

Univerzita Palackého v Olomouci

Bakalářská práce

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra buněčné biologie a genetiky

**Konstrukce molekulárních knihoven pro masivně paralelní
sekvenování**

Bakalářská práce

Marek Pětoň

Studijní program: Biologie

Studijní obor: Molekulární a buněčná biologie

Forma studia: Prezenční

Olomouc 2015

Vedoucí práce: Mgr. Zuzana Macečková

Čestně prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením
Mgr. Zuzany Macečkové, za použití citovaných zdrojů.

V Olomouci dne:

.....

Poděkování

Chtěl bych poděkovat za vedení a rady slečně Mgr. Zuzaně Macečkové, a za pomoc a konzultace panu Mgr. Petru Vojtovi, slečně Mgr. Natálii Táborské, panu Ing. Rastislavu Slavkovskému Ph.D., a také slečně Mgr. Haně Ondryášové.

Souhrn

Molekulární knihovny jsou soubory fragmentovaných částí DNA. Setkáváme se s cDNA knihovnami, genomovými knihovnami nebo randomizovanými mutantními knihovnami. Molekulární knihovny jsou používány při klonování, fyzickém mapování, molekulární cytogenetice, porovnávací genomice a při genomovém sekvenování.

V této bakalářské práci je popsán základní historický vývoj molekulárních knihoven, od knihoven využívajících uměle vytvořené chromozomy po, v dnešní době nejvíce využívané, speciální knihovny pro masivně paralelní sekvenování. Vzhledem k faktu, že se knihovny využívají především k sekvenování jsou v dalších kapitolách popsány sekvenační metody nové generace a u každé je podrobně rozebrána příprava molekulární knihovny pro daný sekvenční systém.

Experimentální část této práce se zaměřuje na optimalizaci přípravy sekvenčních knihoven pro transkriptomové sekvenování u platformy Illumina.

Summary

Molecular libraries are fragmented parts of DNA. Common are cDNA libraries, genomic libraries or randomized mutant libraries. Molecular libraries are used in cloning, physical mapping, molecular cytogenetics, comparative genomics and genomic sequencing.

In this thesis is described the basic historical development of molecular libraries, from the libraries using artificial chromosomes to nowadays the most used special libraries for massively parallel sequencing. Due to the fact that the libraries are primarily used for sequencing, they are described in several chapters. Next chapters are for the methods of new generation sequencing systems and for each of them is described in detail the preparation of molecular libraries.

The experimental part of this work is focused on optimisation of the preparation of sequencing libraries for transcriptomics sequencing on Illumina platform.

Obsah

1. Úvod.....	8
2. Cíl práce	9
3. Současný stav dané problematiky	10
3.1. Umělý chromozom.....	10
3.1.1. Kvasinkový umělý chromozom	10
3.1.2. Bakteriální umělý chromozom	11
3.2. Sekvenování	12
3.2.1. Původní sekvenační metody.....	12
3.2.1.1. Sangerovo sekvenování	12
3.2.2. Sekvenační metody nové generace	15
3.2.2.1. Roche/454 - Pyrosekvenace.....	15
3.2.2.1.1. Postup při sekvenování metodou 454/Roche	15
3.2.2.1.2. Nevýhody platformy Roche/454	18
3.2.2.2. Illumina.....	19
3.2.2.2.1. Příprava RNA knihoven	19
3.2.2.2.2. Příprava DNA knihoven	20
3.2.2.2.3. Další zpracování knihoven	21
3.2.2.2.4. Vytváření klastrů	21
3.2.2.2.5. Sekvenace	23
3.2.2.2.6. Analýza dat.....	25
3.2.2.2.6.1. Analýza sekvenčních dat po transkriptomovém sekvenování.....	26
3.2.2.2.7. Aktuální nabídka přístrojů společnosti Illumina	27
3.2.2.3. Využití sekvenačních metod nové generace	28
3.2.2.3.1. Celogenomové sekvenování.....	28
3.2.2.3.2. Sekvenování transkriptomu	29
3.2.2.3.3. Cílené sekvenování vybraných oblastí genomu	30
3.2.2.3.4. Exomové sekvenování.....	31
4. Materiál a metody	32
4.1. Materiál	32
4.1.1. Biologický materiál	32
4.1.2. Kity.....	32
4.1.3. Použité chemikálie, roztoky a jejich složení	32
4.1.4. Přístrojové vybavení.....	33
4.2. Metody	34
4.2.1. Izolace celkové RNA z lidských buněk	34
4.2.2. Izolace mRNA z celkové RNA	35
4.2.3. Fragmentace mRNA.....	37

4.2.4.	Reverzní transkripce mRNA	37
4.2.5.	Příprava sekvenační knihovny z cDNA	38
4.2.5.1.	Modifikace konců	38
4.2.5.2.	Ligace adaptorů	38
4.2.5.3.	Přečištění	39
4.2.5.4.	Degradace uracilem značeného vlákna	39
4.2.5.5.	Amplifikace knihovny	39
5.	Výsledky	41
5.1.	Izolace celkové RNA	41
5.2.	Izolace mRNA	42
5.3.	Fragmentace mRNA	44
5.4.	Reverzní transkripce	46
6.	Diskuze	53
7.	Závěr	56
8.	Zdroje	57
9.	Seznam zkratk	65

1. Úvod

Sekvenování je proces, pomocí kterého je určováno pořadí nukleotidů ve zkoumané DNA. Jeho počátky sahají do 70. let 20. století, přesněji do roku 1977, kdy byly vyvinuty dvě metody umožňující sekvenování nukleových kyselin. Jejich autory byl tým okolo Fredericka Sangera (Sangerovo sekvenování) a kolektiv Allana Maxama a Waltera Gilberta (Maxam-Gilbertovo sekvenování). Sangerovo sekvenování se stalo na dlouhá desetiletí zlatým standardem, ale i přes modifikace této metody je její využití velmi nákladné a pracné. To vedlo k dalšímu výzkumu v oblasti sekvenačních metod. Postupně se začaly objevovat sekvenační metody nové generace (z anglického názvu „Next-Generation Sequencing“ - NGS). Tato technologie využívá principu masivně paralelního sekvenování, což umožnilo růst kapacity sekvenátorů a dramaticky snížilo cenu sekvenování [1].

Sekvenační metody jsou v dnešní době jedním z nejdůležitějších nástrojů molekulárních biologů. Uplatnění nachází v mnoha vědních oborech, od medicíny po systematickou biologii. Pomáhají při objasňování podstaty mnoha dědičných onemocnění, zkoumání rakoviny, virů, bakterií či umožňují sledovat vnitrodruhovou modifikaci a adaptaci.

Drtivá většina sekvenátorů využívá ke své práci molekulárních knihoven. Prvními molekulárními knihovnami byly takzvané kvasinkové umělé chromozomy, které byly po určité době nahrazeny bakteriálními. Stále větší náročnost na kvalitní a rychlou přípravu knihoven vedla k vývoji sekvenačních knihoven pro metody sekvenování nové generace.

Tato bakalářská práce v teoretické části shrnuje vývoj molekulárních knihoven a také vývoj sekvenačních metod od metody Sangerova sekvenování po sekvenační metody nové generace s největším důrazem na sekvenační platformu Illumina. Praktická část práce je zaměřena na optimalizaci procesu přípravy Illumina knihoven pro sekvenování transkriptomu..

2. Cíl práce

1. Vypracování literární rešerše na téma – Molekulární knihovny se zaměřením na molekulární knihovny pro masivně paralelní sekvenování.
2. Izolace celkové RNA a mRNA z buněčných linií.
3. Reverzní transkripce mRNA do cDNA za použití Strand-Specific metody.
4. Příprava molekulární knihovny ze získané cDNA.
5. Analýza výsledků a vypracování bakalářské práce.

3. Současný stav dané problematiky

3.1. Umělý chromozom

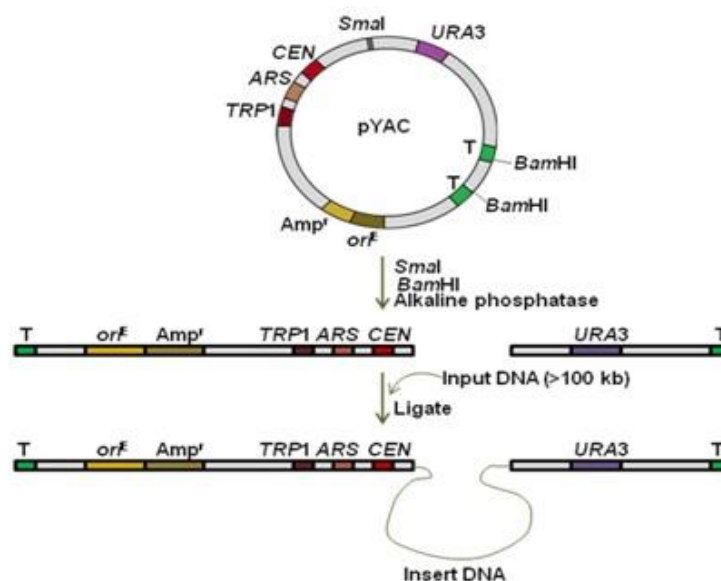
Původní molekulární knihovny využívaly technologie takzvaných umělých chromozomů (z anglického „artificial chromosomes“ (AC)), což jsou uměle upravené hostitelské chromozomy, do kterých jsou vloženy fragmenty DNA. Velké využití nalézají při klonování větších fragmentů DNA, od 300 000 po až 1 milion párů bazí, podle použitého typu umělého chromozomu.

Jsou využívány dva typy umělých chromozomů: Bakteriální umělý chromozom (z anglického „bacterial artificial chromosome“ (BAC)) a kvasinkový umělý chromozom (z anglického „yeast artificial chromosome“ (YAC)). Rozdíly mezi těmito typy jsou rozepsány níže v navazujících kapitolách [2].

Využití nalézají umělé chromozomy při mapování založeném na klonování, fyzickém mapování, molekulární cytogenetice, porovnávací genomice a při genomovém sekvenování. Používání umělých chromozomů ve své době způsobilo revoluci v biologických studiích [3].

3.1.1. Kvasinkový umělý chromozom

V osmdesátých letech byl vytvořen kvasinkový umělý chromozom, který umožnil první na knihovně založeném zkoumání velkých genomů. Do YAC je možné vložit fragmenty o velikosti až 1 Mb DNA [4]. Mapa YAC plazmidu (pYAC) je zobrazena na obrázku č.1.



Obrázek č.1: Mapa YAC plazmidu (pYAC) a vyobrazení YAC klonovacího systému (převzato) [5]

Měl však několik negativních stránek [6]. Docházelo například k tomu, že 50 % klonů YAC bylo chimérických (klon obsahuje fragmenty z více částí genomu), nebo také k vnitřní přestavbě insertu [6-8]. To bylo způsobeno rekombinací uvnitř kvasinkového vektoru. [9] Také vzhledem k časové náročnosti, při manipulaci a izolaci YAC insertu, bylo od této metody upuštěno [10].

3.1.2. Bakteriální umělý chromozom

V devadesátých letech, byla vyvinuta alternativa k YAC a to bakteriální umělý chromozom. Maximální velikost insertu byla menší než u YAC, tedy okolo 500 kb a méně. Běžnější praxí je velikost insertu okolo 80-200 kb [2, 10]. Důvodem používání menšího insertu je velikost genomu bakterií, který je podstatně menší než v případě kvasinek. Pokud by insert byl větší, docházelo by ke zvýšení počtu rekombinací. Hlavní výhodou BAC klonů oproti YAC je relativně lehká manipulovatelnost, která je dána velikostí bakteriálních vektorů a také rychlostí, jakou se bakterie množí [11].

3.2. Sekvenování

V první části této kapitoly se zaměřím na starší metodu, kterou je Sangerovo sekvenování. Následovat budou kapitoly o metodách masivně paralelního sekvenování na platformách Roche/454 a Illumina

Vývoj v oblasti nových sekvenačních metod si vyžádal změny přístupu k přípravě sekvenačních knihoven. NGS technologie umožňuje v řádu několika hodin osekvenovat desítky milionů bazí, a proto se BAC knihovny, které vyžadují zdlouhavou přípravu, se pro tyto nové a především rychlé sekvenační systémy, staly nepraktickými [12].

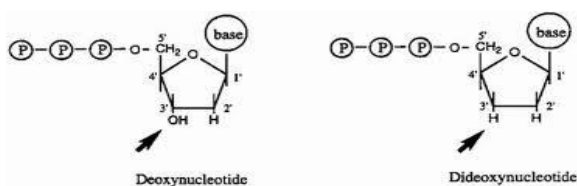
Metody vhodné pro přípravu molekulárních knihoven pro sekvenování nové generace jsou popsány v následujících kapitolách, vždy také s popisem, jak samotná sekvenace na dané platformě probíhá. To z toho důvodu, že podoba knihovny pro daný sekvenční systém je vždy specifická a nepřevoditelná mezi platformami. Je to dáno rozdílnou podobou adaptorů, kterou jsou k přípravě knihoven použity a také způsobem sekvenace na jednotlivých platformách.

3.2.1. Původní sekvenační metody

3.2.1.1. Sangerovo sekvenování

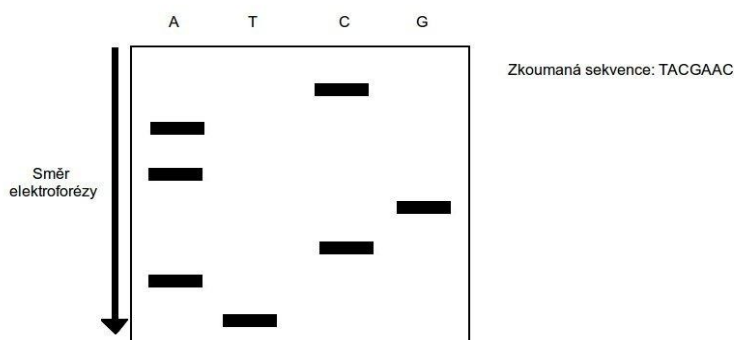
Metoda navržená a publikována britským biochemikem Frederickem Sangerem roku 1977, je založena na principu selektivního zabudovávání dideoxynukleotidu (ddNTP) na syntetizované vlákno DNA. Po desetiletí byla tato metoda zlatým standardem sekvenování [13]. Za tento objev byla Fredericovi Sangerovi v roce 1980 udělena Nobelova cena za chemii.

Nukleotidy používané při syntéze komplementárního vlákna DNA jsou dvojího typu: deoxynukleotidy (dNTP) a ddNTP. Cukernou složkou ddNTP je 2,3-dideoxyribóza, ta znemožňuje dalším nukleotidům se na tento dinukleotid navázat a terminuje tedy syntézu vlákna. Naopak klasické dNTP mají jako cukernou složku 2-deoxyribózu, která navázání dalších nukleotidů nebrání. (Obr. č. 2) Syntéza dle původních specifikací metody probíhá ve čtyřech rozdílných reakčních směsích, kdy každá reakční směs obsahuje 3 typy deoxynukleotidů a čtvrtý nukleotid je dodán v podobě směsi deoxynukleotidu a dideoxynukleotidu (nejčastěji v poměru 99:1) [13 - 16].



Obrázek č.2: Chemická struktura deoxynukleotidu a dideoxynukleotidu (převzato) [17]

Výsledkem byla rozdílně dlouhá vlákna nukleové kyseliny, která se detekovala pomocí agarosové elektroforézy. Z elektroforetogramu se dala vyčíst sekvence zkoumané nukleové kyseliny (Obr. č. 3).

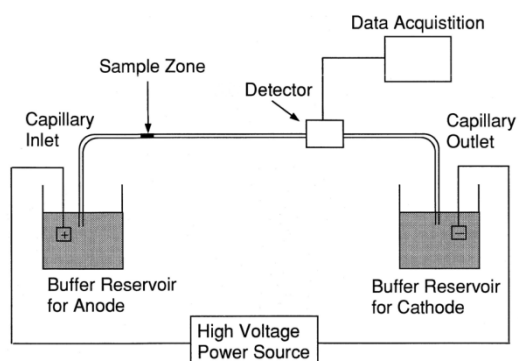


Obrázek č.3: Výsledný elektroforetogram zkoumané sekvence při použití Sangerova sekvenování

Metoda Sangerova sekvenování prošla v průběhu let mnohými vylepšeními. V dnešní době se již nepoužívá agarózový gel na detekci zkoumané sekvence, ani se tato sekvence nevyhodnocuje lidským okem. Celý systém je automatizovaný a je využíváno takzvaného kapilárního sekvenování (kapilární elektroforéza).

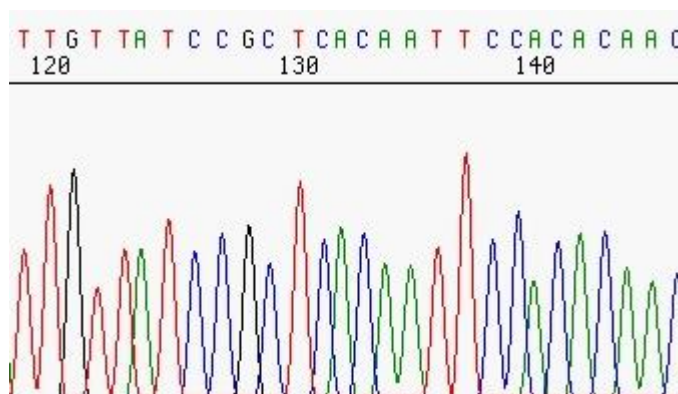
Kapilární sekvenování je akorát jiným uspořádáním Sangerovy metody. Je prováděno v jedné reakční zkumavce, ve které se nachází směs všech čtyř klasických nukleotidů a čtyř fluorescenčně značených dideoxynukleotidů, které slouží jako terminátory. Každý typ z celkových čtyř typů ddNTP ve zkumavce je značen jinou fluorescenční značkou, což znamená, že emituje světlo o jiné vlnové délce.

Následně je využito takzvané kapilární elektroforézy. Při tomto typu elektroforézy je zkumavka s produkty připojena na katodu. Do zkumavky je vložena úzká kapilára, kterou jsou zkoumané molekuly vedeny k anodě. Molekuly se tímto třídí dle velikosti. Kapilára je prosvěcována laserem, čímž dochází k emisi jedné ze čtyř vlnových délek z fluorescenčně značených nukleotidů [18 - 21]. Podoba kapilárního sekvenátoru je zobrazena na Obrázku č.4.



Obrázek č.4: Zjednodušené schéma kapilárního sekvenátoru (převzato) [22]

Data z kapilárního sekvenátoru jsou interpretována na počítači v softwaru dodávaném se sekvenátorem. Jsou v podobě elektroforetogramu, ze kterého určujeme kromě sekvence i kvalitu čtení sekvence. Kvalitu určuje podoba vyobrazených píků pro jednotlivé nukleotidy (obrázek č.5) [18].



Obrázek č.5: Podoba elektroforetogramu (převzato) [23]

3.2.2. Sekvenační metody nové generace

Metody sekvenování nové generace jsou založeny na principu masivně paralelního sekvenování. Název pochází z anglického „Massively Parallel Signature Sequencing“ - MPSS. Tento princip využívá dva základní způsoby sekvenování: sekvenace založena na ligaci („ligation based sequencing“ - LBS) a sekvenaci syntézou („sequencing by synthesis“-SBS) [24 - 25].

V následujících kapitolách budou zmíněny dvě nejpoužívanější a nejrozšířenější z nich: Roche/454 a Illumina.

3.2.2.1. Roche/454 - Pyrosekvenace

Technologie pyrosekvenace vyvinutá švédským biochemikem Pållem Nyrénem, využívá detekce luminiscence při procesu zabudovávání označeného nukleotidu do syntetizovaného vlákna [26 - 27]. Tato metoda je tedy sekvenační metodou založenou na syntéze (z anglického „Sequencing by synthesis“ - SBS).

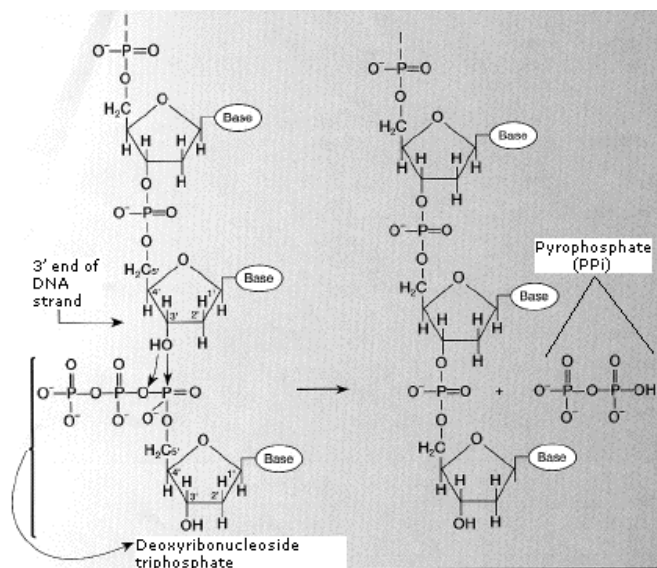
První komerčně využitelný přístroj a první sekvenátor vůbec, založený na principu pyrosekvenace, byl Roche FLX Genome Sequencer v roce 2006 od konsorcia firem 454 Life Sciences a Roche Diagnostics [28].

