

Univerzita Palackého v Olomouci

Bakalářská práce

Olomouc 2016

Gabriela Fryčová

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra buněčné biologie a genetiky



**Změny v methylačním profilu nádorových buněk
asociované s lékovou rezistencí**

Bakalářská práce

Gabriela Fryčová

Studijní program: Biologie

Studijní obor: Molekulární a buněčná biologie

Forma studia: Prezenční

Olomouc 2016

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

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

V Olomouci dne:

.....

Souhrn

Methylace DNA je epigenetická modifikace, ke které nejčastěji dochází na pátém uhlíku v cyklické struktuře cytosinu. Tato kovalentní úprava DNA je esenciální pro správný vývoj organismu a je důležitou součástí procesu inaktivace chromozomu X, imprintingu a regulace exprese genetické informace.

V této bakalářské práci je popsána stručná charakteristika methylace DNA, její vznik a vliv na expresi genů. Dále jsou v této práci zmíněny a popsány příklady epigeneticky modifikovaných genů, které mají významný podíl na vzniku rakovinných onemocnění. Vzhledem k faktu, že methylační profil buněk lze detekovat pomocí sekvenačních technik, jsou v dalších kapitolách popsány principy a využití sekvenační metody nové generace používané na platformě Illumina a dále jsou vysvětleny vybrané metody detekce methylace DNA a formáty souborů pro analýzu sekvenačních dat.

Experimentální část této práce se zaměřuje na celogenomovou detekci methylačních profilů DNA pomocí metody Methyl-Capture sekvenace na platformě Illumina a následném zpracování výstupních dat. Hlavním cílem práce bylo detekovat a identifikovat methylované oblasti v genomech buněčné linie HCT116 odvozené od kolorektálního karcinomu, které mohou mít určitou roli v navozování rezistence buněk vůči demethylační látce decitabinu.

Summary

DNA methylation is an epigenetic modification, which mainly occurs at the fifth position of carbon in the cyclic structure of cytosine. This covalent modification is essential for regular development of organisms and is an important part of processes such as X-chromosome inactivation, imprinting and regulation of gene expression.

In this thesis there are described basic characteristics of DNA methylation, its formation and effect on gene expression. Further, there are mentioned and described examples of epigenetically modified genes, which have significant role in carcinogenesis. Due to the fact that methylation profile of cells can be detected by sequencing techniques, in the next chapters there is also described method of next generation sequencing used on the platform produced by Illumina, several methods of detection of DNA methylation and chosen format files used for analysing sequencing data.

The experimental part of this work is focused on the whole-genome detection of DNA methylation profiles by Methyl-Capture sequencing on platform created by Illumina and data processing. The main object of this work was to detect and identify methylated regions in genomes of the HCT116 cell line derived from cancerous colorectal cells, which can have an important role in inducing resistance to demethylating drug decitabine.

Poděkování

Děkuji svému školiteli Ing. Rastislavu Slavkovskému, Ph.D. za odborné vedení, věcné připomínky a cenné rady, které mi v průběhu zpracování bakalářské práce věnoval. Dále také děkuji kolektivu Ústavu molekulární a translační medicíny za ochotu vždy mi pomoci. V neposlední řadě patří velké poděkování mé rodině za umožnění studia a podporu během něj.

Obsah

1	Úvod	9
2	Cíl práce	10
3	Epigenetika	11
3.1	Methylace DNA	11
3.1.1	Methylace cytosinu	12
3.1.2	Demethylace DNA	17
3.1.3	Imprinting	17
3.1.4	Methylační vzory a profily buněk	18
3.1.5	Methylace DNA a její role v regulaci exprese genetické informace	19
4	Epigenetické modifikace v karcinogenezi	22
4.1	Epigenetické modifikace ve vybraných typech nádorů	22
4.1.1	Hypermethylace genu <i>VHL</i> v rakovinných buňkách ledvin	22
4.1.2	Hypermethylace genu <i>hML1</i> v rakovinných buňkách tlustého střeva, žaludku a endometria	23
4.1.3	Hypermethylace genu <i>E-cadherin</i> v rakovinných buňkách žaludku a prsu	23
4.1.4	Hypermethylace genu <i>Septin 9</i> v rakovinných buňkách kolorekta	24
4.1.5	Hypermethylace genu <i>CDKN2A</i> v rakovinných buňkách pankreatu a tlustého střeva	24
5	Sekvenování nové generace	25
5.1	Systémy MiSeq a HiSeq od firmy Illumina	26
5.2	Detekce methylačního profilu pomocí NGS na platformě Illumina	27
5.2.1	Celogenomové bisulfitové sekvenování	27
5.2.2	Bisulfitové sekvenování se sníženým zastoupením	27
5.2.3	MeDIP sekvenování	27
5.2.4	Tab sekvenování	28
5.2.5	Methyl Capture sekvenování	28
5.3	Formáty sekvenačních dat	28
5.3.1	FASTQ soubory	28
5.3.2	SAM soubory	29
5.3.3	BAM soubory	29
6	Materiál a přístroje	30
6.1	Biologický materiál	30
6.2	Kity	30
6.3	Chemikálie a použité roztoky	30
6.4	Laboratorní přístroje	31
7	Metodika	32
7.1	Příprava DNA	32
7.2	Fragmentace DNA	33
7.3	Vychytávání fragmentů methylované DNA pomocí MethylMiner Methylated DNA Enrichment Kitu	34
7.3.1	Příprava magnetických kuliček	34
7.3.2	Příprava proteinu MBD-biotin	34
7.3.3	Navázání proteinového komplexu MBD-biotin na magnetické kuličky	34

7.3.4	Inkubace konjugátu MBD-biotin a magnetických kuliček s fragmentovanou DNA	35
7.3.5	Eluce nemetylované DNA	35
7.3.6	Eluce metylované DNA	36
7.3.7	Precipitace ethanolem	36
7.4	Kontrola kvality metylované DNA pomocí qPCR	37
7.4.1	Příprava vzorků pro qPCR	37
7.4.2	Příprava kalibrační řady pro qPCR	37
7.4.3	qPCR	38
7.5	Tagmentace s použitím Nextera XT Sample Prep Kitu	38
7.6	Přečistění DNA na magnetických kuličkách SPRIselect	39
7.7	Příprava vzorků na sekvencování	40
7.8	Sekvenování	41
7.9	Zpracování dat	41
8	Výsledky	42
9	Diskuze	56
10	Závěr	59
11	Zdroje	60
12	Seznam zkratk	68

1 Úvod

Methylace DNA je kovalentní epigenetická modifikace, prostřednictvím které je v eukaryotických buňkách methylová skupina přenesena na pátý uhlík v cyklické struktuře cytosinové báze. K takovým reverzibilním úpravám dochází nejčastěji v palindromatické sekvenci 5'-CpG-3'. Tyto CpG dinukleotidy vytvářejí tzv. CpG ostrůvky, což jsou místa v genomu s vysokým obsahem po sobě jdoucích cytosinových a guaninových bází. Oblasti s velkým počtem CpG dinukleotidů se vyskytují v promotorových sekvencích u více než 50 % všech lidských genů, není tedy překvapivé, že methylace DNA má určitý vliv na aktivitu a genovou expresi. K methylaci DNA dochází v buňkách přirozeně enzymatickou aktivitou DNA methyltransferáz, může však vznikat i v důsledku alkylačních reakcí.

Methylace DNA je také velmi důležitá pro správný vývoj organismu, methylové skupiny připojené na cytosinové báze totiž tvoří tzv. methylační profil buněk, který je u různých typů buněk odlišný a je odrazem tkáňově specifické exprese genů. Tyto vzory methylace DNA vznikají v důsledku střídajících se procesů methylace a demethylace a jsou zakládány již brzy v embryogenezi. Methylace DNA má také velký význam v různých buněčných procesech jako je diferenciací buněk, imprinting nebo také inaktivace chromozomu X.

Je již známo, že zvýšená methylace DNA (hypermethylace) vede k potlačení transkripce genetické informace. Methylové skupiny jsou totiž schopny zabraňovat přístupu transkripčních faktorů k vláknům DNA nebo se na tyto modifikace mohou vázat methyl-vazebné proteiny. Methyl-vazebné proteiny jsou následně schopny interagovat s histonovými deacetylázami a samotnými chromatinovými komplexy a tak udržovat chromatin v kondenzovaném stavu, který je transkripčně inaktivní. Naopak hypomethylace DNA, celogenomové snížení methylace, vede často k aktivaci umlčené genetické informace. Aberantní methylace pak může být nástrojem pro nesprávné vypnutí či zapnutí určitých genů zapojených do různých buněčných procesů a tím může indukovat vznik rakovinných onemocnění.

Methylační profil buněk je možno sledovat různými metodami. V praktické části této bakalářské práce byla použita technika Methyl-Capture sekvenace, jejímž principem je vycytávání vysoce methylovaných oblastí DNA prostřednictvím MBD proteinů navázaných na magnetických kuličkách a následném NGS sekvenováním.

2 Cíl práce

Teoretická část

- Seznámit se s problematikou methylace DNA, zejména v souvislosti s nádorovým bujením.
- Sepsat rešerši o základních epigenetických změnách ve vybraných typech nádorů.
- Sepsat rešerši o možnostech detekce methylomového profilu pomocí sekvenování nové generace (NGS).
- Seznámit se s problematikou sekvenování na platformě Illumina MiSeq/HiSeq a možnostech zpracování dat s vysokým sekvenačním pokrytím.

Praktická část

- Zvládnutí izolace DNA z buněk a její fragmentaci.
- Analýza fragmentované DNA pomocí mikročipové elektroforézy pomocí Agilent Bioanalyzer.
- Zvládnutí přípravy obohacené methylované DNA pomocí principu „MethylMiner“ resp. „methyl capture“.
- Zvládnutí přípravy sekvenační knihovny pomocí kitu Illumina Nextera XT.
- Spolupracovat při sekvenování vzorků obohacené methylované DNA.
- S kooperací se spolupracovníky se seznámit s problematikou zpracování sekvenačních dat z methylomových analýz.

3 Epigenetika

Genetika je obor zabývající se dědičnými změnami v genetické aktivitě nebo abnormalitách ve funkci genů, které jsou způsobené změnami v sekvenci DNA. Takovými aberacemi mohou být bodové mutace, delece, inserce nebo například translokace. Naproti tomu epigenetika studuje dědičné změny ve fenotypové expresi genomu, ke kterým dochází, aniž by byla pozměněna sekvence DNA. Epigenomy pak poskytují informace o vzorech, podle kterých se struktury (například methylové skupiny) vážou na DNA a histonové proteiny, a podávají také informace o interakcích mezi různými částmi chromatinu. Zároveň epigenomy obsahují údaje o regulačních elementech v DNA samotné, jednak o elementech, které leží v promotorových oblastech, ale také o těch, které se nacházejí ve vzdálených sekvencích enhancerů.

3.1 Methylace DNA

DNA methylace je jedinou přirozenou modifikací nukleových bází. Jelikož methylační vzor DNA odpovídá expresi genů, může být DNA methylace použita jako prostředek pro analýzu RNA nebo exprese proteinů v rakovinných buňkách.

K modifikaci DNA methylací dochází u prokaryotických i eukaryotických buněk a jejími produkty jsou nejčastěji N6-methyladenin, N4-methylcytosin a 5-methylcytosin (Laird *et Jaenisch*, 1996). V prokaryotických buňkách je methylován adenin i cytosin a methylace zde vykazuje ochrannou funkci vůči cizí DNA (Wilson *et Murray*, 1991). V případě mnohobuněčných eukaryotických organismů probíhá methylace pouze na cytosinových bázích a tento děj je v buňkách spojen s kondenzovanou strukturou chromatinu a inhibicí genové exprese (Bird *et Wolffe*, 1999). Methylační modifikace DNA u savců byly detekovány ihned po objevení DNA jakožto genetického materiálu (McCarty *et Avery*, 1946).

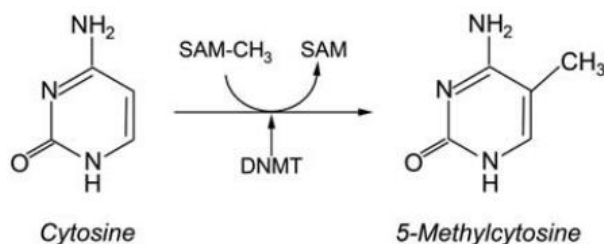
Vědci se domnívají, že jednou z příčin vzniku methylace by mohlo být umlčování repetitivních sekvencí. Přibližně 45 % savčího genomu tvoří pohyblivé a virové elementy, přičemž většina z nich je umlčena methylací nebo mutací 5-methylcytosinu na thymin (Shultz *et al.*, 2006; Walsh *et al.*, 1998). Této umlčovací schopnosti se také využívá v procesech, kde je nutno dosáhnout potlačení transkripce, tedy při imprintingu a inaktivaci chromozomu X.

Methylace DNA byla dlouho považována „pouze“ za hlavní mechanismus represe genové exprese. Další studie ale ukázaly, že správně regulovaná methylace DNA je esenciální pro embryogenezi, diferenciaci buněk a dokonce může zprostředkovávat v určitých případech i expresi genů (Suzuki *et Bird*, 2008).

3.1.1 Methylace cytosinu

Methylace cytosinové báze v palindromické sekvenci 5'-CpG-3' je reverzibilní kovalentní modifikace DNA, kde methylová skupina svou orientací směřuje do velkého žlábků (Jaenisch *et Jähner*, 1984). V genomu savců se vyskytuje přibližně 3×10^7 5-methylcytosinových bází, ze kterých je většina součástí CpG dinukleotidů (Bestor, 2000). Bylo ale také zjištěno, že methylované cytosiny se nacházejí také na místech v sekvencích DNA mimo CpG dinukleotidy, avšak jejich význam a funkce není ještě zcela známá (Woodcock *et al.*, 1987). Lister *et al.* (2009) detekoval v savčích kmenových buňkách velké množství non-CpG methylace a navrhnul, že by tato modifikace mohla být zodpovědná za udržování pluripotence.

Jako první objevil modifikovanou bázi cytosinu v roce 1948 Rollin Hotchkiss, který také předpokládal, že se takto upravený cytosin vyskytuje v DNA zcela přirozeně (Hotchkiss, 1948). Připojení methylové skupiny na cytosinový zbytek může být výsledek dvou odlišných dějů. Prvním z nich je aberantní interakce cytosinových bází s alkylačními činidly, v druhém případě může být methylace DNA přirozeným důsledkem enzymatické reakce, jedná se o tzv. biologickou metylaci (Laird *et Jäenisch*, 1996). Katalyzátory biologické DNA methylace jsou enzymy DNA methyltransferázy (DNMT), které přenášejí methylovou skupinu z S-adenylmethioninu (SAM) na pátý uhlík cytosinu za vzniku 5-methylcytosinu, který tvoří 1 % celé jaderné genetické informace lidské buňky (Moore *et al.*, 2012; Ehrlich *et al.*, 1982). Proces methylace cytosinové báze je znázorněn na **Obrázku 1**.



Obr. 1: Konverze cytosinu na 5-methylcytosin pomocí enzymu DNA methyltransferázy (DNMT) (převzato z: Gibney *et Nolan*, 2010)

3.1.1.1 CpG dinukleotidy

V genomu savců se obecně CpG dinukleotidy vyskytují v malém množství, to je z největší pravděpodobnosti způsobeno velkým potenciálem cytosinu deaminací mutovat na thymin (Coulondre *et al.*, 1978; Bird, 1980). Je také zajímavé, že CpG dinukleotidy jsou po celém lidském genomu rozmístěny asymetricky. Oblasti bohaté na CpG dinukleotidy se

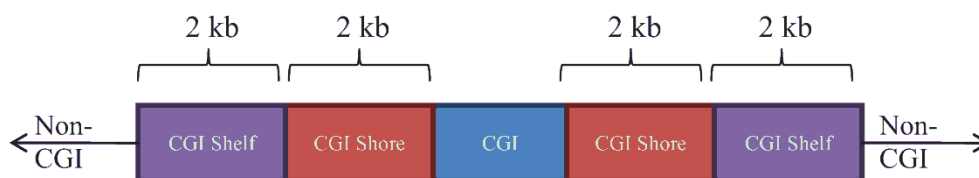
střídají s úseky DNA, které ve svých sekvencích obsahují nízký počet CpG dinukleotidů. Zároveň platí, že ne všechna CpG místa jsou methylována a tyto odlišné methylační vzory jsou pro každý typ tkáně specifické (Bird *et al.*, 1985; Cedar *et Bergman*, 2009).

3.1.1.2 CpG ostrůvky

CpG ostrůvky (CGI) tvoří méně než 1 % celé DNA a jsou preferenčně lokalizovány v promotorových oblastech u více než 50 % všech lidských genů. Nicméně i přes tuto skutečnost bylo zjištěno, že CGI se nevyskytují v oblastech obsahující běžné promotorové elementy jako je například TATA box (Carninci *et al.*, 2006). Vazebných míst pro transkripční faktory, jejichž sekvence jsou bohaté na obsah C a G nukleotidů, je velké množství. Je tedy pravděpodobné, že CGI posilují vazbu transkripčních faktorů na tato iniciační místa (Moore *et al.*, 2012). Navíc promotory vymezující místo transkripčního počátku mohou zahajovat transkripci z několika různých pozic. Naopak promotory bez CGI jsou spojované s jedním iniciačním místem (Sandelin *et al.*, 2007).

