

Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Speciální chemicko-biologické obory
Studijní obor: Molekulární biologie a biochemie organismů



Daniel Hašek

Analýza přestaveb genů pro imunoglobuliny a T-buněčné receptory pomocí sekvenování nové generace

The analysis of immunoglobulin and T-cell receptor gene rearrangements using next generation sequencing

Bakalářská práce

Školitel: MUDr. Eva Froňková, Ph.D.

Praha, 2015

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 5. 5. 2015

Daniel Hašek

Poděkování:

Rád bych poděkoval svojí školitelce MUDr. Evě Froňkové, Ph.D. za odborné vedení a věnovaný čas. Děkuji také svojí rodině za poskytnuté zázemí.

Obsah

Abstrakt.....	6
Abstract.....	7
1. Úvod.....	8
2. Sekvenování DNA.....	9
2.1 Klasická (Sangerova) metoda.....	9
2.1.1 Princip metody.....	9
2.1.2 Technická realizace.....	11
2.1.3 Využití metody.....	12
2.2 Sekvenování principem Roche 454.....	13
2.2.1 Princip metody.....	13
2.2.2 Technická realizace.....	14
2.2.3 Využití metody.....	16
2.3 Applied Biosystems SOLiD.....	17
2.3.1 Princip metody.....	17
2.3.2 Technická realizace.....	19
2.3.3 Využití metody.....	20
2.4 Ion Torrent by Life Technologies.....	20
2.4.1 Princip metody.....	20
2.4.2 Technická realizace.....	20
2.4.3 Využití metody.....	22
2.5 Illumina.....	22
2.5.1 Princip metody.....	23
2.5.2 Technická realizace.....	23
2.5.3 Využití metody.....	27
2.6 Další současné NGS metody a slibné metody ve vývoji.....	27
3. Využití NGS metod při detekci přestaveb genů pro Ig a TCR.....	29
3.1 Úvod do problematiky vývoje antigen-specifických receptorů T a B lymfocytů	29
3.1.1 Vznik imunoglobulinů.....	29
3.1.2 Vznik T-buněčných receptorů.....	31

3.2 Dosavadní metodiky detekce přestaveb Ig/TCR.....	31
3.2.1 Detekce klonality.....	31
3.2.2 Detekce minimální reziduální nemoci (MRN).....	32
3.2.3 Analýza imunitního repertoáru.....	33
3.3 Přínos sekvenování nové generace pro studium přestaveb Ig/TCR.....	33
3.3.1 Využití NGS metod ve výzkumu fyziologických přestaveb Ig/TCR.....	33
3.3.2 NGS v problematice transplantací.....	34
3.3.3 Sekvenování nové generace v souvislosti s výzkumem a léčbou autoimunitních onemocnění...	35
3.3.4 Přínos masivně paralelního sekvenování pro studium hematologických malignit.....	36
4. Závěr.....	38
Seznam použité literatury.....	39

Abstrakt

Sekvenování DNA je molekulárně genetická metoda, jejímž výstupem je údaj o pořadí a druhu nukleotidů ve vzorku deoxyribonukleové kyseliny (DNA), molekuly nesoucí genetickou informaci. Tento údaj má často klíčovou hodnotu pro mnoho odvětví; výzkum, zdravotnictví, průmysl, kriminalistiku a další. Dlouhou dobu byla k sekvenování používána téměř výhradně metoda vynalezená Frederickem Sangerem v 70. letech minulého století, která využívá upravené nukleotidy, jejichž zařazení do řetězce DNA brání jeho dalšímu prodloužení. Syntéza DNA v jejich přítomnosti dá vznik směsi různě dlouhých fragmentů, které jsou elektroforeticky rozděleny podle délky a z výsledného gelu je odečtena sekvence. Protože s sebou princip této metody nese jistá omezení (m.j. nízký výkon a malé pokrytí cílové oblasti), je v poslední době vynakládáno značné úsilí na vývoj alternativních metod sekvenování. Tyto metody, souhrnně nazývané sekvenování nové generace ("next-generation sequencing" - NGS), využívají rozličné technologie k překonání limitů Sangerova sekvenování. Tato práce pojednává o nejvýznamějších NGS metodách a zaměřuje se na možnosti jejich využití při sekvenování přestaveb genů pro imunoglobuliny a T-buněčné receptory, což je oblast se značnou klinickou relevancí v oborech jako je imunologie či hematologie.

Klíčová slova: sekvenování nové generace, přestavby genů pro imunoglobuliny a T-buněčné receptory, masivně paralelní sekvenování, imunogenetika, sekvenování DNA

Abstract

DNA sequencing is a molecular genetic method that results in data about sequence and type of nucleotides present in a given sample of deoxyribonucleic acid (DNA), a molecular carrier of genetic information. These data are frequently of a crucial value for many fields; research, medicine, industry, criminalistics or others. During a long period of time almost all the sequencing was performed using a method invented by Frederick Sanger in the 70's, a technique that uses modified nucleotides that once incorporated into a DNA strand prevent this from further elongation. DNA synthesis in presence of such nucleotides leads to a formation of a mixture of fragments of different length that are electrophoretically separated by length and the sequence is read from the resulting gel. Since the principle of this method entails some inherent drawbacks (e.g. low throughput and coverage) a significant effort is made lately to develop alternative sequencing approaches. These methods collectively referred to as next-generation sequencing (NGS) use several technologies in order to overcome the limitations of the Sanger sequencing.

This thesis discusses the most important NGS methods and focuses on their possible application for sequencing of immunoglobulin and T-cell receptor gene rearrangements, an area of undisputable clinical relevance for fields like immunology or hematology.

Key words: next-generation sequencing, immunoglobulin and T-cell receptor gene rearrangements, massive parallel sequencing, immunogenetics, DNA sequencing

1. Úvod

V posledních několika letech je vědecká komunita svědkem prudkého rozvoje molekulárně genetických technologií. Oblastí, ve které je tento překotný vývoj obzvláště patrný, je sekvenování DNA.

Technologie, umožňující stanovení přesné sekvence neznámého vzorku DNA, mají za sebou bohatou historii, sahající od prvních experimentů v sedmdesátých letech minulého století až k dnešním metodám nové generace. Právě techniky sekvenování nové generace ("Next-generation sequencing" - NGS), alternativně nazývané techniky masivně paralelního sekvenování ("Massive parallel sequencing" - MPS) znamenaly v pravém slova smyslu přerod v sekvenačních technologiích. Zatímco ještě v nedávné době existovala prakticky jediná možnost jak osekvenovat vzorek DNA - použít (postupně vylepšovanou) klasickou (Sangerovu) metodu, dnes je na trhu mnoho platforem, které se liší svými principy, vlastnostmi a možnostmi použití. Současné metody jsou přitom rychle vylepšovány a mnoho nových metod je vyvíjeno, takže se jakékoliv pokusy o přehlednou sumarizaci NGS technologií rychle stávají zastaralými.

Přehled v současnosti dostupných metod sekvenování, nastínění možného vývoje nových metod a popis uplatnění těchto metod při sekvenování přestaveb genů pro imunoglobuliny a T-buněčné receptory je tématem této bakalářské práce.

2. Sekvenování DNA

Sekvenování DNA je molekulárně biologická metoda, která slouží ke zjištění přesné sekvence nukleotidů v daném vzorku DNA. Sekvenování DNA má značný význam, a to jak v základním biologickém a biomedicínském výzkumu, tak v různých aplikovaných oborech, jako je *in vitro* diagnostika (IVD), forenzní genetika či fylogenetika.

Dlouhou dobu byla k dispozici prakticky jediná cesta, jak zjistit cílovou sekvenci - Sangerova metoda. Tato metoda byla postupem času zlepšována a automatizována, takže v moderní podobě přetrvala dodnes. V současné době je však již na trhu řada alternativních, obvykle teprve nedávno vynalezených platforem, které sekvenování DNA umožňují. Za povšimnutí stojí, že zatímco přístroje pro sekvenování Sangerovou metodou byly v průběhu let v různých podobách nabízeny mnoha různými výrobci, dnešní sekvenační technologie a přístroje na nich založené, jsou již obvykle doménou jediné firmy.

Tato část práce si klade za cíl podat přehled dnes dostupných sekvenačních metod.

2.1 Klasická (Sangerova) metoda

Jednou z vůbec prvních metod, které umožnily sekvenování DNA, je metoda vynalezená Frederickem Sangerem v roce 1977 ve Velké Británii (Sanger et al., 1977). Tato dnes již klasická metoda je založena na principu, který v mnohém připomíná přirozený způsob syntézy DNA v buňkách.

Klíčovou ideou Sangerovy metody je použití dideoxynukleotidů (ddNTP), tedy nukleotidů, obsahujících jako cukernou složku 2',3' dideoxyribosu. Zařazení takového nukleotidu do řetězce DNA vede k zastavení syntézy tohoto řetězce.

2.1.1 Princip metody

Základní reakční směs pro Sangerovu sekvenaci je tvořena zkoumaným vzorkem DNA, vyzolovaným z vhodného zdroje a namnoženým pomocí PCR, ke kterému jsou přidány komplementární primery, DNA polymeráza, směs všech čtyř volných dNTP a v nízké koncentraci také směs všech čtyř ddNTP, nesoucích fluorescenční značku (každý druh ddNTP je označen fluoroforem emitujícím jinou barvu). Za vhodných podmínek poté v reakční směsi dojde k syntetické reakci. Molekuly polymerázy postupně inkorporují dostupné dNTP do nově vznikajících vláken DNA na základě principu komplementarity bází, avšak v momentě kdy do řetězce místo dNTP zařadí ddNTP reakce končí (dideoxynukleotidy neobsahují 3' -OH skupinu, potřebnou pro tvorbu fosfodiesterové vazby s následujícím nukleotidem). Metoda pracuje s myšlenkou, že při obrovském množství molekul DNA v reakci, dojde s určitou

pravděpodobností k zastavení syntézy na každém místě řetězce. Zjednodušeně řečeno, do některého řetězce se dideoxynukleotid zařadí už jako první nukleotid, do dalšího jako druhý a do jiného třeba až jako poslední.

Vzniklé dvojitěvláknové molekuly jsou opět denaturovány působením zvýšené teploty a původní (templátová) vlákna se ze směsi odstraní. Výsledkem syntetické reakce je tudíž směs molekul DNA o různé délce, kdy každá na svém 3' konci nese ddNTP s fluorescenční značkou (lišící se emitovanou barvou podle druhu nukleotidu).

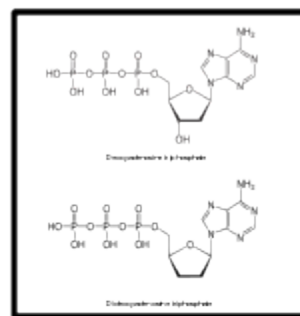
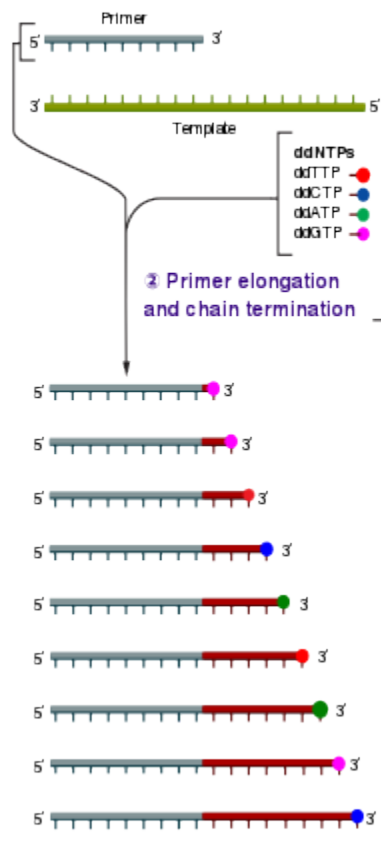
Dalším krokem je separace molekul podle délky. Zatímco původní metoda k tomuto účelu využívala horizontální elektroforézy na velkoplošných agarózových gelech, dnes je dáována přednost variantě kapilárové. Směs je nasáta do speciální kapiláry, ve které proběhne kapilární elektroforéza - obdoba klasické gelové elektroforézy. Protože DNA nese záporný náboj (způsobený zbytky kyseliny fosforečné v cukr - fosfátové kostře), putuje v elektrickém poli k anodě a po určitém čase se tak v kapiláře působením elektrického pole ustaví délkový gradient molekul DNA. Kratší úseky budou umístěny blíže k anodě (gelem procházely rychleji), zatímco s rostoucí délkou budou molekuly od anody vzdálenější. Gelové prostředí v kapiláře je přitom navrženo tak, aby bylo ve vzniklém gradientu možné rozpoznat jednonukleotidové rozdíly v délce molekul (je tedy dostatečně zřetelné, že skupina molekul o délce např. 32 nukleotidů doputovala blíže k anodě než molekuly o délce 33 nukleotidů atp.).

