

Inovativno sekvenciranje tehnikom dugih očitavanja (Long-read WGS) - na pragu ere sveobuhvatne genomske pedijatrijske medicine

Šabić, Magdalena

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, School of Medicine / Sveučilište u Zagrebu, Medicinski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:105:568066>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-13**



Repository / Repozitorij:

[Dr Med - University of Zagreb School of Medicine Digital Repository](#)



**SVEUČILIŠTE U ZAGREBU
MEDICINSKI FAKULTET**

MAGDALENA ŠABIĆ

**Inovativno sekvenciranje tehnikom dugih očitavanja
(Long-read WGS) - na pragu ere sveobuhvatne
genomske pedijatrijske medicine**

DIPLOMSKI RAD



Zagreb, 2023.g.

Ovaj diplomski rad izrađen je na Klinici za pedijatriju Kliničkog bolničkog centra Zagreb pod vodstvom prof. dr. sc. Maria Ćuka, dr. med i predan je na ocjenu u akademskoj godini 2022./2023.

POPIS KORIŠTENIH KRATICA

cDNA- komplementarna deoksiribonukleinska kiselina (*engl.* complementary deoxyribonucleic acid)

DNA- deoksiribonukleinska kiselina (*engl.* deoxyribonucleic acid)

LRS- sekvenciranje tehnikom dugih očitavanja (*engl.* long-read sequencing)

miRNA – mikro-ribonukleinska kiselina (*engl.* microRNA)

NGS – nova generacija sekvenciranja (*engl.* next-generation sequencing)

ONT- *engl.* Oxford Nanopore Technologies

PacBio- *engl.* Pacific Biosciences

PCR- lančana reakcija polimeraze (*engl.* polymerase chain reaction)

RNA- ribonukleinska kiselina (*engl.* ribonucleic acid)

SRS- sekvenciranje tehnikom kratkih očitavanja (*engl.* short-read sequencing)

WGS- cijelogenomsko istraživanje (*engl.* whole-genome sequencing)

SADRŽAJ

SAŽETAK

SUMMARY

1. UVOD	1
2. PRINCIP RADA LRS (LONG-READ SEQUENCING) TEHNOLOGIJA....	2
3. PREDNOSTI KORIŠTENJA LONG- READ SEQUENCING (LRS) U ODNOSU NA OSTALE METODE SEKVENCIRANJA	4
4. NEDOSTATCI LONG-READ SEQUENCING TEHNOLOGIJA.....	20
5. ZAKLJUČAK	22
6. ZAHVALE.....	23
LITERATURA.....	24

SAŽETAK

Inovativno sekvenciranje tehnikom dugih očitavanja (Long-read WGS)- na pragu ere sveobuhvatne genomske pedijatrijske medicine

Autor: Magdalena Šabić

Ovaj pregledni rad pruža uvid u metodologiju Long- Read Sequencing (LRS) tehnologija. One se odlikuju generiranjem dužih očitavanja u usporedbi sa Short- Read Sequencing-om (SRS). Ovakav pristup donosi niz prednosti. Na primjer, duža očitavanja omogućuju izravno određivanje složenih strukturnih varijanti, poput inverzija i translokacija, koje je teže detektirati SRS-om. Osim toga, LRS je koristan u analizi ponavljajućih sekvenci koje često predstavljaju izazov za SRS metode. Prednost LRS-a je i mogućnost dobivanja informacija o metilacijskom profilu iz jednog uzorka. Metilacija je epigenetska modifikacija koja igra važnu ulogu u regulaciji gena i razvoju bolesti. Također, iz samo jednog uzorka moguće je odrediti segregaciju. Prikazane su prednosti ovog pristupa u dijagnostici i njegovi doprinosi u odnosu na trenutne tehnologije. Osim prednosti, rad se dotiče i nedostataka ove tehnologije te ukazuje na područja koja mogu biti poboljšana. Jedan od njih je niža preciznost u određivanju baznih parova u usporedbi sa SRS-om. Postoji prostor za napredak u poboljšanju preciznosti LRS-a i smanjenu troškova tehnologije kako bi postala široko dostupna i primjenjiva u kliničkim postavkama.

Ključne riječi: cijelogenomsko istraživanje, genom, sekvenciranje genoma

SUMMARY

Long-Read WGS- the doorstep of comprehensive genomic pediatric medicine era
Author: Magdalena Sabic

This review provides an insight into the methodology of Long-Read Sequencing (LRS) technologies. They are characterized by generating longer reads compared to Short-Read Sequencing (SRS). This approach brings several advantages. For example, longer reads enable the direct determination of complex structural variants, such as inversions and translocations, which are more challenging to detect with SRS. Additionally, LRS is useful in analyzing repetitive sequences that often pose a challenge for SRS methods. Another advantage of LRS is the ability to obtain information about the methylation profile from a single sample. Methylation is an epigenetic modification that plays an important role in gene regulation and disease development. Furthermore, segregation can be determined from a single sample. The review presents the benefits of this approach in diagnostics and its contributions compared to current technologies. In addition to the advantages, the paper addresses the limitations of this technology and points out areas that can be improved. One of them is the lower accuracy in determining base pairs compared to SRS. There is room for improvement in enhancing the accuracy of LRS and reducing the costs of the technology to make it widely accessible and applicable in clinical settings.

Key words: genome, genome sequencing, whole-genome research

1. UVOD

Gotovo 25 godina nakon što je otkrivena struktura DNA molekule, objavljena je prva metoda za sekvenciranje DNA poznata kao Sangerovo sekvenciranje. Metoda je nastavila napredovati s uvođenjem kapilarne elektroforeze i postala je široko prihvaćena kao „prva generacija sekvenciranja“. Međutim, budući da se samo jedna sekvencijska reakcija mogla analizirati istovremeno, metoda je imala ograničenu propusnost. Bilo je očito da se moraju razviti nove tehnologije kako bi se proširile mogućnosti sekvenciranja (1).

Nova generacija sekvenciranja, poznata kao Short- Read Sequencing (SRS) predstavlja sljedeći napredak nakon tradicionalnog Sangerovog sekvenciranja. Uobičajena značajka SRS-a je masovno sekvenciranje kratkih, klonalno umnoženih DNA molekula koje se sekvenciraju paralelno (2).

Nedugo nakon pojave nove generacije sekvenciranja, pojavila se treća generacija sekvenciranja, poznata kao Long-Read Sequencing (LRS). Osnovne karakteristike treće generacije sekvenciranja uključuju sekvenciranje pojedinačnih molekula i sekvenciranje u stvarnom vremenu, za razliku od nove generacije sekvenciranja gdje se sekvenciranje zaustavlja nakon svakog ugrađenog nukleotida. Dok metode nove generacije sekvenciranja mogu proizvesti očitavanja od nekoliko stotina parova baza, tehnologije dugih očitavanja generiraju očitavanja duljine nekoliko desetaka kilobaza pa sve do nekoliko megabaza. Osim toga, nekorištenje PCR umnožavanja dovodi do manje grešaka vezanih za umnožavanje i homogenije pokrivenosti genoma (3).