Do nedávna se předpokládalo, že místa s transkripčními regulačními oblastmi, ve kterých se objevují CGI, vykazují tkáňově specifické methylační vzory. Další studie ale ukázaly, že CGI asociované s transkripčními iniciačními místy vykazují tkáňově specifickou metylaci velmi vzácně (Maunakea *et al.*, 2010). Místo toho regiony zvané „břehy“ CpG ostrůvků (CpG island shore) lokalizované až 2 kb od samotných CGI, vykazují silně konzervované vzory tkáňově specifické methylace a stejně jako u samotných CGI methylace CpG břehů vysoce koreluje s redukovanou expresí genů. Oblasti DNA vzdálené 2 kb od těchto „břehů“ jsou označovány jako CGI „mělčiny“ (CGI shelves). Jednotlivé zmíněné regiony DNA obsahující vysoký počet CpG dinukleotidů jsou zobrazeny na **Obrázku 2**. Oblasti tkáňově specifických methylovaných břehů se označují jako T-DMR regiony (tissue-specific differentially methylated regions), které se oproti CGI v 96 % případů vyskytují mimo promotory. Jako C-DMR oblasti jsou označovány odlišně methylované oblasti DNA vyskytující se v rakovinných buňkách (Izzarry *et al.*, 2009).

Jak již bylo zmíně, CGI jsou na lidských chromozomech distribuovány nerovnoměrně, kdy na některých chromozomech je počet ostrůvků vyšší než na jiných. Například chromozomy 19 a 22 jsou bohaté na CGI, zatímco naopak na chromozomu 18 bylo detekováno malé množství CGI (Zhu *et al.*, 1993). Tento nepoměr současně odpovídá distribuci genů na jednotlivých chromozomech (Antonarakis, 1994).



Obr. 2: Ilustrace CpG ostrůvku (CGI) s obklopujícími CGI břehy (CGI Shore) a CGI mělčinami (CGI Shelf) (převzato z: Van Dongen *et al.*, 2014)

3.1.1.3 DNA methyltransferázy

Enzymy zodpovědné za metylaci DNA u savců se dělí do dvou kategorií. *De novo* DNA methyltransferázy 3A a 3B (DNMT3A a DNMT3B) katalyzují metylaci na CpG místech, které před tím nebyly tímto způsobem modifikovány. Do druhé skupiny se řadí DNA methyltransferáza 1 (DNMT1), která během replikace kopíruje existující methylační vzory na nově vznikající vlákno. V savčích buňkách se také nachází DNA methyltransferáza 2 (DNMT2), která ale nevykazuje žádnou methylační aktivitu (Okano, 1998a).

DNMT1, DNMT3A a DNMT3B jsou si strukturně velmi podobné. N-konec těchto polypeptidů slouží jako regulační doména, na C-konci se nachází doména s katalytickou aktivitou. I přes tyto podobnosti mají ale všechny tyto tři enzymy v buňce jedinečné funkce a jsou exprimovány odlišně (Yen *et al.*, 1992; Xie *et al.*, 1999).

3.1.1.3.1 DNA methyltransferáza 1

Funkcí DNMT1 je zachování již existujících methylačních vzorů, a proto se tomuto enzymu také říká *udržovací methyltransferáza*. DNMT1 má vysokou afinitu k hemimetylované DNA, což jí umožňuje během replikace katalyzovat metylaci nově syntetizovaného vlákna DNA podle vlákna templátového (Yoder *et al.*, 1997).

DNMT1 je nejintenzivněji transkribována během S fáze buněčného cyklu a je velmi důležitá pro správný zárodečný vývoj a životaschopnost rakovinných buněk. Některé studie ukázaly, že kompletní vypnutí tohoto enzymu způsobuje buněčnou smrt rakovinných buněk kolorektální linie HCT116 (Robertson *et al.*, 2000; Li *et al.*, 1992; Chen *et al.*, 2007). To pravděpodobně souvisí s tím, že DNMT1 je exprimovaná dvěstěkrát více v buňkách kolorektálního karcinomu nežli v normálních buňkách sliznic (El-Deiry *et al.*, 1991). Zároveň je také v buňce přítomná v mnohem větším množství než ostatní DNMT (Bestor, 2000).

3.1.1.3.2 DNA methyltransferázy 3

Proteiny DNMT3A a DNMT3B jsou schopné katalyzovat metylaci CpG dinukleotidů jak metylované, tak i nemetylované DNA, jedná se o tzv. *de novo* metylaci (Okano *et al.*, 1998b). Je tedy logické, že geny pro tyto enzymy vykazují vyšší expresi v kmenových buňkách, ze kterých diferenciací vznikají buňky s odlišným vzorem exprese genetické informace, zatímco v somatických buňkách jsou exprimovány mnohem slaběji (Okano *et al.*, 1999). Oba enzymy, DNMT3A a DNMT3B, vykazují podobné kinetické parametry a jsou tedy schopny se navzájem do jisté míry nahrazovat. I přes tuto skutečnost je načasování jejich exprese a tkáňová specifita u obou proteinů různá (Aoki *et al.*, 2001; Suetake *et al.*, 2003; Watanabe *et al.*, 2002). DNMT3A je exprimována ve všech buňkách, zatímco k expresi DNMT3B v diferenciovaných buňkách s výjimkou štítné žlázy a kostní dřeně dochází zřídka (Xie *et al.*, 1999).

Do proteinové rodiny DNA methyltransferáz 3 také patří protein DNMT3L (DNA (cytosine-5)-methyltransferase 3-like), který je podobný DNMT3ázám v obou N- a C-koncových oblastech, ale sám nevykazuje žádnou enzymatickou funkci. Přesto fyzicky interaguje s DNMT3A a DNMT3B a pomáhá těmto enzymům regulovat jejich katalytickou aktivitu (Suetake *et al.*, 2004). Protein DNMT3L vlastní PHD homeodoménu, kterou mají ve své struktuře i DNMT3A a DNMT3B (Aapola *et al.*, 2000) a komplex DNMT3L-DNMT3A-DNMT3B je tak pomocí této PHD domény schopen interagovat s proteiny, transkripčními faktory, histonovou deacetylázou a histonovou methyltransferázou (Fuks *et al.*, 2000; Fuks *et al.*, 2003). K expresi DNMT3L dochází nejvíce během zárodečného vývoje a v dospělosti se tento protein objevuje pouze v buňkách brzlíku (Aapola *et al.*, 2000).

Methylace je proces, který probíhá vždy pouze na jednom komplementárním vlákně DNA, proto může být *de novo* methylace, která vede k modifikaci obou vláken, považována za proces se dvěma po sobě navazujícími jedno-řetězcovými reakcemi (Cheng, 1995).

3.1.1.3.3 O6-methylguanin DNA methyltransferáza

Některé druhy methyltransferáz se podílejí na opravě míst ve struktuře DNA, které byly chemicky alkylovány. Jedním takovým enzymem je O6-methylguanin DNA methyltransferáza (MGMT), které se také nazývá *opravná DNA methyltransferáza*. MGMT katalyzuje opravu O6-methylguaninu zpět na guanin, přičemž dochází k přesunu methylové skupiny z cyklické struktury guaninové báze na molekulu cysteinu uvnitř MGMT (Olsson *et al.*, 1980). Bylo prokázáno, že tento enzym se v důsledku chyb v jeho aktivitě podílí na vzniku rakoviny (Laird *et al.*, 1996). Studie z roku 2002, ve které byl sledován vliv

represe transkripce *MGMT* genu, ukázala, že hypermethylovaní promotorové oblasti pro *MGMT* dochází k potlačení exprese toho genu, což mělo za následek vznik mikrosatelitové nestability v rakovinných buňkách žaludku (Bae *et al.*, 2002).

3.1.1.4 Methyl-vazebné proteiny

Cytosinové báze modifikované methylovaní jsou rozpoznávány methyl-vazebnými proteiny (MBP), které se třídí do třech proteinových rodin – MBD, UHRF a Zinc finger, které se na tyto místa vážou a následně umožňují nebo naopak zabraňují různým proteinům v přístupu k vláknům DNA (Moore *et al.*, 2012).

3.1.1.4.1 Proteinová rodina MBD

Do rodiny MBD (methyl-binding domain) patří proteiny MeCP1, MeCP2, MBD1, MBD2, MBD3 a MBD4. Proteiny MBD1, MBD2, MeCP1 a MeCP2 (methyl-CpG-binding domain protein) vytváří vazbu přímo s methylovanou DNA a ve své struktuře obsahují TRD doménu (transcriptional repression domain), která umožňuje interagovat s různými represorovými komplexy (Nan *et al.*, 1998; Ng *et al.*, 1999; Sarraf *et al.*, 2004). Příklady represorových komplexů vázajících se na methylovaná místa DNA, které způsobují potlačení transkripce, jsou uvedeny níže v kapitole Vazba methyl-vazebných proteinů.

MBD1, MBD2, MBD4 a MeCP2 rozeznávají na vláknech DNA místa s CpG dinukleotidy, se kterými mohou následně interagovat. Pro tuto vazbu, při které dojde k pokrytí oblasti o velikosti přibližně 12 bp a zabránění přístupu dalším MBD proteinům, je v sekvenci DNA nutná přítomnost aspoň 1 methylovaného CpG dinukleotidu (Nan *et al.*, 1993). Ve struktuře proteinu MBD3 se vyskytují určité substituce aminokyselin, které této vazbě na CpG dinukleotidy zabraňují (Prokhortchouk *et al.*, 2001).

3.1.1.4.2 Proteinové rodiny UHRF a Zinc finger

UHRF1 a UHRF2 (ubiquitin-like with PHD and ring finger domains) jsou multidoménové proteiny, které vytvářejí vazbu s cytosinovými bázemi pomocí jejich DNA vazebných SET a RING domén (Hashimoto *et al.*, 2008; Hashimoto *et al.*, 2009). Přesto UHRF proteiny primárně jako DNA-vazebné proteiny a represory transkripce nefungují (Bostick *et al.*, 2007).

Proteiny patřící do rodiny proteinů s zinc finger motivem se na DNA vážou prostřednictvím zinc-finger domény a do této skupiny patří proteiny Kaiso, ZBTB4 a ZBTB38 (Zinc Finger And BTB Domain Containing) (Prokhortchouk *et al.*, 2001; Filion *et al.*, 2006). Kaiso se na rozdíl od ostatních proteinů vázajících se na methylovou skupinu

cytosinu preferenčně váže na 2 po sobě jdoucí methylované CpG dinukleotidy (Daniel *et al.*, 2002).

3.1.2 Demethylace DNA

Demethylace DNA je charakterizována jako pasivní nebo aktivní děj, přičemž k pasivní demethylaci dochází v dělicích se buňkách. Jelikož enzym DNMT1 aktivně udržuje methylační profil dělicích se buněk během replikace DNA, pak jeho inhibice nebo dysfunkce umožňuje zachování nově zabudovaného cytosinu zůstat v nemethylovaném stavu a tudíž snižuje celkový stupeň methylace v každém následujícím dělení. Aktivní DNA demethylace může probíhat jak v dělicích se buňkách tak i v buňkách nedělicích, tento proces ale vyžaduje enzymatické reakce k navrácení methylovaného cytosinu do jeho nemethylovaného stavu (Mayer *et al.*, 2000; Oswald *et al.*, 2000; Parousch *et al.*, 1990; Zhang *et al.*, 2007).

Aktivní demethylace probíhá prostřednictvím série chemických reakcí, které vedou k dalším modifikacím 5-methylcytosinu a to pomocí deaminačních nebo oxidačních reakcí. Deaminace DNA je zprostředkována enzymy deaminázami, které katalyzují odštěpení aminové skupiny z řetězce nukleotidové báze (Morgan *et al.*, 2004). Dochází tak k neshodě kompatibility protilehlých vláken DNA a musí dojít v této oblasti k opravě, jejímž výsledkem je demethylace. Takový produkt deaminace je pak rozeznáván opravným systémem BER (base excision repair), který nahrazuje pozměněnou bázi za nemodifikovaný cytosin (Bhutani *et al.*, 2011).

Mezi enzymy, které aktivně demethylují DNA, patří také 5-methylcytosin glykosyláza, která odstraňuje methylovaný cytosin z DNA a nechává deoxyribózu neporušenou (Fremont *et al.*, 1997). Dalším takovým enzymem je protein MBD2B, který postrádá glykosylázovou nebo nukleázovou aktivitu a účastní se demethylace DNA prostřednictvím hydrolýzy 5-methylcytosinu za vzniku cytosinu a methanolu (Ng *et al.*, 1999).

3.1.3 Imprinting

V oplodněném vajíčku se vyskytují vždy dvě kopie každého chromozomu, tedy dvě kopie od každého genu, jedna od otce a druhá od matky. U většiny genů funguje exprese obou alel rovnocenně, v genomu se ale vyskytují i tzv. imprintované geny, které mají během vývoje diferenciatní a komplementární funkci.

Imprinting je proces, při kterém je jedna alela v páru vypnuta. Jeho vlivem se tedy schémata exprese genů od Mendelovského vzoru značně liší. O tom, jestli budou jednotlivé

geny inaktivní nebo ne, rozhoduje jejich původ, tedy to po kom byly zděděny. Konkrétní gen je exprimován v případě, že pochází od „správného“ rodiče, a pokud je děděn po druhém z předků, pak je trvale inaktivní. V případě, že je gen ležící na mateřském imprintovaném lokusu exprimován, pak je tato imprintovaná kopie genu v buňkách potomků vypnuta, zatímco exprese kopie genu zděděného po otci probíhá normálně. Stejný princip exprese platí i pro geny imprintované na lokusech otce (Phillips *et Lobo*, 2008).

Obrovskou roli v rozlišování mezi imprintovanými geny děděnými po otci a po matce má methylace DNA, která představuje tzv. genetickou paměť buňky. Jak již bylo zmíněno, některé geny obsahují ve svých sekvencích úseky s CpG dinukleotidy, které mohou být pomocí DNA methyltransferáz modifikovány. Tato methylace, která probíhá ihned po replikaci, se pak epigeneticky dědí na potomky. Imprintované geny jsou tedy „označeny“ prostřednictvím methylačních značek specifických pro genomy rodičů a toto označení umožňuje regulovat zhuštění chromatinu pomocí specifických proteinů, které rozeznávají 5-methylcytosinové báze. Tyto modifikace mohou být během gametogeneze odstraněny a následně znovuobnoveny tak, aby odpovídaly pohlaví rodiče, který tyto gamety produkuje (Ariel *et al.*, 1995; Phillips *et Lobo*, 2008).

Narušení normálního imprintingu může vést k bialelické expresi imprintovaných genů a může mít za následek vznik genetických abnormalit nebo rakoviny (Reik, 1989; Brenton *et al.*, 1995).

3.1.4 Methylační vzory a profily buněk

Methylační vzory vznikají v důsledku dynamických procesů methylace a demethylace DNA (Yoder *et al.*, 1997). Methylační profily buněk u jedinců jednoho druhu jsou v korespondujících typech buněk stejné, naopak tyto vzory jsou odlišné v různých typech buněk jednoho organismu. Jednotlivé methylační strategie jsou založeny během gametogeneze (Razin *et Riggs*, 1980) a k jejich změnám u savců dochází během vývoje (Monk *et al.*, 1987).

V průběhu embryogeneze, v období před nidací, dochází ke globální demethylaci DNA a zániku methylačního vzoru obou gamet. Následně po zahájení vajíčka do děložní sliznice vzniká *de novo* methylací nový methylační profil buňky (Jaenisch *et Jähner*, 1984). Tato celogenomová methylace probíhá také během gametogeneze obou samčích i samičích gamet a podílí se na ustanovení imprintingu. K replikaci těchto methylačních vzorů dochází v S fázi buněčného cyklu (Wiegler *et al.*, 1981).

Methylační vzory DNA v somatických buňkách jsou všeobecně považovány za relativně stabilní a nedochází u nich k velkým změnám. Naproti tomu methylace histonových proteinů se může během buněčného cyklu velmi měnit. Je prokázáno, že i přes tyto odlišnosti jsou methylační mechanismy DNA a histonů navzájem propojeny. Oba procesy během replikace spolupracují na předávání určitých specifických methylačních vzorů z rodičovských buněk na potomstvo (Sarraf *et Stancheva*, 2004).

3.1.5 Methylace DNA a její role v regulaci exprese genetické informace

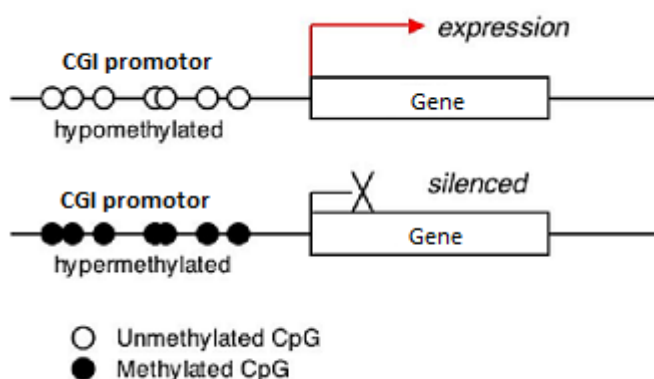
S první hypotézou, podle které methylace a demethylace DNA ovládá expresi genů během vývoje, přišli v roce 1975 pánové Riggs, Holliday a Pugh (Holliday *et Pugh*, 1975; Riggs, 1975). Jak se domnívali, exprese tkáňově specifických genů je založena na reverzibilní methylaci a demethylaci určitých sekvencí DNA, která reguluje interakci regulačních sekvencí s transkripčními faktory (Nan *et al.*, 1998).

Regulace exprese eukaryotických genů je komplexní proces. Iniclace transkripce je vysoce kontrolovaná a jednotná událost, do které jsou zapojeny cis-působící a trans-působící faktory. Cis-působícími faktory jsou sekvence DNA, které souží jako substrát pro faktory trans-působící.

Zvýšená aberantní methylace, hypermethylace, promotorových oblastí genů, je spojena s abnormální inhibicí transkripce těchto sekvencí, zatímco methylace v transkribované oblasti má na expresi daného genu variabilní účinek. Mezi geny, u kterých byla zjištěna hypermethylace promotorů, patří tumor supresorové geny, geny pro regulaci buněčného cyklu a geny podílející se na opravě párování bází (Jones, 1999; Singal *et al.*, 2002). Zároveň platí, že geny s omezeným vzorem exprese jsou pravděpodobně více náchylné k aberantní methylaci, nežli je tomu u silně exprimovaných genů (Feltus *et al.*, 2003).