Gel je následně po jednotlivých vrstvách ozařován laserem. Protože molekuly DNA na svých 3' koncích nesou ddNTP s fluorescenční značkou, emitují po ozáření světlo určité vlnové délky (lišící se podle druhu nukleotidu na 3' konci), toto světlo je snímáno detektorem a vzniklý signál počítačově zpracován. Z barvy, kterou daná skupina molekul vyzáří, je zjištěno, která báze tvoří 3' konec molekul dané délky. Jestliže je tedy gel čten např. od anody ke katodě (od nejkratších molekul po nejdelší), je možné přímo odečítat sekvenci DNA.

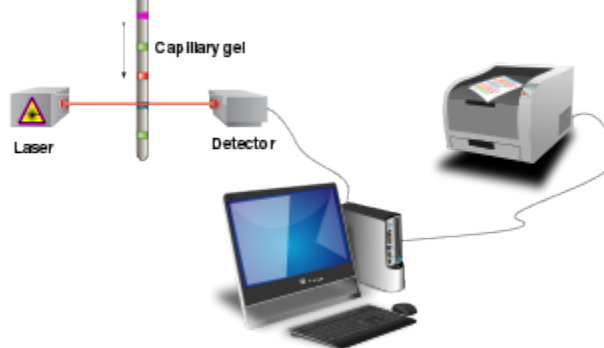
Jednotlivé kroky Sangerovy metody jsou schematicky znázorněny na obr. 1.

1 Reaction mixture

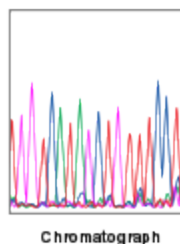
- ▶ Primer and DNA template ▶ DNA polymerase
- ▶ ddNTPs with flouochromes ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



2 Capillary gel electrophoresis separation of DNA fragments



3 Laser detection of flouochromes and computational sequence analysis



obr. 1 - Schéma sekvenace DNA Sangerovou metodou

Zdroj : "Sanger-sequencing" by Estevezj - Own work. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Sanger-sequencing.svg#mediaviewer/File:Sanger-sequencing.svg>. 16.11.2014

2.1.2 Technická realizace

Přístroje pro sekvenaci DNA Sangerovou metodou jsou dnes nabízeny firmou Life Technologies jako tzv. zlatý standard. Firma nabízí přístroje mnoha velikostí a zákazník má tudíž možnost si vybrat vhodný přístroj s ohledem na požadovaný výkon a zamýšlené využití. Přístroje se liší počtem sekvenačních kapilár, množstvím vzorků, které je možno zpracovávat naráz, mírou automatizace, cenou za osekvenovanou megabázi a samozřejmě také rozměry či spotřebou energie. Všechny tyto parametry je třeba před nákupem sekvenátoru zvážit.

Platformy, nabízené firmou Life Technologies, jsou přehledně shrnuty na obr. 2.

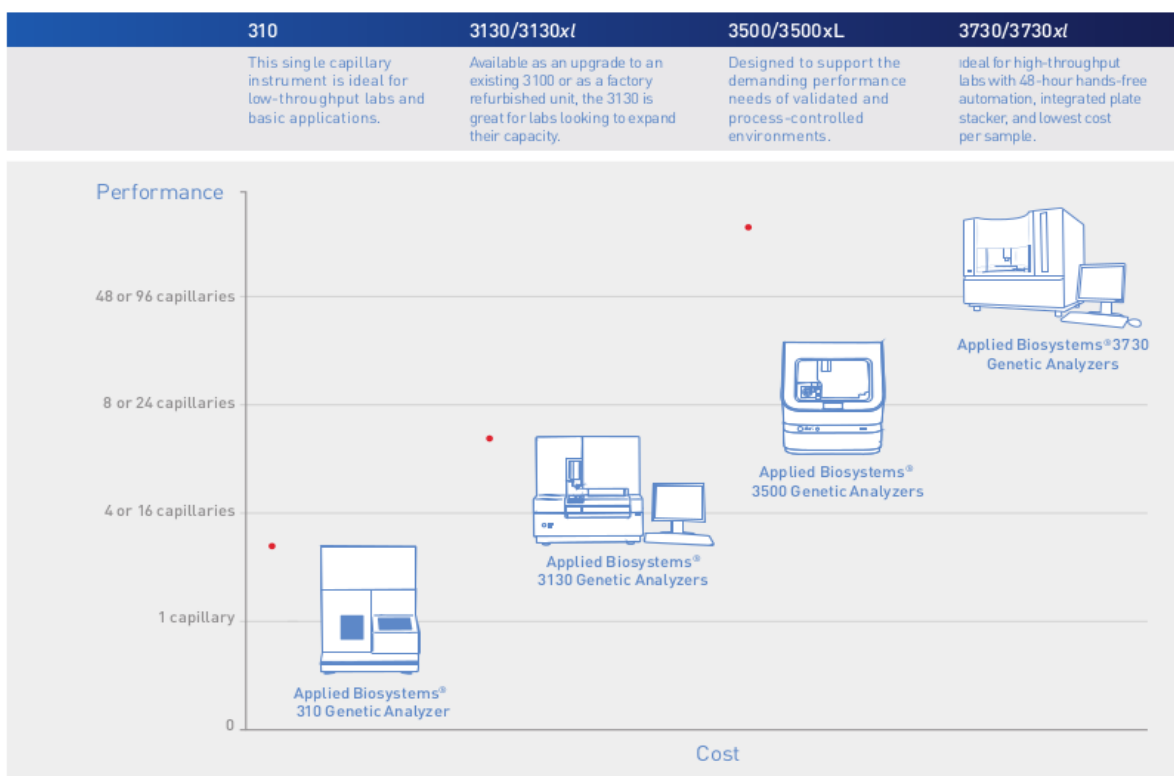


Figure 1. Comparison of performance, throughput and cost of Applied Biosystems® Genetic Analyzers.

Table 1. Features and specifications of Applied Biosystems® Genetic Analyzers

	310	3130/3130xl	3500/3500xL	3730/3730xl
Number of capillaries	1	4 [3130], 16 [3130xl]	8 [3500], 24 [3500xL]	48 [3730], 96 [3730xl]
Number of dyes	5	5	6	5
Capillary array length (cm)	47, 61	22, 36, 50, 80	36*, 50	36, 50
RFID	No	No	Yes	No
Polymer type	POP™ CAP, POP-4™, POP-6™	POP™ CAP, POP-4™, POP-6™, POP-7™	POP-4™, POP-6™, POP-7™	POP™ CAP, POP-6™, POP-7™
Sample capacity	1–96 sample tubes	2 sample plates [96- or 384-well]	2 sample plates [96- or 384-well]	16 sample plates [96- or 384-well]
Hands-free automation	24 hours	24 hours	24 hours	48 hours
Integrated plate stacker	No	No	No	Yes
Availability**	New or refurbished	Upgrade or refurbished	New	New or refurbished

*36 cm capillary array is only used for HID applications on POP-4™.

**Availability of refurbished instruments varies throughout the year, please contact your sales rep for details.

obr. 2 - přehled přístrojů pro sekvenování DNA Sangerovou metodou

Zdroj : <http://tools.lifetechnologies.com/content/sfs/brochures/brochure-ab-genetic-analyzers.pdf>. 16.11.2014

2.1.3 Využití metody

I přes současný rozmach metod NGS je Sangerova metoda stále základem genetického vyšetřování, především díky své rychlosti a cenové nenáročnosti, ale také zásluhou dlouhé délky čteného úseku (tzv. "read") - okolo 800 bp. Uplatňuje se v diagnostice známých mutací, *de novo* sekvenování i dalších aplikacích. Sangerovy metody je také možné využít pro validaci výsledků NGS metod (Judkins

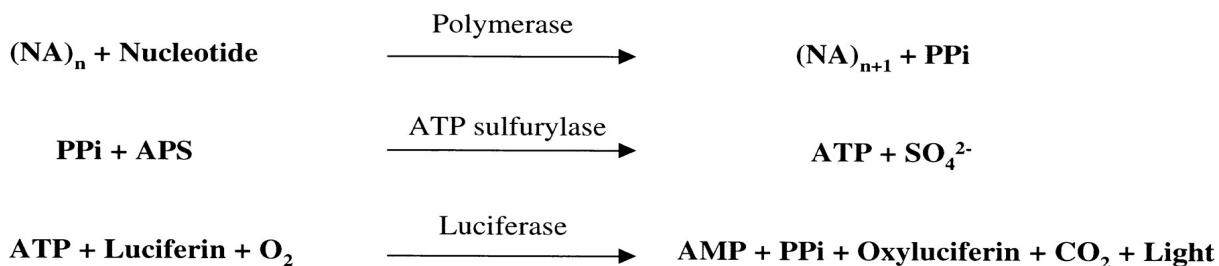
et al., 2015).

2.2 Sekvenování principem Roche 454

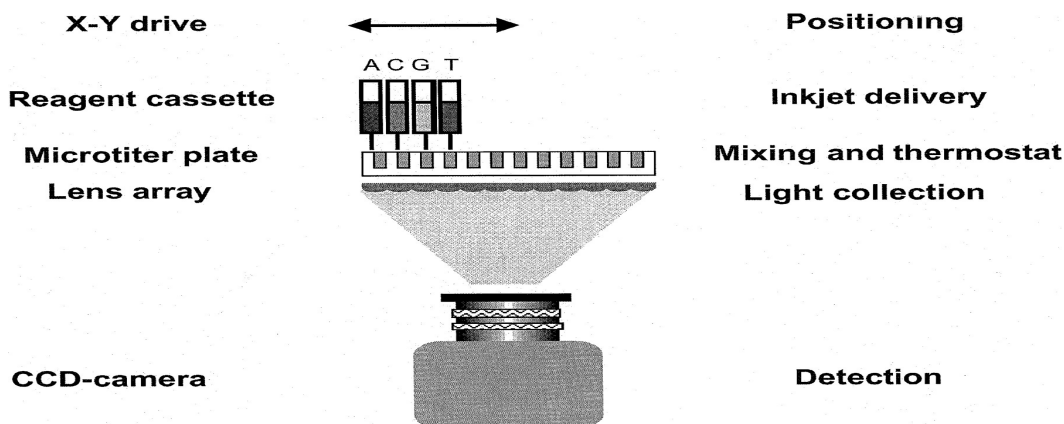
Firma 454, která se nedlouho po svém založení stala dceřinou firmou diagnostické divize společnosti Roche, uvedla v roce 2005 jako první na trh platformu umožňující sekvenování nové generace. Přístroje od 454 jsou založeny na principu pyrosekvenování, který umožňuje zjistit hledanou sekvenci DNA pomocí luciferázové enzymatické kaskády (Ronaghi, 2001).

2.2.1 Princip metody

Metoda, využívaná přístroji firmy 454, je založena na detekci pyrofosfátu (PP_i), který vzniká jako vedlejší produkt při začlenění nukleotidu do řetězce DNA. Uvolněný pyrofosfát je enzymem ATP sulfurylázou přeměněn na ATP, které je spolu s luciferinem substrátem pro enzym luciferázu. Luciferáza oxiduje luciferin na excitovaný oxyluciferin, který při přechodu do základního stavu uvolňuje fotony viditelného světla. Ty jsou systémem čoček směřovány na citlivý CCD detektor, který optický signál převede na elektrický a zesílí jej. Reakce, tvořící luciferázovou enzymatickou kaskádu, jsou shrnuty na obr. 3, schéma jednoduchého zařízení umožňujícího pyrosekvenaci ukazuje obr. 4.