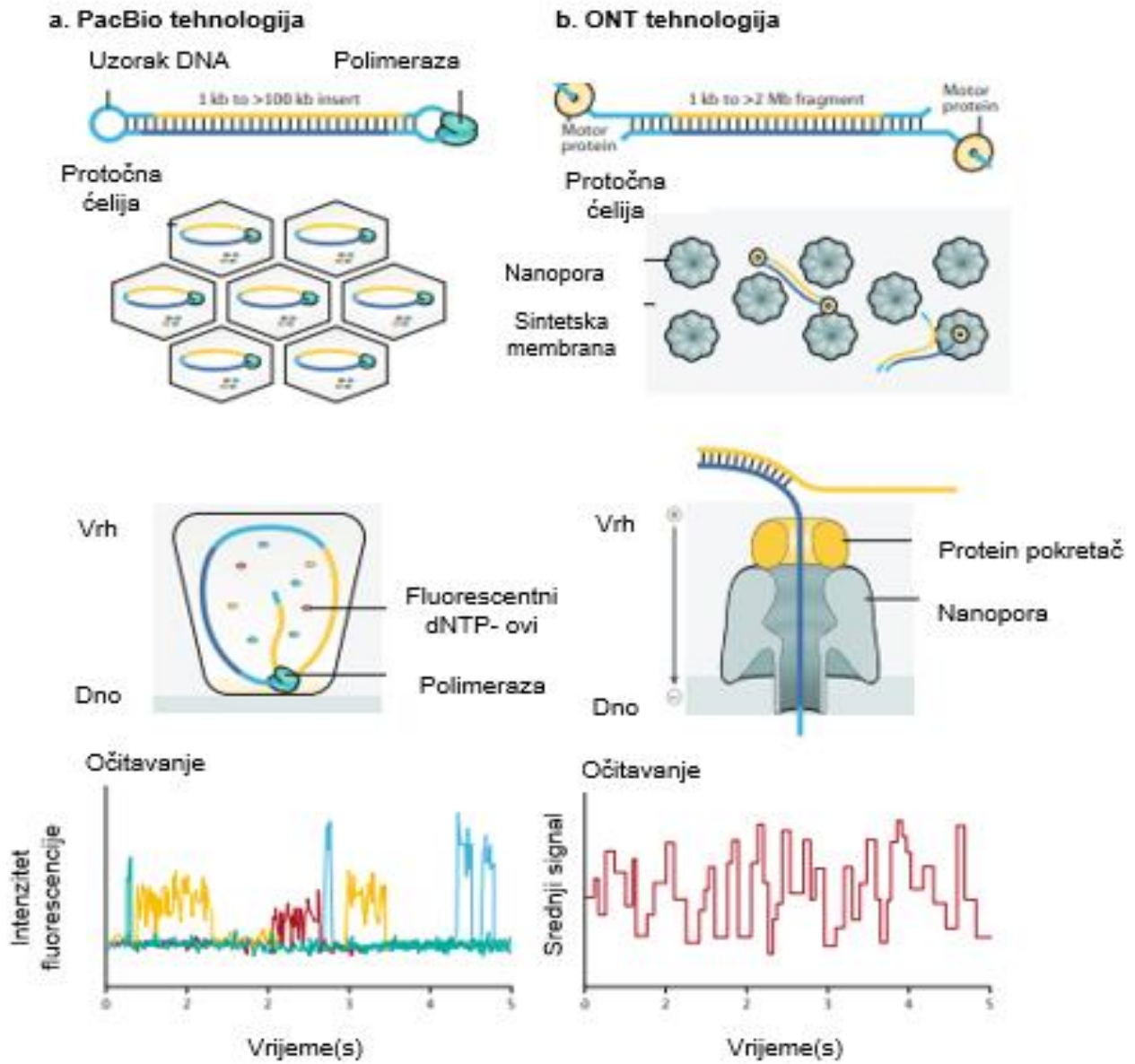
Dok su tehnologije nove generacije sekvenciranja donijele značajna poboljšanja u odnosu na Sangerovo sekvenciranje, njihove ograničenosti poput kratke dužine očitavanja čine ih neprikladnima za sekvenciranje složenih dijelova genoma, određivanje izoformi gena i detekciju metilacije. Long- read sequencing pruža alternativni pristup za prevladavanje tih ograničenja (4).

2. PRINCIP RADA LRS (LONG-READ SEQUENCING) TEHNOLOGIJA

LRS (Long-Read Sequencing) tehnologije omogućuju sekvenciranje fragmenata DNA izravno iz native DNA, što rezultira sekvenciranjem fragmenata dužine od 10 kilobaza pa sve do nekoliko megabaza. Zbog ove karakteristike, LRS tehnologije sve više se primjenjuju. Pacific Biosciences (PacBio) i Oxford Nanopore Technologies (ONT) su dvije glavne tehnologije LRS-a (5).

U PacBio tehnologiji sekvenciranja, uzorak DNA je kružna molekula koja može biti duga od jedne do nekoliko stotina kilobaza (6). Kružna DNA se pričvršćuje za tisuće malih jamica, a svaka jama sadrži DNA polimerazu. Nukleotidi označeni fluorescencijom koriste se za produkciju novog lanca DNA. Kada polimeraza ugradi nukleotid u novi lanac, mjeri se fluorescentni signal. Ovaj proces se odvija u svakoj od tisuću jama istovremeno, omogućavajući sekvenciranje u stvarnom vremenu. Na taj način se generira dugi niz sekvence čitanja (7).

ONT tehnologije koriste linearnu DNA koja je obično duga nekoliko stotina kilobaza, ali može biti i nekoliko megabaza. Sekvenciranje započinje vezanjem DNA polimeraze i proteina pokretača za DNA. Zatim se DNA umeće u protočnu ćeliju koja sadrži tisuće jako malih rupa, nazvanih nanopore, ugrađenih u sintetske membrane. Protein pokretač odmotava dvostruku DNA i s električnom strujom gura DNA kroz poru. Kako se DNA kreće kroz nanoporu, svaka baza mijenja tijek struje na drugačiji način. Promjene u struji se čitaju i prevode kao niz baza u lancu DNA (8).



Slika 1: preuzeto i prilagođeno prema: Logsdan G, Vollger M, Eichler E (2020)(6)

Slikovni prikaz dviju LRS tehnologija- Pacific Biosciences(PacBio) i Oxford Nanopore Technologies(ONT).

3. PREDNOSTI KORIŠTENJA LONG- READ SEQUENCING (LRS) U ODNOSU NA OSTALE METODE SEKVENCIRANJA

Nova generacija sekvenciranja (NGS), koja se pojavila prije više od jednog desetljeća, postala je rutinska dijagnostička metoda u modernoj genetici te je omogućila revolucionarno otkriće brojnih gena uzročnika bolesti. NGS je postao popularan zbog svoje veće učinkovitosti i niže cijene u usporedbi s prethodnom metodom sekvenciranja poznatom kao Sangerovo sekvenciranje (9). Međutim, NGS ima određena ograničenja zbog korištenja kratkih očitavanja koja su nužna za održavanja visoke kvalitete sekvenciranja.

Ljudski genom ima kompleksnu prirodu koja predstavlja izazov za sekvenciranje kratkim očitavanjima. Unatoč inovativnim informatičkim algoritmima, korištenjem NGS-a često je nemoguće precizno mapirati kratka očitavanja koja potječu od strukturnih varijanti, ponavljajućih sekvenci, sadržaja s visokim udjelom GC-nukleotida ili sekvenca s višestrukim homolognim elementima (10). Tehnologija sekvenciranja dugih očitavanja (LRS) pokazala je obećavajuće rezultate zbog identificiranja dotad neidentificiranih dijelova ljudskog genoma te prevladavanja prepreka na koje nailazi NGS.

Glavna prednost LRS-a u odnosu na druge tehnologije sekvenciranja je korištenje dugih očitavanja dobivenih iz jedne molekule DNA (11). Ovo omogućuje sekvenciranje bez potrebe za PCR umnožavanjem DNA, što smanjuje moguće greške vezane uz PCR. Osim toga, budući da se ne koristi PCR, DNA ostaje u svom prirodnom stanju, omogućujući da LRS tehnologije izravno identificiraju modifikacije baza, kao što je metilacija (12).

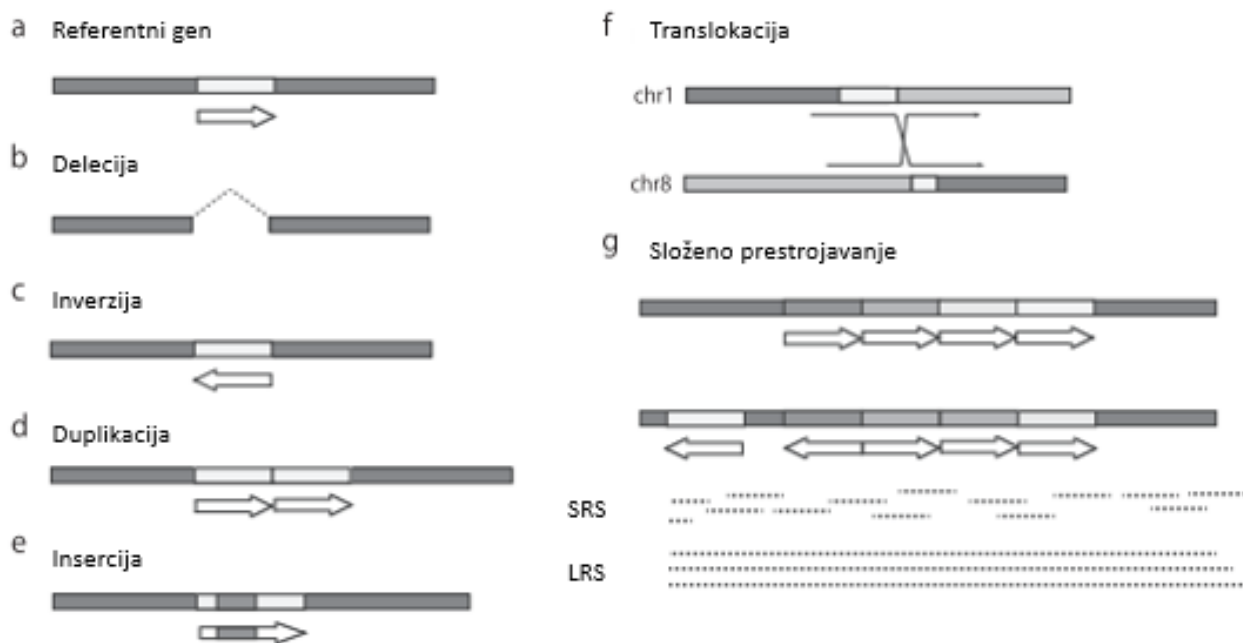
Glavna razlika između LRS-a i konvencionalnog NGS pristupa je značajno povećanje dužine očitavanja. Za razliku od kratkih očitavanja NGS-a dužine između 150 i 300 parova baza, LRS može sekvencirati očitavanja prosječne dužine preko 10 kilobaza. Zahvaljujući ovim značajkama, LRS može sekvencirati one elemente genoma koji su izazovni za konvencionalne metode (13).

