

VILNIAUS UNIVERSITETAS

Žana

KAPUSTINA

Modifikuotų nukleotidų taikymas plataus masto nukleorūgščių analizei

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,
Biologija (N 010)

VILNIUS 2021

Disertacija rengta 2016 – 2020 metais Vilniaus universiteto Gyvybės mokslų centro Biomokslų institute ir UAB „Thermo Fisher Scientific Baltics“ Mokslinių tyrimų ir eksperimentinės plėtros centre.

Mokslinius tyrimus rėmė UAB „Thermo Fisher Scientific Baltics“ ir Vilniaus universitetas.

Mokslinis vadovas:

prof. dr. Arvydas Lubys (Vilniaus universitetas, gamtos mokslai, biologija, N 010).

Gynimo taryba:

Pirmininkas – **prof. habil. dr. Juozas Rimantas Lazutka** (Vilniaus universitetas, gamtos mokslai, biologija, N 010).

Nariai:

prof. dr. Juozas Kupčinskas (Lietuvos sveikatos mokslų universitetas, gamtos mokslai, biologija, N 010).

prof. dr. Linas Mažutis (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).

prof. dr. Rolandas Meškys (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).

prof. dr. Artūras Petronis (Toronto universitetas, gamtos mokslai, biologija, N 010).

Disertacija ginama viešame Gynimo tarybos posėdyje 2021 m. lapkričio mėn. 22 d. 15 val. Gyvybės mokslų centro R-401 auditorijoje. Adresas: Saulėtekio al. 7, Vilnius, Lietuva, tel. +37052234420; el. paštas: info@gmc.vu.lt

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: <https://www.vu.lt/naujienos/ivykiu-kalendarius>

VILNIUS UNIVERSITY

Žana

KAPUSTINA

The utility of modified nucleotides for high-throughput nucleic acid analysis

SUMMARY OF DOCTORAL DISSERTATION

Natural Sciences,
Biology (N 010)

VILNIUS 2021

This dissertation was prepared between 2016 and 2020 at the Institute of Biosciences, Life Sciences Center, Vilnius University and the Research and Development department at UAB Thermo Fisher Scientific Baltics. The research was supported by UAB Thermo Fisher Scientific Baltics and the scholarship from Vilnius University.

Academic supervisor:

Prof. Dr. Arvydas Lubys (Vilnius University, Natural Sciences, Biology, N 010).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Juozas Rimantas Lazutka (Vilnius University, Natural Sciences, Biology, N 010).

Members:

Prof. Dr. Juozas Kupčinskas (Lithuanian University of Health Sciences, Natural Sciences, Biology, N 010).

Prof. Dr. Linas Mažutis (Vilnius University, Natural Sciences, Biochemistry, N 004).

Prof. Dr. Rolandas Meškys (Vilnius University, Natural Sciences, Biochemistry, N 004).

Prof. Dr. Artūras Petronis (University of Toronto, Natural Sciences, Biology, N 010).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 3 p.m. on 22nd of November 2021 in Room R-401 of the Vilnius University Life Sciences Centre.

Address: Saulėtekio av. 7, Vilnius, LT-10257, Lithuania

Tel. +37052234420; e-mail: info@gmc.vu.lt

The text of this dissertation can be accessed at Vilnius University library, as well as on the website of Vilnius University: <https://www.vu.lt/naujienos/ivykiu-kalendorius>