3.2.2.1.1. Postup při sekvenování metodou 454/Roche

Vlákno nukleové kyseliny je v prvním kroku fragmentováno na délku mezi 250-800 bazí. Na získané fragmenty jsou navázány adaptory. Adaptory jsou velmi specifické sekvence oligonukleotidů [29]. V následném kroku je využito amplifikace fragmentů s navázanými adatory metodou zvanou emulzní PCR (emPCR), při které je využíváno mikrokuliček, na které se vážou fragmenty s adaptory [30]. Dalším krokem je centrifugace, během které jsou kuličky s navázanými amplikony přeneseny do jamek na čip tvořený optickými vlákny (PicoTiterPlate). Velikost mikrokuliček (poloměr ~28 μm) zajišťuje, aby se do každé jamky dostala téměř vždy jen jedna mikrokulička. Chyba nastává ve 2-5 % případů [31].

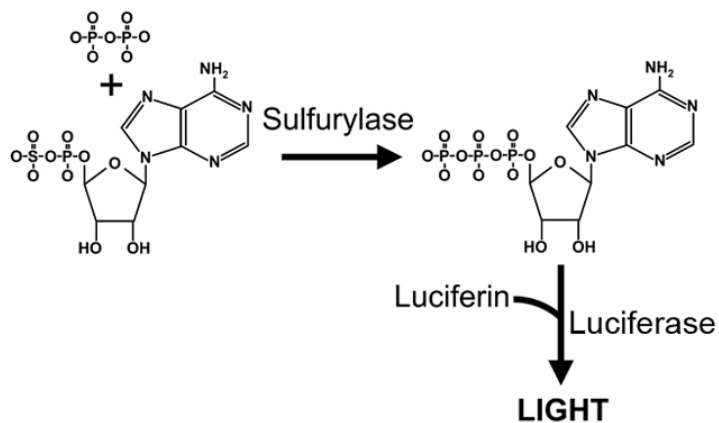
Jakmile jsou jamky na čipu osazeny mikrokuličkami, jsou k nim přidány enzymy: DNA polymeráza, ATP sulfyráza, luciferáza a apyráza. Dále přidaná směs obsahuje adenosin fosfosulfát (APS) a luciferin [32 - 33].

K mikrokuličkám je přidán roztok konkrétního pyrofosfátem značeného nukleotidu. Pyrofosfát je na nukleotid navázán přes fosfátovou skupinu. Pokud je takto značený nukleotid komplementární k templátu syntetizovaného vlákna, tak je zabudován DNA polymerázou. Struktura nukleotidu a pyrofosfátu je zobrazena na obrázku č.6.



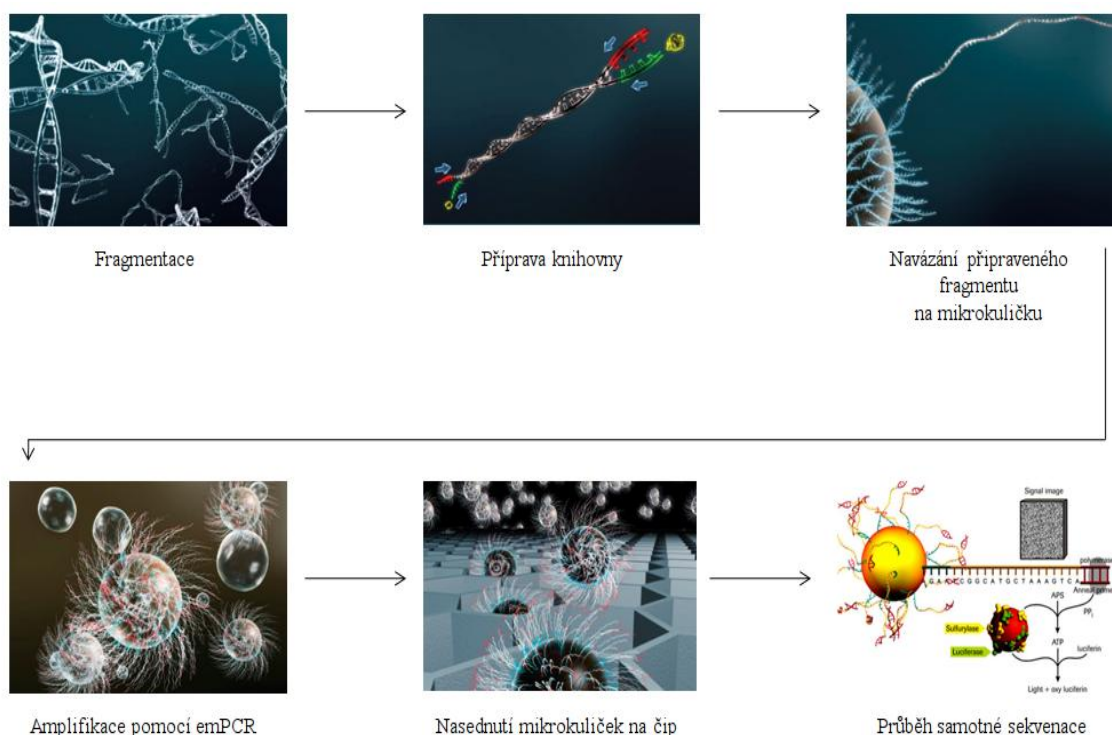
Obrázek č.6: Struktura nukleotidu a pyrofosfátu (převzato) [34]

Uvolněný pyrofosfát reaguje pomocí ATP sulfurylázy v reakci s APS, čímž vznikne ATP. Vzniklá ATP je využita luciferázou při převodu luciferinu na oxyluciferin. Schéma reakce je vidět na obrázku č. 7.



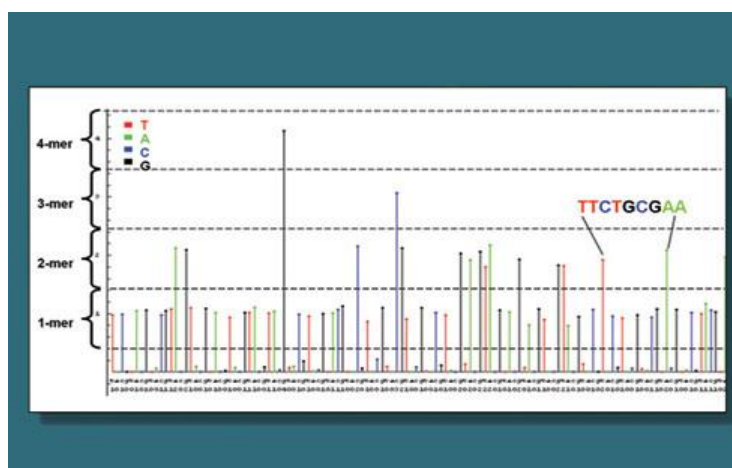
Obrázek č.7: Schéma reakce pyrofosfátu s APS, jehož výsledkem je emise světelného záření. (převzato) [35]

Během tohoto převodu se uvolní světelné záření, které je detekováno (viz. níže). V posledním kroku jsou apyrázou degradovány neinkorporované nukleotidy a ATP. Celý proces je následně opakován s dalšími nukleotidy [36]. Celý postup přípravy a průběhu pyrosekvenování je zobrazen na obrázku č. 8.



Obrázek č.8: Postup při sekvenaci metodou 454/Roche (převzato a upraveno) [37]

Detekce světelného záření probíhá pomocí CCD kamery (z anglického „The Charge-Coupled Device) [38]. Konečným datovým výstupem je graf intenzity světelného záření, pomocí kterého je určena zkoumaná sekvence. V závislosti na intenzitě světelného záření se také určuje, jestli se za sebou nenachází dvě a více stejných bází. Datový výstup je viditelný na Obrázku č.9 [37].



Obrázek č.9: Datový výstup při používání sekvenačního systému od firmy 454/Roche (Převzato) [37]

3.2.2.1.2. Nevýhody platformy Roche/454

Hlavní nevýhodou této platformy je chybovost při vyhodnocování repetitivních nukleotidů, kdy při určitém počtu stejných nukleotidů jdoucích po sobě, je intenzita signálu stejná. Tím pádem není možné určit přesný počet nukleotidů v sekvenci [39].

Mezi další nevýhody této sekvenační platformy patří takzvané negativní či pozitivní posuny čtecího rámce.

Negativní posuny se projevují s četností 0,1 – 0,3 % a projevují se delecemi ve výsledné sekvenci. Jsou způsobovány neúplným začleněním nukleotidu především v homopolymerních oblastech. Na vině je v tomto případě DNA polymeráza, která svou nedostatečnou exonukleázovou aktivitou způsobuje neúplné prodloužení řetězce [40 - 41].

Pozitivní posuny se projevují s větší četností a to 1 – 2 %. Jsou způsobeny inzercí nukleotidů, které pocházejí z předchozího cyklu. Jedná se tedy o nedostatečnou aktivitu enzymu apyrázy [32].

3.2.2.2. Illumina

Sekvenační systémy od společnosti Illumina využívají také metody SBS. Proces sekvenování na platformách Illumina se skládá ze 3 částí: přípravy knihovny, vytváření klastrů a samotné sekvenace.

Příprava sekvenačních knihoven je velmi důležitým stádiem celého procesu. V této části vzniká nejvíce chyb, které při nedůsledné kontrole mohou v konečném důsledku způsobit až znehodnocení výsledků celého sekvenování.

3.2.2.2.1. Příprava RNA knihoven

Po izolaci RNA z organického materiálu dostaneme soubor RNA molekul, z nichž některé ale nekódují žádné proteiny. Největším zástupcem těchto nekódujících RNA molekul je rRNA, která tvoří mnohdy až 95 % celkové RNA v buňce. Pro naše zkoumání je však rRNA nepoužitelná a je nutné se jí tedy zbavit. Pokud bychom ji ve zkoumaném vzorku zanechali, docházelo by k plýtvání kapacity sekvenátoru, jelikož bychom sekvenovali část nukleové kyseliny, která pro nás není předmětem zkoumání.

Izolace mRNA se provádí v zásadě dvěma způsoby. Buď navázáním poly (a) řetězce na 3'-konec mRNA. Druhým způsobem získání mRNA je takzvaná deplece rRNA, pomocí RiboZero kitu. Tento purifikační kit obsahuje specifické sondy, které se navážou na rRNA. Sondy s rRNA jsou posléze navázány na magnetické kuličky, které jsou zachyceny magnetem. Zbylý roztok zbavený rRNA je odpipetován a použit pro další práci [42 - 43].

Po vyzolování mRNA dochází k fragmentaci vzorku na potřebnou velikost. Fragmentací zajišťujeme, aby byl sekvenátor schopen přečíst kompletní zkoumanou sekvenci. Jednotlivé přístroje mají různý rozsah toho, jak dlouhou sekvenci jsou schopny analyzovat. V případě platformy Illumina, přesněji pro přístroj Illumina HiSeq 2500, je ideální délka fragmentů od 200 do 300 bazí bez započtených adaptorů [44 - 45] (Tabulka č.2).

Fragmentace mRNA je prováděna většinou pomocí teploty, kdy je vzorek mRNA smíchan s fragmentačním pufrem, který obsahuje hořečnaté nebo zinečnaté ionty a dále také Tris-HCl. Směs je poté vložena do termocykleru, ve kterém je ponechána při 94°C na dobu potřebnou k nafragmentování na určitou velikost. Dochází k denaturaci vazeb mezi jednotlivými nukleotidy.

Získané fragmenty o požadované velikosti je nutné přepsat z RNA do DNA pomocí reverzní transkripce a následné syntézy komplementárního vlákna. Velmi často je využíváno u přepisu RNA do cDNA takzvané strand-specific metody. Získaná kódující RNA je reverzně

transkribována do podoby cDNA:RNA za použití nespecifických primerů. Na nasyntetizované vlákno cDNA se následně komplementárně začíná vázat druhé syntetizované cDNA vlákno. Při syntéze druhého vlákna je však místo dTTP využito dUTP. Pomocí dUTP se vlákno označí. Toho je využito při následné degradaci tohoto vlákna za pomoci uracil-DNA glykosylázy (UDG). Přípravy specifických knihoven jsou důležité, abychom věděli, že jsme zachovali pouze tu informaci, která nás zajímá, a to sekvenci transkribovaného vlákna [46].

Následující kroky jsou již shodné s přípravou knihoven z DNA a jsou popsány v dalších kapitolách a tabulce č.1.

3.2.2.2. Příprava DNA knihoven

Příprava DNA knihoven vyžaduje jiné metody fragmentace vzorku než je tomu v případě knihoven k sekvenování transkriptomu. Fragmentace je u vzorků DNA prováděna v zásadě třemi metodami: tagmentací (jedná se o fragmentaci pomocí transposomů, které na principu „stříhu a vložení“ nastříhají náhodně DNA a na daný fragment nasyntetizují adaptory [47 - 48]), nebulizací (používá se stlačený dusík nebo vzduch, který svým tlakem mechanicky fragmentuje DNA [49]) a sonifikací (využití ultrazvukových vln k vytvoření plynných kavitačních sil v kapalině, které smykem či zlomem naruší DNA [49]).

Následně je prováděn takzvaný „size selection“, což je proces, při kterém je využito mikrokuliček, například „AMPure Beads“ nebo „SPRI Beads“, které odstraňují nežádoucí malé fragmenty [50]. Velikostní selekce je závislá na koncentraci PEG a solí. Čím vyšší koncentrace těchto složek se v roztoku mikrokuliček a DNA nachází, tím menší fragmenty jsou na kuličkách zachytávány [51].

Následující kroky jsou již shodné s přípravou knihoven z RNA a jsou popsány v následujících kapitolách a tabulce č.1.

Tabulka č.1: Porovnání postupu během přípravy RNA a DNA sekvenačních knihoven

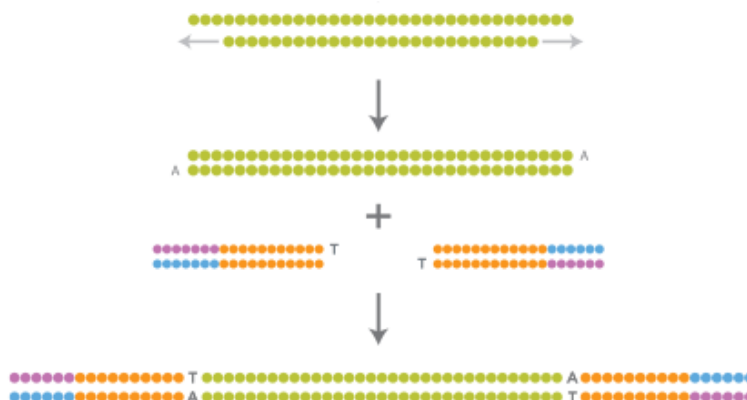
RNA knihovny	DNA knihovny	
Izolace mRNA z totální RNA	-	Rozdílné kroky
Fragmentace mRNA	-	
Přepis mRNA do cDNA	Fragmentace DNA	Shodné kroky
Úprava konců vláken		
Navázání adaptorů		
Přečištění knihovny		

3.2.2.2.3. Další zpracování knihoven

Fragmentace nechává DNA „potrhanou“. Je nutné tedy zarovnat konce fragmentovaných sekvencí. Tento proces se nazývá „end-repairing“ a probíhá v thermocykleru. Do něj jsou přidány vzorky s dNTP a polymerázy spolu s polynukleotid kinásou, která katalyzuje přenos fosfátu z molekuly ATP na 5' konec vlákna. Polymerázy jsou vybírány tak, aby vytvořily tupé konce pomocí svých exonukleázových aktivit a zároveň na 3' konci řetězce vytvořily adeninový přesah. Může být tedy použita T4 DNA polymeráza a Taq DNA polymeráza [52 - 53].

Vytvořeného adeninového přesahu je využito k navázání specifických indexovaných adaptorů. Tyto adaptory jsou dvojího typu: Nextera či TrueSeq. Adaptory jsou v podstatě oligonukleotidy o délce několika desítek bazí. Slouží k navázání vlákna ke komplementárním oligonukleotidům na povrchu amplifikačních destiček.

Postup tvorby sekvenační knihovny od „end-repairingu“ po konečné navázání adaptorů je zobrazen na Obrázku č.10.



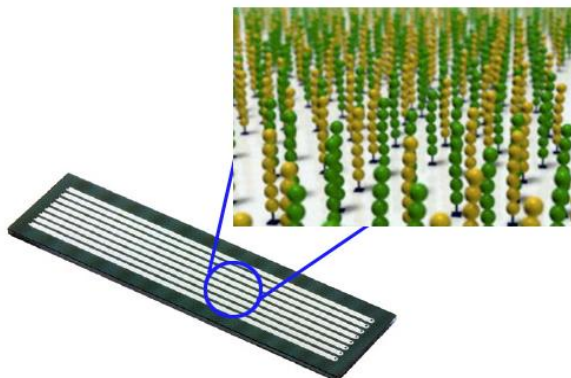
Obrázek č.10: Schéma přípravy DNA knihovny (převzato) [54]

3.2.2.2.4. Vytváření klastrů

Jakmile jsou sekvenační knihovny připraveny dojde k jejich přečištění a následné amplifikaci pomocí takzvané „bridge-PCR/bridge amplification“, česky tedy můstkové amplifikace. Celý tento proces se nazývá vytváření klastrů (z anglického „cluster generation“) [53 - 55].

Amplifikace probíhá na speciálních amplifikačních destičkách („flow cells“). Existují dva typy těchto destiček, a to High Output destičky a Rapid destičky. Na High Output destičkách se nachází osm amplifikačních linií. Na Rapid destičkách jsou tyto linie dvě [56]. Linie amplifikačních destiček jsou pokryty dvěma typy specifických oligonukleotidů [57]. Tyto

oligonukleotidy jsou komplementární k adaptorům navázaných na fragmenty vzorku. Podoba amplifikační destičky je zobrazena na Obrázku č. 11.



Obrázek č.11: Amplifikační destička s vyobrazenými vázanými oligonukleotidy (převzato) [58]

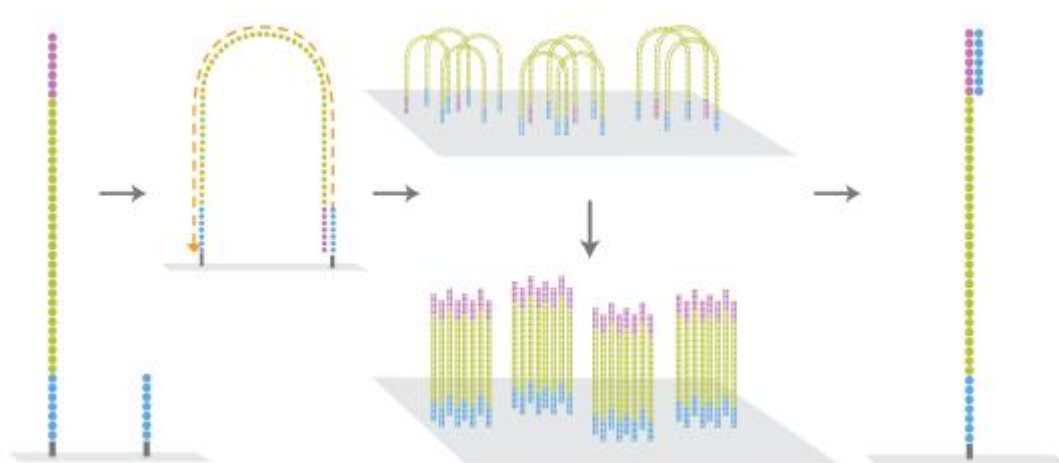
Proces amplifikace probíhá v přístroji cBOT plně automaticky. To však především v případě High OutPut běhu. V případě Rapid běhu můžeme využít takzvaného „on-board cluster generation“. Tento proces probíhá přímo v sekvenátorech HiSeq 2500 a 1500.