Hypomethylace DNA, celogenomová redukce methylace cytosinových bází, je epigenetická aberace, která se podobně jako hypermethylace shromažďuje ve stárnoucích buňkách a je pravděpodobně zapříčiněna postupným nárůstem chyb během methylace DNA. Nejčastěji k ní dochází na CpG dinukleotidech uvnitř mobilních a repetitivních sekvencí (satelitní DNA, retrotranspozóny a sekvence LINE). Bylo zjištěno, že demethylace repetitivních sekvencí v pericentrické oblasti podněcuje vznik chyb v rekombinaci DNA a nepřesnou duplikaci DNA (Suzuki *et al.*, 2006). Demethylace mobilních elementů pak může vést ke změně v organizaci genomu, zvýšení frekvence zlomů vláken DNA a tím k ovlivnění chromozomální stability, což je jedním z rysů objevujících se v rakovinných buňkách (Vaiopoulos *et al.*, 2014). Regulace

exprese genetické informace prostřednictvím změny v methylaci CGI promotorové sekvenci DNA je schematicky znázorněna na **Obrázku 3**.



Obr. 3: Regulace exprese genů prostřednictvím hypermethylace a hypomethylace CGI promotorových oblastí (upraveno podle: Horii *et* Hat, 2013)

K inhibici exprese genů ovlivněné methylací DNA může docházet dvěma způsoby. Prvním je umlčování genů v důsledku inhibice interakcí DNA-vazebných faktorů s příslušnými modifikovanými sekvencemi DNA (Watt *et* Molloy, 1988), nebo mohou methyl-vazebné proteiny (MBP) spolupracovat s transkripčními korepresory a tak přímo zastavit genovou expresi (Nan *et al.*, 1998; Nan *et al.*, 1997).

3.1.5.1 Interference s transkripčními faktory

Jak již bylo zmíněno, methylové skupiny na vláknech DNA mohou způsobovat inhibici transkripci genů. K tomu může docházet rovněž prostřednictvím redukce vazebné afinity transkripčních faktorů vůči DNA a interferencí s iniciací transkripce (Clark *et al.*, 1997; Prendergast *et* Ziff, 1991). Počet transkripčních faktorů, které jsou methylací cytosinových bází ovlivněny, je limitován pouze pro podskupinu faktorů, u kterých se v místech jejich vazby na DNA nacházejí CpG dinukleotidy. I přesto methylace ovlivňuje širokou škálu různých promotorových oblastí (Jones *et* Laird, 1999).

Významnou roli v regulaci vlivu methylace DNA na expresi genů má rovněž vlastní síla promotoru příslušného genu. Slabý promotor je možno umlčet modifikací několika CpG dinukleotidů, zatímco silné promotory by taková methylace potlačit nedokázala (Boyes *et* Bird, 1992). K umlčení genové exprese může docházet i prostřednictvím methylace cytosinových bází ležících mimo promotory (Hug, 1996). Represe samotná je ale mnohem silnější, když jsou CpG dinukleotidové sekvence upraveny přímo v oblasti promotoru (Nan, 1997).

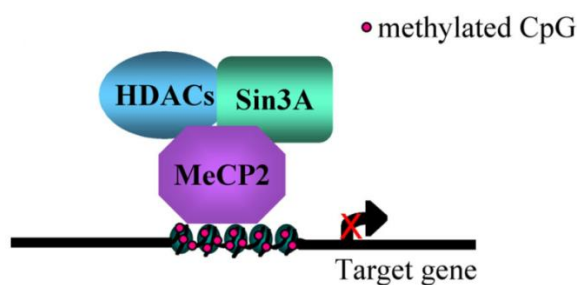
3.1.5.2 Vazba methyl-vazebných proteinů

Regulace genetické exprese může být také založena na vztahu methylace DNA s methyl-vazebnými proteiny. Tyto MBP proteiny jsou schopny se vázat na methylové skupiny a interagovat s histonovou deacetylázou a samotnými chromatinovými komplexy, čímž dochází k stabilizaci kondenzovaného chromatinu a potlačení exprese dané oblasti DNA (Newell – Price *et al.*, 2000).

Jedním takovým methyl-vazebným proteinem indukujícím potlačení transkripce je například protein MeCP2, který na sebe váže multiproteinový potlačující komplex, korepresor mSin3A a histonové deacetylázy HDAC1 a HDAC2 (Pazin *et Kadonoga*, 1997). Tyto enzymy katalyzují odštěpení acetylové skupiny z lyzinových zbytků na histonech H3 a H4. Interakce mezi pozitivně nabitými lysiny s DNA omezuje pohyblivost nukleozomu na DNA, a tak způsobuje nepřístupnost promotoru k transkripčním mechanismům (Ura *et al.*, 1997). Vazba represorového komplexu Sin3A a HDAC proteinů na MeCP2 způsobující potlačení exprese genetické informace je znázorněna na **Obrázku 4**.

K reaktivaci exprese genů je nutná demethylace odpovídajících sekvencí DNA a represivní komplex musí být nahrazen komplexem aktivačním, který naopak katalyzuje acetylaci histonů H3 a H4. Tím dochází k formaci struktury chromatinu, která je pro transkripční faktory přístupná (Wolffe, 1997).

Inaktivace genů, která je způsobena methylací CGI, může být také zprostředkována proteinem MeCP1 (Boyes *et Bird*, 1991). Represorový multiproteinový komplex je tvořen methyl-vazebným proteinem MBD2 (nebo MBD3) a korepresorovým komplexem NuRD, jehož součástí je histonová deacetyláza HDAC1 a HDAC2, protein Mi-2, který se podílí na přestavbách struktury chromatinu, proteiny RbAp46 a RbAp48 (Retinoblastoma-associated proteins 46 a 48) vázající se na histony a protein MTA (Metastasis-associated protein) (Wade *et al.*, 1999).



Obr. 4: Potlačení genové exprese indukované vazbou HDAC-Sin3A komplexu na methylovanou oblast DNA (převzato z: Regione *et al.*, 2012).

4 Epigenetické modifikace v karcinogenezi

Jako rakovina je označována skupina onemocnění, které mohou vznikat z buněk odlišných tkání. Různé typy nádorů se od sebe mohou lišit v rychlosti jejich progresu, některé rostou velmi agresivně, u některých proces karcinogeneze probíhá pomaleji. Jelikož je každý typ nádorového onemocnění jedinečný, musí být léčba jednotlivých typů také odlišná. U některých je možno dosáhnout úplného zastavení rakovinného bujení, u jiných případech se to zatím nedaří (Pecorino, 2002).

Nádorová onemocnění se dělí do dvou skupin, na maligní a benigní. O maligní nádory se jedná tehdy, když se z epicentra nádoru uvolňují buňky a napadají okolní tkáň. V případě benigních nádorů tyto buňky okolní tkáň nenapadají. Pojem *metastazování* označuje situaci, kdy se buňky maligního nádoru rozšiřují do dalších míst v těle a tvoří tzv. sekundární nádory. V obou případech, maligních i benigních nádorů, prokazatelně dochází k poruše v systému kontroly buněčného dělení, což je následek genetických a epigenetických změn v daných buňkách organismu (Pecorino, 2002).

Vznik nádorových onemocnění je iniciován mutacemi nebo epigenetickými aberacemi kritických genů. Tyto mutace jsou příčinou poškození biochemických pochodů a vedou k deregulaci proliferace buněk. Pokud není buněčné dělení regulované, pak se buňky dělí nepřetržitě, rostou jedna přes druhou a tvoří nádor. Existují dvě skupiny genů, které svou mutací nebo modifikací mohou přispívat k vývoji rakoviny. Mutované geny I. typu aktivně podporují buněčný cyklus a označují se jako *onkogeny* (z řeckého slova *ónkos* - „nádor“). Mutované geny II. typu naopak nedokážou potlačit buněčný cyklus a nazýváme je *tumor supresorové geny* (Pecorino, 2002).

4.1 Epigenetické modifikace ve vybraných typech nádorů

4.1.1 Hypermethylace genu *VHL* v rakovinných buňkách ledvin

Mutace *VHL* genu (*von Hippel-Lindau gene*, tumor supresorový gen) v zárodečných buňkách je jednou z příčin vzniku rakoviny ledvin (renální karcinom) a hemangioblastomů, mutace tohoto genu v somatických buňkách jsou zase jedním z rysů sporadické formy renálního karcinomu. Studie sporadických typů rakoviny ledvin navíc ukázaly, že v 20 % případů karcinomů způsobených aberacemi *VHL* genu byla detekována hypermethylace příslušného promotoru (Herman *et al.*, 1994). Zároveň nebyly v těchto vzorcích DNA nalezeny žádné mutace v kódujících oblastech a *VHL* gen byl následně reaktivován působením demethylační látky 5-aza-2-deoxycytidinu (AzaC, decitabin) (Prowse *et al.*, 1997).

4.1.2 Hypermethylace genu *hMLH1* v rakovinných buňkách tlustého střeva, žaludku a endometria

Gen *hMLH1* (*mutL homolog 1*), nacházející se na chromozomu 3p, kóduje protein, který je důležitý pro opravy párování bází v DNA (DNA mismatch repair). K jeho mutaci dochází v zárodečných buňkách velkého počtu jedinců trpících dědičnou formou nepolypózní rakoviny tlustého střeva a je také inaktivován v přibližně 15-25 % případů sporadického karcinomu tlustého střeva (Tycko, 2000). Tyto karcinomy vykazují vysokou frekvenci chyb replikace repetitivní mikrosatelitové sekvence DNA, což je často popisováno jako mikrosatelitová nestabilita. Údaje z několika studií naznačují, že u více než poloviny případů sporadických karcinomů tlustého střeva s mikrosatelitovou nestabilitou nebyla detekována mutace *hMLH1* genu. Naproti tomu byla detekována hypermethylace promotorové oblasti tohoto genu (Herman *et al.*, 1998). Tato aberantní modifikace byla také mnohokrát zaznamenána v řídicích oblastech *hMLH1* genu v rakovinných buňkách žaludku a endometria. Reparační aktivita genu *hMLH1* byla reaktivována prostřednictvím působení demethylační látky AzaC, čímž bylo dokázáno, že umlčení exprese tohoto genu bylo způsobeno hypermethylací (Tycko, 2000).

4.1.3 Hypermethylace genu *E-cadherin* v rakovinných buňkách žaludku a prsu

E-cadherin kóduje protein, který zprostředkovává homotypickou adhezi buněk a podílí se na udržování epiteliální morfologie. Tento polypeptid se také může podílet na kontrole buněčného růstu a diferencí pomocí vlastní interakce s β -kateninem, multifunkčním cytoplasmatickým proteinem (Tycko, 2000). Absence nebo chybné umístění *E-cadherinu* je často zaznamenávána v buňkách určitých typů nekohezních a slabě diferenciovaných karcinomů, zejména v rozptýlených buňkách rakoviny žaludku nebo v lobulárních karcinomech prsu. Asi u poloviny případů těchto typů rakovin žaludku a prsu se vyskytuje inaktivující mutace *E-cadherinu* (Tycko, 2000). Alternativní cestou inaktivace tohoto genu je prostřednictvím promotorové hypermethylace, ke které velmi často dochází ve sporadických typech rakoviny žaludku (Suzuki *et al.*, 1999), ale také v rakovinných buňkách prsu, prostaty a tlustého střeva. Bylo prokázáno, že působením demethylační látky AzaC docházelo k odstranění methylové skupiny na promotorových sekvencích *E-cadherinu* a tím byla umožněna jeho reaktivace (Yoshiura *et al.*, 1995, Graff *et al.*, 1995).

4.1.4 Hypermethylace genu *Septin 9* v rakovinných buňkách kolorekta

Septin 9 (*SEPT9*) je gen kódující protein septin 9, který je členem skupiny proteinů zvané septiny. Tyto proteiny mají důležitou roli při cytokinezi, tedy dělení mateřské cytoplasmy do dvou buněk dceřiných. Protein septin 9 také vykazuje tumor supresorovou aktivitu a pomáhá tedy regulovat rychlost dělení buněk a brání jim v přechodu do nekontrolovatelného stavu. (National Library of Medicine, 2016). Exprese genu *SEPT9* bývá abnormálně potlačena v rakovinných buňkách tlustého střeva a konečníku (CRC, rakovina kolorekta), k čemuž dochází brzy v průběhu karcinogeneze v důsledku methylace promotorové oblasti. Byla vyvinuta metoda detekce přítomnosti tohoto biomarkeru v krvi, která je již komerčně dostupná a využívána v klinické praxi. Princip Epi proColon testu je založen na třech PCR reakcích, které z krevního séra zjišťují přítomnost promotoru *SEPT9* v methylovaném stavu. Dalšími podobnými testy pro detekci modifikovaného *septinu 9* jsou například ColoVantage a Abbott RealTime mS9 (Vaiopoulos *et al.*, 2014; Pentagen, n.d).

4.1.5 Hypermethylace genu *CDKN2A* v rakovinných buňkách pankreatu a tlustého střeva

Gen *CDKN2A* (*cyclin-dependent kinase inhibitor 2A*), kódující tumor supresorový protein p16, je na chromozomu 9p umístěn v oblasti, která je často podrobena delecím nebo ztrátě heterozygoty v mnoha typech rakovinných onemocnění vyskytujících se u člověka. Produktem exprese tohoto genu je inhibitor cyklin dependentní kinázy, který omezuje průběh buněčného cyklu tím, že zabraňuje fosforylaci a inaktivaci Rb proteinů (Tycko, 2000). K expresi genu *CDKN2A* dochází v místě komplexním lokusu s alternativním promotorem a různým čtecím rámcem, jejichž prostřednictvím dochází taktéž ke vzniku proteinu p19^{ARF}, což je pozitivní regulátor tumor supresoru *TP53*. Gen *CDKN2A* zároveň leží v bezprostřední blízkosti sekvence DNA pro protein p15, což je další inhibitor enzymů CDK (Tycko, 2000). Gen *CDKN2A* se v rakovinných buňkách často vyskytuje v inaktivovaném stavu v důsledku hypermethylace odpovídající sekvence DNA (Tycko, 2000). Tato modifikace promotoru pro *CDKN2A* byla zaznamenána například v rakovinných buňkách tlustého střeva (Myohanen *et al.*, 1998). V další studii byla detekována hypermethylace pouze jedné alely genu *CDKN2A* v rakovinných buňkách pankreatu a zároveň nebyla nalezena žádná mutace v kódujících sekvencích. K deregulaci exprese *CDKN2A* methylací ale také dochází v rakovinných buňkách plic, močového měchýře, jater nebo děložního čípku (Baylin *et al.*, 1998) a bylo dokonce zjištěno, že tato hypermethylace *CDKN2A* se objevuje i v bronchiálních lézích vyskytujících se u kuřáků (Belinsky *et al.*, 1998).

5 Sekvenování nové generace

Sekvenování nukleových kyselin je metoda využívána pro určování přesného pořadí nukleotidů v molekule DNA nebo RNA. První sekvenování, známé také jako Sangerovo sekvenování, bylo vyvinuto v roce 1975 Edwardem Sangerem (Sanger *et* Coulson, 1975) a stalo se pilířem všech sekvenování v následujícím čtvrtstoletí.

V roce 2005 bylo provedeno první masivně paralelní sekvenování DNA, které odstartovalo novou éru vysoce výkonných analýz dnes označovaných jako sekvenování nové generace (NGS). Všechny NGS platformy spolu sdílejí jeden společný technologický rys – masivně paralelní sekvenování amplifikovaných nebo jednotlivých molekul DNA rozdělených v průtokové reakční komůrce (flow cell), přičemž toto vysokovýkonné sekvenování je schopno produkovat stovky megabází až gigabází nukleotidových sekvencí v průběhu jednoho sekvenačního běhu (run) podle typu platformy, na kterém je analýza prováděna.

Princip NGS na platformě Illumina je založen na syntéze různě dlouhých fragmentů nukleotidových řetězců za pomoci reverzibilního terminátoru. Jednotlivé báze jsou pak detekovány v takové pozici, v jaké byly začleněny do prodlužovaného vlákna DNA. Terminátor, který je fluorescenčně označen, je vizualizován po každém přidání modifikovaných dNTP. Následně dochází k odstěpení této značky a terminátorové skupiny a tím je pak umožněno začlenění další báze do syntetizovaného řetězce DNA. Odlišností od běžného sekvenování je, že při NGS dochází k paralelní inkorporaci dNTP až na milionech fragmentů DNA. Pracovní postup NGS na platformě Illumina je rozdělen do čtyř základních kroků – příprava knihovny, vytváření klastrů, sekvenování a analýza dat (Illumina, 2016a).

Sekvenování nové generace je možno využívat pro široké spektrum metod, které analyzují genom, transkriptom nebo epigenom jakéhokoliv organismu. Tyto metody se rozlišují primárně podle způsobu opatření daného vzorku a podle zvolené analýzy dat. Mezi metody analýz genomu patří celogenomové sekvenování, sekvenování exomu, *de novo* sekvenování a cílené sekvenování. Metodami studia transkriptomu je celkové sekvenování RNA a mRNA, cílené RNA sekvenování a sekvenování malých a nekódujících RNA. Epigenetické vlastnosti genomu lze studovat například celogenomovým bisulfitovým sekvenováním (WGSB) a bisulfitovým sekvenováním se sníženým zastoupením (RRBS), ChIP sekvenováním nebo profilováním ribozomu (Illumina, 2016a).