TURINYS

| | |
|---|----|
| SANTRUMPOS | 7 |
| ĮVADAS | 9 |
| 1. LITERATŪROS APŽVALGA | 14 |
| 1.1 Funkcionalizuotos nukleorūgštys | 15 |
| 1.2 Fermentinis modifikuotų nukleotidų įterpimas | 18 |
| 1.3 Funkcionalizuotų nukleorūgščių replikacija | 18 |
| 1.4 Didelio našumo nukleorūgščių analizė..... | 20 |
| 1.4.1 DNR sekoskaita..... | 20 |
| 1.4.2 RNR sekoskaita..... | 22 |
| 2. MEDŽIAGOS IR METODAI | 24 |
| 2.1 Oligonukleotidu modifikuotų dideoksinukleotidų sintezė | 24 |
| 2.2 Įterpimo ir replikacijos tyrimai..... | 26 |
| 2.3 Fragmentų bibliotekų paruošimas taikant <i>in vitro</i> transkripciją 28 | |
| 2.4 Pusiau taikininė 16S rRNR geno sekoskaita | 29 |
| 2.5 Genų raiškos analizė..... | 30 |
| 2.6 NKS bibliotekų kokybės kontrolė ir sekoskaita | 31 |
| 2.7 Duomenų analizė..... | 32 |
| 3. PAGRINDINIAI REZULTATAI | 33 |
| 3.1 ddN ^{ON} TP panaudojimas fermentinėse reakcijose..... | 33 |
| 3.1.1 ddN ^{ON} TP kaip fermentų substratas..... | 33 |
| 3.1.2 ddN ^{ON} TP žymėtų nukleorūgščių replikacija | 36 |
| 3.1.3 Linijinio amplifikavimo produktų žymėjimas..... | 37 |
| 3.2 Pusiau taikininė 16S rRNR geno sekoskaita | 39 |

| | | |
|-----|-------------------------------|----|
| 3.3 | Genų raiškos analizė..... | 44 |
| | IŠVADOS..... | 48 |
| | PUBLIKACIJŲ SĄRAŠAS | 50 |
| | <i>CURRICULUM VITAE</i> | 52 |
| | LITERATŪROS SĄRAŠAS..... | 53 |

SANTRUMPOS

| | |
|----------------------|---|
| ACN | acetonitrilas |
| AMV | paukščių mieloblastozės virusas (<i>angl.</i> avian myeloblastosis virus) |
| AT | atvirkštinė transkripcija |
| AUC | plotas po kreive (<i>angl.</i> area under the curve) |
| CuAAC | variu katalizuojama azido-alkino grupių ciklizacija |
| ddN ^{N3} TP | 5- arba 7-deaza-7-(3-(2-azidoacetamido)prop-1-inil)-2', 3'-dideoksinukleozidų 5'-trifosfatai |
| ddN ^{ON} TP | oligonukleotidais modifikuoti 2',3'-dideoksinukleozidų 5'-trifosfatai |
| dgDNR | dvigrandinė DNR |
| DMSO | dimetilsulfoksidas |
| dNTP | 2'-deoksinukleozidų 5'-trifosfatai |
| DTT | ditiotreitolis |
| EDTA | etilendiamintetraacto rūgštis |
| ERCC | išorinių RNR kontrolių konsorciumas (<i>angl.</i> External RNA Controls Consortium) |
| exo- | 3'-5' egzonukleazinio aktyvumo trūkumas |
| HPLC | efektyvioji skysčių chromatografija (<i>angl.</i> high-performance liquid chromatography) |
| iRNR | informacinė RNR |
| IVT | <i>in vitro</i> transkripcija |
| kDNR | kopijinė DNR |
| kPGR | kiekybinė PGR |
| LC-MS | skysčių chromatografija – masių spektrometrija (<i>angl.</i> liquid chromatography – mass spectrometry) |
| M-MLV | Moloney pelių leukemijos virusas (<i>angl.</i> Moloney murine leukemia virus) |
| MTAS-seq | iRNR sekoskaita taikant ddN ^{ON} TP (<i>angl.</i> mRNA sequencing by terminator-assisted synthesis) |
| NKS | naujos kartos sekoskaita |
| NTP | ribonukleozidų 5'-trifosfatai |

| | |
|-----------|--|
| ON | oligonukleotidas |
| PAAG | poliakrilamidinis gelis |
| PBS | fosfatinis buferis (<i>angl.</i> phosphate-buffered saline) |
| PEX | pradmens pratęsimas (<i>angl.</i> primer extension) |
| PGR | polimerazės grandininė reakcija |
| PUP | poliU polimerazė |
| RNRP | RNR polimerazė |
| rRNR | ribosominė RNR |
| SELEX | sisteminė ligandų evoliucija taikant eksponentinį praturtinimą (<i>angl.</i> systematic evolution of ligands by exponential enrichment) |
| st16S-seq | pusiau taikininė 16S rRNR geno sekoskaita (<i>angl.</i> semi-targeted 16S rRNA gene sequencing) |
| TBE | tris-boratinis EDTA buferis |
| TdT | terminalinė deoksinukleotidil transferazė |
| TEAAc | trietilamonio acetatas |
| THPTA | tris-hidroksipropiltriazolilmetilaminas |
| UHRR | universali žmogaus referentinė RNR (<i>angl.</i> universal human reference RNA) |
| UTR | netransliuojamas regionas (<i>angl.</i> untranslated region) |
| vgDNR | viengrandinė DNR |
| ŽIV | žmogaus imunodeficito virusas |

