

STŘEDOŠKOLSKÁ ODBORNÁ ČINNOST

Obor č. 06: Zdravotnictví

Analýza genetického profilu leukemických buněk

Markéta Protivánková
Jihomoravský kraj

Brno 2021

STŘEDOŠKOLSKÁ ODBORNÁ ČINNOST

Obor č. 06: Zdravotnictví

Analýza genetického profilu leukemických buněk

Analysis of the genetic profile of leukemic cells

Autoři: Markéta Protivánková

Škola: Gymnázium Brno, Křenová 36, 602 00

Kraj: Jihomoravský

Konzultant: Mgr. Kristýna Závacká

Brno 2021

Prohlášení

Prohlašuji, že jsem svou práci SOČ vypracovala samostatně a použila jsem pouze prameny literaturu uvedené v seznamu bibliografických záznamů.

Prohlašuji, že tištěná verze a elektronická verze soutěžní práce SOČ jsou shodné.

Nemám závažný důvod proti zpřístupnění této práce v souladu se zákonem č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších předpisů.

V Brně

.....
Markéta Protivánková

Poděkování

Mé velké poděkování patří především Mgr. Kristýně Závacké, mé konzultantce, za možnost vypracovat svoji Středoškolskou odbornou činnost pod jejím laskavým vedením. Velice si vážím příležitosti s ní spolupracovat, její obětavé pomoci, ochoty a úsilí, se kterými se mi vždy věnovala.

Dále bych chtěla poděkovat celému kolektivu z výzkumné skupiny Lékařská genomika v CEITEC MU za vstřícnost a přátelský přístup. Zvláště bych chtěla poděkovat Mgr. Ing. Jakubu Pawłu Porcovi, kterému vděčím za zpracování a analýzu získaných dat. Dále bych chtěla poděkovat Mgr. Karle Plevové, Ph.D. za cenné rady při závěrečném sepisování práce a Mgr. Haně Svozilové za poskytnutí buněčné linie a pomoc s experimentální částí mé práce. Závěrem bych chtěla poděkovat své rodině, bez jejíž podpory, pochopení a pomoci by tato práce nevznikla.

Anotace

Chronická lymfocytární leukémie (CLL) je nejčastěji diagnostikovanou leukémií dospělých. Vyznačuje se velmi variabilním průběhem, a proto je po celém světě vyvíjena snaha identifikovat genetické změny předpovídající vývoj onemocnění. Cílem projektu bylo provést genomickou analýzu buněčné linie HG-3, která je odvozená od CLL, pomocí metody Hi-C. Jedná se o revoluční přístup v oblasti nádorové genomiky, který umožňuje odhalit interakce chromatinu napříč celým genomem. Metoda je vhodná pro studium regulace genové exprese a umožňuje detekovat strukturní změny genomu a jejich funkční dopad. Tím poskytuje nové informace a lepší pochopení komplexity nádorových genomů.

Klíčová slova

genom; interakce chromatinu; Hi-C; HG-3 buněčná linie; chronická lymfocytární leukémie

Annotation

Chronic lymphocytic leukemia (CLL) is the most common adult leukemia. It is characterised by a very variable disease course. Therefore, efforts are being made around the world to identify genetic changes associated with the development of the disease. The aim of the project was to perform the genomic analysis of the HG-3 cell line derived from CLL using a Hi-C method. Hi-C is a revolutionary approach in the field of tumor genomics, which enables to detect chromatin interactions across the whole genome. It is suitable for the study of regulatory mechanisms of gene expression and the detection of genomic structural changes with their functional impact. This provides new information and better understanding of the complexity of tumor genomes.

Keywords

genome; chromatin interactions; Hi-C; HG-3 cell line; chronic lymphocytic leukemia

Obsah

1	Úvod.....	8
2	Cíle práce.....	9
3	Studium genomu	10
3.1	Genomika.....	10
3.2	Struktura genomu	10
3.2.1	Geny	11
3.2.2	Genové mutace	12
3.3	Struktura chromatinu	12
3.4	Remodelace chromatinu	13
3.5	Topologicky asociované domény	13
4	Metody studia chromatinových interakcí	15
4.1	3C a její modifikace.....	16
4.2	Hi-C	17
4.3	ChIP-Loop	18
4.4	ChIP-seq	18
4.5	ChIA-PET	19
5	Chronická lymfocytární leukémie	20
5.1	Charakteristika chronické lymfocytární leukémie, diagnóza a výskyt.....	20
5.2	Patogeneze onemocnění	21
5.3	Prognóza	22
5.4	Léčba	23
5.5	Buněčná linie HG-3	24
6	Materiál	26
7	Metody	27
7.1	Použité chemikálie:.....	27
7.2	Použité přístroje:.....	29
7.3	EpiTect Hi-C Kit	29
7.3.1	Krok 1: Stabilizace chromatinu a buněčná lyze	29
7.3.2	Krok 2: Značení biotinem a rozrušení crosslinků	29
7.3.3	Krok 3: Příprava Hi-C knihovny	29
7.4	Dovetail™ Micro-C Kit.....	30
7.4.1	Krok 1: Stabilizace chromatinu a buněčná lyze	30

7.4.2	Krok 2: Značení biotinem a rozrušení crosslinků	30
7.4.3	Krok 3: Příprava Hi-C knihovny	30
7.5	Dovetail™ Omni-C Kit	31
7.5.1	Krok 1: Stabilizace chromatinu a buněčná lyze	31
7.5.2	Krok 2: Značení biotinem a rozrušení crosslinků	31
7.5.3	Krok 3: Příprava Hi-C knihovny	31
7.6	Sekvenování.....	32
7.7	Analýza dat	32
8	Výsledky.....	33
8.1	Kontrola fragmentačních profilů	33
8.2	Kontrola kvality sekvenačních knihoven a kvantifikace na Qubit® fluorometru	35
8.3	Kontrola kvality sekvenačních dat	36
8.4	Vizualizace chromatinových interakcí	37
8.5	Strukturní varianty	41
9	Diskuze.....	43
10	Závěr.....	44
11	Seznam použité literatury	45
12	Seznam použitých zkratk.....	50

1 ÚVOD

Nádorová onemocnění patří mezi nejčastější příčiny úmrtí na celém světě a jejich incidence stále stoupá. Široké spektrum nádorových onemocnění představují hematologické malignity, pod které spadají mimo jiné také leukémie. Chronická lymfocytární leukémie (CLL) je nejčastějším druhem leukémie dospělých pacientů v západním světě. Charakteristickým znakem onemocnění je velmi variabilní průběh, který znemožňuje pacienty léčit univerzálním přístupem. Cílem mnoha výzkumných projektů, včetně našeho, je pochopení abnormálních molekulárních procesů probíhajících v leukemických buňkách CLL.

Pro analýzu leukemických buněk byla zvolena metoda Hi-C, která umožňuje popsat interakce chromatinu a strukturní aberace v rámci celého genomu. Hlavním cílem této práce bylo použít metodu Hi-C na modelové buněčné linii HG-3, odvozené od CLL, a určit její genetické abnormality. V rámci experimentů jsem pracovala se třemi různými komerčně dostupnými sadami chemikálií (kity), které byly využity ke zpracování vzorků. Vytvořené sekvenční knihovny byly podrobeny sekvenování a rozsáhlé bioinformatické analýze. Samotná práce je důležitou součástí většího grantového projektu výzkumné skupiny Lékařská genomika v CEITEC MU a budou na ni navazovat další experimenty zaměřené na optimalizaci metody přímo na patientských vzorcích. Výsledky tohoto projektu by v budoucnu mohly napomoci pochopení genetické a klinické heterogenity CLL, což by umožnilo optimalizaci léčebných postupů.

2 CÍLE PRÁCE

Na různorodém průběhu CLL se do velké míry podílí množství různých genových mutací a strukturních aberací. To komplikuje stanovení prognózy a volbu potenciálně účinné terapie. Komplexní analýza genomu pacientů může poskytnout novou úroveň informace, která přispěje k objasnění molekulárních procesů v nádorových buňkách a pomůže při volbě vhodné léčby.

Pro analýzu leukemických buněk jsme se rozhodli využít moderní metodu Hi-C, která umožňuje stanovit interakce chromatinu a strukturní aberace. Naše cíle byly následující:

- Na buněčné linii HG-3, odvozené od CLL, vyzkoušet přípravu Hi-C sekvenční knihovny pomocí tří různých komerčně dostupných kitů:
 - EpiTect Hi-C Kit (Qiagen)
 - Dovetail™ Micro-C Kit (Dovetail Genomics)
 - Dovetail™ Omni-C Kit (Dovetail Genomics).
- Metodou Hi-C potvrdit genetické abnormality buněčné linie HG-3, zjištěné cytogenetickými metodami.
- Porovnat mezi sebou výsledky získané pomocí jednotlivých kitů.
- Zvolit nejvhodnější přístup pro budoucí analýzy a optimalizovat Hi-C protokol.

Experimenty prováděné v rámci této práce jsou důležitou součástí většího grantového projektu výzkumné skupiny Lékařská genomika v CEITEC MU. Vybraný postup bude v budoucnu využíván pro studium většího souboru patientských vzorků.

3 STUDIUM GENOMU

3.1 Genomika

Genomika je obor genetiky, jež se zabývá studiem genomů různých organismů, konkrétně například sekvenací DNA a genetickým mapováním. Mapování obnáší například zjišťování počtu chromozomů, lokace genů na chromozomech a posloupnosti genů. Genomické přístupy se využívají zejména při studiu komplexních chorob, protože při hledání odpovědí zvažují účast více genů a faktorů. Mezi typické komplexní choroby patří nádorová onemocnění, kde zvrát ve zdravé buňce může být zapříčiněn složitými interakcemi mezi geny navzájem a mezi geny a vnějším prostředím.

Genomika využívá řadu biochemických a molekulárně biologických metod. Mezi nejvíce využívané patří sekvenování DNA, které umožňuje zjistit pořadí nukleových bází v molekulách DNA. Tyto poznatky napomáhají porozumět základním biologickým procesům. S rostoucím zájmem o sekvenační metody roste také jejich nabídka na trhu. Metodické postupy se vylepšují nebo se objevují zcela nové přístupy a mechanismy. Sekvenační metody můžeme rozdělit na klasické metody, metody nové generace a metody třetí generace. Cílem vývoje je snaha zrychlit sekvenování a zároveň snížit náklady na jeho provozování.

Pod klasické sekvenační metody spadá Maxam-Gilbertova metoda sekvenování. Tato chemická metoda je založena na specifickém štěpením řetězce DNA po modifikaci jednotlivých bází chemickými činidly. Naopak Sangerova metoda je metoda enzymatická, založená na ukončení procesu replikace DNA na specifických místech pomocí tzv. terminátorů replikace. V praxi se z klasických metod v současnosti využívá už jen Sangerova metoda.

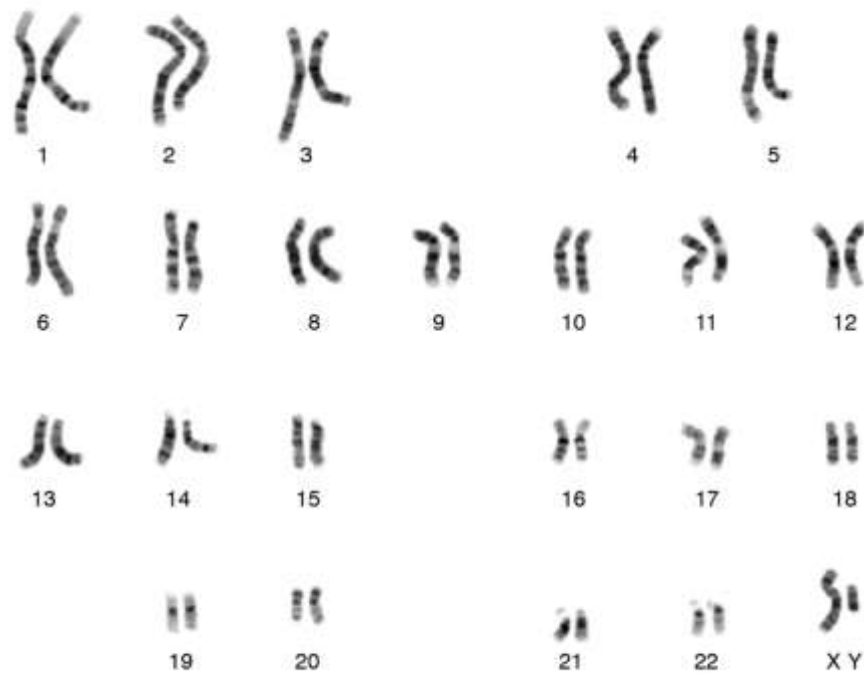
Metody nové generace sekvenování (NGS) byly vyvíjeny s cílem zrychlit a zlevnit proces sekvenování, automatizovat postupy a minimalizovat nevýhody, které měly klasické metody. NGS je založeno na rozdělení genomu na kratší fragmenty, ty se následně masivně paralelně sekvenují, což snižuje náklady na provoz a zvyšuje množství získaných dat v kratším časovém úseku. Patří sem například platformy od firem Roche nebo Illumina.

Nejnovější v praxi používané metody se řadí mezi metody třetí generace sekvenování (TGS). V současnosti neexistuje jednotná definice TGS, avšak dají se rozdělit dle techniky. Množství metod a systémů je stále vyvíjeno. Mezi nejznámější patří HeliscopeTM (Helicos), SMRT (Single molecule real-time) a Oxford nanopore sequencing.

3.2 Struktura genomu

Soubor veškeré genetické informace člověka je označován jako lidský genom. Genetická informace je uložena v DNA, která se nachází především v buněčném jádře, kde tvoří komplexy s proteiny označované jako chromatin. Jaderný genom je uložen v souboru chromozomů, který nazýváme karyotypem. Je tvořen 23 páry chromozomů, které lze rozdělit na 22 párů autozomů a jeden pár heterochromozomů/gonozomů, tedy pohlavních chromozomů. Vedle jaderného

genomu existuje mimojaderný genom, který je tvořen DNA ze semiautonomních organel, v případě lidského genomu z mitochondrií.



Obrázek 1: Lidský karyotyp, 46, XY. Převzato z [1].

3.2.1 Geny

Geny lze chápat jako základní jednotky dědičnosti nebo také jako základní funkční a informační genetické jednotky, protože se skládají ze sekvencí nukleotidů, které kódují primární strukturu RNA nebo proteinů [2]. Ve skutečnosti až 97 % DNA u člověka tvoří nekódující DNA, která neobsahuje informace o RNA ani proteinech, a pouhá 3 % tvoří protein-kódující geny. Určité úseky nekódující DNA mají funkci regulační, avšak u většiny nekódujících oblastí je funkce stále neznámá. Tyto oblasti se často označují jako tzv. „junk“ DNA, česky volně přeloženo jako odpadní DNA. Nicméně její odstranění z genomu mívá negativní následky. Můžeme se však setkat i s nekódující DNA uvnitř genu, tzv. introny, naopak oblasti kódující protein se nazývají exony. Během transkripce dochází k eliminaci intronů při tzv. sestřihu, což je termín pro postsyntetickou úpravu mediátorové RNA (mRNA). Sestřižená mRNA je poté tvořena už jen exony.

Všechny buňky v lidském těle nesou totožnou sadu genů. Buněčné diferenciaci neboli strukturní a funkční specializace je dosaženo aktivací pouze některých genů, které se označují jako tzv. transkripční profil buňky. Proces, při kterém se genetická informace převádí v reálnou buněčnou strukturu nebo funkci, se nazývá genová exprese. Diferencovaná buňka je tedy schopná exprimovat pouze určité geny.

3.2.2 Genové mutace

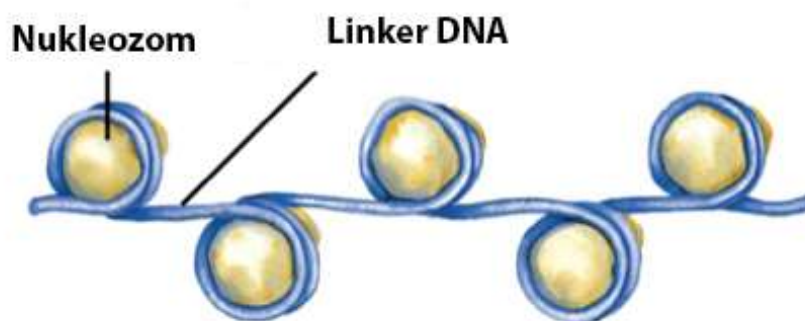
Obecně mutace znamenají změny v genetické informaci. Můžeme je dělit na genomové, chromozomové a genové [3]. Mutace ovlivňující počet některých chromozomů (aneuploidie) nebo celých chromozomálních sad (změny ploidie, např. polyploidie) se nazývají genomové mutace. Chromozomové mutace, respektive aberace, se týkají mutací na chromozomech, které mohou měnit tvar a strukturu chromozomu [4]. Genové mutace probíhají pouze na úrovni jednoho genu a nenarušují stavbu chromozomu.