Do přístroje jsou vloženy amplifikační destičky s navázanými oligonukleotidy. Jak již výše bylo popsáno, tyto oligonukleotidy jsou komplementární k adaptorům na fragmentech vzorku. Kromě vzorku jsou přidány také dNTP, DNA polymeráza a další reaktanty potřebné k úspěšné amplifikaci fragmentů.

Fragment DNA se naváže adaptorem na komplementární oligonukleotid. Je nasyntetizováno komplementární vlákno k fragmentu DNA a vytvořena dsDNA, která je následně denaturována a původní templát je odmyt. Poté dochází k ohnutí vlákna získané DNA k druhému typu adaptorů. Takto je vytvořen můstek mezi oběma typy oligonukleotidů na amplifikační destičce – odtud pojem můstková amplifikace.

K ohnutému vláknu DNA je dosyntetizováno komplementární vlákno. dsDNA je opět denaturována a celý můstek je renaturován. Tímto jsme získali dvě ssDNA vlákna ve směru 5' – 3' a 3' – 5'. Celý proces je s těmito vlákny mnohonásobně opakován, dokud nedojde k vytvoření miliónů kopií obou vláken.

Po dokončení amplifikace jsou reverzní vlákna odštěpena a odmyta. Zachovány jsou pouze přední vlákna. Tímto krokem je ukončen celý proces vytváření klastrů. Jeho zjednodušená podoba je znázorněna na Obrázku č. 12.



Obrázek č.12: Zjednodušené schéma vytváření klastrů (převzato) [59]

3.2.2.2.5. Sekvence

Jakmile je proces generování klastrů dokončen, je amplifikační destička vložena do sekvenátoru, ve kterém probíhá samotná sekvence.

Sekvenační platformy Illumina používají sekvenační metodu SBS („Sequenced by synthesis“), během které probíhá syntéza komplementárního vlákna templátu. Během této syntézy jsou využívány speciálně značené nukleotidy. Po inkorporaci těchto nukleotidů se uvolní záření, jehož vlnová délka odpovídá určitému nukleotidu.

Sekvenování začíná, jakmile je nasyntetizována komplementární část sekvenačního primeru a je tak započato takzvané „první čtení“ (first read) vlákna. Jsou přidány fluorescenčně značené nukleotidy, které se postupně vážou na vlákno začínající sekvenačním primerem.

Jakmile je nukleotid zabudován do syntetizujícího vlákna, uvolní se charakteristické záření o vlnové délce odpovídající danému nukleotidu. Toto záření je detekováno přístrojem a na jeho základě je určována sekvence prvního čtení vlákna. Přístroj je schopen detekovat záření díky vytvořeným klastrům. Ty v dané oblasti obsahují vždy shodné fragmenty, na které jsou navazovány stejně značené nukleotidy ve stejnou chvíli. To má za následek zesílení záření, které je pak pro přístroj zjistitelné.

Po dokončení prvního čtení sekvence je nasyntetizované vlákno odmyto. Poté je provedeno takzvané „čtení indexu“ (index read), kdy je nasyntetizováno komplementární vlákno k prvnímu

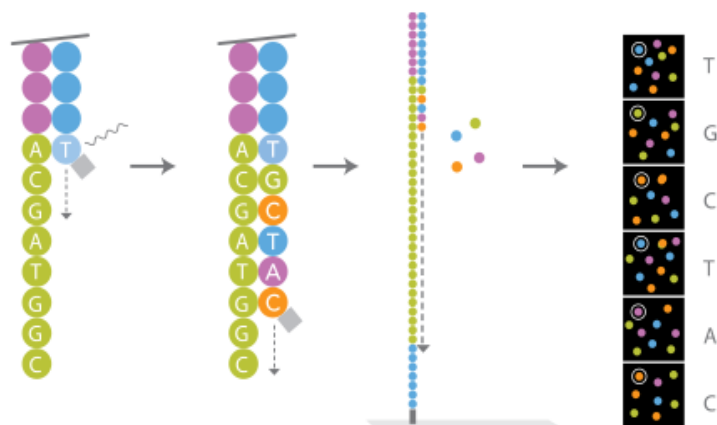
indexu a sekvenačnímu primeru, který je schodný s tím na počátku sekvenování prvního čtení. Nasyntetizovaný produkt je poté opět odmyt [60].

Indexy hrají velmi významnou roli. Umožňují totiž takzvanou multiplexaci, která je nezbytná pro co nejchopodárnější využití sekvenátorů. Díky multiplexaci můžeme sekvenovat najednou několik rozdílných vzorků na jedné amplifikační destičce. Každý vzorek vložený na amplifikační destičku je označen specifickým indexem, který umožňuje jeho následné odlišení od ostatních vzorků. Indexace má však nevýhodu v možné chybné detekci indexu a tedy špatnému přiřazení dané sekvence [60].

Templát je ohnut a navázán k druhému oligonukleotidu na amplifikační destičce. Je nasyntetizováno komplementární vlákno druhého indexu a poté je odmyto. Polymeráza nasyntetizuje komplementární vlákno k celému fragmentu. Během této syntézy je využito běžných neznačených dNTP. Dojde k vytvoření dsDNA můstku. dsDNA je denaturována, čímž se získají dvě vlákna, forward a reverse. Původní forward vlákno je odmyto a zůstává je reverzní vlákno, které je připraveno na takzvané „pair-end“ sekvenování.

To začíná po přečtení a nasyntetizování komplementárního vlákna sekvenačního primeru – „druhé čtení“ (second read). Stejně jako v prvním čtení se do syntetizovaného vlákna vkládají specificky značené nukleotidy.

Díky tomu, že je vlákno sekvenováno z obou stran získáme kompletní sekvenci u dlouhých fragmentů. Získaná čtení jsou poté použita k analýze, která je popsána v další kapitole. Proces sekvenování je zobrazen na Obrázku č. 13.



Obrázek č. 13: Zjednodušené schéma sekvenace (převzato) [61]

Pair-endové sekvenování není jediným typem sekvenace. Můžeme využít takzvaného single-end sekvenování, u kterého na rozdíl od pair-endového sekvenování dochází pouze ke čtení

z jednoho konce sekvence. Dalším typem je takzvané mate-pair sekvenování umožňuje sekvenovat knihovny o velikosti až několik kilobází (kb) [62].

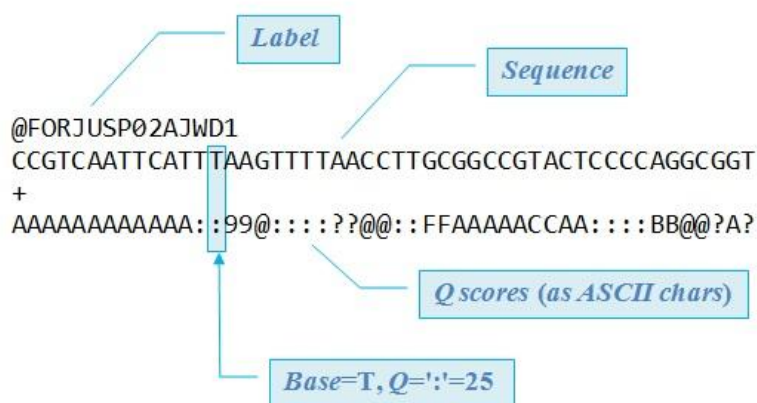
3.2.2.2.6. Analýza dat

Po ukončení sekvenace je nutné přejít k analýze získaných dat. Výstupem, který dostaneme z přístroje je textový soubor obsahující text ve formátu FASTq. Jedná se o modifikaci známého FASTA formátu, tedy souboru písmen značících jednotlivé nukleotidy. Obsahem textového souboru jsou jednotlivé „ready“, kterých může být několik stovek tisíc až miliónů v závislosti na použité platformě.

Na rozdíl od klasického FASTA formátu obsahuje FASTq formát i údaje o kvalitě získaných sekvenčních dat (q – z anglického „quality“). Každý FASTq soubor se skládá ze čtyř řádků. První řádek obsahuje identifikátor sekvence, na druhém je vypsána samotná sekvence. Na třetím řádku je znak „+“, který je identifikátorem následujícího čtvrtého řádku, na kterém je již vypsána kvalita samotného čtení pomocí tzv. phred skóre. Kvalita (Q) se vypočítá dle následujícího vzorce

$$Q = -10 * \log(P_e),$$

kde P_e udává pravděpodobnost, s jakou při volání báze došlo k chybě. Pojem „volání báze“ pochází z anglického „base calling“ [63]. Kvalita správné identifikace báze je zapisována pomocí amerického standardního kódu pro výměnu informací (ASCII znaky, z anglického „American Standard Code for Information Interchange“) [64]. Čím vyšší hodnota ASCII znaku (33-126: !"#\$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcefghij klmnopqrstuvwxyzfjg~), tím s větší přesností byla zkoumaná sekvence přečtena. Podoba FASTq formátu je zobrazena na obrázku č. 14.



Obrázek č.14: Podoba FASTq formátu s vysvětlením jednotlivých řádků (převzato) [65]

Možností dalšího zpracování sekvence ve formátu FASTq je nepřehledné množství a slouží v mnoha odvětvích výzkumu. Manipulace se sekvenčními daty a jejich vyhodnocování však vyžaduje alespoň základní znalost bioinformatiky.

Jedním ze základních bioinformatických nástrojů pro pre-analýzu sekvenčních dat je software FastQC, který byl vyvinut Babrahamovým Institutem a je volně šířen pod GPL licencí jako open source software. Tento nástroj analyzuje jednotlivé zjištěné sekvence a provádí kontrolu jejich kvality. Zkontrolování kvality sekvenčních dat by mělo být provedeno před jakýmkoli dalším zpracováním dat. Software poskytuje základní statistiku počtu readů, délce readů, přítomnosti adaptorů, Phred kvalita bází, GC obsah nebo přítomnost duplikovaných sekvencí [66].

Platformy MiSeq do výstupních dat automaticky nezapisují sekvence adaptorů a sekvenačních primerů, což ale například platformy HiSeq neumí a je tedy nutné jejich sekvence ručně odfiltrovat. K tomuto účelu se používá knihovna nástrojů zvaná FASTX-Toolkit, která umožňuje ořezat ready od nechtěných sekvencí. Dále také umí filtrovat nekvalitní data, filtrovat data atd [67].

3.2.2.2.6.1. Analýza sekvenčních dat po transkriptomovém sekvenování

Jak již bylo výše zmíněno, získaná data jsou uložena do souboru s formátem FASTq. Fragmenty uložené v tomto souboru je nutné seskládat tak, aby byla získána sekvence původního genomu. Využívá se takzvaného skládání neboli kompletace (anglicky „assembling“). Ten může být dvojího typu: referenční nebo *de novo*. Při referenčním assemblingu je využíváno známé referenční sekvence (například v současné době je využívána při sekvenování lidského genomu sekvence HG38). K této referenční sekvenci jsou fragmenty přikládány a na základě porovnání je skládána výsledná sekvence, která může být buď plně shodná či podobná té referenční. Při *de novo* assemblingu jsou jednotlivé fragmenty skládány k sobě bez jakékoli referenční sekvence.

V případě *de novo* sekvenování neobsahují knihovny pouze fragmenty přesně navazující na sebe, ale především fragmenty, které se z ať už větší či menší části překrývají. Přístupuje se tedy k takzvanému sekvenčnímu přiložení (anglicky „alignment“). Při něm jsou jednotlivé sekvence přikládány k sobě tak, aby byly odhaleny shodné sekvence (překryvy), pomocí kterých se zjistí, zda-li fragmenty náleží vedle sebe. Takto jsou jednotlivé fragmenty sestavovány do větších celků, kterými jsou kontigy. Kontigy jsou dále spojovány do takzvaných scaffolds.

Při analýze výstupu ze sekvenování transkriptomu se využívá například volně dostupného softwaru s názvem TopHat, případně jeho novější verze TopHat2, který má oproti původní verzi několik vylepšení. Software mapuje jednotlivé ready - hledá spoje mezi jednotlivými přečtenými

sekvencemi a spojuje je například na základě porovnání s referenční sekvencí. [68]. TopHat2 však využívá ještě před samotným mapováním, takzvaného „předmapování“. K tomu je používán software Bowtie2. Tento program vytváří indexovanou genomickou sekvenci využitím Burrowsovy Wheelerovy transformace (BWT) [69-70]. Tato transformace převádí data do podoby mnohem vhodnější ke kompresi. Předmapované ready jsou posléze využity programem TopHat, který sekvence následně zmapuje a také identifikuje mezery mezi exony (anglicky splice junction). Výstupem je soubor ve formátu BAM, který obsahuje namapované ready [70].

Další analýza je prováděna pomocí softwaru HTSeq [71], což je python knihovna využívající pro svou práci data spojená s genomovými souřadnicemi. Obsahuje analyzátoři pro referenční sekvence (FASTA), krátká čtení (FASTQ), seřazená krátká čtení (SAM/BAM formát) a analyzátoři pro genomické funkce, anotace a skórování dat [72].

3.2.2.2.7. Aktuální nabídka přístrojů společnosti Illumina

První platformou, kterou tato společnost uvedla na trh, byl sekvenátor Illumina Genome Analyzer, který byl vyvinut společností Solexa. Illumina tuto společnost koupila na začátku roku 2007 [73]. Dokázal přečíst sekvence dlouhé 36 bází. Běh přístroje trval několik dní a bylo během něj vyprodukováno 1 Gb dat určených ke zpracování [74 - 77].

Vývoj v této oblasti však již velmi postoupil a v dnešní době přístroje společnosti Illumina poskytují portfolio, které obsahuje nástroje umožňující sekvenovat, jak malé části genomů, amplikonů, exonů, provádět celogenomová i transkriptomová sekvenování.

Aktuální nabídka společnosti Illumina zahrnuje následující sekvenční platformy: MiSeq, NextSeq 500, HiSeq 2500, HiSeq X. Specifikace sekvenátorů jsou uvedeny v Tabulce č.2.

Tabulka č.2: Sekvenační systémy společnosti Illumina. (převzato a upraveno) [78]

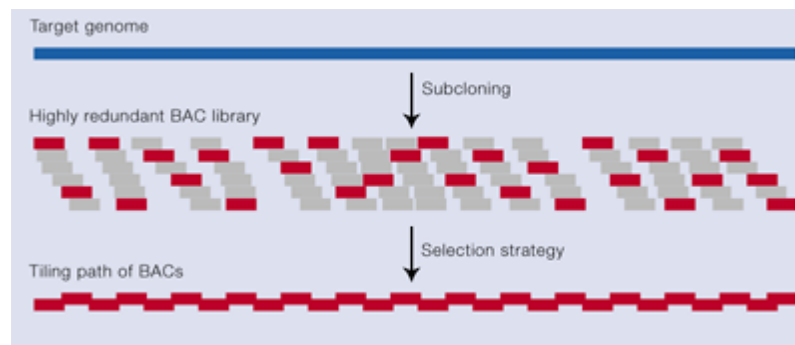
Systém:	MiSeq	NextSeq 500		HiSeq 2500		HiSeq X
Vhodný pro	Sekvenování malých genomů, amplikonů a jiné.	Mezistupeň mezi systémy MiSeq a HiSeq		Např. exomové či transkriptomové sekvenování		Celogenomové sekvenování
Mód běhu	-	Střední výkon	Vysoký výkon	Rychlý běh	Vysoký výkon	-
Počet využitých amplifikačních linií během běhu	1	1	1	1-2	1-2	1-2
Objem výstupních dat	0,3-15 Gb	20 – 39 Gb	30-120 Gb	10-180 Gb	50-1000 Gb	1,6-1,8 Tb
Doba běhu	5-55 h	15-26 h	12-30h	7-40 h	1-6 d	do 3 d

3.2.2.3. Využití sekvenačních metod nové generace

Sekvenační metody nové generace nachází díky svým výhodám, především krátké době trvání samotné sekvenace a nízké ceně za přečtenou bázi, uplatnění v celé řadě aplikacích. Od *de novo* genomového sekvenování, celogenomového či cíleného resekvenování, analýzy transkriptomu, analýzy DNA metylovaných oblastí, mapování DNA-proteinových interakcí chromatinovou precipitací. Umožňují charakterizovat molekulární podstatu dědičných chorob a také infekčních onemocněních. Také nachází uplatnění v prenatální diagnostice, molekulární diagnostice nádorů či ve farmakogenomice [79].

3.2.2.3.1. Celogenomové sekvenování

Během projektu „Human Genome“, který si dával za cíl zmapovat kompletní lidský genom, bylo využito v prvotní fázi přípravy BAC sekvenčních knihoven. DNA byla fragmentována na délku okolo 150-200 kb. Tyto fragmenty byly klonovány do bakterií. Tím byla vytvořena BAC knihovna s 20 tisíci různými BAC klony nesoucí fragmenty DNA. Pro sekvenování byly tyto BAC klony dále fragmentovány na fragmenty o délce 2 kb – takzvané „subklony“. Bylo tím dosaženo usnadnění sekvenace daných částí lidského genomu. Na obrázku č.15 je vidět jakým způsobem probíhalo následné skládání sekvence subklonů [80].



Obrázek č.15: Vyobrazení skládání zkoumané sekvence z jednotlivých subklonů (převzato) [81]

Náklady na osekvenování lidského genomu pomocí této metody byly však astronomické. Posun v sekvenčních technologiích, který snížil náklady na přípravu a samotné sekvenování však umožnil vznik projektu „1000 Genome“, který na základě mezinárodní spolupráce si dává za cíl osekvenovat co největší počet lidí, aby našel a poskytl komplexní zdroj genetických variant a to i takových, které mají v populaci frekvenci menší než 1 % [82].

Celogenomové sekvenování můžeme rozdělit na dva druhy a to *de novo* sekvenování, během kterého je získávána kompletní sekvence genomu ještě nesekvenovaného organismu. Druhým druhem celogenomového sekvenování je takzvané resekvenování, které se od *de novo* sekvenování liší tím, že je k dispozici referenční sekvence, podle které se nová sekvence mapuje. Tento proces se nazývá „read mapping“ [83].

Tento typ sekvenování se zaměřuje na celou chromozomální DNA a pokrývá tedy i promotorové či regulační sekvence. Tím poskytuje kompletní informace o úplném DNA profilu daného genomu. Získáme informace o jednonukleotidových polymorfismech (z anglického „single-nucleotide polymorphism“), což jsou jednobodové záměny nukleotidů v různých oblastech genomu [84]. Dále také dochází k detekci inzercí a delecí, chromozomálních změn a strukturních variant [79].

Poprvé bylo celogenomového sekvenování využito roku 2008 u pacienta s akutní myeloidní leukémií (AML) [85].

3.2.2.3.2. Sekvenování transkriptomu

Pod takzvaný transkriptom spadá veškerá RNA v buňce (mRNA, tRNA, rRNA, miRNA a další nekódující RNA molekuly). Sekvenací a analýzou transkriptomu můžeme analyzovat míru genové exprese, dále také získáváme informace o alternativních sestřihových variantách, somatických mutacích a mezigenových fúzí [86].

Jednou z možností uplatnění sekvenování transkriptomu je například ve výzkumu Diamond-Blackfanovy anémie (DBA). DBA je vrozená erythroidní hypoplazie způsobená funkční haploinsuficiencí genů kódující ribozomální proteiny. V téměř 25 % zahrnuje ribozomální protein S19 [87]. Sekvenování transkriptomu může například sloužit v tomto případě k analýze transkripčních profilů u RPS19-deficientních embryí modelových organismů a vymezení funkce p53 u těchto embryí [88].

Další využití nalézá sekvenování transkriptomu například ve srovnávací analýze genové exprese lidí s normální mozkovou tkání a pacientů postižených Alzheimerovou chorobou. Analýzou mRNA z dvou částí mozku (čelní a spánkový lalok) byla zjištěna různá genová exprese izoform genů, alternativní využití promotorů atd. [89]. Výzkum v oblasti neurodegenerativních onemocnění všeobecně velmi často využívá transkriptomového sekvenování, jelikož analýza genové exprese pomáhá analyzovat mechanismy zapojené do molekulární patogeneze neurodegenerativních onemocnění a výzkum v této oblasti může vést k novým terapeutickým léčivům. Kromě Alzheimerovy choroby se výzkum týká například Parkinsonova onemocnění či

Huntingtonovy choroby a mnoha dalších [90].

Sekvenování RNA přináší mnoho výhod oproti klasicky používaným DNA microarray technologiím. Při kvantifikaci exprese je u RNA sekvenování postrádán její horní limit a záleží tedy pouze na množství zmapovaných readů. Microarray technologie oproti tomu nejsou citlivé na velmi malé či velmi vysoké úrovně exprese. RNA sekvenování se využívá také při *de novo* sekvenování, jelikož na rozdíl od čipových technologií není nutná znalost referenčního genomu [91-92].