ĮVADAS

Natūralios nukleotidų modifikacijos, tokios kaip metil, hidroksimetil, formil, karboksi grupės, atlieka svarbų vaidmenį epigenetinėje reguliacijoje (Bilyard *et al.*, 2020). Sintetiniai modifikuoti nukleotidai reikšmingai praplėtė nukleorūgščių pritaikomumo spektrą tokiose srityse kaip terapija, bioanalizė, cheminė biologija, katalizė, biojutikliai, ir kitose (Dhuri *et al.*, 2020; Xu *et al.*, 2017; Lapa *et al.*, 2016; Hollenstein, 2015; Hollenstein *et al.*, 2008). Nukleotidų bazių modifikacijos dažniausiai įvedamos į pirimidinų C5 padėtį arba 7-deazapurinų C7 padėtį (Jäger *et al.*, 2005). Sintetinės modifikacijos gali varijuoti nuo mažų funkcinių grupių iki didelių konjuguotų junginių, tokių kaip baltymai. Gamtinės DNR ir RNR polimerazės pasižymi stebėtinu substrato atpažinimo plastiškumu ir neretai geba naudoti sintetinius nukleotidų analogus kaip substratus. Tai atveria galimybes fermentinei funkcionalizuotų nukleorūgščių sintezei.

Praktinis modifikuotų nukleorūgščių pritaikomumas daugeliu atvejų remiasi jų suderinamumu su fermentinėmis reakcijomis. Nors žinoma, kad fosfodiesteriniai, amidiniai ir triazoliniai DNR karkasai funkcionalūs *in vitro* ir net *in vivo* (Ciafrè *et al.*, 1995; Kuwahara *et al.*, 2009; El-Sagheer *et al.*, 2011; Birts *et al.*, 2014), modifikuotų nukleorūgščių replikacija tiriama sąlyginai mažiau nei nukleotidų analogų įterpimas. Tyrimai rodo, kad fosfatinės grupės pačios savaime nėra būtinos sintetinių matricų kopijavimui, kas leido pritaikyti CuAAC „click“ reakciją fermentinių sistemų toleruojamų modifikuotų nukleorūgščių gamybai. Aukšto tikslumo nenatūralių nukleorūgščių karkasų replikacija įgalintų tirti funkcionalizuotų nukleorūgščių sekas. Pastebėta, jog DNR polimerazės, pasižyminčios klaidų taisymo aktyvumu, sulėtėja ties modifikacijos vieta matricoje, kas veda prie delecijų įvedimo, tuo tarpu fermentai be egzonukleazinio aktyvumo turi mažesnę tikimybę klaidingai prašokti sintetinės modifikacijos sritį (Shivalingam *et al.*, 2017). Modifikuotos nukleorūgštys kartu su suderinamais fermentais nukleotidų analogų

įterpimui ir modifikuotų matricų replikacijai sudarytų patrauklų įrankių komplektą naujų molekulinės biologijos metodų vystymui.