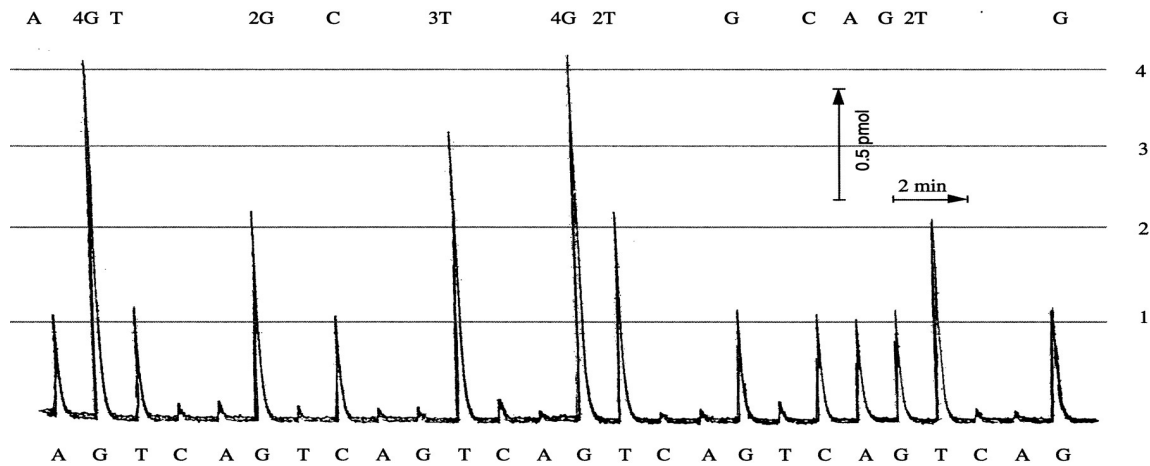


obr. 3 - Luciferázová enzymatická kaskáda. APS = adenosinfosfosulfát
Zdroj : (Ronaghi, 2001)



obr. 4 - Schéma pyrosekvenačního zařízení
Zdroj : (Ronaghi, 2001)

Výstupem z pyrosekvenátoru je graf intenzity detekovaného světla v závislosti na počtu přidávaných nukleotidů (tzv. pyrogram - obr. 5), z něhož je možné odečíst sekvenci bází ve zkoumaném řetězci.



obr. 5 - Pyrogram
Zdroj : (Ronaghi, 2001)

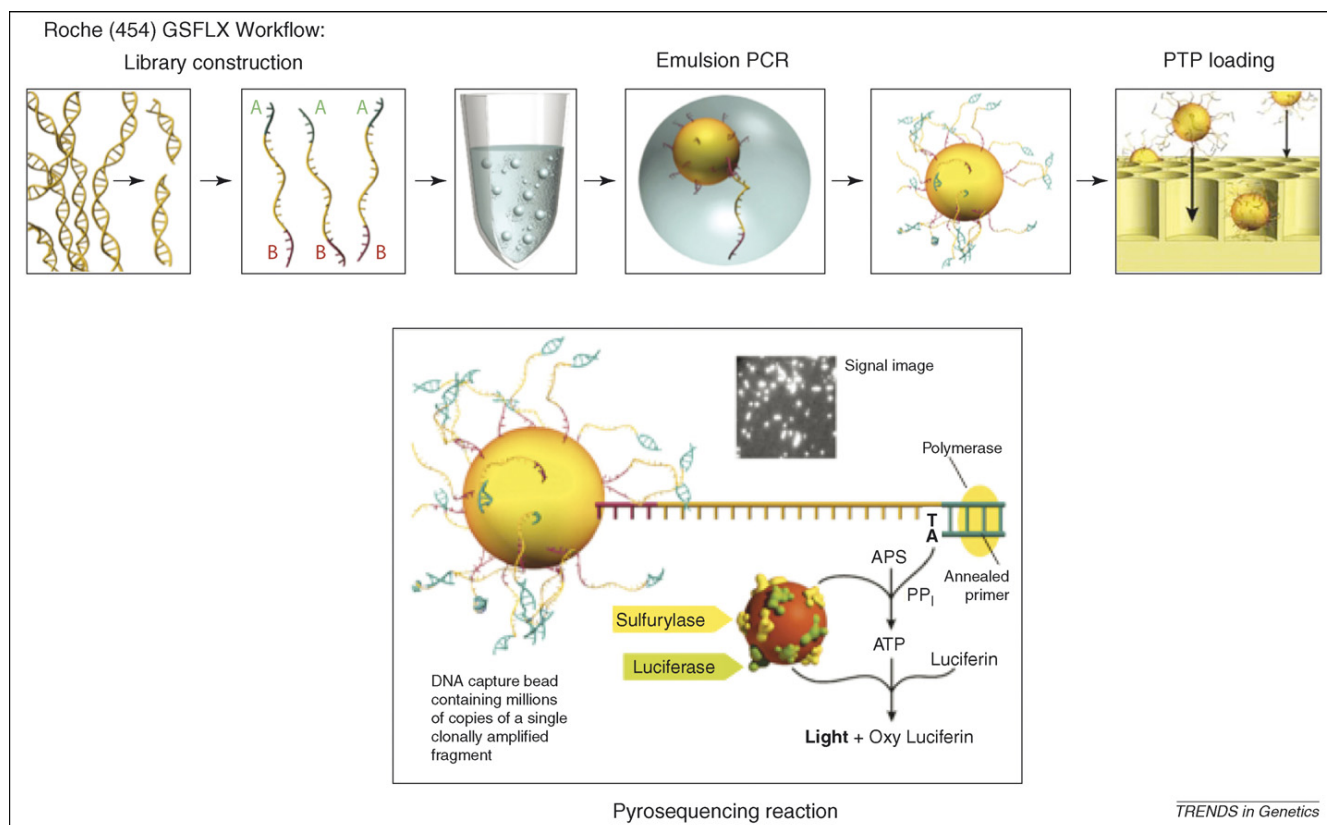
Jak je z výše uvedeného patrné, celý proces pyrosekvenace se obejde bez fluorescenčně značených primerů či nukleotidů, odpadá rovněž nutnost provádět elektroforetickou separaci vzniklých molekul. To s sebou nese zjednodušení a zlevnění celé metody. Kvůli chemiluminiscenční povaze vznikajícího světla ale není možné zvýšit intenzitu signálu např. použitím výkonnějších laserů, jako je tomu u metod pracujících s fluorescencí (Ronaghi, 2011). S intenzitou vznikajícího světla také souvisí asi nejpalčivější problém metod založených na pyrosekvenování, který velmi komplikuje hlavně sekvenování delších homopolymerních úseků (sekvencí DNA, tvořených pouze jedním druhem nukleotidu). Množství uvolněného záření při zařazení více stejných nukleotidů totiž narůstá lineárně jen do 5. - 6. nukleotidu homopolymeru a v úsecích delších než je tato hranice může mít pyrosekvenační metoda problém určit, kolik přesně nukleotidů se vlastně do řetězce inkorporovalo. Problémy se sekvenováním homopolymerních úseků jsou ovšem typické i pro jiné NGS metody a proto se intenzivně pracuje na jejich softwarové kompenzaci.

2.2.2 Technická realizace

Proces sekvenace na přístrojích firmy 454 sestává z několika kroků.

Na začátku je zkoumaný vzorek DNA rozdělen na krátké fragmenty (o délce v rozmezí 50 - 900 bp), čehož se dosáhne nebulizací - "rozprášením" vzorku aplikací stlačeného plynu. Na fragmenty jsou posléze navázány dva typy krátkých (44 bp) molekul DNA, které se nazývají sekvenační adaptory. Adaptory jsou dvou typů, A a B, které se kromě sekvence od sebe liší také tím, že adaptor B je na 5'

konci biotinylován. Po navázání adaptorů jsou dvojitě vláknové molekuly alkalicky denaturovány a jednovláknové fragmenty DNA s navázanými adaptory tvoří tzv. sekvenační knihovnu. Za pomoci adaptorů je poté každý fragment přichycen na mikroskopickou, streptavidinem obalenou magnetickou kuličku, jejíž povrch je pokryt velkým množstvím krátkých úseků DNA komplementárních k těmto adaptorům. V další fázi je fragment, přichycený na kuličce, amplifikován pomocí tzv. emulzní PCR (emPCR). Jde o amplifikační metodu, při které jsou ke kuličkám s navázanými fragmenty přidány reagenty, potřebné pro PCR a specifická směs olejů. Speciálně navrženým procesem třepání je následně docíleno vytvoření miniaturních bublinek okolo každé kuličky (vznikají tzv. mikroreaktory). V každé bublince poté proběhne potřebný počet reakčních cyklů a kuličky se postupně obalí kopiemi původního fragmentu. Když je DNA amplifikována, jsou mikroreaktory rozbity působením isopropylalkoholu a kuličky umístěny na destičku s drobnými jamkami (tzv. PicoTiterPlate), pod kterou je umístěn detekční systém, tvořený soustavou čoček a citlivou CCD kamerou. Díky velikosti kuliček a průměru jamek je zaručeno umístění právě jedné kuličky do každé jamky. Ke kuličkám s DNA templáty se ještě přidávají další kuličky, nesoucí na svém povrchu enzymy, potřebné pro pyrosekvenační reakci - ATP sulfurylázu a luciferázu. V destičce probíhá samotná sekvenační reakce, která se skládá z cyklicky opakovaných dějů - zalití destičky roztokem s jedním druhem nukleotidu, detekce případně vzniklých signálů a odmytí nezařazených nukleotidů (Margulies et al., 2005). Průběh sekvenace na přístrojích Roche 454 vyobrazen na obr. 6.



obr. 6 - schéma sekvenace přístroji Roche 454
Zdroj : (Mardis, 2007)

Sekvenátory, založené na popsaném mechanismu, nabízí firma Roche 454 ve dvou provedeních. Větší přístroj, GS FLX+ dokáže zpracovávat až 1 000 000 reads na jeden běh a je tedy určen spíše pro větší laboratoře, pracující na rozsáhlejších projektech. "Benchtop" varianta, GS Junior je naopak určena převážně pro použití v menších laboratořích a na klinických pracovištích.

2.2.3 Využití metody

Díky schopnosti tvorby poměrně dlouhých reads jsou přístroje Roche 454 použitelné pro *de novo* sekvenování (včetně sekvenování celých genomů - např. bakteriálních), uplatnění nalézají rovněž při transkriptomovém sekvenování či v metagenomických aplikacích. GS FLX+ i GS Junior rovněž umožňují tzv. amplikonové sekvenování, tedy sekvenování předem připraveného PCR produktu bez nutnosti amplifikace. Výhodou této platformy je také její rychlost - celá sekvenační reakce zabere pouhých 10 hodin. Nevýhodou Roche 454 zůstává cena reagentů, dlouhý čas přípravy běhu a zvýšená chybovost v homopolymerních úsecích delších než 6 bází (Glenn, 2011, Liu L et al., 2012).

Ačkoliv byla použitelnost sekvenačních řešení firmy Roche 454 prokázána mnoha publikacemi, přístroje se nakonec neukázaly jako dostatečně konkurenceschopné na trhu NGS technologií a firma

oznámila ukončení podpory Roche 454 platformy ke konci roku 2016.

2.3 Applied Biosystems SOLiD

Přístroj od firmy Applied Biosystems (ABI), součásti konsorcia Life Technologies, byl představen v roce 2007 a je založen na sekvenování ligací.

2.3.1 Princip metody

Sekvenování na platformě SOLiD (Sequencing by Oligonucleotide Ligation and Detection) je v úvodních krocích podobné pyrosekvenování, samotná sekvenace se však výrazně liší. Namísto sekvenování syntézou, které nalézáme u přístrojů Roche 454 (ale i u dalších NGS technologií), je tento systém založen na sekvenování ligací krátkých nukleotidových oligomerů.

V prvním kroku dochází k tvorbě sekvenační knihovny. Sekvenovaný vzorek DNA je fragmentován na krátké úseky a k fragmentům jsou z obou stran přidány odlišné sekvenační adaptory (P1, P2).

Následuje amplifikační reakce, která se opět uskutečňuje pomocí emulzní PCR. Fragменты knihovny jsou smíchány s 1 µm velkými, magnetickými kuličkami, jejichž povrch je pokryt krátkými oligonukleotidy, komplementárními k jednomu typu sekvenačních adaptorů (např. k P1). Do směsi jsou přidány reagenty, potřebné pro PCR, všechny komponenty se po přilítí speciálního oleje emulsifikují třepáním a za potřebných podmínek proběhne amplifikace fragmentů, analogicky Roche 454 metodě. Další krok se nazývá "bead-enrichment" a spočívá v selekci kuliček, na kterých skutečně došlo k amplifikaci. Selekcce se dosáhne použitím jiného druhu kuliček, které na svém povrchu nesou oligonukleotidy, komplementární k adaptorům P2. K těmto kuličkám se tedy navážou pouze kuličky s amplifikovanou DNA (protože pouze ty nesou P2 adaptor). Vzniklé agregáty se poté separují od zbytku směsi a následně jsou kuličky s fragmenty odděleny od nosných kuliček. Tímto způsobem se výrazně zvyšuje výtěžek amplifikační reakce.

Na "bead-enrichment" navazuje fáze, která spočívá v kovalentní vazbě kuliček s fragmenty na skleněnou destičku. Aby toto bylo možné, je před navázáním upraven 3' konec molekul DNA na kuličkách. Kuličky jsou na destičce náhodně rozmístěné, přičemž každá destička může být tvořena jedním, čtyřmi nebo osmi políčky pro vazbu kuliček. Destička je následně umístěna do sekvenátoru. Do reakční směsi je poté přidána DNA ligasa a směs speciálních, fluorescenčně označených oktamerů. Oktamery jsou tvořeny třemi degenerovanými nukleotidy na 3' konci, dvěma specifickými nukleotidy na pozicích 4 a 5 a znovu třemi univerzálními nukleotidy na 5' konci. 5' konec oktamerů také nese jeden ze čtyř druhů fluorescenčních značek. Platforma SOLiD využívá unikátní systém barevného

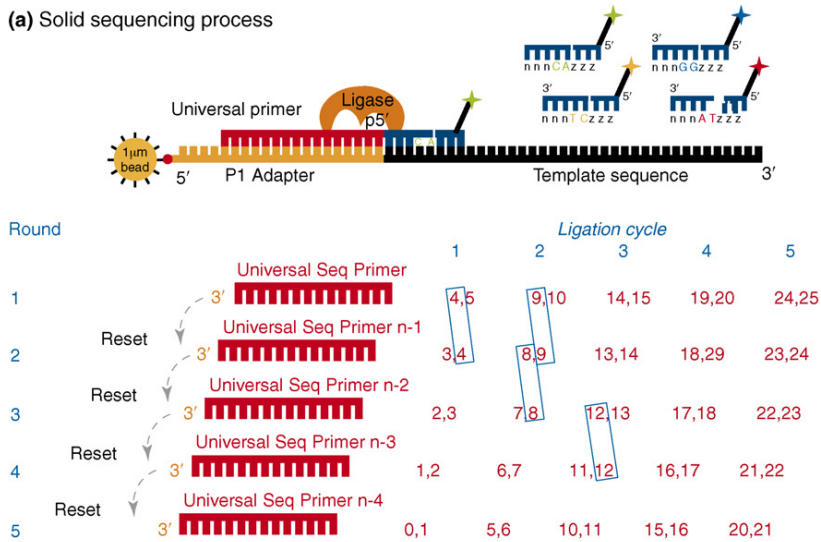
kódování, který se nazývá "two base encoding", neboť stejným fluoroforem je vždy označena čtveřice párů bazí (např. modrá barva kóduje dvojice AA, TT, CC a GG).