Genové mutace mohou vnikat samovolně, ale výrazně vyšší je výskyt mutací indukovaných, které vznikají působením různých faktorů (tzv. mutageny). Jako mutagen může vystupovat například UV záření, planární aromatické uhlovodíky, radikálové iniciátory nebo viry. Podstatou genových mutací je nukleotidová substituce, delece nebo inserce v řetězci DNA. Často se jedná o mutace bodové (mutace v rozsahu několika nukleotidů) [3]. U mutací nacházejících se v protein kódující části genu můžeme hovořit o synonymních (silent) a nesynonymních (missense a nonsense) mutacích. U tichých (silent) mutací dochází k substituci jedné báze za jinou v exonu genu kódujícího protein, avšak mutace se neprojeví na aminokyselinové sekvenci bílkoviny. Naopak missense mutace způsobují změnu v kódované aminokyselině. Kvůli nonsense mutacím dochází k vytvoření stop kodonu, a tedy i k předčasnému ukončení translace, takže zkrácený proteinový produkt nemusí být funkční [5]. Genové mutace také můžeme dělit podle typu buněk, které jsou mutací zasaženy. Somatické mutace postihují tělní buňky, nejsou zděděné a zároveň se ani nepředávají potomkům. Mutace zárodečných buněk neboli germinální mutace mohou být zděděny od rodičů, nebo vznikají v gametách, a následně jsou předány do dalších generací.

3.3 Struktura chromatinu

Chromatin je tvořen vlákny DNA a proteiny, především histony, které se podílejí na spiralizaci. Spiralizace chromozomu nastává před dělením buňky, kdy se vlákno DNA začne smotávat do spirály, ta vytvoří struktury vyšších řádů a vzniká chromozom. Na tvorbě chromatinu se podílí především 5 druhů histonů: H1, H2A, H2B, H3 a H4 [6]. Volná vlákna DNA jsou namotána na histony a tvoří útvary zvané nukleozomy, ty jsou považovány za základní jednotku chromatinu. Každý nukleozom je tvořen 145-147 nukleotidy DNA, jež jsou navinuty na osm histonových proteinů [7]. Vlákno s nukleozomy připomíná korálky na niti, protože mezi jednotlivými nukleozomy se nachází volné úseky DNA, které se nazývají linkery, jejich délka však není konstantní [8].

Jaderný genom se skládá z heterochromatinu, což je část obsahující hustě sbalený neaktivní chromatin, který je nepřístupný transkripčním mechanismům, a euchromatinu, který obsahuje rozvolněná chromatinová vlákna, jež jsou přístupná mechanismům transkripce. Chromatin má schopnost dynamicky přecházet mezi oběma stavy, a tím ovlivňovat aktivitu genů. Rozvolněný chromatin si ponechává alespoň primární strukturu nukleozomů.



Obrázek 2: Struktura chromatinu. Chromatin je tvořen vlákny DNA a proteiny, především histony. Jeho struktura připomíná korálky na niti. Převzato z [9], upraveno.

3.4 Remodelace chromatinu

Chromatin se přeskupuje podle potřeb buňky a v závislosti na fázi buněčného cyklu. Toto přeskupování je souhrnně nazýváno remodelace chromatinu. Dle dosavadních výzkumů se na remodelaci podílí především dva mechanismy [10].

První mechanismus je založen na kovalentní modifikaci konců histonů vlivem enzymů, které ovlivňují přenos chemických funkčních skupin v rámci histonů. V důsledku toho dochází ke změně chemických vlastností. Histony jsou malé bazické proteiny a rozlišujeme u nich N-konec, C-konec a globulární doménu.[6] Při kovalentní modifikaci hrají klíčovou roli právě konce histonů, zejména N-konce, které jsou k posttranslačním modifikacím nejnáchylnější, protože vyčnívají z histonů volně do prostoru. Dochází na nich k acetylaci, methylaci, ubiquitinaci a fosforylaci[11], které mají vliv na interakce mezi nukleovými bázemi a histony. Tím je ovlivněna úroveň sbalení DNA. Dochází tedy k přímé regulaci genové exprese [12]. Kovalentní modifikace konců histonů souvisí s hypotézou histonového kódu [13]. Tato hypotéza tvrdí, že specifická modifikace histonu, resp. jejich kombinace, vytváří kód, který definuje současnou nebo potenciální transkripční aktivitu.

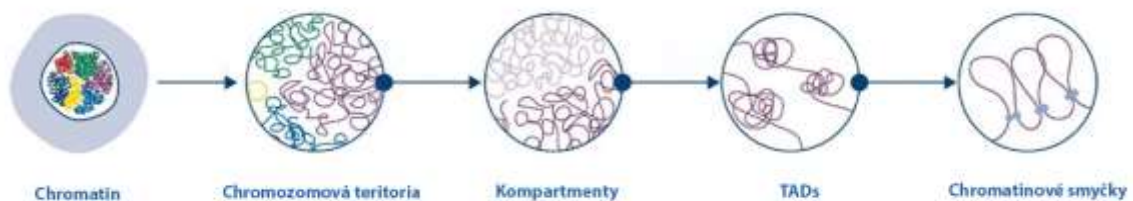
Druhý mechanismus se uplatňuje při remodelaci samotných nukleozomů. Uplatňují se zde ATP-dependentní remodelační komplexy, jejichž činností dochází k fyzickému přesouvání jednotlivých nukleozomů nebo k úplnému rozrušení struktur chromatinu[14], avšak pro uskutečnění mechanismu je nutné dodat energii ve formě ATP. Součástí těchto komplexů je podjednotka ATPáza [15], která se váže na histonové jádro a zároveň na DNA. Jakmile se podjednotka naváže, dojde k hydrolýze ATP a uvolní se energie. Uvolněná energie poté může rozrušit vazbu mezi DNA a histonovým jádrem, což umožňuje nukleozomu posunout se po vláknu DNA. Dochází tedy ke zpřístupnění určité části šroubovice.

3.5 Topologicky asociované domény

Pro objevení topologicky asociovaných domén byl klíčový rok 2012, kdy tři nezávislé vědecké skupiny objevily zvláštní jev, který popsaly jako prostorově vymezené interakce v rámci chromatinu [16]. Později byl tento objev nazván topologicky asociované domény (TADs).

TADs lze popsat jako oblasti chromozomů, které jsou velké několik megabází (Mb) a je pro ně charakteristické, že sekvence DNA uvnitř těchto domén upřednostňují více interakce mezi sebou než interakce se sekvencemi mimo tyto domény [17]. Můžeme zde zároveň rozlišovat i podjednotky, tzv. sub-TADs [18], které jsou menší než 1 Mb a spadají pod TADs [19].

Chromatin je po většinu času rozvolněný, takže chromozomy nejsou spiralizované, pouze jsou členěny do tzv. chromozomových teritorií, které jsou dále tvořeny kompartmenty. Chromatinová teritoria budou nejspíše hrát významnou epigenetickou roli při expresi genů, avšak tato myšlenka zatím není vědecky podložena. Kompartmenty se skládají z TADs, jež si lze představit jako shluky chromatinových smyček (chromatin loops), viz Obrázek 3. Chromatinová smyčka vznikne, jakmile se úseky genomové sekvence, které leží na stejném chromozomu (konfigurace *cis*), dostanou do užší fyzické blízkosti. Smyčky vznikají v rámci domén, ale přesný mechanismus jejich tvorby zatím není znám. Procesu se účastní několik transkripčních faktorů a též bylo zjištěno, že forma i rozdělení smyček může mít vliv na regulaci genové exprese, což lze vidět například na faktoru GAGA u *Drosophila*, který indukuje remodelaci heterochromatinu, a aktivuje tak expresi mnoha genů [20].



Obrázek 3: Členění rozvolněného chromatinu. Převzato z příručky EpiTect® Hi-C, upraveno.

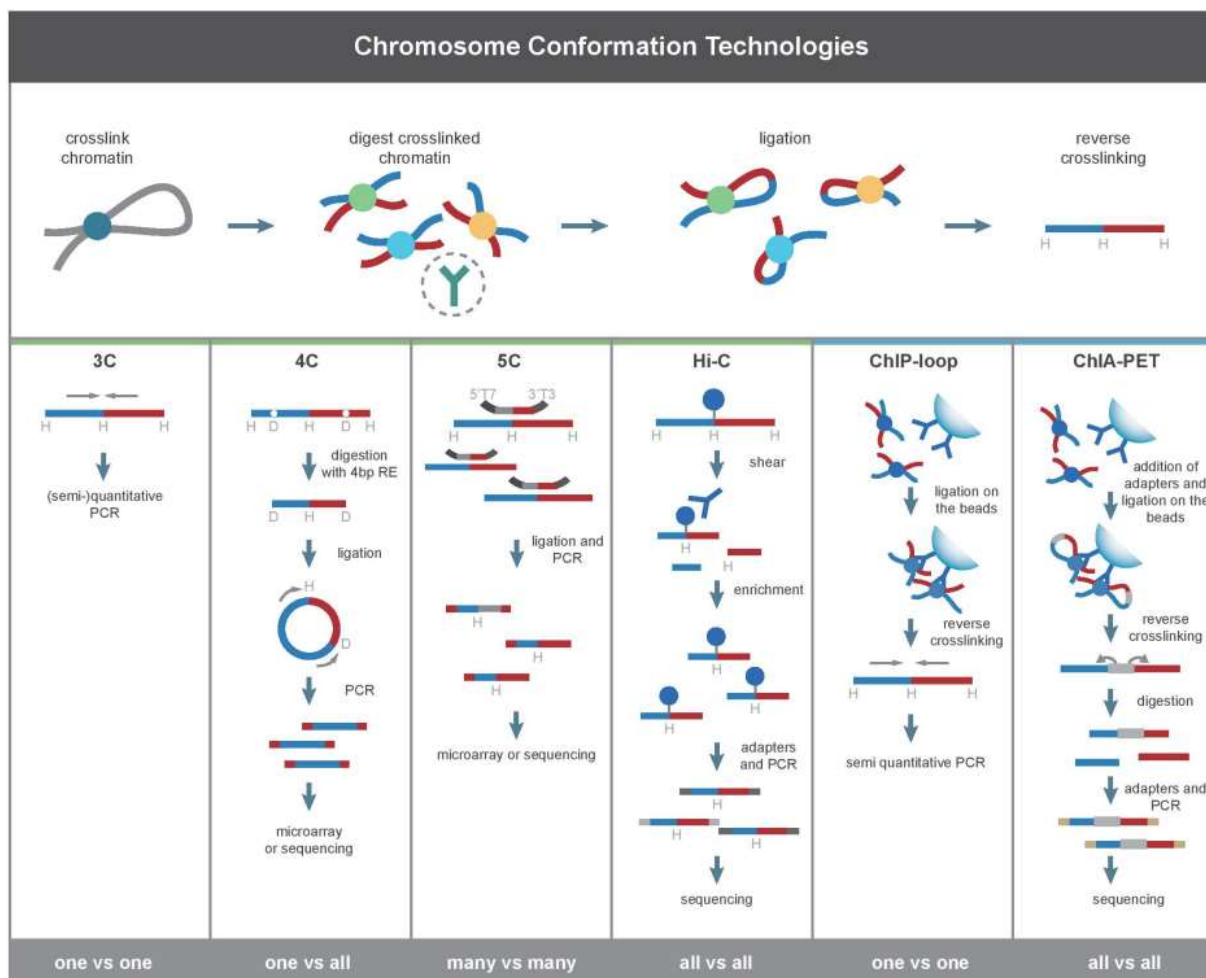
TADs lze z hlediska jejich funkce vnímat jako základní regulační jednotky genomu [16]. Mnoho odborných prací se zabývá jejich vlivem na regulaci transkripce a translace [21], [22]. Právě díky tomu, že jsou TADs prostorově vymezené, je umožněna kompartmentalizace těchto procesů. Toto také podporuje skutečnost, že geny patřící pod určitou stejnou doménu podléhají koregulaci, tedy jsou regulovány stejnými mechanismy. Hranice mezi TADs vymezují oblasti, ve kterých mohou působit regulační mechanismy, takže je málo pravděpodobné, aby regulační mechanismy jedné domény zasahovaly do oblasti podléhající druhé doméně. Lze tedy říci, že hranice mají klíčovou roli pro funkce TADs, a proto může jejich narušení způsobit deregulaci genové exprese, a tedy zapříčinit rozvoj onemocnění [23].

4 METODY STUDIA CHROMATINOVÝCH INTERAKCÍ

Organizace genomu není náhodná a uspořádání ovlivňuje mnoho funkcí včetně transkripce, replikace a oprav poškození DNA. Specifické genomové oblasti obsahující stejné nebo různé chromozomy často fyzicky asociují navzájem a s jadernými strukturami, což vede ke složitě rozčleněnému jádru [24]. Sekvenační technologie zmíněné v úvodu práce (viz kapitola 3.1) v kombinaci s dalšími metodami umožnily zkoumat donedávna těžko pozorovatelné chromatinové interakce. Každá metoda má však svá specifika a limity.

Rozmach studií zaměřených na chromatinové interakce přišel s rozvojem tzv. 3C-technologií (Chromosome Conformation Capture). Tyto metody umožňují detekci interakcí mezi oblastmi genomu, které se v 3D prostoru nachází v těsné blízkosti, avšak v lineární sekvenci DNA mohou být odděleny mnoha nukleotidy. První 3C metoda byla vyvinuta již v roce 2002, od té doby vzniklo mnoho jejích modifikací. V roce 2005 byla představena metoda ChIP-Loop jako kombinace klasické chromatinové imunoprecipitace (ChIP) a metody 3C. V roce 2006 byla vyvinuta metoda 4C, která byla následně vylepšena na metodu 5C. O rok později byla představena technologie ChIP-seq. Rok 2009 přinesl hned dva nové přístupy: metodu Hi-C a ChIA-PET, které v současnosti poskytují nejlepší rovnováhu mezi rozlišením a pokrytím celého genomu [25].

Technologie C (3C, 4C, 5C, Hi-C) se liší způsobem detekce a rozsahem interakcí, které mohou určit. Metoda 3C testuje interakci mezi dvěma známými místy v genomu, 4C umožňuje zkoumání více neznámých interaktorů a jedné známé sekvence, 5C identifikuje všechny oblasti interakce v dané doméně genomu a Hi-C umožnila studovat organizaci chromatinu v rámci celého genomu. Další varianty (ChIA-PET, ChIP-Loop) obsahují navíc krok srážení proteinu, který umožňuje identifikaci genomových interakcí zahrnujících konkrétní sledovaný protein [24].



Obrázek 4: Přehled metod studia chromatinových interakcí. Převzato z [26].

4.1 3C a její modifikace

3C-technologie vždy zahrnují pět kroků: (1) fixace formaldehydem za účelem zesíťování chromatinu v místech fyzické interakce, (2) štěpení chromatinu restriktivním enzymem nebo ultrazvukem, (3) ligace mezi konci DNA zachycenými na stejný komplex, (4) detekce ligačních spojů v závislosti na zvolené metodě, (5) výpočetní analýza ke stanovení frekvencí interakcí zachycených při ligaci zesíťovaného chromatinu [27]. Zesíťování chromatinu (tzv. crosslinking) je pojem označující proces, při kterém dochází k reakci nějaké látky, např. formaldehydu, se dvěma nukleotidy DNA a vzniká mezi nimi kovalentní vazba. Ke vzniku crosslinků může docházet na stejném řetězci nebo mezi protilehlými řetězci dvouvláknové DNA.

Podstatou 3C metody je detekce frekvencí interakcí jediné dvojice genových lokusů [27]. Genový lokus je termín označující přesnou oblast na chromozomu, ve které je kódován nebo se vyskytuje příslušný gen/geny. Metoda jako vůbec první umožnila porozumět 3D organizaci chromatinu [27]. Využívá se například k testování interakcí mezi promotorem a zesilovačem. Produkty ligace jsou detekovány pomocí PCR a následné gelové elektroforézy nebo kvantitativní PCR (qPCR) [28]. Základní princip 3C metody je v současnosti kombinovaný s dalšími metodami, je tedy zvětšován rozsah i specifita metod.

Od metody 3C je odvozena metoda 4C, která se liší především širším rozsahem. Lze díky ní detekovat interakce mezi jedním specifickým lokusem a všemi ostatními genomovými lokusy. 4C zahrnuje standardní přípravu knihovny 3C a knihovna je poté amplifikována z místa zájmu. Toho je dosaženo použitím inverzní PCR se specifickými primery. Produkty PCR jsou následně identifikovány pomocí sekvenování nebo DNA čipu. Metoda se využívá k odhalení řady různých biologických jevů, např. vlivu faktorů pluripotence na formování genomu embryonálních kmenových buněk a dopadu duplikací na organizaci chromatinu [28].