NGS je mnohem levnější a výkonnější alternativa tradičního Sangerova sekvenování se spoustou výhod, jako je například schopnost sekvenování malého genomu v průběhu

jediného dne. Celogenomové a exomové NGS sekvenování umožňuje objevovat geny a případně také regulační elementy, jejichž změny vedou ke vzniku nemocí. Cílené sekvenování je jedním ze způsobů identifikace mutací, které taktéž způsobují vznik chorob, což je využíváno při diagnostice patologických stavů organismu (Grada *et* Weinbrecht, 2013). RNA sekvenování poskytuje informace o celém transkriptomu a pomocí těchto dat je uživatel schopen popsat aktuální biologické procesy. Transkripční profil je velmi variabilní a citlivý na vnější podněty. Pokud jsou buňky vystaveny účinku určité chemické látky, pak tyto buňky ihned reagují na tento podnět změnou v expresi genů (Slavkovský, 2016).

5.1 Systémy MiSeq a HiSeq od firmy Illumina

MiSeq sekvenátor

Systém MiSeq je platforma vyrobená firmou Illumina, která umožňuje nejjednodušším způsobem NGS analyzovat vygenerovaná data už za 8 hodin od začátku sekvenování. Tento sekvenátor NGS je jediný, který dovoluje integraci amplifikace, sekvenování a analýzy dat v jednom přístroji. Zároveň svými rozměry nevyžaduje ke své funkci velké laboratorní prostory. MiSeq využívá sekvenační chemii, která vytváří ideální podmínky pro rychlou a relativně levnou analýzu genetických vzorků a pro širokou škálu aplikací (Divoký, 2016a).

Tato technologie je využívána pro analýzu malých genomů, ampliconů a cíleného sekvenování panelů genů, přičemž v průběhu každého sekvenačního běhu lze provést až 25 milionů čtení (reads), která umožňují produkci až 15 Gb dat (Divoký, 2016a).

HiSeq sekvenátory

Systém HiSeq 2500 je stejně jako platforma MiSeq produktem firmy Illumina. Tento sekvenátor využívá stejné technologie, avšak oproti MiSeq systému nabízí svým uživatelům dva režimy sekvenování, High Output Mode a Rapid Run Mode. Další výhodou tohoto přístroje je jeho schopnost současně pracovat ve dvou průtokových reakčních komůrkách (flow cell), čímž umožňuje velmi flexibilní analýzy se širokou škálou aplikací a tím také rozsáhlejší studie sekvenovaných dat (například celogenomové sekvenování, cílené sekvenování, *de novo* sekvenování a analýzy epigenomu) (Divoký, 2016b).

Režim Rapid Run je ideální pro rychlé zpracování omezeného počtu vzorků s podporou delšího čtení (až 2 x 250 bp). Naproti tomu režim High Output je přijatelnější pro rozsáhlejší studie, nebo tehdy, když je požadována větší hloubka pokrytí. Tento režim je zároveň schopen zpracovat šestkrát více vzorků než režim Rapid Run (Divoký, 2016b).

HiSeq sekvenátory 3000 (1 průtoková reakční komůrka) a 4000 (2 průtokové reakční komůrky) jsou rychlejší a výkonnější verzí Systému HiSeq 2500. Na těchto platformách lze

pomocí 5 miliard čtení vygenerovat až 750 Gb/1500 Gb dat, přičemž je čteno více než 75 % sekvenovaných bází s hodnotou kvality nad Q30 (Divoký, 2016b). Tímto způsobem lze sekvenovat až 12 lidských genomů (Illumina, 2016b).

Úplně nejnovější a nejvýkonnější série sekvenátorů od firmy Illumina je řada zvaná HiSeq X, která umožňuje provést 6 miliard sekvenčních čtení a vyprodukovat tak data o velikost až 1800 Gb (Illumina, 2016c).

5.2 Detekce methylačního profilu pomocí NGS na platformě Illumina

5.2.1 Celogenomové bisulfitové sekvenování

Principem detekce methylovaných cytosinových bází metodou celogenomového bisulfitového sekvenování (WGBS-Seq, Whole-Genome Bisulfite Sequencing) je konverze nemethylovaných cytosinů na uracilové báze za použití hydrogensířičitanu sodného. Tyto konvertované báze se pak při sekvenování čtou jako thyminy, zatímco 5-methylcytosin a 5-hydroxycytosin jsou před touto přeměnou chráněny a čteny jako cytosiny. Při sekvenování jsou čteny sekvence o velikosti 75-100 bp, které jsou předem amplifikovány za použití primerů umožňující vznik různě dlouhých úseků DNA. Tato metoda je vhodná pro analýzu většiny druhů a v lidských buňkách umožňuje pokrytí všech genů a více než 38 miliard CpG (Illumina, 2015).

5.2.2 Bisulfitové sekvenování se sníženým zastoupením

Principem metody bisulfitového sekvenování se sníženým zastoupením (RRBS-Seq, Reduced Bisulfite Sequencing) je enzymatické štěpení pomocí restričních endonukleáz a následné bisulfitové sekvenování. Tato kombinace dvou různých metod umožňuje obohacení oblastí genomu s vysokým obsahem CGI, čímž je umožněna analýza celého methylačního profilu buňky. Takto modifikované WGBS sekvenování fragmentů DNA o velikosti 100-150 bp je schopno analyzovat více než 85 % všech CGI, přičemž délka jednotlivých čtení je přibližně 1 x 75 bp (Illumina, 2015).

5.2.3 MeDIP sekvenování

Metoda MeDIP sekvenování (MeDIP-Seq, Methylated DNA Immunoprecipitation) je založena na sonikaci DNA na fragmenty o velikosti 100 až 300 bp, opravě obou konců těchto úseků, ligaci adapterů a následné denaturaci vláken DNA. Poté jsou k vychytávání fragmentů DNA obsahující methylované cytosinové báze použity magnetické kuličky obsahující na svém povrchu protilátky anti-5-methylcytosin. Nakonec je provedeno sekvenování, kde každé čtení má velikost přibližně 75-100 bp. Tato technika, kterou lze kombinovat s bisulfitovou

konverzí, se především využívá, s ohledem na její rozlišení, k analýze hypermethylovaných oblastí (Illumina, 2015).

5.2.4 Tab sekvenování

Technika Tab sekvenování (Tab-Seq, Tet-Associated Bisulfite Sequencing) využívá schopnosti enzymu β -glucosyltransferázy přidávat molekulu glukózy na 5-hydroxymethylcytosin. 5-methylcytosinové báze nejsou katalytickou funkcí tohoto enzymu ovlivněny, avšak vlivem enzymu Tet (ten-eleven translocation enzyme) dochází k jejich oxidaci na 5-karboxylcytosiny, které jsou následně vlivem hydrogensířičitanu sodného konvertovány na uracily. Glykosylované 5-methylcytosinové báze nelze oxidovat, a proto budou při sekvenování čteny jako cytosiny, zatímco methylcytosinové báze budou čteny jako thyminy. Tato metoda je využívána k detekci 5-hydroxymethylcytosinů a to v celém genomu s rozlišením až na 1 bázi (Illumina, 2015).

5.2.5 Methyl Capture sekvenování

Methyl Capture sekvenování (MethylCap-Seq, Methyl-Capture Sequencing) je robustní metoda celogenomového profilování DNA methylace. Tato technika se skládá z vycytávání methylované DNA pomocí MBD domény proteinu MeCP2 a NGS sekvenování. Použití MBD domén poskytuje výhodu vysoce kontrolované eluce methylovaných fragmentů DNA, která je provedena prostřednictvím několika kroků za využití gradientu roztoků solí, což umožňuje DNA do frakcí s odlišnou hustotou CpG nukleotidů. Získané frakce fragmentů DNA neobsahují oproti celému genomu velké množství genetické informace, proto může být dosaženo vysokého sekvenačního pokrytí už za použití jedné dráhy (lane) v sekvenátoru Illumina (Brinkman *et al.*, 2010).

5.3 Formáty sekvenačních dat

Při sekvenování je generováno obrovské množství dat uložené v podobě různých typů souborů. Různé formáty jsou kompatibilní s jinými programy, které jsou dále používány k hodnocení výsledků sekvenování, jako je například kontrola kvality nebo alignment (zarovnání).

5.3.1 FASTQ soubory

FASTQ je běžný formát používaný pro výměnu dat mezi programy. Je to rozšíření FASTA formátu, které má navíc schopnost uchovávat číselnou hodnotu kvality pro každou osekvenovanou nukleotidovou bázi. FASTQ formát byl prvně vyvinut Jimem Mullikinem, ale

vlastní FASTQ soubory si posléze vytvořily také dvě velké společnosti Solexa a Illumina (Cock *et al.*, 2009).

Ve FASTQ souborech jsou rozlišovány 4 typy řádků. První řádek nese informace o identifikátoru záznamu. Toto pole není nijak délkově ohraničeno, což umožňuje uživateli doplnit tento řádek o libovolnou anotaci a komentáře. Poté se ve FASTQ souborech nachází tzv. sekvenační řádky, které mohou být stejně jako ve FASTA formátu zabalené. Sekvenační řádky jsou následovány řádkem „+“, který označuje konec sekvenačního pole a začátek řádků s hodnotami kvality (Cock *et al.*, 2009).

5.3.2 SAM soubory

Soubor SAM (Sequence Alignment/Map) většinou vzniká zarovnáním FASTQ souboru k referenčnímu genomu. Je složený ze dvou částí, v první z nich se vyskytuje záhlaví neboli „hlavičková“ část, druhý úsek zahrnuje informace přímo o zarovnání nukleotidových sekvencí. Hlavičková část nemusí být ve všech SAM souborech přítomná, pokud se ale v souboru nachází, pak je uvedena před částí se zarovnáním a jsou na začátku řádku uvedeny znakem „@“. Každý řádek alignmentu má 11 polí, jejichž přítomnost v souboru je nutná, a variabilní počet fakultativních polí, které jsou klíčové a uchovávají dodatečnou informaci o použité platformě nebo o programu zprostředkovávající zarovnání (Li *et al.*, 2009; The SAM/BAM Format Specification, 2015).

5.3.3 BAM soubory

BAM soubory (*.bam) jsou komprimovanou binární verzí SAM souborů, které se používají k zobrazení zarovnaných nukleotidových sekvencí v programu IGV nebo UCSC databázi. Jako u souborů SAM je tento BAM formát rozdělen na dvě části, hlavičku a sekci obsahující informace o zarovnání sekvencí nukleotidů. V hlavičce jsou zakódovány informace o celém souboru (název a délka vzorku), ale zároveň jsou tyto informace propojeny s částí pro alignment (BAM File Format, n. d.; The SAM/BAM Format Specification, 2015).

Indexované BAM soubory (*.bam.bai) odpovídající formátu BAM jsou využívány pro rychlejší prohlížení BAM souborů, aniž by bylo potřeba prohledávat alignment veškerých dat (BAM File Format, n. d.; The SAM/BAM Format Specification, 2015).

6 Materiál a přístroje

6.1 Biologický materiál

Nezveřejněné, důvěrné.

6.2 Kity

- Qubit dsDNA HS Assay Kit (ThermoFisher Scientific)
- Agilent High Sensitivity DNA Kit (Agilent Technologies)
- MethylMiner Methylated DNA Enrichment Kit (Invitrogen)
- Nextera XT Sample Prep Kit (Illumina)
- Nextera XT Index Kit 96 indexes – 192 Samples (Illumina)
- HiSeq Rapid SBS Kit (Illumina)
- HiSeq Rapid SR Cluster Kit v2 (Illumina)

6.3 Chemikálie a použité roztoky

- HT1 (Hybridization Buffer) (Illumina)
- SPRIselect (Beckman Coulter)
- Hot Star Taq Plus Master Mix 2x (Qiagen)
- Primers for Non-Methylated Controls (Invitrogen) – součást MethylMinerTM Methylated DNA Enrichment Kitu
- Primers for Methylated Controls (Invitrogen) – součást MethylMinerTM Methylated DNA Enrichment Kitu
- Non-Methylated DNA (Invitrogen) – součást MethylMinerTM Methylated DNA Enrichment Kitu
- Methylated DNA (Invitrogen) – součást MethylMinerTM Methylated DNA Enrichment Kitu
- DEPC H₂O (Sigma-Aldrich)

- 100% ethanol (Sigma-Aldrich)
- 80% ethanol (Sigma-Aldrich)
- 70% ethanol (Sigma-Aldrich)
- Eva Green (Biotium)
- 5M NaOH

6.4 Laboratorní přístroje

- Qubit 2.0 Fluorometer (Invitrogen)
- Sonikátor BioRuptor Plus (Diagenode)
- Agilent 2100 Bioanalyzer (Agilent Technologies)
- Chip Priming Station (Agilent Technologies)
- C1000 Thermal Cycler (Bio-Rad)
- Real Time PCR Cycler CFX96 (Bio-Rad)
- Minicentrifuge MiniSpin (Eppendorf)
- Centrifuge Z 383 K (Hermle)
- Ika MS3 Vortexer (IKA)
- Vortex Mixer PV-1 (Pro LAB, P-LAB)
- Vortex Genie 2 (Scientific Industries)
- Sekvenátor HiSeq 2500 (Illumina)

7 Metodika

7.1 Příprava DNA

1. DNA z buněčných linií HCT116 byla izolována studentkou Mgr. Khushboo Agrawal. Izolace byla provedena prostřednictvím DNeasy Blood & Tissue Kitu a postupovalo se podle protokolu určeného pro izolaci DNA z kultivovaných buněk.
 - a. Byla provedena centrifugace nejvíce 5×10^6 buněk (300 g/5 min). Buňky byly následně resuspendovány v 200 μ l PBS a posléze bylo k suspenzi přidáno 20 μ l proteinázy K.
 - b. Do mikrozkušavky bylo přidáno 200 μ l Buffer AL. Obsah mikrozkušavky byl promíchán na vortexu a poté byly vzorky inkubovány při 56 °C po dobu 10 minut.
 - c. Bylo přidáno 200 μ l ethanolu (96-100%) a obsah mikrozkušavky byl vortexováním promíchán.
 - d. Směs byla napipetována na DNeasy Mini spin kolonu umístěnou ve sběrací mikrozkušavce (2 ml). Byla provedena centrifugace (min. 6000g/1 min) a po jejím ukončení byla sběrací mikrozkušavka odstraněna.
 - e. Spin kolona byla umístěna do nové sběrací mikrozkušavky (2 ml), bylo přidáno 500 μ l Buffer AW1. Následně byla provedena centrifugace (20 000g/3 min) a po jejím ukončení byla sběrací mikrozkušavka opět odstraněna.
 - f. Spin kolona byla přenesena do nové mikrocentrifugační zkumavky.
 - g. DNA byla z kolony eluována přidávkem 200 μ l Buffer AE (přidáno na střed membrány v koloně).
 - h. Obsah mikrozkušavek byl inkubován po dobu 1 minuty při pokojové teplotě. Poté byla provedena centrifugace (min. 6 000g/1 min).
2. Změřila jsem koncentraci DNA ve vzorcích na přístroji Qubit 2.0 Fluorometer.
 - a. Do mikrozkušavky jsem pro 1 měření připravila 200 μ l reakční směsi (199 μ l Qubit dsDNA HS Buffer + 1 μ l Qubit dsDNA HS Reagent)
 - b. K 199 μ l reakční směsi jsem přidala 1 μ l DNA vzorku a obsah mikrozkušavky promíchala na vortexu a následně stočila na minicentrifuze.
 - c. Změřila jsem koncentraci na přístroji Qubit 2.0 Fluorometer.

3. Vypočítala jsem objem DNA a DEPC vody potřebné pro přípravu 1 μg DNA s celkovým objemem 82 μl dle následujících vzorců:

$$\text{Objem DNA } [\mu\text{l}]: V_{\text{DNA}} = \frac{1000 \text{ ng}}{\text{nameřená koncentrace v ng}/\mu\text{l}}$$

$$\text{Objem DEPC vody } [\mu\text{l}]: V_{\text{H}_2\text{O}} = 82 - V_{\text{DNA}}$$

7.2 Fragmentace DNA

1. Do mikrozkmavek pro sonikaci jsem připravila 82 μl roztoku DNA (1 μg).
2. DNA ve vzorcích jsem štěpila sonikací na přístroji Bioruptor Plus s programem LOW 30s ON, 90s OFF, 12-18 cyklů. Počet cyklů jsem nastavila pro jednotlivé vzorky odlišně podle velikosti fragmentů DNA na elektroforetogramu tak, aby počet párů bází po štěpení byl přibližně v rozmezí 300 až 500.
3. Velikost fragmentů DNA ve vzorcích jsem určila pomocí gelové elektroforézy na čipu na přístroji Agilent 2100 Bioanalyzer, u kterého byl použit DNA High Sensitivity Kit a test DNA High Sensitivity.
 - a. Nastavila jsem preparační nástavec tak, aby stříkačka byla uzamčena v pozici, která je používána pro vtlačování gelu pro analýzu DNA.
 - b. High Sensitivity DNA čip jsem vložila na preparační nástavec.
 - c. Napipetovala jsem 9 μl gelu na dno jamky pro gel, píst jsem natáhla do pozice označené na stupnici jako 1 ml a uzavřela jsem preparační nástavec.
 - d. Stlačením pístu jsem vtlačila gel do všech kapilár uvnitř čipu a píst jsem v této pozici uzamkla.
 - e. Po 60 sekundách jsem píst uvolnila, vyčkala jsem 5 sekund a pak jsem píst pomalým pohybem opět natáhla do pozice označené na stupnici jako 1 ml.
 - f. Otevřela jsem preparační nástavec a následně jsem napipetovala 9 μl gelu do 2 dalších jamek pro gel.
 - g. Do všech jamek pro vzorky a marker jsem napipetovala 5 μl High Sensitivity DNA marker.
 - h. Do jamky pro ladder jsem napipetovala 1 μl High Sensitivity DNA ladder.
 - i. Do jamek pro vzorky jsem napipetovala 1 μl DNA.
 - j. Čip jsem položila na míchačku a nechala 1 minutu promíchávat.
 - k. Poté jsem čip vložila do přístroje Agilent 2100 Bioanalyzer a spustila analýzu.