3.2.2.3.3. Cílené sekvenování vybraných oblastí genomu

Jedná se o metody, pomocí které je sekvenována pouze definovaná oblast v genomu. Tímto je ušetřen čas a výsledná sekvenační data nevyžadují tak velký skladovací prostor. Náklady jsou také v případě cíleného sekvenování mnohem nižší. Využití nachází ve screeningu u velkého počtu pacientů či při zkoumání genetických variant v populaci [93].

Jak bylo výše zmíněno, cílené sekvenování se velice často používá při diagnostice v medicíně. V poslední době se jej velmi využívá při detekci nejrůznějších biomarkerů. Onkologické biomarkery jsou užitečné při posuzování stádia rakoviny, diagnóze pacienta s rakovinou. Pomocí nich lze utvářet prognózy postupu onemocnění, určovat šanci na přežití daného pacienta, či predikovat odpověď na jednotlivá protinádorová léčiva [94-95].

Vybranými příklady z nepřeberného množství biomarkerů mohou být geny proteinové rodiny RAS (NRAS, KRAS, HRAS, ...), nebo také genů pro receptor epidermálního růstového faktoru (EGFR) [96-97].

EGFR, který je také známý jako HER1 nebo c-ErbB1 se v běžné tkáni vyskytuje ve velmi nízkých koncentracích. Jeho exprese je důležitá při regulaci buněčné proliferace a přežití buňky. Za zvýšením genové exprese EGFR může být amplifikace kopií genu, naprodukce EGFR ligandů nebo také aktivace genu mutací. Všechny tyto procesy poté vedou k maligní transformaci buňky [98].

Důvodem, proč je EGFR velmi často cílem cíleného sekvenování je, že je exprimován u velkého množství lidských nádorů, jako je karcinom prostaty (40-80 %), rakovina vaječníků (35-70 %), kolorektální karcinom (80-100 %), a u mnoha dalších [99].

Pokud se zaměříme na biomarkery RAS (tyto proteiny navazují na signální dráhu EGFR), přesněji například na typ KRAS zjistíme, že se nachází v normálních buňkách a jedná se intracelulární signální transduktor, který odvádí signály z buněčného povrchu do jádra přes

cytoplazmu. Jeho mutovaná forma byla nalezena u mnoha typů rakoviny. Například u karcinomu pankreatu (okolo 90 %), karcinomu tlustého střeva (okolo 40 %), či rakoviny plic (15-30 %). Mutaci genu KRAS má za příčinu záměna jednoho nukleotidu (nejčastěji v kodonu 12 a 13 exonu – až 95 % všech mutací) [100-103].

Biomarkery nejen v případě EGFR a RAS slouží kromě diagnostiky rakoviny také k tomu, aby bylo možné zjistit, zda-li aplikovaná léčba pacientovi pomůže, či jestli mu spíše nemůže uškodit. Například v případě medikace pacientů s indikací biomarkeru KRAS, která se provádí cetuximabem, je léčba účinná pouze u wild-type typu KRAS genu, kdežto v případě mutantního typu genu KRAS může tato léčba pacientovi uškodit, což může mít za následek zkrácení doby dožití pacienta [104-106].

3.2.2.3.4. Exomové sekvenování

Při exomovém sekvenování jsou sekvenovány pouze proteiny kódující oblasti DNA – exomy, které představují přibližně 1-2 % velikostí celého genomu, jedná se tedy o formu cíleného sekvenování. Vzhledem k tomu, že se zaměřuje jen na určitou část genetické informace, je levnější a dosahuje vyššího pokrytí (coverage) [107-108]. Při exomovém sekvenování je nutné zachytit exony z celkové DNA. To se dělá pomocí hybridizací, které jsou dvojího typu. Na „poli“ založené hybridizaci [109] a nebo kapalinové hybridizaci [110].

Sekvenování exonu se stalo jedním z hlavních nástrojů pro studium nejen genetické příčiny mendelovsky dědičných onemocnění, ale také způsobu dědičnosti tohoto onemocnění, jeho závažnost a frekvenci v obyvatelstvu. Můžeme detekovat patogenní mutace mezi tisíci až miliony genomickými variantami [111-113]. Další uplatnění exomového sekvenování je v oblasti evoluční biologie, kdy bývají alignmentovány sekvence a na základě výsledku může být například zjištěn původ určité exprimující části genomu [114-115].

4. Materiál a metody

4.1. Materiál

4.1.1. Biologický materiál

Jako biologický materiál byla v praktické části použita buněčná linie U2OS z Ústavu molekulární a translační medicíny v Olomouci.

4.1.2. Kity

Nextera XT index kit 96 indices – 192 samples	Illumina
Pico RNA Analysis Kit	Agilent Technologies
High Sensitivity DNA Analysis Kit	Agilent Technologies
Dynabeads [®] mRNA Purification Kit	Life Technologies
MinElute [®] PCR purification Kit 250	Qiagen

4.1.3. Použité chemikálie, roztoky a jejich složení

5X reaction buffer for RT	Thermo scientific
Revert AID H minus reverse transcriptase	Thermo scientific
DTT	Qiagen
Random Primers	Promega
dNTP	Bioline
Actinomycin D	Orphan Europe
NEBuffer 2	New England BioLabs
dUTP mix (50mM)	Bioline
RNase H	New England BioLabs
DNA polymerase I. (<i>E. Coli</i>)	New England BioLabs
Thermo-Start Taq DNA polymerase	Thermo scientific
T4 DNA polymerase	New England BioLabs
T4 Polynukleotide Kinase	New England BioLabs
T4 DNA ligase	New England BioLabs
10x NEB T4 Ligase Buffer (with 10 mM ATP)	New England BioLabs
Thermo Start PCR Buffer	Thermo scientific
Magnesium chloride solution	Thermo scientific
DMSO	Roche
SPRI Select	Beckman Coulter
TRIzol	Life Technologies

Chloroform
Isopropanol
75% etanol
Eva Green

Sigma-Aldrich
Sigma-Aldrich
Sigma-Aldrich
Biotium

4.1.4. Přístrojové vybavení

Minispin
VortexMixer
Thermostat Plus
Centrifuge 5430
C1000 Thermo Cycler
TopSafe 1.2
Agilent 2100 Bioanalyzer
Ika MS3 Vortexer
Qubit 2.0 Fluorometer
Chip Priming Station
Nanodrop Spectrophotometer ND 1000
Hermle 2323K
LightCycler[®] 480

Eppendorf
Labnet
Eppendorf
Eppendorf
Biorad
Euroclone
Agilent Technologies
IKA
Invitrogen
Agilent Technologies
Thermo Scientific
Hermle LaborTechnik
Roche

4.2. Metody

4.2.1. Izolace celkové RNA z lidských buněk

Pro izolaci celkové RNA z 100 μ l buněčné suspenze buněčné linie U2OS jsme k ní přidali 1 ml Trizol reagentu a směs intenzivně zvortexovali. Ke směsi jsme přidali 200 μ l chloroformu a opět zvortexovali. Inkubovali jsme 10 minut za pokojové teploty. Následně jsme mikroskopickou skleničku se směsí centrifugovali v předchlazené centrifuzě (Hermle 2323K) na 4 °C, při 12 000 rpm po dobu 15 minut.

Stočením došlo k separaci fází, z nichž vrchní průhledná fáze obsahovala RNA. Odebrali jsme 500 μ l této průhledné fáze a přenesli ji do nové mikroskopickou skleničky. Dále jsme do mikroskopickou skleničky přidali 500 μ l isopropanolu a obsah jsme promýchali několikanásobným otočením mikroskopickou skleničky. Následovala inkubace po dobu 5 minut za pokojové teploty. V dalším kroku jsme centrifugovali mikroskopickou skleničku za stejných podmínek jako v předchozím kroku, tentokrát po dobu 10 minut.

Po centrifugaci jsme pozorovali bílou peletu na stěně mikroskopickou skleničky. Odstranili jsme supernatant a k peletě přidali 1.5 ml 75% etanolu. Mikroskopickou skleničku jsme poté opět centrifugovali za stejných podmínek po dobu 5 minut. Opět jsme z mikroskopickou skleničky odstranili supernatant a peletu sušili po dobu 6 minut.

Poté jsme přidali k peletě 25 μ l RNase-free vody a směs několikrát propipetovali. Pro lepší rozpuštění pelety jsme směs inkubovali 10 minut při 60 °C a následně chladili 10 vteřin v ledové tříšti.

Po izolaci jsme změřili koncentraci, míru integrity a strukturu vzorku na přístrojích Nanodrop Spectrophotometer ND 1000 (Thermo Scientific) a Agilent 2100 Bioanalyzer (Agilent technologies), u kterého jsme použili RNA Pico Chip a test mRNA Pico.

4.2.2. Izolace mRNA z celkové RNA

Nejprve jsme celkovou RNA ředili s RNase-free vodou v mikrozkuvkách tak, aby celkový objem odpovídal objemu kalibrovaných mikrokuliček Dynabeads®. Rozpis objemů reagiencí použitých při izolaci mRNA je uveden v tabulce č.3. Mikrozkuvky s naředěnou celkovou RNA jsme inkubovali 2 minuty při 65 °C a minutu chladili na ledu.

Mezitím jsme nakalibrovali mikrokuličky Dynabeads®. Požadovaný objem mikrokuliček jsme napipetovali do mikrozkuvek a vložili je na magnet. Po přichycení mikrokuliček jsme odstranili supernatant a mikrozkuvky vyjmuli z magnetu. K mikrokuličkám jsme přidali poloviční množství Binding Buffer oproti původnímu objemu mikrokuliček a kuličky v něm resuspendovali. Opět jsme mikrozkuvky umístili na magnet a supernatant, po usednutí mikrokuliček na stěnu mikrozkuvky, odsáli. Poté jsme přidali stejné množství Binding Buffer, jako v předchozím kroku a mikrokuličky v něm opět resuspendovali.

Obsah mikrozkuvek jsme smíchali s kalibrovanými mikrokuličkami Dynabeads® a inkubovali za mírného třepání 5 minut při pokojové teplotě. Mikrozkuvky jsme následně umístili na magnet a po usazení všech mikrokuliček odsáli supernatant. Mikrokuličky jsme dvakrát promyli Washing Buffer B v množství, které odpovídalo objemu kalibrovaným mikrokuliček. Mezi každým promytím jsme mikrozkuvku vyjmuli z magnetu a mikrokuličky v Washing Buffer B resuspendovali.

Po odstranění promývacího pufru jsme mikrokuličky resuspendovali v 10 µl RNase-free vody a 2 minuty inkubovali při 75 °C. Následně jsme přenesli mikrozkuvky na magnet a po usazení mikrokuliček odsáli supernatant, který obsahoval mRNA, a přenesli jej do čistých mikrozkuvek.

Vzorky jsme poté analyzovali na přístroji Agilent 2100 Bioanalyzer, který analyzoval kontaminaci vzorku zbytkovou rRNA, koncentraci mRNA a její strukturu. Opět jsme použili RNA Pico Chip a test mRNA Pico. Ze získaných hodnot jsme vypočetli skutečné množství vyizolované mRNA a také podíl této mRNA k množství celkové RNA použité k izolaci

Výpočet podílu mRNA na množství celkové RNA vstupující do izolace jsme vypočetli dle následující rovnice:

$$\text{podíl mRNA} = \frac{\text{získané množství mRNA}}{\text{množství celkové RNA použité k izolaci}} * 100$$

Tabulka č.3: Rozpis množství a objemů použitých při izolaci mRNA.

Vzorek	Množství celkové RNA [μg]	Objem vzorku po ředění [μl]	Množství mikrokuliček před/po kalibraci [μl]
A	2	100	200/100
B	6	100	200/100
C	15	100	200/100
D	30	100	200/100
E	2	50	100/50
F	6	50	100/50
G	15	50	100/50
H	30	50	100/50
I	2	25	50/25
J	6	25	50/25
K	15	25	50/25
L	30	25	50/25
M	2	12,5	25/12,5
N	6	12,5	25/12,5
O	15	12,5	25/12,5
P	30	12,5	25/12,5

4.2.3. Fragmentace mRNA

Reakční směs k fragmentaci mRNA jsme připravili smícháním 4 μl mRNA, 4 μl 5x Reverse Transcriptase Reaction Buffer (Thermo scientific) a 0.5 μl 100 mM dithiotreitol (DTT, Qiagen). Mikrozkumavky s reakční směsí jsme inkubovali 2, 4 a 8 minut při 94 °C. Pro kontrolu jsme nechali jeden vzorek nefragmentovaný.

Vzorky fragmentované mRNA jsme opět analyzovali na přístroji Agilent 2100 Bioanalyzer, kterým jsme zjistili velikost fragmentů a koncentraci vzorku. Použili jsme stejný chip a test, jako v předchozích dvou krocích.

4.2.4. Reverzní transkripce mRNA

Reverzní transkripce byla prováděna podle protokolu „Strand-specific libraries“ [46] (protokol č.1a) a zároveň podle modifikace tohoto protokolu (protokol č.2a).

Reverzní transkripci jsme začali nasedáním primerů. K fragmentové mRNA jsme dle protokolu č.1a přidali 1 μl Random primers (3 $\mu\text{g}/\mu\text{l}$, Promega), 7 μl DEPC vody a 1,5 μl 100mM DTT. V případě protokolu č.2a jsme k fragmentované mRNA přidali 2 μl Random primers (0,5 $\mu\text{g}/\mu\text{l}$, Promega), 3 μl DEPC vody a 1,5 μl DTT. Mikrozkumavky se směsí jsme inkubovali 3 minuty při teplotě 65 °C.

Po inkubaci jsme k vzorku z protokolu č.1a přidali 1 μl 10mM dNTP (Bioline), 0,5 μl 100mM DTT, 1 μl Reverse Transcriptase H(-) (Thermo Scientific) a 8 μl Actinomycin D 0,5 $\mu\text{g}/\mu\text{l}$ (Orphan Europe). U vzorku z protokolu č.2a byl rozdíl jen v přidaném množství Actinomycinu D, kterého bylo přidáno 3 μl . Mikrozkumavky s reakční směsí jsme inkubovali 10 minut při 25°C, následně 60 minut při 42 °C a celá reakce byla ukončena 10 minutovou inkubací při teplotě 70 °C.

Vzorek z protokolu č.1a obsahující komplex cDNA:mRNA jsme poté přečistili pomocí 32 μl SPRISelect magnetických mikrokuliček (Beckman Coulter). Přečištěný komplex cDNA:mRNA jsme eluovali do 22 μl DEPC vody a přenesli jej do nové mikrozkumavky. V případě vzorku připravovaném podle protokolu č.2a jsme komplex čistili s 45 μl SPRISelect a eluovali jej do 13 μl DEPC vody.

Následující část byla shodná pro oba dva protokoly. K vzorku v mikrozkumavce jsme dále přidali 3 μl NEB 2 Buffer (New England BioLabs), 2 μl dUTP mix (20 mM dUTP, 10 mM dATP, dCTP, DTP; Bioline), abychom zajistili specifičnost druhého vlákna cDNA, dále pak 1 μl RNase H (2 U/ μl ; New England BioLabs), 2 μl DNA Polymerase I (*E.Coli*) (10 U/ μl ; New England

BioLabs) a 0,5 μ l 100 mM DTT. Celou směs jsme poté inkubovali 2,5 hodiny při teplotě 16 °C.

Po inkubaci jsme směs přečistili pomocí mikrokuliček SPRISelect. Pro vzorek z protokolu č.1a jsme použili 45 μ l mikrokuliček a poté jej eluovali do 33 μ l. Vzorek zpracováváný podle protokolu č.2a jsme přečistili 45 μ l SPRISelect mikrokuliček a přečištěnou dvouvláknovou cDNA eluovali do 15 μ l RNase-free vody. cDNA jsme poté analyzovali přístrojem Qubit 2.0 Fluorometer (Invitrogen), kterým jsme zjišťovali koncentraci vzorku a přístrojem Agilent 2100 Bioanalyzer, kterým jsme získali data o velikosti fragmentů cDNA. V tomto kroku jsme použili High Sensitivity DNA Chip a test High Sensitivity DNA.

4.2.5. Příprava sekvenační knihovny z cDNA

K přípravě knihovny jsme použili dva alternativní protokoly. Protokol č.1b byl původně optimalizován k přípravě knihoven z amplikonů RAS, zatímco protokol č.2b se v laboratoři ÚMTM používá k přípravě knihoven pro celogenomové sekvenování.

4.2.5.1. Modifikace konců

K modifikaci konců jsme u protokolu č.1b použili 5 μ l dvouvláknové cDNA, ke které jsme přidali reakční směs, složenou z 1,2 μ l 10x NEB T4 Ligase Buffer (with 10 mM ATP) (New England BioLabs), 1,2 μ l ATP (10 mM; New England BioLabs), 1 μ l T4 polynucleotide kinase (3 U/ μ l; New England BioLabs), 0,4 μ l Taq Polymerase (5 U/ μ l; Thermo Scientific) a 3,6 μ l DEPC vody. Směs byla inkubována 30 minut při 37 °C a poté 20 minut při 72 °C.

V případě protokolu č.2b jsme k 10 μ l dvouvláknové cDNA přidali 2 μ l 10x NEB T4 Ligase Buffer (with 10 mM ATP), 0,5 μ l T4 polynucleotide kinase 3 U/ μ l, 0,4 μ l Taq Polymerase (5 U/ μ l; Thermo Scientific), 0,4 μ l T4 DNA polymerase (New England BioLabs) a 7,5 μ l DEPC vody. Takto připravenou směs jsme inkubovali 20 minut při teplotě 25 °C a následně 20 minut za teploty 72 °C.

4.2.5.2. Ligace adaptorů

Adaptory pro ligaci jsme připravili stejným postupem pro oba dva protokoly. Smíchali jsme 10 μ l forward a 10 μ l reverse adaptoru. Forward i5 s reverse i5 adaptorem a stejně tak v případě i7 adaptorů. Následně jsme směsi inkubovali 10 minut při teplotě 95 °C a poté snižovali inkubační teplotu o 0,5 °C každou minutu po dobu 40 minut. Po inkubaci jsme adaptory naředili do konečné koncentrace 10 μ M a smíchali v poměru 1:1.

Protokol č.1b pokračoval přidáním 0,5 μ l směsi adaptorů a 0,5 μ l T4 DNA Ligase (New England BioLabs). Směs jsme poté inkubovali 20 minut při teplotě 25 °C, následně 10 minut při

16 °C a inkubace byla ukončena teplotou 65 °C po dobu 10 minut.

V případě protokolu č.2b jsme ke vzorku z předchozího kroku přidali 1 µl směsi adaptorů a 1 µl T4 DNA Ligase (New England BioLabs). Takto připravenou směs jsme inkubovali 20 minut při 25 °C.

4.2.5.3. Přečištění

U obou protokolů jsme přistoupili k přečištění vzorku pomocí mikrokuliček SPRISelect. Vzorky jsme přečistili dvojnásobným objemem mikrokuliček a po přečištění je v obou případech eluovali do 10 µl DEPC vody.

4.2.5.4. Degradace uracilem značeného vlákna

Abychom degradovali z dvouvláknové cDNA vlákno, které jsme značili uracilem, přidali jsme k přečištěnému vzorku 2 µl Uracil-DNA Glykosylase (UDG, 5 U/µl; Bionline). Směs jsme inkubovali 30 minut při teplotě 37 °C.

4.2.5.5. Amplifikace knihovny

K amplifikaci knihovny jsme v případě protokolu č.1b odebrali ze vzorku 5 µl a v případě protokolu č.2b 8 µl, ke kterému jsme přidali reakční směs a následně doplnili DEPC vodou na celkový objem 20 µl. Složení reakční směsi je uvedeno v tabulce č. 4.

Amplifikaci jsme provedli na přístroji LightCycler[®] 480 (Roche). Program, který jsme použili je rozepsán v tabulce č.5. Po amplifikaci jsme vzorky přečistili na kolonách z MinElute[®] PCR purification Kit 250 (Qiagen). Přečištěné vzorky jsme eluovali do 10 µl DEPC vody.