LRS ima prednost u identifikaciji i karakterizaciji velikih strukturnih varijanti (SV) uzrokovanih, primjerice, velikim inverzijama ili translokacijama. Također, LRS je učinkovitiji u sekvenciranju dugih tandemskih ponavljanja i regija s visokim udjelom GC-nukleotida. Ova tehnologija poboljšava segregaciju genetskih varijanti na homologne očeve ili majčine kromosome te omogućuje određivanje uzorka nasljeđivanja. Također je korisna u razlikovanju između klinički relevantnih gena i pseudogena. S obzirom na to da korištenje LRS-a omogućuje sekvenciranje izazovnih regija DNA bez potrebe za PCR umnožavanjem, ova tehnologija igra važnu ulogu u otkrivanju novih patogenih mutacija u ljudskim bolestima s nepoznatim genetskim uzrokom (14).

Kod detekcije strukturnih varijanti, kao što su velike delecije, duplikacije, inverzije, translokacije, SRS tehnologije pokazuju manjak osjetljivosti, višak lažno pozitivnih rezultata te krivu interpretaciju kompleksnih strukturnih varijanti (14). Razlog tome su kratka, umnožena očitavanja te nemogućnost rekonstrukcije originalne sekvence koja je veća od dužine čitanja (15). Suprotno tome, uporaba dugih očitavanja(LRS), omogućava lakšu interpretaciju strukturnih varijanti, kao i otkrivanje novih varijanti (16). U jednoj studiji znanstvenici su istraživali razliku SRS i LRS tehnologija u identifikaciji strukturnih varijanti. Pokazali su kako je korištenje LRS tehnologija rezultiralo 32% više identificiranih strukturnih varijanti u usporedbi sa SRS tehnologijama (12).

Strukturne varijante odgovorne su za velik dio genske raznolikosti među pojedincima. Tijekom zadnja dva desetljeća, istraživanja su pokazala da su strukturne varijante povezane s brojnim medicinskim stanjima, osobito s neurorazvojnim poremećajima kao što su autizam, shizofrenija, epilepsija te Alzheimerova bolest s ranim početkom. Istraživanja uloge strukturnih varijanti u patogenezi različitih bolesti uglavnom se oslanjaju na funkcionalne jedinice, posebice egzone. Novija istraživanja naglašavaju važnost funkcionalno značajnih elemenata u nekodirajućim regijama te njihov doprinos različitim bolestima poput raka, neurorazvojnih poremećaja i neurodegenerativnih bolesti (17).

Autori G.Begum, A.Albanna i suradnici (16) u svojoj studiji su istraživali obitelj u kojoj trojke imaju poremećaj spektra autizma. Budući da je rezultat SRS testiranja kodirajućih gena na patogene strukturne varijante bio negativan, željeli su istražiti strukturne varijante u nekodirajućim regijama. U tu svrhu usporedili su SRS i LRS tehnologije kako bi pokazali njihovu osjetljivost i specifičnost za detekciju strukturnih varijanti. Kao rezultat, LRS tehnologije otkrile su veći broj strukturnih varijanti (97.9%) u odnosu na SRS (91.1%). Većina varijanti identificiranih LRS tehnologijom bila je izvan kodirajućih regija DNA. Jedna od nekodirajućih RNA, miRNA, upravo je povezana sa spektrom autizma. Stoga, LRS predstavlja obećavajuću tehnologiju koja može prevladati nedostatke SRS tehnologija te olakšati analizu i interpretaciju strukturnih varijantu u humanom genomu.



Slika 2 : Preuzeto i prilagođeno prema: Mitsuhashi S, Matsumoto N (2019)(18)

Slika prikazuje različite oblike strukturnih varijanti. SRS može detektirati neke prijelomne točke, međutim ne može dobiti kompletnu sliku kompleksnih strukturnih varijanti kao LRS.

Strukturne varijante nalaze se u pozadini brojnih rijetkih bolesti, a s obzirom na nedostatke SRS metoda u njihovom otkrivanju, nekoliko studija ukazuje na superiornost korištenja LRS-a.

Studija Dutta i sur. (19) analizirala je pacijenta s intelektualnim poteškoćama i konvulzijama koji je imao balansiranu kromosomsku translokaciju između kromosoma X i 20. Korištenjem LRS tehnologije, identificirali su prijelomnu točku na intronu gena ARFGEF9 i između gena RASSF2 i SLC23A2. SRS analiza nije uspjela detektirati ovu translokaciju.

Mizuguchi i sur. (20) su pronašli deleciju dugu 12 kilobaza koja je bila negativna nakon SRS analize. Ova delecija, koja je sadržavala višestruke ponavljajuće elemente i bila bogata GC nukleotidima, nije bila detektirana SRS tehnologijom.

Merker i sur. (21) identificirali su heterozigotnu deleciju od 2184 parova baza u egzonu PRKAR1A koja je povezana s autosomno-dominantnim Carneyevim kompleksom. Iako je tu deleciju moguće identificirati SRS tehnologijom, autori su je smatrali nepraktičnom zbog visokog broja lažno pozitivnih rezultata.

Reiner i sur. (22) su precizno identificirali deleciju od 72.8 kilobaza kod pacijenta s autosomno- recesivnim Bardet-Biedl sindromom korištenjem LRS tehnologije.

Miao i sur. (23) su identificirali patogenu strukturnu varijantu u recesivnoj bolesti pohrane glikogena. Sanchis- Juan i sur. analizirali su duplikaciju-inverziju-duplikaciju u CDKL5 genu kod pacijenta s neonatalnom- hipoksemijskom encefalopatijom te uspješno odredili konfiguraciju ove kompleksne strukturne varijante, što je pothvat koji ne bi bio moguć korištenjem SRS tehnologija.

Svi ovi primjeri pokazuju da LRS tehnologije mogu biti učinkovitije u otkrivanju i identifikaciji strukturnih varijanti kod rijetkih bolesti u usporedbi sa SRS metodama. Uzimajući u obzir nedostatke SRS tehnologija za detekciju strukturnih varijanti, korištenje LRS-a može pružiti bolju identifikaciju i dijagnozu rijetkih bolesti (18).

Stanice raka sadrže mutacije u svojim genima, od kojih neke utječu na protoonkogene, a druge na tumor-supresorske gene što rezultira abnormalnom proliferacijom i inicijacijom ili progresijom karcinogeneze. Moderne tehnologije sekvenciranja omogućavaju lakšu identifikaciju i karakterizaciju mutacija gena raka. Dosad su SRS tehnologije uglavnom bile usmjerene na točkaste mutacije i jednonukleotidne polimorfizme, međutim identifikacija strukturnih varijanti u genomu raka predstavlja izazov za te tehnologije. LRS se sve češće koristi za identifikaciju i dešifriranje kompleksnih genoma raka (24). Upravo strukturne varijante koje se javljaju u nestabilnim genomima raka utječu na funkciju nekoliko onkogenih i tumor-supresorskih gena. Nedavna istraživanja pokazuju uspješnost LRS tehnologija u detekciji strukturnih varijanti u različitim vrstama raka što je prikazano u tablici 1.

Kategorija	Tehnologija sekvenciranja	Rak	Godina objavljivanja
Strukturna varijanta	ONT	Rak gušterače	2016.
	ONT	Tumor mozga	2017.
	PacBio	Rak dojke	2018.
	ONT	Rak pluća	2019.
	ONT	Rak dojke	2020.
	ONT	Rak jajnika, Rak prostate	2021.
	ONT	Rak jetre	2021.

Tablica 1: Nedavna istraživanja strukturnih varijanti u genomu raka koristeći LRS tehnologije.

