského genetického kódu získá lidstvo jednou pro vždy nástroj, jak diagnostikovat a léčit dědičná onemocnění ještě před jejich vznikem a stanovovat životní prognózy právě na základě rozdílů v DNA jednotlivců. Bohužel skutečnost je jiná a jak jsme se mohli přesvědčit z mnoha prováděných studií, je dnes například nacházeno mnoho genetických variant, kterou vedou jen k velmi mírnému zvýšení rizika vzniku určitého onemocnění. U častých multifaktoriálních onemocnění jakým je například diabetes se genetický vliv odhaduje pouze na několik procent.

A právě nevhodná interpretace získaných výsledků může být příčinou zbytečné emoční zátěže pacienta, která sama o sobě může vést v konečném důsledku ke zhoršení jeho zdravotního stavu. Přehnaný důraz na významnost nalezených genetických variant v souvislosti s vývojem klinického stavu pacienta, může uvést pacienta v názor, že ho s největší pravděpodobností čeká boj se závažným onemocněním. To může napáchat neodčinitelné škody nejen v jeho psychice. Naopak jakýsi přehnaný pocit bezstarostnosti vyplývající z jeho pocitu, že není nositelem určité genetické predispozice, může vést dokonce až ke vzniku či prohlubování rizikového chování, jakým jsou různé závislosti na alkoholu či tabáku. V rámci genetické konzultace je tak nezbytně nutné představit pacientovi široké spektrum okolností, za nichž jsou studie prováděny a důkladně vysvětlit informace, které z výzkumu vyplývají přímo pro pacientův aktuální klinický stav.

Velmi důležité jsou také etické aspekty prenatalního testování a testování dětí. Je otázkou, jak a koho informovat v případě nálezu genetické predispozice u plodu či dítěte a klinický genetik stojí před rozhodnutím, který ze tří postupů zvolit. Může tuto informaci sdělit ihned rodině testovaného dítěte, sdělit tuto informaci dítěti po dovršení plnoletosti či nesdělit tuto informaci vůbec. Vždy by se však měl řídit tím, aby případná sdělená informace měla pro příjemce co největší benefit a současně, aby bylo minimalizováno nebezpečí, které z této informace plyne. Pokud je například nalezena mutace v kauzálním genu pro vznik nádorového onemocnění, tato informace má vliv na možnou prevenci a léčbu jak potomka, tak i rodičů, kteří mohou být přenašeči mutace a měla by být sdělena v rámci genetického poradenství. Opačná situace nastává v případě, že je nalezena kauzální příčina onemocnění, které není možné ovlivnit prevencí, ani terapií či onemocněním jakým je například porucha pozornosti dítěte. V tomto případě by měl klinický genetik velmi dobře uvážit, zda je tato informace pro rodiče a jejich potomka přínosná, či pouze zatěžující.

8. PRAKTICKÁ ČÁST

Všechny metody popsané v této části diplomové práce byly prováděny na Oddělení molekulární genetiky, které je součástí Ústavu biologie a lékařské genetiky 2. LF UK a FN Motol (dále jen ÚBLG). Do konce roku 2013 bylo na ÚBLG vyšetřováno více než 90 dědičných onemocnění především pomocí kapilárního sekvenování. Již několik let byla vybírána nová platforma, která by umožnila částečnou a postupnou náhradu klasické sekvenační metody a současně by umožnila výzkum na fakulním pracovišti, které je také součástí největší nemocnice ve střední Evropě. Nakonec byla zvolena platforma společnosti Illumina a v roce 2014 byly zakoupeny dva sekvenátory nové generace. Prvním přístrojem byl stolní sekvenátor MiSeq, jehož význam by měl spočívat v rychlé a levné diagnostice cílených genových panelů či samostatných genů a do něhož se vkládá naděje postupného nahrazení kapilární sekvenace. Druhým přístrojem byl HiSeq 1500, který by měl sloužit převážně výzkumným účelům a největší roli by měl plnit v exomovém sekvenování. Právě na těchto přístrojích byla provedena praktická část této diplomové práce, která zahrnuje přípravu DNA knihoven a její optimalizaci na tomto konkrétním pracovišti, sekvenování na přístrojích MiSeq a HiSeq 1500 (obrázek 25) a pokus o analýzu získaných dat.

8.1 Materiál a pomůcky

K experimentální části byly ve spolupráci s Kardiologickou klinikou Fakultní nemocnice Motol a IKEM vybrány vzorky pro detekci zárodečných mutací u pacientů s diagnózou hereditární kardiomyopatie. Celkem bylo vybráno 191 pacientů. Z toho bylo 65 pacientů vyšetřeno cíleným panelem TruSight Cardiomyopathy, který umožňuje analýzu 46 kandidátních genů. U 126 vzorků ze 42 rodin, u nichž byl zaznamenán výskyt hereditární kardiomyopatie, byl vyšetřen tzv. klinický exom (mendeliom) pomocí kitu TruSight One. Mimo pacientů s hereditární kardiomyopatií byla také vyzkoušena účinnost sekvenování celého exomu na skupině 8 pacientů z 3 rodin pomocí kitu Nextera Rapid Capture Exome (všechny kity Illumina, USA). K sekvenaci byly použity NGS přístroje MiSeq a HiSeq 1500 (Illumina, USA)

TruSight Cardiomyopathy kit byl vyvinut ve spolupráci s Laboratory for Molecular Medicine a Partners Healthcare Center for Personalized Genetic Medicine, Harvard Medical School. Obsahuje 46 genů spojených s hereditární kardiomyopatií uvedených v tabulce 5. Kit

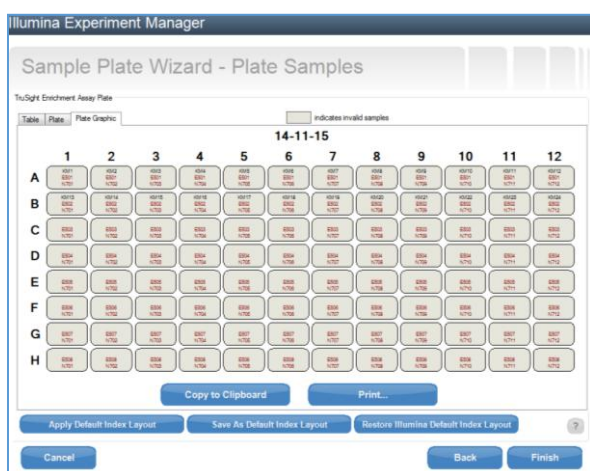
TruSight One umožňuje vyšetření genů, u nichž byla popsána souvislost s genetickým onemocněním (mendeliom) o délce přibližně 12 Mb a kit Nextera Rapid Capture Exome umožňuje vyšetření všech kódujících oblastí DNA o délce asi 37 Mb.

Tab. 5: Seznam genů vyšetřovaných kitem TruSight Cardiomyopathy (Illumina, USA)

<i>ABCC9</i>	<i>CSRP3</i>	<i>EMD</i>	<i>LMNA</i>	<i>MYOZ2</i>	<i>SGCD</i>	<i>TPM1</i>
<i>ACTC1</i>	<i>CTF1</i>	<i>FHL2</i>	<i>MYBPC3</i>	<i>NEXN</i>	<i>TAZ</i>	<i>TTN</i>
<i>ACTN2</i>	<i>DES</i>	<i>GLA</i>	<i>MYH6</i>	<i>PKP2</i>	<i>TCAP</i>	<i>TTR</i>
<i>ANKRD1</i>	<i>DSC2</i>	<i>JUP</i>	<i>MYH7</i>	<i>PLN</i>	<i>TMEM43</i>	<i>VCL</i>
<i>CASQ2</i>	<i>DSG2</i>	<i>LAMA4</i>	<i>MYL2</i>	<i>PRKAG2</i>	<i>TNNC1</i>	
<i>CAV3</i>	<i>DSP</i>	<i>LAMP2</i>	<i>MYL3</i>	<i>RBM20</i>	<i>TNNI3</i>	
<i>CRYAB</i>	<i>DTNA</i>	<i>LDB3</i>	<i>MYLK2</i>	<i>RYR2</i>	<i>TNNT2</i>	

8.2 Příprava experimentů

Jednotlivé experimenty byly navrženy pomocí softwaru Illumina Experiment Manager v 1.8 (IEM, Illumina, USA), který umožňuje rozvržení vzorků a použitých indexů, které jsou během přípravy knihovny připojeny k naštěpeným fragmentům DNA a umožňují tak identifikaci jednotlivých vzorků. Každý použitý kit obsahuje 2x index i5 a 12x index i7, např. indexy E501, E502 a indexy N701-N712. Jejich kombinací lze tak duálně označit maximálně 48 vzorků. V IEM je pak připraveno správné kombinování jednotlivých indexů pro vyšetřované vzorky a výstupem je grafické znázornění pozic vzorků v pipetovacím plátíčku spolu s příslušnými indexy (obr. 20). Druhým výstupem z programu je Sample Sheet, v němž před sekvenací byly zadány informace o použitých chemikáliích a tento soubor byl importován přímo do počítače, který je součástí sekvenátoru.



Obr. 20: Příprava experimentu k sekvenaci 24 vzorků v programu Illumina Experiment Manager

8.3. Izolace a ředění genomické DNA

Koncentrace a čistota vstupní DNA hrají zásadní roli v získání kvalitní knihovny pro následné sekvenování. DNA pro přípravu knihoven byla izolována z 5 ml čerstvé periferní krve odebrané do K₃EDTA. K izolaci byl použit automatický izolátor Flex Star (AutoGen, USA). Přístroj umožňuje izolaci z 1-10 ml plné krve i jiných biologických materiálů. Pracuje ve čtyřstupňovém režimu: lýza a odstranění erytrocytů, lýza leukocytů, denaturace a odstranění proteinů, precipitace, purifikace a rozpouštění DNA. Maximálně kapacita přístroje činí 40 vzorků/cyklus. Délka izolačního cyklu je cca 3 hodiny a výtěžnost tímto přístrojem se pohybuje mezi 100–210 µg DNA. Čistota DNA reprezentovaná poměrem 260/280 je u takto izolovaných vzorků 1,7 – 1,9.

Izolovaná DNA byla po izolaci postupně ředěna na koncentraci 100 ng/µl, 10 ng/µl až na finální koncentraci 5 ng/µl. Aby bylo zamezeno křížové kontaminaci, bylo veškeré ředění prováděno v laminárním boxu se zapnutým odsáváním vzduchu. K ředění byl použit 10 mM roztok Tris-HCl o pH 8,5. Pro změření koncentrace 100 ng/µl byl použit UV spektrofotometr Nanodrop 2000 (Thermo Fisher Scientific, USA), který zároveň umožňuje detekci případných kontaminantů, jakými jsou soli či proteiny a to měřením při vlnových délkách 230, 260 a 280 nm. Na základě hodnot získaných měřením spektrofotometrem Nanodrop 2000 byla DNA opět naředěna, tentokrát již na koncentraci 10 ng/µl. V takto nízké koncentraci již byla DNA měřena citlivějším přístrojem Qubit 2.0 (Invitrogen, USA) pracujícím na principu fluorometrie. Na základě tohoto přesného měření byla DNA dále naředěna na finální koncentraci 5 ng/µl s povolenou odchylkou +/- 10 %. Tato maximální odchylka byla zvolena na začátku všech experimentů, aby nedocházelo k velkým reakčním nepoměřům mezi jednotlivými vzorky. Citlivost přístrojů Nanodrop 2000 a Qubit 2.0 byla před prvním experimentem testována na skupině 10 vzorků (tabulka 6).

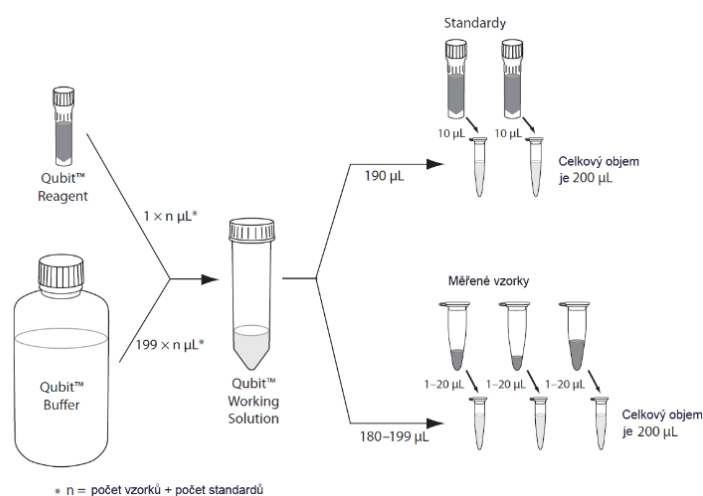
Tab. 6: Porovnání měření koncentrace DNA na přístrojích NanoDrop 2000 a Qubit 2.0

Vzorek	Nanodrop 2000 (ng/µl)	Qubit 2.0 (ng/µl)	Odchylka (ng/µl)	Odchylka (%)
1	11,2	6,8	4,4	64,71
2	10,8	5,3	5,5	103,77
3	9,7	5,1	4,6	90,20
4	12,5	6,2	6,3	101,61
5	8,9	6,7	2,2	32,84
6	10,3	7,5	2,8	37,33
7	9,6	5,8	3,8	65,52
8	11,3	7,0	4,3	61,43
9	12,0	6,3	5,7	90,48
10	8,9	5,9	3,0	50,85
Průměr			4,3	69,90

8.3.1 Práce s přístrojem Qubit 2.0

Přístroj Qubit 2.0 je velmi citlivým přístroj pro kvantifikaci DNA, RNA a proteinů. Pro měřenou složku je na rozdíl od jiných kvantifikačních přístrojů natolik specifický, že znemožňuje interferenci s jinými složkami v roztoku. Pro snadné ovládání je přístroj vybaven dotykovou grafickou obrazovkou a výstupem měření je číselná hodnota spolu s jejím grafickým vyjádřením na kalibrační křivce získané měřením standardů. Naměřenou hodnotu lze pomocí přístroje automaticky převést na požadované jednotky. V přípravě NGS knihoven byl přístroj využíván k měření kvantity DNA před samotnou přípravou a dále ke kontrole úspěšnosti jednotlivých reakčních kroků v průběhu přípravy a před samotnou sekvenací. Pro měření DNA přístrojem Qubit lze použít dva různé kity, které umožňují měření v různých koncentračních rozpětích. Qubit® dsDNA BR Assay Kit umožňující měření DNA v koncentračním rozmezí 100 pg/μl až 1 000 ng/μl a Qubit® dsDNA HS Assay Kit umožňující měření v koncentracích 10 pg/μl až 100 ng/μl.