Siekdami gerinti mėginių paruošimo naujos kartos sekoskaitai (NKS) procesus bei kurti naujus taikymus mokslininkai žvalgosi į chemofermentinius metodus. „Click“ reakcija, dar vadinama cheminiu ligavimu, leidžia specifiškai prijungti alkinio grupę modifikuotą sekoskaitos adapterį prie azido grupę funkcionalizuotų terminatorių, fermentiniu būdu įterptų į sekoskaitai ruošiamas molekules (Miura *et al.*, 2018; Routh *et al.*, 2015). Šis principas turi pranašumų: tinka tiek DNR, tiek kDNR fragmentų bibliotekų paruošimui, pasižymi žemesniu nepageidaujamos rekombinacijos dažniu, taip pat suteikia galimybę ruošti NKS bibliotekas iš viengrandinės DNR. Vis dėlto, vario jonų sąlygota DNR degradacija, žemas konversijos efektyvumas ir daugiastadijiniai protokolai mažina chemofermentinių metodų patrauklumą. Paprasti ir efektyvūs nukleorūgščių žymėjimo sintetiniais oligonukleotidais metodai, nepasižymintys dabartinių chemofermentinių technologijų trūkumais, tačiau išsaugoję jų privalumus, galėtų tapti patraukliu įrankiu nukleorūgščių analizei.

Tikslas ir uždaviniai

Šio darbo **tikslas** – ištirti oligonukleotidais modifikuotų 2',3'-dideoksiribonukleozidų 5'-trifosfatų (ddN^{ON}TP) savybes ir pritaikomumą plataus masto nukleorūgščių analizei, ypatingą dėmesį skiriant DNR ir kDNR molekulių žymėjimui naujos kartos sekoskaita paremtiems taikymams. Tikslui pasiekti iškelti šie **uždaviniai**:

1. Ištirti ddN^{ON}TP, kaip substrato, savybes ir nustatyti fermentus, tinkamus DNR ir RNR žymėjimui.
2. Ištirti fermentų inžinerijos galimybes, siekiant sukurti naujus metodus nukleorūgščių, žymėtų ddN^{ON}TP, sintezei.

3. Iširti nukleorūgščių, žymėtų ddN^{ON}TP, matricines savybes ir nustatyti DNR polimerazes, gebančias vykdyti sintezę per nenatūralią triazolo žiedą turinčią jungtį.
4. Įvertinti ddN^{ON}TP pritaikomumą DNR ir RNR sekoskaitos taikymams, tiriant tiek visą genomą arba transkriptomą, tiek tikslius regionus.
5. Pritaikyti DNR žymėjimą ddN^{ON}TP mikrobiomo charakterizavimui.
6. Iširti kDNR žymėjimo ddN^{ON}TP potencialą genų raiškos analizėje.

Mokslinis naujumas ir praktinė reikšmė

Šiame darbe polimerizacijos metu įterpto nukleotido sudėtyje esanti oligonukleotidinė modifikacija buvo panaudota kaip universali pradmens hibridizacijos vieta komplementarios grandinės sintezei, taip sukuriant galimybę specifiskai padauginti šiuo modifikuotu nukleotidu pažymėtus DNR arba kDNR fragmentus. Svarbu ir tai, kad dideoksinukleotidų naudojimas įgalino atsitiktinę sintetinės DNR grandinės terminaciją suformuojant tinkamo sekoskaitai ilgio fragmentus.

Darbo metu parodyta, kad efektyvų fermentinį ddN^{ON}TP įterpimą vykdo Thermo Sequenase, CycleSeq, Sequenase V2.0 ir TdT polimerazės, o kopijuojant iRNR sintetinama kDNR gali būti efektyviai pažymėta naudojant Maxima, SuperScript IV, SuperScript II, RevertAid ir ŽIV atvirkštines transkriptazes. RNR žymėjimui tinka poliU polimerazė, tuo tarpu šiame darbe *in vitro* evoliucijos pagalba atrinktas T7 RNR polimerazės mutantas V783M, gebantis sintetinti iš dTTP, dCTP, dATP ir 2'-F-dGTP sudarytus vgDNR transkriptus, taip pat geba įjungti ddN^{ON}TP atsitiktinėse transkripto vietose. Tai leidžia pažymėti transkripcijos produktų 3' galus universalisia seka, o tuo pačiu leidžia linijiškai padauginti NKS tinkamas DNR sekas – tai yra ypač aktualu dirbant su itin mažais DNR kiekiais. Nustatyta, kad triazolo žiedą turinti netipinė jungtis tarp bazės ir prie jos prijungto

oligonukleotido (ddN^{ON}TP sudėtyje) komplementarios grandinės sintezės metu toleruojama Phusion exo-, Klenow fragmento exo-, Thermo Sequenase ir SuperScript IV DNR polimerazių, iš kurių Phusion exo- pasižymėjo geriausiu perskaitymo efektyvumu. Sėkmingas įterpimo ir netipinės jungties perskaitymo fermentų identifikavimas įgalino pritaikyti ddN^{ON}TP technologiją fragmentų bibliotekų paruošimui naujos kartos sekoskaitai. Be to, parodyta, kad oligonukleotidinė modifikacija gali turėti randomizuotos sekos regionus arba afininius ligandus, kurie vėliau tarnauja atitinkamai kaip molekuliniai barkodai arba afininiai žymenys patogiam tikslinių molekulių gryninimui.