Samotná sekvenace sestává z pěti cyklů a každý cyklus je tvořen sedmi ligačními kroky. Sekvenační reakce začíná vazbou univerzálních sekvenačních primerů, komplementárních k adaptorům, ohraničujícím sekvenované fragmenty DNA. V prvním cyklu je k univerzálnímu primeru na základě principu komplementarity hybridizován vhodný oktamer, který posléze DNA ligasa připojí k sekvenačnímu primeru. Po ligaci následuje odečtení fluorescenční značky a ta je poté spolu s posledními třemi nukleotidy na 5' konci odštěpena. Na volný 5' konec hybridizuje další oktamer a celý krok se opakuje. Když proběhne všech sedm ligačních kroků, je zahájen další cyklus sekvenace, který začíná tzv. resetováním vlákna, při němž je celé vlákno denaturováno a doposud hybridizované úseky odmyty. Na uvolněný templát je znovu hybridizován sekvenační primer, tentokrát však o jednu pozici blíže k 5' konci sekvenovaného řetězce. Celý cyklus hybridizací, ligací a detekce fluorescence se znovu opakuje (McKernan et al., 2009).

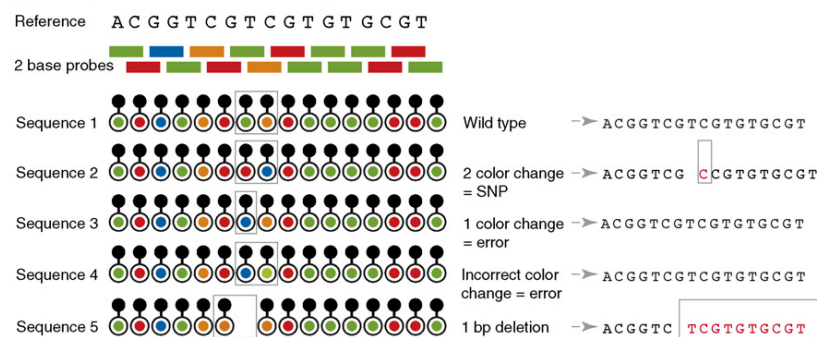
Sekvenace končí po pěti cyklech a z principu metody je tedy patrné, že délka readu je u platformy SOLiD 35 bází, ačkoliv tuto základní délku je možné prodloužit.

Kroky, zahrnuté v sekvenaci systémem SOLiD spolu s barevným kódováním shrnuje obr. 7.

(a) Solid sequencing process



(b) Principles of two base encoding



TRENDS in Genetics

obr. 7 - Sekvenace ligací přístroji SOLiD a principy barevného kódování využívaného těmito přístroji.

Zdroj : (Mardis, 2007)

2.3.2 Technická realizace

Firma Applied Biosystems dnes nabízí přístroj 5500 W Series Genetic Analyzer, který oproti svým předchůdcům poskytuje výrazně rychlejší a levnější sekvenaci. Využívá totiž tzv. Wildfire technologii, díky níž přístroj pro svůj běh nepotřebuje žádné kuličky a amplifikace probíhá přímo na sekvenační destičce procesem zvaným template walking. Tato technologie také zahrnuje zvýšenou hustotu vzorků na destičce (> 500 000), délka readu přitom dosahuje až 75 bází. Větší varianta přístroje, 5500 xl W, rovněž umožňuje zpracovávat dvě destičky naráz, což přináší další úsporu času a prostředků. V současné době je přístroj schopen za jeden běh nashromáždit až 160 Gb (320 Gb u xl varianty), osekvenovat až 18 exomů (36 u xl varianty) či 12 transkriptomů (24 u xl varianty).

Platforma SOLiD nemá žádnou "benchtop" variantu ani menší verzi standardního přístroje.

2.3.3 Využití metody

Hlavní předností systému SOLiD je jeho vysoká přesnost, které je dosaženo specifickým způsobem kódování bazí ("two base encoding"). Díky němu je každá báze čtena dvakrát a po přiřazení k referenční sekvenci je tak snadno možné odlišit pravý SNP (záměna obou barev oproti referenční sekvenci) od chyby v sekvenování. Platforma SOLiD tak nalézá uplatnění při celogenomovém resekvenování a sekvenování transkriptomu či malých RNA (Liu L et al., 2012).

2.4 Ion Torrent by Life Technologies

Kromě přístrojů pro sekvenování Sangerovou metodou a platformy SOLiD, nabízí společnost Life Technologies ještě třetí typ sekvenátorů, založený na polovodičové technologii, původně vyvíjené firmou Ion Torrent, která přešla pod Life Technologies v roce 2010.

2.4.1 Princip metody

Sekvenování na platformě Ion Torrent je založeno na doprovodném jevu přirozené syntézy DNA, kterým je uvolnění protonu (H^+) při tvorbě fosfodiesterové vazby během připojování nového nukleotidu do řetězce DNA. Jde o první metodu sekvenace DNA, která se zcela obejde bez optických komponent či chemicky modifikovaných nukleotidů nebo jiných speciálních reagensů. Pokud je do vznikajícího řetězce polymerázou začleněn další nukleotid, je toto doprovázeno odštěpením vodíkového iontu, což vyvolá změnu (snížení) pH okolního roztoku. Změna pH, která je úměrná počtu zařazených nukleotidů, může být registrována a převedena na změnu napětí. Pokud současně víme, který ze čtyř nukleotidů byl přítomen v roztoku v době inkorporace, můžeme ze změn pH (a tedy i napětí) odvodit sekvenci zkoumané DNA.

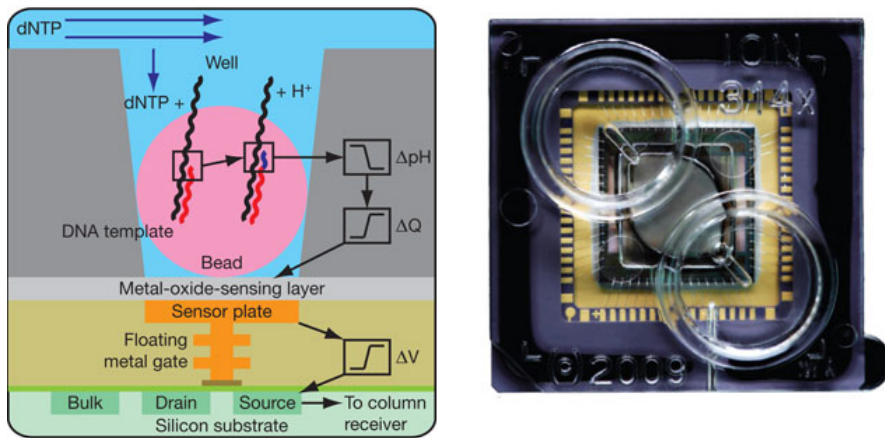
2.4.2 Technická realizace

Průběh sekvenace u přístrojů Ion Torrent sestává z následujících kroků.

Prvním krokem je tvorba sekvenační knihovny. Vzorek DNA je fragmentován na molekuly o délce 100-160 bazí a na každý fragment jsou ligovány forward a reverse adaptory. Proces pokračuje amplifikací knihovny prostřednictvím emPCR. Molekuly, tvořící knihovnu, jsou smíchány s 2 μ m akrylamidovými kuličkami, jejichž povrch je pokryt krátkými oligonukleotidy o sekvenci komplementární k jednomu ze sekvenačních adaptorů. DNA knihovna je amplifikována v olejové emulzi, podobně jako u jiných NGS systémů a navazuje na ni "bead-enrichment", využívající magnetické kuličky. Amplifikovaný templát je

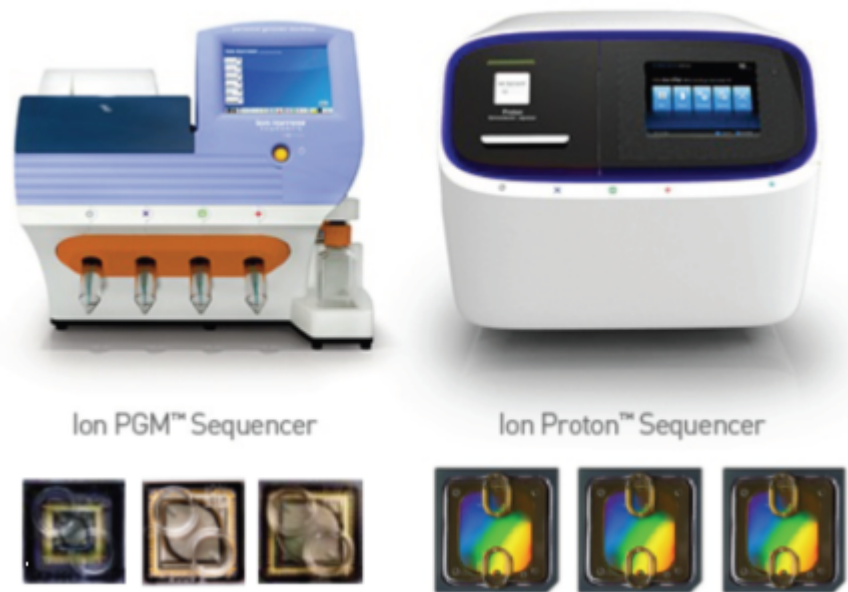
poté doplněn o sekvenační primery spolu s DNA polymerázou a sekvenován.

Sekvenační reakce probíhá v mikroskopických komůrkách na povrchu speciálně navrženého polovodičového čipu. Každá komůrka, do které se vejde právě jedna kulička, má na svém dně vrstvu citlivou na změnu pH. Tato vrstva, obsahující oxidy tantalu, odpovídá na změnu pH roztoku změnou vlastního elektrického potenciálu. Pod pH citlivou vrstvou se nachází vrstva polovodičových tranzistorů (ISFET - ion-sensitive field-effect transistor) která změnu potenciálu zesílí a převede k dalšímu zpracování na elektronické součásti ležící mimo čip. Čip je v cyklech naplňován vždy jedním druhem nukleotidů, případná změna pH je odečtena a nezařazené nukleotidy odmyty pufrům, který zároveň ustálí pH na výchozí hodnotě (Rothberg et al., 2011). Schéma sekvenování přístroji Ion Torrent je na obr. 8 vlevo, vpravo je fotografie sekvenačního čipu.



obr. 8 - Princip polovodičového sekvenování (vlevo) a sekvenační čip (vpravo).
Zdroj : (Rothberg et al., 2011)

Přístroje, založené na popsané metodě, nabízí firma Life Technologies ve dvou variantách. Menší přístroj se nazývá Ion PGM (Personal Genome Machine) a je opět určen spíše pro menší laboratoře či klinická pracoviště. Velmi vhodný je například pro identifikaci mikrobiálních patogenů nebo k ampikonovému sekvenování panelů vybraných genů, např. v diagnostice známých mutací u nádorů nebo dědičných onemocnění (Beadling et al., 2013). Větší přístroj, Ion Proton je v současné konfiguraci schopen sekvenovat dva lidské exomy najednou a nalezne proto uplatnění na pracovištích provádějících komplexní genetickou diagnostiku. Oba přístroje mají odlišný design a používají jiné sekvenační čipy (viz. obr. 9), ale pracují na stejném principu a používají stejný způsob zpracování dat.



obr. 9 - Přístroje Ion Torrent s příslušnými sekvenačními čipy
 Zdroj : http://www.genomics.cn/uploadfile/cms/image/20130717/ion%20torrent%20_932.png, 16.11.2014

2.4.3 Využití metody

Sekvenátory Ion Torrent podporují aplikace jako *de novo* sekvenování, transkriptomové sekvenování či cílené resekvenování, možnosti použití těchto přístrojů jsou však ještě širší. Díky nízké pořizovací i provozní ceně (přístroje jsou založeny na běžné polovodičové technologii a nevyžadují pro svůj běh žádné speciální reagenty) je technologie Ion Torrent dostupná pro široké spektrum pracovišť. Výkon přístrojů (a tedy cenu za jeden běh) je navíc možné přizpůsobit potřebě konkrétního pokusu - volbou vhodného sekvenačního čipu (Loman et al., 2012).

Asi největší nevýhodou sekvenátorů, založených na polovodičové technologii, zůstává zvýšená chybovost v homopolymerních úsecích (typická i pro přístroje 454). Přístroje Ion Torrent rovněž nejsou příliš vhodné pro sekvenování velmi A-T bohatých genomů (Quail et al., 2012).