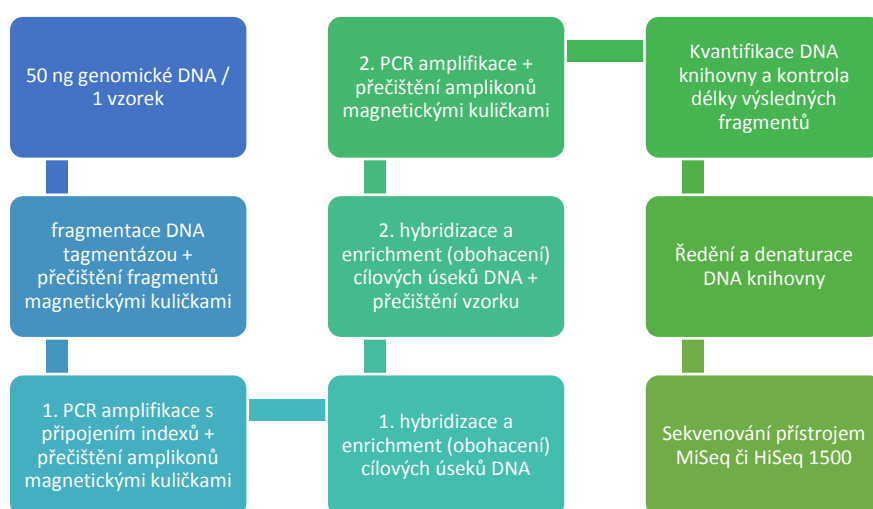
Prvním krokem při měření je příprava Qubit Working Solution (QWS) v objemu 200 μl / vzorek. Ta se připraví smísením 199 μl Qubit Buffer (QB) a 1 μl HS či BR Qubit Reagent (QR) pro 1 vzorek. Připravený QWS se rozpipetuje do speciálních jednorázových zkumavek, které umožňují fluorescenční měření a je k němu přidávána měřená DNA. Objem měřené DNA lze zvolit v rozmezí 1-20 μl, u standardů je objem zvolen přesně uprostřed tohoto rozmezí, tedy 10 μl. Tento objem se odečte od výsledného objemu 200 μl tak, že v našem případě byly ředěny 2 μl DNA se 198 μl QWS, u standardů bylo ředěno 10 μl standardu se 190 μl QWS. Postup přípravy je znázorněn na obrázku 21.



Obr. 21: Schéma přípravy vzorků pro měření na přístroji Qubit 2.0

8.4 Příprava DNA knihovny

Příprava DNA knihoven byla prováděna podle standardního protokolu dodávaného s kity TruSight a Nextera Rapid Capture pouze s tím rozdílem, že v několika případech bylo nutné provést úpravy pro získání optimálních výsledků, které jsou popsány v poznámce daného kroku. Celý postup přípravy je znázorněn na obrázku 22 a doba přípravy knihovny pro jeden experiment (max. 48 vzorků) trvala 4 dny. V rámci přípravy DNA knihovny jsou začleněny 3 kontrolní body, během nichž je kontrolována kvalita a kvantita připravované DNA.



Obr. 22: Postup přípravy DNA knihovny kity TruSight a Nextera Rapid Capture

8.4.1. Fragmentace genomické DNA a přečištění fragmentů

V tomto prvním kroku přípravy knihovny byla genomická DNA štěpena na kratší fragmenty pomocí hyperaktivní varianty Tn5 transposázy. DNA byla štěpena na fragmenty v rozmezí 150 bp – 1 kb. V optimálním případě by největší zastoupení měly mít fragmenty o velikosti kolem 250 bp. Následné přečištění je důležité pro odstranění transpozomů, které by mohly rušit následné reakce. DNA fragmenty byly při něm připojeny na magnetické kuličky, omyty čerstvě připraveným 80% etanolem a po omytí opět z kuliček eluovány do roztoku.

8.4.1.1 Postup tagmentace DNA a čištění fragmentů

Do odpovídajících pozic v platíčku bylo napipetováno po 10 μl genomické DNA v předem připravené koncentraci 5 ng/ μl \pm 10 % (celkem 50 ng v 1 reakci). Ke genomické DNA bylo přidáno 25 μl pufru Tagment DNA Buffer, 7,5 μl Tagment DNA enzymu a 7,5 μl dH₂O. Platíčko bylo zalepeno ochranou folií a obsah byl promísen na vortexu po dobu 1 minuty při 1 800 rpm. Následně bylo platíčko stočeno v centrifuze při 280 g po dobu 1 minuty a zahříváno v termocykleru 10 minut při teplotě 58 °C. Po skončení zahřívání byla reakce ukončena přidáním 15 μl Stop Tagment Bufferu. Platíčko bylo opět zalepenou folií a směs promísená na vortexu při 1 800 rpm po dobu 1 minuty. Nakonec bylo platíčko stočeno v centrifuze při 280 g 1 minutu a inkubováno při laboratorní teplotě po dobu 4 minut.

Ke každému tagmentovanému vzorku bylo přidáno 65 μl dobře rozpuštěné a propipetované směsi magnetických kuliček (Sample Purification Beads), platíčko bylo přelepeno ochranou folií, protřepáno při 1 800 rpm po dobu 1 minuty, inkubováno po dobu 8 minut při laboratorní teplotě a stočeno při 280 g 1 minutu. Po centrifugaci bylo platíčko na 2 minuty vloženo do magnetického stojánku. Magnety stojánku umožňují přichytit čistící kuličky k jedné straně jamky a je tak možné z opačné strany jamky pipetou odsát supernatant. Po odsátí veškeré kapaliny bylo ke kuličkám připipetováno 200 μl čerstvě naředěného 80% etanolu, který byl po 30 sekundách opět odsán. Celý postup omytí etanolem byl ještě 1x opakován. Po odsátí veškerého etanolu bylo platíčko ponecháno na magnetickém stojánku po dobu 10 minut při laboratorní teplotě, aby došlo k úplnému usušení magnetických kuliček s nachytanou DNA. Po usušení bylo platíčko sejmuto z magnetického stojánku, do každé jamky bylo přidáno 22,5 μl elučního roztoku Resuspension Buffer, následně bylo platíčko zalepeno, vortexováno 1 minutu při 1 800 rpm a inkubováno 2 minuty při laboratorní teplotě. Po inkubaci bylo centrifugováno 1 minutu při 280 g. Po centrifugaci bylo opět platíčko na 2 minuty vloženo na magnetický stojánek a 20 μl supernatantu bylo přeneseno do odpovídajících jamek v novém platíčku. Byla provedena kontrola přečištěné DNA pomocí čipové elektroforézy v přístroji 2100 Bioanalyzer (Agilent Technologies, USA).

8.4.1.2 Optimalizace kroku

Tento krok se v rámci opakované přípravy knihoven projevil jako nejproblematictější. I při dodržení všech kroků stanovených výrobcem kitu, nedocházelo při několika prvních pokusech k optimální fragmentaci DNA. Bylo zjištěno, že zásadní význam má počet otáček

třepačky, která mísí tagmentační enzym se vzorkem a při práci v platíčku nedocházelo i při standardním nastavení otáček k dostatečnému protřepání všech vzorků. Zvýšení otáček třepačky spolu s vizuální kontrolou vedlo ke zlepšení výsledku tagmentace. Po vyřešení dostatečného promísění vzorku s enzymem se dalším problémem ukázala délka výsledných fragmentů. Ty byly zpočátku kratší, než bylo potřeba, bylo tedy vyzkoušeno zkrácení doby štěpení z 10 na 5 minut. Toto zkrácení doby štěpení sice mělo tížený efekt, ovšem ne u všech vzorků a v rámci reprodukovatelnosti mezi jednotlivými experimenty se ukázalo jako nevhodné. Nakonec bylo vyzkoušeno naředění enzymu v poměru 1:1 dH₂O při zanechání standardní doby štěpení. Až tento postup se během několika experimentů ukázal jako nejvhodnější.

8.4.1.3 Postup měření DNA na přístroji Agilent 2100 Bioanalyzer

K měření DNA na přístroji Agilent 2100 lze použít dva typy čipů – čip s označením DNA HS pro měření 11 DNA vzorků o koncentraci 5-500 pg/μl a délce 50-7000 bp nebo čip s označením DNA 1000 pro měření 12 DNA vzorků o koncentraci 0,1-50 ng/μl a délce 25-1000 bp. V tomto kroku byl použit čip DNA 1000 a ke kontrole tagmentace bylo náhodně vybráno 12 připravovaných vzorků. Všechny chemikálie používané k měření jsou uchovávány v lednici při teplotě 4-8 °C a 30 minut před analýzou byly vyjmuty z lednice a vytemperovány na laboratorní teplotu. Při prvním experimentu musel být připraven roztok Gel-Dye Mix. Pro přípravu mixu je potřeba gel matrix (gel) a DNA dye concentrate (barva). Obě zkumavky byly protřepány na míchačce a krátce stočeny. Poté bylo 25 μl barvy připipetováno do zkumavky s gelem a obě látky byly promíseny na třepačce po dobu 10 sekund. Poté byl celý objem zkumavky přenesen na zkumavku s filtrem, která je součástí kitu a pomocí centrifugace při 2240 g po dobu 15 minut byl Gel-Dye Mix přefiltrován a připraven k použití.

Čip DNA 1000 byl vložen do pipetovací stanice, píst stříkačky byl nastaven na objem 1 ml. Do jamky označené **G** bylo napipetováno 9 μl připravené směsi Gel-Dye Mix. Stanice s čipem byla uzavřena a píst injektoru byl stlačen směrem dolů, dokud nebyl secvaknut svorkou. Po 1 minutě byl píst uvolněn ze svorky, 5 sekund po uvolnění byl píst manuálně vrácen na hodnotu 1 ml a pipetovací stanice otevřena. Gel byl napipetován do dalších 2 jamek označených G. Poté bylo do všech prázdných jamek napipetováno 5 μl markeru a do jamky označené symbolem žebříku byl napipetován 1 μl ladderu. Do číselně označených pozic byl

Tab. 7: Podmínky 1. PCR reakce v přípravě DNA knihovny

Krok PCR reakce:	Teplota:	Čas:	Počet cyklů:
Počáteční denaturace	98 °C	30 sek.	10x
Denaturace	98 °C	10 sek.	
Připojení primerů	60 °C	30 sek.	
Extenze	72 °C	30 sek.	
Konečná extenze	72 °C	5 min.	
Chlazení	10 °C	∞	

Po PCR amplifikaci opět následovalo přečištění vzorku od nežádoucích produktů. K tomu byly opět použity magnetické kuličky (Sample Purification Beads) a jejich omytí v čerstvě připraveném 80% etanolu na magnetickém stojánku podobně jako v předchozím kroku. Oba promývací kroky se liší pouze objemem přidávaných roztoků. V tomto kroku bylo přidáno 90 µl magnetických kuliček k 50 µl PCR amplikonům a v posledním kroku bylo ke každému vzorku připipetováno 27 µl Resuspension bufferu, z něhož bylo po eluci do nového plátíčka přepipetováno 25 µl eluátu přečištěných amplikonů. V případě, že nemůžeme z nějakého důvodu ihned pokračovat v přípravě knihovny, lze po přečištění vzorky zamrazit při teplotě -15 až -25 °C po dobu maximálně 14 dnů. V rámci přípravy knihoven pro tuto diplomovou práci byly v této fázi vzorky vždy zamraženy do druhého dne. Před zamražením či po něm byla změřena délka fragmentů a koncentrace DNA. K tomu byly použity přístroje 2100 Bioanalyzer a Qubit 2.0.

8.4.3 První hybridizace a obohacení

V tomto kroku dochází k hybridizaci knihoven se sondami cílenými ke specifickým oblastem DNA s navázaným biotinem. Ke správnému přisednutí sond je bezpodmínečně nutné dodržet teplotu a délku hybridizace. Ta je výrobcem u této první hybridizace stanovena v rozmezí 90 minut až 24 hodin, přičemž optimální délka se nachází v druhé polovině tohoto rozmezí. V rámci diplomové práce byly vzorky hybridizovány 18-20 hodin. Před samotnou hybridizací je však nutno ještě provést poolování vzorků, neboli jejich smísení tak, aby k hybridizaci bylo použito stejné množství DNA od každého vzorku. Po hybridizaci následuje enrichment neboli obohacení vzorku pomocí streptavidinových kuliček, na něž jsou navázány pouze cílené úseky a ostatní úseky jsou naopak odstraněny. Na streptavidin kuliček jsou biotinem přichyceny specifické próby připojené k cílovým oblastem a v následných dvou promývacích krocích jsou za zvýšené teploty odstraněny próby s nespecifickou vazbou. Po eluci z kuliček je pak knihovna připravena k druhé hybridizaci.

8.4.3.1 Postup poolování a první hybridizace

Při současném zpracování 12 - 48 vzorků byly vždy tvořeny 1-4 pooly po 12 vzorcích. Do každého poolu byl přidán objem každého vzorku odpovídající 500 ng DNA, celkový obsah poolu tak byl 6 000 ng DNA. Pro hybridizační reakci bylo potřeba zajistit, aby objem každého poolu byl 40 μ l, Proto v případě, že pool měl po smísení 12 vzorků objem překračující 40 μ l, byl tento objem snížen pomocí centrifugačního zakoncentrování. V opačném případě byl objem doplněn na 40 μ l pomocí Resuspension buffer.