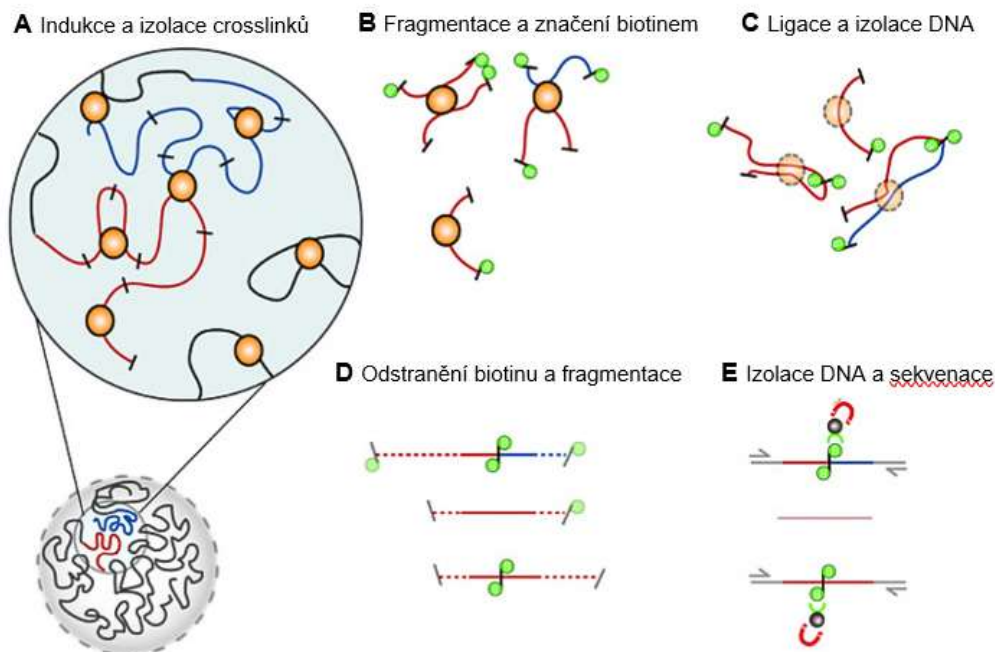
Metoda 5C umožňuje analýzu všech interakcí v konkrétním lokusu až do několika megabází. Samotná metoda napomohla k objevení TADs a nadále je využívána k jejich zkoumání. Opět zahrnuje standardní přípravu 3C knihovny, avšak produkty jsou inkubovány se směsí univerzálních primerů [28]. Po amplifikaci pomocí PCR následuje detekce metodou sekvenování nebo DNA čipu [29].

4.2 Hi-C

Metoda Hi-C umožnila zkoumat 3D organizaci celých genomů. Hi-C protokoly vyžadují jiný postup přípravy knihovny než výše zmíněné metody, viz Obrázek 5. Před ligací fragmentů obsahujících crosslinky se kohezní konce fragmentů označí nukleotidy s navázaným biotinem. Po ligaci dochází k rozrušení crosslinků a případné další fragmentaci sonikací. Biotinylované fragmenty jsou následně selektovány prostřednictvím kuliček potažených streptavidinem. Získané fragmenty jsou analyzovány pomocí vysoce výkonného paired-end sekvenování, které umožňuje sekvenování z obou stran fragmentu DNA, a tím poskytuje informaci o sekvenci. Následné bioinformatické vyhodnocení určí, kam mají být v rámci genomu jednotlivá čtení umístěna [30].

Rané studie využívající metodu Hi-C odhalily do té doby neznámé vlastnosti chromatinu, například rozdělení genomu do kompartmentů, TADs a sub-TADs [31]. Hi-C data s vyšším rozlišením byla představena v roce 2014 v návaznosti na vývoj *in situ* Hi-C, kde se ligace provádí v neporušených jádrech. To umožňuje vizualizaci struktur malého rozsahu včetně genomové analýzy regulačních interakcí a vizualizace kontaktních domén příliš malých na to, aby byly dříve detekovány [32]. Metoda Hi-C je jedinečná ve své schopnosti zkoumat rozsáhlé struktury a kontakty velmi dlouhého dosahu. Zároveň není schopna tak vysokého rozlišení jako jiné metody, počet kontaktů pro jakýkoli daný lokus je zhruba stokrát nižší než u 5C [33]. Nejnovější soubory dat získané pomocí Hi-C mohou poskytovat rozlišení přibližně kilobáze (kb) [32]. Zvýšení citlivosti testů Hi-C je přímo úměrné hloubce sekvenování a další snahy jsou do značné míry omezeny finanční náročností projektů.

Jedním z aspektů experimentů Hi-C je obrovský objem a složitost generovaných sekvenčních dat. Pro zpracování a analýzu dat jsou tedy nutné výpočetní nástroje a dovednosti v oblasti bioinformatiky. Hlavní strategie analýzy dat nejčastěji spočívá ve vizualizaci do tzv. heat map pomocí softwaru pro analýzu dat.



Obrázek 5: Princip metody Hi-C. A) Interakce chromatinu uvnitř i mezi chromozomy jsou stabilizovány tvorbou crosslinků (= komplexů protein/DNA) indukovaných formaldehydem. B) Chromatin se štěpí restričním enzymem a konce jsou označeny biotinem. C) Fragmenty DNA v crosslinkovaných komplexech jsou ligovány za vzniku chimérických molekul DNA. D) Biotin je odstraněn z konců lineárních fragmentů a molekuly jsou dále fragmentovány, aby se snížila jejich celková velikost. E) Molekuly s navázaným biotinem jsou vyzolovány pomocí magnetických kuliček potažených streptavidinem a upraveny pro masivně paralelní hluboké sekvenování. Převzato z [30], upraveno.

4.3 ChIP-Loop

ChIP-Loop je technika, která je založena na chromatinové imunoprecipitaci a 3C-technologiích. Začátek protokolu je totožný s přípravou 3C knihovny, chromatin je formaldehydem crosslinkován a naštěpen. Následně je přidána protilátka proti sledovanému proteinu, pomocí které dochází k obohacení (tzv. imunoprecipitaci) komplexů protein-DNA [34]. Ligace probíhá, zatímco je k chromatinu stále vázána protilátka. K identifikaci hledaných interakcí se pak používá qPCR. Hlavní výhodou ChIP-Loop je snížení šumu na pozadí v experimentech 3C a vyšší specifická výběrem známého proteinu zprostředkujícího interakci DNA-DNA.

4.4 ChIP-seq

ChIP-sequencing poskytuje genomové profilování oblastí DNA, kde dochází k modifikaci histonů nebo vazbě transkripčních faktorů a dalších proteinů. Kombinuje selektivitu ChIP k získání specifických komplexů protein-DNA s výkonem sekvenování NGS. Navíc, protože se komplexy protein-DNA získávají ze živých buněk, lze vazebná místa srovnávat v různých typech buněk a tkáních nebo za různých podmínek. Hlavní výhodou je vyšší rozlišení a snížení šumu oproti metodě ChIP-chip (chromatinová imunoprecipitace na čipu) [35]. Nevýhodou

ChIP-seq jsou finanční náklady, navíc metoda vyžaduje větší množství tkáně, což může být u některých vzácných vzorků problém [36].

4.5 ChIA-PET

Chromatin interaction analysis by paired-end tag sequencing neboli ChIA-PET kombinuje ChIP s technologií analýzy konformace chromatinu. Umí detekovat vzájemné interakce vzdálených oblastí DNA prostřednictvím vychytávání proteinu zájmu [37]. Metoda ChIA-PET se dá použít například k odhalení kontrolních mechanismů v genomu během procesů, jako je buněčná diferenciaci. Crosslinkovaný chromatin se štěpí, vzniklé komplexy DNA-protein jsou značeny protilátkami namířenými proti proteinu zprostředkávajícímu interakci. Na konce DNA jsou připojeny tzv. linkerové sekvence s navázaným biotinem, což usnadňuje jejich vzájemnou ligaci a následnou izolaci pomocí streptavidinových magnetických kuliček. Vznikají hybridní fragmenty DNA ze dvou různých oblastí genomu. Připravená knihovna je následně analyzována masivně paralelním sekvenováním [38]. Největší nevýhodou je, že ChIA-PET testuje interakce jednoho známého proteinu. Pro studování složitějších komplexů jsou tím pádem vyžadovány opakované experimenty [37].

5 CHRONICKÁ LYMFOCYTÁRNÍ LEUKÉMIE

Nádorová onemocnění patří mezi nejčastější příčiny úmrtí na celém světě a jejich výskyt neustále narůstá. Mohou postihnout téměř každou část lidského těla. Přes mnohaletou snahu porozumět mechanismům vzniku těchto onemocnění neumíme často vysvětlit, proč se u konkrétních pacientů objeví. Nádorová buňka je velice podobná ostatním buňkám v těle, nepřichází zvenčí, ale vzniká z buněk těla vlastním procesem postupného hromadění mutací. V určitém okamžiku dochází ke ztrátě kontroly nad buněčnými procesy, buňky začínají autonomně růst a vzniká životaschopný novotvar, který může celý organismus ohrozit na životě. Pochopení molekulárních procesů vedoucích ke vzniku mutací a nalezení konkrétních mutací typických pro jednotlivé diagnózy je cílem mnoha výzkumných projektů.

Speciální pozornost si zaslouží hematologické malignity, pod které spadají mimo jiné leukémie – nádorová onemocnění krvetvorby, která se nejčastěji vyznačují zmnožením leukocytů. Dle rychlosti průběhu onemocnění je dělíme na akutní a chronické, dále podle krvetvorných buněk, ve kterých došlo ke zvratu, na lymfocytární, myeloidní a jejich podtypy.

5.1 Charakteristika chronické lymfocytární leukémie, diagnóza a výskyt

Chronická lymfocytární leukémie (CLL) je nejčastějším druhem leukémií dospělých pacientů v Evropě a Severní Americe [39]. Jedná se o onemocnění s nízkou malignitou, avšak s velmi různorodým klinickým průběhem. Vzniká z B-lymfocytů, které se díky schopnosti unikát apoptóze mohou hromadit v krvi, kostní dřeni, lymfatických uzlinách, ale i jiných orgánech. To má za následek utlumení krvetvorby, tedy nedostatečnou produkci zdravých erytrocytů, leukocytů a trombocytů [39].

Pacienti s CLL mohou mít roky až desetiletí dlouhé bezpříznakové období. Onemocnění se rozvíjí velmi pomalu, takže jediným detekovatelným příznakem je abnormalita v krevním obraze, konkrétně lymfocytóza v periferní krvi. K diagnóze v časných stádiích často dochází náhodně na preventivních prohlídkách při rutinních odběrech. Mezi další příznaky patří zvětšení lymfatických uzlin, sleziny nebo jater. Za časté příznaky, tzv. B-symptomy typické pro B-lymfoproliferaci, lze považovat noční pocení, horečky, úbytek hmotnosti, který může přejít až do kachektizace, anémie či trombocytopenie s krvácivými příznaky. Pacienti mají také vyšší riziko infekčních onemocnění [39].

Stanovit diagnózu lze podle několika kritérií a řadou vyšetření. Mezi základní vyšetření patří laboratorní vyšetření vzorku krve. Pro stanovení diagnózy CLL je vyžadováno minimálně 5×10^9 lymfocytů v 1 l periferní krve, které jsou malé, vyžralé, mají uniformní morfologii a charakteristický imunofenotyp dle flow-cytometrie. Dalším vyšetřením je imunofenotypizace, která je prováděna metodou průtokové cytometrie, a umožňuje určení příslušnosti buněk podle antigenních molekul na jejich povrchu. Pro CLL buňky je typická zvýšená exprese CD5, CD19 a CD23 a naopak nižší exprese CD20, CD79b a povrchových

imunoglobulinů [40]. Kromě toho byla u většiny CLL buněk prokázána exprese povrchového receptoru ROR1. Tyto základní metody doplňuje cytogenetické vyšetření pro zjištění chromozomálních aberací.

CLL tvoří 30 % všech diagnostikovaných leukémií. Výjimečně se objevuje u osob mladších padesáti let. Je dvakrát častější u mužů než u žen [41]. U žen dochází k propuknutí onemocnění přibližně o 6 let později. Ve výsledku nastává vyrovnání incidencí mezi ženami a muži staršími 50 let. Medián věku pacientů v době diagnózy je 70 až 72 let [40]. I přes velké pokroky v léčbě dochází u mnoha pacientů po ukončení terapie k opětovnému propuknutí onemocnění a CLL zůstává nevléčitelná.

5.2 Patogeneze onemocnění

Přesnou patogenezi CLL stále neznáme, avšak probíhají studie se snahou zjistit roli antigenních stimulací, narušení apoptózy a vlivu mikroprostředí. Pacienti s tímto onemocněním se dají rozdělit do dvou skupin s rozdílným klinickým průběhem v závislosti na mutačním statusu genů variabilní části těžkého řetězce imunoglobulinů (IGHV) [39]. Také chromozomové aberace mají významnou roli u tohoto onemocnění. Mezi aberace se známým prognostickým významem patří delece oblastí 11q, 13q, 17p a trizomie chromozomu 12. Dalšími potenciálními změnami v leukemických buňkách může být např. zmnožení 2p, delece 6q21, duplikace 8q24, přestavby 14q32 a komplexní karyotyp [42]. K somatickým mutacím dochází s větší četností například v genech *ATM*, *TP53*, *SF3B1*, *NOTCH1*, *MYD88*, *ATM*, *BIRC3*, *XPO1*, *CHD2*, *POT1* nebo *KRAS*.

Mutační status IGHV se zjišťuje během standardních vyšetření při diagnóze CLL. IGHV se podílí na tvorbě B-buněčného receptoru (BCR). Jako nemutovaný je označován gen IGHV, který se shoduje v minimálně 98 % sekvence se zárodečnou linií. Mutace v IGHV vznikají přirozeným vývojem B lymfocytů, nelze je označit za patogenní. S mutovaným IGHV se setkáme přibližně u 60 % pacientů s CLL. Naopak u zbylých 40 % nemocných se vyskytuje IGHV nemutovaný, protože B lymfocyty neprojdou procesem somatické hypermutace genů kódujících IGHV. Vyšetření mutačního statusu IGHV je důležité pro určení prognózy pacientů [43].

Somatické mutace u CLL buněk byly objeveny především v genech, které se účastní apoptózy, mechanismů opravy DNA a kontroly buněčného cyklu. K nejvýznamnějším mutacím a delecím dochází v genech *TP53* a *ATM*, které se podílí na opravě DNA. Gen *TP53* patří mezi důležité tumor supresorové geny a kóduje protein p53 fungující jako senzor poškození DNA. Produkt genu *ATM* je důležitý pro opravu dvouřetězcových zlomů DNA [44].

Chromozomální aberace jsou diagnostikovány pomocí metody FISH (fluorescenční *in situ* hybridizace), která patří mezi molekulárně cytogenetická vyšetření. Mezi nejčastěji vyšetřované vady na chromozomech patří delece na chromozomu 11q (del11q), delece na chromozomu 13q (del13q), delece na chromozomu 17p (del17p) a trizomie chromozomu 12.

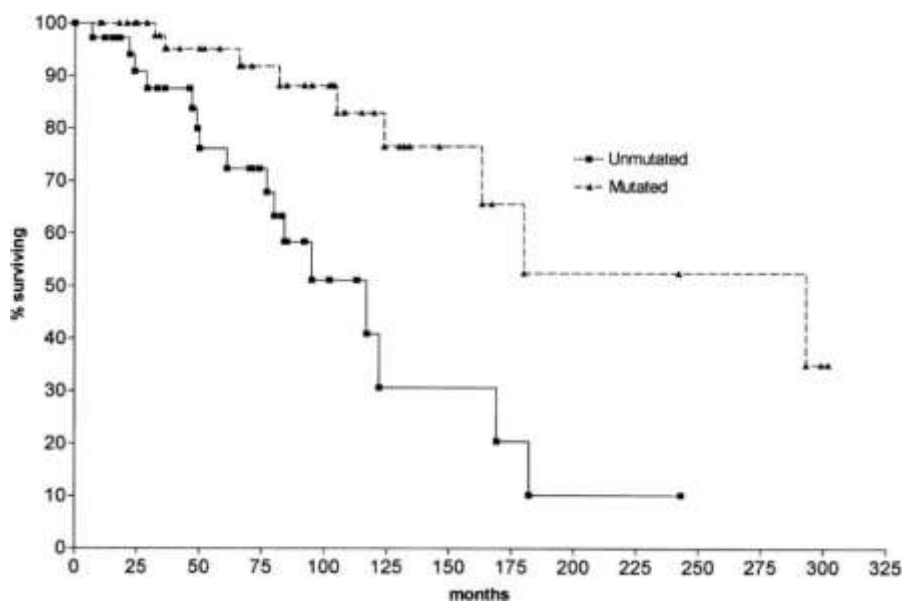
5.3 Prognóza

Průběh CLL bývá velmi heterogenní. Klinické stádium se stanovuje na základě vyhodnocení rizika progresu nemoci. Pro klasifikaci stupně pokročilosti onemocnění se využívají dva systémy klinického stagingu zavedené Raiem a Binetem. Prognostický systém dle Raie z roku 1975 měl původně 5 prognostických skupin, ale systém byl nedávno modifikován pouze na tři skupiny, stejně jako klinická klasifikace CLL dle Bineta, která pochází z r 1981.