7.3 Vychytávání fragmentů methylované DNA pomocí MethylMiner Methylated DNA Enrichment Kitu

7.3.1 Příprava magnetických kuliček

1. Připravila jsem 1x Wash Buffer pro 12 reakcí (360 μ l 5x Wash Buffer x N + 1,44 ml DEPC vody x N; N = počet reakcí).
2. Suspenzi magnetických kuliček Dynabeads M-280 Streptavidin jsem pipetováním důkladně resuspendovala.
3. Do prázdné mikrokumavky jsem napipetovala N x 10 μ l magnetických kuliček.
4. Mikrokumavku s magnetickými kuličkami jsem vložila na 1 minutu do magnetického stojanu.
5. Po shromáždění kuliček na vnitřní straně mikrokumavky jsem opatrně odstranila čirý roztok tak, abych neporušila pelet.
6. Vyjmula jsem mikrokumavku z magnetického stojanu.
7. Kuličky v mikrokumavce jsem resuspendovala pipetováním v N x 10 μ l 1x Wash Buffer.
8. Mikrokumavku jsem na 1 minutu vložila do magnetického stojanu.
9. Opatrně jsem z mikrokumavky odebrala čirý roztok.
10. Mikrokumavku s promytými kuličkami jsem vyjmula z magnetického stojanu a rychle jsem obsah magnetických kuliček resuspendovala jemným pipetováním v N x 10 μ l 1x Wash Buffer.

7.3.2 Příprava proteinu MBD-biotin

1. Do nové mikrokumavky jsem napipetovala N x 7 μ l roztoku MBD-biotin.
2. Objem jsem doplnila do 200 μ l roztokem 1x Wash Buffer.

7.3.3 Navázání proteinového komplexu MBD-biotin na magnetické kuličky

1. Ke 200 μ l roztoku MBD-biotin jsem přidala 100 μ l magnetických kuliček.
2. Obsah mikrokumavky jsem nechala inkubovat 1 hodinu na třepačce při pokojové teplotě.
3. Mikrokumavku jsem vložila na 1 minutu do magnetického stojanu.
4. Po shromáždění kuliček na vnitřní straně mikrokumavky jsem opatrně odstranila čirý roztok tak, abych neporušila pelet.
5. Obsah mikrokumavky jsem promyla 100 μ l roztoku 1x Wash Buffer.
6. Směs v mikrokumavce jsem nechala promíchávat na třepačce po dobu 5 minut.

7. Kroky 3-6 jsem provedla ještě dvakrát.
8. Mikrozkušavku jsem vložila na 1 minutu do magnetického stojanu.
9. Opět jsem opatrně odebrala z mikrozkušavky čirý roztok.
10. Přidala jsem 100 μl 1x Wash Buffer a resuspendovala pipetováním, vznikl koncentrovaný roztok magnetických kuliček s navázaným proteinem MBD-biotin.

7.3.4 Inkubace konjugátu MBD-biotin a magnetických kuliček s fragmentovanou DNA

1. Pro 12 vzorků fragmentované DNA jsem připravila 1200 μl zředěného roztoku magnetických kuliček v 1x Wash Buffer. Smíchala jsem 100 μl magnetických kuliček s navázaným proteinem MDB-biotin a 1100 μl 1x Wash Buffer.
2. Připravila jsem si zředěné roztoky methylované a nemethylované DNA o koncentraci 10 $\text{pg}/\mu\text{l}$, které byly v dalším postupu použity jako *spike-in* kontroly pro ověření správnosti vychytávání fragmentů methylované DNA. 1 μl příslušné methylované/nemethylované DNA o koncentraci 1 $\text{ng}/\mu\text{l}$ jsem zředila v 99 μl DEPC vody.
3. Do nové mikrozkušavky jsem k 80 μl vzorku fragmentované DNA přidala 20 μl 5x Wash Buffer, 100 μl zředěného roztoku magnetických kuliček s navázaným komplexem MBD-biotin, 1 μl zředěné kontrolní methylované DNA (10 $\text{pg}/\mu\text{l}$) a 1 μl zředěné kontrolní nemethylované DNA (10 $\text{pg}/\mu\text{l}$). Tímto způsobem jsem připravila všech 12 vzorků.
4. Obsah mikrozkušavek se vzorky jsem nechala 1 hodinu promíchávat na třepačce při pokojové teplotě.

7.3.5 Eluce nemethylované DNA

1. Mikrozkušavku s fragmentovanou DNA a magnetickými kuličkami jsem vložila na 1 minutu do magnetického stojanu.
2. Po zkoncentrování všech magnetických kuliček na vnitřní straně mikrozkušavky jsem odebrala čirý roztok a tuto frakci uschovala do nové mikrozkušavky (**frakce nemethylované DNA**).
3. Do mikrozkušavky s magnetickými kuličkami jsem přidala 200 μl 1x Wash Buffer.
4. Obsah mikrozkušavky jsem nechala promíchávat po dobu 3 minut na třepačce.
5. Mikrozkušavku jsem opět vložila na 1 minutu do magnetického stojanu, magnetické kuličky se opět zkoncentrovaly na vnitřní straně mikrozkušavky.

6. Odebrala jsem čirý roztok a ten přepipetovala do nové mikroskopické kuličky (**frakce 1x Wash Buffer**).
7. Do mikroskopické kuličky s koncentrovanými magnetickými kuličkami jsem přidala 200 μl 1x Wash Buffer.
8. Obsah mikroskopické kuličky jsem nechala promíchávat na třepačce po dobu 3 minut.
9. Mikroskopickou kuličku jsem opět vložila na 1 minutu do magnetického stojanu, magnetické kuličky se opět zkoncentrovaly na vnitřní straně mikroskopické kuličky.
10. Odebrala jsem čirý roztok a přidala do mikroskopické kuličky s frakcí 1x Wash Buffer.

7.3.6 Eluce methylované DNA

1. Do mikroskopické kuličky s koncentrovanými magnetickými kuličkami jsem přidala 200 μl High Salt Elution Buffer.
2. Obsah mikroskopické kuličky jsem nechala promíchávat 3 minuty na třepačce.
3. Mikroskopickou kuličku jsem vložila na 1 minutu do magnetického stojanu.
4. Z mikroskopické kuličky jsem odebrala čirý roztok tak, abych neporušila usazené kuličky. Tuto frakci methylované DNA jsem přepipetovala do nové mikroskopické kuličky (**frakce methylované DNA**).
5. Do mikroskopické kuličky s magnetickými kuličkami jsem opět přidala 200 μl High Salt Elution Buffer.
6. Obsah mikroskopické kuličky jsem nechala 3 minuty míchat na třepačce.
7. Mikroskopickou kuličku jsem vložila na 1 minutu do magnetického stojanu.
8. Z mikroskopické kuličky jsem odebrala čirý roztok a přidala do mikroskopické kuličky s již odebranou frakcí methylované DNA.

7.3.7 Precipitace ethanolem

1. Ke každé frakci v mikroskopických kuličkách jsem přidala 1 μl glykogenu (20 $\mu\text{g}/\mu\text{l}$).
2. Přidala jsem 1/10 objemu 3M acetátu sodného (pH = 5,2), 20 nebo 40 μl , a 2x objem 100% ethanolu, 400 nebo 800 μl .
3. Obsah zkumavek jsem důkladně promíchala vortexováním a nechala inkubovat minimálně 2 hodiny při -20 až -80 $^{\circ}\text{C}$.
4. Provedla jsem centrifugaci (12 000 g/15 min/4 $^{\circ}\text{C}$).
5. Odstranila jsem supernatant tak, abych nedošlo k porušení peletu na dně mikroskopické kuličky.
6. Přidala jsem 500 μl chlazeného 70% ethanolu.

7. Provedla jsem centrifugaci (12 000 g/5 min/4 °C).
8. Opět jsem odstranila supernatant tak, aby nedošlo k porušení peletu na dně mikrozkuřavky.
9. Provedla jsem centrifugaci (12 000 g/15 min/4 °C).
10. Opatrně jsem odstranila veřkerý supernatant.
11. Supernatant na dně mikrozkuřavky jsem nechala 5 minut vysychat na vzduchu.
12. Obsah mikrozkuřavky jsem resuspendovala v 30 µl DEPC vody.
13. Změřila jsem koncentraci DNA ve vzorcích na přístroji QUBIT 2.0 Fluorometer High Sensitivity (postup viz **Příprava DNA**, krok č. 2).
14. Vzorky jsem uschovala v mrazáku při -80 °C.

7.4 Kontrola kvality methylované DNA pomocí qPCR

7.4.1 Příprava vzorků pro qPCR

1. Připravila jsem reakční směs pro methylovanou a nemethylovanou DNA (viz **Tab. 1 a 2**).
2. Do mikrozkuřavky pro PCR jsem napipetovala 18 µl reakční směsi pro methylovanou DNA a přidala 1 µl fluorescenčního barviva Eva Green a 1 µl vzorku DNA z frakce s methylovanou DNA. Tímto způsobem jsem připravila všech 12 vzorků.
3. Do nové PCR mikrozkuřavky jsem napipetovala 18 µl reakční směsi pro nemethylovanou DNA a přidala 1 µl fluorescenčního barviva Eva Green a 1 µl vzorku DNA z frakce nemethylované DNA. Tímto způsobem jsem připravila všech 12 vzorků.

7.4.2 Příprava kalibrační řady pro qPCR

1. Ze zásobních roztoků methylované a nemethylované *spike-in* DNA jsem připravila zředěné roztoky o koncentracích 1×10^7 , 1×10^6 , 1×10^5 , 1×10^4 , 1×10^3 , 1×10^2 , 1×10^1 a 1.
2. Do nové mikrozkuřavky pro PCR jsem napipetovala 18 µl reakční směsi pro methylovanou DNA. Přidala jsem 1 µl fluorescenčního barviva Eva Green a 1 µl roztoku kalibrační řady methylované DNA nebo 1 µl DEPC vody (negativní kontrola).
3. Do nové mikrozkuřavky pro PCR jsem napipetovala 18 µl reakční směsi pro nemethylovanou DNA. Přidala jsem 1 µl fluorescenčního barviva Eva Green a 1 µl roztoku kalibrační řady methylované DNA nebo 1 µl DEPC vody (negativní kontrola).

7.4.3 qPCR

1. Vzorky jsem vložila do termocykléru CFX 96 a spustila program PCR Methylation Capture Kit v ovládacím programu termocykléru (viz **Tab. 3**).
2. Z naměřených SQ hodnot (starting quantity, množství DNA vstupující do PCR reakce) a kalibrační řady jsem vypočítala počet kopií *spike-in* DNA a procentuální zastoupení methylované a nemethylované *spike-in* kontroly v jednotlivých frakcích každého analyzovaného vzorku.

Tab. 1: Složení reakční směsi pro qPCR (methylovaná DNA)

Reagencie	Objem pro 1 reakci [μl]
Hot Star <i>Taq</i> Plus Master Mix 2x	10
Primery 10 μM – methylovaná DNA	0,5
H ₂ O	7,5
Celkový objem	18

Tab. 2: Složení reakční směsi pro qPCR (nemethylovaná DNA)

Reagencie	Objem pro 1 reakci [μl]
Hot Star <i>Taq</i> Plus Master Mix 2x	10
Primery 10 μM – nemethylovaná DNA	0,5
H ₂ O	7,5
Celkový objem	18

Tab. 3: Podmínky qPCR

Cyklus	Teplota	Čas	Opakování
Denaturace	95 °C	15 min	1x
Denaturace	94 °C	15 s	40x
Hybridizace	55 °C	15 s	
Elongace	68 °C	30 s	
Konečná elongace	72 °C	5 min	1x

7.5 Tagmentace s použitím Nextera XT Sample Prep Kitu

1. Pro každý vzorek jsem do PCR mikrozkušavek připravila 5 μl 1 ng DNA z frakce methylované DNA (0,2 ng/μl). Přidala jsem 10 μl TD a 5 μl ATM.
2. Mikrozkušavky jsem vložila do termocykléru a spustila jsem program tagmentace (viz **Tab. 4**).
3. Ke každému vzorku jsem přidala 5 μl NT, 15 μl NPM a 5 μl příslušných indexů (viz **Tab. 5**).
4. Mikrozkušavky jsem umístila do termocykléru a spustila jsem program PCR (viz **Tab. 6**).

- Změřila jsem koncentraci DNA ve vzorcích na přístroji QUBIT 2.0 Fluorometer High Sensitivity (postup viz **Příprava DNA**, krok č. 2).

Tab. 4: Podmínky tagmentace

Cyklus	Teplota	Čas	Opakování
1	55 °C	5 min	1x
2	10 °C	60 min	1x

Tab. 5: Indexy použité při tagmentaci vzorků

Vzorky	Index i5	Index i7
M1	1	1
M2	2	1
M3	3	1
M4	4	1
M5	5	1
M6	6	1
M7	7	1
M8	8	1
M9	1	2
M10	2	2
M11	3	2
M12	4	2

Tab. 6: Podmínky PCR

Cyklus	Teplota	Čas	Opakování
1	72 °C	3 min	12x
2	95 °C	30 s	
3	95 °C	10 s	
4	55 °C	30 s	
5	72 °C	30 s	
6	72 °C	5 min	1x
7	72 °C	nekonečno	1x

7.6 Přečistění DNA na magnetických kuličkách SPRIselect

- K 48 µl tagmentovaného vzorku jsem přidala 72 µl suspenze magnetických kuliček SPRIselect a pipetováním jsem obsah zkumavky jemně promíchala (10x).
- Vzorky jsem nechala 3-5 minut inkubovat při pokojové teplotě.
- Mikrozukavky jsem položila na magnet a nechat 1 minutu magnetizovat do vyčíření.
- Opatrně jsem odsála čirý supernatant.
- Ke každému vzorku jsem přidala 150 µl 80% ethanolu.
- Z každé mikrozukavky jsem pipetou odsála ethanol.
- Nechala jsem z mikrozukavek odpařit ethanol (minimálně 2 minuty).

8. Mikrozkušavky jsem sundala z magnetu.
9. Do každé zkumavky jsem přidala 30 μl H_2O a dokonale pipetováním promíchala, počkala jsem 3 minuty.
10. Mikrozkušavky jsem položila na magnet a nechala magnetizovat do vyčíření.
11. Odebrala jsem čirý vzorek (25 μl).
12. Změřila jsem koncentraci DNA ve vzorcích na přístroji Qubit 2.0 Fluorometer High Sensitivity (postup viz **Příprava DNA**, krok č. 2).
13. Pro všechny vzorky jsem provedla gelovou elektroforézu na čipu na přístroji Agilent 2100 Bioanalyzer, u kterého byl použit DNA High Sensitivity Chip a test DNA High Sensitivity (postup viz **Fragmentace DNA**, krok č. 3).
14. Vzorky jsem uschovala do mrazáku při $-80\text{ }^\circ\text{C}$.

7.7 Příprava vzorků na sekvencování

1. Z naměřených koncentrací DNA (viz **Přečištění DNA na magnetických kuličkách**, krok č. 12) jsem vypočítala objemy potřebné pro poolování. Pool vzorků jsem připravila tak, aby byl každý vzorek zastoupen ve stejném množství (3 ng/vzorek).
2. Koncentraci poolu jsem 3x zkontrolovala na přístroji Qubit 2.0 Fluorometer High Sensitivity (postup viz **Příprava DNA**, krok č. 2) a následně jsem vypočítala jejich průměrnou hodnotu.
3. Ze zjištěné koncentrace poolu jsem podle vzorce vypočítala molaritu DNA v poolu.

$$c [\text{nM}] = \frac{\rho_i [\text{ng}/\mu\text{l}] \times 10^6}{660 \times 350}$$

Legenda: 660 g/mol je průměrná molární hmotnost jedné báze (1 bp); 350 je orientační průměrná velikost molekuly DNA po indexaci v bp

4. Naředila jsem pool DEPC vodou na koncentraci 1,6 nM (10 μl poolu a 18 μl vody).
5. Smícháním 10 μl 5M NaOH a 240 μl DEPC vody jsem připravila 0,2M NaOH.
6. K 10 μl 1,6nM poolu jsem přidala 10 μl 0,2M NaOH a 5 minut inkubovala při pokojové teplotě.
7. Denaturovanou DNA jsem zředila ledovým hybridizačním pufrem. K 10 μl poolu jsem napipetovala 990 μl HT1.

7.8 Sekvenování

1. Za asistence svého školitele, Ing. Rastislava Slavkovského, PhD., jsem do sekvenační kazety vložila reagencie pro sekvenování, jednalo se o chemikálie pro HiSeq sekvenování (Illumina HiSeq SBS Rapid Kit).
2. Obsluhu HiSeq 2500 sekvenátoru zajišťoval pracovník ÚMTM, Mgr. Vojta Petr, který do sekvenátoru vložil 600 µl připraveného poolu v HT1 a poté spustil sekvenační program (Illumina HiSeq software).

7.9 Zpracování dat

Data získaná po sekvenaci následně upravil a zpracoval vedoucí mé bakalářské práce Ing. Rastislav Slavkovský, PhD.

1. Pro data ze sekvenátoru v podobě FASTQ souborů byl proveden alignment do formátu SAM, při kterém byly sekvence DNA přirovnány k referenčnímu lidskému genomu (hg19). Pro tohle mapování byl použit program BWA.
2. Následně byly tyto SAM soubory převedeny do podoby BAM. Pro tuto konverzi byly v programu SAMtools použity tři příkazy: view, sort a index.
 - a. `samtools view -b -s $fp.sam>$fp.bam`
 - b. `samtools sort $fp.bam $(fp)_sort`
 - c. `samtools index $(fp)_sort.bam`
3. Data jednotlivých vzorků byla v experimentálních skupinách sloučena.
4. Data byla analyzována v programu MACS2 pomocí funkce peak calling. Tak byly zjištěny informace o poloze píku odpovídající metylované oblasti v jednotlivých vzorcích samostatně a poté i v rámci experimentální skupiny, kdy bylo provedeno porovnání methylačních profilů mezi skupinami (skupina A a C).
5. Pomocí skriptovacího jazyka Python byl vytvořen program, pomocí kterého byly metylované oblasti anotovány.
6. V programu IGV byly vizualizovány a porovnány methylační profily jednotlivých skupin.