Tab. 8: Příklad poolování 12 vzorků před první hybridizací

Označení poolu:	Číslo vzorku:	Koncentrace DNA (ng/ μ l):	Objem vzorku obsahující 500 ng DNA (μ l)	Objem poolu (μ l) :	Objem, který je potřeba odpařit (μ l)
A	KM1	103,5	4,8	61,3	21,3
	KM2	101,1	4,9		
	KM3	105,7	4,7		
	KM4	93,8	5,3		
	KM5	95,7	5,2		
	KM6	109,1	4,6		
	KM7	78,6	6,4		
	KM8	102,2	4,9		
	KM9	86,7	5,8		
	KM10	104,5	4,8		
	KM11	99,01	5,0		
	KM12	105,24	4,8		

Po zpoolování vzorků a dosažení objemu 40 μ l bylo ke každému poolu připipetováno 50 μ l Enrichment Hybridization Buffer a 10 μ l Content Set oligos (CSO), který obsahuje specifické próby lišící se podle experimentu. Následně byla celá směs promísena 1 minutu na třepačce při 1 200 rpm, stočena 1 minutu při 280 g a vložena do termocyklieru s přednastaveným teplotním programem – viz tabulka 9.

Tab. 9: Teplotní program hybridizace

Krok reakce:	Teplota:	Čas:	Počet cyklů:
Počáteční denaturace	95 °C	10 min.	
Denaturace	94 °C a pokles 2 °C/cyklus	1 min.	18x
Hybridizace	58 °C	90 min. – 24 hodin	

8.4.3.1 Postup přečištění a eluce

Každý pool byl ihned po vyjmutí z termocykleru a stočení přepipetován do 1,5 ml zkumavky a k němu bylo připipetováno 250 μ l streptavidinových magnetických kuliček (Streptavidin Magnetic Beads). S těmi byl každý pool promísen na třepačce po dobu 5 minut při 1 200 rpm. Následovala 25 minutová inkubace při laboratorní teplotě. Po inkubaci byly vzorky stočeny 1 minutu při 280 g a umístěny na 2 minuty na magnetický stojánek. Po usazení vzorků ke stěně zkumavky byl odsán supernatant a ke každému vzorku bylo připipetováno 200 μ l Enrichment Wash Solution. Po přidání byly vzorky protřepávány 4 minuty při 1 800 rpm a inkubovány 30 minut při teplotě 50 °C ve zkumavkovém inkubátoru. Po inkubaci byly vzorky na 2 minuty vloženy na magnetický stojánek a poté byl veškerý supernatant odsán.

Postup s 200 μ l Enrichment Wash Solution byl ještě jednou opakován a po odsátí supernatantu bylo ke každému vzorku přidáno 23 μ l eluční směsi, která byla pro jeden vzorek připravena v nadbytku z 28,5 μ l elučního pufru (Enrichment Elution Buffer 1) a 1,5 μ l 2N NaOH. Streptavidinové kuličky byly s eluční směsí protřepány 2 minuty při 1 800 rpm, vzorky inkubovány 2 minuty při laboratorní teplotě a následně stočeny 1 minutu při 1 200 rpm. Po centrifugaci byly vzorky vloženy po dobu 2 minut na magnetický stojánek a poté bylo 21 μ l každého vzorku přeneseno do nové popsané zkumavky. K 21 μ l přečištěného vzorku byly přidány 4 μ l neutralizačního pufru Elute Target Buffer 2. Vzorky byly s pufrem protřepány 2 minuty při 1 200 rpm a poté byl obsah zkumavky stočen 1 minutu při 1 200 rpm.

8.4.4 Druhá hybridizace a obohacení

Tento krok je v podstatě pouze zopakování kroku předchozího. Tato druhá hybridizace zajišťuje vysokou specifitnost cílených úseků. Také zde je nutné dodržet teplotu a dobu hybridizace, která je v tomto kroku 14,5 – 24 hodin. Po obohacení streptavidinovými kuličkami byla DNA knihovna přečištěna pomocí magnetických kuliček (Sample Purification Beads) a tím připravena k druhé PCR amplifikaci.

8.4.4.1 Postup druhé hybridizace a obohacení

Ke každému přečištěnému vzorku bylo přidáno 15 μ l Resuspension Buffer, 50 μ l Enrichment Hybridization Buffer a 10 μ l Content Set oligos (CSO) a opět byly vzorky vloženy do termocykleru s předvoleným programem (viz tabulka 9). Doba hybridizace byla u provedených experimentů 18-20 hodin. Po vyjmutí z termocykleru byly vzorky obohaceny pomocí streptavidinových kuliček a eluovány postupem popsáním v předchozím kroku. K 25 μ l knihovny bylo přidáno 45 μ l Sample Purification Beads, vzorky protřepány 1 minutu při 1 800 rpm a inkubovány 10 minut při laboratorní teplotě. Během této doby byl nově naředěn 80% etanol ze zásobního absolutního etanolu. Po inkubaci byly vzorky stočeny 1 minutu při 1 200 rpm a vloženy na magnetický stojánek na 2 minuty.

Po usazení kuliček byly na magnetickém stojánku opláchnuty 200 μ l etanolu, který byl po 30 vteřinách odsán a postup ještě jednou zopakován. Po dvojitém oplachu 80% etanolem byl veškerý etanol odsán a vzorky ponechány k usušení 10 minut při laboratorní teplotě. Po usušení vzorků byly sejmuty z magnetického stojánku, ke každému vzorku bylo přidáno 27,5 μ l Resuspension Buffer, vzorky protřepány 1 minutu při 1 800 rpm a inkubovány 2 minuty při laboratorní teplotě. Po inkubaci byly stočeny 1 minutu při 1 200 rpm a opět vloženy na magnetický stojánek. Po 2 minutách usazování kuliček bylo opatrně přeneseno 25 μ l přečištěného vzorku do nových zkumavek o objemu 0,2 ml k provedení druhé PCR reakce.

8.4.5 Druhá PCR reakce

V tomto kroku jsou cílové úseky DNA, které byly v předchozím kroku vychytány, amplifikovány pomocí 10-12 PCR cyklů. Během PCR reakce jsou připojeny sekvenční primery, kterými se jednotlivé fragmenty přichycují k průtokové komůrce sekvenátoru.

8.4.5.1 Postup druhé PCR reakce

Do každé zkumavky s 25 μ l knihovny bylo přidáno 5 μ l směsi primerů PCR Primer Cocktail a 20 μ l reakční směsi Enrichment Amplification Mix. Amplifikační zkumavky se vzorky byly promíseny 1 minutu na třepačce při 1 200 rpm a poté centrifugovány 1 minutu při 280 g. Po centrifugaci byly zkumavky vloženy do termocykleru s nastaveným PCR programem (tabulka 10). Po proběhlé PCR reakci byla PCR opět přečištěna pomocí

magnetických kuliček (Sample Purification Beads). Ke každému vzorku (50 µl) bylo přidáno 90 µl roztoku kuliček a zkumavky byly promíseny 1 minutu při 1 800 rpm. Poté byly vzorky inkubovány 10 minut při laboratorní teplotě, stočeny 1 minutu při 280 g a vloženy na 2 minuty na magnetický stojánek. Po usazení kuliček byl odsán supernatant a kuličky opatrně 2x opláchnuty 200 µl 80% etanolu. Po oplachu byl veškerý etanol odsán a vzorky ponechány 10 minut při laboratorní teplotě k úplnému odpaření zbytků alkoholu. Poté bylo ke vzorkům připipetováno 32 µl Resuspension Buffer, zkumavky protřepány 1 minutu při 1 800 rpm a inkubovány 2 minuty při laboratorní teplotě. Nakonec byly zkumavky stočeny 1 minutu při 280 g a vloženy na 2 minuty na magnetický stojánek. Po důkladném usazení všech magnetických kuliček bylo 30 µl hotové knihovny z každé zkumavky pipetou přeneseno do nové označené zkumavky a ihned dále zpracováno či zamraženo do dne sekvenování na teplotu -20 °C.

Tab. 10: Teplotní podmínky 2. PCR reakce při přípravě DNA knihovny

Krok PCR reakce:	Teplota:	Čas:	Počet cyklů:
Počáteční denaturace	98 °C	30 sek.	10-12x
Denaturace	98 °C	10 sek.	
Připojení primerů	60 °C	30 sek.	
Extenze	72 °C	30 sek.	
Konečná extenze	72 °C	5 min.	
Chlazení	10 °C	∞	

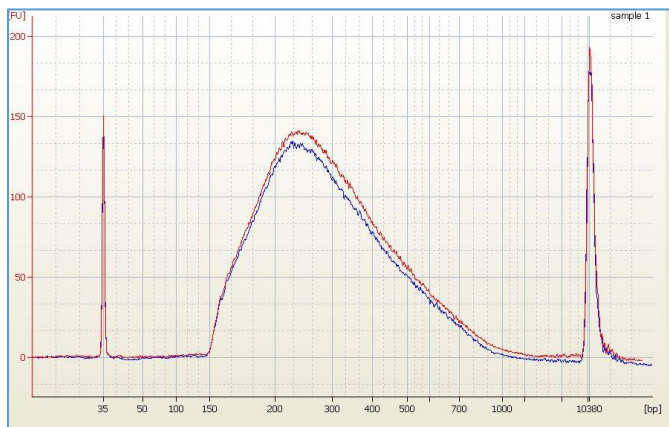
8.4.1.2 Optimalizace kroku

I přesto, že výrobce u tohoto kroku zmiňuje rizikovost snižování počtu amplifikačních kroků, byl tento počet po zkušenostech snížen při přípravě exomových knihoven na 10 cyklů z původně doporučených 12 cyklů.

8.5 Kvantifikace, ředění a denaturace finální DNA knihovny

Pro dosažení co nejvyšší kvality dat při sekvenování na platformě Illumina je nutné vytvořit podmínky, aby při tvorbě klastrů docházelo k rovnoměrnému a dostatečně hustému pokrytí promývací komůrky vzniklými klastry. Toho se docílí vhodnou kvantifikací připravené knihovny, převodem její koncentrace z ng/µl na molární koncentraci v nM a dostatečným naředěním na požadovanou finální koncentraci. Každý pool připravené knihovny tak byl kvantifikován pomocí přístroje Qubit 2.0 a na přístroji Bioanalyzer byla

změřena průměrná délka vzniklých fragmentů (obr. 24). Před měření přístrojem Bioanalyzer bylo dle doporučení výrobce použito ředění knihovny destilovanou vodou 1: 10, které se však ukázalo potřebné pouze u vzorků s koncentrací nad 10 ng/μl. Při nižších koncentracích získaných knihoven naopak mělo toto naředění nevhodný vliv při hodnocení výsledků elektroforézy z důvodu nízkého fluorescenčního signálu.



Obr. 24: Dva pooly finální knihovny o průměrné délce fragmentů 250 bp

Získané údaje byly použity do následujícího výpočtu:

$$\frac{\text{změřená koncentrace knihovny [ng/ul]}}{\text{molekulová hmotnost 1 bp [g/mol]} * \text{průměrná délka fragmentů [bp]}} \times 10^6 = \text{koncentrace knihovny [nM]}$$

Například tedy při průměrné délce fragmentů 250 bp a koncentraci 26,4 ng/μl bylo dosažením do rovnice získáno:

$$\frac{26,4 \text{ ng/}\mu\text{l}}{660 \text{ g/mol} * 250 \text{ bp}} \times 10^6 = \mathbf{160 \text{ nM}}$$

Výpočtové vzorce byly pro další experimenty uchovány a používány jako vzorce v programu Excel. Po výpočtu molární koncentrace byla knihovna ještě dále naředěna na koncentraci, která byla experimentálně zjištěna jako optimální a zdenaturována pomocí hydroxidu sodného. Jako optimální sekvenační koncentrace byla pro přístroj HiSeq zvolena koncentrace 20 pM a pro přístroj MiSeq 10 pM. V prvním kroku byl každý pool naředěn na přibližně stejnou koncentraci a následně byly pooly smíseny tak, aby pro HiSeq byla v max. 10 μl knihovna o koncentraci 2 nM a pro MiSeq v 5 μl knihovna o koncentraci 4 nM. K DNA knihovně byla přidána kontrolní DNA PhiX o koncentraci 2 nM pro HiSeq či 0,2 nM

pro Miseq v množství, které odpovídalo 2 % z koncentrace celé knihovny. Jednořetězcová DNA z fáze PhiX byla použita jako kontrola klastrování i chybovosti při sekvenování. V případě nižšího objemu celkové knihovny byl objem doplněn do 10 resp. 5 μ l destilovanou vodou. Následně bylo přidáno 10 μ l NaOH v koncentraci 0,1 N k HiSeq knihovně resp. 5 μ l NaOH o koncentraci 0,2 N k MiSeq knihovně. Knihovna s NaOH byla protřepána 1 minutu při 2 100 rpm a 5 minut denaturována při laboratorní teplotě na ssDNA. Následně bylo k 20 μ l zdenaturované HiSeq knihovny přidáno 980 μ l vychlazeného HT1 pufru, čímž došlo k naředění na koncentraci 20 pM a knihovna byla vložena do sekvenátoru či zamražena do dne sekvenace. K 10 μ l knihovny pro MiSeq bylo přidáno 990 μ l HT1 pufru a po promíslení byla knihovna ještě dále naředěna v poměru 300 μ l naředěné knihovny + 300 μ l HT1 pufru k dosažení koncentrace 10 pM. Poté byla knihovna sekvenována či zamražena.

8.6 Sekvenace na přístroji HiSeq 1500

Před spuštěním sekvenačního běhu byl vždy celý systém promyt destilovanou vodou (Water Wash) či po delším stání roztokem Maintenance Wash Solution (Maintenance Wash). Tím byly odstraněny zbytky chemikálií z předchozího běhu a současně případné vzduchové bubliny, které do systému vnikají během nasávání chemikálií. K tomuto promytí byly do chladicího prostoru sekvenátoru vloženy dvě cartridge s destilovanou vodou – menší cartridge s 10 ml zkumavkami bez víčka a větší cartridge s 200 ml lahvemi opatřenými děrovanými víčky. Z gumových hadiček vycházejících z přístroje byly dvě vloženy do separátních odměrných zkumavek, aby mohl být zkontrolován objem vody vypuštěné přístrojem při promývání. Ten by v ideálním případě měl činit 9,5 ml pro jednu zkumavku. Ke kontrole průtoku chemikálií systémem byla použita průtoková komůrka z předchozího běhu a průchodnost systému byla zkontrolována viditelným průchodem bublin oběma drahami komůrky v průběhu promývání.