Atsitiktinis ddN^{ON}TP įterpimo pobūdis ir tuo pat metu vykstantis sintetinės grandinės žymėjimas universalio seka tapo naujo sekoskaitos bibliotekų konstravimo metodo vystymo pagrindu. Parodyta, kad atsitiktinių pradmenų pratęsimo produktų žymėjimas leidžia paruošti visą genomą arba visą transkriptomą dengiančias NKS bibliotekas, o tuo tarpu specifinių pradmenų pratęsimas ir žymėjimas konstruojant pusiau taikines (*angl.* semi-targeted) bibliotekas leidžia tirti nežinomas sekos regionus šalia pasirinktų lokusų. Įrodyta, kad ši strategija yra naudinga mikrobiomo charakterizavimui – sukurtas naujas st16S-seq metodas, įgalinęs tirti *a priori* nežinomus genominius lokusus, esančius prieš 16S rRNR geną. Naujas metodas turi rimtų pranašumų prieš tradicinius: tikslus 16S rRNR geno kopijų nustatymas, tikslesnis bakteriųjų rūšių charakterizavimas ir mažesnė metodo priklausomybė nuo pradmenų dizaino, kadangi st16S-seq metodui reikia tik į vieną pusę orientuotų specifinių pradmenų.

RNR sekoskaitos srityje buvo sukurtas naujas genų raiškos analizės metodas MTAS-seq, kuris įgalino ženkliai supaprastinti mėginio paruošimo procesą ir užtikrino aukštos kokybės NKS duomenis dirbant tiek su išgryninta RNR, tiek ir ruošiant mėginius tiesiogiai eukariotinių ląstelių lizatuose. Be to, 3'UTR regionų praturtinimas leidžia analizuoti alternatyvaus poliadenilinimo profilius eukariotiniuose transkriptomuose. Įdomu tai, kad ddN^{ON}TP su modifikacija, atitinkančia pilno ilgio sekoskaitos adapterį, įgalino

sekvenuoti kDNR fragmentus be PGR amplifikacijos, kas anksčiau nebuvo įmanoma su standartinėmis NKS sistemomis.

Šiame darbe išvystytoms technologijoms pateikta tarptautinė patentinė paraiška.

Ginamieji teiginiai

1. ddN^{ON}TP yra substratai DNR ir RNR polimerazėms su natūraliu arba baltymų inžinerijos būdu įgytu gebėjimu inkorporuoti nukleotidų analogus.
2. Į DNR įjungtas ddN^{ON}TP tarnauja kaip matrica komplementarios grandinės sintezei. Egzonukleazinio aktyvumo neturinčios DNR polimerazės geba vykdyti sintezę per triazolo žiedą turinčią jungtį.
3. Pirmos grandinės sintezės pradmenų pratęsimas, ddN^{ON}TP įterpimas ir antros grandinės sintezė naudojant iniciacijai pradmenį, kuris yra komplementarus įjungtajam, įgalina DNR sekoskaitos taikymus. Matricos padengimas priklauso nuo pirmos grandinės sintezėje naudojamų pradmenų sekų.
4. Pusiau taikininė 16S rRNR geno sekoskaita yra naujo tipo mikrobiomo charakterizavimo strategija, įgalinanti nustatyti 16S rRNR geno kopijų skaičių ir gerinanti bakterijų klasifikacijos tikslumą rūšių lygmenyje.
5. Pradmenų pratęsimas ir ddN^{ON}TP įterpimas naudojant SuperScript IV atvirkštinę transkriptazę įgalina RNR sekoskaitos taikymus. Matricos padengimas priklauso nuo naudojamų pradmenų sekų.
6. Kopijinės DNR žymėjimas ddN^{ON}TP patobulina genų raiškos analizės metodiką, įgalindama vieno mėgintuvėlio protokolą, paprastą molekulinį barkodų įvedimą ir suderinamumą su tiesiogine atvirkštine transkripcija nuo eukariotinių ląstelių lizatų.