8 Výsledky

V buněčné linii HCT116 odvozené od kolorektálního karcinomu byl studován methylační profil genomu, se zaměřením na celogenomový screening methylačního profilu buněk rezistentních vůči 5-aza-2-deoxycytidinu (DAC, decitabin). Detekce epigenetických profilů jednotlivých vzorků byla provedena metodou vychytávání fragmentů methylované DNA na magnetických kuličkách a jejich následným sekvenováním na přístroji HiSeq 2500 (Illumina) (MethylCap-Seq).

Před samotným sekvenováním byla provedena kontrola kvality eluovaných frakcí analyzovaných vzorků. Tato validace byla provedena metodou qPCR s pomocí specifických primerů pro methylovanou a nemethylovanou *spike-in* DNA kontrolu. Tyto kontroly byly ke vzorkům přidány ještě před vychytáváním methylovaných fragmentů DNA na magnetických kuličkách, přičemž po provedení qPCR amplifikace bylo z detekovaných kopií těchto kontrol vypočítáno procentuální zastoupení methylované a nemethylované DNA ve všech frakcích jednotlivých vzorků (viz **Tab. 7**). U uvedených hodnot lze usoudit, že odebrané frakce nemethylované DNA u všech vzorků, obsahují téměř vždy pouze fragmenty nemethylované DNA. Ve frakci 1x Wash Buffer se v různých podílech vyskytuje methylovaná i nemethylovaná DNA, zatímco ve frakci obsahující DNA, která byla eluována pomocí High Salt Elution Buffer, byla téměř v 100% zastoupení detekována pouze methylovaná DNA.

Tab. 7: Procentuální zastoupení methylované a nemethylované DNA v odebraných frakcích jednotlivých vzorků

Identifikace		Počet detekovaných kopií <i>spike-in</i> kontroly		Podíl z celkového počtu detekovaných kopií <i>spike-in</i> kontroly	
Vzorek	Frakce	NM	M	NM	M
NC1	Nemethylovaná DNA	6,7E+05	2,6E+03	99,6%	0,4%
NC2		4,7E+06	5,1E+04	98,9%	1,1%
NC3		5,1E+06	4,1E+05	92,5%	7,5%
NC4		1,1E+07	2,3E+04	99,8%	0,2%
NC5		7,9E+06	3,1E+04	99,6%	0,4%
NC6		1,4E+07	2,3E+05	98,4%	1,6%
NC7		4,7E+06	2,8E+04	99,4%	0,6%
NC8		1,6E+07	4,1E+04	99,7%	0,3%
NC9		9,5E+06	1,0E+05	98,9%	1,1%
NC10		9,2E+06	8,1E+04	99,1%	0,9%
NC11		5,3E+06	1,4E+04	99,7%	0,3%
NC12		1,2E+07	5,4E+04	99,6%	0,4%

Identifikace		Počet detekovaných kopií <i>spike-in</i> kontroly		Podíl z celkového počtu detekovaných kopií <i>spike-in</i> kontroly	
Vzorek	Frakce	NM	M	NM	M
W1	1x Wash Buffer	1,2E+04	1,5E+04	44,9%	55,1%
W2		8,5E+03	7,7E+03	52,4%	47,6%
W3		6,6E+04	1,0E+05	39,8%	60,2%
W4		1,9E+04	2,7E+04	41,9%	58,1%
W5		1,8E+04	7,5E+03	70,1%	29,9%
W6		1,5E+04	9,4E+03	61,6%	38,4%
W7		1,5E+04	7,0E+03	68,0%	32,0%
W8		2,8E+04	1,2E+04	71,0%	29,0%
W9		4,5E+04	1,7E+04	72,8%	27,2%
W10		4,0E+04	3,1E+04	56,5%	43,5%
W11		1,1E+04	7,8E+02	93,1%	6,9%
W12		4,3E+04	1,3E+04	77,0%	23,0%
M1	Methylovaná DNA	2,2E+03	3,5E+06	0,1%	99,9%
M2		4,7E+02	3,5E+06	0,0%	100,0%
M3		4,0E+03	2,9E+06	0,1%	99,9%
M4		1,9E+03	2,2E+06	0,1%	99,9%
M5		2,0E+03	2,1E+06	0,1%	99,9%
M6		5,3E+02	2,7E+06	0,0%	100,0%
M7		3,3E+02	1,8E+06	0,0%	100,0%
M8		2,3E+03	2,6E+06	0,1%	99,9%
M9		3,6E+03	2,0E+06	0,2%	99,8%
M10		4,2E+02	1,9E+06	0,0%	100,0%
M11		2,2E+03	2,9E+06	0,1%	99,9%
M12		2,4E+03	3,1E+06	0,1%	99,9%

Legenda: NC (1 – 12) = frakce nemethylované DNA; W (1 – 12) = frakce 1x Wash Buffer (methylovaná i nemethylovaná DNA); M (1- 12) = frakce methylované DNA; NM = nemethylovaná *spike-in* DNA kontrola; M = methylovaná *spike-in* DNA kontrola

Tabulka 8 obsahuje údaje o identifikaci vzorků zařazených do sekvenční analýzy, jejich označení, zařazení do experimentální skupiny a počtu sekvenčních čtení (reads), která u nich proběhla. Vzorky DNA označeny jako HCT116 byly izolovány z rakovinných buněk kolorekta a z této mateřské linie byly následně připraveny buněčné linie HCT116 DAC, R1.4 a R1.4 JQ1. Buněčné linie označené jako HCT116 DAC byly ošetřeny hypomethylační látkou decitabinem. Vzorky označené jako R1.4 obsahují DNA izolovanou z buněk rezistentních vůči decitabinu a vzorky s označením R1.4 JQ1 obsahují DNA z buněk rezistentních vůči decitabinu poté, co na ně působila inhibiční látka thienotriazolodiazepin, která má schopnost opět nastolit senzitivu buněk vůči decitabinu. Celkem byly analyzovány 4 experimentální skupiny obsahující vždy 3 replikáty.

Tab. 8: Vzorky použité při analýze

ID vzorku	Pořadové číslo	Experimentální skupina	Buněčná linie a ošetření	Počet identifikovaných sekvenčních čtení
M1	1	A	HCT116A	28 M
M2	2	B	HCT116A DAC	27 M
M3	3	C	R1.4A	30 M
M4	4	D	R1.4 JQ1	33 M
M5	5	A	HCT116B	33 M
M6	6	B	HCT116B DAC	37 M
M7	7	C	R1.4B	30 M
M8	8	D	R1.4 JQ1	29 M
M9	9	A	HCT116C	28 M
M10	10	B	HCT116B DAC	32 M
M11	11	C	R1.4C	30 M
M12	12	D	R1.4 JQ1	39 M
Celkový počet sekvenčních čtení				376 M

Legenda: HCT116 = buněčná linie rakovinných buněk kolorekta; DAC = decitabin, 5-aza-2-deoxycytidin; JQ1 = thienotriazolodiazepin; R = rezistentní; A, B, C = identifikátor replikátoru; M (1 – 12) = frakce methylované DNA

Graf 1 zobrazuje statistické přiřazení identifikovaných čtení, která prošla interním filtrem kvality, k analyzovaným vzorkům. Z údajů v grafu je patrné, že počet čtení u jednotlivých vzorků je u všech přibližně stejný, můžeme tedy usuzovat, že sekvenování probíhalo rovnocenně a nedocházelo během něj ke vzniku artefaktů.

Graf 1: Identifikovaná sekvenční čtení u jednotlivých vzorků



Legenda: Index number = číslo indexu přiřazené ke každému vzorku (pořadové číslo viz **Tab. 7**); % Reads Identified (PF) = procento sekvenčních čtení mapovaných k jednotlivým vzorkům, která prošla vnitřním filtrem kvality

Tabulka 9 obsahuje hodnoty parametrů sekvenování vypovídající o jeho kvalitě a výkonnosti. Celkový počet cyklů v sekvenčním čtení bylo i po započítání sekvencí indexů 67. Hustota vytvořených klastrů byla přibližně 1 100 K klastrů na 1 mm², což je asi o 20 % více, než je uvedeno ve specifikaci pro použité chemikálie (HiSeq Rapid v2 Kit). Nicméně i přes tzv. překlastrování nedošlo k poklesu kvality sekvenovaných bází. Sekvenčních čtení, která prošla vnitřním filtrem kvality (PF), bylo zhruba 386 M. Z tohoto počtu mělo 97,28 % čtení kvalitu vyšší než 30, což znamená, že tyto báze byly s 99,9% pravděpodobností osekvenovány správně. Celkový výtěžek sekvenování byl 24,74 Gbp.

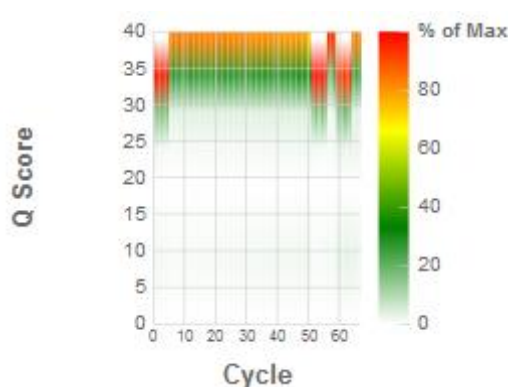
Celkový počet čtení pro všechny vzorky v **Tabulce 8** se od údajů v **Tabulce 9** mírně liší. Je to způsobeno tím, že tyto hodnoty započítávají pouze sekvenční čtení, která byla identifikována k jednotlivým indexům, zatímco v **Tabulce 9** jsou zahrnuta čtení všechna.

Tab. 9: Výstupní hodnoty ze sekvenování

Počet cyklů	Hustota klastrů (K / mm ²)	Klastr PF (%)	Počet čtení	Počet čtení PF	% > Q30	Výtěžek (Gbp)
67	1100 ±118	94,565 ± 1,715	409 480 512	386 519 312	97,28	24,74

Graf 2 znázorňuje rozložení hodnoty kvality (Q score) v průběhu všech cyklů sekvenčního čtení. Z grafu je patrné, že během sekvenování nedocházelo k poklesům kvality, výjimkou jsou pouze první a poslední cykly, ve kterých probíhalo čtení sekvencí adapterů (indexů) na okrajových částech analyzovaných fragmentů DNA.

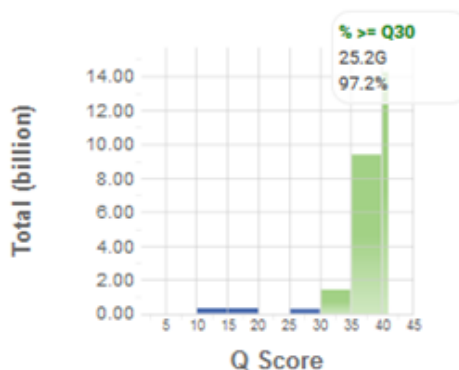
Graf 2: Hodnota kvality v průběhu cyklů sekvenčního čtení



V **Grafu 3** je zobrazena distribuce počtu přečtených bází v závislosti na jejich hodnotě kvality. Z tohoto grafu vyplývá, že 97,2 % všech osekvenovaných bází, které prošly interním filtrem, mají hodnotu kvality vyšší než 30. Celkový výtěžek sekvenovaných bází (25,2 Gbp) v grafu nekorresponduje s hodnotou uvedenou v **Tabulce 8**. To je pravděpodobně způsobeno

zaokrouhlováním při výpočtech, tato nesrovnalost je avšak minimální a nemá žádný význam pro další analýzy.

Graf 3: Rozložení přečtených bází podle jejich skóre kvality



Data vyprodukovaná sekvenováním v podobě FASTQ souborů byla převedena do formátu BAM a následně sloučena do 4 velkých souborů obsahující data jednotlivých experimentálních skupin.

Pomocí programu MACS2 byl proveden tzv. peak calling, čímž byly získány informace o přesné poloze píků, které korespondovaly s jednotlivými detekovanými methylovanými oblastmi genomu. V programu MACS2 byly nejdříve analyzovány jednotlivé vzorky samostatně a pro další hodnocení byly porovnávány mezi sebou experimentální skupiny, tedy jedna experimentální skupiny vůči druhé. V tomto případě byla provedena analýza rozdílů v methylaci na základě odlišného pokrytí.

V **Tabulce 10** jsou uvedena data získaná z analýzy jednotlivých vzorků v programu MACS2. Ke každému vzorku v experimentální skupině je přiřazen celkový počet čtení použitý pro detekci píků znázorňující methylované oblasti. Dále se v tabulce nachází hodnoty průměrného počtu detekovaných píků ve vzorcích a v experimentálních skupinách. Z těchto dat byly následně vypočítány normalizované hodnoty počtu píků na 1 M sekvenčních čtení. Z hodnot normalizovaného počtu píků na 1 M čtení vyplývá, že neexistuje přímá spojitost mezi počtem čtení a počtem píků. Nižší počet píků v některých vzorcích avšak může souviset s méně efektivním obohacením methylované DNA v preanalytické fázi.

Ve vzorcích experimentální skupiny B, u kterých byla použita demethylační látka decitabin, byl oproti očekávání detekován velký počet píků. To je pravděpodobně způsobeno schopností decitabinu inkorporovat se do DNA a chovat se jako tzv. falešná „methylovaná“ DNA. Tato DAC-DNA pak byla nejspíše stejně jako methylovaná DNA vychytávána pomocí proteinů MBD, a proto bylo u těchto vzorků detekováno více píků, nežli tomu bylo u ostatních experimentálních skupin.

Tab. 10: Celkový počet píků vygenerovaných v programu MACS2

Experimentální skupina	Vzorek	Počet čtení ve vzorku	Počet píků ve vzorku	Průměrný počet píků ve vzorku	Normalizace počtu píků na 1 M čtení
A (HCT116)	1	25 249 388	246 146	222 664	9749
	5	28 756 059	184 235		6407
	9	25 039 065	237 611		9490
B (HCT116 DAC)	2	23 660 159	215 317	304 001	9100
	6	31 801 912	384 235		12082
	10	27 562 875	312 450		11336
C (R1.4)	3	26 863 776	163 395	201 158	6082
	7	27 121 064	318 540		11745
	11	26 553 947	121 540		4577
D (R1.4 JQ1)	4	27 919 812	174 130	254 916	6237
	8	25 408 255	260 566		10255
	12	33 662 964	330 052		9805

Legenda: HCT116 = buněčná linie rakovinných buněk kolorekta; DAC = decitabin, 5-aza-2-deoxycytidin; JQ1 = thienotriazolodiazepin; R = rezistentní; A, B, C = identifikátor replikátoru

V **Tabulce 11** jsou zobrazeny veškeré informace uvedené v hlavičkové části výstupního souboru z programu MACS, která obsahuje údaje o identifikaci vzorků (název, formát) a klíčových parametrech provedené analýzy.

Tab. 11: Hlavička výstupního souboru z programu MACS – porovnání skupiny C a A

```
# This file is generated by MACS version 2.0.9 20111102 (tag:alpha)
# ARGUMENTS LIST:
# name = M_CvsAq001
# format = BAM
# ChIP-seq file = M3711_sort.bam
# control file = M159_sort.bam
# effective genome size = 2.70e+09
# band width = 300
# model fold = 10,30
# qvalue cutoff = 1.00e-02
# Larger dataset will be scaled towards smaller dataset.
# Range for calculating regional lambda is: 1000 bps and 10000 bps
# Broad region calling is off

# tag size is determined as 50 bps
# total tags in treatment: 86736043
# tags after filtering in treatment: 77686680
# maximum duplicate tags at the same position in treatment = 2
# Redundant rate in treatment: 0.10
# total tags in control: 86193534
# tags after filtering in control: 75970322
# maximum duplicate tags at the same position in control = 2
# Redundant rate in control: 0.12
# d = 51
```

Legenda: name = název; formát = formát, ve kterém je soubor vkládán; ChIP-seq file = analyzovaný soubor; control file = kontrolní soubor, se kterým je vkládán soubor porovnáván; effective genome size = velikost genomu po odstranění repetitivních elementů; bandwidth = velikost fragmentů DNA po sonikaci; model fold = parametr definující charakteristiku píku; total tags = počet čtení; tag size = velikost čtení; qvalue cutoff = hodnoty qvalue (-log10), které jsou z analýzy vyřazeny; redundant rate = část čtení, která jsou nadbytečná a proto vyřazena z analýzy

V **Tabulce 12** jsou uvedena částečná data získaná prostřednictvím funkce peak calling v programu MACS. Do tabulky byla vybrána data prvních deseti záznamů z výstupního souboru, který obsahuje porovnání methylace DNA experimentální skupiny C a A, kde skupina A sloužila jako kontrola, vůči které byly detekovány rozdíly v methylaci. V tabulce jsou shromážděny informace o přesné poloze píku znázorňující methylovanou oblast na chromozomu, délce píku a místě, kde se nachází nejvyšší bod píku (abs summit). Dále jsou v tabulce uvedeny hodnoty pileup, které udávají počet detekovaných čtení přiřazených k jedné oblasti DNA a fold enrichment charakterizující obohacení methylace dané oblasti DNA oproti DNA v kontrolní experimentální skupině. Se zvyšující se hodnotou pileup roste i významnost píku. P-value a q-value udávají míry statistické věrohodnosti detekovaných píků. Hodnoty q-

value jsou stanovovány z hodnot p-value pomocí Benjamini-Hochbergova postupu. Existuje tedy jistá provázanost mezi q- a p-value a lze tedy říci, že se zvyšující se hodnotou p roste i hodnota q . Do tohoto vztahu lze zahrnout i závislost q-value na hodnotách fold enrichment a pileup, kdy lze pozorovat, že píky vykazující vyšší hodnoty q mají taktéž vyšší hodnoty pileup a fold enrichment.