V posledních letech došlo k řadě nových objevů, které napomohly posunu v léčbě CLL natolik, že výše zmíněné systémy už nestačí, protože neumí uspokojivě predikovat vývoj onemocnění. Rai a Binet opřeli své klasifikační systémy o klasické prognostické faktory (koncentrace hemoglobinu, počet trombocytů, postižení lymfatických orgánů a lymfocytóza). Prognostický význam má také dynamika nárůstu počtu lymfocytů v krevním obraze, delší zdvojovací čas je spojen s delším přežitím. Pokud má pacient počet lymfocytů nad 50×10^9 , je pravděpodobnost přežití přibližně 4-5 let, s lymfocyty pod 50×10^9 pak 6 a více let [45].

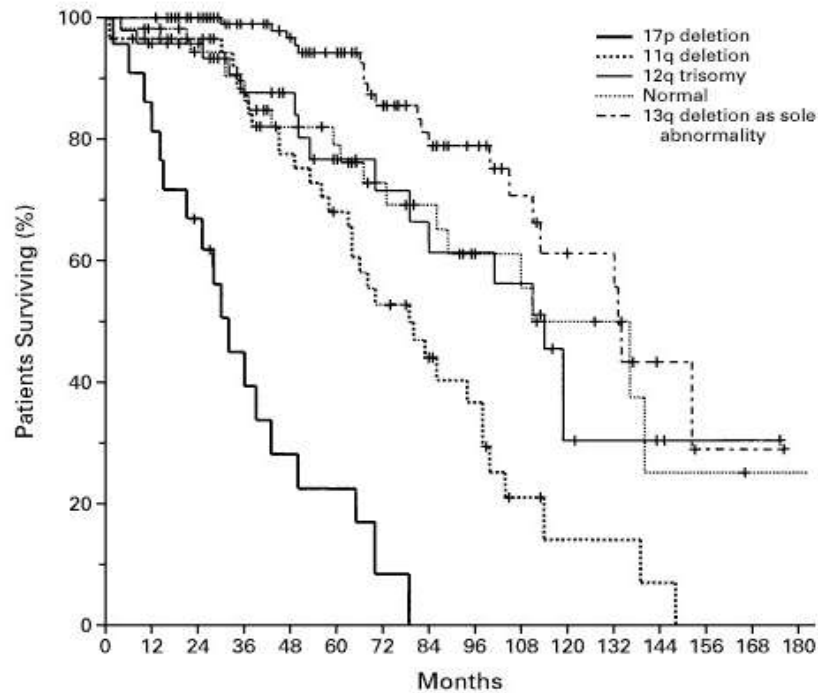
Se zavedením molekulárně biologických metod, jako je např. metoda FISH či sekvenování nové generace, začal být kladen důraz na molekulárně genetické markery jako na nové prognostické faktory, které umožňují stanovit individuální prognózu. Standardní klinická vyšetření jsou běžně doplněna o vyšetření somatických hypermutací v genu IGHV, chromozomálních aberací a genových mutací.

Pomocí mutační analýzy IGHV můžeme CLL pacienty rozdělit do dvou skupin. U první skupiny je onemocnění odvozeno z B-lymfocytů, u kterých došlo k somatické hypermutaci, tedy k přeskupení genu IGHV v reakci na určitý antigen. Druhá skupina nese nemutovaný gen IGHV, což bývá často spojeno s nepříznivými cytogenetickými abnormalitami, agresivním průběhem onemocnění a horší prognózou.



Obrázek 6: Kaplan-Meierovy křivky přežití u CLL na základě mutačního statusu IGHV. Převzato z [45], upraveno.

Pacienti s normálním karyotypem mají pravděpodobnost přežití i 15 let.[46] Doba přežití však bývá výrazně zkrácena v případě delece 11q22-q23 (oblast kódující gen *ATM*), která je detekována u 10-20 % pacientů s CLL. Nejzávažnější chromozomová aberace je delece 17p13 (oblast kódující gen *TP53*), která je detekována u 10-15 % pacientů [39]. Samostatně či v kombinaci s del 17p se často může vyskytovat i mutace *TP53*, což má také negativní vliv na přežití. Pacienti s tímto poškozením obvykle špatně reagují na léčbu imunochemoterapií [42].



Obrázek 7: Kaplan-Meierovy křivky přežití u CLL podle typu chromozomálních aberací. Převzato z [42], upraveno.

5.4 Léčba

CLL je onemocnění s nízkou malignitou. Pacienti bez klinických příznaků jsou pouze sledováni a aktivní léčba je zahájena, pokud onemocnění výrazně progreduje, to může být spojeno např. s klesajícím množstvím erytrocytů a trombocytů, hepatosplenomegalií vytvářející mechanické potíže, výraznou únavou, úbytkem hmotnosti apod. [39].

CLL se většinou podaří zachytit v počátečních stádiích díky běžným vyšetřením krevního obrazu u praktického lékaře. Mnohdy ani nedojde k zahájení terapie, asi v polovině případů, není na základě klinického stavu nutné zahájit terapii a následuje strategie watch-and-wait. Podle individuálního posouzení klinického stavu pacienta a prognostických faktorů se stanovuje vhodná terapie. U velké části pacientů však dochází k relapsu a onemocnění se stává rezistentním na terapii. Proto je aktuálně kladen důraz na vývoj léčiv, jež budou cílena na určité proteiny a signální dráhy související s patogenezí CLL. V současné době se hojně využívá léčba látkami inhibujícími signalizaci přes B-buněčný receptor či inhibitor antiapoptického proteinu Bcl-2 [39].

Pacientům s CLL, kteří mají velmi špatnou prognózu, může být nabídnuta alogenní transplantace krvetvorné tkáně, je však nutné zvážit celkový klinický obraz pacienta. Transplantace přináší mnoho komplikací, takže je doporučovaná pouze lidem v dobré kondici.

5.5 Buněčná linie HG-3

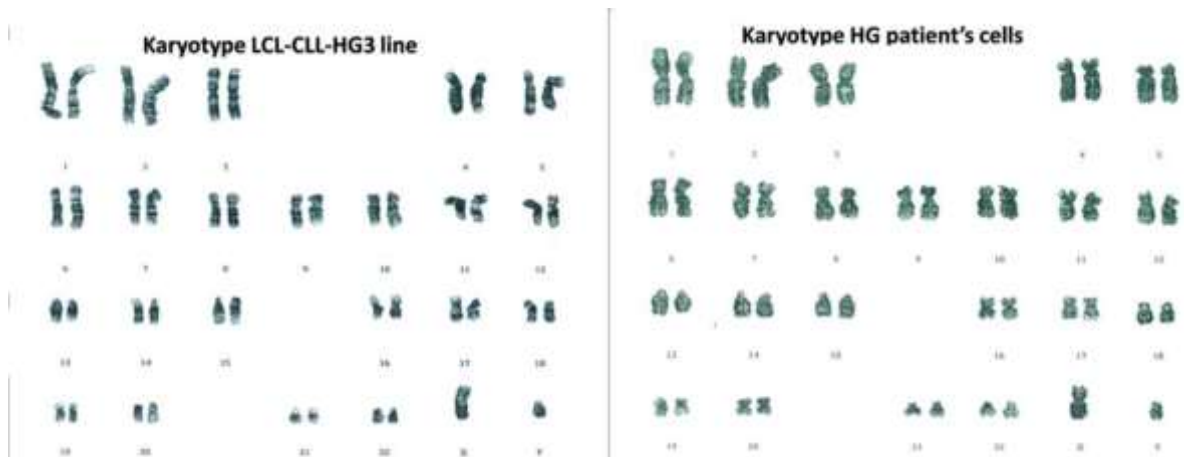
Počátkem 20. století se poprvé povedlo udržet odebrané buňky naživu i mimo organismus. Jako klíčové se ukázalo nastavení optimálních podmínek pro jejich množení. Kultivované buňky můžeme z hlediska stáří dělit do tří skupin: primární kultury (primokultury), buněčné kmeny a buněčné linie.

Primární kultury tvoří čerstvě odebrané buňky z organismu. Vyznačují se krátkou životností v řádu maximálně několika dní, poté je potřeba buňky převést do nového média - tzv. pasážovat. Z primární kultury, která podstoupila minimálně jednu pasáž, může vycházet buněčný kmen. Zde už mluvíme o stabilní kultuře normálních diploidních buněk, které přežívají 40-50 dělení, než zaniknou. Nakonec buněčné linie jsou již adaptovány na podmínky *in vitro* a mají schopnost se neomezeně dlouho množit, a tím se podobají nádorovým buňkám.

Buněčné linie mohou vytvářet buňky izolované z nádorů, anebo vznikají transformací buněčných kmenů. Snad nejznámější nádorovou buněčnou linií je HeLa, pocházející z buněk izolovaných z karcinomu děložního hrdla Henrietty Lacksové. Nádorové buněčné linie jsou velmi využívané při výzkumných studiích, protože jsou málo náročné na růstové faktory a hormony, zároveň je lze neomezeně pasážovat, neboť u nich nedochází ke zkracování telomer, obvykle se snadno množí a kultivují. Nepodléhají tzv. Hayflickovu limitu (maximální počet buněčných dělení).

Dalším příkladem buněčné linie izolované z nádorového onemocnění je linie HG-3, která je odvozená z chronické lymfocytární leukémie. HG-3 je složená ze tří subklonů, z nichž každý nese klonově specifické aberace, genovou expresi a vzory methylace DNA. Zatímco všechny leukemické buňky dárce exprimovaly zvýšeně CD5, dva ze tří subklonů HG-3 tento marker ztratily. Právě variabilita těchto subklonů nesoucích jednotlivé mutace a charakteristické sady odlišně exprimovaných genů mohou být užitečné pro různé regulační a funkční studie [47].

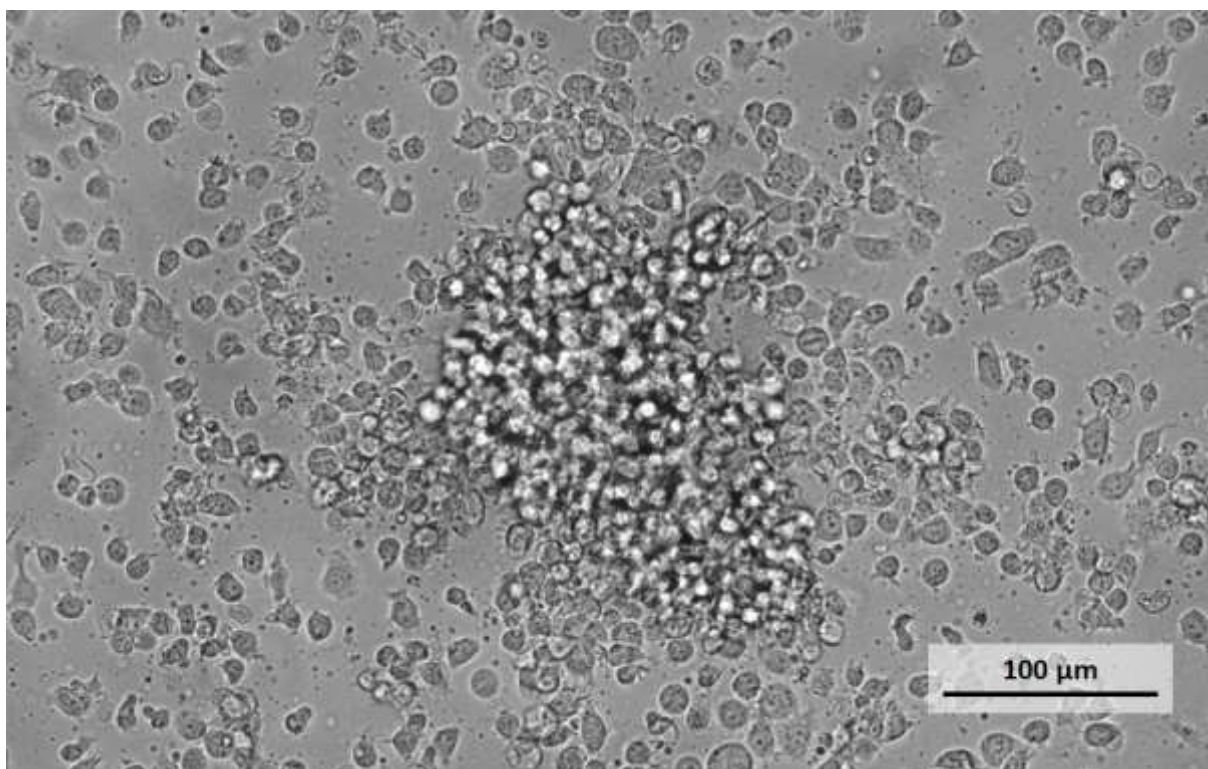
HG-3, která patří mezi lymfoblastoidní buněčné linie (LCL), byla vytvořena z klonu pacienta s CLL, jehož leukemické buňky byly *in vitro* infikovány částicemi viru EBV (virus Epstein-Barrové). Virové proteiny fungují jako transkripční faktory kontrolující růst a přežívání, zajišťují tak imortalizaci (= možnost nekonečného dělení) leukemických buněk. Srovnání HG-3 linie se dvěma LCL vytvořenými z normálních B lymfocytů ukázalo, že 32 genů je u HG-3 exprimováno ve více než dvojnásobných hladinách. Mezi nimi byly například *LHX2* a *LILRA*. Tyto geny mohou hrát roli ve vývoji onemocnění [48].



Obrázek 8: Karyotyp buněčné linie HG-3 a buněk dárce. Převzato z [48].

6 MATERIÁL

Experimenty byly prováděny na buněčné linii HG-3, kterou pracovišti laskavě poskytli prof. Richard Rosenquist Brandell z Uppsala University ve Švédsku. Jedná se o kulaté buňky rostoucí v suspenzi, které tvoří menší shluky a lehce přisedají ke dnu. Buňky byly kultivovány v médiu RPMI 1640 se stabilním glutaminem a 25mM HEPES pufrům s přidavkem fetálního bovinního séra (finální koncentrace 10 %), penicilinu (finální koncentrace 100 U/ml) a streptomycinu (finální koncentrace 100 ng/ml). Buňky byly pasážovány dvakrát týdně s ředěním v poměru 1:2 až 1:6. Kultivace probíhala v inkubátoru s řízenou 5% koncentrací oxidu uhličitého, která zajišťuje optimální pH kultivačního média.



Obrázek 9: Suspenze HG-3 buněk v RPMI 1640 médiu zachycená světelným mikroskopem. (Foceno Mgr. Hanou Svozilovou v CEITEC MU).

7 METODY

Pro analýzu leukemických buněk jsme se rozhodli využít moderní metodu Hi-C, která umožňuje stanovit interakce chromatinu a strukturní aberace. Na buněčné linii HG-3, odvozené od CLL, jsme vyzkoušeli přípravu Hi-C sekvenační knihovny pomocí tří různých komerčně dostupných kitů. EpiTect Hi-C Kit od firmy Qiagen umožňuje zachytit především interakce na delší vzdálenost. Toho je docíleno využitím restričního enzymu, který štěpí chromatin v místech obsahujících specifickou sekvenci, což je obecně používaný přístup pro generování Hi-C knihoven. U enzymu dodaného v rámci EpiTect Hi-C kitu šlo konkrétně o sekvenci GATC. Naproti tomu firma Dovetail Genomics využívá enzymy, které štěpí genom v náhodných místech. Tím je zajištěno sekvenační pokrytí celého genomu a získání přesnější informace o jeho organizaci. Dovetail™ Micro-C Kit od firmy Dovetail Genomics používá při přípravě sekvenační knihovny MNázu. Jedná se o enzym, který štěpí DNA mezi jednotlivými nukleozomy, tedy v místě linkerů. Vznikají malé fragmenty s přibližně stejnou délkou. To umožňuje vysledovat interakce na velmi krátkou vzdálenost, ale naopak nemusí být po analýze dat viditelné interakce vzdálenější. Dovetail™ Omni-C Kit od firmy Dovetail Genomics používá DNázu, která štěpí náhodně s nepravidelnými rozestupy, po štěpení tedy vznikají fragmenty různé délky. Kit umožňuje zachytit jak interakce s krátkým, tak i s delším rozsahem. Měl by tedy mít nejširší využití, od sestavení kompletní genomické sekvence a identifikaci interakcí až po genotypizaci a určení haplotypu.