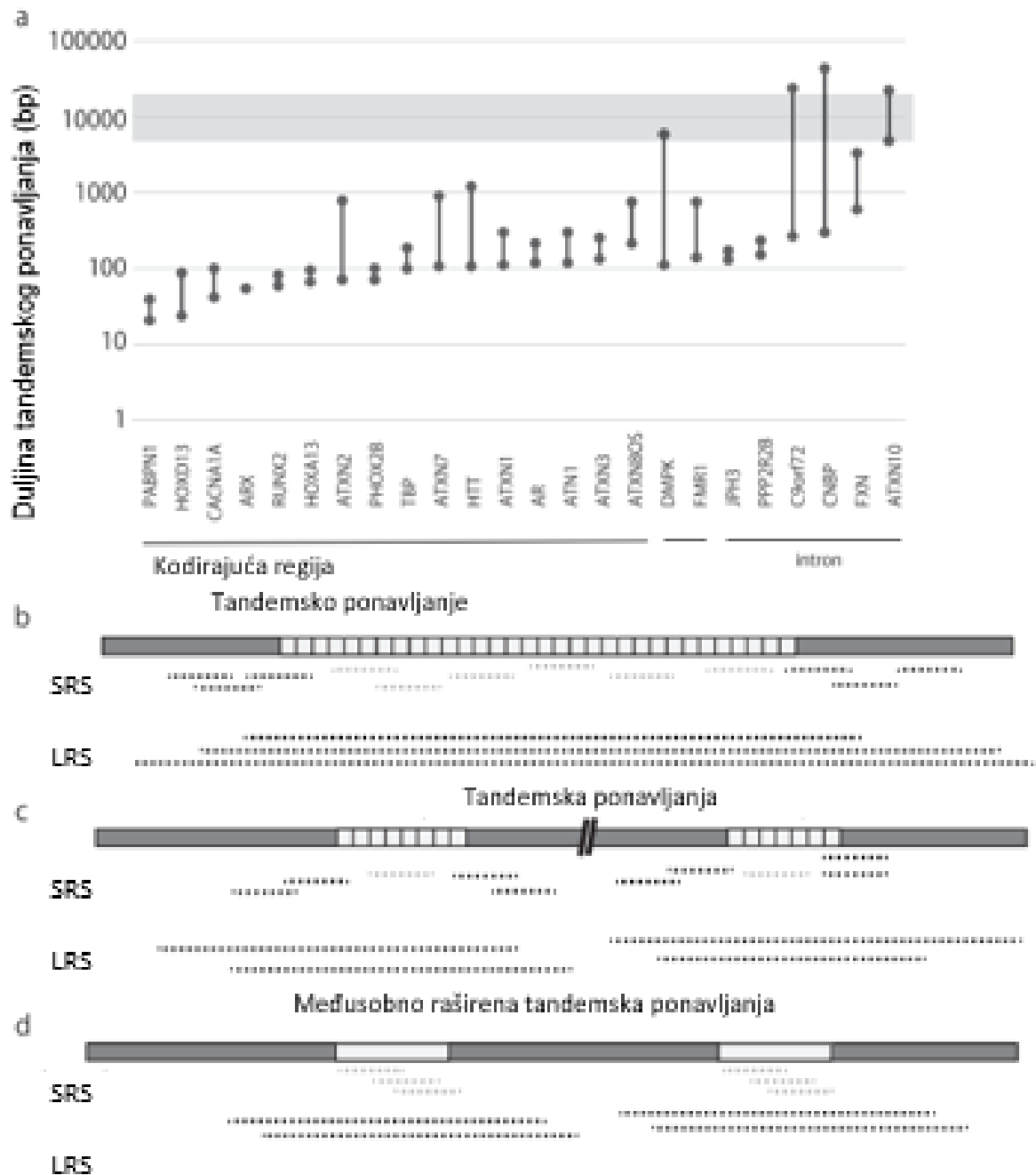
Preuzeto i prilagođeno prema: Sakamoto Y, Sereewattanawoot S, Suzuki A (2019) (ONT-Oxford nanopore technology; PacBio-Pacific Biosciences)(24)

Tablica prikazuje tehnologije sekvenciranja koje su korištene za identifikaciju strukturnih varijanti u genomima različitih vrsta raka, te godinu objavljivanja tih studija.

Tandemska ponavljanja su kratke regije DNA koje se višestruko ponavljaju unutar ljudskog genoma. Većina ovih ponavljanja nalazi se u nekodirajućim regijama genoma, no neka od njih mogu biti prisutna i u kodirajućim regijama. Istraživanja su pokazala da čak do 9% gena sadrži tandemska ponavljanja u kodirajućim regijama, dok ih oko 12% gena sadrži u svojim promotorskim regijama. Tandemska ponavljanja su poznata po tome što su vrlo osjetljiva na mutacije. One su često povezane s nekoliko genskih bolesti i imaju ulogu u forenzici i populacijskoj genetici. U usporedbi s ostalim dijelovima genoma, stopa mutacija u tandemskim ponavljanjima može biti čak 10 do 10000 puta veća (14).

Ranije metode istraživanja tandemskih ponavljanja, kao što su PCR i SRS, imale su određene nedostatke i ograničenja. Na primjer, većina bolesti uzrokovana ekspanzijama tandemskih ponavljanja često je povezana s duljim ponavljanjima od onih koje je moguće precizno očitati pomoću SRS tehnologija. Stoga, precizno identificiranje tih dugih ponavljanja nije bilo moguće tim metodama.

Međutim, različite studije su pokazale da su LRS tehnologije dobro prilagođene za identifikaciju ovih dugih ponavljanja. LRS tehnologije omogućuju generiranje dužih očitavanja DNA sekvenci, što je ključno za identifikaciju i karakterizaciju dugih tandemskih ponavljanja. Također, takva ponavljanja često sadrže visok udio GC- nukleotida, a LRS tehnologije mogu bolje obraditi takve sekvence bogate GC- sadržajem (25).



Slika 3: Preuzeto i prilagođeno prema: Mitsuhashi S, Matsumoto N (2019) (18)

Na slici su prikazana tandemsko ponavljanje poznatih genetskih bolesti i njihova duljina u parovima baza, te razlika SRS-a i LRS-a u otkrivanju različitih oblika tandemskih ponavljanja.

LRS može pokriti većinu poznatih patogenih ponavljanja. Treba uočiti da SRS ne može adekvatno otkriti različite oblike tandemskih ponavljanja. Za razliku od toga, LRS, zbog duljine očitavanja, može razlikovati tandemska ponavljanja na različitim genskim lokusima, kao i ona koja su međusobno raširena.

Na slici je također prikazano kako su tandemska ponavljanja u intronskim i netranslantirajućim regijama obično duža, te kako LRS tehnologije pokrivaju i njih. Ovo ukazuje na to da će se neotkrivene bolesti uzrokovane ponavljanjima pronaći korištenjem LRS tehnologija.

Poznato je da tandemska ponavljanja u više od 40 gena uzrokuju genske poremećaje od kojih su većina neurološki ili neuromuskularni. Primjeri takvih bolesti su Huntigtonova bolest, sindrom fragilnog X kromosoma, nasljedna cerebralna ataksija, mišićne distrofije, mioklone epilepsije, frontotemporalne demencije i amiotrofična lateralna skleroza (26). Svaka od ovih bolesti uzrokovana tandemskim ponavljanjima pogađa približno 1 do 10 od 100000 ljudi. Osim toga, otkrivaju se novi geni s tandemskim ponavljanjima koji su uzročnici bolesti, što dovodi do rasta broja poznatih bolesti povezanih s tim ponavljanjima. Tandemska ponavljanja pokazuju veliku varijabilnost i raznolikost u veličini, zbog čega je velik broj bolesti uzrokovan upravo njima. Zbog svega navedenog, Southern blot i PCR nisu adekvatne metode za njihovu identifikaciju jer precizne, spore su te zahtijevaju novu probu za svako tandemsko ponavljanje (27).

LRS ima sposobnost odabrati sve relevantne gene, uključujući i novootkrivene, odrediti veličinu ponavljanja te tražiti dodatne dijagnostičke ili prognostičke markere kao što su prekidi DNA ili metilacija. Zbog svega navedenog, LRS nadvladava ograničenja SRS tehnologija (28). Prednost LRS tehnologija leži u mogućnosti sekvenciranja regija bogatih GC- nukleotidima koje je teško sekvencirati koristeći SRS. Nekoliko studija pokazalo je kako LRS može sekvencirati FMR1 gen koji sadrži CGG ponavljanja te je odgovoran za fragilni X sindrom. Koristeći LRS tehnologije, određena je duljina i struktura ponavljanja te je time potvrđena veza između ponavljajućih struktura i kliničkog fenotipa bolesti (29). Ishiura i suradnici (30) koristili su LRS kako bi identificirali strukturu pentanukleotidnih ponavljanja kod pacijenata s benignom obiteljskom mioklonom epilepsijom.