Tabulka č.4: Reakční směs pro amplifikaci cDNA knihovny

Reagencie	Objem [µl]
10x Thermo Start PCR Buffer (Thermo Scientific)	2
MgCl ₂ 25 mM(Thermo Scientific)	1,6
Index N712 (Illumina)	1
Index S501 (Illumina)	1
dNTP 10mM (Bionline)	0,4
Eva Green 20X (Biotium)	1
Thermo Start Taq DNA Polymerase 5U/µl (Thermo Scientific)	0,4

Tabulka č.5: Podmínky amplifikace knihovny

Teplota [°C]	Čas	Opakování
95	5 min	-
95	10 sek	30x
63	30 sek	
72	1 min	
4	-	-

4.2.6. Ověření knihovny

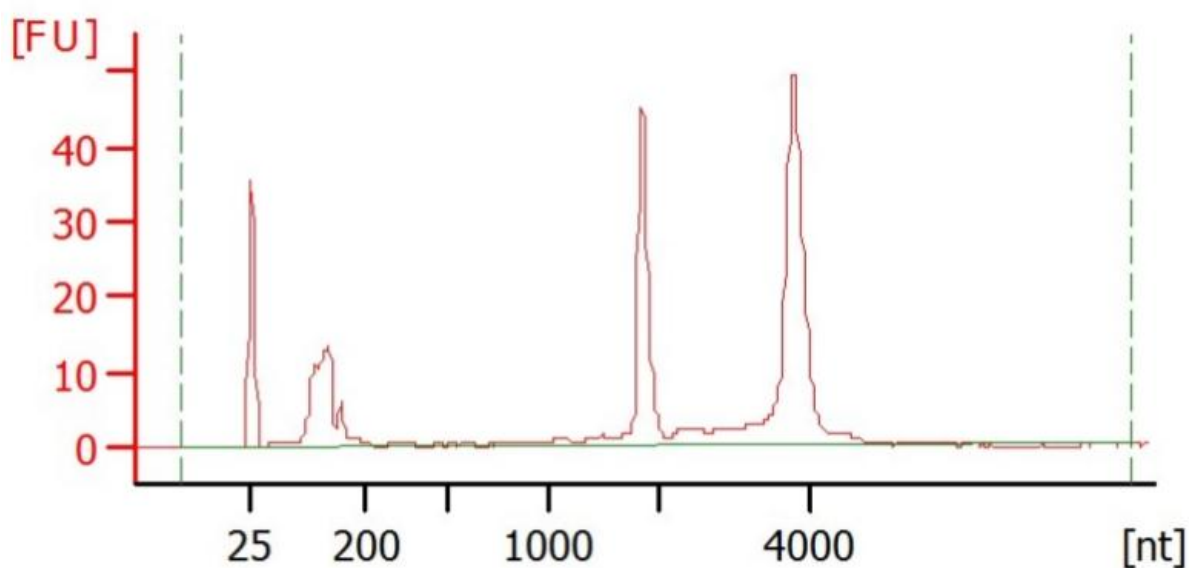
Abychom lépe analyzovali průběh přípravy knihovny, otestovali jsme vzorky z jednotlivých kroků na přístroji Agilent 2100 Bioanalyzer. K této analýze jsme použili High Sensitivity DNA Chip a test High Sensitivity DNA.

5. Výsledky

5.1. Izolace celkové RNA

Tabulka č.6 obsahuje hodnoty koncentrací a čistoty u jednotlivých izolovaných vzorků celkové RNA.

Na obrázku č.16 je vyobrazena podoba elektroforetogramu vyizolované celkové RNA vzorku H001. Elektroforetogramy zbývajících vzorků celkové RNA vykazovali totožnou křivku s křivkou v elektroforetogramu vzorku H001.



Obrázek č.16: Podoba elektroforetogramu celkové RNA vzorku H001

Tabulka č.6: Naměřené hodnoty koncentrací a čistoty vzorků izolované celkové RNA

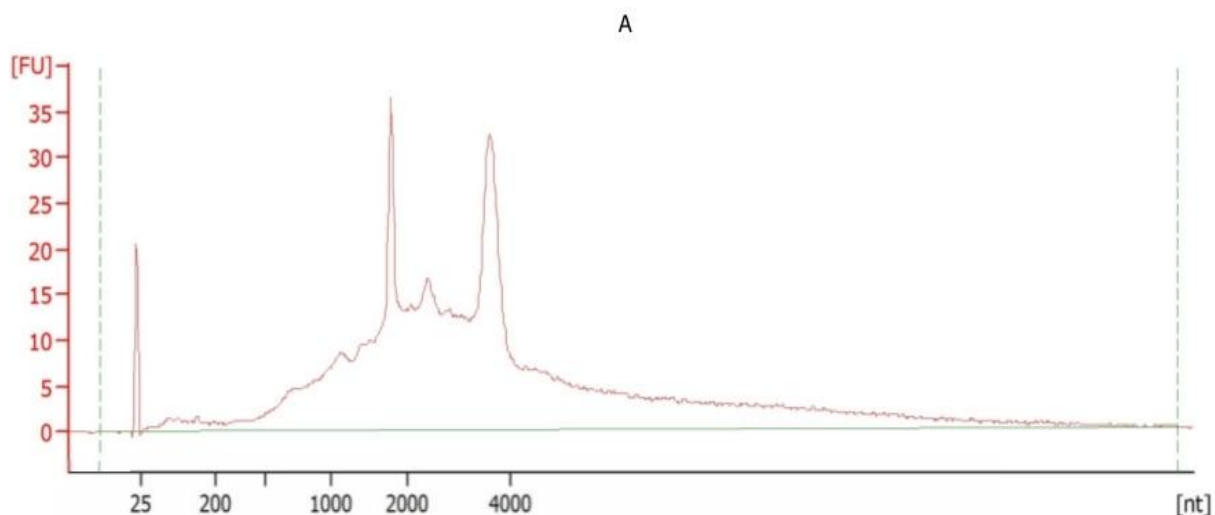
Vzorek	ng/ μ g	260/280	260/230
H001	1916	1,96	2,01
H002	2001	1,95	2,01
H003	2361	1,93	2,02
H004	2174	1,92	2,20
H005	1798	1,92	2,19
H006	2464	1,90	2,07

5.2. Izolace mRNA

Na obrázku č.17 je vyobrazena podoba elektroforetogramu vyizolované mRNA. Celkem takto bylo analyzováno 48 vzorků vyizolované mRNA. Struktura křivky elektroforetogramu zbývajících 47 vzorků je téměř totožná se strukturou křivky u vzorku A.

Každý řádek v Tabulce č.7 je průměrem 3 oddělených izolací za stejných podmínek a obsahuje údaje o průměrném množství vyizolovaného materiálu a průměrné naměřené hodnotě kontaminace zbytkovou rRNA. Další data, která tabulka obsahuje, jsou průměrné hodnoty, které byly vypočteny ze zjištěných údajů o množství vyizolovaného materiálu a míře kontaminace. Tyto hodnoty reprezentují skutečné množství vyizolované mRNA a podíl této mRNA k množství celkové RNA použité k izolaci.

Graf č.1 zobrazuje vliv vstupního množství celkové RNA a objemu použitých mikrokuliček Dynabeads[®] na výsledný podíl vyizolované mRNA k celkové RNA vstupující do reakce.

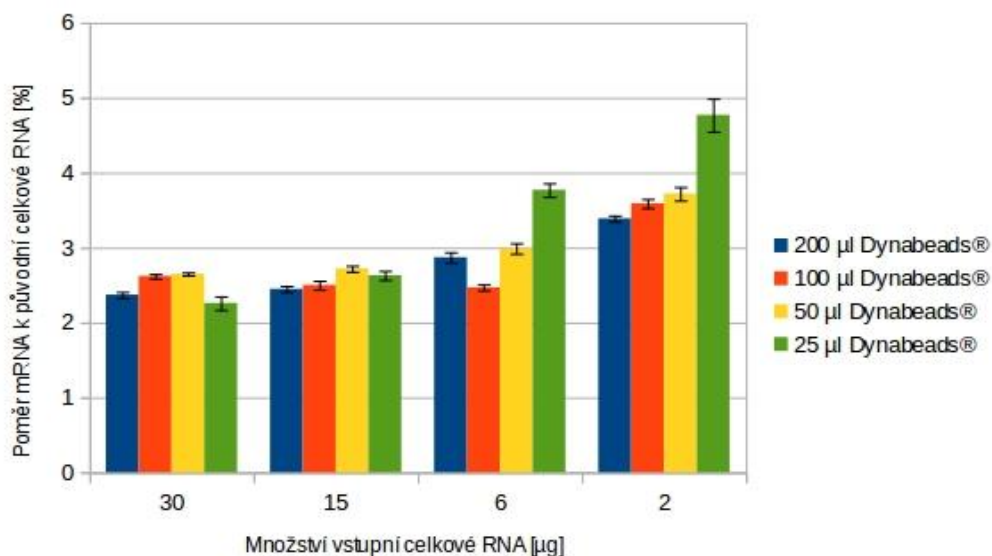


Obrázek č.17: Podoba elektroforetogramu mRNA vzorku A

Tabulka č.7: Naměřené a vypočtené hodnoty u vzorků izolované mRNA

Vzo- rek	Celkem po izolaci [ng]	Směro- datná odchylka	Míra kontami- nace zbytko- vou rRNA [%]	Směrodatná odchylka	Celkem vyizolova- né mRNA [ng]	Směro- datná odchylka	Poměr mRNA k původní celkové RNA [%]	Směro- datná odchyl- ka
A	77,0	2,16	12,0	0,36	67,76	1,76	3,39	0,09
B	197,0	2,83	12,4	0,36	172,80	1,44	2,87	0,02
C	422,0	4,50	12,8	0,37	367,41	4,62	2,45	0,03
D	802,0	13,14	11,5	0,24	709,77	12,57	2,37	0,04
E	80,4	1,30	10,8	0,08	71,72	1,11	3,59	0,06
F	173,0	2,94	14,2	0,29	148,43	2,31	2,47	0,04
G	432,0	8,99	13,2	0,42	375,29	8,99	2,50	0,06
H	901,0	16,27	12,9	0,37	784,77	11,80	2,62	0,04
I	87,0	2,12	14,5	0,09	74,90	1,86	3,72	0,09
J	200,0	5,10	10,3	0,16	179,40	4,26	2,99	0,07
K	456,0	7,79	10,5	0,16	408,12	6,32	2,72	0,04
L	902,0	21,60	11,9	0,29	794,66	21,39	2,65	0,07
M	109,0	4,97	12,6	0,24	95,27	4,37	4,77	0,22
N	268,0	5,66	15,6	0,29	226,19	5,02	3,77	0,09
O	454,0	11,58	13,1	0,16	394,52	9,93	2,63	0,06
P	782,0	14,14	13,2	0,29	678,74	10,06	2,26	0,04

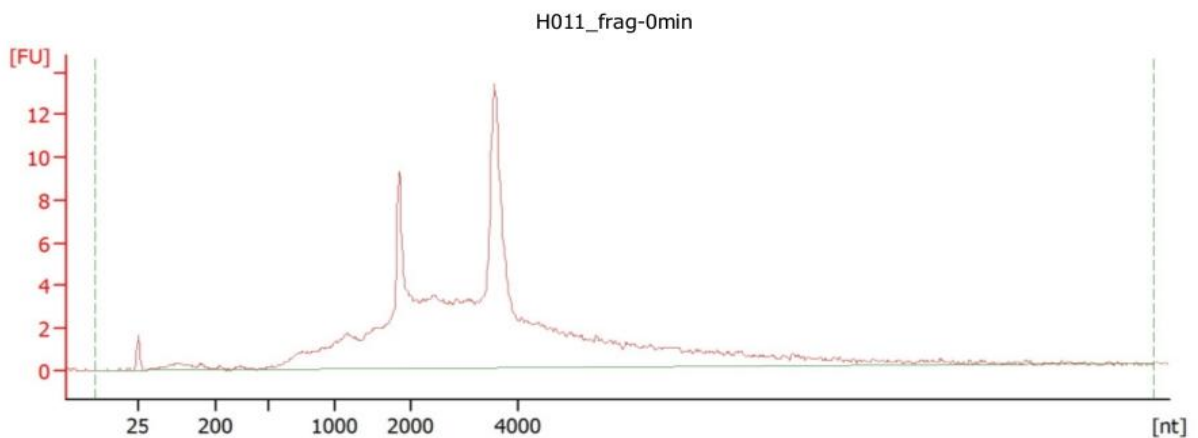
Graf č.1: Porovnání vlivu množství mikrokuliček Dynabeads® a vstupního množství celkové RNA na účinnost izolace.



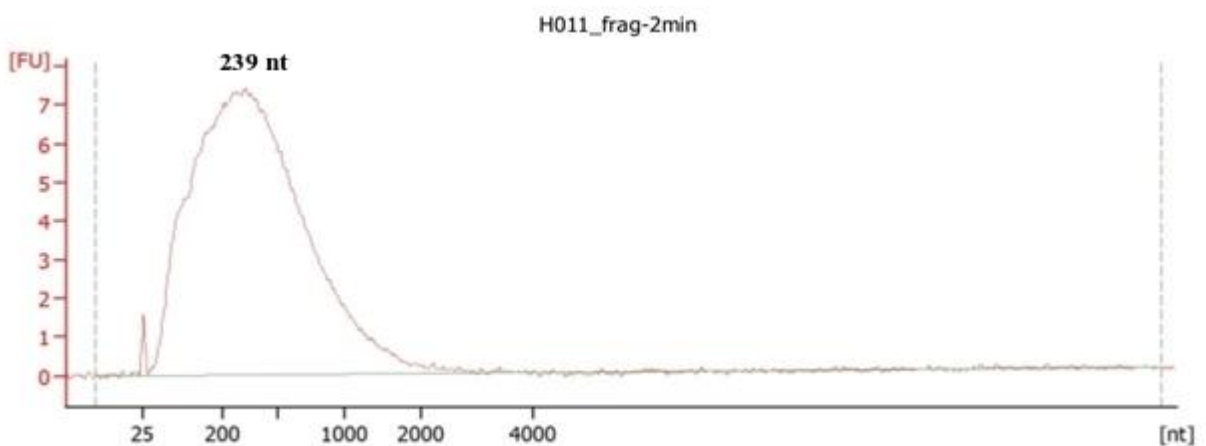
5.3. Fragmentace mRNA

Obrázky č.18-21 vyobrazují podobu elektroforetogramů vzorků mRNA. V případě obrázku č.18 se jedná o vzorek nefragmentovaný, ovšem obrázky č.19-21 vyobrazují elektroforetogramy fragmentovaných vzorků. Velikost fragmentů se lišila podle dob fragmentace vzorku, kdy vzorek fragmentovaný 2 minuty vykazoval fragmenty o průměrné velikosti 239 nt, vzorek fragmentovaný 4 minuty měl průměrnou velikost fragmentů 195 nt a vzorek fragmentovaný 6 minut měl průměrnou velikost fragmentů nejmenší a to 164 nt.

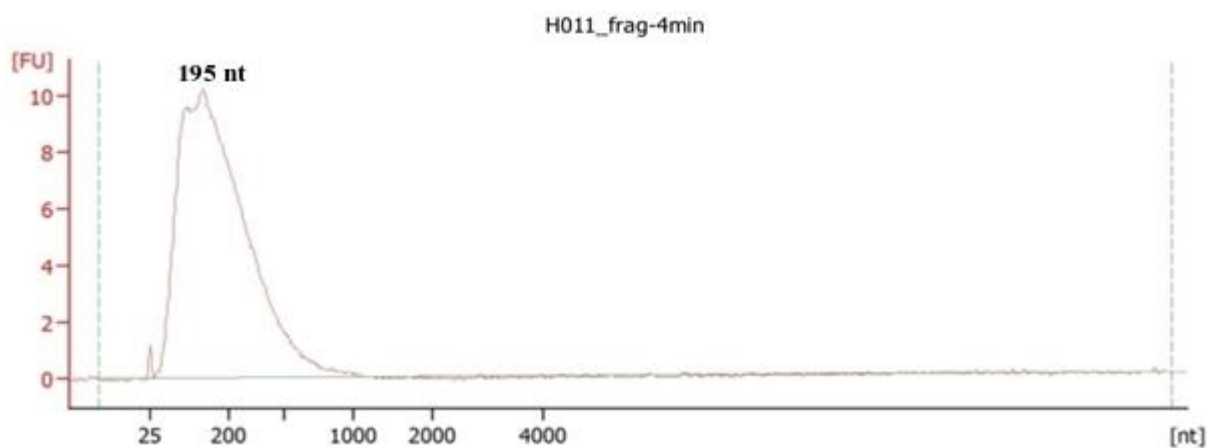
Tabulka č. 8 obsahuje naměřené průměrné velikosti fragmentů u jednotlivých fragmentovaných vzorků.



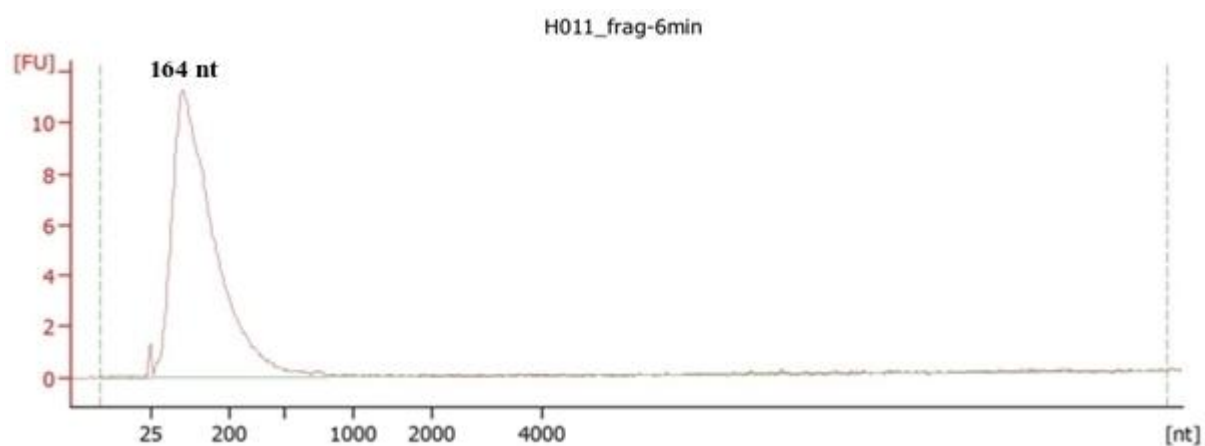
Obrázek č.18: Vzorek nefragmentované mRNA, který však obsahoval fragmentační směs



Obrázek č.19: Vzorek fragmentované mRNA, která byla fragmentována po dobu 2 minut



Obrázek č.20: Vzorek fragmentované mRNA, která byla fragmentována po dobu 4 minut



Obrázek č.21: Vzorek fragmentované mRNA, která byla fragmentována po dobu 6 minut

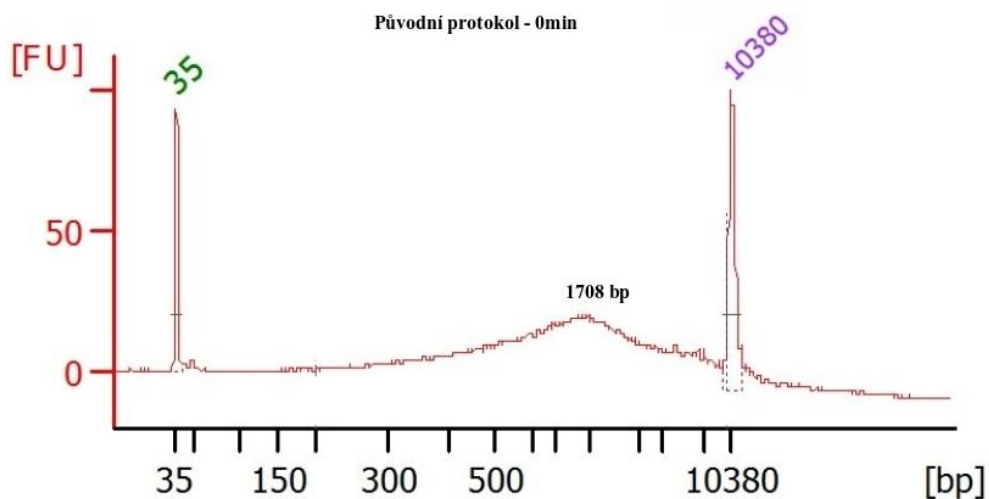
Tabulka č.8: Průměrné velikosti fragmentů fragmentovaných vzorků mRNA

Vzorek	Průměrná velikost fragmentů mRNA [nt]
H011_frag-2min	239
H011_frag-4min	195
H011_frag-6min	164

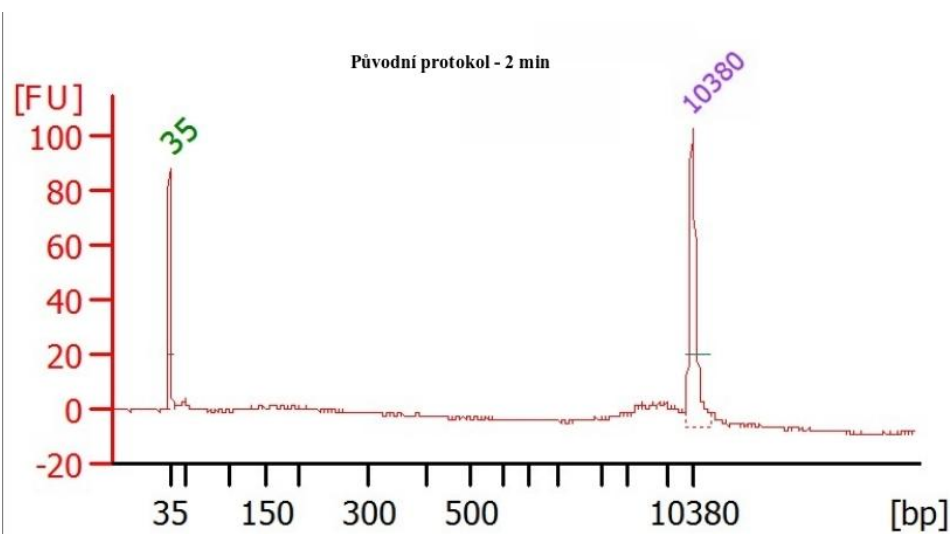
5.4. Reverzní transkripce

Obrázky č.22-25 zobrazují podobu elektroforetogramů vzorků cDNA, které byly připraveny podle protokolu č.1a. V případě obrázku č.22 se jedná o vzorek, který vznikl reverzní transkripcí nefragmentované mRNA a můžeme vidět, že průměrná velikost vzorku cDNA je 1708 bp. Molarita tohoto vzorku byla 1,4 nmol/l. Fragmentované vzorky, které jsou na obrázcích č. 23-25 nebyly detekovatelné. Obrázky č.26-29 zobrazují elektroforetogramy vzorků cDNA připravených podle protokolu č.2a.