2.5 Illumina

Kalifornská společnost Illumina, dříve známá především pro svoje produkty z oblasti technologií založených na microarrays, jako je genotypizace SNP nebo studie genové exprese, vstoupila do světa sekvenování DNA v roce 2007 zakoupením firmy Solexa. Illumina převzala princip sekvenování syntézou (SBS - Sequencing By Synthesis) vyvíjený Solexou a implementovala jej do svých přístrojů.

2.5.1 Princip metody

Sekvenování syntézou přístroji Illumina je založeno na modifikované syntéze DNA, která je prováděna cyklicky po jednom nukleotidu. V cyklu jsou k syntetizovanému vlákně přidány reagenty - DNA polymeráza a speciálně upravené nukleotidy, které mají reverzibilně zablokovaný 3' konec, takže neumožňují polymeraci. Do vznikajícího vlákna se tudíž zařadí právě jeden nukleotid (na základě principu komplementarity s templátovým vláknem). Modifikované nukleotidy navíc na svém 3' konci nesou fluorescenční značku, charakteristickou pro každou bázi. V momentě, kdy je do nascentního řetězce inkorporován komplementární nukleotid, je vzorek ozářen laserem a vzniklá fluorescence odečtena. Následně se ze 3' konce řetězce, spolu s fluorescenční značkou, odstraní blokující skupina, jsou přidány další reagenty a probíhá nový cyklus.

Díky tomu, že jsou do reakce přidávány všechny čtyři druhy nukleotidů najednou, jsou oproti jiným metodám mnohem vzácnější případy, kdy je do řetězce zařazen nesprávný nukleotid (tzv. misincorporation). V reakci totiž dochází k přirozené kompetici nukleotidů o vazebné místo polymerázy a je tudíž vyšší šance, že bude polymerázou zařazen právě komplementární nukleotid. Další výhodou s sebou nese syntéza vlákna po jednom nukleotidu, díky níž by (teoreticky) neměly nastat žádné problémy v homopolymerních úsecích, charakteristické pro jiné metody využívající sekvenování syntézou.

2.5.2 Technická realizace

Sekvenovaný vzorek DNA je nejprve transposasou fragmentován na krátké úseky, opatřené adaptorovými oligonukleotidy, odlišnými pro 3' a 5' konce. Adaptory jsou dále upraveny a výsledně se skládají ze tří úseků - směrem od konce molekuly k sekvenovanému úseku DNA jde o úsek komplementární k oligonukleotidu na povrchu reakční destičky (tzv. "flow cell"), Index pro multiplexování vzorků (typ 1 nebo 2) a jeden ze dvou typů míst pro uchycení sekvenačního primeru. Dvojvláknové úseky s ligovanými adaptory jsou posléze denaturovány a jednotlivá vlákna uchycena na povrch reakční destičky na základě komplementarity k jednomu ze dvou druhů oligonukleotidů, kterými je povrch destičky pokryt. Na destičce poté dochází k mnohonásobnému namnožení vláken (aby mohla později sloužit jako templáty pro sekvenační reakci) v procesu nazvaném můstková amplifikace ("bridge amplification"). Ke každému vlákně je nejprve dosyntetizován komplementární řetězec, přičemž jako primer pro syntézu je využíván oligonukleotid, ukotvený na povrchu reakční destičky, který je komplementární k adaptoru. Takto vzniklá dvojvláknová molekula je denaturována a

volný konec uchyceného vlákna spojen s komplementárním oligonukleotidem na povrchu destičky (s opačným druhem oligonukleotidu než v prvním kroku). Vzniká tak můstek, který umožňuje syntézu komplementárního vlákna. Tento krok se v procesu můstkové amplifikace mnohokrát zopakuje a na povrchu destičky tak dojde k vytvoření tzv. "clusters" - shluků identických molekul. Protože jsou ale clusters tvořeny jak vlákny s původní sekvencí ("forward strands") tak vlákny se sekvencí k forward strands komplementární ("reverse strands"), je na úplný konec amplifikace zařazen krok spočívající v odmytí reverse strands. Volné konce forward strands jsou blokovány proti nechtěné hybridizaci s oligonukleotidy na povrchu destičky.

Když je dokončena amplifikace, přichází na řadu samotné sekvenování. Do sekvenační reakce jsou přidány primery, komplementární k primer-vazebné části 3' koncového adaptoru (tedy k adaptoru na volném konci vláken). Polymeráza poté dosyntetizuje komplementární řetězec v procesu sekvenace syntézou a během syntézy je odečtena sekvence prvního readu. Nasyntetizovaný produkt je odmyt a k vláknu je hybridizován Index 1 primer, který se váže na druhý typ primer-vazebného místa. Index 1 je pak stejným způsobem osekvenován a vzniklý produkt rovněž odmyt. 3' konec vlákna je následně odblokován a hybridizován k druhému typu oligonukleotidu ukotvenému na povrchu destičky, který je k němu komplementární (opět tedy vznikne můstek). Jelikož na místo, komplementární k povrchovému oligonukleotidu, bezprostředně navazuje Index 2, je tento osekvenován a sekvenační produkt znovu odmyt. Následně je polymerázou dosyntetizováno celé druhé vlákno můstku. Dvojvláknová molekula je poté denaturována a původní vlákno odmyto, takže v reakci zůstává pouze reverse strand. K úseku, který je komplementární k sekvenačnímu primeru (tentokrát druhého typu), se naváže příslušný primer a stejným způsobem jako v případě prvního readu je přečten i druhý read (Syed et al., 2009, Bentley et al., 2008). Přípravnou fázi sekvenace přístroji Illumina zobrazuje obr. 10.

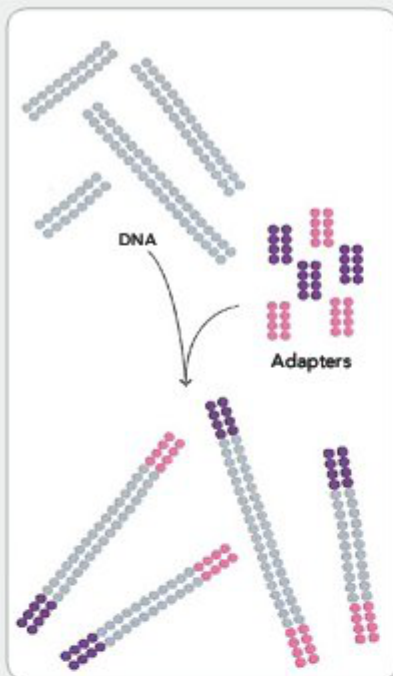
Po sekvenaci následuje analýza dat, která umožňuje provést identifikaci přečtených reads na základě sekvenačních indexů a ze vzniklých dat je vytvořena cílová sekvence, kterou je možné následně porovnat s referenční sekvencí za účelem nalezení sekvenčních variant ve zkoumaném vzorku.

Spektrum přístrojů, nabízené společností Illumina, zahrnuje platformy s různým výkonem. Nejmenším přístrojem je MiSeq, malá "benchtop" platforma vhodná pro cílené sekvenování a sekvenování malých genomů. Illumina rovněž nabízí variantu MiSeqDx, NGS sekvenátor určený pro klinickou diagnostiku, který jako první získal certifikát FDA. Prostředním přístrojem z hlediska velikosti i výkonu je platforma NextSeq, nabízející využití v exomovém sekvenování a dalších aplikacích. Největším z nabízených přístrojů je HiSeq, určený svým výkonem a velikostí do laboratoří, zaměřených na

sekvenování ve velkém rozsahu. Fotografie všech tří platforem jsou na obr. 11.

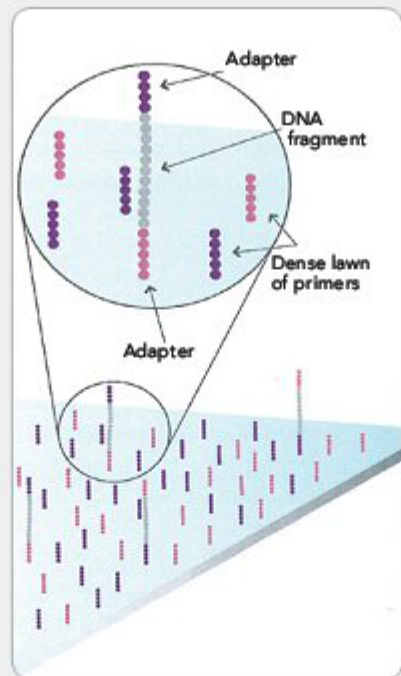
Kromě těchto tří běžných přístrojů nabízí Illumina také platformu HiSeq X Ten, což je soubor deseti přístrojů HiSeq. Toto řešení umožňuje provádění těch největších sekvenačních projektů a jako první na trhu NGS technologií nabízí osekvenování lidského genomu za cenu menší než 1000 dolarů.

1. PREPARE GENOMIC DNA SAMPLE



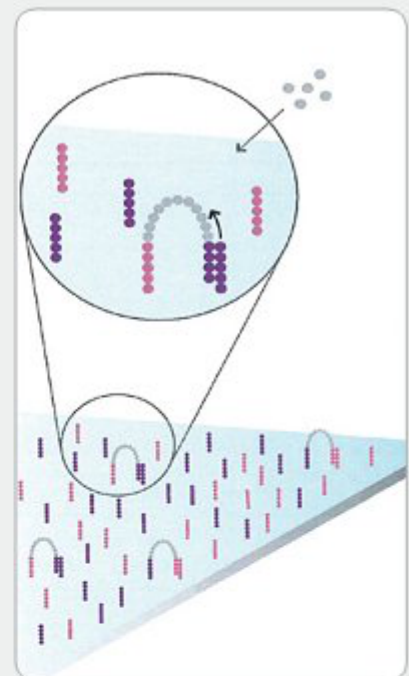
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



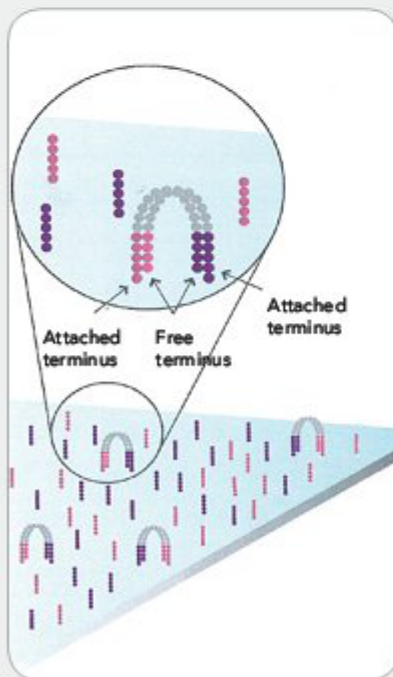
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



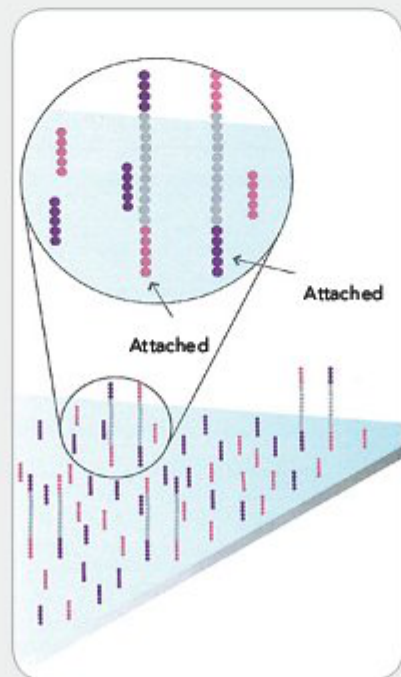
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE STRANDED



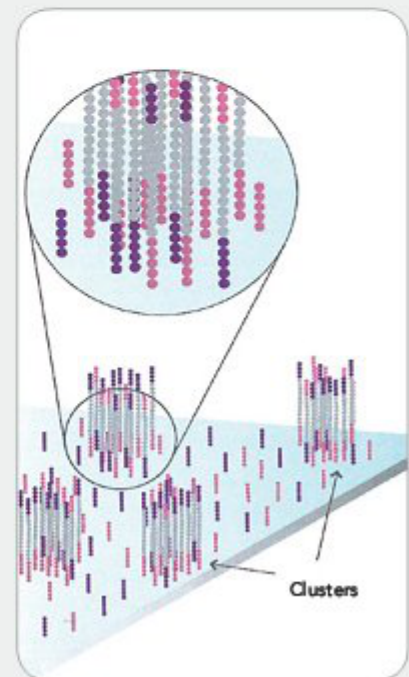
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

Obr. 10 - Příprava sekvenace na platformě Illumina
 Zdroj : <http://seqanswers.com/forums/images/content/ilmn-step1-6.jpg>, 21.4.2015

2.5.3 Využití metody

Díky kvalitní technologii a širokému spektru nabízených řešení nalézají přístroje společnosti Illumina uplatnění prakticky ve všech aplikacích souvisejících s NGS, počínaje *de novo* sekvenováním, přes sekvenování transkriptomu či cílené resekvenování až například po metagenomický výzkum. Pro svoje výhodné vlastnosti je platforma Illumina v současné době lídrem trhu s NGS technologiemi a její přístroje nalezneme v mnoha různých laboratořích po celém světě (Glenn, 2011, Quail et al., 2012).



obr. 11 - Sekvenátory společnosti Illumina. Zleva : MiSeq, NextSeq a HiSeq.
Zdroj : http://res.illumina.com/images/home/_v2/updated_ngs_section_v3.jpg. 16.11.2014

2.6 Další současné NGS metody a slibné metody ve vývoji

Kromě platform zmíněných výše figuruje na trhu NGS technologií ještě i několik dalších výrobců. Jejich přístroje ale zpravidla nejsou tolik rozšířené (popřípadě se vůbec neprodávají), nebo jsou zaměřeny pouze na jednu určitou aplikaci. Z firem, které vyrábí méně zastoupené NGS systémy se zmíním o Pacific Biosciences, Complete Genomics a Oxford Nanopore.