Ta je bolest uzrokovana tandemskim ponavljanjima u SAMD12 genu što je također potvrđeno korištenjem LRS tehnologija. Mitsuhashi i suradnici (31) su koristili LRS tehnologije kako bi otkrili novo patogeno tandemsko ponavljanje u pacijenata s neuralnom intranuklearnom inkluzijskom bolesti. Lokus uzročnog gena bio je bogat GC- nukleotidima te ga SRS tehnologije nisu mogle identificirati kod pacijenata s neuralnom intranuklearnom inkluzijskom bolesti. Ispitivanje tandemskih ponavljanja koristeći LRS mora se uzeti u obzir kada uzročna mutacija u nasljednim bolestima nije pronađena koristeći SRS.

LRS je omogućio bolje razumijevanje neurodegenerativnih bolesti uzrokovanih tandemskim ponavljanjima zbog mogućnosti identifikacije novih mutacija, karakterizacije uzročnih ponavljanja te direktne detekcije epigenetskih modifikacija native DNA te će omogućiti daljnje razvijanje znanja o neurodegenerativnim bolestima (32). U tablici 2 prikazan je sažetak primjene LRS tehnologija u neurodegenerativnim bolestima.

Bolest	Vrsta mutacije	Lokus gena	Referenca/God. objavljivanja
Tremor/ataksija sindrom povezan s fragilnim X	CGG ekspanzija	FMR1	2018. (29)
Neuralna intranuklearna inkluzijska bolest	GGC ekspanzija	NOTCH2NLC	2019. (33)
Okulofaringodistalna miopatija	GCC ekspanzija	GIPC1	2020. (34)
Spinocerebralna ataksija tip 10	ATTCT ekspanzija	ATXN10	2017. (35)
Parkinsonova bolest	ATTCT ekspanzija	ATXN10	2017. (36)
Huntingtonova bolest	CAG Ekspanzija	HTT	2018. (37)
Benigna adultna familijarna mioklona epilepsija tip 2	TTTCA/TTTTA ekspanzija	SAMD12	2019. (38)
Familijarna adultna mioklona epilepsija tip 2	ATTTC ekspanzija	STARD7	2019. (39)
Familijarna adultna mioklona epilepsija tip 3	TTTCA/TTTTA ekspanzija	MARCH6	2019. (40)
Benigna adultna familijarna mioklona epilepsija tip 4	TTTCA/TTTTA ekspanzija	YEATS2	2019. (40)
Amiotrofična lateralna skleroza/ Frontotemporalna demencija	GGGGCC ekspanzija	C9ORF72	2018. (41)

Tablica 2: Preuzeto i prilagođeno prema: Su Y, Fan L, Shi C, Wang T, Zheng H, Luo H, i sur.(32)

Sažeti prikaz primjene LRS tehnologija u neurodegenerativnim bolestima.

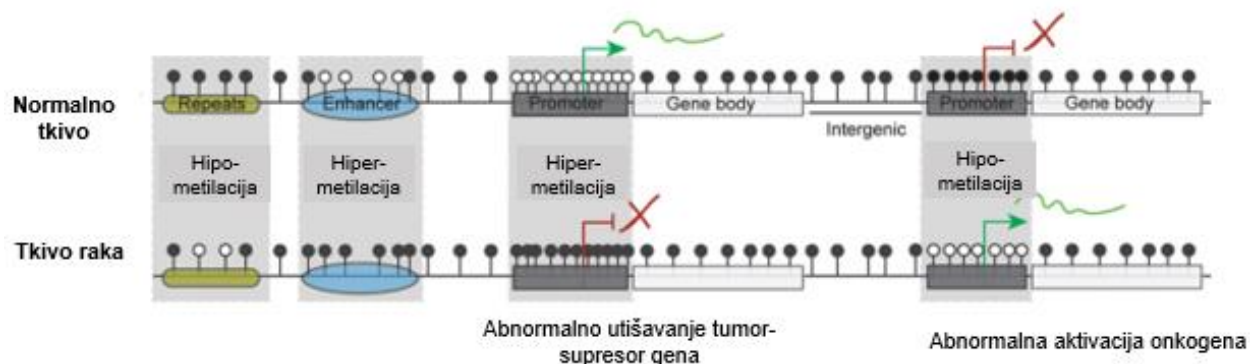
Metilacija DNA u stanicama sisavaca odvija se dodavanjem metilne grupe na citozin bazu, stvarajući 5-metilcitozin (5mC). Ovaj proces katalizira enzim poznat kao DNA-metiltransferaza. Metilacija DNA ima važnu ulogu u regulaciji ekspresije gena, utišavanju gena, stabilnosti centromere, kromosomskoj segregaciji tijekom mitoze, inaktivaciji X-kromosoma i genomskom imprintingu (42). Također, metilacija DNA ima važne implikacije u rijetkim bolestima poput Rettovog sindroma i Rubinstein-Taybijeovog sindroma. Kada se sumnja na genomski imprinting, analiza metilacijskog profila može pružiti važne informacije. Genomski imprinting je fenomen u kojem je samo jedan od dva roditeljska kromosoma izražen zbog metilacije koja se događa tijekom gametogeneze (43).

SRS je ograničen zbog nemogućnosti izravnog sekvenciranja epigenetskih i metilacijskih markera (14). Kod tradicionalnih metoda određivanja metilacijskih profila, poput disulfitnog sekvenciranja, DNA se prvo mora fragmentirati, što može dovesti do varijabilnih rezultata. Disulfitno sekvenciranje i slične tehnologije mogu identificirati modificirane citozine, ali ne mogu razlikovati različite vrste modifikacija citozina ili detektirati druge modificirane baze.

Najveća prednost LRS-a je mogućnost sekvenciranja native DNA bez prethodne pripreme i amplifikacije. LRS također omogućuje simultanu detekciju različitih vrsta modificiranih baza, uključujući 4-metilcitozin, 5-metilcitozin, 5-hidroksimetilcitozin, N-metiladenin i 8-oksogvanin (44). Ova metoda je jednostavnija i lakša jer ne zahtijeva tretiranje uzoraka disulfitima.

Primjena LRS sekvenciranja dovela je do otkrića razlika u metilacijskim profilima bolesnih i zdravih pojedinaca te je korištena za identifikaciju novih hipermetiliranih regija u genomu. Na primjer, Ishiura i suradnici (45) su otkrili da je nova CGG ekspanzija povezana s neuralnom intranuklearnom inkluzijskom bolešću hipermetilirana u usporedbi s neekspandiranom regijom. Također, Suzuki i suradnici (46) su identificirali nove metilirane elemente jezgre koji prethodno nisu bili identificirani korištenjem disulfitnog sekvenciranja.)

Poznato je da se uzorak metilacije DNA obično mijenja u stanicama raka, što rezultira hipometilacijom na centromerama i onkogenima te hipermetilacijom regulatornih elemenata poput pojačivača i promotora. Ove promjene u konačnici dovode do aktivacije onkogenih i supresije tumor-supresorskih gena što pokreće proces tumorogeneze (42).

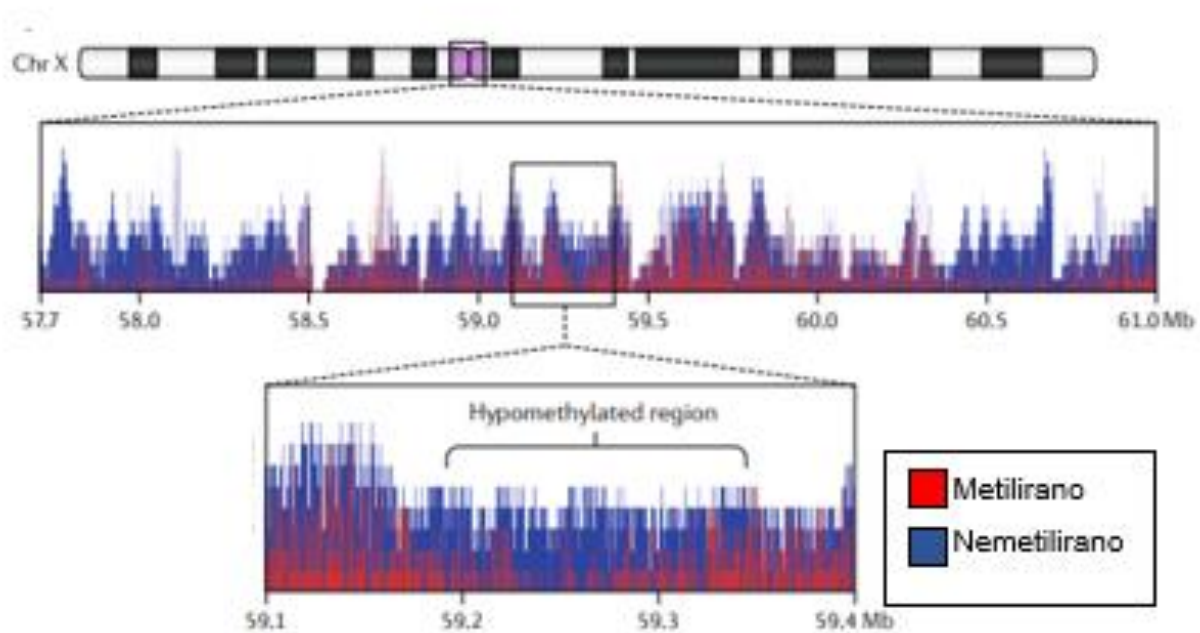


Slika 4: Preuzeto i preuređeno prema: Skvortsova K, Strizaker C, Taberlay P (2019) (42)
Usporedba metilacije zdravog tkiva i tkiva raka.