Tab. 12: Částečný výstup porovnání experimentální skupiny C vůči skupině A z programu MACS2

Chromosome	Start	End	Length	Abs summit	Pileup	p-value	Fold enrichment	q-value
chr1	8005358	8005408	51	8005384	20.00	12.61	8.44	6.67
chr1	8087006	8087062	57	8087053	34.00	11.77	4.28	6.40
chr1	9471225	9471317	93	9471270	72.00	15.15	3.20	8.60
chr1	15869872	15869922	51	15869914	19.00	10.84	7.21	5.36
chr1	16898270	16898396	127	16898292	14.00	10.35	10.22	4.94
chr1	18270572	18270638	67	18270602	19.00	6.46	3.91	2.28
chr1	23719060	23719156	97	23719106	22.00	8.41	4.50	3.56
chr1	23776029	23776092	64	23776079	16.00	11.60	10.19	5.89
chr1	23832247	23832320	74	23832290	30.00	12.31	5.50	6.45
chr1	25384486	25384562	77	25384488	28.00	7.30	3.15	2.67

Legenda: Chromosome = chromozom; Start = místo počátku píku; End = místo konce píku; Length = délka píku; Abs summit = místo, ve kterém se nachází nejvyšší bod píku; Pileup = počet detekovaných čtení přiřazených k jedné oblasti DNA; p-value = míra statistické věrohodnosti detekovaných píků; Fold enrichment = bohatost methylace dané oblasti DNA oproti DNA v kontrolní experimentální skupině; q-value = míra statistické věrohodnosti detekovaných píků

V **Tabulkách 13 a 14** jsou obsažena anotovaná data získaná funkcí peak calling. Jednotlivé píky, které představují methylovanou oblast, byly pomocí skriptovacího jazyka Python přiřazeny ke genům, které u těchto míst leží nejbližší. Jednotlivé anotace zahrnují informace o genu, který je asociován s příslušným píkem, tedy údaje o identifikaci genu, jeho umístění na chromozomech a orientaci vlákna DNA. Jsou zde také uvedeny informace o poloze start kodonu, transkripčním počátku (Tx start) a transkripčním konci (Tx end) v genomu. *Absolutní vzdálenost píku od transkripčního počátku* je hodnota udávající vzdálenost okraje methylované oblasti od první báze v transkripčním počátku nebo transkripčním konci v závislosti na orientaci vlákna DNA. V případě, že gen leží na vláknech DNA orientovaném „+“, pak místem začátku transkripce je Tx start. Pokud ale gen leží na vláknech DNA orientovaném opačně („-“), pak se počítá s tím, že transkripce reálně začíná v místě Tx end. Dále jsou v tabulkách uvedeny informace o methylované oblasti, tedy její lokalizaci v rámci

genu (před genem/uvnitř genu), a hodnoty peak pileup, peak fold enrichment a peak q-value, které korespondují s hodnotami vygenerovanými prostřednictvím funkce peak calling.

Nezveřejněné, důvěrné.

9 Diskuze

V bakalářské práci byla provedena analýza methylačního profilu buněk se zaměřením na celogenomový screening methylace DNA u buněčných linií HCT116 rezistentních vůči decitabinu. Pro tyto účely byla použita mateřské buněčná linie HCT116 odvozená od kolorektálního karcinomu a její dceřiné buněčné linie HCT116 DAC, R1.4 a R1.A JQ1. V experimentu byla pro detekci methylované DNA použita metoda Methyl-Capture sekvenování, kde pro enrichment methylovaných oblastí byl využit MethylMiner Methylated DNA Enrichment Kit (Invitrogen) a samotné sekvenování bylo provedeno na platformě Illumina pomocí HiSeq Rapid SBS Kitu.

Před samotným vychytáváním methylovaných fragmentů DNA byly ke vzorkům v poměru 1 : 1 přidány methylované a nemethylované *spike-in* DNA kontroly, které byly následně použity pro ověření efektivity vychytávání methylovaných fragmentů DNA, tzv. methylmining. Tato validace byla provedena prostřednictvím metody qPCR, přičemž z množství detekovaných kopií těchto *spike-in* DNA kontrol bylo vypočítáno procentuálního zastoupení methylované a nemethylované DNA v jednotlivých eluovaných frakcích.

Informace ze sekvenátoru ukázaly, že jednotlivé vzorky v analýze byly osekvenovány rovnocenně a nedocházelo tak k tvorbě artefaktů, přičemž 97,28 % všech sekvenčních čtení mělo hodnotu kvalitu vyšší než Q30 a celkový výtěžek byl 24,74 Gb. V programu MACS2 byl proveden peak calling pro porovnání dat vzorků vůči kontrolním hodnotám, pomocí kterého byly detekovány všechny methylované oblasti. Z těchto výsledků lze usuzovat, že započítáním počtu a délky všech sekvenčních čtení, by byl celý genom osekvenován celkem 1/2x.

Následně bylo v programu MACS2 provedeno i porovnání jednotlivých experimentálních skupin mezi sebou, čímž byly sledovány rozdíly v methylaci na základě rozdílného pokrytí. Pro tento krok byla data získaná ze sekvenování jednotlivých vzorků sloučena do větších souborů v rámci příslušné skupiny. Z těchto hodnot vyplývá, že v každém vzorku bylo detekováno nejméně 200 000 oblastí DNA modifikovaných methylací, přičemž ve skupině B bylo těchto míst detekováno nejvíce. Buněčná linie, ze které byla tato DNA izolována, byla ošetřena pomocí decitabinu, demethylační látky, a očekávaným výsledkem byla tedy detekce nižšího počtu píků. Tento jev lze pravděpodobně vysvětlit tím, že decitabin je schopen se inkorporovat do struktury DNA, a tak může fungovat jako falešně „methylovaná“ DNA (Veselý *et* Čihák, 1977). Během vychytávání methylované DNA pomocí MBD proteinů se na tyto proteiny vedle methylovaných cytosinů vázala i DNA s inkorporovaným 5-aza-2-

deoxycytidinem. Tato hypotéza je dále podpořena faktem, že decitabin zabudovaný do DNA může také vytvářet komplex s DNA methyltransferázou (Santi *et al.*, 1984).

Pomocí programu naskriptovaném v jazyce Python byly detekované methylované oblasti přiřazeny ke genům ležícím v nejbližší oblasti. Tímto způsobem byly anotovány všechny píky vygenerované v programu MACS2 prostřednictvím funkce peak calling. V bakalářské práci byly uvedeny příklady anotace 10 piků ze skupiny A a C po vzájemném porovnání. Tyto methylované oblasti byly z dat vyfiltrovány podle dvou kritérií. Prvním z podmínek bylo, že methylovaná oblast leží ve vzdálenosti menší než 2000 bp. Píky pak byly seřazeny podle velikosti q-value, respektive od největší hodnoty po nejmenší, přičemž byly odfiltrovány všechny modifikované oblasti s nižší hodnotou q než 6. Dalším stupněm generování genomických dat by pak mohlo být filtrování dle biologických procesů, do kterých se příslušné geny zapojují. Tímto způsobem anotace genů se zabývá sdružení zvané Genová ontologie (Gene Ontology, GO), které se snaží definovat koncepty používané pro popisování funkce genů a charakterizovat vztahy mezi nimi. Genová ontologie klasifikuje funkce genů podle třech aspektů: (1) *molekulární funkce* – molekulární aktivita produktů genů, (2) *buněčné komponenty* – místo, kde jsou produkty genů aktivní, (3) *biologické procesy* – dráhy a větší procesy, které jsou tvořeny aktivitou několika genových produktů. Všechny tyto informace jsou shromažďovány v databázi GeneCards, která automaticky propojuje data přibližně ze 125 internetových zdrojů.

Nezveřejněné, důvěrné.

Detekce vzorů methylace DNA je velmi rychle rozvíjenou oblastí výzkumu rakovinných onemocnění, přičemž metody profilování methylace mají velký potenciál pro identifikaci a charakterizaci aberací způsobující tyto choroby a umožňují studium vlivu chemoterapeutických látek na genom buněk při jejich léčbě. Jednotlivé metody vyvinuté pro analýzu methylomu lze rozdělit do 3 skupin: (1) chemická modifikace prostřednictvím hydrogensířičitanu sodného, (2) digesce restrikčními enzymy a (3) izolace methylované DNA založená na afinitě 5-methylcytosinu. Metoda Methyl-Capture sekvenace (MC-Seq), jejíž

princepem je afinita 5-methylcytosinu k proteinům, kombinuje techniku precipitace methylované DNA pomocí rekombinantní methyl-CpG-vazebné domény proteinu MBD2 s NGS. Jelikož velké množství metod má nevýhodu nízkého pokrytí genomu (například microarray), MC-Seq se zdá jako atraktivní řešení umožňující zvýšení rozsahu analýzy, které je zároveň oproti dalším metodám detekce celogenomových profilů methylace DNA (například WGBS) časově i finančně méně náročná. Stejně jako MC-Seq také metody RRBS a MeDIP-Seq nevyžadují vysoké finanční náklady, avšak tyto postupy jsou oproti MC-Seq limitovány na analýzu takové části methylované DNA, která je vyprodukovaná pomocí použitých restričních enzymů a protilátek (Teh *et al.*, 2016).

10 Závěr

Teoretická část této bakalářské práce byla vypracována na základě dostupné odborné literatury a byly v ní shrnuty poznatky z oblasti methylace DNA s důrazem na roli epigenetických modifikací v karcinogenezi. Popsány byly také základní principy a možnosti sekvenování nové generace na platformě Illumina, vybrané metody detekce methylačních profilů buněk a závěrem byla věnována pozornost charakteristice vybraných formátů výstupních souborů, které slouží k analýze sekvenačních dat.

V experimentální části byla zkoumána možnost detekce methylačních profilů buněk prostřednictvím metody Methyl-Capture sekvenování, kde pro vychytávání methylovaných oblastí byl použit MethylMiner Methylated DNA Enrichment Kit (Invitrogen) a pro samotné sekvenování byly použity reagentie od firmy Illumina (HiSeq Rapid SBS Kit). K analýze byla použita mateřská buněčná linie rakoviny kolorekta HCT116 a její odvozené linie HCT116 DAC, R1.4 a R1.4 JQ1. Výsledky ukázaly, že během sekvenování nedocházelo ke vzniku artefaktů a sekvenace probíhala u všech vzorků rovnocenně, přičemž více než 97.28 % všech čtení mělo kvalitu vyšší než Q30. Celkový výtěžek sekvenační analýzy byl 24,74 Gb. V programu MACS2, prostřednictvím funkce peak calling, byly z těchto čtení vygenerovány píky odpovídající methylovaným oblastem DNA. Z normalizovaných hodnot počtu detekovaných píků na 1 M sekvenčních čtení bylo zjištěno, že mezi těmito hodnotami v analyzovaných vzorcích neexistuje přímá spojitost. Pro souhrnná data všech experimentálních skupin bylo také v programu MACS2 provedeno porovnání jednotlivých skupin vůči sobě. Tyto vygenerované oblasti methylace byly následně v programu naskriptovaném v jazyce Python anotovány k nejbližše ležícím genům, kdy okraj píků byl přiřazen k transkripčnímu počátku nejbližšího genu. V práci je prezentováno 10 vybraných genů, které jsou asociovány s píky vygenerovanými během porovnání skupiny A a C a pro které platí, že se vyskytují v nejbližší blízkosti od transkripčního počátku daného genu a zároveň hodnota kvality píku je vyšší než 6.

Nezveřejněné, důvěrné.

V této práci byly pomocí metody Methyl-Capture sekvenování detekovány methylační profily buněčných linií HCT116. Geny, které byly pomocí tohoto screeningu identifikovány, jsou předmětem pro validaci, na které spolupracuje kolektiv pracovníků Ústavu molekulární a translační medicíny a která prověřuje relevantnost těchto genů pro navozování a udržování rezistence buněčných linií HCT116 vůči demethylační látce decitabinu.

11 Zdroje

- Aapola, U., Shibuya, K., Scott, H. S., Ollila, J., Vihinen, M., Heino, M., ... Peterson, P. (2000). Isolation and initial characterization of a novel Zinc finger gene, DNMT3L, on 21q22.3, Related to the Cytosine-5-Methyltransferase 3 Gene Family. *Genomics*, 65(3), 293-298.
- Agrawal, K., Holub, D., Vojta, P., Frydrych, I., Maceckova, Z., Otmar, M., Džubák, P., Hajdúch, M. (2014). 5-azacytidine nucleosides and their derivatives: Molecular hallmarks of drug resistance & alternative therapeutic regimen. In: *DDPEO* [online] Olomouc, 2014. [cit. 25. 11. 2015]. Dostupné z: http://www.ddpeo.cz/img/abstracts/pdf/ddpeo_2014.pdf
- Agrawal, K., Dzubak, P., Hajduch, M., Frydrych, I. (2015). Method of predicting the tumor response to DNA methylation inhibitors and alternative therapeutic regimens for overcoming resistance. Palacky university olomouc. *CZ Patent EP2924125 (A1)*.
- Antonarakis, S. E. (1994). Genome linkage scanning: systematic or intelligent? *Nature Genetics*, 8(3), 211-212.
- Aoki, A., Suatake, I., Miyagawa, J., Fujio, T., Sasaki, H., & Tajima, S. (2001). Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Research*, 29(17), 3506-3512.
- Ariel, M., Robinson, E., McCarrey, J. R., & Cedar, H. (1995). Gamete-specific methylation correlates with imprinting of the murine Xist gene. *Nature Genetics*, 9(3), 312-315.
- Bae, S. I., Lee, H. S., & Kim, S. H. (2002). Inactivation of O6-methylguanine-DNA methyltransferase by promoter CpG island hypermethylation in gastric cancers. *British Journal of Cancer*, 86, 1888-1892.
- BAM File Format. (n. d.) In: *BaseSpace Support* [online]. [cit. 2016-03-25]. Dostupné z: <https://support.basespace.illumina.com/knowledgebase/articles/144843-bam-file-format>
- Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M., & Issa, J. (1997). Alterations in DNA methylation: A fundamental aspect of neoplasia. *Advances in Cancer Research*, 72, 141-196.
- Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., ... Herman, J. G. (1998). Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proceedings of the National Academy of Sciences*, 95(20), 11891-11896.
- Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Human Molecular Genetics*, 9(16), 2395-2402.
- Bhutani, N., Burns, D. M., & Blau, H. M. (2011). DNA demethylation dynamics. *Cell*, 146(6), 866-872.
- Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucl Acids Res*, 8(7), 1499-1504.
- Bird, A. P., & Wolffe, A. P. (1999). Methylation-induced repression - belts, braces, and chromatin. *Cell*, 99(5), 451-454.
- Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1), 91-99.
- Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian Cells. *Science*, 317(5845), 1760-1764.
- Boyes, J., & Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell*, 64(6), 1123-1134.
- Boyes, J., & Bird, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J*, 11, 327-33.

- Brenton, J. D., Viville, S., & Surani, M. A. (1995). Genomic imprinting and cancer. *Cancer Surv*, 25, 161-171.
- Brinkman, A. B., Simmer, F., Ma, K., Kaan, A., Zhu, J., & Stunnenberg, H. G. (2010). Whole-genome DNA methylation profiling using MethylCap-seq. *Methods*, 52(3), 232-236.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., ... Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, 38(6), 626-635.
- Cattanach, B. M., & Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*, 315(6019), 496-498.
- Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*, 10(5), 295-304.
- Cheng, X. (1995). Structure and function of DNA methyltransferases. *Annual Review of Biophysics and Biomolecular Structure*, 24(1), 293-318.
- Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., ... Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature Genetics*, 39(3), 391-396.
- Clark, S. J., Harrison, J., & Molloy, P. L. (1997). Sp1 binding is inhibited by mCpG methylation. *Gene*, 195(1), 67-71.
- Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6), 1767-1771.
- Coulondre, C., Miller, J. H., Farabaugh, P. J., & Gilbert, W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. *Nature*, 274(5673), 775-780.
- Daniel, J. M., Spring, C. M., Crawford, H. C., Reynolds, A. B., & Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Research*, 30(13), 2911-2919.
- Divoký, K. (2016a). MiSeq sekvenátor – Illumina. [online]. [cit. 2016-03-25]. Dostupné z: <http://www.genetica.cz/miseq-sekvenator#.VtNJhoSzXIU>
- Divoký, K. (2016b). HiSeq 2500 sekvenátor – Illumina. [online]. [cit. 2016-03-25]. Dostupné z: <http://www.genetica.cz/hiseq-2500-sekvenator#.VtmsAYSzXIV>
- Ehrlich, M., Gama-Sosa, M. A., Huang, L., Midgett, R. M., Kuo, K. C., McCune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, 10(8), 2709-2721.
- El-Deiry, W. S., Nelkin, B. D., Celano, P., Yen, R. W., Falco, J. P., Hamilton, S. R., & Baylin, S. B. (1991). High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer. *Proceedings of the National Academy of Sciences*, 88(8), 3470-3474.
- Feltus, F. A., Lee, E. K., Costello, J. F., Plass, C., & Vertino, P. M. (2003). Predicting aberrant CpG island methylation. *Proceedings of the National Academy of Sciences*, 100(21), 12253-12258.
- Filion, G. J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., & Defossez, P. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and Cellular Biology*, 26(1), 169-181.