V průběhu promývacího kroku byla ve vodní lázni rozmražena sekvenační a klastrovací chemie. U všech chemikálií byla zvážena jejich hmotnost na digitálních předvážkách a hodnoty zaznamenány do sekvenačního protokolu. Po rozmražení byly lahve a zkumavky s chemikáliemi vloženy do nových cartridge, v číselně i barevně odpovídajících pozicích. Po skončení promývacího kroku byl spuštěn sekvenační proces a dále bylo postupováno podle softwarové navigace. Promývací komůrka byla vyměněna za novou, jejíž čárový kód byl zaznamenán elektronickou čtečkou kódů. Nová komůrka byla vyjmuta

z pufru, opláchnuta destilovanou vodou, 70% etanolem a následně osušena a vyleštěna pomocí bezvláknového papírového ubrousku. Před vložením komůrky do systému byly také vyměněny obě gumičky (gaskety), které umožňují podtlakové přichycení komůrky. Obě cartridge s vodou v přístroji byly vyměněny za cartridge obsahující sekvenční a klastrovací chemii. V dalším kroku byl nahrán seznam vzorků s odpovídajícími indexy, který byl na začátku vytvořen v programu Illumina Experiment Manager a zadány přístupové údaje k online nástroji BaseSpace, kam jsou v průběhu sekvenačního runu odesílána data. Ve finálním kroku před samotným spuštěním sekvenace dochází k amplifikačnímu cyklu a k přečtení první báze. Kvalita celého sekvenačního procesu v této fázi závisí na hustotě pokrytí sekvenační komůrky vzniklými klastry. Přístroj kamerou snímá vzniklé shluky a porovnává jejich hustotu v obou dráhách průtokové komůrky. Ideální hodnota pokrytí je výrobcem uváděna mezi 500 tisíci až 1 milionem klastrů na mm^2 . Této hodnoty bylo u experimentů zpracovávaných v rámci této diplomové práce dosaženo.

8.7 Sekvenace na přístroji MiSeq

Práce na sekvenátoru MiSeq je oproti HiSeq značně zjednodušená. Celý systém byl před analýzou promyt roztokem Tween 20 o 0,5% koncentraci. Sekvenační chemie je uložena v reagenční cartridge, která byla před analýzou rozmrazena a do vyznačeného otvoru bylo napipetováno 600 μl naředěné DNA knihovny o koncentraci 10 μM . Poté byla cartridge vložena do přístroje spolu s lahví pufru. Průtoková komůrka byla omyta destilovanou vodou, 70% etanolem, dosucha očištěna a vložena do přístroje. Poté, co byl přístroj připraven k analýze, byl nahrán sample sheet s informacemi o vzorcích a spuštěn sekvenační běh.



Obr. 25: Sekvenační přístroje MiSeq (81, vlevo) a HiSeq 1500 (82, vpravo) společnosti Illumina

9. VÝSLEDKY

Cílem praktické části bylo zavedení a optimalizace NGS metod na klinickém genetickém pracovišti a ukázka jejich účinnosti na skupině vybraných vzorků. Během diplomové práce byla úspěšně zavedena metoda sekvenování na přístrojích MiSeq a HiSeq 1500 společnosti Illumina v ÚBLG FN Motol a při prvních experimentech bylo vyšetřeno 199 DNA vzorků. U všech vyšetřovaných pacientů byl před zahájením testování získán informovaný souhlas s vyšetřením a použitím DNA pro výzkumné účely.

9.1 Rozdělení analyzovaných vzorků

Pro otestování účinnosti nových sekvenačních metod bylo na přístroji MiSeq vyšetřeno 65 DNA vzorků pacientů s hereditární kardiomyopatií pomocí kitu TruSight Cardiomyopathy. Na přístroji HiSeq 1500 bylo vyšetřeno 126 DNA vzorků osob ze 42 rodin s předpokládaným výskytem hereditární kardiomyopatie. Pro plné exomové sekvenování kitem Nextera Rapid Capture Exome na přístroji HiSeq 1500 bylo vyšetřeno 8 pacientů z 3 rodin se závažnou genetickou zátěží. Ve dvou rodinách se jednalo o výskyt spinocerebelární ataxie (SCA) a v jedné rodině se jednalo o nejasný syndrom. V rodině A byla SCA diagnostikována celkem u 3 jedinců – matky, její sestry a u dcery. V rodině B u 2 jedinců – u matky a u dcery. V případě rodiny C se jednalo o nejasný syndrom s projevy mikrocefalie, intrauterinní růstové retardace, s nálezem chromozomových zlomů a neprospíváním. Rodiče v této rodině nejevili známky onemocnění. První potomek zemřel již v kojeneckém věku, u druhého byly pozorovány stejné projevy onemocnění.

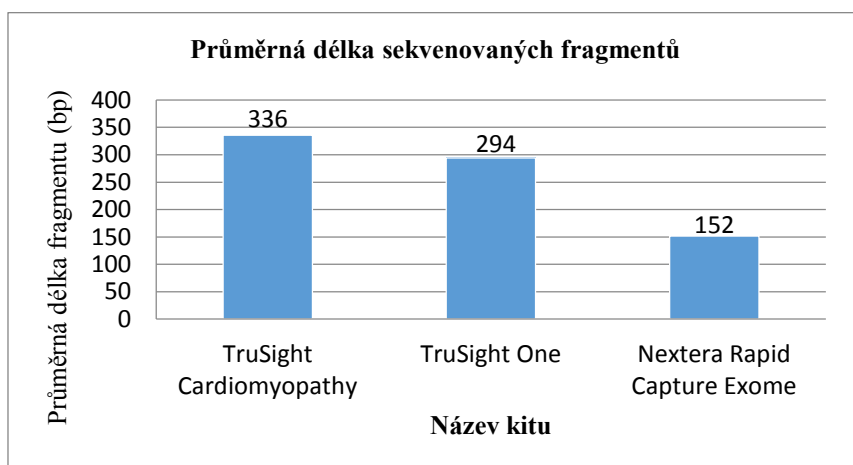
9.2 Optimalizace potřebných kroků

Při přípravě DNA knihoven byla potřebná především optimalizace při fragmentaci DNA, kde bylo docíleno optimálního množství štěpícího enzymu a počet otáček při mísení velkého množství vzorků s enzymem. Oproti standardnímu protokolu bylo množství tagmentačního enzymu sníženo na polovinu a byl také zajištěn postup, aby docházelo k rovnoměrnému přenosu otáček při zpracovávání velkého množství vzorků. Tím bylo dosaženo optimálních a reprodukovatelných délek fragmentů DNA. Při přípravě PCR pro exomové sekvenování byl také snížen počet PCR cyklů z 12 na 10, čímž došlo k optimální koncentraci namnožených fragmentů. Tyto změny oproti standardnímu protokolu odstranily

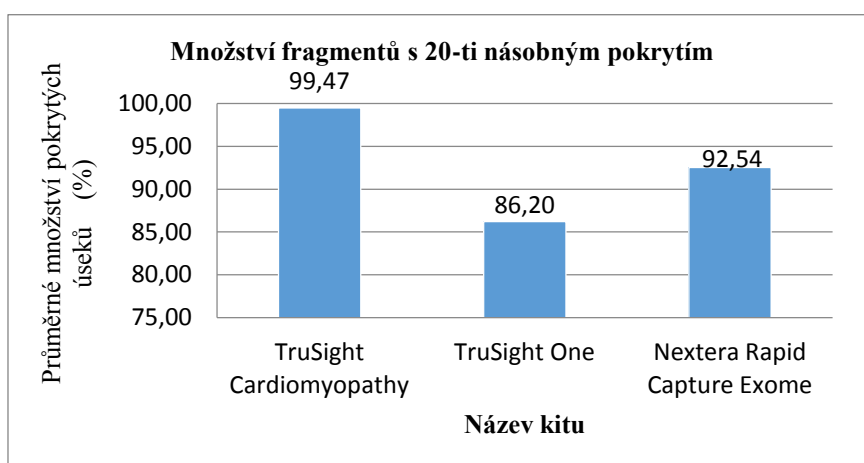
problémy, které provázely několik prvních experimentů a v následujících experimentech se ukázaly jako dostatečně reprodukovatelné. Jako optimální koncentrace finální knihovny byla experimentálně zjištěna a používána 10 pM pro MiSeq a 20 pM pro HiSeq 1500.

9.3 Kvalita sekvenačních dat

Při přípravě sekvenačních knihoven různými druhy kitů bylo dosaženo průměrné délky sekvenovaných fragmentů 336 bp u kitu TruSight Cardiomyopathy, 294 bp u kitu TruSight One a 152 bp u kitu Nextera Rapid Capture Exome (obrázek 26). Při sekvenování na přístroji HiSeq 1500 bylo dosaženo pokrytí průtokové komůrky u všech experimentů mezi 500 tisíci – 1 milionem klastrů/mm². V provedených experimentech bylo dosaženo 20 násobné pokrytí přibližně u 85-99 % čtených fragmentů (obrázek 27). Všechny tyto hodnoty plně odpovídají doporučením výrobce a splňují podmínky kvality pro následné kroky analýzy.



Obr. 26: Průměrná délka sekvenovaných fragmentů po přípravě knihoven různými kity



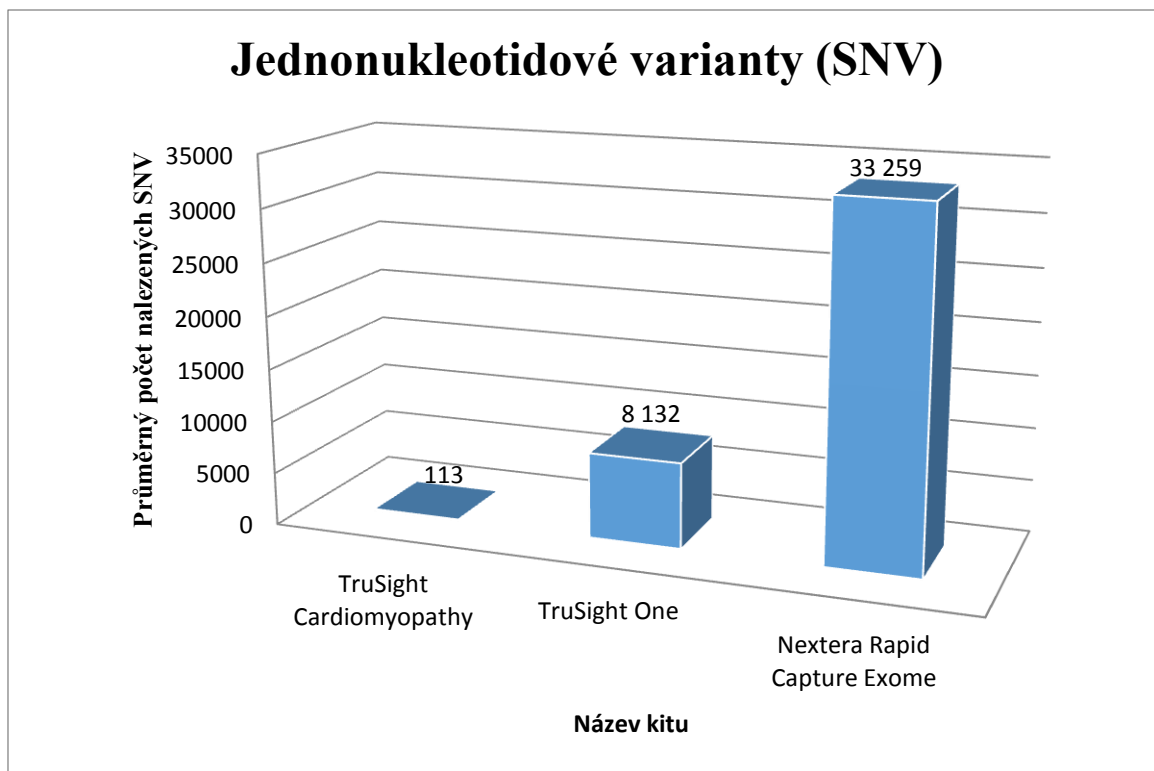
Obr. 27: Průměrné množství fragmentů, u nichž bylo dosaženo pokrytí 20x