U studiji autora Y.Sakamoto, S.Zaha i suradnika (47), korištenjem LRS tehnologije detektirane su nemetilirane promotorske regije gena ERBB2 i PGR koje su korištene za klasifikaciju subtipova tumora dojke. Također, primijećene su različito metilirane regije u genima povezanim s rakom, kao što su CMYA5, TSLP, ZND503 i ZNF217. Ovi nalazi sugeriraju da metilacijski status ovih gena može biti povezan s procesom tumorogeneze.

LRS tehnologije omogućile su otkrivanje metilacijskog statusa u prethodno nedostupnim regijama genoma, kao što je centromera ili geni uključeni u tumorigenezu.

Upravo pronalazak dijela centromere koji nije metiliran upućuje na razlike u epigenetskoj regulaciji. Također, otkriće strukturnih varijanti koje su različito metilirane u stanicama raka dovode do uvida u nova epigenetska saznanja o strukturnim varijantama povezanih s rakom (48).



Slika 5: Preuzeto i prilagođeno prema: Logsdan G, Vollger M, Eichler E (2020) (6)

Slika prikazuje kromosom X. Korištenjem LRS tehnologije, pokazano je kako je većina centromere metilirana, osim jednog dijela koji pokazuje hipometilaciju.

Prednosti LRS-a nisu ograničene na proučavanje DNA, već je također koristan u proučavanju sekvenci RNA. Osobito je važan za otkrivanje raznolikosti i razine ekspresije izoformi RNA koje su rezultat alternativnog izrezivanja, što je često teško prikazati SRS metodama (49). Korištenje dugih očitavanja za analizu izoformi RNA ima jasne prednosti u odnosu na kratka očitavanja jer nema potrebe za mapiranjem kako bi se otkrile strukture tih izoformi. Umjesto toga, struktura se može uočiti samim pogledom na kompletnu sekvencu (50).

LRS tehnologije jedine su metode sekvenciranja koje izravno mogu sekvencirati molekulu RNA. Konvencionalno sekvenciranje RNA SRS tehnologijama, poznato kao "RNA-seq", zahtijeva reverznu transkripciju RNA u cDNA koja se potom umnaža PCR-om. Direktnim sekvenciranjem RNA LRS tehnologijama izravno se identificiraju ribonukleotidi, bez potrebe za reverznom transkripcijom ili PCR-om, što eliminira pogreške vezane za amplifikaciju. Ovaj jednostavniji pristup omogućava proučavanje složenih i novih izoformi transkripta, razlikovanje njihovih haplotipova te identifikaciju dužine poli-A repa (51). Također, još jedna prednost LRS-a u odnosu na SRS kod sekvenciranja RNA je detekcija modifikacija nukleotida. Najčešća modifikacija RNA koja je otkrivena je N6- metiladenozin koja je uključena u metabolizam i regulaciju RNA (52).

Nadopunjavanje sekvenciranja DNA i RNA može dovesti do otkrića novih uzročnih varijanti (43).

TEHNOLOGIJA SEKVENCIJANJE	PREDNOSTI	NEDOSTATCI
SRS- cDNA	<ul style="list-style-type: none"> • Visoka propusnost: mogućnost 100-1000x više čitanja u jednom nizu • Dobro proučene pogreške • Dostupnost velikog broja kompatibilnih metoda i računalnih programa za analizu • Mogućnost analize degradirane RNA 	<ul style="list-style-type: none"> • Priprema uzorka zahtijeva reverznu transkripciju i PCR amplifikaciju što dovodi do varijabilnih rezultata • Detekcija izoformi je ograničena • Metode otkrivanja transkripta zahtijevaju složeni de novo transkript
LRS- RNA	<ul style="list-style-type: none"> • Duga očitavanja identificiraju cijele transkripte • Pojednostavljene računalne metode za analizu de novo transkriptoma • Priprema uzorka ne zahtijeva reverznu transkripciju niti PCR umnažanje • Detekcija modifikacija baza RNA • Detekcija dužine poli-A repa iz jedne molekule RNA 	<ul style="list-style-type: none"> • Niska propusnost: 500000-1000000 čitanja u jednom nizu • Priprema uzorka i greške u sekvenciranju nisu dobro proučene • Ne preporučuje se analiza degradirane RNA

Tablica 3: preuzeto i prilagođeno prema: Stark R, Grzelak M, Hadfield J (2019) (53)

U tablici su prikazane glavne prednosti i nedostatci korištenja SRS i LRS tehnologija u sekvenciranju RNA.

Haplotip se definira kao slijed međusobno povezanih alela na kromosomu koji se zajedno nasljeđuju od jednog roditelja. Rekonstrukcija pojedinačnih haplotipova ima važnu primjenu u interpretaciji rizika od nastanka bolesti. Budući da SRS proizvodi kratka očitavanja koja pokrivaju samo male regije genoma, omogućuje samo nepotpunu rekonstrukciju haplotipa. S druge strane, LRS, zbog dugih očitavanja, omogućuje izravnu rekonstrukciju haplotipa (54).

Informacije o haplotipu su važne za medicinsku i populacijsku genetiku jer pojedinačni haplotipovi mogu pomoći u otkrivanju iznimno složenih varijanti, poput ugniježđenih strukturnih varijanti, inverzija i druga složenih preuređenja, te pristupu širokom spektru rijetkih naslijeđenih varijanti i de novo mutacija. Na primjer, informacije o haplotipu su korisne za otkrivanje rijetkih slučajeva sindroma keratitis-ichtioza-gluhoća. Osim toga, fenomen kombinirane heterozigotnosti na homolognim kromosomima odgovoran je za recesivne Mendelove poremećaje. Haplotipovi na razini kromosoma također imaju funkcionalnu važnost. Raspodjela varijanti u cis i trans položaju, odnosno faza varijanti, može utjecati na izražavanje gena. Prisutnost obje varijante na istom alelu (cis) dovodi do specifičnog fenotipa, dok prisutnost tih varijanti na odvojenim alelima (trans) ne dovode do istog. Još jedan vrlo važan primjer haplotipizacije na razini kromosoma je razumijevanje konteksta aneuploidije u genomima karcinoma (55).

LRS tehnologija može olakšati brzo i izravno haplotipiziranje zbog proizvodnje dugih sekvenci očitavanja i izravne analize DNA molekule u stvarnom vremenu. Međutim, LRS ima veću stopu pogrešaka od drugih SRS metoda, zbog čega su potrebna bioinformatička prilagođavanja (56).

4. NEDOSTATCI LONG-READ SEQUENCING TEHNOLOGIJA

Iako LRS (Long-Read Sequencing) tehnologije imaju mnogo istraživačkih i kliničkih primjena, postoje određena ograničenja koja je potrebno prevladati kako bi se potpuno iskoristio njihov potencijal. Jedan od glavnih nedostataka LRS tehnologija je visoka stopa pogrešaka tijekom sekvenciranja. Ovo je osobito istaknuto u točnoj identifikaciji točkastih mutacija i jednonukleotidnih polimorfizama, za koje se trenutno preporučuje korištenje SRS (Short-Read Sequencing) tehnologija (57). SRS tehnologije pružaju veću preciznost u analizi ovih genetskih varijacija. Visoka stopa pogrešaka u LRS tehnologijama može se pripisati slaboj osjetljivosti nanopora i nemogućnosti precizne kontrole prolaska DNA kroz njih. Unatoč stalnom napretku u poboljšanju LRS tehnologija tijekom proteklih godina, još uvijek nisu postigle razinu točnosti koju imaju SRS tehnologije (58).