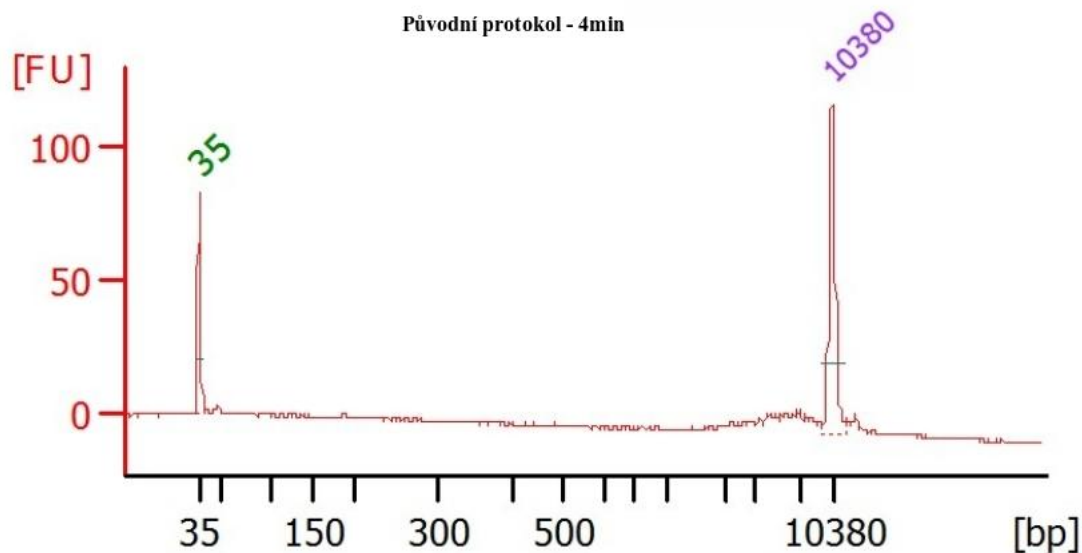
Tabulka č. 9 obsahuje data vzorků připravených protokolem č.2a o naměřeném množství cDNA, velikosti fragmentů jednotlivých vzorků cDNA a také jejich molaritu naměřenou v oblasti od 35 do 1000 bp.



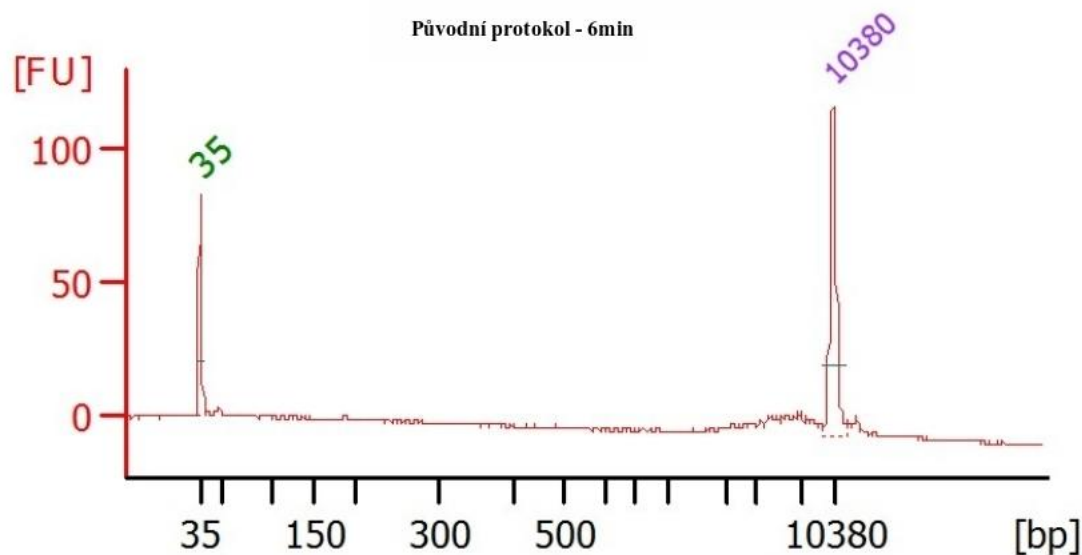
Obrázek č.22: Vzorek nefragmentované cDNA (protokol č.1a)



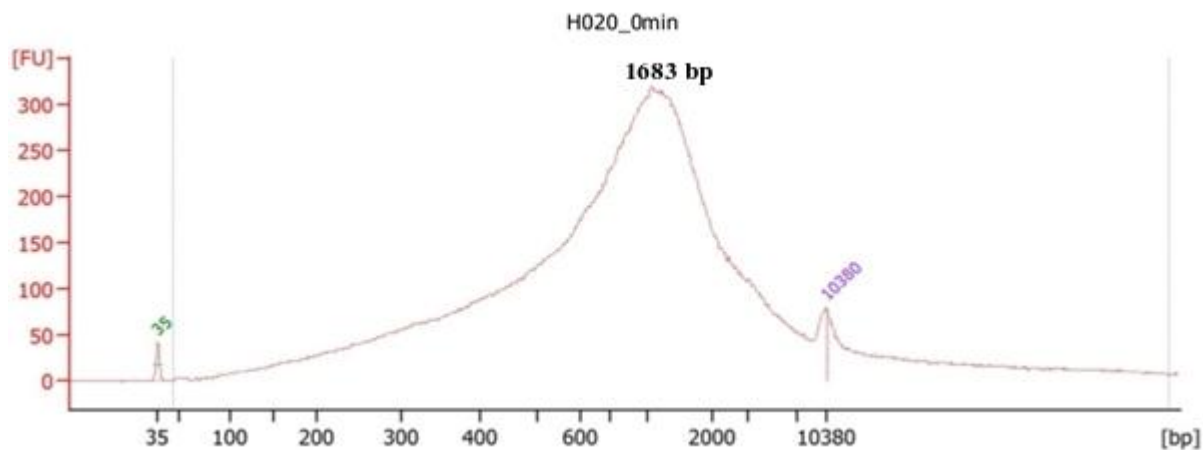
Obrázek č.23: Vzorek cDNA fragmentovaný 2 minuty (protokol č.1a)



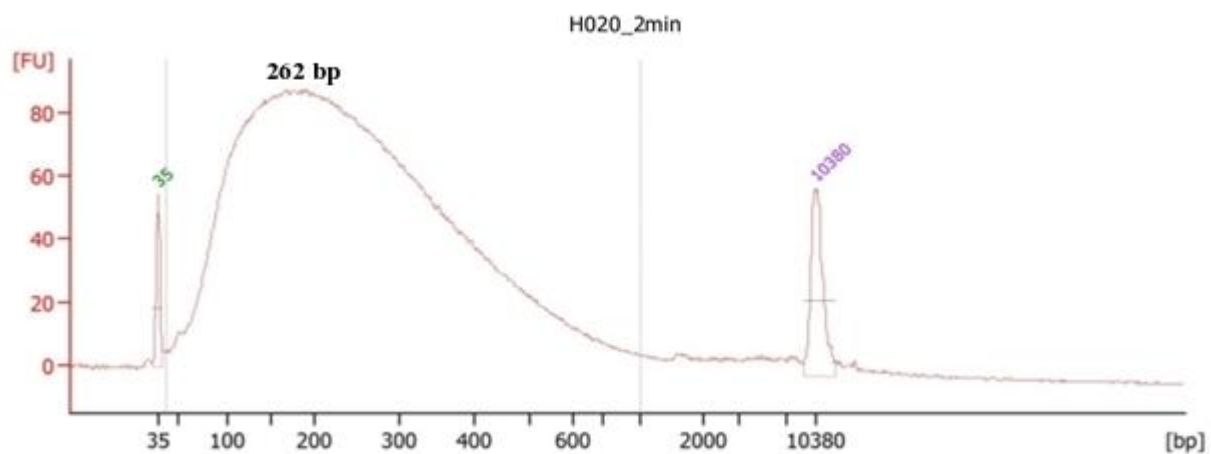
Obrázek č.24: Vzorek cDNA fragmentovaný 4 minuty (protokol č.1a)



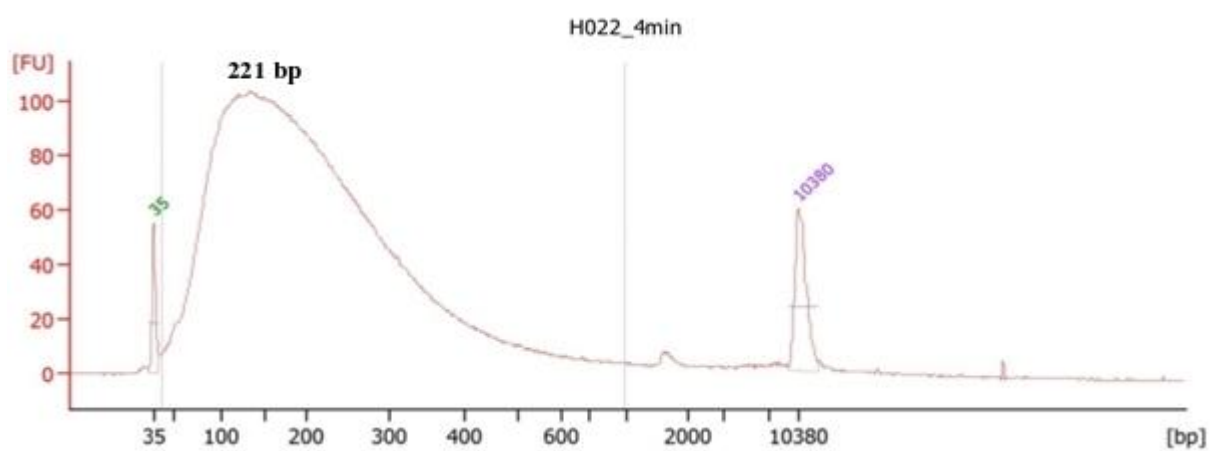
Obrázek č.25: Vzorek cDNA fragmentovaný 6 minut (protokol č.1a)



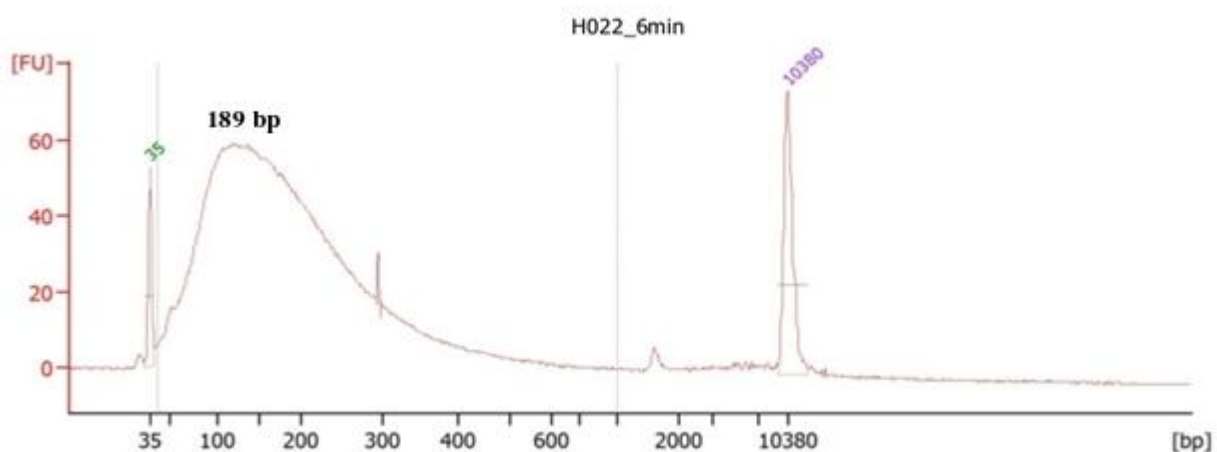
Obrázek č.26: Vzorek nefragmentované cDNA (protokol č.2a)



Obrázek č.27: Vzorek cDNA fragmentované 2 minuty (protokol č.2a)



Obrázek č.28: Vzorek cDNA fragmentované 4 minuty (protokol č.2a)



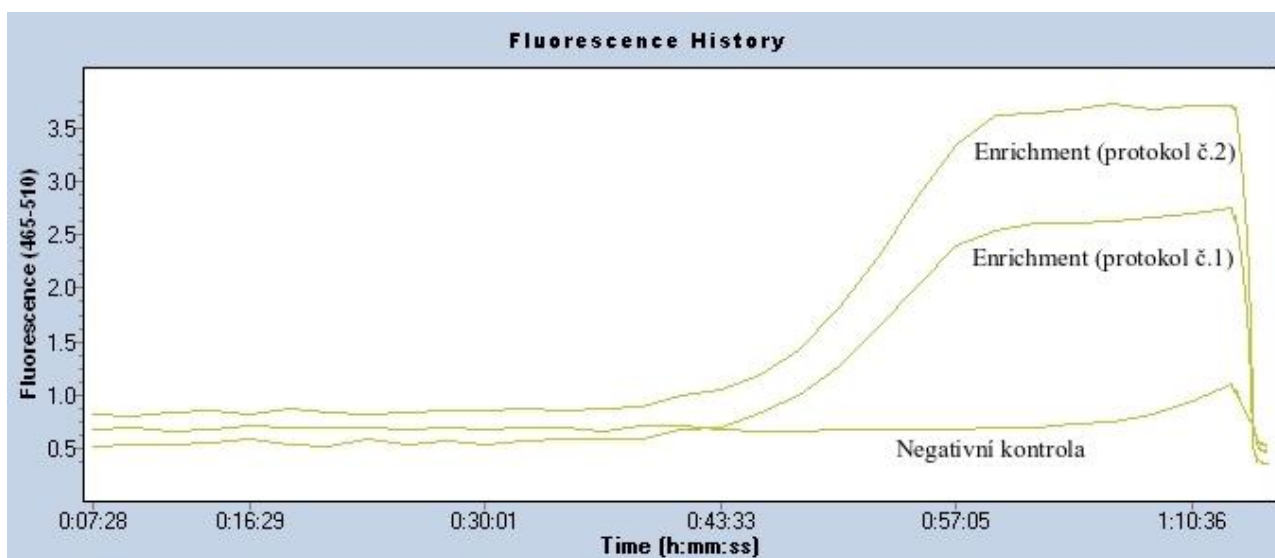
Obrázek č.29: Vzorek cDNA fragmentované 6 minut (protokol č.2a)

Tabulka č.9: Množství cDNA, průměrná velikost fragmentů a molarita vzorků po reverzní transkripci protokolem č.2a.

Vzorek	Množství cDNA [ng]	Průměrná velikost fragmentů cDNA	Molarita [nmol/l]
H020_0min	114	1683	12,5
H020_2min	36,9	262	10,8
H022_4min	31,9	221	9,4
H022_6min	13,7	189	8,6

5.5. Enrichment

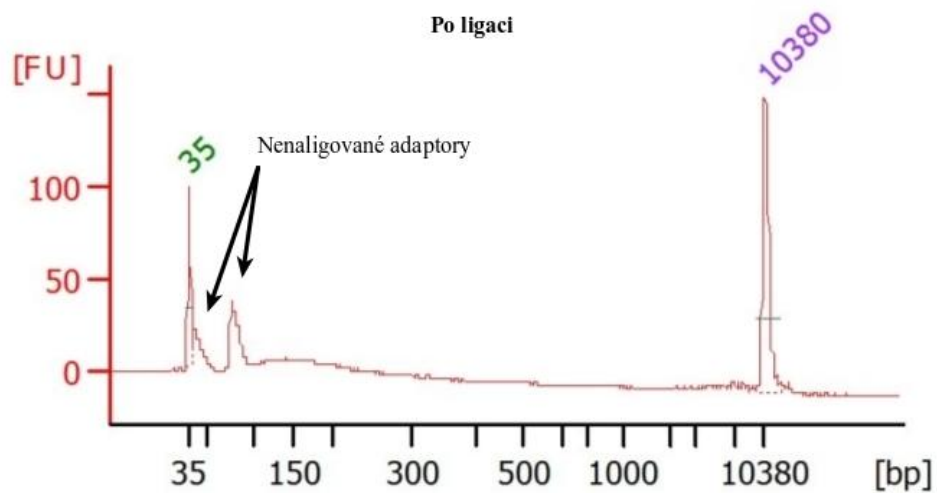
Na obrázku č.30 je podoba křivky míry fluorescence detekované během qPCR u vzorků připravených podle protokolu č.1b, 2b a negativní kontroly. Nejvíce se amplifikovala knihovna připravena podle protokolu č.2b. O něco méně účinněji probíhala amplifikace u knihovny připravené protokolem č.1. U obou vzorků byla amplifikace detekovatelná od 15 cyklu. Amplifikace u negativní kontroly neprobíhala.



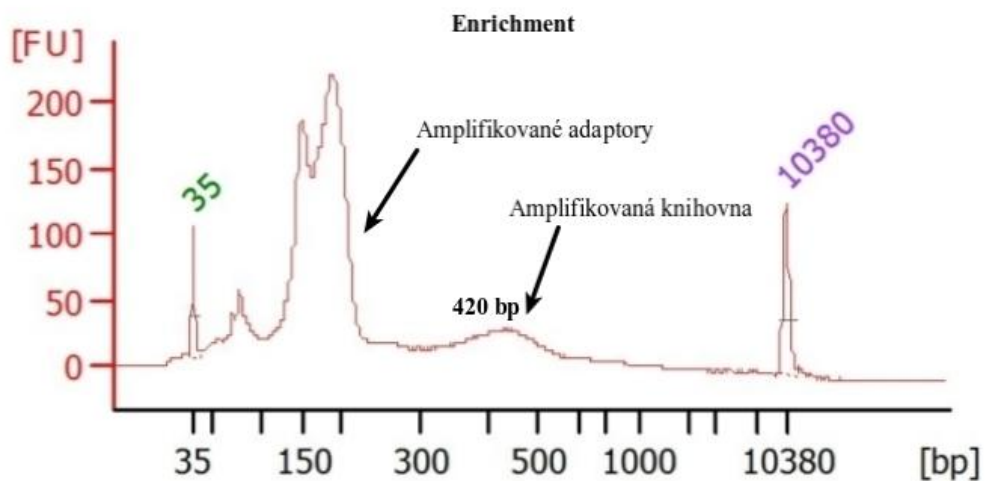
Obrázek č.30: Graf s křivkami fluorescence vzorků knihoven a negativní kontroly z qPCR.