9.4 Výsledky sekvenování

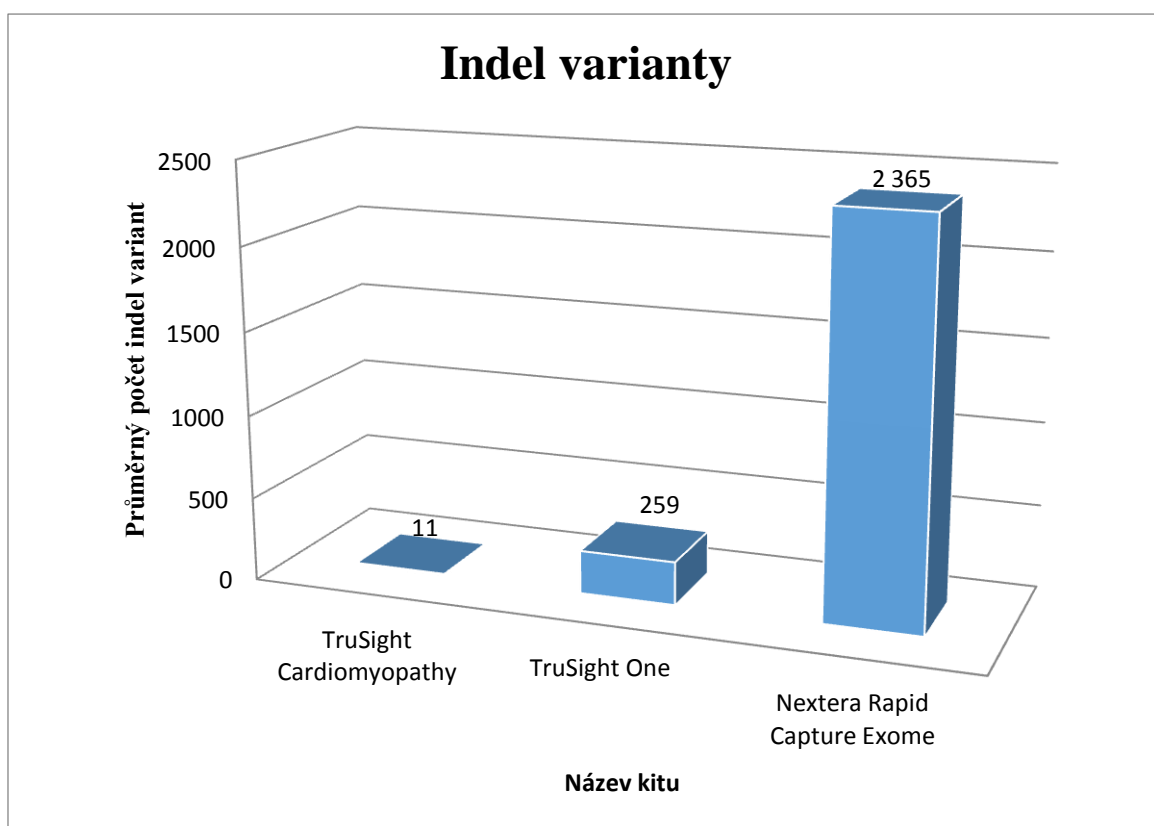
Kitem TruSight Cardiomyopathy bylo u každého vzorku vyšetřeno 46 genů v odpovídající délce 246 kb. U každého pacienta bylo tímto sekvenováním zjištěno průměrně 113 jednonukleotidových variant a 11 indel variant, z čehož byly vždy 0-4 varianty potenciálně škodlivé. Kitem TruSight One byla vyšetřena oblast DNA o délce 12 Mb a v průměru bylo u každého vzorku nalezeno 8 132 jednonukleotidových variant a 259 indel variant. Z tohoto počtu bylo vždy přibližně 500 variant potenciálně škodlivých. Kitem Nextera Rapid Capture Exome byla u každého vzorku vyšetřena oblast DNA o délce 37 Mb. V této oblasti bylo průměrně nalezeno 33 259 jednonukleotidových variant a 2 365 indel variant. Z tohoto počtu bylo vždy přibližně 3 500 variant potenciálně škodlivých. Výsledky sekvenování různými kity jsou shrnuty v tabulce 11 a rozdíly v záchytnosti kitů jsou znázorněny na obrázcích 28 a 29.

Tab. 11: Výsledky sekvenační analýzy s použitím tří rozdílných přístupů

	Použitý kit		
	TruSight Cardiomyopathy	TruSight One	Nextera Rapid Capture Exome
Sekvenátor	MiSeq	HiSeq 1500	HiSeq 1500
Vyšetřená oblast	200-300Kb	12Mb	37Mb
Počet sekvenovaných vzorků	65	126	8
Průměrný počet nalezených SNP	113	8 132	33 259
Poměr het/hom	1,1	1,8	1,8
Průměrný počet nalezených indel variant	11	259	2 365
Poměr het/hom	4,4	1,6	2,2
Přibližný počet potenciálně škodlivých variant	0-4	~500	~3 500



Obr. 28: Průměrný počet nalezených jednonukleotidových genetických variant u jednoho pacienta

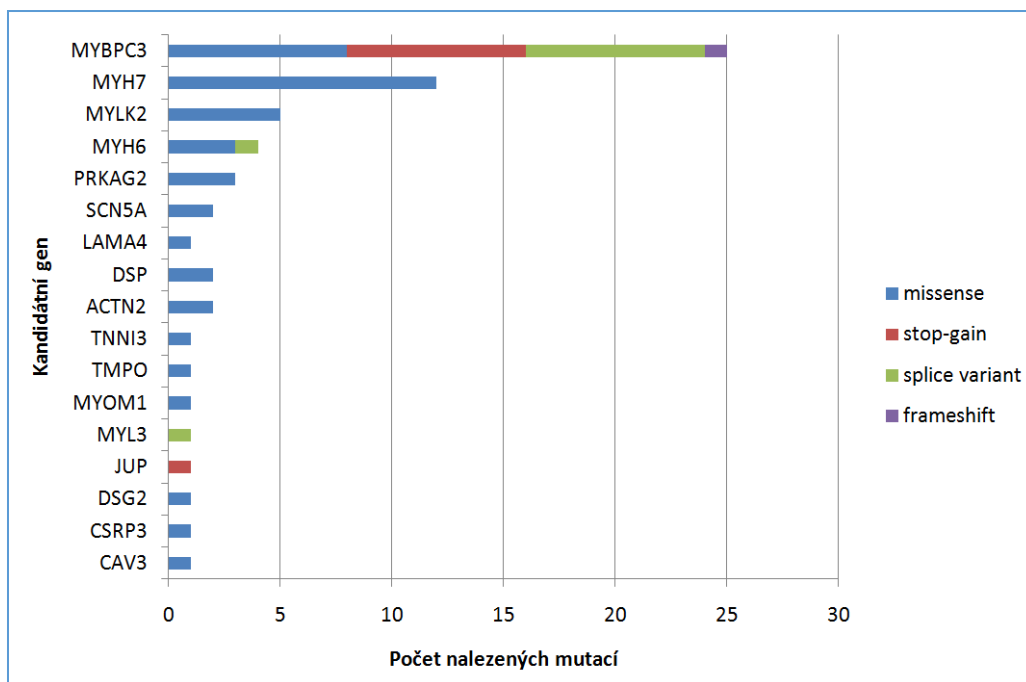


Obr. 29: Průměrný počet nalezených variant typu indel u jednoho pacienta

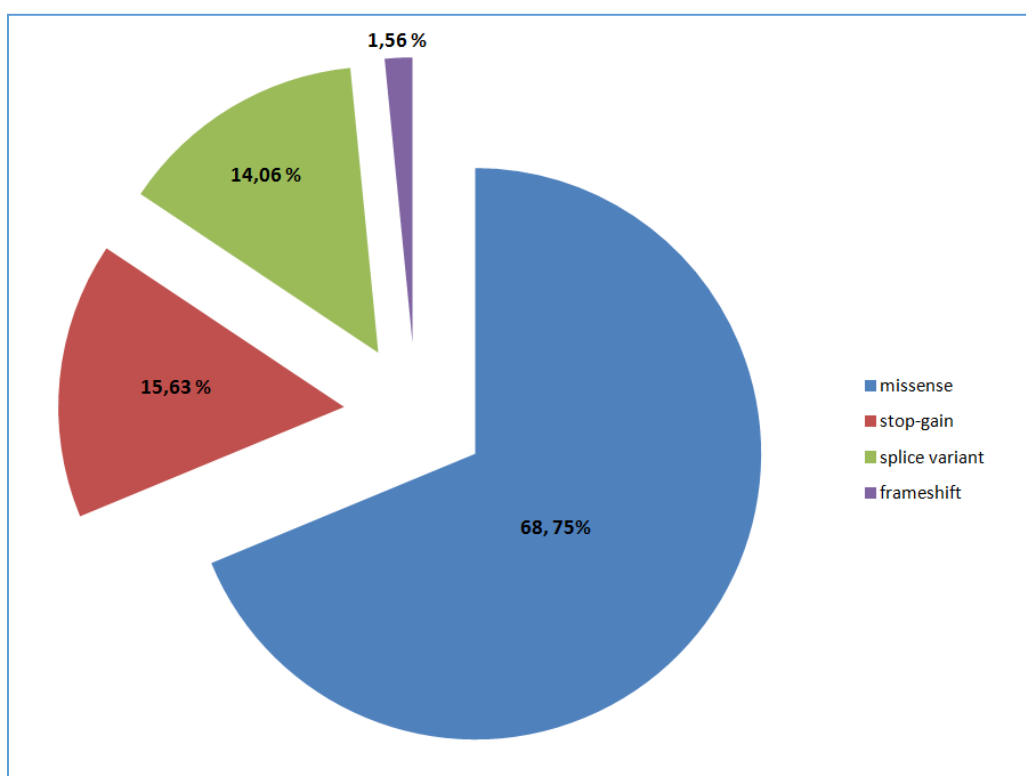
9.5 Zpracování dat ze sekvenování panelu

Varianty zjištěné při cíleném sekvenování kitem TruSight Cardiomyopathy byly filtrovány v několika krocích především v programu Illumina Variant Studio verze 2.2. V prvním kroku byla data filtrována podle kvality a k další filtraci byla připuštěna pouze ta, u nichž byla hloubka čtení >10 . V další filtraci byly odfiltrovány alternativní varianty v množství < 30 %. V případě, že chceme zachytit mozaicismus, je vhodné nastavit citlivější filtr přibližně kolem 20 %. U ideálního heterozygota by měl být výskyt alternativní alely 50 %, v případě homozygota 100 %. V následném kroku byly odfiltrovány alely, jejichž zastoupení v populaci je vyšší než 2 % v případě autozomálně recesivního přenosu, resp. vyšší než 1 % v případě předpokládané autozomálně dominantní dědičnosti. Další filtr byl zaměřen na patogenitu případné mutace. Byly zobrazeny pouze mutace, v jejichž případě je jasný dopad na tvořenou aminokyselinu, tedy stop-gain a frameshift mutace. Pokud byla po této filtraci nalezena mutace v některém genu, bylo možné předpokládat, že se jedná právě o kauzální mutaci. V opačném případě byly naopak zobrazeny missense a sestřihové varianty. Nalezené missense varianty byly podrobeny *in silico* testování predikce patogenity v programech Sift (<http://sift.jcvi.org/>), PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) a Mutation taster (<http://www.mutationtaster.org/>).

Pomocí filtrování bylo u 48 pacientů z 65 (73,85 %) vytipováno 64 mutací v 17 genech (obrázek 30), u nichž lze předpokládat kauzální vztah s hereditární kardiomyopatií. Nejčastěji byly mutace nacházeny v genech *MYBPC3* (39,06 %), *MYH7* (18,75 %), *MYLK2* (7,81 %), *MYH6* (6,25 %) a *PRKAG2* (4,69 %). V 1-4 % se vyskytovaly mutace v genech *CAV3*, *CSRP3*, *DSG2*, *JUP*, *MYL3*, *MYOM1*, *TMPO*, *TNNI3*, *ACTN2*, *DSP*, *LAMA4* a *SCN5A*. V naprosté většině případů se jednalo o missense mutace (68,75 %), dále pak o stop-gain mutace (15,63 %) a sestřihové varianty (14,06 %). V jednom případě (1,56 %) byla zachycena posunová mutace frameshift (obrázek 31). Kompletní tabulka nalezených variant jako možných kauzálních mutací na sekvenátoru MiSeq je uvedena v příloze 1.



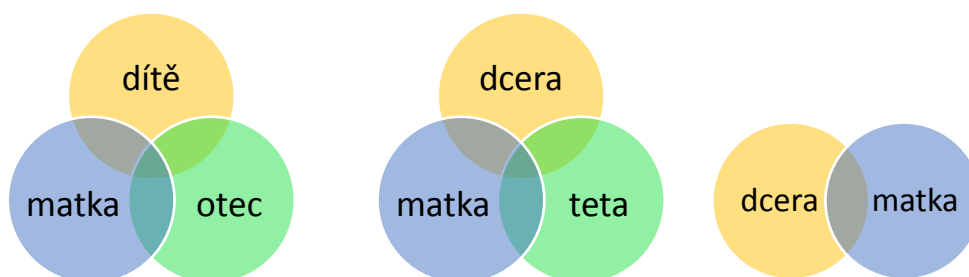
Obr. 30: Mutované geny u pacientů s hereditární kardiomyopatií nalezené cíleným panelem



Obr. 31: Typy nalezených mutací přístrojem MiSeq u pacientů s hereditární kardiomyopatií

9.6 Zpracování dat z exomového sekvenování

Filtrace dat u exomového sekvenování probíhala obdobně jako u cíleného sekvenování panelu genů s několika málo rozdíly. Exomovým sekvenováním jsme hledali dědičné kauzální mutace v rodinách, přičemž byla testována většinou tzv. tria – matka, otec a dítě. V jednom případě celoexomového sekvenování byly u rodiny A testováni 3 postižení SCA – dcera, matka a teta a v jednom případě u rodiny B byli testováni 2 postižení SCA – dcera a matka. Použité sestavy jsou znázorněny na obrázku 32. V těchto sestavách jsme hledali shodné varianty, proto filtrace shodných variant předcházela dalšímu filtrování popsanému u analýzy panelu.



Obr. 32: Filtrování společných variant využitá při analýze exomových dat

Další rozdíl při filtrování dat získaných exomovým sekvenováním oproti výsledkům panelu bylo, že nebyly přesně stanoveny geny, u nichž kauzální mutaci očekáváme. Tyto geny bylo tedy nutné zvolit podle daného experimentu. V případě známých onemocnění, jakým je skupina hereditárních kardiomyopatií či SCA, lze seznam asociovaných popsaných genů získat z několika zdrojů. V našem případě byl získán z databáze OMIM (<http://www.omim.org/>). U rodiny C, při níž byla hledána kauzální mutace v rámci neznámého syndromu, byly vhodné geny pro filtraci získány na základě popisu fenotypových známek onemocnění v programu Phenomizer (<http://compbio.charite.de/phenomizer/>).

Vyfiltrováním dat po sekvenování tzv. klinického exomu kitem TruSight One byly dodnes předpokládané kauzální mutace nalezeny ve 23 genech u 27 rodin s hereditární kardiomyopatií (tabulka 12). Ve 12 případech (52,17 %) se jedná o geny, které nejsou zahrnuty v panelu TruSight (vyznačeny tučně). Při filtrování dat získaných celoexomovým sekvenováním se u rodiny A podařilo filtrováním snížit počet možných kauzálních variant ze 7 875 shodných variant na 12 variant v 6 genech, u rodiny B z 12 307 shodných variant na

5 variant v 5 genech a u rodiny C z 6 675 shodných variant na 3 varianty ve 3 genech. Seznam genů, u nichž byly nalezeny možné kauzální mutace, jsou uvedeny v tabulkách 13-15. Tyto mutace budou i nadále studovány v dostupné literatuře.