7.1 Použité chemikálie:

EpiTect Hi-C Kit (Qiagen)

- Phosphate buffered saline (PBS)
- 37% formaldehyde
- 3M Tris, pH 7.5
- 1% SDS
- 10% Triton X-100
- 100% isopropanol
- 80% ethanol
- 3M sodium acetate, pH 5.2
- RNase-free water
- Buffer C1
- Hi-C Digestion Buffer
- Hi-C Ligation Buffer
- Ultralow Input Ligase
- Hi-C digestion enzyme
- Hi-C End Labeling Mix
- Hi-C End Labeling Enzyme
- MinElute® DNA Spin Columns
- Buffer EB
- Buffer SB1
- ER/A-Tailing Buffer
- ER/A-Tailing Enzyme Mix
- Adapter Ligation Buffer
- 2x HiFi PCR Master Mix
- Primer Mix Illumina Library Amp
- Streptavidin Beads
- Bead Wash Buffer 1
- Bead Wash Buffer 2
- Bead Resuspension Buffer
- Adapter Ligation Buffer
- Illumina Adapter
- QIAseq Beads

Dovetail™ Micro-C Kit a Dovetail™ Omni-C Kit (Dovetail Genomics)

- TE Buffer pH 8.0
- 10x Wash Buffer
- TWB Solution
- NTB Solution
- LWB Solution
- NWB Solution
- Chromatin Capture Beads
- Crosslink Reversal Buffer
- Streptavidin Beads
- RBC Lysis Buffer
- 20% SDS
- MNase Enzyme Mix / Nuclease Enzyme Mix
- Nuclease Digest Buffer
- 100 mM MgCl₂
- 0.5 M EGTA (Egtazic Acid)
- 0.5 M EDTA (Ethylenediaminetetraacetic Acid)
- End Polishing Enzyme Mix
- End Polishing Buffer
- Bridge Ligation Buffer
- Bridge Ligase
- Bridge 0
- Intra-Aggregate Ligation Enzyme Mix
- Intra-Aggregate Ligation Buffer
- Proteinase K
- 250 mM DTT
- HotStart PCR Ready Mix
- 37 % Formaldehyde Solution
- Phosphate buffered saline (PBS)
- 100% ethanol
- UltraPure™ DNase/RNase-Free Distilled Water
- DNA Clean & Concentrator™-5
- DSG (Disuccinimidyl Glutarate)
- DMSO (Dimethyl Sulfoxide, Anhydrous ≥ 99.99%)

Kontrola kvality a kvantity DNA

- Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific)
- High Sensitivity D1000/D5000 ScreenTape (Agilent Technologies)
- High Sensitivity D1000/D5000 Sample Buffer (Agilent Technologies)
- DNF-474 High Sensitivity NGS Fragment Analysis Kit, (1-6000 bp) (Agilent Technologies)

Sekvenování

- NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®
- SPRIselect® Beads
- NextSeq 500/550 Mid Output Kit v2.5 (150 cycles) (Illumina)

7.2 Použité přístroje:

- Vortex Genie®2 / 2T / 2Digi (Scientific Industries)
- Memmert WNE 14 (Verkon)
- Eppendorf Thermomixer comfort (Eppendorf)
- Eppendorf Centrifuge 5810/5810 R (Eppendorf)
- Eppendorf MiniSpin Plus (Eppendorf)
- Dual Fluorescence Cell Counter LUNA-FL™ (Logos Biosystems)
- Qubit® 2.0 Fluorometer (Invitrogen)
- 2200 TapeStation System (Agilent Technologies)
- Fragment Analyzer™ Automated CE System (Advanced Analytical)
- S220 Focused-ultrasonicator (Covaris®)
- Biometra TAdvanced thermal cycler (Analytik Jena AG)
- NextSeq 500 Sequencing System (Illumina)

7.3 EpiTect Hi-C Kit

7.3.1 Krok 1: Stabilizace chromatinu a buněčná lyze

HG-3 buňky byly nejdříve vystaveny působení 37 % formaldehydu, aby byl indukován vznik crosslinků chromatinu. Po inkubaci 10 minut byly jeho účinky neutralizovány přidáním 3M Tris pufru, pH 7.5. Buňky byly vystaveny teplotnímu šoku a následně přepočítány pomocí fluorescenčního hemocytometru Dual Fluorescence Cell Counter LUNA-FL™. Pro další krok bylo použito 5×10^5 buněk. Následovala buněčná lýza a štěpení chromatinu pomocí enzymu, který štěpí specificky v místech obsahujících sekvenci GATC.

7.3.2 Krok 2: Značení biotinem a rozrušení crosslinků

Ke vzorku byly přidány biotinylované nukleotidy a ligační mix. Po 2hodinové inkubaci následovalo rozrušení vazeb v crosslinkovaném chromatinu pomocí proteinázy K. DNA byla přečištěna za použití přečišťovacích kolonek MinElute® DNA Spin Columns.

7.3.3 Krok 3: Příprava Hi-C knihovny

Prvním krokem přípravy knihovny byla mechanická fragmentace provedená na ultrazvukovém sonikátoru S220 Focused-ultrasonicator (Covaris®). Cílem bylo získat fragmenty o průměrné délce 400 až 600 bp. Purifikovaná DNA byla smíchána se streptavidinovými kuličkami, aby došlo k vyvázání biotinylovaných fragmentů. Dále byly upraveny konce fragmentů, aby mohlo dojít k navázání adaptérů. Dalším krokem byla amplifikace knihovny pomocí chemie dodané výrobcem. Amplifikace byla provedena v přístroji Biometra TAdvanced, který byl nastaven na uvedený program, viz Tabulka 1.

Fáze	Teplota	Čas	Počet cyklů
Aktivace enzymů	98 °C	2 minuty	1
Denaturace	98 °C	20 sekund	7
Hybridizace	60 °C	30 sekund	
Elongace	72 °C	30 sekund	
Extenze	72 °C	1 minuta	1
	12 °C	Podržet do vyzvednutí vzorku	1

Tabulka 1: Nastavený program přístroje Biometra TAdvanced.

Ověření kvality připravené knihovny proběhlo pomocí Qubit® fluorometru a 2200 TapeStation System.

7.4 Dovetail™ Micro-C Kit

7.4.1 Krok 1: Stabilizace chromatinu a buněčná lyze

Nejdříve byla změřena koncentrace HG-3 buněk v suspenzi pomocí fluorescenčního hemocytometru Dual Fluorescence Cell Counter LUNA-FL™. Pro vstupní krok protokolu Micro-C bylo použito 1×10^6 buněk, což odpovídá doporučení výrobce. Buňky byly vystaveny teplotnímu šoku uložením buněčného peletu na -80 °C. Po cca 60 minutách byl indukován crosslinking působením 37% formaldehydu. Následovalo štěpení chromatinu pomocí MNázy – enzymu, který štěpí nespecificky v celém genomu, avšak v pravidelných rozestupech. Vznikají tak fragmenty s přibližně stejnou délkou. Po 15 minutách byla enzymatická reakce zastavena přidáním kyseliny egtázové (EGTA). Poté proběhla buněčná lyze. Část lyzátu byla použita pro kontrolu kvality, zbylý materiál byl uchován pro použití v dalším kroku za předpokladu, že kontrola kvality poskytla uspokojivé výsledky.

Kontrola kvality spočívala v izolaci a přečištění DNA pomocí DNA Clean & Concentrator™-5 kitu, který využívá centrifugační kolonky. Následně byla purifikovaná DNA kvantifikována na Qubit® fluorometru.

7.4.2 Krok 2: Značení biotinem a rozrušení crosslinků

Pro přípravu Hi-C knihovny bylo použito 1000 ng buněčného lyzátu získaného v předchozím kroku protokolu. Po izolaci crosslinkovaného chromatinu proběhla úprava konců DNA fragmentů a značení biotinem. Konce fragmentů byly spojeny ligací. Následně došlo k rozrušení crosslinků působením proteinázy K. DNA byla přečištěna za použití magnetických kuliček. Bylo získáno 50 µl supernatantu. Purifikovaná DNA byla kvantifikována pomocí Qubit® fluorometru. Pro další krok jsme použili 150 ng DNA dle doporučení výrobce.

7.4.3 Krok 3: Příprava Hi-C knihovny

Tento krok zahrnoval přípravu sekvenační knihovny. K tomu byl využit kit Ultra™ II DNA Library Prep Kit for Illumina® od firmy NEBNext®. Nejdříve proběhla úprava konců DNA a navázání adaptérů – oligonukleotidových sekvencí potřebných pro přichycení fragmentů DNA k sekvenační komůrce v průběhu sekvenování. Adaptéry obsahovaly specifické

sekvence, tzv. indexy, které při sekvenování více knihoven naráz umožňují rozpoznat, která čtení patří jakému vzorku. Následovalo přečištění DNA pomocí magnetických kuliček.

Biotinylované fragmenty byly následně zachyceny na kuličkách potažených streptavidinem. Takto připravený vzorek byl amplifikován pomocí Hot Start DNA polymerázy v termocycleru Biometra TAdvanced, kde se nastavil následující program, viz Tabulka 2.

Fáze	Teplota	Čas	Počet cyklů
Aktivace enzymů	98 °C	3 minuty	1
Denaturace	98 °C	20 sekund	12
Hybridizace	65 °C	30 sekund	
Elongace	72 °C	30 sekund	
Extenze	72 °C	1 minuta	1
	12 °C	Podržet do vyzvednutí vzorku	1

Tabulka 2: Nastavení programu termocycleru Biometra TAdvanced.

Následně proběhlo odstranění krátkých fragmentů pomocí magnetických kuliček. Kvantifikace připravené knihovny byla opět provedena na Qubit® fluorometru. Pomocí přístroje Fragment Analyzer™ Automated CE System jsme ověřili distribuci velikostí fragmentů, která se pohybovala v doporučeném rozmezí 350 až 1000 bp.

7.5 Dovetail™ Omni-C Kit

Vzhledem k velké podobnosti protokolů Dovetail Omni-C kit a Dovetail Micro-C zde budou z protokolu Omni-C vybrány pouze odchylky, které však nemění princip přípravy knihovny.

7.5.1 Krok 1: Stabilizace chromatinu a buněčná lyze

Nejpodstatnějším rozdílem mezi kity od firmy Dovetail je použití různých štěpicích enzymů. Součástí Omni-C kitu je DNáza, která štěpí náhodně v celém genomu s nepravidelnými rozestupy. Po štěpení tedy očekáváme fragmenty různé délky. Inkubace buněk s enzymem trvá 30 minut, je tedy dvakrát delší oproti Micro-C protokolu. Následovala kvantifikace vzorku na přístroji Qubit®. Délka získaných fragmentů měřená přístrojem Fragment Analyzer™ Automated CE System.

7.5.2 Krok 2: Značení biotinem a rozrušení crosslinků

Postup při tomto kroku se neodlišuje od Micro-C protokolu. Použili jsme opět 1000 ng materiálu získaného v předchozím kroku. Po purifikaci rozrušených crosslinků jsme pomocí Qubit® fluorometru naměřili koncentraci. Pro další krok jsme použili 150 ng DNA dle doporučení výrobce.

7.5.3 Krok 3: Příprava Hi-C knihovny

Omni-C knihovna byla připravena stejně jako Micro-C knihovna. Kvantifikace připravené knihovny byla provedena na Qubit® fluorometru. Pomocí přístroje Fragment Analyzer™

Automated CE System jsme ověřili distribuci velikostí fragmentů, která se pohybovala v doporučeném rozmezí 350 až 1 000 bp.

7.6 Sekvenování

Sekvenování připravených knihoven proběhlo na přístroji NextSeq™ 500 Sequencing System v Centrální laboratoři Genomika, CEITEC MU. Pro knihovnu připravenou pomocí EpiTect Hi-C kitu byla použita sekvenační chemie NextSeq 500/550 Mid Output Kit v2.5 (150 cycles) (Illumina) dle doporučení výrobce. Bylo získáno 8,7 milionů sekvenačních čtení. Zbylé dvě knihovny připravené pomocí chemie od firmy Dovetail se sekvenovaly kitem NextSeq 500/550 Mid Output Kit v2.5 (300 cycles) (Illumina). Výsledný počty čtení byly 4,7 milionů čtení (Micro-C) a 3,6 milionů čtení (Omni-C).

7.7 Analýza dat

Zpracování a vyhodnocení dat provedl bioinformatik Mgr. Ing. Jakub Paweł Porc. Nejdříve bylo nutné získaná data zkontrolovat a upravit pro další analýzu. Standardní kontrola kvality surových sekvenačních dat proběhla pomocí nástroje FastQC. Další kontrola kvality byla zaměřena na ověření spolehlivosti detekovaných interakcí. Pro knihovnu EpiTect byl použit nástroj HiC-Pro, pro knihovny připravené chemií od firmy Dovetail program vyvinutý výrobcem kitů. Knihovny EpiTect a Micro-C měly velmi dobré statistiky, avšak data získaná z Omni-C knihovny vykazovala nízkou rozmanitost, což svědčí o hroší kvalitě sekvenační knihovny. Na základě hodnoty tzv. MAPQ skóre (= ukazatel kvality mapování získaných sekvenačních čtení na genomickou sekvenci) byly odstraněny falešně pozitivní interakce.

Pomocí nástroje HiC-Pro a softwaru od firmy Dovetail byly vygenerovány tzv. interakční matice. Tyto matice poskytují informace o tom, kolik interakcí bylo detekováno mezi konkrétními genomickými oblastmi. Interakční matice byly poté upraveny pomocí speciálního algoritmu (tzv. Knight-Ruiz matrix balancing).

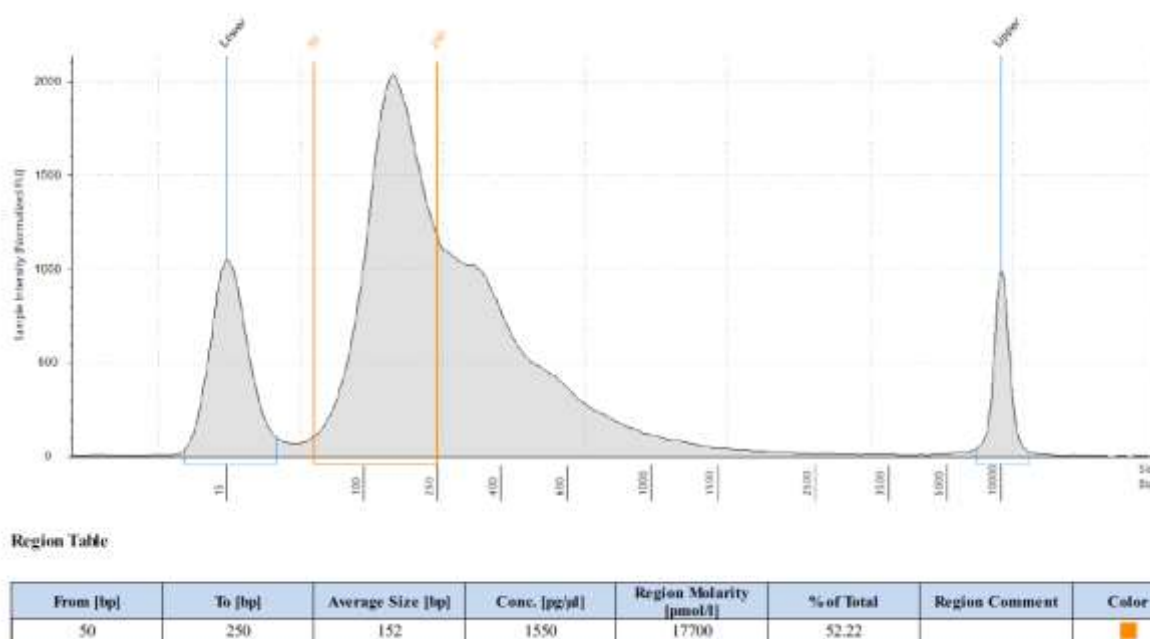
Data z interakčních matic byla promítnuta do heat map pomocí nástroje Juicebox. Nástroj umožňuje tvorbu heat map s různým rozlišením, přidání různých anotací (například domén) a porovnání 2 vzorků navzájem. Následovalo statistické vyhodnocení významnosti interakcí promítnutých do heat map za použití nástroje Fit-Hi-C. Hodnocení bylo provedeno zvlášť pro inter- a intrachromozomální interakce, aby se zvýšila statistická významnost. Pomocí programu hic_breakfinder byly detekovány strukturní varianty. K předběžné identifikaci TADs byl použit nástroj TopDom. Domény byly detekovány v rozlišení 50 kb. Pro přesnou analýzu však bude nutné získat rozsáhlejší soubor dat.

Na závěr bylo nutné zhodnotit reprodukovatelnost dat. K tomu byl využit nástroj IDR2D, který počítá tzv. Irreproducible Discovery Rate (IDR) hodnotu pro všechny genomické interakce. Určuje, kolik interakčních oblastí mezi 2 interakčními maticemi se statisticky významně liší.