5.6. Ověření připravené knihovny

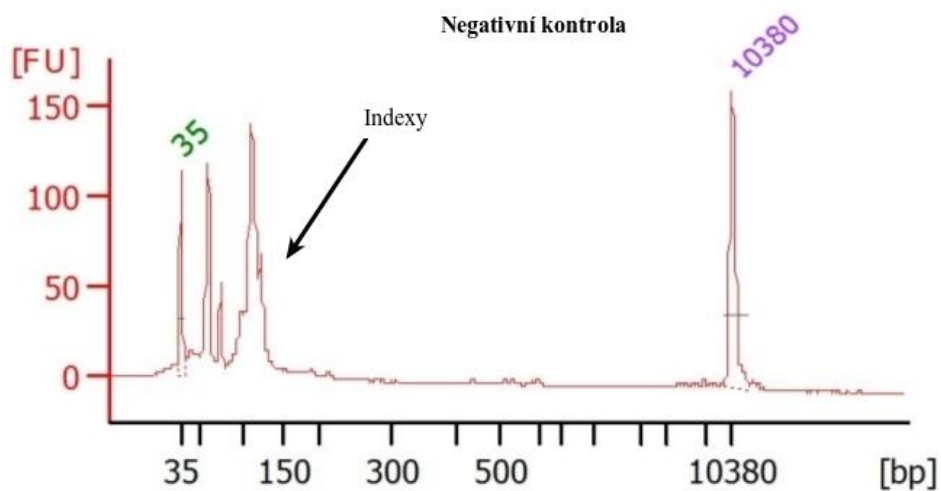
Na obrázcích č. 31-33 jsou vyobrazeny elektroforetogramy vzorků připravených pomocí protokolu č.1b. Na obrázku č.32 je viditelná amplifikovaná knihovna po nerichmentu o průměrné velikosti 420 bp. Obrázky č.34-36 jsou elektroforetogramy vzorků připravených podle protokolu č.2b. Na obrázku č.35 je vidět amplifikovaná knihovna po enrichmentu o průměrné velikosti 416 bp. Tabulka č. 10 obsahuje naměřené hodnoty molarity v oblasti od 300 do 1000 párů bází.



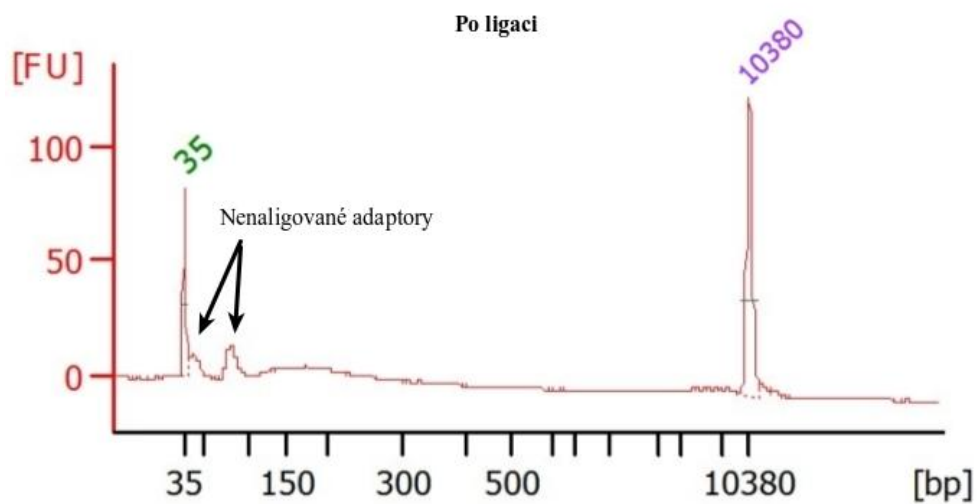
Obrázek č.31: Elektroforetogram vzorku po ligaci a přečištění na mikrokuličkách (protokol č.1b).



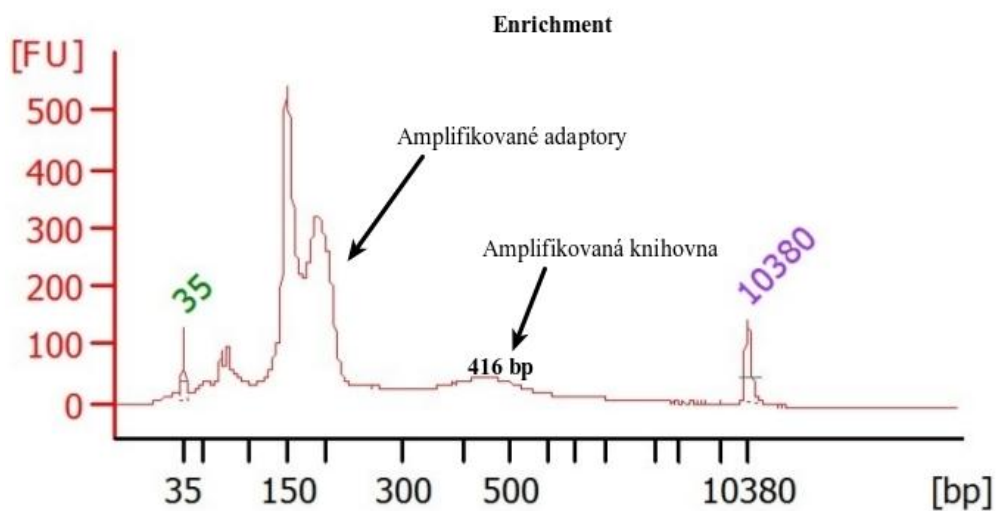
Obrázek č. 32: Elektroforetogram vzorku po enrichmentu a přečištění na kolonách (protokol č.1b).



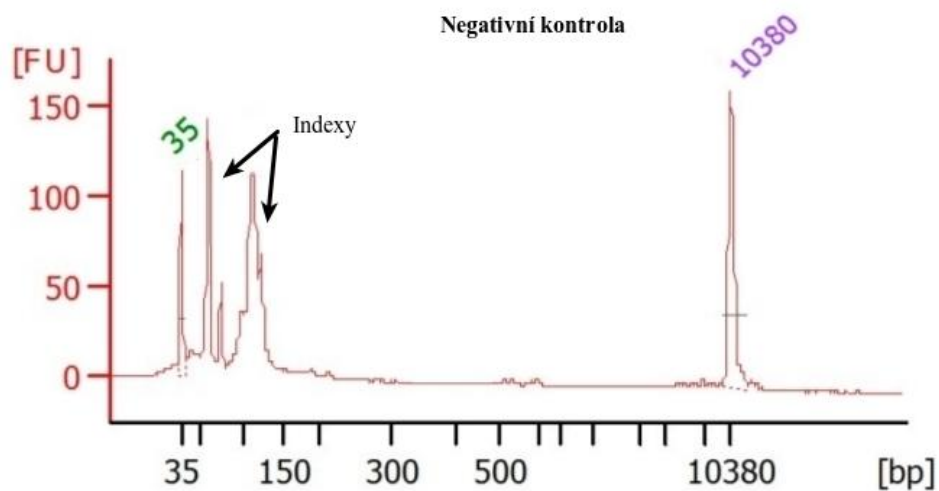
Obrázek č.33: Negativní kontrola enrichmentu (protokol č.1b).



Obrázek č.34: Elektroforetogram vzorku po ligaci a přečištění na mikrokuličkách (protokol č.2b).



Obrázek č.35: Elektroforetogram vzorku po enrichmentu a přečištění na kolonách (protokol č.2b).



Obrázek č.36: Negativní kontrola enrichmentu (protokol č.2b).

Tabulka č.10: Naměřené hodnoty molarity v jednotlivých krocích přípravy knihovny.

Protokol č.	Ligace [pmol/l]	Enrichment [pmol/l]
1	91,8	2282
2	123,5	2951

6. Diskuze

Cílem experimentální části této bakalářské práce bylo optimalizovat alternativní postup přípravy sekvenačních knihoven pro platformu Illumina, tak aby nebylo nutné využívat komerčně dodávaný kit na přípravu těchto knihoven.

Hodnoty koncentrace a čistoty vzorků z izolace celkové RNA odpovídaly našim požadavkům. Hodnota koncentrace celkové RNA se u všech vzorků pohybovala okolo 2000 ng/μl. Dále byly používány vzorky již pod označením odpovídající způsobu následné izolace mRNA.

Protokol výrobce doporučoval jako vstupní množství při izolaci mRNA 75 μg celkové RNA. Bylo nutné jej optimalizovat pro naše podmínky. Nejlepší výsledky byly dosaženy při izolaci mRNA za použití 25 μl mikrokuliček Dynabeads[®], ke kterým byly přidány 2 μg celkové RNA. mRNA získaná touto upravenou verzí protokolu tvořila 4,76 % původního vzorku, což bylo o 1,37 % více, než při použití 200 μl mikrokuliček Dynabeads[®] a zachování stejného vstupního množství celkové RNA. Důvodem proč bylo možné použít mnohem menší množství mikrokuliček je jejich kapacitance, kdy tyto mikrokuličky jsou schopny v objemu 200 μl navázat až 1 μg mRNA. Tato úprava izolačního protokolu přinesla kromě snížení nároků na vstupní množství celkové RNA také možnost využití izolačního kitu na větší počet vzorků, přesněji namísto 10 vzorků jej s touto úpravou lze použít na 80 izolací. Tento fakt má především velmi výrazný vliv na cenu izolace jednoho vzorku. S původní úpravou protokolu by cena izolace jednoho vzorku vyšla na 1217,- Kč. S naší úpravou byla částka snížena na 152,- Kč [116].

Následně byla provedena fragmentace mRNA pomocí teploty a hořčičných kationtů. Tento způsob je nejběžnější metodou fragmentace RNA [117]. Výsledná hodnota průměrné velikosti fragmentů byla pro vzorek fragmentovaný 2 minuty 239 nukleotidů, pro vzorek fragmentovaný 4 minuty 195 nukleotidů a vzorek, který byl fragmentován 6 minut, měl průměrnou velikost fragmentů 164 nukleotidů. Pro další práci byly použity všechny fragmentované vzorky a také vzorek nefragmentované mRNA, aby bylo možné zjistit, zda-li jednotlivé kroky při syntéze cDNA nemají vliv na průměrnou velikost fragmentů.

Reverzní transkripce vycházela z protokolu „Strand-specific libraries“ [46]. Fragmentované vzorky, které jsme připravovali tímto protokolem, nebyly detekovatelné. Nefragmentovaný vzorek však detekovatelný byl, což znamená, že samotná reverzní transkripce proběhla. S největší pravděpodobností bylo příčinou těchto nespecifických výsledků čištění mezi a po syntéze vláken cDNA, během kterých se fragmenty odmyly ze vzorku. Přistoupili jsme tedy k modifikaci tohoto protokolu. Při modifikaci bylo především změněno množství SPRI mikrokuliček použitých při čistících krocích a množství vody, do které se vzorky po čištění eluují. Dále pak množství použitých univerzálních primerů a Actinomycinu D. Bylo použito 1 μg univerzálních primerů namísto 3 μg a také bylo sníženo množství vody, která se přidává v kroku nasedání univerzálních

primerů, z 7 μ l na 3 μ l. Změna při tomto kroku byla kompromisem mezi původním protokolem a protokolem „First Strand cDNA Synthesis“ [118]. Tento protokol byl originálním protokolem pro Reverse Transcriptase H(-), kterou jsme použili k syntéze prvního vlákna namísto SuperScript III Reverse transcriptase. Pro použitou reverzní transkriptázu bylo nutné upravit také teploty inkubace, aby odpovídaly ideálním podmínkám pro tuto reverzní transkriptázu. Bylo použito 1,5 μ g Actinomycinu D místo protokolem doporučovaných 4 μ g. Snížení množství Actinomycinu bylo kompromisem mezi původním protokolem a protokolem TotalScript™ [119]. Odborná literatura uvádí, že přítomnost Actinomycinu D není nutná a doporučuje se jen pro lepší reprodukovatelnost výsledků [120]. Při čištění po syntéze prvního vlákna bylo použito 45 μ l SPRI mikrokuliček, čímž byl zvýšen poměr čisticí mikrokuličky:objem vzorku z 1,2:1 na 1,6:1. Toto zvýšení nám zabezpečilo, že nebudou odmyvány menší fragmenty, které chceme zachovat, ale zároveň budou odmyty nejmenší fragmenty, které by mohly způsobovat různé nespecifity v dalších krocích. Eluce probíhala do 13 μ l namísto 22 μ l DEPC vody. U syntézy druhého vlákna cDNA nebyly změněny objemy reagensů ani čas či teplota inkubace. Čištění po syntéze druhého vlákna probíhalo do stejného množství mikrokuliček, jaké je uvedeno v protokolu. Jedinou změnou byla eluce do 15 μ l DEPC vody. Průměrná velikost fragmentů po reverzní transkripci byla u vzorku, který byl fragmentován 2 minuty, byla 269 bp, u vzorku fragmentovaného 4 minuty měly fragmenty průměrnou velikost 221 bp a u vzorku fragmentovaného 6 minut byla hodnota průměrné velikosti fragmentů 189 bp. Za ideální dobu, která je nutná pro fragmentaci vzorku mRNA jsme tedy určily dobu 4 minut. Takto fragmentovaná mRNA vykazuje ideální velikost, která při napojení adaptorů a indexů bude nejbližší schopnosti čtení sekvenátoru a plně tedy využije jeho kapacity [75]. Po přípravě cDNA tímto změněným protokolem byly získány průkazné výsledky, které potvrdily, že náš modifikovaný protokol funguje správně a bude tedy používán k další práci.

Pro přípravu cDNA knihovny byl tedy použit vzorek získaný z mRNA fragmentované po dobu 4 minut. Pro tuto přípravu však již nebyl použit předlohou výše zmíněný protokol, ale dva postupy, které se v Laboratoři experimentální medicíny Ústavu molekulární a translační medicíny používají v prvním případě k přípravě knihoven z DNA amplikonů (protokol č.1b) a v případě druhém k přípravě celogenomových knihoven (protokol č.2b).

Ligace adaptorů fungovala v případě obou použitých protokolů správně, což potvrdil následný enrichment. Analýzou fluorescenčních křivek získaných při kvantitativní PCR bylo zjištěno, že k amplifikaci dochází v případě obou vzorků. Větší nárůst fluorescence v případě vzorku připravovaného podle protokolu č.2b byl dán větším vstupním množstvím cDNA, která byla dána do reakce. Analýzou vzorků po ligaci a po enrichmentu bylo zjištěno, že velkou část naměřené amplifikace při kvantitativní PCR tvořila amplifikace nenavázaných adaptorů. Zároveň ale bylo potvrzeno, že zvolené postupy modifikace konců, ligace adaptorů a enrichmentu fungují,

jelikož byly v elektroforetogramu patrné také amplifikované knihovny o průměrné velikosti 420 bp.

Do budoucna bude tedy nutné optimalizovat množství použitých mikrokuliček v čistícím kroku po ligaci tak, aby byly odstraněny nenavázané adaptory, ale zároveň došlo k zachování fragmentů vzorku. I přes nutnou optimalizaci čištění po ligaci je možné konstatovat, že příprava knihoven pomocí modifikace alternativního postupu funguje. Dalším krokem, kromě optimalizace bude porovnání kvality knihovny připravené podle protokolu č.1b a protokolu č.2b s knihovnou připravenou pomocí komerčního kitu. Analýza bude provedena přípravou těchto knihoven ze stejného vzorku a jejich následné sekvenaci. Analýzou získaných dat budeme moci potvrdit správnost postupu přípravy knihovny alternativním protokolem.

V neposlední řadě přináší tento alternativní postup významný posun ve finanční náročnosti přípravy sekvenačních knihoven. Cena přípravy jednoho vzorku za použití originálního kitu je 1770,- Kč [121]. Použitím našich alternativních protokolů se cena snížila na přibližně 620,-Kč.

7. Závěr

Teoretická část této práce byla vypracována na základě dostupné odborné literatury a byly v ní shrnuty poznatky z oblasti sekvenčních technologií, s důrazem na přípravu sekvenančních knihoven pro sekvenanční platformu Illumina a analýzu získaných sekvenčních dat. Popsány byly také možnosti využití sekvenčních technologií v oblasti výzkumu a medicíny.

V experimentální části práce byla zkoumána možnost přípravy sekvenační knihovny pro sekvenování transkriptomu bez použití komerčního kitu. K přípravě byla použita buněčná linie U2OS. Hlavní pozornost byla věnována optimalizaci reverzní transkripce mRNA do cDNA a následně optimalizaci ligace adaptorů a enrichmentu sekvenanční knihovny. Byla optimalizována izolace mRNA, pomocí Dynabeads[®] mRNA Purification Kitu, tak aby měla co nejnižší nároky na množství vstupního materiálu. Dalším kritériem pro optimalizaci byla co nejnižší cena za izolaci jednoho vzorku. Optimalizací izolace, a následné reverzní transkripce, bylo dosaženo požadovaného množství získané cDNA o odpovídající velikosti fragmentů. K optimalizaci přípravy knihovny z cDNA byly použity dva odlišné používané protokoly, které byly následně modifikovány pro potřeby přípravy knihovny z cDNA. Výsledky ukázaly, že je ještě nutná optimalizace čištění knihoven po ligaci adaptorů. Tato modifikace bude předmětem dalšího zkoumání.

8. Zdroje

- [1] Pospíšilová Š, Tichý B, Mayer J.. Sekvenování lidského genomu – Technologie nové generace aneb budeme rutinně sekvenovat lidské genomy?. Čas. Lék. čes. 2009;148:296-302
- [2] Shizuya H, Birren B, Kim UJ, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(18):8794-8797.
- [3] Schulte D, Ariyadasa R, Shi B, et al. BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.). BMC Genomics. 2011;12:247.
- [4] Takahashi R, Ueda M. Generation of transgenic rats using YAC and BAC DNA constructs. Methods Mol Biol. 2010;597:93-108..
- [5] Mapa YAC plazmidu (pYAC) a vyobrazení YAC klonovacího systému. [obrázek] [online] [cit. 2015-03-25] Dostupné z: <http://nptel.ac.in/courses/102103045/module3/lec19/images/s2.jpg>
- [6] Christopher Anderson. YAC Troubles Run Deep. Science. 1993; 259:1684
- [7] Neil DL, Villasante A, Fisher RB, et al. Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucleic Acids Research. 1990;18(6):1421-1428.
- [8] Larionov V, Kouprina N, Nikolaishvili N, et al. Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). Nucleic Acids Research. 1994;22(20):4154-4162.
- [9] Haldi, M., Perrot, V., Saumier, M., et al. Large human YACs constructed in a rad52 strain show a reduced rate of chimerism. Genomics. 1994;24(3):478-84.
- [10] Woo SS, Jiang J, Gill BS, et al. Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Research. 1994;22(23):4922-4931.
- [11] Marra MA, Kucaba TA, Dietrich NL, et al. High Throughput Fingerprint Analysis of Large-Insert Clones. Genome Research. 1997;7(11):1072-1084.
- [12] Yu-Hui Rogers and J. Craig Venter. Massively parallel sequencing. Nature. 2005;437:326-327
- [13] F. Sanger, S. Nicklen et A. R. Coulson. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 1997;74:5463-5467
- [14] F. Sanger, A.R. Coulson, T. Friedmann et al. The nucleotide sequence of bacteriophage φX174. Journal of Molecular Biology 1978;125:225–246
- [15] DNA sequencing - the Sanger method [online] [cit. 2015-03-25] Dostupné z: <http://www.wellcome.ac.uk/Education-resources/Education-and-learning/Resources/Animation/WTDV026689.htm>