Tab. 12: Mutované geny u pacientů s hereditární kardiomyopatií zjištěné kitem TruSight One

Název genu:	Pozice genu:	Postižený protein:
<i>TTN</i>	chr2:76703980-76982547 (-)	titin
<i>MYH7</i>	chr14:54970688-54994634 (-)	myosin, heavy chain 7, cardiac muscle, beta
<i>PDLIM3</i>	chr8:45885485-45919546 (+)	alpha-actinin-2-associated LIM protein
<i>TPM1</i>	chr9:67022590-67049406 (-)	tropomyosin 1 (alpha)
<i>FLNC</i>	chr6:29433256-29461849 (+)	filamin C
<i>TNNI3</i>	chr7:4518305-4524229 (-)	troponin I type 3 (cardiac)
<i>DSG2</i>	chr18:20558074-20604521 (+)	desmoglein 2
<i>JPH2</i>	chr2:163336242-163397949 (-)	junctophilin 2
<i>MYBPC3</i>	chr2:91118144-91136516 (+)	myosin binding protein C, cardiac
<i>OBSCN</i>	chr11:58994259-59136402(-)	obscurin
<i>PKP2</i>	chr16:16213318-16272712 (+)	plakophilin-2
<i>LAMP2</i>	chrX:38401357-38456454 (-)	lysosome-associated membrane protein 2
<i>MURC</i>	chr4:48663514-48673502 (+)	muscle-restricted coiled-coil rotein
<i>VCL</i>	chr14:20929433-21033673 (+)	vinculin
<i>KLF10</i>	chr15:38294413-38300707 (-)	krueppel-like factor 10
<i>BAG3</i>	chr7:128523583-128546977 (+)	BAG family molecular chaperone regulator 3
<i>CTNNA3</i>	chr10:63430098-65003667 (+)	alpha-3-catenin (also called alpha-T-catenin)
<i>ACTC1</i>	chr2:114047282-114053548 (-)	actin, alpha cardiac muscle 1
<i>SCN5A</i>	chr9:119483410-119579016 (-)	sodium channel, voltage-gated, type V, alpha
<i>DOLK</i>	chr2:30284231-30286334 (-)	dolichol kinase
<i>CHRNB4</i>	chr9:55028156-55048544 (-)	neuronal acetylcholine receptor subunit beta-4
<i>RBM20</i>	chr19:53677306-53867080 (+)	RNA binding motif protein 20
<i>PIK3C3</i>	chr18:30272747-30348126 (+)	phosphatidylinositol 3-kinase, catalytic subunit type 3

Tab. 13: Geny, u nichž byla v rodině A zjištěna možná kauzální mutace celoxomovým sekvenováním

Rodina A		
Název genu:	Pozice genu:	Postižený protein:
<i>FLVCRI</i>	chr1:212858255-212899363 (+)	feline leukemia virus subgroup C receptor-related protein 1
<i>ITPRI</i>	chr3:4493348-4847840 (+)	inositol 1,4,5-trisphosphate receptor, type 1
<i>SYNE1</i>	chr6:152121684-152637399 (-)	spectrin repeat containing, nuclear envelope 1
<i>SETX</i>	chr9:132261440-132356726 (-)	senataxin
<i>TTBK2</i>	chr15:42738734-42921007 (-)	tau tubulin kinase 2
<i>CACNA1A</i>	chr19:13206442-13506460 (-)	calcium channel, voltage-dependent, P/Q type, alpha 1A subunit

Tab. 14: Geny, u nichž byla v rodině B zjištěna možná kauzální mutace celoxomovým sekvenováním

Rodina B		
Název genu:	Pozice genu:	Postižený protein:
<i>RELN</i>	chr7:103471784-103989516 (-)	reelin
<i>SETX</i>	chr9:132261440-132356726 (-)	senataxin
<i>ATXN3</i>	chr14:92058552-92106621 (-)	ataxin 3
<i>RPGRIP1L</i>	chr16:53599906-53703859 (-)	RPGRIP1-like
<i>KARS</i>	chr16:75627724-75647687 (-)	lysyl-tRNA synthetase

Tab. 15: Geny, u nichž byla v rodině C zjištěna možná kauzální mutace celoxomovým sekvenováním

Rodina C		
Název genu:	Pozice genu:	Postižený protein:
<i>RAI1</i>	chr17:17681473-17811453 (+)	retinoic Acid-Induced Protein
<i>PHF8</i>	chrX:53936680-54048935 (-)	PHD finger protein 8
<i>SLC9A6</i>	chrX:135985424-136047269 (+)	sodium/hydrogen exchanger 6

10. DISKUZE

Nové metody sekvenování dnes pomalu nahrazují klasické kapilární sekvenování, které si v molekulárně genetických laboratořích zachovávalo své výsadní postavení od konce 70. let 20. století. Během uplynulého desetiletí jsme svědky nebývalého rozvoje sekvenačních technologií a snižování nákladů potřebných k přečtení lidské genetické informace. I přes to, že se v minulých letech podařilo snížit náklady na osekvenování lidského genomu pod magickou hranici tisíc amerických dolarů, celogenomové sekvenování stále zůstává minoritní oblastí zájmu při diagnostice onemocnění NGS metodami. Technická náročnost přípravy a především složitost vyhodnocení dat získaných celogenomovým sekvenováním, vedou na řadě vědeckých pracovišť k tomu, že je častěji sekvenován lidský exom, který zahrnuje pouze kódující úseky DNA o velikosti přibližně 1 % genomu. Exomové studie tak umožnily rozsáhlý výzkum genetických příčin mnoha dnes stále nedostatečně prostudovaných onemocnění, jakými jsou nádorová onemocnění či skupina hereditárních kardiomyopatií.

Především výstupy exomových studií tak umožňují zaměřit se na diagnostiku nových genů, u nichž se předpokládá kauzální vztah s dědičným onemocněním a převést tyto poznatky do rutinní praxe v rámci testování tzv. genových panelů, tedy skupin genů, jejichž mutace se častěji vyskytují u pacientů s daným onemocněním. Na trhu je dnes mnoho hotových kitů, které laboratořím umožňují velmi snadný přechod na tuto vyšetřovací metodu a některé z nich již dokonce získaly validaci pro rutinní klinickou diagnostiku (CE-IVD). Na druhou stranu je však vzhledem k rostoucímu rozsahu a složitosti sekvenačních panelů pro rutinní genetické testování velmi důležité, aby také jejich vyhodnocení a interpretace byla prováděna odborníky v dané oblasti. Do této debaty začínají vstupovat také zdravotní pojišťovny, které mají za cíl prověřovat, zda je vyšetřovaným pacientům následně poskytována odpovídající péče, ať už v oblasti genetického poradenství či péče preventivní a terapeutická. Řada nových poznatků, které NGS přináší, vede na části genetických pracovišť ke snaze zavádět do rutinního a screeningového testování osob v riziku také geny, u nichž však nebyla dostatečně prokázána souvislost se vznikem onemocnění.

Typickým příkladem je snaha o rozšíření panelu pro prediktivní testování nádorů prsu a ovarií u žen v riziku o další geny, než jsou známé geny *BRCA1* a *BRCA2*. Zatímco oba tyto geny jsou zkoumány více než 20 let, na některých pracovištích je dnes testováno zhruba dalších 15 genů, jejichž prevalence, penetrace a vliv na vznik nádoru ještě nejsou dostatečně

pochopeny. Proti tomuto testování hovoří také nízká váha případného negativního výsledku. Velké populační studie prokázaly, že ženy, které nezdědily mutace v *BRCA1/2* genech, nemají zvýšené riziko vzniku rakoviny oproti průměrné populaci a výstupy těchto studií mají uklidňující vliv na negativně testované ženy. V případě nově testovaných genů jako je např. *CHEK2* však dnes podobné studie neexistují a tato skutečnost výrazně snižuje hodnotu negativního výsledku vyšetření. Na druhou stranu testování těchto panelů u *BRCA1/2* negativních žen může do budoucna přispět k poznatkům o vlivu těchto genů na vznik nádorového onemocnění. Při interpretaci případného pozitivního výsledku je však potřeba postupovat velmi obezřetně.

Každé genetické pracoviště by mělo před zavedením NGS metod důkladně zvážit nejen to, jakou platformu zvolit, ale také jaký zvolit přístup k sekvenování a k samotným pacientům. Každá z dnes používaných NGS platform má své výhody a své nevýhody. Žádnou z nich tedy nelze označit za nejlepší na trhu a je nutné správně zvážit všechny faktory, které jsou k výběru nového přístroje pro dané pracoviště důležité. Častým rozhodovacím faktorem bývá samozřejmě cena sekvenátoru. Z hlediska ceny by však neměla být brána v potaz pouze cena pořizovací, nýbrž by mělo být počítáno i s náklady na přípravu sekvenačních knihoven a provoz přístroje. Hlavním rozhodovacím kritériem by mělo být také to, zda bude sekvenace využívána převážně pro rutinní diagnostiku jako náhrada klasického sekvenování, k výzkumným účelům, či dokonce k celogenomovému sekvenování. Také analýza dat získaných sekvenováním nové generace je však náročným úkolem, který na laboratoře klade nároky z hlediska dostatečného personálního i technického vybavení.

Náklady na NGS testování tak zvyšují také nákupy nových počítačů a uložišť pro objemná data, stejně jako nové analytické softwary. Daleko větší význam, než v minulosti, v molekulárně genetických provozech získávají bioinformatiči, kteří se spolu s klinickými pracovníky podílejí na analýze získaných dat. S rostoucím množstvím dat, která jsou od každého pacienta získávána, roste význam bezpečnosti a zacházení s těmito daty, stejně jako význam kvalifikované interpretace výsledků analýzy zkušenými klinickými genetiky. Při zavádění nové sekvenační metody pro rutinní testování na klinickém pracovišti je mimo jiné důležité rozvrhnout, jakými pracovníky budou jednotlivé části vyšetření prováděny. Pracovníci, kteří například provádí přípravy knihoven, jsou často plně touto prací vytíženi a měli by být vyčleněni z běžného rutinního provozu. Vyčlenění laboratorních pracovníků pro přípravu knihoven je důležité také z hlediska možné kontaminace během přípravy DNA

vzorků. Na vyhodnocení dat by se měl podílet bioinformatik spolu s klinickým genetikem. Je optimální, když kliničtí pracovníci, kteří se léta zabývají určitým onemocněním a znají jeho genetické i klinické aspekty, využijí NGS pouze jako nástroj pro rozšíření poznatků o genetickém pozadí nemoci svých pacientů.

Přínos NGS technologií pro diagnostiku byl prokázán v praktické části této diplomové práce. Po úspěšném zavedení a optimalizaci potřebných kroků se pomocí cíleného panelu genů již při prvních experimentech podařilo odhalit předpokládanou příčinu hereditární kardiomyopatie u 73,85 % testovaných pacientů. To ukazuje velkou účinnost tohoto způsobu sekvenování na předem vybrané skupině osob. I přes některé technické problémy s přístrojem HiSeq 1500, který musel být opakovaně opravován, se díky výsledkům sekvenování klinického exomu kitem TruSight One dodnes podařilo objasnit příčinu dědičné kardiomyopatie u 27 rodin ze 42 testovaných (64,29 %) nálezem patogenních genetických variant ve 23 genech. Velká část těchto genů není obsažena v sekvenačním panelu TruSight Cardiomyopathy a při jeho využití by tak nebyla odhalena. To ukazuje jednoznačný přínos tohoto typu sekvenování a současně hlavní omezení sekvenování vybraného panelu genů.

Z 13 rodin s diagnózou HCM (hypertrofická kardiomyopatie) byla kauzální mutace identifikována ve 12 případech (92,31 %). Z 22 rodin s diagnózou DCM (dilatační kardiomyopatie) byly specifické mutace nalezeny u 11 rodin (50 %). Navíc u jedné rodiny byla diagnostika LVNC (nonkompaktní kardiomyopatie) vyřešena nálezem mutace v genu *OBSCN*. Diagnostika jedné rodiny s ARVD (arytmogenní pravostranná dysplasie srdce) byla vyřešena díky nalezení mutace v genu *PKP2*. U jedné rodiny byla vyřešena diagnostika kombinovaného syndromu LQT/CPVT (syndrom dlouhého QT intervalu/katecholaminergní polymorfní komorová tachykardie) nálezem mutace v genu *SCN5A*. Z dvou rodin s kombinovanou poruchou RCM/HCM (restriktivní kardiomyopatie/hypertrofická kardiomyopatie) byla kauzální mutace odhalena zatím u jedné z nich. Na druhou stranu v 15 rodinách sekvenovaných kitem TruSight One stále nebyla odhalena kauzální mutace a získaná sekvenační data budou i nadále vyhodnocována.

Podobně tomu je i v případě celoexomového sekvenování. Dodnes se podařilo díky mnohonásobnému filtrování získaných dat snížit počet možných kauzálních variant u rodiny A na 6 (v genech *FLVCRI*, *ITPRI*, *SYNE1*, *SETX*, *TTBK2* a *CACNA1A*), u rodiny B na 5 (v genech *RELN*, *SETX*, *ATXN3*, *RPGRIP1L* a *KARS*) a u rodiny C dokonce na 3 (v genech

RAI1, *PHF8* a *SLC9A6*). K této filtraci byl použit nejen software společnosti Illumina, ale také on-line nástroje pro prediktivní testování patogenity. Právě analýza dat se ukázala jako skutečně nejsložitější fází celého procesu. Zatímco příprava sekvenačních knihoven se v podstatě nelišila, ať již byla připravována knihovna pro panel několika desítek genů, klinický exom obsahující několik tisíc genů či celý exom obsahující přibližně 20 tisíc genů, následná analýza dat se výrazně lišila především počtem vstupních variant pro analýzu. Při testování panelu, kdy vstupem do analýzy bylo přibližně 100-200 nalezených variant, byla situace poměrně snadná a v optimálním případě bylo možné nalézt kauzální mutaci v řádu několika dní. Jiná situace však nastala v případě sekvenování exomů, kdy bylo nalezeno přibližně 8, resp. 30 tisíc variant u jednoho pacienta. V takovém případě byla doba potřebná pro filtrování variant i několik měsíců a v případě výsledků celých exomů stále zbývá důkladné zkoumání zbylých variant.