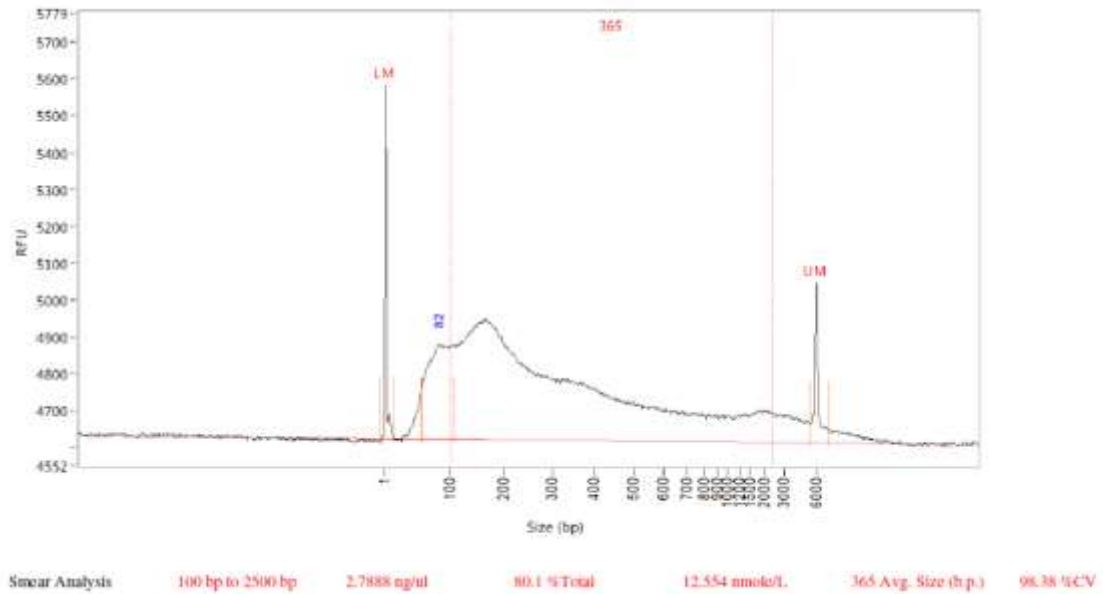
8 VÝSLEDKY

8.1 Kontrola fragmentačních profilů

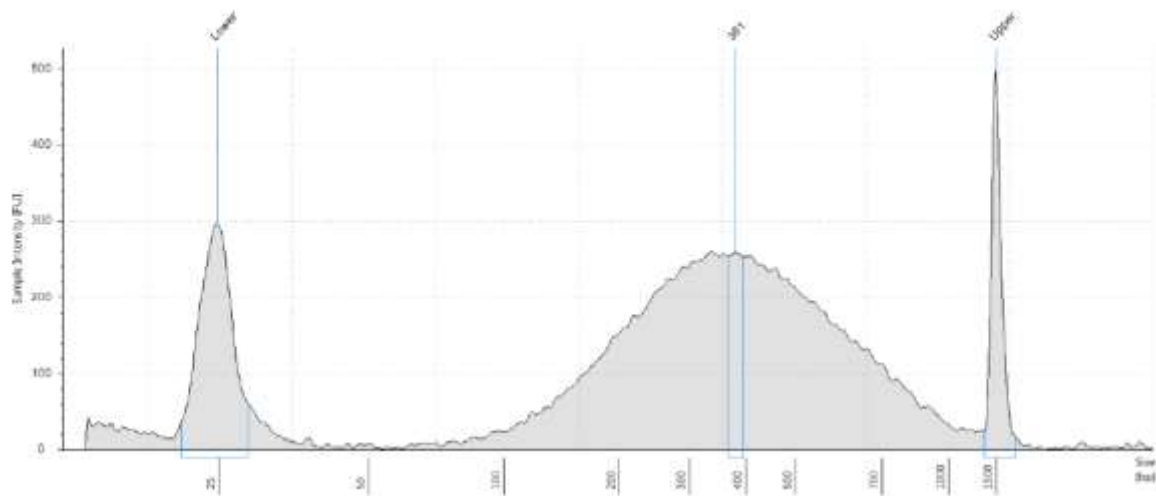
Délky získaných fragmentů jsme ověřili na přístroji 2200 TapeStation System a Fragment Analyzer™ Automated CE System. Pro kity Dovetail™ Micro-C Kit a Dovetail™ Omni-C Kit jsme zjišťovali délky fragmentů po chemické fragmentaci. EpiTect Hi-C Kit byl následně podroben ještě mechanické fragmentaci, fragmentační profil jsme tedy kontrolovali až v pozdějším kroku.



Obrázek 10: Kontrola fragmentačního profilu po štěpení chemií Micro-C na přístroji TapeStation. Dle výrobce by po štěpení měly být detekovány fragmenty dlouhé přibližně 146 bp a 40-70 % všech získaných fragmentů by mělo být v rozsahu 50-250 bp. Námi získané fragmenty měly průměrnou délku 152 bp a rozsah 50-250 bp zaujímalo 52,22 % všech získaných fragmentů. Štěpení tedy proběhlo úspěšně.



Obrázek 11: Kontrola fragmentačního profilu po štěpení chemií Omni-C na přístroji Fragment Analyzer. Dle výrobce by po štěpení měly být detekovány fragmenty dlouhé 50 až 5000 bp a více než 50 % všech získaných fragmentů by mělo být v rozsahu 100 až 2500 bp. Námi získané fragmenty měly průměrnou délku 365 bp a rozsah 100-2500 bp zaujímalo 80,1 % všech získaných fragmentů. Štěpení tedy proběhlo úspěšně.



Obrázek 12: Kontrola fragmentačního profilu po štěpení chemií EpiTect Hi-C a mechanickém štěpení na přístroji Covaris. Dle výrobce je po štěpení očekávána průměrná délka fragmentů 400 až 600 bp. Průměrná délka námi získaných fragmentů byla o něco menší, nicméně fragmenty v očekávaném rozmezí jsme získali.

8.2 Kontrola kvality sekvenačních knihoven a kvantifikace na Qubit® fluorometru

V závěru všech protokolů bylo potřeba opět ověřit kvalitu všech připravených knihoven pomocí přístroje 2200 TapeStation System a Fragment Analyzer™ Automated CE System. Zde jsme získali velmi podobné hodnoty, které splňovaly doporučené rozmezí 350 až 1000 bp.

Dovetail™ Micro-C Kit a Dovetail™ Omni-C Kit vyžadovaly kvantifikaci na Qubit® fluorometru jak po stabilizaci chromatinu a buněčné lyzi (krok 1), tak po značení biotinem a rozrušení crosslinků (krok 2), ale i na konci protokolu při ověřování kvality připravených knihoven (krok 3). Pro EpiTect Hi-C Kit se stanovovala kontrola až na konci celé přípravy sekvenční knihovny.

	po 1. kroku	po 2. kroku	po 3. kroku
Dovetail™ Micro-C Kit	15,5 ng/μl	7,74 ng/μl	32,6 ng/μl
Dovetail™ Omni-C Kit	2,54 ng/μl	9,99 ng/μl	52 ng/μl
EpiTect Hi-C Kit	-	-	0,483 ng/μl

Tabulka 3: Výsledné koncentrace pro jednotlivé kity po každém kroku.

Výsledná koncentrace EpiTect Hi-C knihovny byla poměrně nízká, 0,483 ng/μl, nicméně pro sekvenační reakci bylo toto množství dostačující.

8.3 Kontrola kvality sekvenačních dat

Získaná data bylo potřeba následně zkontrolovat. Z celkového počtu párových čtení nás zajímala mapovaná párová čtení, která byla úspěšně mapována na referenční genomickou sekvenci. Jako výsledná párová čtení jsou pak myšlena mapovaná čtení, která zůstala po odfiltrování falešně pozitivních interakcí.

	EpiTect Hi-C Kit	Dovetail™ Micro-C Kit	Dovetail™ Omni-C Kit
Celkový počet párových čtení	$8,71 \times 10^6$ (100 %)	$4,72 \times 10^6$ (100 %)	$3,59 \times 10^6$ (100 %)
Mapovaná párová čtení	$8,58 \times 10^6$ (98 %)	$4,44 \times 10^6$ (94 %)	$3,25 \times 10^6$ (91 %)
Výsledná párová čtení	$3,77 \times 10^6$ (40 %)	$3,63 \times 10^6$ (76 %)	$2,76 \times 10^6$ (72 %)

Tabulka 4: Výsledná statistika kontroly kvality sekvenačních dat pro jednotlivé kity.

Následně jsme stanovili podíly jednotlivých druhů párových čtení pro každý kit. *Cis* čtení pochází z intrachromozomálních interakcí, tedy interakcí v rámci jednoho chromozomu, *trans* z interchromozomálních interakcí, což jsou interakce mezi dvěma různými chromozomy.

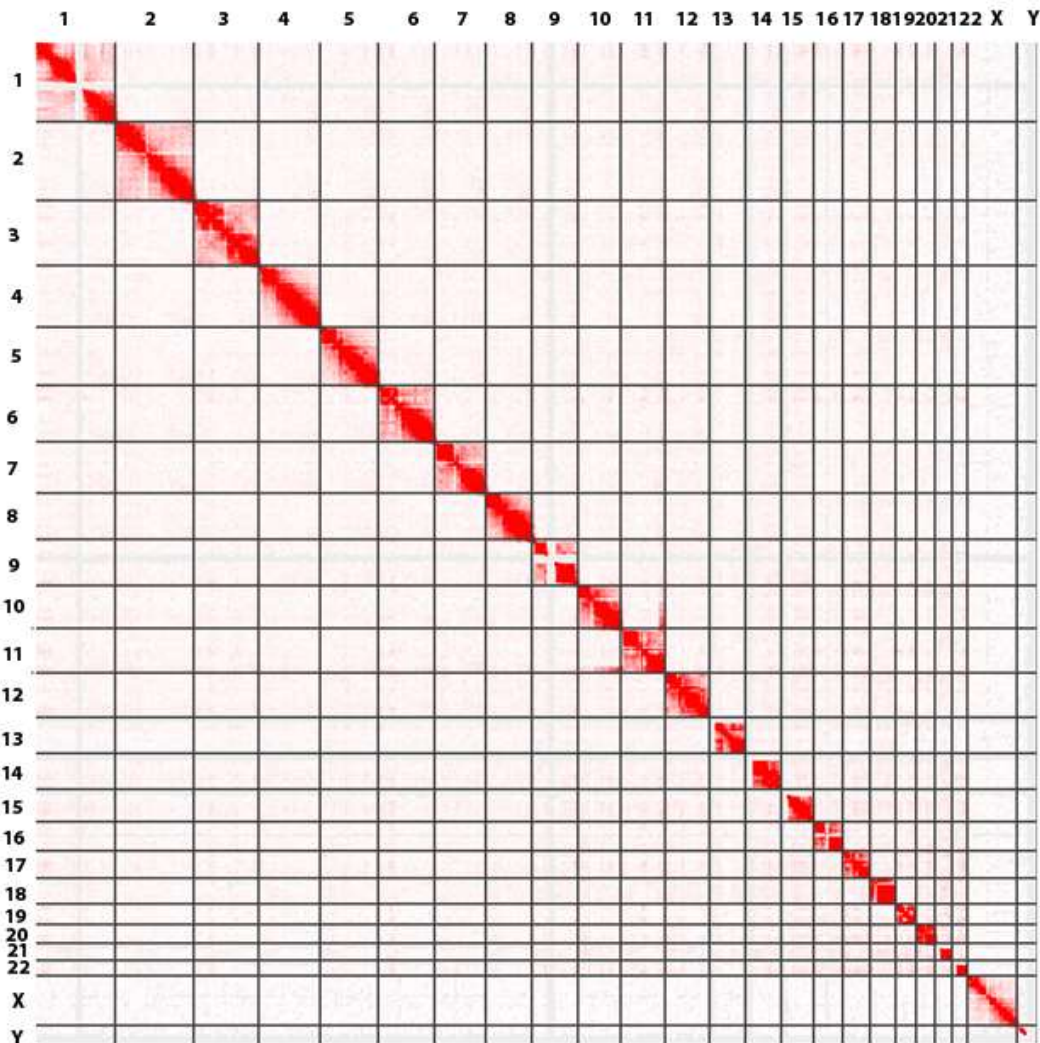
	EpiTect Hi-C Kit	Dovetail™ Micro-C Kit	Dovetail™ Omni-C Kit
<i>Cis</i> párové čtené	$2,86 \times 10^6$ (83 %)	$2,76 \times 10^6$ (76 %)	$2,19 \times 10^6$ (85 %)
<i>Cis</i> < 1kb	-	$0,34 \times 10^6$ (9 %)	$0,81 \times 10^6$ (31 %)
<i>Cis</i> < 20 kb	$0,50 \times 10^6$ (14 %)	-	-
<i>Trans</i> párové čtení	$0,61 \times 10^6$ (17 %)	$0,83 \times 10^6$ (23 %)	$0,39 \times 10^6$ (15 %)

Tabulka 5: Podíl jednotlivých druhů párových čtení.

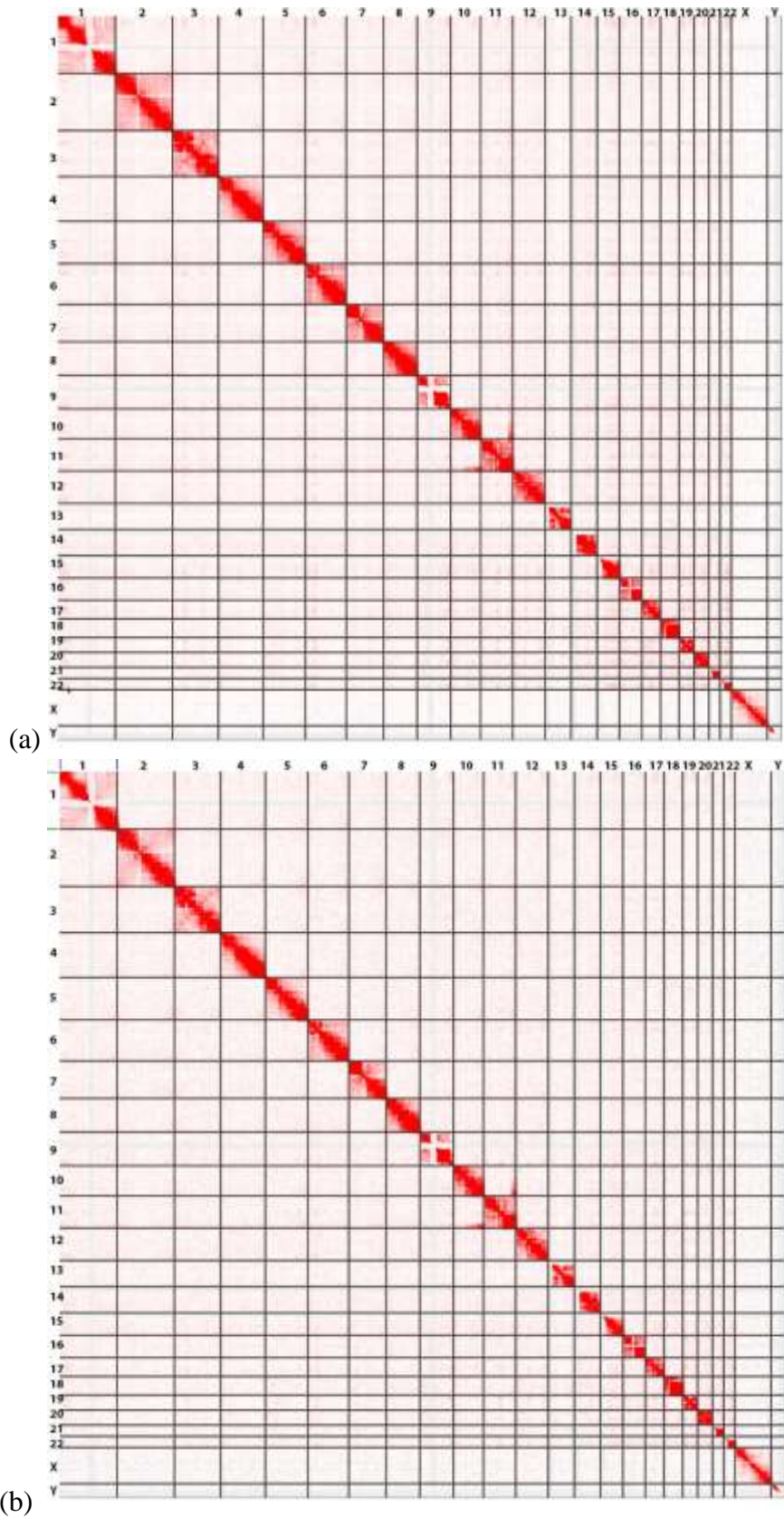
8.4 Vizualizace chromatinových interakcí

Typickým výstupem získaným z Hi-C dat je heat mapa. Osy x a y znázorňují postupně řazené pozice celého genomu, jednoho či více chromozomů či pouze konkrétní chromozomální oblasti. Každý bod v heatmapě představuje prostor pro znázornění možné interakce. Můžeme zde vidět intrachromozomální interakce na diagonále a interchromozomální interakce mimo diagonálu. Frekvence interakce je znázorněna intenzitou červené barvy, takže bílé oblasti neobsahují žádné interakce.

První heat mapa znázorňuje genomické interakce v buněčné linii HG-3 pro všechny chromozomy a je získaná pomocí EpiTect Hi-C kitu. Další dvě heat mapy jsou získané pomocí Micro-C a Omni-C kitu.