Společnost Pacific Biosciences (PacBio) vyrábí přístroje PacBio RS II, které jsou založeny na technologii nazvané SMRT (single molecule real time). Tato technologie je jednou z metod sekvenování syntézou a spočívá v real-time sledování tvorby DNA řetězce z fluorescenčně označených prekurzorů. Sekvenační reakce probíhá v mikroskopických jamkách na povrchu reakční destičky, nazvané SMRT Cell. Tyto jamky, pojmenované "zero-mode waveguides" (ZMW) mají nepatrný objem (20 zeptolitřů - tedy 2×10^{-20} l) a jemu odpovídající rozměry (v desítkách nanometrů). Jejich dno je tvořeno průsvitným materiálem a pod SMRT Cell jsou umístěny zdroj světla a detektor fluorescence. Sekvenování je zprostředkováno molekulou DNA polymerázy, ukotvenou na dně každé ZMW. Když polymeráza syntetizuje vlákno DNA, je ozařována zdrojem světla a fluorescence z každého zařazeného nukleotidu je odečítána v reálném čase (Eid et al., 2009).

Protože metoda produkuje velmi dlouhé reads, dosahující i několika desítek tisíc bází, je s výhodou používána pro *de novo* sekvenování. Uplatnění ovšem může najít například i při studiu epigenetických modifikací DNA (Quail et al., 2012, Flusberg et al., 2010).

Kalifornská firma Complete Genomics oproti tomu ve svých sekvenátorech implementuje variantu sekvenování ligací, nazvanou DNA nanoball sequencing. Sekvenovaná DNA je nejprve fragmentována a každý fragment je namnožen metodou zvanou "rolling circle replication". Z jednotlivých fragmentů tak vzniknou dlouhá vlákna DNA, obsahující stovky kopií původní sekvence. Vlákna jsou posléze sbalena do podoby malých kuliček ("DNA nanoballs") a uchycena k drobným ploškám (tzv. "sticky spots") na povrchu speciálního křemíkového čipu. Sekvenování samotné je prováděno technikou nazvanou Combinatorial Probe-Anchor Ligation (cPAL), spočívající v ligaci krátkých, fluorescenčně označených oligonukleotidů (Drmanac et al., 2009).

Complete Genomics zastává v porovnání s ostatními NGS výrobci poněkud netradiční způsob poskytování svých technologií. Namísto distribuce přístrojů a reagensů přímo do výzkumných či klinických zařízení firma od klientů pouze přijme vzorky DNA, které její zaměstnanci sami osekvenují ve vlastním sekvenačním centru v Kalifornii a klienti obdrží jen výsledky sekvenace.

Metoda již byla využita například k celogenomovému a transkriptomovému sekvenování nádorů plic (Liu J et al., 2012).

Pokrok v oblasti NGS ovšem s dnešními technologiemi neskončil. Řada firem pracuje na vývoji nových metod sekvenování, založených na odlišných přístupech. Mezi metody ve fázi výzkumu patří například sekvenování s využitím biologických pórů (nanopórové sekvenování). Tento princip používá např. firma Oxford Nanopore, která v roce 2014 představila miniaturní přístroj MinION o velikosti přenosného USB disku. Tento přístroj je v současnosti testován prvními uživateli, výhodou by měla být velká délka reads, nevýhodou je podle prvních zpráv značná chybovost čtení (Goodwin et al., 2015). Budoucnost ukáže, zda některé z těchto nových metod nabídnou dostatečné kvality a prosadí se mezi zavedenými NGS platformami.

3. Využití NGS metod při detekci přestaveb genů pro Ig a TCR

3.1 Úvod do problematiky vývoje antigen-specifických receptorů T a B lymfocytů

Imunoglobuliny (Ig) a T-buněčné receptory (TCR) jsou molekuly s klíčovým významem ve fyziologii imunitního systému. Jejich společným jmenovatelem je rozpoznávání antigenu, ať už v nativní formě (Ig) nebo v komplexu s MHC glykoproteiny (většina TCR).

Protože je množství možných antigenních struktur závratné a jejich variabilita značná, musí být podobně obsáhlý i repertoár antigenně specifických receptorů. Mechanismy, kterými je tohoto dosaženo, budou stručně popsány.

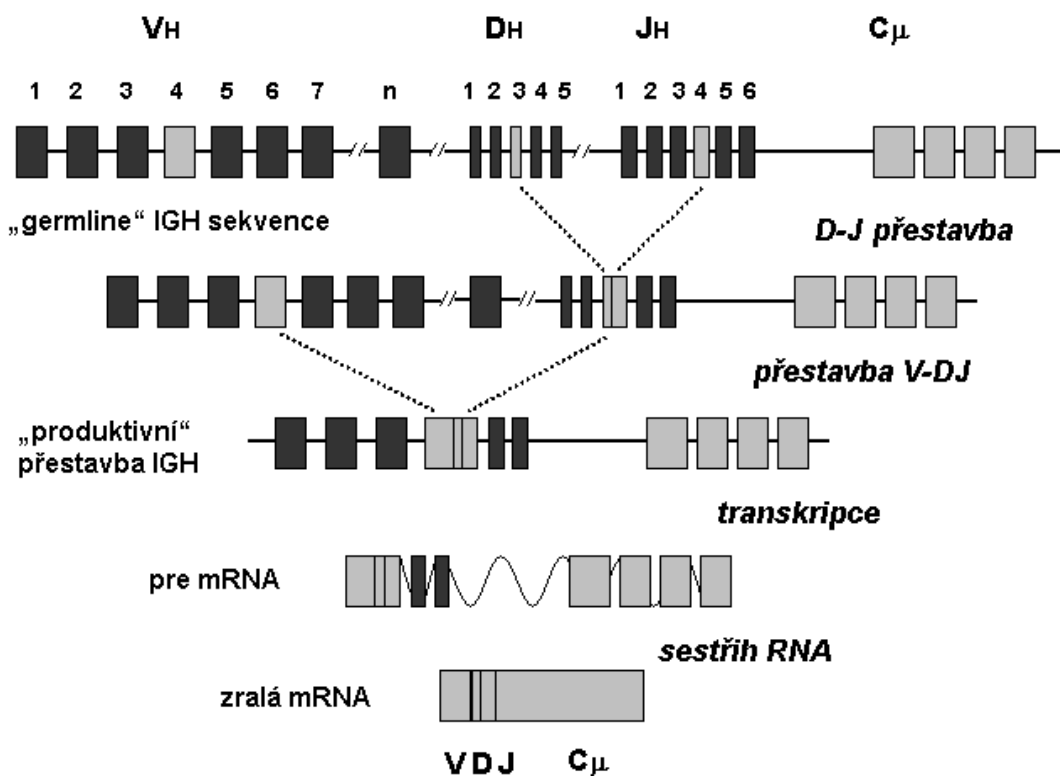
3.1.1 Vznik imunoglobulinů

Imunoglobuliny (protilátky) jsou antigenně specifické receptory lymfocytů B. Molekula imunoglobulinu má schematicky tvar písmene "Y" a je tvořena dvěma těžkými (IgH) a dvěma lehkými (IgL) řetězci, které jsou drženy pohromadě prostřednictvím disulfidových vazeb. Rozlišujeme pět typů těžkých řetězců (α , γ , δ , ϵ , μ) a dva typy lehkých řetězců (κ , λ) a platí, že jednu molekulu tvoří dva páry řetězců stejného typu (tedy např. dva těžké řetězce typu μ a dva lehké řetězce typu λ). Přítomnost konkrétního typu těžkého řetězce určuje příslušnost daného imunoglobulinu k jedné z pěti tříd (tzv. izotypů - IgA, IgG, IgD, IgE a IgM). Těžké řetězce se v závislosti na třídě, ke které náleží, skládají ze čtyř nebo pěti soudečkovitých domén (každá o velikosti cca 110 - 120 AA), lehké řetězce mají podobně velké domény dvě. N-koncové domény těžkých i lehkých řetězců se označují jako variabilní (ostatní domény jsou tzv. konstantní) a nesou antigen-vazebná místa, charakteristická vysokou variabilitou. Různorodosti antigen-vazebných míst imunoglobulinů je dosaženo specifickými procesy provázejícími jejich vznik. Geny, kódující těžké řetězce, jsou lokalizovány na chromosomu 14, kde tvoří genový komplex. Tento komplex nese na 5' konci několik desítek segmentů nazývaných V ("variability"), následovaných několika desítkami úseků D ("diversity") a několika segmenty J ("joining"), 3' konec komplexu obsahuje geny pro konstantní části těžkých řetězců (C).

Variabilní domény těžkých řetězců vznikají v procesu tzv. V-D-J rekombinace, kdy jsou jednotlivé genové úseky náhodně přeskupovány specializovanými enzymy na základě charakteristických krátkých sekvencí, které segmenty ohraničují ze 3' i 5' konce. V-D-J rekombinace probíhá současně na obou chromosomech.

Prvním krokem tvorby IgH je vyštěpení úseku mezi některým D segmentem a některým J segmentem

(D-J přeskupení) a toto je následováno vyštěpením úseku mezi některým V segmentem a D-J segmentem vzniklým v předchozím kroku (V-DJ přeskupení). Již touto rekombinací je dosaženo značné variability, která je ještě umocněna dvěma mechanismy. První zvýšení variability je způsobeno nepřesným odstrižením konců vyštěpených segmentů během D-J i V-DJ rekombinace, druhé vzniká náhodným prodloužením těchto konců enzymem terminální deoxynukleotidyltransferázou, čímž vznikají tzv. N-segmenty. Jako první se tvoří těžké řetězce typu μ , protože C_μ segment leží v genovém komplexu nejbližší 5' konci. Schéma tvorby těžkých řetězců ukazuje obr. 12.



Obr. 12 - Přestavby genových segmentů zahrnuté ve vývoji těžkých řetězců imunoglobulinů.

Zdroj : Detekce přestaveb genů pro imunoreceptory u akutní lymfoblastické leukémie a u zdravých krevních buněk. Disertační práce. MUDr. Eva Froňková (2006)

Jakmile dojde na jednom chromosomu k úspěšnému přeskupení genů, je přeskupování na druhém chromosomu zastaveno (tomuto jevu se říká alelická exkluze), produkt rekombinace je přepsán do mRNA a jsou vyštěpeny oblasti vzniklé přepisem intronů. Takto vzniklý těžký řetězec se dočasně asociuje s tzv. náhradními L řetězci (ψ L) a celý tento komplex je spolu se signálními proteiny CD79 α a CD79 β vystaven na povrchu pre-B lymfocytu. Úspěšná exprese tohoto tzv. pre-B receptoru je

pro buňku signálem k zahájení přeskupování genů pro lehké řetězce.

Lehké řetězce jsou kódovány genovými komplexy na chromosomu 2 (Ig κ) respektive 22 (Ig λ) a jejich struktura se podobá genovým komplexům IgH, chybí zde však D segmenty. Tvorba lehkých řetězců začíná rekombinací segmentů κ , kde dochází k vyštěpování segmentů mezi některým V a některým J segmentem tak dlouho, než vznikne produkt, který je po přepsání do polypeptidu schopný asociovat s těžkým řetězcem (typu μ) a dát tak vznik kompletnímu povrchovému IgM, schopnému nahradit pre-B receptor. Pokud dojde k úspěšnému přeskupení segmentů κ na jednom chromosomu, je přeskupování na druhém chromosomu zastaveno alelickou exkluzí. Jestliže není možné ani na jednom chromosomu získat přeskupováním κ segmentů funkční produkt schopný asociace s IgH, je zahájeno přeskupování segmentů λ . Když není úspěšné ani to, pre-B lymfocyt umírá. Expresí povrchového IgM se pre-B lymfocyt mění v nezralý B lymfocyt.

3.1.2 Vznik T-buněčných receptorů

Receptory lymfocytů T jsou, podobně jako imunoglobuliny, tvořeny dvěma řetězci, propojenými disulfidovou vazbou. Rozeznáváme dva typy TCR: TCR typu $\alpha\beta$ (majoritní typ) a TCR typu $\gamma\delta$ (obvykle výrazně méně zastoupený typ). TCR β a TCR δ obsahují D segmenty a jsou tedy obdobou IgH, TCR α a TCR γ jsou složeny jen z V a J úseků a v tomto se tudíž podobají IgL.