Na základě výsledků a získaných zkušeností dnes můžeme prohlásit, že cílené testování genových panelů dnes může bez problému nahradit klasické Sangerovo sekvenování a zajistit tak kvalitní testování osob v riziku na všech molekulárně genetických pracovištích. Skutečnost, že při našich prvních experimentech byly všechny mutace nalezeny pouze v 17 ze 46 testovaných genů (36,96 %), na jednu stranu ukazuje na kapacitní rezervu testovaného panelu pro záchyt mutací v dalších genech. Na druhou stranu naopak výsledky exomového sekvenování svědčí o tom, že řada potřebných genů pro hereditární kardiomyopatii není v panelu zahrnuta. Je tak na zvážení každého pracoviště, zda nevyužít současných možností, které někteří výrobci nabízí a nevytvořit si pro testování vlastní panel složený z genů, které si zvolí na základě vlastních zkušeností. Pokud si pracoviště naopak zvolí pro diagnostiku sekvenování klinického exomu, vyhne se tak možnosti, že nezachytí mutace v genech, které nejsou součástí panelu. Stále však, kromě složitosti analýzy takto získaných dat, zbývá vyřešit etický aspekt tohoto vyšetření. Během analýzy klinického exomu nejsou totiž testovány pouze geny související s onemocněním, pro něž je pacient testován, ale jsou vyšetřeny všechny geny související s některým dědičným onemocněním. Nastává tak otázka, zda je toto testování z etického pohledu vhodné a to, jak a případně zda vůbec pacienta informovat o případných nálezech v dalších genech, které vedou ke vzniku predispozice pro jiné závažné onemocnění.

11. ZÁVĚR

V této diplomové práci byly popsány principy sekvenačních technologií nové generace u dostupných platforem a jednotlivé přístupy k sekvenování. Na základě porovnání jednotlivých platforem byla jako nejvhodnější pro rutinní diagnostiku vybrána platforma společnosti Illumina, jejíž přístroje byly použity i v rámci praktické části. Byla přiblížena specifika NGS metod a problémy, se kterými se laboratoře při vyšetřování setkávají. Byl také nastíněn postup analýzy dat při resekvenování na platformě Illumina. V rámci praktické části byla úspěšně zavedena metoda sekvenování nové generace přístroji MiSeq a HiSeq 1500 na pracovišti Ústavu biologie a lékařské genetiky 2. LF UK a FN Motol a účinnost sekvenování byla prokázána na vybrané skupině vzorků. Díky prvním provedeným experimentům se podařilo cíleným panelem genů zachytit kauzální mutace u 48 pacientů a vyšetřením klinického exomu se podařilo nalézt kauzální mutace ve 27 rodinách s hereditární kardiomyopatií.

Díky této diplomové práci byla také optimalizována příprava DNA knihoven a tato vylepšení budou součástí nově připravovaného standardního operačního postupu na pracovišti. Byly také vytvořeny podmínky pro testování dalších vyšetřovacích metod a již dnes jsou v ÚBLG testovány kity pro sekvenování genů spojených s nádorovými onemocněními, cystickou fibrózou či kity pro preimplantační diagnostiku. Zavedené metody tak již dnes velmi rozšířily diagnostické možnosti pracoviště a jsou velkým přínosem pro jeho pracovníky i samotné pacienty, kterým je díky přesným výsledkům poskytována adekvátní péče.

12. SEZNAM POUŽITÝCH ZKRATEK

a-CGH	Array Comparative Genomic Hybridization, komparativní genomová hybridizace
ARVD	Arrhythmogenic Right Ventricle Cardiomyopathy, arytmogenní dysplazie pravé komory
ASCII	American Standard Code for Information Interchange, americký standardní kód pro výměnu informací
ATP	Adenosine Triphosphate, adenosintrifosfát
bp	Base pair, pár bází
CCD	Charge-Coupled Device, zařízení s vázanými náboji
cDNA	complementary DNA, komplementární DNA
CE-IVD	„Conformité Européenne“ - In Vitro Diagnostic, certifikováno pro in vitro diagnostiku
CMOS	Complementary Metal–Oxide–Semiconductor
CPVT	Catecholaminergic Polymorphic Ventricular Tachycardia, katecholaminergní polymorfni ventrikulární tachykardie
ČLS JEP	Česká lékařská společnost Jana Evangelisty Purkyně
DCM	Dilated Cardiomyopathy, dilatační kardiomyopatie
DNA	Deoxyribonucleic Acid, deoxyribonukleová kyselina
dNTP	Deoxyribonucleotide Triphosphate, deoxyribonukleotid trifosfát
dsDNA	double-strand DNA, dvouvláknová DNA
Gb	Gigabase, miliarda bází
HCM	Hypertrophic Cardiomyopathy, hypertrofická kardiomyopatie
HGP	Human Genome Project, Projekt lidského genomu
IKEM	Institut klinické a experimentální medicíny
IUPAC	International Union of Pure and Applied Chemistry, Mezinárodní unie pro čistou a užitou chemii
kb	Kilobase, tisíc bází
LQT	Long QT syndrome, syndrom dlouhého QT intervalu
LVNC	Left Ventricular Noncompaction, nonkompaktní kardiomyopatie
Mb	Megabase, milion bází

MLPA	Multiplex Ligation-dependent Probe Amplification
MPSS	Massively Parallel Signature Sequencing, masivní paralelní sekvenování
mRNA	messenger ribonucleic acid, mediátorová ribonukleová kyselina
NGS	Next Generation Sequencing, sekvenování nové generace
OMIM	Online Mendelian Inheritance in Man, online databáze lidských genů a dědičných onemocnění
PCR	Polymerase Chain Reaction, polymerázová řetězová reakce
PGD	Preimplantation Genetic Diagnostic, preimplantační genetická diagnostika
pH	potential of hydrogen, vodíkový potenciál
QFPCR	Quantitative Fluorescence PCR, kvantitativní fluorescenční PCR
RCM	Restrictive Cardiomyopathy, restriktivní kardiomyopatie
RNA	Ribonucleic Acid, ribonukleová kyselina
rpm	Revolutions per minute, otáček za minutu
rRNA	ribosomal ribonucleic acid, ribozomální ribonukleová kyselina
SCA	Spinocerebellar Ataxia, Spinocerebelární ataxie
SCD	Sudden Cardiac Death, náhlá srdeční smrt
SMRT	Single Molecule Real Time Sequencing, jednomolekulové sekvenování v reálném čase
SNP	Single Nucleotide Polymorphism, jednonukleotidový polymorfismus
SNV	Single Nucleotide Variant, jednonukleotidová varianta
ssDNA	single-strand DNA, jednovláknová DNA
tRNA	transfer ribonucleic acid, transferová ribonukleová kyselina
USB	Universal Serial Bus, univerzální sériová sběrnice
ZMW	Zero-Mode Waveguide

13. POUŽITÁ LITERATURA A PRAMENY

1. DAHM, R. Friedrich Miescher and the discovery of DNA. *Developmental Biology*. 2005, vol. 278, issue 2, s. 274-288.
2. WATSON, J.D, CRICK F.H.C. A Structure for Deoxyribose Nucleic Acid. *Nature*. 1953. vol. 171, s. 737-738.
3. HUTCHISON, C.A. DNA sequencing: bench to bedside and blond. *Nucleic Acids Research*. 2007. vol. 35, s. 6227–6237.
4. DAVIES, K. Cracking the genome: inside the race to unlock human DNA. New York: Free Press, 2001, 310 s.
5. SIKKEMA-RADDATZ, B. et al. Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics. *Human Mutation*. 2013, vol. 34, issue 7, s. 1035-1042.
6. SANGER, F. et al. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977, vol. 74, issue 12, s. 5463-5467.
7. ANDERSON, S. et al. Sequence and organization of the human mitochondrial enome. *Nature*. 1981, vol. 290, issue 5806, s. 457-465.
8. FLEISCHMANN, R. et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*. 1995, vol. 269, issue 5223, s. 496-512.
9. GOFFEAU, A. et al. Life with 6000 Genes. *Science*. 1996, vol. 274, issue 5287, s. 546-567.
10. POSPÍŠILOVÁ, Š. et al. Sekvenování lidského genomu – technologie nové generace aneb budeme rutinně sekvenovat lidské genomy?. *Časopis lékařů českých*. 2009, roč. 7, č. 148, s. 296-302.
11. *Human Genome Project's Five-Year Plan (1991-1995)* [online]. Poslední revize 01.2015 [cit. 2015-01-14]. Dostupné z: <http://www.genome.gov/10001477>
12. STRETTON, A.O.W. The First Sequence: Fred Sanger and Insulin. *Genetics*. 2002, vol. 162, issue 2, s. 527-532.
13. KARGER, B., GUTTMAN A. DNA sequencing by CE. *Electrophoresis*. 2009, vol. 30, issue 1, s. 196-202.
14. PRIMROSE, S, TWYMAN R. *Principles of genome analysis and genomics*, 3rd ed. Malden, Mass : Blackwell Pub., 2002, 288 s. ISBN 978-1-4051-0120-2
15. BRDIČKA, R. Kam směřuje molekulárně-genetická analýza. *Klin. Biochem. Metab.* 2011, č. 19, s. 143-148.
16. STADEN, R. A Strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*. 1979, vol. 6, issue 7, s. 2601-2610.

17. REECE, J.B. et al. *Campbell Biology*, 9th ed. San Francisco: Benjamin Cummings Publishing Company, 2012, 1464 s. ISBN 978-0321831545.
18. COLLINS, F. S. The Human Genome Project: Lessons from Large-Scale Biology. *Science*. 2003, vol. 300, issue 5617, s. 286-290.
19. *About the Human Genome Project: What was the Human Genome Project?. Human Genome Project Information Archive 1990-2003* [online]. Poslední revize 02.2015 [cit. 2015-02-19]. Dostupné z: http://web.ornl.gov/sci/techresources/Human_Genome/project/index.shtml
20. OLSON, M.V. The maps: Clone by clone by clone. *Nature*. 2001, vol. 409, issue 6822, s. 816-818.
21. LANDER, E.S. Initial impact of the sequencing of the human genome. *Nature*. 2011, vol. 470, issue 7333, s. 187-197.
22. VENTER, J. C. The Sequence of the Human Genome. *Science*. 2001, vol. 291, issue 5507, s. 1304-1351.
23. ZOBANÍKOVÁ, M., ČEJKOVÁ D. Genové patenty. *Universitas: revue Masarykovy univerzity*. 2013, č. 3, s. 22-27.
24. PLATENIK, J. *Sequencing Genomes* [online]. Poslední revize 12.2014 [cit. 2014-12-10]. Dostupné z: <https://www.yumpu.com/en/document/view/30918070/sequencing-genomes>
25. VOELKERDING, K. V. et. al. Next-Generation Sequencing: From Basic Research to Diagnostics. *Clinical Chemistry*. 2009, vol. 55, issue 4, s. 20-40.
26. *HiSeq X Ten System* [online]. Poslední revize 02.2015 [cit. 2015-02-26]. Dostupné z: <http://www.illumina.com/systems/hiseq-x-sequencing-system/system.html>
27. WEISS, M.M. et al. Best Practice Guidelines for the Use of Next-Generation Sequencing Applications in Genome Diagnostics: A National Collaborative Study of Dutch Genome Diagnostic Laboratories. *Human Mutation*. 2013, vol. 34, issue 10, s. 1313-1321.
28. *GS FLX+ System : 454 Life Sciences, a Roche Company* [online]. Poslední revize 12.2014 [cit. 2014-12-12]. Dostupné z: <http://454.com/products/gs-FLX-system/index.asp>
29. *Specifications for HiSeq 2500* [online]. Poslední revize 01.2015 [cit. 2015-01-21]. Dostupné z: http://www.illumina.com/systems/hiseq_2500_1500/performance_specifications.html
30. KNIERIM, E. et al. Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing. *PLoS ONE*. 2011, vol. 6, issue 11.
31. PICELLI, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*. 2014, vol. 24, issue 12, s. 2033-2040.
32. LENNON, N.J. et al. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biology*. 2010, vol. 11, issue 2.

33. ZHOU, X. et al. The next-generation sequencing technology and application. *Protein*. 2010, vol. 1, issue 6, s. 520-536.
34. SHOKRALLA, SHADI et al. Next-generation sequencing technologies for environmental DNA research: New Grants Support Development of Faster, Cheaper DNA Sequencing. *Molecular Ecology*. 2012, vol. 21, issue 8, s. 285-302.
35. XUAN, J. et al. Next-generation sequencing in the clinic: Promises and challenges. *Cancer Letters*. 2013, vol. 340, issue 2, s. 284-295.
36. *Next-Gen Sequencing 2014 Update* [online]. Poslední revize 12.2014 [cit. 2014-12-19]. Dostupné z: <http://www.biocompare.com/Editorial-Articles/155411-Next-Gen-Sequencing-2014-Update/>
37. *Pacific Biosciences Releases New DNA Sequencing Chemistry to Enhance Read Length and Accuracy for the Study of Human and Other Complex Genomes* [online]. Poslední revize 12.2014 [cit. 2014-12-21]. Dostupné z: <http://investor.pacificbiosciences.com/releasedetail.cfm?ReleaseID=876252>
38. LIU, L. et al. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*. 2012, vol. 2012, s. 61-70.
39. *Quality (Phred) scores* [online]. Poslední revize 01.2015 [cit. 2015-01-10]. Dostupné z: http://drive5.com/usearch/manual/quality_score.html
40. *FASTQ files* [online]. Poslední revize 01.2015 [cit. 2015-01-10]. Dostupné z: http://drive5.com/usearch/manual/fastq_files.html
41. NYRÉN, P. Enzymatic method for continuous monitoring of DNA polymerase activity. *Analytical Biochemistry*. 1987, vol. 167, issue 2, s. 235-238
42. *FASTX-Toolkit - Galaxy Usage* [online]. Poslední revize 12.2014 [cit. 2014-12-09]. Dostupné z: http://hannonlab.cshl.edu/fastx_toolkit/galaxy.html
43. MAJEWSKI, J., ROSENBLATT D.S. Exome and whole-genome sequencing for gene discovery: The future is now!. *Human Mutation*. 2012, vol. 33, issue 4, s. 591-592.
44. GUAN, Y.F. et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chinese Journal of Cancer*. 2012, vol. 31, issue 10.
45. KOUBKOVÁ, L. et al. Next Generation Sequencing – Application in Clinical Practice: Sanger and Next-Generation Platforms. *Klinická onkologie*. 2014, vol. 27, issue 1, s. 81-85.
46. CASTÉRA, L. et al. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *European Journal of Human Genetics*. 2014, vol. 22, issue 11, s. 1305-1313.
47. WEI, X. et al. Targeted next-generation sequencing as a comprehensive test for patients with and female carriers of DMD/BMD: a multi-population diagnostic study. *European Journal of Human Genetics*. 2013, vol. 22, issue 1, s. 110-118.