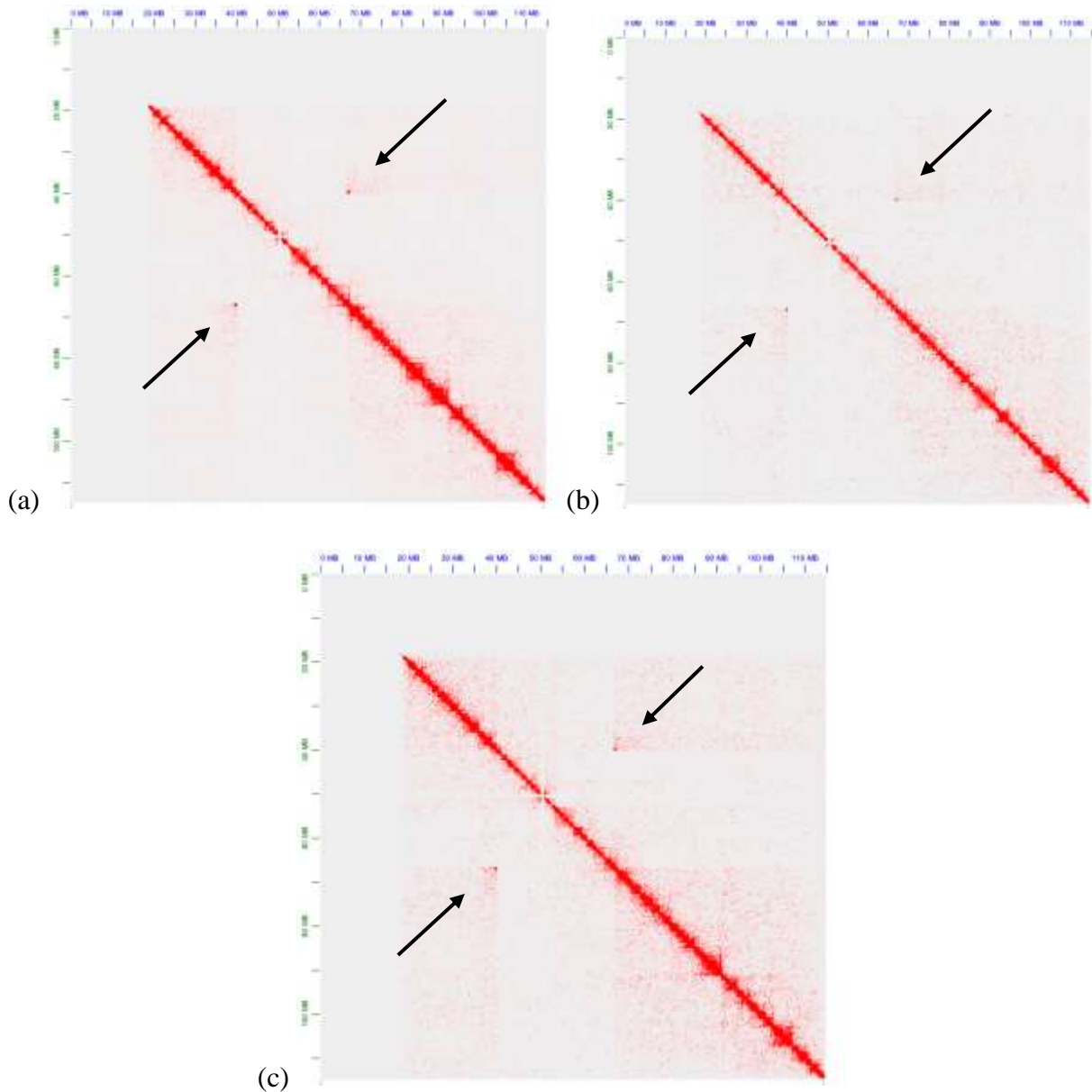


Obrázek 13: Genomické interakce v buněčné linii HG-3 zjištěné pomocí EpiTect Hi-C kitu. Každý čtyřúhelník v heatmapě představuje prostor pro znázornění možné interakce v genomu. Na osách jsou zleva doprava a shora dolů seřazeny pozice na chromozomech (od chromozomu 1 až po Y).



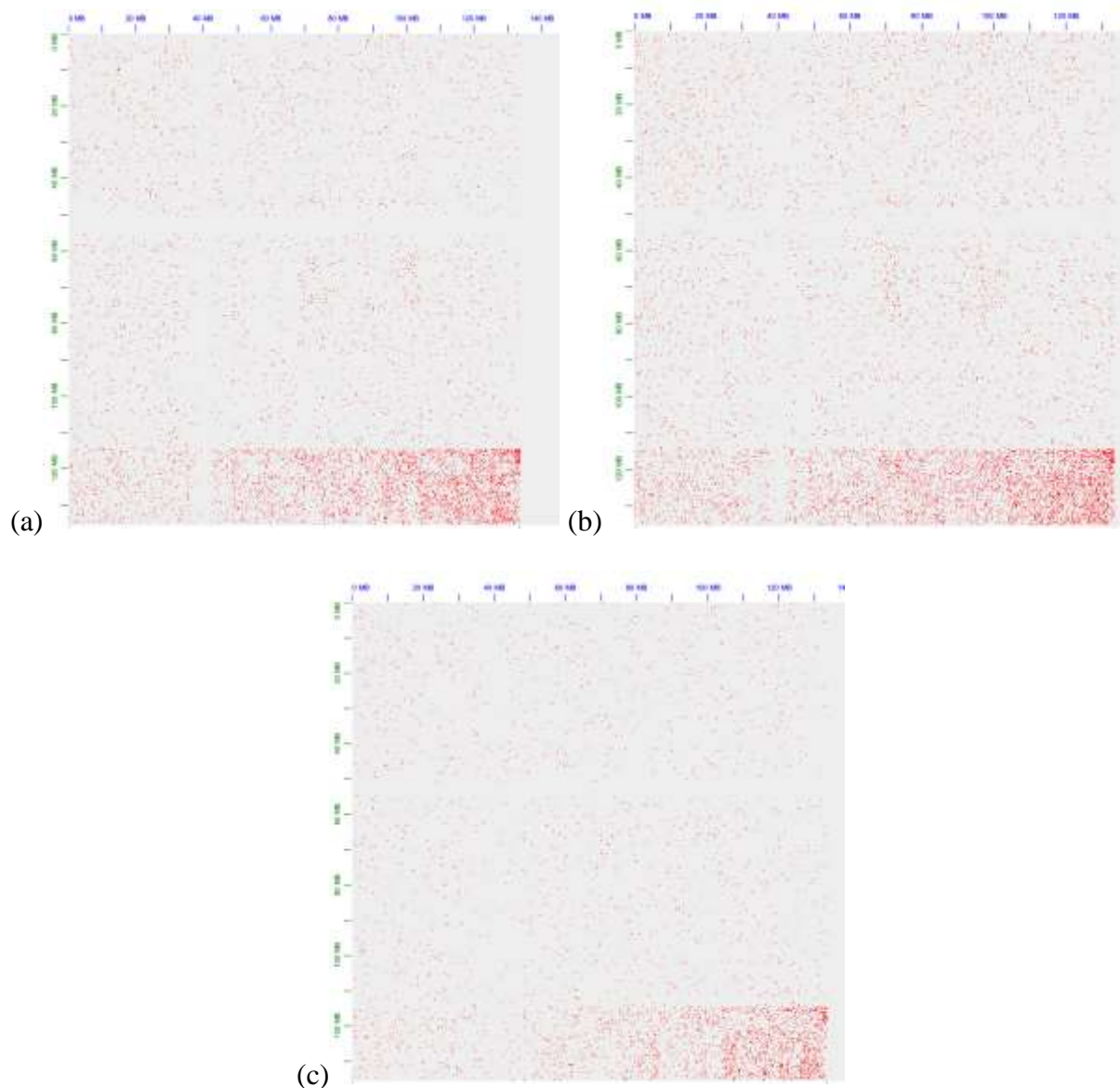
Obrázek 14: Genomické interakce v buněčné linii HG-3 zjištěné pomocí Micro-C (a) a Omni-C (b) kitu.
 Na osách jsou zleva doprava a shora dolů seřazeny pozice na chromozomech (od chromozomu 1 až po Y).

Následující heat mapy znázorňují konkrétní zjištěné aberace. První tři heat mapy vizualizují intrachromozomální interakce na chromozomu 13. Je zde dobře viditelná silná interakce mezi oblastmi kolem 40 MB a 67 MB, což naznačuje jejich fyzické přiblížení v důsledku možné ztráty části chromozomu mezi těmito oblastmi. Delece byla následně prokázána cytogenetickými metodami.



Obrázek 15: Intrachromozomální interakce chromozomu 13 zjištěné pomocí EpiTect Hi-C kitu (a), Omni-C kitu (b) a Micro-C kitu (c). Heat mapa porovnává interakce v rámci chromozomu 13, na obou osách jsou seřazeny pozice na chromozomu 13. Je viditelná interakce mezi genomovými oblastmi kolem 40 MB a 67 MB.

Poslední sada heat map znázorňuje interchromozomální interakce mezi chromozomem 10 na ose x a chromozomem 11 na ose y. Oblast chromozomu 11 mezi 115–135 MB má vyšší intenzitu interakcí s koncem chromozomu 10. Interakce je slabší s nižší genomickou pozicí na chromozomu 10. To naznačuje možnou translokaci oblasti z chromozomu 11 na chromozom 10.



Obrázek 16: Interchromozomální interakce mezi chromozomem 10 (osa x) a chromozomem 11 (osa y) zjištěné pomocí EpiTect Hi-C kitu (a), Omni-C kitu (b) a Micro-C kitu (c). Oblast chromozomu 11 mezi 115-135 MB má vyšší intenzitu interakcí s koncem chromozomu 10. To naznačuje možnou translokaci oblasti z chromozomu 11 na chromozom 10.

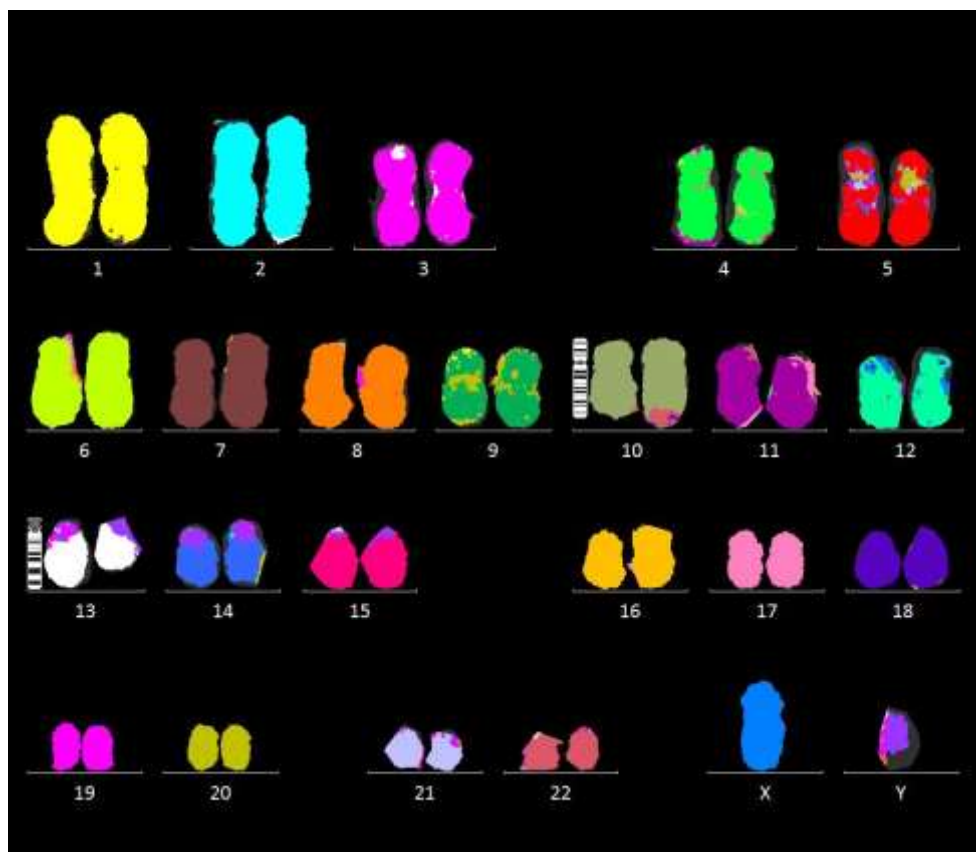
8.5 Strukturní varianty

V následující tabulce jsou shrnuty strukturní varianty nalezené ve vzorcích pomocí programu hic_breakfinder. Pro jednotlivé kity jsou uvedeny konkrétní pozice v genomu, kde byly strukturní varianty detekovány.

EpiTect Hi-C Kit	Dovetail™ Micro-C Kit	Dovetail™ Omni-C Kit
chr10:128Mb-134Mb a chr11:114Mb-119Mb	chr10:125Mb-132Mb a chr11:122Mb-134Mb	chr10:126Mb-132Mb a chr11:126Mb-135Mb
chr13:39,76Kb-39,95Kb a chr13:66,56Kb-66,67Kb	-	chr13:39,7Kb-40Kb a chr13:66,5Mb-67,5Mb

Tabulka 6: Strukturní varianty v buňkách HG-3 zjištěné pomocí jednotlivých kitů. S Dovetail™ Micro-C Kitem jsme nebyli schopni zachytit aberaci na 13 chromozomu, což mohlo být zapříčiněno příliš krátkými fragmenty získanými při přípravě knihovny. Všechny kity umožnily detekci přestavby mezi chromozomy 10 a 11.

Zjištěné aberace jsme porovnali s karyotypem majoritního klonu linie HG-3 sestaveným pomocí metody M-FISH. Výsledek laskavě poskytly pracovnice Sekce nádorové cytogenetiky z Centra molekulární biologie a genetiky z FN Brno. Pro hybridizaci byly použity malovací sondy MetaSystems 24Xcyte. Snímky byly získány na fluorescenčním mikroskopu Axio Imager Z2. Výsledek cytogenetiky potvrdil delecii na chromozomu 13 a translokaci mezi chromozomy 10 a 11, které jsme nezávisle pozorovali metodou Hi-C.



46, XY, der(10)t(10;11)(q?26;q?24.1),del(13)(q?12q32)

Obrázek 17: Karyotyp hlavního klonu linie HG-3 zjištěný metodou M-FISH. Oblasti zlomů byly určeny metodou chromozomálního G pruhování. Výsledek cytogenetiky potvrdil delecii na chromozomu 13 a translokaci mezi chromozomy 10 a 11.

9 DISKUZE

Co se týče samotné přípravy Hi-C knihoven, narazili jsme na několik překážek, které bylo nutné překonat. Jako nejobtížnější se jeví optimalizace koncentrace štěpícího enzymu používaného pro fragmentaci crosslinkovaného chromatinu. Zjistili jsme, že je potřeba důsledně dodržování množství enzymu, neboť i malá odchylka v ředění enzymu může způsobit nadměrnou nebo naopak nedostatečnou fragmentaci. Důležité je rovněž přesně dodržet dobu inkubace. Je vhodné připravit různá ředění enzymu a pro další krok poté zvolit vzorek s optimálním fragmentačním profilem. Takto bude v budoucnu potřeba přistupovat i k analýze jednotlivých patientských vzorků.

Všechny kity poskytly vysoký počet mapovaných čtení (> 90 %). Při kontrole kvality byl nastaven velmi přísný limit hodnoty MAPQ skóre ($MAPQ > 40$), aby byl odstraněn šum a jakékoli nesprávné interakce, což výrazně snížilo počet čtení vybraných pro další analýzu. Správně připravená Hi-C knihovna by měla vykazovat nízký počet interakcí *cis* na krátkou vzdálenost (u kitů od firmy Dovetail je to méně než 20 % *cis* párových čtení, která jsou menší než 1 kB) a ne více než 25 % *trans* párových čtení. Aktuálně pokračujeme na vylepšování a optimalizaci protokolu. Také plánujeme zopakovat sekvenování všech knihoven, abychom získali další data.

Promítnutím interakcí do heat map jsme byli schopni vizualizovat frekvence interakcí mezi všemi chromozomy. Pozorovali jsme vysokou frekvenci interakcí na diagonále matice a nízkou frekvenci mimo ni, což znamená, že většina interakcí probíhá mezi oblastmi na stejném chromozomu. Toto je očekávaný jev v lidských buňkách, který poukazuje na věrohodnost našich dat. Interakční signály ve vzorcích připravených pomocí Micro-C a EpiTect Hi-C kitu byly intenzivnější ve srovnání s Omni-C, což je způsobeno nižší kvalitou Omni-C knihovny. Kromě toho jsme byli schopni identifikovat aberaci na chromozomu 13, která se jevila jako delece, a aberaci mezi chromozomy 10 a 11, kde jsme předpokládali, že se jedná o translokaci. Přítomnost těchto aberací byla potvrzena cytogenetickými metodami.

Analýzou strukturních variant pomocí nástroje `hic_breakfinder` jsme přesně určili polohu aberace na chromozomu 13 pro všechny kity. Interakce mezi chromozomy 10 a 11 nebyla detekována s takovou přesností, což ovšem má logické vysvětlení. Dle tvůrců programu Fit-Hi-C, který hodnotí statistickou významnost nalezených interakcí, je obtížné identifikovat významné interchromozomální interakce, protože celkové množství intrachromozomálních interakcí je mnohem větší. K nalezení těchto interakcí je navíc zapotřebí velmi vysoké rozlišení dat. Taková data v tuto chvíli bohužel nemáme k dispozici.