Osim visoke stope pogrešaka, još jedan nedostatak LRS tehnologija je njihova visoka cijena sekvenciranja. Trošak korištenja LRS tehnologija može biti značajno veći u usporedbi sa SRS tehnologijama, što otežava širu primjenu ovih tehnologija, osobito u laboratorijskim postavkama s ograničenim proračunom. Visoka cijena sekvenciranja predstavlja jedan od razloga zašto je usvajanje LRS tehnologija relativno sporije, unatoč činjenici da su se pojavile prije više od jednog desetljeća (59).

Da bi se LRS tehnologije proširile i postale dominantna metoda sekvenciranja, bitno je poboljšati njihovu točnost i smanjiti trošak. Kad bi LRS tehnologije postigle visoku razinu točnosti koju trenutno imaju SRS tehnologije, laboratoriji ne bi trebali koristiti druge tehnologije sekvenciranja, poput SRS-a ili Sangerovog sekvenciranja. Kombinacija prednosti LRS tehnologija, kao što su veća dužina očitavanja i sposobnost otkrivanja strukturnih varijanti, uz povećanu točnost i smanjenu cijenu, mogla bi dovesti do šire primjene LRS tehnologija kao jedine metode sekvenciranja u budućnosti (60).

PREDNOSTI	NEDOSTATCI
Detekcija strukturnih varijanti	Točnost po očitaju manja od SRS
Detekcija složenih preostrojanja	Cijena viša od SRS
Detekcija tandemskih ponavljanja	Manja propusnost uzorka nego u SRS
Sekvenciranje gena s pseudogenima	Bioinformatički pristup manje je razvijen nego za SRS
Detekcija modifikacija baza- metilacija	DNA treba biti duga što je izazov za standardne laboratorijske metode izolacije DNA
Segregacija	
Sekvenciranje transkripta RNA za određivanje alternativnog izrezivanja	
Primjene u kojima je potreban prijenosni uređaj	

Tablica 4: Preuzeto i preuređeno prema: Nurk S, Koren S, Phillipy A i sur. (2022) (61)

Sažeti prikaz prednosti i nedostataka LRS tehnologija.

5. ZAKLJUČAK

Long-read sequencing (LRS) je revolucionarna tehnologija koja ima značajan doprinos u razumijevanju genoma i identifikaciji genetskih uzroka bolesti. Jedan od najznačajnijih doprinosa je mogućnost detaljnijeg razumijevanja genoma, posebno složenih dijelova koji sadrže ponavljajuće sekvence. Ove sekvence često predstavljaju izazov za tradicionalne tehnike sekvenciranja kratkih očitavanja, ali LRS omogućava rekonstrukciju tih dijelova s visokom točnošću. Još jedan značajan doprinos LRS-a je identifikacija strukturnih varijanti u genomu. Ove varijante, poput inverzija, delecija, translokacija i duplikacija, igraju važnu ulogu u razvoju bolesti, uključujući razvoj raka. Tradicionalne metode sekvenciranja kratkih očitavanja često su ograničene u otkrivanju i preciznom karakteriziranju ovih strukturnih varijanti, dok LRS omogućava njihovo detaljno mapiranje i razumijevanje.

Međutim, LRS ima i izazove s kojima se susreće. Bitno je istaknuti nižu stopu točnosti očitavanja u usporedbi sa SRS-om. Niža točnost može uzrokovati greške u rekonstrukciji genoma i tumačenju rezultata. Također, analiza velike količine podataka generiranih LRS-om zahtijeva napredne računalne algoritme i resurse, što predstavlja izazov za istraživače.

Unatoč tim izazovima, očekuje se sve šira upotreba u budućnosti. Tehnološki napredak i kontinuirano poboljšanje analitičkih metoda omogućit će povećanje preciznosti i učinkovitosti ove tehnologije. Očekuje se da će LRS postati sveprisutna tehnika u istraživanjima genoma i kliničkoj praksi. Njegova primjena proširit će se na područja poput personalizirane medicine, gdje će se koristiti za pružanje preciznije dijagnostike i terapije temeljene na genetskim karakteristikama pojedinca.

U konačnici, LRS je tehnologija koja donosi važne doprinose u razumijevanju genoma, identifikaciji genetskih uzroka bolesti i napretku personalizirane medicine te otvara nove perspektive za unapređenje zdravstvene skrbi u budućnosti.

6. ZAHVALE

Zahvaljujem svom mentoru prof.dr.sc. Mariu Ćuku, dr. med što mi je ukazao povjerenje i prepoznao moj interes i želju te svojim znanjem i iskustvom pomogao da ovim radom uspješno završim svoje studiranje.

Hvala mojim roditeljima i cijeloj obitelji što su bili uz mene kad mi je bilo najteže, pružali mi neizmjernu podršku i ljubav tijekom cijelog školovanja.

Ujedno zahvaljujem i svim profesorima što su svoje znanje nesebično prenosili svima nama.

Ovim putem želim zahvaliti i svojim prijateljima i kolegama na kolegijalnosti, razmjeni iskustava i druženjima.

Hvala dragom Bogu što me usmjerio na ovaj humani poziv i dao mi snage da ostvarim svoj životni cilj!

LITERATURA

1. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol.* 2021 Nov;82(11):801-11. doi: 10.1016/j.humimm.2021.02.012.
2. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif).* 2013;6:287-303. doi: 10.1146/annurev-anchem-062012-092628.
3. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet.* 2018 Sep;34(9):666-81. doi: 10.1016/j.tig.2018.05.008.
4. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics.* 2015 Oct;13(5):278-89. doi: 10.1016/j.gpb.2015.08.002.
5. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. *Hum Mol Genet.* 2018 Aug 1;27(R2):R234-41. doi: 10.1093/hmg/ddy177.
6. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet.* 2020 Oct;21(10):597-614. doi: 10.1038/s41576-020-0236-x.
7. Jain M, Koren S, Miga KH. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* 2018 Apr;36(4):338-45. doi: 10.1038/nbt.4060.
8. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021 Nov;39(11):1348-65. doi: 10.1038/s41587-021-01108-x.
9. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed.* 2013 Dec;98(6):236-8. doi: 10.1136/archdischild-2013-304340.
10. Petersen BS, Fredrich B, Hoepfner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet.* 2017 Dec;18(1):14.
11. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020 Feb 7;21(1):30. doi: 10.1186/s13059-020-1935-5.
12. Cretu Stancu M, van Roosmalen MJ, Renkens I, Nieboer MM, Middelkamp S, de Ligt J, i sur. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun.* 2017 Nov 6;8(1):1326. doi: 10.1038/s41467-017-01343-4.
13. Sanford Kobayashi E, Batalov S, Wenger AM, Lambert C, Dhillon H, Hall RJ, i sur. Approaches to long-read sequencing in a clinical setting to improve diagnostic rate. *Sci Rep.* 2022 Oct 9;12(1):16945. doi: 10.1038/s41598-022-20113-x.
14. Midha MK, Wu M, Chiu KP. Long-read sequencing in deciphering human genetics to a greater depth. *Hum Genet.* 2019 Dec;138(11-12):1201-15. doi: 10.1007/s00439-019-02064-y.

15. Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, i sur. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol.* 2010;11(5):R52. doi: 10.1186/gb-2010-11-5-r52.
16. Begum G, Albanna A, Bankapur A, Nassir N, Tambi R, Berdiev BK, i sur. Long-Read Sequencing Improves the Detection of Structural Variations Impacting Complex Non-Coding Elements of the Genome. *Int J Mol Sci.* 2021 Feb 19;22(4):2060. doi: 10.3390/ijms22042060.
17. Liaci C, Prandi L, Pavinato L, Brusco A, Maldotti M, Molineris I, i sur. The Emerging Roles of Long Non-Coding RNAs in Intellectual Disability and Related Neurodevelopmental Disorders. *Int J Mol Sci.* 2022 May 30;23(11):6118. doi: 10.3390/ijms23116118.
18. Mitsuhashi S, Matsumoto N. Long-read sequencing for rare human genetic diseases. *J Hum Genet.* 2020 Jan;65(1):11-9. doi: 10.1038/s10038-019-0671-8.
19. Dutta UR, Rao SN, Pidugu VK, V S V, Bhattacharjee A, Bhowmik AD, i sur. Breakpoint mapping of a novel de novo translocation t(X;20)(q11.1;p13) by positional cloning and long read sequencing. *Genomics.* 2019 Sep;111(5):1108-14. doi: 10.1016/j.ygeno.2018.07.005.
20. Mizuguchi T, Suzuki T, Abe C, Umemura A, Tokunaga K, Kawai Y, i sur. A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J Hum Genet.* 2019 May;64(5):359-68. doi: 10.1038/s10038-019-0569-5.
21. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, i sur. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med.* 2018 Jan;20(1):159-63. doi: 10.1038/gim.2017.86.
22. Reiner J, Pisani L, Qiao W, Singh R, Yang Y, Shi L, i sur. Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a Bardet-Biedl Syndrome 9 (*BBS9*) deletion. *NPJ Genom Med.* 2018 Jan 22;3:3. doi: 10.1038/s41525-017-0042-3.
23. Miao H, Zhou J, Yang Q, Liang F, Wang D, Ma N, i sur. Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis. *Hereditas.* 2018 Sep 28;155:32. doi: 10.1186/s41065-018-0069-1.
24. Sakamoto, Y., Sereewattanawoot, S. & Suzuki, A. A new era of long-read sequencing for cancer genomics. *J Hum Genet* 65, 3–10 (2020). <https://doi.org/10.1038/s10038-019-0658-5>
25. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet.* 2019 May 7;10:426. doi: 10.3389/fgene.2019.00426.
26. Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet.* 2021 May 6;108(5):764-85. doi: 10.1016/j.ajhg.2021.03.011.
27. Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK, i sur. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv.* 2022 Mar 4;8(9):eabm5386. doi: 10.1126/sciadv.abm5386.