48. NIBA, E.T.E. et al. Validation of ambiguous MLPA results by targeted next-generation sequencing discloses a nonsense mutation in the DMD gene. *Clinica Chimica Acta*. 2014, vol. 436, s. 155-159.
49. SIMON, R., ROYCHOWDHURY S. Implementing personalized cancer genomics in clinical trials. *Nature Reviews Drug Discovery*. 2013, vol. 12, issue 5, s. 358-369.
50. *What kinds of gene mutations are possible? - Genetics Home Reference*[online]. Poslední revize 10.2014 [cit. 2014-10-11]. Dostupné z: <http://ghr.nlm.nih.gov/handbook/mutationsanddisorders/possiblemutations>
51. ÚSTAV ZDRAVOTNICKÝCH INFORMACÍ A STATISTIKY ČR. *Novotvary 2011* [online]. Praha, 2015 [cit. 2015-01-14]. ISBN 978-80-7472-097-0. Dostupné z: <http://www.uzis.cz/publikace/novotvary-2011>
52. LEDFORD, H. Gene reading steps up a gear. *Nature*. 2011, vol. 470, issue 7333, s. 155-155.
53. DUŠEK, L. et al. Epidemiology of Screening-targeted Cancers According to New Data of the Czech National Cancer Registry. *Klinická onkologie*. 2014, vol. 27, issue 2, s. 2519-2539
54. *Overview of SOLiD™ Sequencing Chemistry* [online]. Poslední revize 01.2015 [cit. 2015-01-21]. Dostupné z: <http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html>
55. MUZNY, D.M. et al. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012, vol. 487, issue 7407, s. 330-337.
56. WARREN, R. et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome*. 2013, vol. 1, issue 1.
57. SESHAGIRI, S. et al. Recurrent R-spondin fusions in colon cancer. *Nature*. 2012, vol. 488, issue 7413, s. 660-664.
58. BAGNALL, R. D. et al. Molecular Diagnostics of Cardiomyopathies: The Future Is Here. *Clinical Cardiovascular Genetics*. 2011, vol. 4, issue 2, s. 103-104.
59. EASTON, D.F. Breast and ovarian cancer incidence in BRCA1-mutation carriers. *Human Genetics*. 1995, volume 56, issue 1, s. 265-271.
60. FORD, D. et al. Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families. *The American Journal of Human Genetics*. 1998, vol. 62, issue 3, s. 676-689.
61. KURIAN, A. et al. Clinical Evaluation of a Multiple-Gene Sequencing Panel for Hereditary Cancer Risk Assessment. *Journal of Clinical Oncology*. 2014, vol. 32, issue 19, s. 2001-2009.
62. TRIPATHY, D. et al. Next generation sequencing and tumor mutation profiling: are we ready for routine use in the oncology clinic?. *BMC Medicine*. 2014, vol. 12, issue 1.

63. VESELKA, J. *Hypertrofická kardiomyopatie* [online]. 2012 [cit. 2015-01-13]. Dostupné z: http://www.kardio-cz.cz/resources/upload/data/473_veselka-hypertrofickakardiomyopatie.pdf.
64. TOMAŠOV, P. *Genetika kardiomyopatii* [online]. 2012 [cit. 2015-01-13]. Dostupné z: http://www.kardio-cz.cz/resources/upload/data/466_tomasov-genetikakardiomyopatii.pdf
65. AWAD, M. et al. Mechanisms of Disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Nature Clinical Practice Cardiovascular Medicine*. 2008, vol. 5, issue 5, s. 258-267.
66. YIN, X. et al. Massively Parallel Sequencing for Chromosomal Abnormality Testing in Trophectoderm Cells of Human Blastocysts. *Biology of Reproduction*. 2013, vol. 88, issue 3, s. 69-69.
67. *VeriSeq PGS* [online]. Poslední revize 02.2015 [cit. 2015-02-20]. Dostupné z: http://www.illumina.com/products/veriseq_pgs.html
68. ROMMENS, J.M. Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Trends in Genetics*. 1989, vol. 5, s. 1059-1065
69. *CFTR MASTR Dx: A simple and robust CE-IVD validated diagnostic assay for MPS-based mutation analysis of the complete CFTR gene* [online]. Poslední revize 04.2015 [cit. 2015-04-13]. Dostupné z: <http://www.multiplicom.com/products/cftr-mastr-dx>
70. CHIEN-YUEH, L. Common applications of next-generation sequencing technologies in genomic research. *Translational Cancer Research*. 2013, vol. 2, issue 1. s. 33-45.
71. *Too good to be true?! What can Nextera do for you?* [online]. Poslední revize 02.2015 [cit. 2015-02-20]. Dostupné z: <http://bitesizebio.com/13567/too-good-to-be-true-what-can-nextera-do-for-you/>
72. SHENDURE, J., Hanlee J. Next-generation DNA sequencing. *Nature Biotechnology*. 2008, volume 26, issue 10, s. 1135-1145
73. The Illumina HiSeq 2000 Sequencing Technology [online]. Poslední revize 02.2015 [cit. 2015-02-20]. Dostupné z: http://www.dkfz.de/gpcf/hiseq_technology.html
74. METZKER, M. 2009-12-8. Sequencing technologies - the next generation. *Nature Reviews Genetics*. 2009, vol. 11, issue 1, s. 31-46.
75. *How the Ion Torrent Sequencer works - Bitesize Bio* [online]. Poslední revize 02.2015 [cit. 2015-02-20]. Dostupné z: <http://bitesizebio.com/13547/how-the-ion-torrent-sequencer-works/>
76. *Oxford Nanopore Technologies* [online]. Poslední revize 02.2015 [cit. 2015-02-22]. Dostupné z: <https://nanoporetech.com/news/movies#movie-24-nanopore-dna-sequencing>
77. BRUNECKY, Z. The incidence and genetics of cystic fibrosis. *Journal of Medical Genetics*. 1972, volume 9, issue 1, s. 33-37.

78. *Cystická fibróza - Novorozenecký screening* [online]. Poslední revize 05.2015 [cit. 2015-05-10]. Dostupné z: <http://www.novorozeneckyscreening.cz/cysticka-fibroza>
79. MACEK, M. *Doporučení týkající se informovaného souhlasu pro genetická laboratorní vyšetření* [online]. Poslední revize 05.2013 [cit. 2015-02-10]. Dostupné z: <http://www.slg.cz/system/files/informovany-souhlas-doporuceni-2013.pdf>
80. *Cystická fibróza - Novorozenecký screening* [online]. Poslední revize 05.2015 [cit. 2015-05-10]. Dostupné z: <http://www.novorozeneckyscreening.cz/cysticka-fibroza>
81. *MiSeq Gene & Small Genome Sequencer* [online]. Poslední revize 05.2015 [cit. 2015-05-01]. Dostupné z: <http://www.illumina.com/systems/miseq.html>
82. *HiSeq 2500/1500* [online]. Poslední revize 05.2015 [cit. 2015-05-01]. Dostupné z: http://genomeinformaticsalliance.org/systems/hiseq_2500_1500.ilmn

14. PŘÍLOHY

Příloha 1: Nalezené předpokládané kauzální mutace na přístroji MiSeq

Vzorek:	Nalezená mutace:	Mutovaný gen:	Chromozom:	Typ mutace:
KM2	NM_000256.3:c.3130C>T	<i>MYBPC3</i>	11	stop gain
KM3	NM_000256.3:c.3697C>T	<i>MYBPC3</i>	11	stop gain
KM5	NM_000256.3:c.821+1G>A	<i>MYBPC3</i>	11	stop gain
KM8	NM_003803.3:c.4222G>A	<i>MYOM1</i>	18	missense
	NM_001099404.1:c.1673A>G	<i>SCN5A</i>	3	missense
KM9	NM_000257.2:c.3109T>C	<i>MYH7</i>	14	missense
KM10	NM_000256.3:c.3413G>C	<i>MYBPC3</i>	11	missense
KM13	NM_001943.3:c.877A>G	<i>DSG2</i>	18	missense
	NM_003276.2:c.1243T>G	<i>TMPO</i>	14	missense
KM15	NM_001105206.2:c.2419C>A	<i>LAMA4</i>	6	missense
KM16	NM_000256.3:c.2687T>G	<i>MYBPC3</i>	11	missense
KM17	NM_001103.2:c.1822C>T	<i>ACTN2</i>	1	missense
KM19	NM_002471.3:c.345+1G>T	<i>MYH6</i>	14	splice variant
KM20	NM_002471.3:c.1876G>T	<i>MYH6</i>	14	missense
	NM_000258.2:c.307+1G>T	<i>MYL3</i>	3	splice variant
	NM_001099404.1:c.1673A>G	<i>SCN5A</i>	3	missense
KM21	NM_000256.3:c.3190+5G>A	<i>MYBPC3</i>	11	splice variant
KM23	NM_000256.3:c.3697C>T	<i>MYBPC3</i>	11	stop gain
KM24	NM_000256.3:c.1484G>A	<i>MYBPC3</i>	11	missense
	NM_000256.3:c.1879G>T	<i>MYBPC3</i>	11	missense
	NM_000257.2:c.5774G>T	<i>MYH7</i>	14	missense
	NM_016203.3:c.298G>T	<i>PRKAG2</i>	7	missense
KM25	NM_000256.3:c.1420G>T	<i>MYBPC3</i>	11	stop gain
KM27	NM_000256.3:c.1484G>A	<i>MYBPC3</i>	11	missense
	NM_000256.3:c.1558G>T	<i>MYBPC3</i>	11	stop gain
KM29	NM_000256.3:c.1558G>T	<i>MYBPC3</i>	11	stop gain
KM30	NM_000256.3:c.3490+1G>T	<i>MYBPC3</i>	11	splice variant
KM31	NM_000256.3:c.1790G>A	<i>MYBPC3</i>	11	missense
	NM_002230.2:c.130G>T	<i>JUP</i>	17	stop gain
KM32	NM_004415.2:c.88G>A	<i>DSP</i>	6	missense
KM33	NM_002471.3:c.4822C>T	<i>MYH6</i>	14	missense
KM35	NM_000256.3:c.406+3G>T	<i>MYBPC3</i>	11	splice variant
	NM_000257.2:c.2081G>A	<i>MYH7</i>	14	missense
	NM_016203.3:c.350C>T	<i>PRKAG2</i>	7	missense
	NM_000257.2:c.2081G>A	<i>MYH7</i>	14	missense
KM36	NM_000257.2:c.428G>A	<i>MYH7</i>	14	missense
KM37	NM_000257.2:c.3286G>T	<i>MYH7</i>	14	missense
	NM_000256.3:c.649A>G	<i>MYBPC3</i>	11	missense
	NM_033118.3:c.834T>A	<i>MYLK2</i>	20	missense
KM38	NM_000257.2:c.1207C>T	<i>MYH7</i>	14	missense
	NM_033118.3:c.430C>G	<i>MYLK2</i>	20	missense
KM39	NM_016203.3:c.337T>A	<i>PRKAG2</i>	7	missense
KM41	NM_033118.3:c.1769C>A	<i>MYLK2</i>	20	missense

KM42	NM_002471.3:c.292G>A	<i>MYH7</i>	14	missense
KM44	NM_000257.2:c.1436A>G	<i>MYH7</i>	14	missense
KM45	NM_000256.3:c.724T>C	<i>MYBPC3</i>	11	missense
KM46	NM_000257.2:c.1979C>A	<i>MYH7</i>	14	missense
KM48	NM_033118.3:c.1768G>A	<i>MYLK2</i>	20	missense
KM49	NM_000256.3:c.2309-3C>G	<i>MYBPC3</i>	11	splice variant
KM50	NM_003476.4:c.208G>T	<i>CSRP3</i>	11	missense
KM51	NM_000256.3:c.3697C>T	<i>MYBPC3</i>	11	stop gain
KM52	NM_000256.3:c.373_374delGC	<i>MYBPC3</i>	11	frameshift
KM53	NM_000256.3:c.3697C>T	<i>MYBPC3</i>	11	stop gain
KM54	NM_033118.3:c.430C>G	<i>MYLK2</i>	20	missense
KM55	NM_004415.2:c.2438A>T	<i>DSP</i>	6	missense
KM56	NM_000256.3:c.906-7G>T	<i>MYBPC3</i>	11	splice variant
KM57	NM_000256.3:c.821+1G>A	<i>MYBPC3</i>	11	splice variant
KM59	NM_033337.2:c.233C>T	<i>CAV3</i>	3	missense
KM60	NM_000257.2:c.4472C>G	<i>MYH7</i>	14	missense
KM61	NM_001103.2:c.1304C>T	<i>ACTN2</i>	1	missense
KM62	NM_002471.3:c.166G>A	<i>MYH6</i>	14	missense
KM64	NM_000257.2:c.5561C>T	<i>MYH7</i>	14	missense
	NM_000363.4:c.497C>T	<i>TNNI3</i>	19	missense
KM65	NM_000256.3:c.3490+1G>T	<i>MYBPC3</i>	11	splice variant