10 ZÁVĚR

Práce se zabývá analýzou genetického profilu CLL buněk. Praktická část byla zaměřena především na provedení metody Hi-C, porovnání tří různých komerčních kitů a optimalizaci protokolu. Analýza dat získaných sekvenováním připravených Hi-C knihoven potvrdila genetické abnormality buněčné linie HG-3, zjištěné cytogenetickými metodami. Co se týče genetických abnormalit, podařilo se nám identifikovat aberaci na chromozomu 13 a aberaci mezi chromozomy 10 a 11. Veškeré stanovené cíle práce tímto byly splněny.

V následujících měsících navážeme experimenty zaměřenými na optimalizaci metody na vzorcích CLL pacientů. Postup bude poté využit ke studiu většího souboru patientských vzorků uložených v biobance Centra molekulární biologie a genetiky, FN Brno.

11 SEZNAM POUŽITÉ LITERATURY

- [1] Vztah struktury a funkce nukleových kyselin. Replikace, transkripce [online]. no date [accessed. 2021-03-07]. Dostupné z: http://old.vscht.cz/anl/oppa-sem-bio/PDFs/II_11_OPPA.pdf
- [2] ROSYPAL, Stanislav and et. al. *Úvod do molekulární biologie*. 2. vydání. 1997.
- [3] URBAN, T. and T. VYHNÁNEK. Genové mutace [online]. no date [accessed. 2021-03-14]. Dostupné z: http://user.mendelu.cz/urban/vsg1/molekul/mol_mutace1.html
- [4] URBAN, T. and T. VYHNÁNEK. Cytogenetika - struktura chromozomu [online]. no date [accessed. 2021-03-14]. Dostupné z: http://user.mendelu.cz/urban/vsg1/mendel/klas_cytogen.html
- [5] *Mutace* [online]. [accessed. 2021-03-14]. Dostupné z: <http://biocentrum.zf.jcu.cz/docs/ruzne/ruz-8.-Mutace-50637c3a1e.pdf>
- [6] TAKIZAWA, Takumi and Eran MESHORER. Chromatin and nuclear architecture in the nervous system. *Trends in Neurosciences* [online]. 2008, **31**(7). ISSN 01662236. Dostupné z: doi:10.1016/j.tins.2008.03.005
- [7] LUGER, Karolin, Armin W. MÄDER, Robin K. RICHMOND, David F. SARGENT and Timothy J. RICHMOND. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* [online]. 1997, **389**(6648). ISSN 0028-0836. Dostupné z: doi:10.1038/38444
- [8] BESHNOVA, Daria A., Andrey G. CHERSTVY, Yevhen VAINSHTEIN and Vladimir B. TEIF. Regulation of the Nucleosome Repeat Length In Vivo by the DNA Sequence, Protein Concentrations and Long-Range Interactions. *PLoS Computational Biology* [online]. 2014, **10**(7). ISSN 1553-7358. Dostupné z: doi:10.1371/journal.pcbi.1003698
- [9] BARBARA BOWLES BIESECKER, Ph.D. Karyotype [online]. no date [accessed. 2021-03-07]. Dostupné z: <https://www.genome.gov/genetics-glossary/Karyotype>
- [10] TEIF, Vladimir B. and Karsten RIPPE. Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities. *Nucleic Acids Research* [online]. 2009, **37**(17). ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkp610
- [11] MORALES, Violette and Hélène RICHARD-FOY. Role of Histone N-Terminal Tails and Their Acetylation in Nucleosome Dynamics. *Molecular and Cellular Biology* [online]. 2000, **20**(19). ISSN 1098-5549. Dostupné z: doi:10.1128/MCB.20.19.7230-7237.2000
- [12] TAN, Minjia, Hao LUO, Sangkyu LEE, Fulai JIN, Jeong Soo YANG, Emilie MONTELLIER, Thierry BUCHOU, Zhongyi CHENG, Sophie ROUSSEAUX, Nisha RAJAGOPAL, Zhike LU, Zhen YE, Qin ZHU, Joanna WYSOCKA, Yang YE, Saadi KHOCHBIN, Bing REN and Yingming ZHAO. Identification of 67 Histone Marks and Histone

Lysine Cronylation as a New Type of Histone Modification. *Cell* [online]. 2011, **146**(6). ISSN 00928674. Dostupné z: doi:10.1016/j.cell.2011.08.008

- [13] STRAHL, Brian D. and C. David ALLIS. The language of covalent histone modifications. *Nature* [online]. 2000, **403**(6765). ISSN 0028-0836. Available at: doi:10.1038/47412
- [14] VIGNALI, Marissa, Ahmed H. HASSAN, Kristen E. NEELY and Jerry L. WORKMAN. ATP-Dependent Chromatin-Remodeling Complexes. *Molecular and Cellular Biology* [online]. 2000, **20**(6). ISSN 1098-5549. Dostupné z: doi:10.1128/MCB.20.6.1899-1910.2000
- [15] LUSSER, Alexandra and James T. KADONAGA. Chromatin remodeling by ATP-dependent molecular machines. *BioEssays* [online]. 2003, **25**(12). ISSN 0265-9247. Available at: doi:10.1002/bies.10359
- [16] DIXON, Jesse R., David U. GORKIN and Bing REN. Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell* [online]. 2016, **62**(5). ISSN 10972765. Available at: doi:10.1016/j.molcel.2016.05.018
- [17] GIBCUS, Johan H. and Job DEKKER. The Hierarchy of the 3D Genome. *Molecular Cell* [online]. 2013, **49**(5). ISSN 10972765. Available at: doi:10.1016/j.molcel.2013.02.011
- [18] WEINREB, Caleb and Benjamin J. RAPHAEL. Identification of hierarchical chromatin domains. *Bioinformatics* [online]. 2016, **32**(11). ISSN 1367-4803. Available at: doi:10.1093/bioinformatics/btv485
- [19] ROCHA, Pedro P, Ramya RAVIRAM, Richard BONNEAU and Jane A SKOK. Breaking TADs: insights into hierarchical genome organization. *Epigenomics* [online]. 2015, **7**(4). ISSN 1750-1911. Available at: doi:10.2217/epi.15.25
- [20] KADAUKE, Stephan and Gerd A. BLOBEL. Chromatin loops in gene regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* [online]. 2009, **1789**(1). ISSN 18749399. Available at: doi:10.1016/j.bbagrm.2008.07.002
- [21] GHAVI-HELM, Yad, Aleksander JANKOWSKI, Sascha MEIERS, Rebecca R. VIALES, Jan O. KORBEL and Eileen E. M. FURLONG. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics* [online]. 2019, **51**(8). ISSN 1061-4036. Available at: doi:10.1038/s41588-019-0462-3
- [22] SMITH, Emily M., Bryan R. LAJOIE, Gaurav JAIN and Job DEKKER. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *The American Journal of Human Genetics* [online]. 2016, **98**(1). ISSN 00029297. Available at: doi:10.1016/j.ajhg.2015.12.002

- [23] LUPIÁÑEZ, Darío G., Malte SPIELMANN and Stefan MUNDLOS. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends in Genetics* [online]. 2016, **32**(4). ISSN 01689525. Available at: doi:10.1016/j.tig.2016.01.003
- [24] HAKIM, Ofir and Tom MISTELI. SnapShot: Chromosome Conformation Capture. *Cell* [online]. 2012, **148**(5). ISSN 00928674. Available at: doi:10.1016/j.cell.2012.02.019
- [25] DE WIT, E. and W. DE LAAT. A decade of 3C technologies: insights into nuclear organization. *Genes & Development* [online]. 2012, **26**(1). ISSN 0890-9369. Available at: doi:10.1101/gad.179804.111
- [26] LI, Guoliang, Liuyang CAI, Huidan CHANG, Ping HONG, Qiangwei ZHOU, Ekaterina v KULAKOVA, Nikolay A KOLCHANOV and Yijun RUAN. Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics* [online]. 2014, **15**(S12). ISSN 1471-2164. Available at: doi:10.1186/1471-2164-15-S12-S11
- [27] HAN, Jinlei, Zhiliang ZHANG and Kai WANG. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Molecular Cytogenetics* [online]. 2018, **11**(1). ISSN 1755-8166. Available at: doi:10.1186/s13039-018-0368-2
- [28] MCLAUGHLIN, Kathryn. Chromosome Conformation Capture. *Materials and Methods* [online]. 2018, **8**. ISSN 2329-5139. Available at: doi:10.13070/mm.en.8.2662
- [29] GAVRILOV, Alexey, Elvira EIVAZOVA, Iryna PIROZHKOVA, Marc LIPINSKI, Sergey RAZIN and Yegor VASSETZKY. Chromosome Conformation Capture (from 3C to 5C) and Its ChIP-Based Modification. In: [online]. 2009. Available at: doi:10.1007/978-1-60327-414-2_12
- [30] BELTON, Jon-Matthew, Rachel Patton MCCORD, Johan Harmen GIBCUS, Natalia NAUMOVA, Ye ZHAN and Job DEKKER. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* [online]. 2012, **58**(3). ISSN 10462023. Available at: doi:10.1016/j.ymeth.2012.05.001
- [31] DIXON, Jesse R., Siddarth SELVARAJ, Feng YUE, Audrey KIM, Yan LI, Yin SHEN, Ming HU, Jun S. LIU and Bing REN. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* [online]. 2012, **485**(7398). ISSN 0028-0836. Available at: doi:10.1038/nature11082
- [32] RAO, Suhas S.P., Miriam H. HUNTLEY, Neva C. DURAND, Elena K. STAMENOVA, Ivan D. BOCHKOV, James T. ROBINSON, Adrian L. SANBORN, Ido MACHOL, Arina D. OMER, Eric S. LANDER and Erez Lieberman AIDEN. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* [online]. 2014, **159**(7). ISSN 00928674. Available at: doi:10.1016/j.cell.2014.11.021

- [33] DAVIES, James O J, A Marieke OUDELAAR, Douglas R HIGGS and Jim R HUGHES. How best to identify chromosomal interactions: a comparison of approaches. *Nature Methods* [online]. 2017, **14**(2). ISSN 1548-7091. Available at: doi:10.1038/nmeth.4146
- [34] KIM, Tae Hoon and Job DEKKER. Generation of ChIP-Loop Libraries. *Cold Spring Harbor Protocols* [online]. 2018, **2018**(8). ISSN 1940-3402. Available at: doi:10.1101/pdb.prot097857
- [35] MASSIE, Charles E. and Ian G. MILLS. Mapping Protein–DNA Interactions Using ChIP-Sequencing. In: [online]. 2012. Available at: doi:10.1007/978-1-61779-376-9_11
- [36] GILFILLAN, Gregor D, Timothy HUGHES, Ying SHENG, Hanne S HJORTH AUG, Tobias STRAUB, Kristina GERVIN, Jennifer R HARRIS, Dag E UNDLIEN and Robert LYLE. Limitations and possibilities of low cell number ChIP-seq. *BMC Genomics* [online]. 2012, **13**(1). ISSN 1471-2164. Available at: doi:10.1186/1471-2164-13-645
- [37] ZHANG, Jingyao, Huay Mei POH, Su Qin PEH, Yee Yen SIA, Guoliang LI, Fabianus Hendriyan MULAWADI, Yufen GOH, Melissa J. FULLWOOD, Wing-Kin SUNG, Xiaoan RUAN and Yijun RUAN. ChIA-PET analysis of transcriptional chromatin interactions. *Methods* [online]. 2012, **58**(3). ISSN 10462023. Available at: doi:10.1016/j.ymeth.2012.08.009
- [38] FULLWOOD, Melissa J., Mei Hui LIU, You Fu PAN, Jun LIU, Han XU, Yusoff bin MOHAMED, Yuriy L. ORLOV, Stoyan VELKOV, Andrea HO, Poh Huay MEI, Elaine G. Y. CHEW, Phillips Yao Hui HUANG, Willem-Jan WELBOREN, Yuyuan HAN, Hong Sain OOI, Pramila N. ARIYARATNE, Vinsensius B. VEGA, Yanquan LUO, Peck Yean TAN, Pei Ye CHOY, K. D. Senali Abayratna WANSA, Bing ZHAO, Kar Sian LIM, Shi Chi LEOW, Jit Sin YOW, Roy JOSEPH, Haixia LI, Kartiki v. DESAI, Jane S. THOMSEN, Yew Kok LEE, R. Krishna Murthy KARUTURI, Thoreau HERVE, Guillaume BOURQUE, Hendrik G. STUNNENBERG, Xiaoan RUAN, Valere CACHEUX-RATABOUL, Wing-Kin SUNG, Edison T. LIU, Chia-Lin WEI, Edwin CHEUNG and Yijun RUAN. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* [online]. 2009, **462**(7269). ISSN 0028-0836. Available at: doi:10.1038/nature08497
- [39] STARÝ, Jan and Jiří MAYER. *Leukemie*. Praha: Grada, 2002.
- [40] HALLEK, Michael. Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. *American Journal of Hematology* [online]. 2019, **94**(11). ISSN 0361-8609. Available at: doi:10.1002/ajh.25595
- [41] MOLICA, Stefano. Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. *Leukemia & Lymphoma* [online]. 2006, **47**(8). ISSN 1042-8194. Available at: doi:10.1080/10428190600555819
- [42] DÖHNER, Hartmut, Stephan STILGENBAUER, Axel BENNER, Elke LEUPOLT, Alexander KRÖBER, Lars BULLINGER, Konstanze DÖHNER, Martin BENTZ and Peter LICHTER. Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia. *New*

England Journal of Medicine [online]. 2000, **343**(26). ISSN 0028-4793. Available at: doi:10.1056/NEJM200012283432602

- [43] FABBRI, Giulia and Riccardo DALLA-FAVERA. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nature Reviews Cancer* [online]. 2016, **16**(3). ISSN 1474-175X. Available at: doi:10.1038/nrc.2016.8
- [44] KIPPS, Thomas J., Freda K. STEVENSON, Catherine J. WU, Carlo M. CROCE, Graham PACKHAM, William G. WIERDA, Susan O'BRIEN, John GRIBBEN and Kanti RAI. Chronic lymphocytic leukaemia. *Nature Reviews Disease Primers* [online]. 2017, **3**(1). ISSN 2056-676X. Available at: doi:10.1038/nrdp.2016.96
- [45] HAMBLIN, Terry J., Zadie DAVIS, Anne GARDINER, David G. OSCIER and Freda K. STEVENSON. Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. *Blood* [online]. 1999, **94**(6). ISSN 1528-0020. Available at: doi:10.1182/blood.V94.6.1848
- [46] O'MALLEY, D. P., C. GIUDICE, A. S. CHANG, D. CHANG, T. S. BARRY, M. K. HIBBARD, R. CHEN and S.-T. CHEN. Comparison of array comparative genomic hybridization (aCGH) to FISH and cytogenetics in prognostic evaluation of chronic lymphocytic leukemia. *International Journal of Laboratory Hematology* [online]. 2011, **33**(3). ISSN 17515521. Available at: doi:10.1111/j.1751-553X.2010.01284.x
- [47] QUENTMEIER, Hilmar, Claudia POMMERENKE, Ole AMMERPOHL, Robert GEFFERS, Vivien HAUER, Roderick AF MACLEOD, Stefan NAGEL, Julia ROMANI, Emanuela ROSATI, Anders ROSÉN, Cord C UPHOFF, Margarete ZABORSKI and Hans G DREXLER. Subclones in B-lymphoma cell lines: isogenic models for the study of gene regulation. *Oncotarget* [online]. 2016, **7**(39). ISSN 1949-2553. Available at: doi:10.18632/oncotarget.11524
- [48] ROSÉN, Anders, Ann-Charlotte BERGH, Peter GOGOK, Chamilly EVALDSSON, Anna Lanemo MYHRINDER, Eva HELLQVIST, Abu RASUL, Magnus BJÖRKHOLM, Mattias JANSSON, Larry MANSOURI, Anquan LIU, Bin Tean TEH, Richard ROSENQUIST and Eva KLEIN. Lymphoblastoid cell line with B1 cell characteristics established from a chronic lymphocytic leukemia clone by in vitro EBV infection. *OncoImmunology* [online]. 2012, **1**(1). ISSN 2162-402X. Available at: doi:10.4161/onci.1.1.18400

12 SEZNAM POUŽITÝCH ZKRATEK

3C technologie	technologie pro studium 3D struktury chromatinu (tzv. Chromosome Conformation Capture)
ATP	adenosintrifosfát
CLL	chronická lymfocytární leukémie
ChIP	chromatinová imunoprecipitace
ChIP-seq	ChIP-sequencing
IGHV	variabilní těžký řetězec imunoglobulinu
LCL	lymfoblastoidní buněčné linie
Mb	megabáze
mRNA	mediátorová RNA
NGS	metody nové generace sekvenování
PCR	polymerázová řetězová reakce
qPCR	kvantitativní polymerázová řetězová reakce
TADs	topologicky asociované domény
TGS	metody třetí generace sekvenování