Uspořádání genových segmentů, které TCR kódují je analogické imunoglobulinům a velmi podobně probíhá také přeskupování těchto segmentů (účastní se jej i stejné enzymy).

3.2 Dosavadní metodiky detekce přestaveb Ig/TCR

V-(D)-J genová rekombinace, popsána výše, probíhá v každém B respektive T lymfocytu individuálně, díky čemuž je v organismu vytvářeno velké množství lymfocytů s různou specifitou. Tato skutečnost je podstatou antigenně specifické imunitní odpovědi. Přirozená skladba T respektive B lymfocytů v organismu je tedy vysoce polyklonální a převažující výskyt jednoho lymfocytárního klonu může být známkou probíhajícího lymfoproliferativního onemocnění.

Použitím vhodných metod je možné přestavby genů pro variabilní domény Ig respektive TCR detekovat a získat tak klinicky cenné informace s širokými možnostmi využití.

3.2.1 Detekce klonality

Nejprve se provádí amplifikace zkoumaných Ig/TCR segmentů, obvykle pomocí screeningového panelu multiplexní PCR cílené na kombinace jednotlivých rodin V a J segmentů. V klinické praxi se

nejčastěji používají systémy popsané skupinou BIOMED-2 zahrnující kompletní a nekompletní (D-J) přestavby IGH (těžkého řetězce imunoglobulinů), IGK (lehkého řetězce imunoglobulinů typu kappa), IGL (lehkého řetězce imunoglobulinů typu lambda), TCRD (řetězce T-buněčného receptoru typu delta), TCRG (řetězce T-buněčného receptoru typu gamma) a TCRB (řetězce T-buněčného receptoru typu beta) (van Dongen et al., 2003). Analýza PCR produktů na obyčejném agarosovém gelu ale nepodává jasnou informaci o klonalitě vzorku, protože výsledné produkty mají velmi podobnou velikost. Proto se používá tzv. analýza heteroduplexů na polyakrylamidovém gelu.

Princip heteroduplexové analýzy spočívá v ohřátí, které způsobí denaturaci PCR produktů, následném ochlazení a analýze na polyakrylamidovém gelu. Pokud jsou všechna vlákna ve vzorku dokonale komplementární (tzn. vzorek je monoklonální pro jedinou sekvenci Ig nebo TCR), vytvoří spolu tzv. homoduplex, jestliže spolu párují jen v některých úsecích, dojde naopak k vytvoření tzv. heteroduplexu. Směs dvojlákových fragmentů je následně elektroforeticky analyzována. Vzhledem k tomu, že molekula homoduplexu představuje pravidelnější útvar, prostupuje snáze nosným médiem a proto v něm za stejnou dobu doputuje dále než heteroduplex. Analýzou výsledného média (gelu) je tak možné zjistit, zda je sekvence monoklonální, oligoklonální (více menších klonů) nebo polyklonální (Bottaro et al., 1994, Langerak et al., 1997). Analýza klonality se stala nedílnou součástí diferenciální diagnostiky zejména lymfoproliferativních onemocnění ze zralých lymfocytů. Její výsledek je ovšem nutné vždy posuzovat s ohledem na závěry patologického a klinického vyšetření.

3.2.2 Detekce minimální reziduální nemoci (MRN)

Termínem minimální reziduální nemoc označujeme stav hematoonkologického pacienta po provedené léčbě (chemoterapii, radioterapii, HSCT či jiné), kdy v pacientově těle stále přetrvávají zbytkové nádorové buňky. Monitorování MRN dává specialistům šanci sledovat další vývoj onemocnění a v případě potřeby včas přistoupit k adekvátnímu léčebnému postupu. Hodnota MRN je tudíž významným prognostickým markerem, a proto se jejímu stanovení věnuje značná pozornost (van Dongen et al., 1998, Fronkova et al., 2008).

V případě stanovování MRN u akutní lymfoblastické leukémie (ALL) doposud převažovalo použití průtokové cytometrie (FC) a alelově specifická oligonukleotidová PCR (ASO-PCR). Vyhodnocení dat z průtokové cytometrie je ovšem velmi náročné na odbornou úroveň pracovníků a nedosahuje takové senzitivity jako ASO-PCR, která je sice citlivější, ale velmi pracná (vyšetření je třeba navrhnout na míru každému pacientovi). Ani jedna z těchto metod navíc téměř neumožňuje sledovat změny v

zastoupení jednotlivých leukemických klonů v průběhu léčby, což může vyústit ve falešně negativní výsledek. (Faham et al., 2012)

3.2.3 Analýza imunitního repertoáru

Repertoár přestaveb lymfocytů lze stávajícími metodami analyzovat pouze nepřímo, podle rozložení délek úseků naamplifikovaných pomocí PCR, které by u fyziologických vzorků mělo být přibližně Gaussovské. K tomu se používá metoda Immunoscope/Spectratyping, využívající mRNA izolovanou z lymfocytů, která je po přepsání do cDNA amplifikována pomocí PCR. Primery, použité při amplifikaci, jsou zacíleny na oblast CDR3 (complementarity determining region 3), která se vyznačuje nejširší sekvenční variabilitou v rámci variabilních domén Ig/TCR. Tato metoda ovšem neposkytuje informaci o nukleotidové sekvenci daných úseků. (Pannetier et al., 1995)

3.3 Přínos sekvenování nové generace pro studium přestaveb Ig/TCR

Metody masivního paralelního sekvenování daly imunologickému výzkumu zcela nový rozměr. Protože umožňují současné zpracování velkého množství vzorků, není s jejich použitím problém analyzovat značnou část antigenně specifického repertoáru jednotlivců i skupin subjektů. V následující kapitole práce stručně pojedná o některých doposud publikovaných studiích z dané oblasti.

3.3.1 Využití NGS metod ve výzkumu fyziologických přestaveb Ig/TCR

S použitím masivního paralelního sekvenování byla zavedena technologie, umožňující sekvenování variabilních úseků těžkých a lehkých řetězců imunoglobulinů (a také $\alpha\beta$ řetězců TCR), které zároveň zachovává informaci o párování těchto řetězců. Vzhledem k tomu, že skutečná diverzita B- a T-buněčných receptorů se skrývá právě v těchto párech, spíše než v jednotlivých variabilních monomerech, naleznou tyto údaje, které nebylo možné získat aplikací dosavadních postupů, jistě využití v dalším výzkumu antigenně specifických receptorů B- a T-lymfocytů a mohou být prospěšné například pro vývoj klinicky využitelných protilátek (DeKosky et al., 2013).

NGS technologie přispěly rovněž k poznání, že některé varianty TCR β řetězce u CD8⁺ T lymfocytů se vyskytují výrazně častěji než jiné. Konkrétně se jedná o řetězce, které obsahují krátké N-úseky (sekvence nukleotidů, které se nacházejí mezi V a D resp. D a J segmenty a jejichž tvorba je katalyzována enzymem terminální deoxynukleotidyl transferázou), t.j. N-úseky o délce kratší nebo rovné šesti nukleotidům. Bylo rovněž zjištěno, že při tvorbě β řetězců TCR jsou preferovány určité V a J segmenty a celkově tak vyplývá najevo, že antigenně specifické repertoáry jedinců v populaci jsou si

mnohem podobnější, než se předpokládalo dříve (Robins et al., 2010).

Výzkum v oblasti Ig/TCR přestaveb, využívající metod sekvenování nové generace, byl zaměřen i na přestavby TCR β v souvislosti s procesem, kdy se daná buňka stává buďto $\alpha\beta$ T lymfocyttem nebo $\gamma\delta$ T lymfocyttem (" $\alpha\beta/\gamma\delta$ commitment"). Výsledky ukazují, že zatímco geny pro TCR γ řetězec se přeskupují před $\alpha\beta/\gamma\delta$ přechodem, geny pro β řetězec procházejí přeskupováním až po rozhodnutí, že se buňka vyvine v $\alpha\beta$ T lymfocyt. Tento objev může dostát praktického využití například při sledování MRN u hematologických malignit, kdy na základě něj budou jako sledovaný cíl použity přestavby v genech kódující řetězec TCR γ , neboť ty jsou (narozdíl od přestavěných genů pro TCR β) přítomny u naprosté většiny leukemií či lymfomů (Sherwood et al., 2011).

Jedním z výzkumných směrů, pro které bude NGS patrně velkým přínosem, je studium změn v imunitním systému, provázejících stárnutí organismu. V pozdějších etapách života je u zkoumaných osob možné pozorovat snížení schopnosti reagovat na nové antigenní podněty (dokladem toho je nízká odpovědávost na vakcinaci a zvýšená náchylnost k infekcím), spojené se zvýšeným výskytem nežádoucí imunitní aktivace (nárůst hladiny autoprotilátek a zánětlivých mediátorů). Výzkum těchto změn je zatím v počátcích, využití NGS již však pomohlo nastínit jeho možné budoucí směřování (Boyd et al., 2013).

3.3.2 NGS v problematice transplantací

Výzkum v této oblasti byl zaměřen převážně na vývoj nových technik, umožňujících efektivnější sledování posttranslační obnovy repertoáru TCR. Jedna z nových metod, která kombinuje RACE PCR (jeden z druhů polymerázové řetězové reakce, využívající reverzní transkripci) s masívním paralelním sekvenováním (konkrétně Roche/454), přispěla k lepšímu pochopení vývoje repertoáru TCR β u pacientů po transplantaci kmenových buněk krvetvorby (HSCT). Pacienti byli sledováni jak v čase (odběry krve v různých časových bodech po transplantaci), tak s ohledem na použitou metodu transplantace (konvenční HSCT, pupečnicková krev, periferní krev zbavená T-lymfocytů), zvláštní zřetel byl také kladen na souvislost postransplantačního výskytu GvHD s vývojem TCR β , či na vliv systémového podání steroidů na tento vývoj. Byl tak získán informativní obraz vývoje TCR β repertoáru dotyčných osob, a tato metoda tudíž může nalézt uplatnění v klinické praxi při sledování pacientů po transplantaci kmenových buněk krvetvorby (van Heijst et al., 2013).

Jiná studie, zaměřená na sledování TCR β repertoáru po transplantaci HSC odhalila, že množství klonů T-lymfocytů přežívá transplantaci a některé z nich se krátce po provedené HSCT množí a pomáhají

bojovat s infekcemi. Výsledky této práce jistě podníčí další výzkum v této oblasti, jehož cílem by mohla být úprava léčebných režimů u pacientů vyžadujících transplantaci, která bude zohledňovat důležitost zachování těchto lymfocytárních klonů (Mamedov et al., 2011).

Další technologie, využívající NGS, byla vyvinuta pro sledování vývoje repertoáru T-buněčných receptorů po transplantaci solidních orgánů. Častou komplikací transplantací solidních orgánů jsou virové infekce (mimo jiné infekce lidským cytomegalovirem - CMV), které mohou být až život ohrožující. Technika využívá PCR amplifikace TCR β úseků lymfocytární genomické DNA a jejich následného sekvenování na platformě Illumina, a nabízí tak možnost sledovat virově specifické klony T-lymfocytů, což přináší klinicky cenné údaje, které mohou posloužit pro úpravu léčby každého konkrétního pacienta, například pokud by hrozila reaktivace viru, spojená s rizikem pro danou osobu (Dziubianau et al., 2013).

3.3.3 Sekvenování nové generace v souvislosti s výzkumem a léčbou autoimunitních onemocnění

Výzkum v oblasti studia autoimunitních chorob cílil například na podobu repertoáru T-buněčných receptorů při revmatoidní artritidě. Vědci srovnávali klonalitu periferní krve a synovie postižených kloubů s využitím sekvenační platformy Roche/454 a z výsledků plyne, že ačkoliv krev i synovie obsahovaly polyklonální populaci T-buněk, byla v synoviálních vzorcích patrná převaha několika vysoce zastoupených klonů. Vysoce zastoupené klony byly nejpočetnější v synoviích pacientů s počínající revmatoidní artritidou, při déletrvajícím onemocnění jejich podíl klesal. Zajímavostí také je, že populace vysoce zastoupených klonů si byly navzájem velice podobné mezi jednotlivými postiženými klouby, nikoliv však již mezi klouby a periferní krví. Kloubní prostředí tak zřejmě představuje určitou "niku", která je využívána k preferenční proliferaci určitých T-lymfocytárních klonů. (Klarenbeek et al., 2012).

Jiná studie byla zaměřena na objasnění stavu TCR repertoáru u pacientů trpících různými formami aplastické anémie. Výzkumný tým ustanovil novou metodu nazvanou NGS-spectratyping, která kombinuje PCR amplifikaci TCR β na úrovni RNA (cDNA) a NGS na platformě Roche/454 a vyvinul také software pro analýzu získaných dat. Tato nová metoda je schopná rychle a přesně určit diverzitu T-buněčných klonů ve sledovaném vzorku (s ohledem na senzitivitu významně předčí standardní spectratyping) a s jejím využitím mohli autoři m.j. vysledovat sníženou komplexitu repertoáru T-lymfocytárních receptorů u pacientů s variantami aplastické anémie v porovnání se zdravými kontrolami. (Krell et al., 2013).