28. Chintalaphani SR, Pineda SS, Deveson IW, Kumar KR. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol Commun.* 2021 May 25;9(1):98. doi: 10.1186/s40478-021-01201-x.
29. Ardui S, Race V, de Ravel T, Van Esch H, Devriendt K, Matthijs G, i sur. Detecting AGG Interruptions in Females With a FMR1 Premutation by Long-Read Single-Molecule Sequencing: A 1 Year Clinical Experience. *Front Genet.* 2018 May 16;9:150. doi: 10.3389/fgene.2018.00150.
30. Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, i sur. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet.* 2018 Apr;50(4):581-90. doi: 10.1038/s41588-018-0067-2.
31. Mitsuhashi S, Frith MC, Mizuguchi T, Miyatake S, Toyota T, Adachi H, i sur. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol.* 2019 Mar 19;20(1):58. doi: 10.1186/s13059-019-1667-6.
32. Su Y, Fan L, Shi C, Wang T, Zheng H, Luo H, i sur. Deciphering Neurodegenerative Diseases Using Long-Read Sequencing. *Neurology.* 2021 Aug 31;97(9):423-33. doi: 10.1212/WNL.00000000000012466.
33. Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, i sur. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat Genet.* 2019 Aug;51(8):1215- 21. doi: 10.1038/s41588-019-0459-y.
34. Deng J, Yu J, Li P, Luan X, Cao L, Zhao, i sur. Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. *Am J Hum Genet.* 2020 Jun 4;106(6):793-804. doi: 10.1016/j.ajhg.2020.04.011. Epub 2020 May 14.
35. McFarland KN, Liu J, Landrian I, Godiska R, Shanker S, Yu F, i sur. SMRT Sequencing of Long Tandem Nucleotide Repeats in SCA10 Reveals Unique Insight of Repeat Expansion Structure. *PLoS One.* 2015 Aug 21;10(8):e0135906. doi: 10.1371/journal.pone.0135906.
36. Schüle B, McFarland KN, Lee K, Tsai YC, Nguyen KD, Sun C, i sur. Parkinson's disease associated with pure *ATXN10* repeat expansion. *NPJ Parkinsons Dis.* 2017 Sep 5;3:27. doi: 10.1038/s41531-017-0029-x.
37. Höijer I, Tsai YC, Clark TA, Kotturi P, Dahl N, Stattin EL, i sur. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. *Hum Mutat.* 2018 Sep;39(9):1262-72. doi: 10.1002/humu.23580.
38. Zeng S, Zhang MY, Wang XJ, Hu ZM, Li JC, Li N, i sur. Long-read sequencing identified intronic repeat expansions in *SAMD12* from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J Med Genet.* 2019 Apr;56(4):265-270. doi: 10.1136/jmedgenet-2018-105484.
39. Corbett MA, Kroes T, Veneziano L, Bennett MF, Florian R, Schneider AL, i sur. Intronic ATTTTC repeat expansions in *STARD7* in familial adult myoclonic epilepsy linked to chromosome 2. *Nat Commun.* 2019 Oct 29;10(1):4920. doi: 10.1038/s41467-019-12671-y.

40. Florian RT, Kraft F, Leitão E, Kaya S, Klebe S, Magnin E, i sur. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. *Nat Commun.* 2019 Oct 29;10(1):4919. doi: 10.1038/s41467-019-12763-9.
41. Ebbert MTW, Farrugia SL, Sens JP, Jansen-West K, Gendron TF, Prudencio M, i sur. Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol Neurodegener.* 2018 Aug 21;13(1):46. doi: 10.1186/s13024-018-0274-4.
42. Skvortsova K, Stirzaker C, Taberlay P. The DNA methylation landscape in cancer. *Essays Biochem.* 2019 Dec 20;63(6):797-811. doi: 10.1042/EBC20190037.
43. Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome Med.* 2022 Feb 28;14(1):23. doi: 10.1186/s13073-022-01026-w.
44. Lucas, M.C., Novoa, E.M. Long-read sequencing in the era of epigenomics and epitranscriptomics. *Nat Methods* 20, 25–9 (2023). <https://doi.org/10.1038/s41592-022-01724-8>
45. Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, i sur. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat Genet.* 2019 Aug;51(8):1222-32. doi: 10.1038/s41588-019-0458-z.
46. Suzuki Y, Korfach J, Turner SW, Tsukahara T, Taniguchi J, Qu W, i sur. AgIn: measuring the landscape of CpG methylation of individual repetitive elements. *Bioinformatics.* 2016 Oct 1;32(19):2911-9. doi: 10.1093/bioinformatics/btw360.
47. Sakamoto Y, Zaha S, Suzuki Y, Seki M, Suzuki A. Application of long-read sequencing to the detection of structural variants in human cancer genomes. *Comput Struct Biotechnol J.* 2021 Jul 28;19:4207-16. doi: 10.1016/j.csbj.2021.07.030.
48. Lee I, Razaghi R, Gilpatrick T, Molnar M, Gershman A, Sadowski N, i sur. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat Methods.* 2020 Dec;17(12):1191-9. doi: 10.1038/s41592-020-01000-7.
49. Ergin S, Kherad N, Alagoz M. RNA sequencing and its applications in cancer and rare diseases. *Mol Biol Rep.* 2022 Mar;49(3):2325-33. doi: 10.1007/s11033-021-06963-0.
50. Bayega A, Wang YC, Oikonomopoulos S, Djambazian H, Fahiminiya S, Ragoussis J. Transcript Profiling Using Long-Read Sequencing Technologies. *Methods Mol Biol.* 2018;1783:121-47. doi: 10.1007/978-1-4939-7834-2_6.
51. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ.* 2019 Jun;61(5):316-26. doi: 10.1111/dgd.12608.
52. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics.* 2015 Feb;13(1):4-16. doi: 10.1016/j.gpb.2015.01.009.

53. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet.* 2019 Nov;20(11):631-56. doi: 10.1038/s41576-019-0150-2.
54. Snyder MW, Adey A, Kitzman JO, Shendure J. Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet.* 2015 Jun;16(6):344-58. doi: 10.1038/nrg3903.
55. Garg S. Computational methods for chromosome-scale haplotype reconstruction. *Genome Biol.* 2021 Apr 12;22(1):101. doi: 10.1186/s13059-021-02328-9.
56. Maestri S, Maturo MG, Cosentino E, Marcolungo L, Iadarola B, Fortunati E, et al. A Long-Read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings. *Int J Mol Sci.* 2020 Dec 1;21(23):9177. doi: 10.3390/ijms21239177.
57. Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. *Life (Basel).* 2021 Dec 26;12(1):30. doi: 10.3390/life12010030.
58. Adewale BA. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr J Lab Med.* 2020 Nov 26;9(1):1340. doi: 10.4102/ajlm.v9i1.1340.
59. Conlin LK, Aref-Eshghi E, McEldrew DA, Luo M, Rajagopalan R. Long-read sequencing for molecular diagnostics in constitutional genetic disorders. *Hum Mutat.* 2022 Nov;43(11):1531-44. doi: 10.1002/humu.24465.
60. Marx V. Method of the year: long-read sequencing. *Nat Methods.* 2023 Jan;20(1):6-11. doi: 10.1038/s41592-022-01730-w.
61. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science.* 2022 Apr;376(6588):44-53. doi: 10.1126/science.abj6987.