- Fremont, M., Siegmann, M., Gaulis, S., Matthies, R., Hess, D., & Jost, J. (1997). Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA. *Nucleic Acids Research*, 25(12), 2375-2380.
- Fuks, F., Godin, W. Q., Kasai, N., & Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *The EMBO Journal*, 20(10), 2536-2544.
- Fuks, F., Hurd, P. J., Deplus, R., & Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research*, 31(9), 2305-2312.
- Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, 105(1), 4-13.
- Grada, A., & Weinbrecht, K. (2013). Next-generation sequencing: Methodology and application. *Journal of Investigative Dermatology*, 133(8), 1-4.
- Graff, J. R., Herman, J. G., Lapidus, R. G., Chopra, H., Xu, R., Jarrard, D. F., ... Isaacs, W. B. (1995). E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Research*, 55, 5195-5199.
- Hashimoto, H., Horton, J. R., Zhang, X., Bostick, M., Jacobsen, S. E., & Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, 455(7214), 826-829.
- Hashimoto, H., Horton, J. R., Zhang, X., & Cheng, X. (2009). UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics*, 4(1), 8-14.
- Herman, J. G., Latif, F., Weng, Y., Lerman, M. I., Zbar, B., Liu, S., ... Linehan, W. M. (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proceedings of the National Academy of Sciences*, 91(21), 9700-9704.
- Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. J., ... Baylin, S. B. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proceedings of the National Academy of Sciences*, 95(12), 6870-6875.
- Horii, T., & Hat, I. (2013). Epigenetic instability in embryonic stem cells. *Pluripotent Stem Cells*. [online] [cit. 2016-01-7]. Dostupné z: <http://www.intechopen.com/books/pluripotent-stem-cells/epigenetic-instability-in-embryonic-stem-cells>
- Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *Biol Chem*, 175, 315-332.
- Hug, M., Silke, J., Georgiev, O., Rusconi, S., Schaffner, W., & Matsuo, K. (1996). Transcriptional repression by methylation: cooperativity between a CpG cluster in the promoter and remote CpG-rich regions. *FEBS Letters*, 379(3), 251-254.
- Illumina. (2009). FASTQ files. [online] [cit. 2016-01-7] Dostupné z: http://support.illumina.com/help/SequencingAnalysisWorkflow/Content/Vault/Informatics/Sequencing_Analysis/CASAVA/swSEQ_mCA_FASTQFiles.htm
- Illumina. (2015). Field guide methylation. Files. [online] [cit. 2016-01-7] Dostupné z: http://www.illumina.com/content/dam/illumina-marketing/documents/products/other/field_guide_methylation.pdf

- Illumina. (2016a). An introduction to next-generation sequencing technology [online] [cit. 2016-01-7] Dostupné z: http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf
- Illumina. (2016b). HiSeq 3000/HiSeq 4000 applications. [online] [cit. 2016-01-7] Dostupné z: <http://www.illumina.com/systems/hiseq-3000-4000/applications.html>
- Illumina. (2016c). Next-Generation Sequencing Platforms. Compare NGS Platforms. [online] [cit. 2016-01-7] Dostupné z: <http://www.illumina.com/systems/sequencing-platform-comparison.html>
- Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., ... Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, 41(2), 178-186.
- Jaenisch, R., & Jähner, D. (1984). Methylation, expression and chromosomal position of genes in mammals. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 782(1), 1-9.
- Jones, P. A. (1999). The DNA methylation paradox. *Trends in Genetics*, 15(1), 34-37.
- Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nature Genetics*, 21(2), 163-167.
- Laird, P. W., & Jaenisch, R. (1996). The role of DNA methylation in cancer genetics and epigenetics. *Annual Review of Genetics*, 30(1), 441-464.
- Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6), 915-926.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., ... Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315-322.
- Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., ... Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 466(7303), 253-257.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Embryogenesis: Demethylation of the zygotic paternal genome. *Nature*, 403(6769), 501-502.
- McCarty, M. (1946). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: II. effect of deoxyribonuclease on the biological activity of the transforming substance. *Journal of Experimental Medicine*, 83(2), 89-96.
- Monk, M., Boubelik, M., & Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, 99, 371-382.
- Moore, L. D., Le, T., & Fan, G. (2012). DNA Methylation and its basic function. *Neuropsychopharmacology*, 38(1), 23-38.
- Myohanen, S. K., Baylin, S. B., & Herman, J. G. (1998). Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia. *Cancer Research*, 58, 591-593.
- Nan, X., Campoy, F., & Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, 88(4), 471-481.

- Nan, X., Meehan, R. R., & Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research*, *21*(21), 4886-4892.
- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, *393*(6683), 386-389.
- National Library of Medicine. (2016). *SEPT9 - septin 9 - Genetics Home Reference*. [online] [cit. 2016-01-7] Dostupné z: <http://ghr.nlm.nih.gov/gene/SEPT9>
- Newell-Price, J., Clark, A. J., & King, P. (2000). DNA methylation and silencing of gene expression. *Trends in Endocrinology & Metabolism*, *11*(4), 142-148.
- Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., ... Tempst, P. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature Genetics*, *23*(1), 58-61.
- Okano, M. (1998a). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Research*, *26*(11), 2536-2540.
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b Are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247-257.
- Okano, M., Xie, S., & Li, E. (1998b). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, *19*(3), 219-220.
- Olsson, M., & Lindahl, T. (1980). Repair of alkylated DNA in Escherichia coli. Methyl group transfer from O6-methylguanine to a protein cysteine residue. *The Journal of Biological Chemistry*, *255*(22), 10569-71.
- Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., ... Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Current Biology*, *10*(8), 475-478.
- Paroush, Z., Keshet, I., Yisraeli, J., & Cedar, H. (1990). Dynamics of demethylation and activation of the α -actin gene in myoblasts. *Cell*, *63*(6), 1229-1237.
- Pazin, M. J., & Kadonaga, J. T. (1997). What's up and down with histone deacetylation and transcription? *Cell*, *89*(3), 325-328.
- Pecorino, L. (2005). *Molecular biology of cancer: Mechanisms, targets, and therapeutics* (3rd ed.). Oxford, CZ: Oxford University Press.
- Pentagen. (n.d). Epi proColon 2.0 Druhá generace testů pro Septin 9 [cit. 2016-10-02] Dostupné z: http://www.pentagen.cz/files/letak_navod_epigenomics/proColon-Epi.pdf
- Philips, T., & Lobo, I. (2008). Genetic Imprinting and X Inactivation. *Nature Education*, *1*(1), 117.
- Prendergast, G., & Ziff, E. (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science*, *251*(4990), 186-189.
- Prokhortchouk, A., Hednrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Geogiev, G., ... Bird, A. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & Development*, *15*(13), 1613-1618.
- Prowse, A. H., Webster, A. R., Richards, F. M., Richard, S., Olschwang, S., Resche, F., ... Affara, N. A. (1997). Somatic inactivation of the VHL gene in Von Hippel-Lindau disease tumors. *H. Genetics*, *60*, 765-771.
- Razin, A., & Riggs, A. (1980). DNA methylation and gene function. *Science*, *210*(4470), 604-610.

- Ragione, F. D., Filosa, S., Scalabri, F., & D'Esposito, M. (2012). MeCP2 as a genome-wide modulator: the renewal of an old story. *Front. Gene*, 3.
- Reik, W. (1989). Genomic imprinting and genetic disorders in man. *Trends in Genetics*, 5, 331-336.
- Robertson, K. D., Keyomarsi, K., Gonzales, F. A., Velicescu, M., & Jones, P. A. (2000). Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G0/G1 to S phase transition in normal and tumor cells. *Nucleic Acids Research*, 28(10), 2108-2113.
- Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L., & Almouzni, G. (2015). Epigenomics: Roadmap for regulation. *Nature*, 518(7539), 314-316.
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, 8(6), 424-436.
- Sanger, F., & Coulson, A. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441-448.
- Santi, D. V., Norment, A., & Garrett, C. E. (1984). Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proceedings of the National Academy of Sciences*, 81(22), 6993-6997.
- Sarraf, S. A., & Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell*, 15(4), 595-605.
- Schulz, W. A., Steinhoff, C., & Florl, A. R. (n.d.). Methylation of Endogenous Human Retroelements in Health and Disease. *Current Topics in Microbiology and Immunology*, 211-250.
- Singal, R., Wang, S. Z., Sargent, T., Zhu, S. Z., & Ginder, G. D. (2002). Methylation of promoter proximal-transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. *Journal of Biological Chemistry*, 277, 1897-1905.
- Slavkovský, R. (2016). Ústní sdělení. [Ústav molekulární a translační medicíny, Hněvotínská 1333/5, 779 00 Olomouc, Česká republika] [cit. 2016-04-24]
- Suetake, I., Miyazaki, J., Murakami, C., Takeshima, H., & Tajima, S. (2003). Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. *Journal of Biochemistry*, 133(6), 737-744.
- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., & Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *Journal of Biological Chemistry*, 279(26), 27816-27823.
- Suzuki, H., Itoh, F., Toyota, M., Kikuchi, T., Kakiuchi, H., Hinoda, Y., & Imai, K. (1999). Distinct methylation pattern and microsatellite instability in sporadic gastric cancer. *International Journal of Cancer*, 83(3), 309-313.
- Suzuki, K., Suzuki, I., Leodolter, A., Alonso, S., Horiuchi, S., Yamashita, K., & Perucho, M. (2006). Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. *Cancer Cell*, 9(3), 199-207.
- Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, 9(6), 465-476.
- Tazi, J., & Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell*, 60(6), 909-920.

- Teh, A. L., Pan, H., Lin, X., Lim, Y. I., Patro, C. P., Cheong, C. Y., ... Karnani, N. (2016). Comparison of methyl-capture sequencing vs. Infinium 450K methylation array for methylome analysis in clinical samples. *Epigenetics*, *11*(1), 36-48.
- The SAM/BAM Format Specification. (2015). Sequence Alignment/Map format specification. [online] [cit. 25. 11. 2015] Dostupné z: <http://samtools.github.io/hts-specs/SAMv1.pdf>
- Tycko, B. (2000). Epigenetic gene silencing in cancer. *Journal of Clinical Investigation*, *105*(4), 401-407.
- Ura, K., Kurumizaka, H., Dimitrov, S., Almouzni, G., & Wolffe, A. P. (1997). Histone acetylation: influence on transcription, nucleosome mobility and positioning, and linker histone-dependent transcriptional repression. *The EMBO Journal*, *16*(8), 2096-2107.
- Vaiopoulos, A. G., Athanasoula, K. C., & Papavassiliou, A. G. (2014). Epigenetic modifications in colorectal cancer: Molecular insights and therapeutic challenges. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1842*(7), 971-980.
- Van Dongen, J., Ehli, E., Sliker, R., Bartels, M., Weber, Z., Davies, G., ... Boomsma, D. (2014). Epigenetic variation in monozygotic twins: A genome-wide analysis of DNA methylation in buccal cells. *Genes*, *5*(2), 347-365.
- Veselý, J., & Čihák, A. (1977). Incorporation of a potent antileukemic agent, 5-Aza-2'- deoxycytidine, into DNA of cells from leukemic mice. *Cancer Research*, *37*, 3684-3689.
- Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: From basic research to diagnostics. *Clinical Chemistry*, *55*(4), 641-658.
- Wade, P. A., Geronzi, A., Jones, P. L., Ballestar, E., Aubry, F., & Wolffe, A. P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nature Genetics*, *23*(1), 62-66.
- Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, *20*(2), 116-117.
- Watanabe, D., Suetake, I., Tada, T., & Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of Development*, *118*(1-2), 187-190.
- Watt, F., & Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & Development*, *2*(9), 1136-1143.
- Wigler, M., Levy, D., & Perucho, M. (1981). The somatic replication of DNA methylation. *Cell*, *24*(1), 33-40.
- Wilson, G. G., & Murray, N. E. (1991). Restriction and modification Systems. *Annual Review of Genetics*, *25*(1), 585-627.
- Wolffe, A. P. (1997). Sinful repression. *Nature*, *387*(6628), 16-17.
- Woodcock, D., Crowther, P., & Diver, W. (1987). The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide. *Biochemical and Biophysical Research Communications*, *145*(2), 888-894.
- Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W., ... Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, *236*(1), 87-95.
- Yen, R. C., Vertino, P. M., Nelkin, B. D., Yu, J. J., El-Deiry, W., Cumaraswamy, A., ... Baylin, S. B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucl Acids Res*, *20*(9), 2287-2291.

- Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, *13*(8), 335-340.
- Yoshiura, K., Kanai, Y., Ochiai, A., Shimoyama, Y., Sugimura, T., & Hirohashi, S. (1995). Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proceedings of the National Academy of Sciences*, *92*(16), 7416-7419.
- Zhang, F., Pomerantz, J. H., Sen, G., Palermo, A. T., & Blau, H. M. (2007). Active tissue-specific DNA demethylation conferred by somatic cell nuclei in stable heterokaryons. *Proceedings of the National Academy of Sciences*, *104*(11), 4395-4400.
- Zhu, Y., Cantor, C. R., & Smith, C. L. (1993). DNA Sequence analysis of human chromosome 21 NotI linking clones. *Genomics*, *18*(2), 199-205.

12 Seznam zkratek

ALAD	Aminolevulinatová dehydratase
ALS2CL	ALS2 C-terminal like
ATM	Amplicon Tagment Mix, 96 RXN
AzaC	5-Aza-2'-deoxycytidine
BER	Base excision repair
bp	Páry bází
BWA	Burrows-Wheeler alignment
C	Cytosin
C-DMR	Cancer-specific differentially methylated region
CGI	CpG ostrůvky
CFAP58	Cilia and flagella associated protein 58
CpG	Cytosine-phosphate-guanine
CRC	Colorectal cancer
CSTF2	Cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa
DAC	Decitabine
ddNTP	Dideoxynukleotid
DEPC	Diethyl pyrokarbonát
DNA	Deoxyribonukleová kyselina
DNMT	DNA methyltransferáza
DNMT1	DNA methyltransferáza 1
DNMT2	DNA methyltransferáza 2
DNMT3A	DNA methyltransferáza 3A
DNMT3B	DNA methyltransferáza 3B
DNMT3L	(cytosine-5)-methyltransferase 3-like
dNTP	Deoxynukleotid
ERRFI	ERBB Receptor feedback inhibitor 1
G	Guanin
Gb	Giga báze
Gbp	Giga basepair
GO	Genová ontologie
ChIP-Seq	Chromatin immunoprecipitation sequencing
H ₂ O	Voda
H3	Histon 3
H4	Histon 4
HCT	Human colorectal cancer cells
HDAC	Histonová deacetyláza
HDAC1	Histonová deacetyláza 1
HDAC2	Histonová deacetyláza 2

hg19	Human Genome 19
hMLH1	Human mutL homolog 1
HT1	Hybridization buffer
IGV	Integrative Genomics Viewer
IYD	Iodotyrosine deiodinase
JQ1	Thienotriazolodiazepine
KISS1	KiSS-1 metastasis-suppressor
KRT39	Keratin 3, Type I
LINE	Long interspersed element
LOC400794	Uncharacterized LOC400794
M	Mega (10^6)
MACS	Model-based analysis of ChIP-Seq
MC-Seq	Methyl-capture sequencing
MBD	Methyl-CpG-binding domain
MBD1	Methyl-CpG-binding domain 1
MBD2	Methyl-CpG-binding domain 2
MBD3	Methyl-CpG-binding domain 3
MBD4	Methyl-CpG-binding domain 4
MBP	Methyl-binding protein
MeCP1	Methyl-CpG-binding domain protein 1
MeCP2	Methyl-CpG-binding domain protein 2
MeDIP	Methylated DNA immunoprecipitation sequencing
MGMT	O6-methylguanin DNA methyltransferáza
MetCap-Seq	Methyl capture sequencing
MethylCap-Seq	Methyl capture sequencing
MC-Seq	Methyl capture sequencing
min	Minuta
min.	minimálně
mSin3A	SIN3 transcription regulator family member A
MTA	Metastasis-associated protein
MTERF1a	Mitochondrial transcription termination factor 1
NaOH	Hydroxid sodný
ng	Nanogram
NGS	Sekvenování nové generace
NPM	Nextera PCR master mix
NT	Neutralize tagment buffer
NTN4	Netrin 4
Obr.	Obrázek
PCR	Polymerázová řetězcová reakce
PF	Passing Filters

pg	Pikogram
PHD	Plant homeodomain
qPCR	Kvantitativní polymerázová řetězcová reakce
R	Rezistentní
Rb	Retinoblastoma
RbAp46	Retinoblastoma-associated proteins 46
RbAp48	Retinoblastoma-associated proteins 48
RING	Really interesting new gene
RNA	Ribonukleová kyselina
RRAGC	Ras-related GTP binding C
RRBS	Reduced representation bisulfite sequencing
RRBS-Seq	Reduced representation bisulfite sequencing
SAM	S-Adenosylmethionine
SBS	Sequencing by synthesys
SDHAF3	Succinate dehydrogenase complex assembly factor 3
SEPT9	Septin 9
SQ	Starting quantity
Tab.	Tabulka
Tab-Seq	Tet-assisted bisulfite sequencing
TD	Tagment DNA buffer
T-DMR	Tissue-specific differentially methylated region
Tet	Ten-eleven translocation enzyme
TGM2	Transglutaminase 2
TMEM229	Transmembrane protein 229
TREML3P	Triggering receptor expressed on myeloid cells-like 3, pseudogene
TRD	Transcriptional repressor domain
Tx	Transcription
UCSC	The University of California Santa Cruz
UHFR1	Ubiquitin-like with PHD and ring finger domains 1
UHRF2	Ubiquitin-like with PHD and ring finger domains 2
VHL	Von Hippel-Lindau tumor suppressor
WGSB	Whole-genome bisulfite sequencing
WGSB-Seq	Whole-genome bisulfite sequencing
ZBTB4	Zinc finger and BTB domain containing 4
ZBTB38	Zinc finger and BTB domain containing 38
ZNF812	Zinc finger protein 812