- [16] The Sanger Method [online] [cit. 2015-03-25] Dostupné z: http://www.bio.davidson.edu/courses/molbio/molstudents/spring2003/obenrader/sanger_method_page.htm
- [17] DNA Sequencing Module [online] [cit. 2015-03-25] Dostupné z: http://hshgp.genome.washington.edu/teacher_resources/modules-view.htm
- [18] Karger BL, Guttman A. DNA Sequencing by Capillary Electrophoresis. *Electrophoresis*. 2009;30(Suppl 1):S196-S202.
- [19] Marie C. Ruiz-Martinez, Jan Berka, J Alexei Belenkii et al. DNA Sequencing by Capillary Electrophoresis with Replaceable Linear Polyacrylamide and Laser- Induced Fluorescence Detection. *Anal. Chem.* 1993;65:2851-2858
- [20] A. Guttman, A. S. Cohen , D. N. Heiger and Barry L. Karger. Analytical and micropreparative ultrahigh resolution of oligonucleotides by polyacrylamide gel high-performance capillary electrophoresis. *Anal. Chem.* 1990;62(2):137–141
- [21] Kan CW, Fredlake CP, Doherty EA and Barron AE. DNA sequencing and genotyping in miniaturized electrophoresis systems. *Electrophoresis*. 2004;25(21-22):3564-88.
- [22] Capillary_Gel_Electrophoresis_Instrument_Schematic [obrázek] [online] [cit. 2015-03-25] Dostupné z: http://upload.wikimedia.org/wikipedia/commons/9/93/Capillary_Gel_Electrophoresis_Instrument_Schematic.png
- [23] DNA Sequencing Core [obrázek] [online] [cit. 2015-03-25] Dostupné z: <http://seqcore.brcf.med.umich.edu/doc/dnaseq/normal3730.gif>
- [24] Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Research*. 2006;34(3):e22.
- [25] Bentley D.R., Balasubramanian S., Swerdlow H.P., et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53–59.
- [26] Nyrén P. The history of pyrosequencing. *Methods Mol. Biol.* 2007;373:1-14.
- [27] Mostafa Ronaghi, Mathias Uhlén, Pål Nyrén. A Sequencing Method Based on Real-Time Pyrophosphate. *Science*. 1998;281:363-365
- [28] 454 Life Sciences Ships First Genome Sequencer FLX System(TM) to Roche Diagnostics [online] [cit. 2015-03-25] Dostupné z: <http://454.com/resources-support/news.asp?display=detail&id=63>
- [29] Shale Dames, Jacob Durtschi, Katherine Geiersbach et al. Comparison of the Illumina Genome Analyzer and Roche 454 GS FLX for Resequencing of Hypertrophic Cardiomyopathy-Associated Genes. *J Biomol Tech.* 2010;21(2):73–80.
- [30] 454 Sequencing System Guidelines for Amplicon Experimental Design [online] [cit. 2015-03-25] Dostupné z: http://my454.com/downloads/my454/applications-info/454SequencingSystem_GuidelinesforAmpliconExperimentalDesign_July2011.pdf

- [31] Jarvie T. Next generation sequencing technologies. *Drug Discov Today Technol.* 2005;2(3):255-60
- [32] Marcel Margulies, Michael Egholm, William E. Altman. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature.* 2005;437:376-380.
- [33] Od Sangera ke klinickým aplikacím sekvenování lidského genomu [online] [cit. 2015-03-25] Dostupné z: http://web2.stapro.cz/fons/archiv/2010/after/prednasky/5_1.pdf
- [34] DNAsynthesis [obrázek] [online] [cit. 2015-03-25] Dostupné z: <http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAsynthesis5eL3.png>
- [35] 454_reaction [obrázek] [online] [cit. 2015-03-25] Dostupné z: http://cdn.arstechnica.net/Science/April10/454_reaction.png
- [36] Edward David Hyman. A new method of sequencing DNA. *Analytical Biochemistry.* 1988;174:423-436
- [37] System Workflow [online] [cit. 2015-03-25] Dostupné z: <http://454.com/products/technology.asp>
- [38] Mostafa Ronaghi. Pyrosequencing Sheds Light on DNA Sequencing. *Genome Res.* 2001;11:3-11
- [39] Zagordi O, Däumer M, Beisel C, Beerenwinkel N. Read length versus Depth of Coverage for Viral Quasispecies Reconstruction. Poon AFY, ed. *PLoS ONE.* 2012;7(10):e47046.
- [40] Maureen J. Donlin, Smita S. Patel et Kenneth A. Johnson. Kinetic Partitioning between the Exonuclease and Polymerase Sites in DNA Error Correction. *Biochemistry* 1991;30:538-546
- [41] Liu L, Li Y, Li S, et al. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology.* 2012;2012:251364.
- [42] Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat) [online] [cit. 2015-03-25] Dostupné z: <http://www.illumina.com/products/ribo-zero-rrna-removal-human-mouse-rat.html>
- [43] FAQ Ribo-Zero Kits [online] [cit. 2015-03-25] Dostupné z: <http://www.illumina.com/products/ribo-zero-rrna-removal-kit-faqs.html>
- [44] Sambrook J, Russell DW. Fragmentation of DNA by nebulization. *CSH Protoc.* 2006;2006(4)
- [45] Bronner IF, Quail MA, Turner DJ, Swerdlow H. Improved Protocols for Illumina Sequencing. *Current protocols in human genetics / editorial board, Jonathan L Haines . [et al].* 2009;0 18:10.1002/0471142905.hg1802s62.
- [46] Zhang Z, Theurkauf WE, Weng Z, Zamore PD. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence.* 2012;3(1):9.
- [47] Picelli S, Björklund ÅK, Reinius B, et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research.* 2014;24(12):2033-2040.

- [48] Adey A, Morrison HG, Asan, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*. 2010;11(12):R119.
- [49] Knierim E, Lucke B, Schwarz JM, et al. Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing. Gilbert MTP, ed. *PLoS ONE*. 2011;6(11):e28240.
- [50] Erwin L. van Dijka, Yan Jaszczyszynb and Claude Thermesa. Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research*. 2014;322:12-20
- [51] Hawkins TL, O'Connor-Morin T, Roy A, Santillan C. DNA purification and isolation using a solid-phase. *Nucleic Acids Research*. 1994;22(21):4543-4544.
- [52] Bowman SK, Simon MD, Deaton AM, et al. Multiplexed Illumina sequencing libraries from picogram quantities of DNA. *BMC Genomics*. 2013;14:466.
- [53] Illumina [online] [cit. 2015-03-25] Dostupné z: http://ycga.yale.edu/sequencing/Illumina/240_21592_2.pdf
- [54] Illumina_fragments [obrázek] [online] [cit. 2015-03-25] Dostupné z: http://www.dkfz.de/gpcf/fileadmin/_migrated/RTE/RTEmagicC_Illumina_fragments.png
- [55] Datasheet cbot [online] [cit. 2015-03-25] Dostupné z: http://res.illumina.com/documents/products/datasheets/datasheet_cbot.pdf
- [56] Illumina Sequencing [online] [cit. 2015-03-25] Dostupné z: <https://rtsf.natsci.msu.edu/genomics/sequencing-services/ngs/illumina/>
- [57] Kozarewa I, Ning Z, Quail MA, et al. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of GC-biased genomes. *Nature methods*. 2009;6(4):291-295.
- [58] ILLUMINA Genome Analyser Sequencing technology: how it works [online] [cit. 2015-03-25] Dostupné z: http://www.giga.ulg.ac.be/jcms/prod_26859/illumina-genome-analyser-sequencing-technology-how-it-works
- [59] Illumina cluster generation [obrázek] [online] [cit. 2015-03-25] Dostupné z: http://www.dkfz.de/gpcf/fileadmin/_migrated/pics/Illumina_cluster_generation_02.png
- [60] Martin Kircher, Susanna Sawyer and Matthias Meyer. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 2012;40(1):e3.
- [61] Illumina sequencing [obrázek] [online] [cit. 2015-03-25] Dostupné z: http://www.dkfz.de/gpcf/fileadmin/_migrated/pics/Illumina_sequencing_02.png
- [62] Data Processing of Nextera[®] Mate Pair Reads on Illumina Sequencing Platforms [online] [cit. 2015-03-25] Dostupné z: http://www.illumina.com/documents/products/technotes/technote_nextera_matepair_data_processing.pdf

- [63] Cock PJA, Fields CJ, Goto N, et al. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*. 2010;38(6):1767-1771.
- [64] FASTQ Files [online] [cit. 2015-03-25] Dostupné z: http://support.illumina.com/help/SequencingAnalysisWorkflow/Content/Vault/Informatics/Sequencing_Analysis/CASAVA/swSEQ_mCA_FASTQFiles.htm
- [65] Fastq figure [obrázek] [online] [cit. 2015-03-25] Dostupné z: http://drive5.com/usearch/manual/fastq_fig.jpg
- [66] FastQC: a quality control tool for high throughput sequence data. [online]. [cit. 2015-03-23]. Dostupné z: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- [67] FASTX-TOOLKIT. [online]. [cit. 2015-03-23]. Dostupné z: http://hannonlab.cshl.edu/fastx_toolkit/index.html
- [68] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105-1111.
- [69] Benjamin AM, Nichols M, Burke TW, Ginsburg GS, Lucas JE. Comparing reference-based RNA-Seq mapping methods for non-human primate data. *BMC Genomics*. 2014;15(1):570.
- [70] TopHat .[online]. [cit.2015-03-23]. .Dostupné z: <http://ccb.jhu.edu/software/tophat/index.shtml>
- [71] HTSeq: Analysing high-throughput sequencing data with Python [online] [cit. 2015-03-25] Dostupné z: <http://www-huber.embl.de/HTSeq/doc/overview.html>
- [72] Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169.
- [73] History of Illumina Sequencing [online] [cit. 2015-03-25] Dostupné z: <http://www.illumina.com/technology/next-generation-sequencing/solexa-technology.html>
- [74] Farrer RA1, Kemen E, Jones JD, Studholme DJ. De novo assembly of the *Pseudomonas syringae* pv. *syringae* B728a genome using Illumina/Solexa short sequence reads. *FEMS Microbiol Lett*. 2009;291(1):103-111.
- [75] Illumina systems [online] [cit. 2015-03-25] Dostupné z: <http://systems.illumina.com/systems.ilmn>
- [76] Solexa technology [online] [cit. 2015-03-25] Dostupné z: <http://technology.illumina.com/technology/next-generation-sequencing/solexa-technology.ilmn>
- [77] Illumina (Solexa) Genome Analyzer II DNA Sequencer [online] [cit. 2015-03-25] Dostupné z: <http://genomics.ucr.edu/facility/genomics/instruments/illumina.html>
- [78] Sequencing systems [online] [cit. 2015-03-25] Dostupné z: <http://systems.illumina.com/systems/sequencing.ilmn>
- [79] Guan YF, Li GR, Wang RJ et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer* 2012;31(10):463-470.

- [80] The Human Genome Project Completion: Frequently Asked Questions [online] [cit. 2015-03-25] Dostupné z: <http://www.genome.gov/11006943>
- [81] Jared C. Roach, Andrew F. Siegel, Ger van den Engh et al. Gaps in the Human Genome Project. *Nature*. 1999;401:843-845
- [82] 1000 Genome Project [online] [cit. 2015-03-25] Dostupné z: <http://www.1000genomes.org/about>
- [83] Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*. 2013;41(10):e108.
- [84] Kreimer A, Pe'er I. Co-regulated Transcripts Associated to Cooperating eSNPs Define Bifan Motifs in Human Gene Networks. *Cotsapas C, ed. PLoS Genetics*. 2014;10(9):e1004587. doi:10.1371/journal.pgen.1004587.
- [85] Ley TJ, Mardis ER, Ding L et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*. 2008;456(7218):66-72.
- [86] Maher, C. A., Kumar-Sinha, C., Cao, X., et al. Transcriptome Sequencing to Detect Gene Fusions in Cancer. *Nature*. 2009; 458(7234): 97–101.
- [87] Jaako P, Debnath S, Olsson K, et al. Gene therapy cures the anemia and lethal bone marrow failure in a mouse model of RPS19-deficient Diamond-Blackfan anemia. *Haematologica*. 2014;99(12):1792-1798.
- [88] Jia Q, Zhang Q, Zhang Z, et al. Transcriptome Analysis of the Zebrafish Model of Diamond-Blackfan Anemia from RPS19 Deficiency via p53-Dependent and -Independent Pathways. *Sabaawy HE, ed. PLoS ONE*. 2013;8(8):e71782.
- [89] Twine NA, Janitz K, Wilkins MR, Janitz M. Whole Transcriptome Sequencing Reveals Gene Expression and Splicing Differences in Brain Regions Affected by Alzheimer's Disease. *Preiss T, ed. PLoS ONE*. 2011;6(1):e16266.
- [90] Courtney E, Kornfeld S, Janitz K, Janitz M. Transcriptome profiling in neurodegenerative disease. *J Neurosci Methods*. 2010; 193(2):189-202
- [91] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews Genetics*. 2009;10(1):57-63.
- [92] Castellanos-Martínez S, Arteta D, Catarino S, et al. De Novo Transcriptome Sequencing of the Octopus vulgaris Hemocytes Using Illumina RNA-Seq Technology: Response to the Infection by the Gastrointestinal Parasite *Aggregata octopiana*. *Knight M, ed. PLoS ONE*. 2014;9(10):e107873.
- [93] Kiiialainen A, Karlberg O, Ahlford A, et al. Performance of Microarray and Liquid Based Capture Methods for Target Enrichment for Massively Parallel Sequencing and SNP Discovery. *Tan P, ed. PLoS ONE*. 2011;6(2):e16486.
- [94] MATTOS-ARRUDA, L. D., DIENSTMANN, R. a TABERNERO, J. Development of Molecular Biomarkers in Individualized Treatment of Colorectal Cancer. *Clin Colorectal Cancer*. 2011; 10(4):279-289

- [95] Madu CO, Lu Y. Novel diagnostic biomarkers for prostate cancer. *Journal of Cancer*. 2010;1:150-177.
- [96] Nikiforova MN, Wald AI, Roy S, et al. Targeted Next-Generation Sequencing Panel (ThyroSeq) for Detection of Mutations in Thyroid Cancer. *The Journal of Clinical Endocrinology and Metabolism*. 2013;98(11):E1852-E1860.
- [97] Vincent MD, Kuruvilla MS, Leighl NB, et al. Biomarkers that currently affect clinical practice: EGFR, ALK, MET, KRAS. *Current Oncology*. 2012;19(Suppl 1):S33-S44.
- [98] Liang Z, Zhang J, Zeng X, et al. Relationship between EGFR expression, copy number and mutation in lung adenocarcinomas. *BMC Cancer*. 2010;10:376.
- [99] HARARI, P. M. Epidermal growth factor receptor inhibition strategies in oncology. *Endocr.Relat Cancer*. 2004;11(4):689-708.
- [100] Yachida S, White CM, Naito Y, et al. Clinical Significance of the Genetic Landscape of Pancreatic Cancer and Implications for Identification of Potential Long Term Survivors. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2012;18(22):6339-6347.
- [101] VAUGHN, C. P., ZOBELL, S. D., FURTADO, L. V., BAKER, C. L. a SAMOWITZ, W. S. Frequency of KRAS, BRAF, and NRAS mutations in colorectal cancer. *Genes Chromosomes.Cancer*. 2011;50(5): 307-312.
- [102] HINODA, Y. KRAS mutation test. *Rinsho Byori*. 2011; 59(6): 598-601.
- [103] Loupakis F, Ruzzo A, Cremolini C, et al. KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *British Journal of Cancer*. 2009;101(4):715-721.
- [104] Dunn EF, Iida M, Myers RA, et al. Dasatinib sensitizes KRAS mutant colorectal tumors to cetuximab. *Oncogene*. 2011;30(5):561-574.
- [105] Misale S, Yaeger R, Hobor S, et al. Emergence of KRAS mutations and acquired resistance to anti EGFR therapy in colorectal cancer. *Nature*. 2012;486(7404):532-536.
- [106] Jones C, Taylor MA, McWilliams B. The role of cetuximab as first-line treatment of colorectal liver metastases. *HPB: The Official Journal of the International Hepato Pancreato Biliary Association*. 2013;15(1):11-17.
- [107] Ng SB, Turner EH, Robertson PD, et al. Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes. *Nature*. 2009;461(7261):272-276.
- [108] Kiezun A, Garimella K, Do R, et al. Exome sequencing and the genetic basis of complex traits. *Nature genetics*. 2012;44(6):623-630.
- [109] Choi, M., Scholl, U. I., Ji, W., et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(45):19096–19101.
- [110] Bainbridge MN, Wang M, Burgess DL, et al. Whole exome capture in solution with 3 Gbp of data. *Genome Biology*. 2010;11(6):R62.

- [111] Wright, A. F.. DNA analysis in human disease. *Journal of Clinical Pathology*. 1986;39(12): 1281–1295.
- [112] Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., ... Bamshad, M. J. (2010). Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics*, 42(1), 30–35.
- [113] Gilissen C, Hoischen A, Brunner HG, et al. Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics*. 2012;20(5):490-497.
- [114] Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Human Molecular Genetics*. 2010;19(R2):R145-R151.
- [115] Zhang S-J, Liu C-J, Yu P, et al. Evolutionary Interrogation of Human Biology in Well-Annotated Genomic Framework of Rhesus Macaque. *Molecular Biology and Evolution*. 2014;31(5):1309-1324.
- [116] Dynabeads® mRNA Purification Kit (for mRNA purification from total RNA preps) [online] [cit. 2015-07-28] Dostupné z: <https://www.lifetechnologies.com/order/catalog/product/61006>
- [117] Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*. 2014;56(2):61-passim.
- [118] Product Information Thermo Scientific RevertAid H Minus Reverse Transcriptase [online] [cit. 2015-07-28] Dostupné z: https://tools.lifetechnologies.com/content/sfs/manuals/MAN0012886_RevertAid_H_Minus_Reverse_Transcript_ep0451_UG.pdf
- [119] TotalScript™ RNA-Seq Kit [online] [cit. 2015-07-28] Dostupné z: <http://www.epibio.com/docs/default-source/protocols/totalscript-rna-seq-kit.pdf?sfvrsn=8>
- [120] Parkhomchuk D, Borodina T, Amstislavskiy V, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*. 2009;37(18):e123.
- [121] TruSeq Stranded Total RNA Library Prep Kit [online] [cit. 2015-07-28] Dostupné z: http://www.illumina.com/products/truseq_stranded_total_rna_library_prep_kit.html

9. Seznam zkratek

AC	Umělý chromozom (z anglického Artificial chromosome)
AML	Akutní myeloidní leukémie
APS	Adenosin fosfosulfát
ASCII	Americký standardní kód pro výměnu informací (z anglického American Standard Code for Information Interchange)
ATP	Adenosintrifosfát
BAC	Bakteriální umělý chromozom (z anglického Bacterial artificial chromosome)
bp	Páry bází (z anglického Base pair)
BWT	Burrowsova Wheelerova transformace
CCD	Charge-coupled device
cDNA	Komplementární deoxyribonukleová kyselina (z anglického Complementary deoxyribonucleic acid)
d	den
DBA	Diamond-Blackfanova anémie (z anglického Diamond-Blackfan anemia)
ddNTP	Dideoxynukleotid
DEPC	Dietyl pyrokarbonát (z anglického Diethylpyrocarbonate)
DMSO	Dimethylsulfoxid (z anglického Dimethyl sulfoxide)
DNA	Deoxyribonukleová kyselina (z anglického Deoxyribonucleic acid)
dNTP	Deoxynukleotid
dsDNA	Dvouvláknová deoxyribonukleová kyselina (z anglického Double-strand deoxyribonucleic acid)
DTT	Dithiothreitol
dTTP	Deoxythymidin trifosfát
dUTP	Deoxyuridine trifosfát
EGFR	Epidermální růstový receptorový faktor (z anglického Epidermal growth factor receptor)

emPCR	Emulzní polymerázová řetězová reakce (z anglického emulsion polymerase chain reaction)
Gb	Giga báze (z anglického Giga base)
GPL	Všeobecná veřejná licence (z anglického General Public License)
h	Hodina
HG38	Lidský genom 38 (z anglického Human Genome 38)
HRAS	z anglického Harvey rat sarcoma
KRAS	z anglického Kirsten rat sarcoma
LBS:	Sekvenace založena na ligaci (z anglického ligation based sequencing)
Mb	Mega báze (z anglického Mega base)
miRNA	mikro RNA (z anglického micro RNA)
MPSS	Masivně paralelní sekvenování (z anglického Massively Parallel Signature Sequencing)
mRNA	Mediátorová ribonukleová kyselina (z anglického Messenger ribonucleic acid)
NEB	New England BioLabs
NGS	Sekvenování nové generace (z anglického Next generation sequencing)
NRAS	z anglického Neuroblastoma rat sarcoma
nt	nukleotidů
Obr.	Obrázek
PCR	Polymerázová řetězová reakce (z anglického Polymerase chain reaction)
PEG	Polyethylenglykol (z anglického Polyethylene-glycol)
pYAC	Plasmid kvasinkového umělého chromozomu
qPCR	Kvantitativní polymerázová řetězová reakce (z anglického Quantitative polymerase chain reaction)
RNA	Ribonukleová kyselina (z anglického Ribonucleic acid)
rRNA	Ribozomální ribonukleová kyselina (z anglického Ribosomal ribonucleic acid)
SBS	Sekvenování na bázi syntézy (z anglického Sequencing by synthesis)

ssDNA	Jednovláknová deoxyribonukleová kyselina (z anglického Single-strand deoxyribonucleic acid)
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
tRNA	Transferová ribonukleová kyselina (z anglického Transfer ribonucleic acid)
UDG	Uracil-DNA-glykosyláza (z anglického Uracil-DNA-Glycosylase)
YAC	Kvasinkový umělý chromozom (z anglického yeast artificial chromosome)