3.3.4 Přínos masivně paralelního sekvenování pro studium hematologických malignit

Pozornost vědeckých týmů, které zkoumaly možné využití NGS při diagnostice a léčbě nádorových onemocnění krve, byla především soustředěna na hledání nových metod sledování MRN, které by dokázaly obejít některé nedostatky dosavadních postupů.

Snaha o odstranění problémů, zmíněných výše (kap. 3.2.2) vedla k návrhu techniky, založené na amplifikaci Ig/TCR přestaveb leukemické DNA pomocí univerzálních primerů a následném sekvenování získaných vzorků pomocí masivně paralelního sekvenování (konkrétně technologií LymphoSIGHT firmy Sequentia, která je založena na platformě Illumina). Tato metoda byla testována na množství vzorků a porovnávána s dosavadními přístupy. Z výsledků vyplývá, že nová technika dosahuje senzitivity 10^{-6} (tzn. je schopná nalézt jednu leukemickou buňku ve vzorku o milionu lymfocytů - vykazuje tedy citlivost o jeden řád vyšší než ASO-PCR a o dva řády vyšší než průtoková cytometrie) a navíc je velmi spolehlivá, což spolu s možností sledování klonální evoluce v průběhu léčby a univerzálním použitím (odpadá nutnost přizpůsobovat vyšetření na míru jednotlivým pacientům) naznačuje, že by se u pacientů s ALL mohla prosadit jako hlavní metoda pro sledování MRN (Faham et al., 2012).

Technologie LymphoSIGHT byla testována i v další studii, která se opět zabývala srovnáním ASO-PCR a NGS, tentokrát u tří různých nemocí odvozených od B-lymfocytů (ALL, lymfomu z plášťových buněk (MCL) a mnohočetného myelomu (MM)). Diagnostickým cílem byly klonálně specifické přestavby těžkého řetězce imunoglobulinů (IgH). Výzkumný tým také v tomto případě dospěl k podobným závěrům: NGS má alespoň srovnatelnou citlivost jako ASO-PCR a díky univerzálnímu použití představuje perspektivní diagnostický nástroj (Ladetto et al., 2014).

Pro výzkum kvantifikace MRN u chronické lymfocytární leukémie (CLL) bylo zvoleno použití sekvenačního přístupu, založeného na technologii 454. Ta poskytuje ve srovnání s jinými metodami sekvenování syntézou dlouhé reads (viz výše) a autoři studie navíc vyvinuli vlastní bioinformatické nástroje, schopné kompenzovat některé její slabé stránky (m.j. poměrně běžné inserce, delece či substituce v oblastech delších homopolymerních úseků). Vědci srovnávali výsledky stanovení MRN pomocí sekvenování nové generace, ASO-PCR a průtokové cytometrie u celkem 28 vzorků od šesti pacientů po transplantaci HSC. Sledovaným objektem byly rekombinované úseky IgH a to opět s využitím univerzálních primerů. Závěry prokázaly vysokou senzitivitu NGS metody (10^{-5}), jejíž výstupy navíc solidně korelovaly s výsledky ASO-PCR (avšak už méně s výsledky FC). Podobně jako jiné metody detekce MRN, využívající NGS, přináší i tato přidanou hodnotu v podobě možnosti

sledovat vývoj zastoupení lymfocytárních klonů v průběhu léčby. Použití univerzálních primerů navíc usnadňuje rozšíření této techniky a tak z jejích přínosů budou snad moci těžit jak pacienti s CLL, tak i nemocní trpící jinými chorobami s výskytem klonálně-specifických přestaveb imunoreceptorových genů (Logan et al., 2011).

4. Závěr

Z výše uvedeného je patrné, jak rozsáhlý dopad má a určitě i nadále bude mít masivně paralelní sekvenování na podobu moderní molekulární genetiky. Ustavičný vývoj v této oblasti s sebou kromě nových technologií, zlepšování přesnosti, usnadnění obsluhy a dalších pozitiv nese rovněž snižování ceny sekvenování. Jistým mezníkem v tomto ohledu byla hranice 1000\$ za kompletní osekvenování jednoho lidského genomu, kterou se v nedávné době podařilo prolomit firmě Illumina. Jde o úspěch, který je možná prvním krokem na cestě k období personalizované medicíny, kdy bude celogenomové sekvenování pacientů běžné. Tento mocný diagnostický nástroj přinese řadu nových možností, s jeho použitím bude však zajisté spojena velická zodpovědnost. Především vzhledem k některým etickým aspektům by se totiž mohl velmi snadno stát dvojsečnou zbraní a je tedy jisté důležité, aby se příslušní odborníci v předstihu začali touto problematikou zabývat.

V rámci studia přeskupování genů pro variabilní úseky imunoreceptorů přineslo NGS jak vylepšení současných metod stanovování MRN, tak rovněž nezanedbatelné objevy v jiných oblastech. Lze očekávat, že vzhledem k rapidnímu vývoji na poli sekvenačních technologií, budeme v nedaleké budoucnosti moci sledovat další klíčový pokrok.

Seznam použité literatury

- Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, Andersen M and Corless ChL. (2013). Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. *J Mol Diagn*, **15**, 171-76.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53-59.
- Bottaro M, Berti E, Biondi A, Migone N, Crosti L. (1994). Heteroduplex analysis of T-cell receptor gamma gene rearrangements for diagnosis and monitoring of cutaneous T-cell lymphomas. *Blood*, **83**, 3271–3278.
- Boyd SD, Liu Y, Wang C, Martin V and Dunn-Walters DK. (2013). Human lymphocyte repertoires in ageing. *Curr Opin Immunol*, **25**, 511-15.
- DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, Varadarajan N, Giesecke C, Dörner T, Andrews SF, Wilson PC, Hunicke-Smith SP, Willson CG, Ellington AD and Georgiou G. (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol*, **31**, 166-69.
- Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A et al. (2009). Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science*.
- Dziubianau M, Hecht J, Kuchenbecker L, Sattler A, Stervbo U, Rödelsperger C, Nickel P, Neumann AU, Robinson PN, Mundlos S, Volk HD, Thiel A, Reinke P, Babel N. (2013). TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell-related pathology. *Am J Transplant*, **13**, 2842-54.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F et al. (2009). Real-Time DNA sequencing from Single Polymerase Molecules. *Science*, **323**, 133-138.
- Faham M, Zheng J, Moorhead M, Carlton VE, Stow P, Coustan-Smith E, Pui CH and Campana D. (2012). Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. *Blood*, **120**, 5173-80.
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J and Turner SW. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat methods*, **7**, 461-465.
- Fronkova E, Mejstrikova E, Avigad S, Chik KW, Castillo L, Manor S, Reznickova L, Valova T, Zdrahalova K, Hrusak O, Jabali Y, Schrappe M, Conter V, Izraeli S, Li CK, Stark B, Sary J and Trka J. (2008). Minimal residual disease (MRD) analysis in the non-MRD-based ALL IC-BFM 2002 protocol for childhood ALL: is it possible to avoid MRD testing? *Leukemia*, **22**, 989-97.

- Glenn TC. (2011). Field guide to next-generation DNA sequencers. *Mol Ecol Resour*, **11**, 759-69.
- Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC and McCombie WR. (2015). Oxford Nanopore Sequencing and *de novo* Assembly of a Eukaryotic Genome. *BioRxiv*.
- Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, Bhatnagar S, Murray A, Craft J, Wardell B, Bastian M, Mitchell J, Chen J, Tran T, Williams D, Potter J, Jammulapati S, Perry M, Morris B, Roa B and Timms K. (2015). Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. *BMC Cancer*, **15**, 215.
- Klarenbeek PL, de Hair MJ, Doorenspleet ME, van Schaik BD, Esveldt RE, van de Sande MG, Cantaert T, Gerlag DM, Baeten D, van Kampen AH, Baas F, Tak PP and de Vries N. (2012). Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Ann Rheum Dis*, **71**, 1088-93.
- Krell PF, Reuther S, Fischer U, Keller T, Weber S, Gombert M, Schuster FR, Asang C, Stepensky P, Strahm B, Meisel R, Stoye J and Borkhardt A. (2013). Next-generation-sequencing-spectratyping reveals public T-cell receptor repertoires in pediatric very severe aplastic anemia and identifies a β chain CDR3 sequence associated with hepatitis-induced pathogenesis. *Haematologica*, **98**, 1388-96.
- Ladetto M, Bruggemann M, Monitillo L, Ferrero S, Pepin F, Drandi D, Barbero D, Palumbo A, Passera R, Boccadoro M, Ritgen M, Gokbuget N, Zheng J, Carlton V, Trautmann H, Faham M and Pott C. (2014). Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. *Leukemia*, **28**, 1299-307.
- Langerak AW, Szczepanski T, van der Burg M, Wolvers-Tettero ILM, van Dongen JJM. (1997). Heteroduplex PCR analysis of rearranged T cell receptor genes for clonality assessment in suspect T cell proliferations. *Leukemia*, **11**, 2192-2199.
- Liu J, Lee W, Jiang Z, Chen Z, Jhunjhunwala S, Haverty PM, Gnad F, Guan Y, Gilbert HN, Stinson J, Klijn Ch, Guillory J, Bhatt D, Vartanian S et al. (2012). Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res*, **22**, 2315-27.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L and Law M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, **2012**, 251364.
- Logan AC, Gao H, Wang C, Sahaf B, Jones CD, Marshall EL, Buno I, Armstrong R, Fire AZ, Weinberg KI, Mindrinos M, Zehnder JL, Boyd SD, Xiao W, Davis RW and Miklos DB. (2011). High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment. *Proc Natl Acad Sci U S A*, **108**, 21194-9.
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J and Pallen MJ. (2012).

Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*, **30**, 434-9.

- Mamedov IZ, Britanova OV, Bolotin DA, Chkalina AV, Staroverov DB, Zvyagin IV, Kotlobay AA, Turchaninova MA, Fedorenko DA, Novik AA, Sharonov GV, Lukyanov S, Chudakov DM and Lebedev YB. (2011). Quantitative tracking of T cell clones after haematopoietic stem cell transplantation. *EMBO Mol Med*, **3**, 201-7.
- Mardis ER. (2007). The impact of next-generation sequencing technology on genetics. *Trends in genet*, **24**, 133-141.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho ChH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF and Rothberg JM. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **441**, 376-380.
- McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser ChR, Duncan C, Ichikawa JK, Lee CC, Zhang Z, Ranade SS, Dimalanta ET, Hyland FC, Sokolsky TD, Zhang L, Sheridan A, Fu H, Hendrickson CL, Li B, Kotler L, Stuart JR, Malek JA, Manning JM, Antipova AA, Perez DS, Moore MP, Hayashibara KC, Lyons MR, Beaudoin RE, Coleman BE, Laptewicz MW, Sannicandro AE, Rhodes MD, Gottimukkala RK, Yang S, Bafna V, Bashir A, MacBride A, Alkan C, Kidd JM, Eichler EE, Reese MG, De La Vega FM and Blanchard AP. (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res*, **19**, 1527-1541.
- Pannetier C, Even J and Kourilsky P. (1995). T-cell repertoire diversity and clonal expansions in normal and clinical samples. *Immunol today*, **16**, 176-81.
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP and Gu Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.
- Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR, Carlson ChS and Warren EH. (2010). Overlap and effective size of the human CD8⁺ T-cell receptor repertoire. *Sci Transl Med*, **2**, 47ra64.
- Ronaghi M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome Res*, **11**, 3-11.
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Shorn M, Alanjary M, Dimalanta

- E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT and Bustillo J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, **475**, 348-52.
- Sanger F, Nicklen S and Coulson AR. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, 5463-67.
- Sherwood AM, Desmarais C, Livingston RJ, Andriesen J, Haussler M, Carlson CS and Robins H. (2011). Deep sequencing of the human TCR γ and TCR β repertoires provides evidence that TCR β rearranges after $\alpha\beta$, $\gamma\delta$ T-cell commitment. *Sci Transl Med*, **3**, 90ra61.
- Syed F, Grunenwald H and Caruccio N. (2009). Next-generation sequencing library preparation: simultaneous fragmentation and tagging using *in vitro* transposition. *Nat Methods application note*, **10**, 782.
- van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuurin E, Garcia-Sanz R, van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M and Macintyre EA. (2003). Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*, **17**, 2257-317.
- van Dongen JJ, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L, Stolz F, Schrappe M, Masera G, Kamps WA, Gadner H, van Wering ER, Ludwig WD, Basso G, de Bruijn MA, Cazzaniga G, Hettinger K, van der Does-van den Berg A, Hop WC, Riehm H and Bartram CR. (1998). Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet*, **352**, 1731-8.
- van Heijst JW, Ceberio I, Lipuma LB, Samilo DW, Wasilewski GD, Gonzales AM, Nieves JL, van den Brink MR, Perales MA and Pamer EG. (2013). Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation. *Nat Med*, **19**, 372-7.