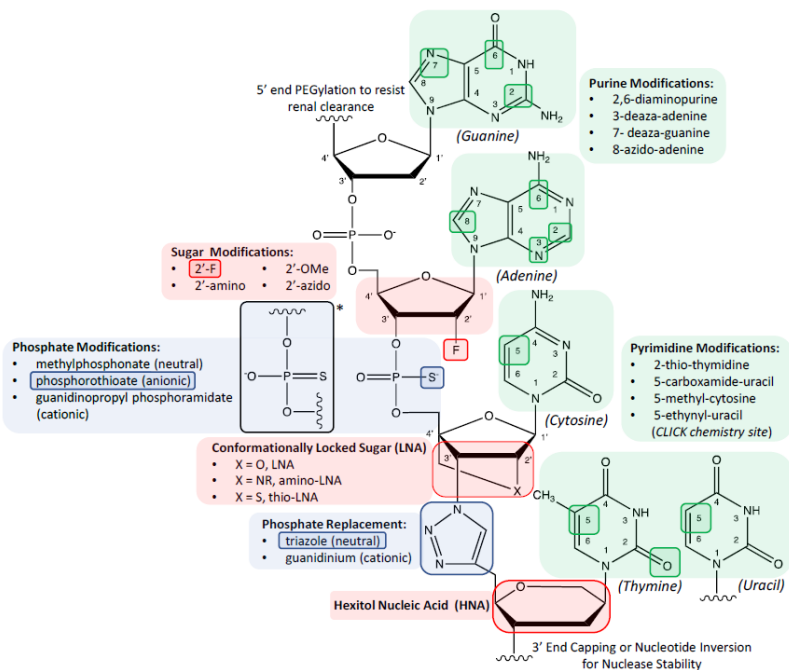
1. LITERATŪROS APŽVALGA

Natūralūs nekanoniniai nukleotidai praturtina cheminę DNR ir RNR įvairovę ir įgalina efektyvų sudėtingų funkcijų atlikimą. Modifikuotų nukleotidų pozicijos ir cheminė struktūra sudaro papildomą genetinės informacijos sluoksnį, tyrinėjamą epigenetikos disciplinos. Spektras modifikacijų, aptinkamų RNR molekulėse, yra stebėtinai platus, tuo tarpu DNR modifikacijų žinoma mažiau. Tai siejama su svarbia, bet ribota DNR funkcija, susijusia su genetinės informacijos saugojimu (Carell *et al.*, 2012).

Pastaraisiais dešimtmečiais susidomėjimas nekanoniniais nukleotidais ir funkcionalizuotais DNR ir RNR polimerais auga (Nakatani & Tor, 2016) sparčiai vystantis jų pritaikomumo sritims cheminėje biologijoje, bioanalizėje, terapijoje. Tarp pritaikomumo pavyzdžių yra molekuliniai reporteriai (Xu *et al.*, 2017), aptamerų (Lapa *et al.*, 2016; Zhou & Rossi, 2017), katalizuojančių nukleorūgščių (Hollenstein, 2015), biojutiklių (Hollenstein *et al.*, 2008) ir kitų molekulinų įrankių kūrimas.

Nukleorūgščių cheminis pakeitimas apima nukleobazės, cukraus arba fosfato grupės modifikavimą ir įvairias šių modifikacijų kombinacijas (1.1 pav.). Norint išnaudoti visą nukleotidų analogų potencialą, būtina turėti polimerazių, pasižyminčių platesniu priimtinų substratų diapazonu. Įdomu tai, kad gamtinės polimerazės pasižymi tam tikru substrato atpažinimo plastiškumu – nukleotidų modifikacijos kai kuriose pozicijose yra gerai toleruojamos, be to, įmanoma sukurti nenatūralias bazių poras ir jas replikuoti, kas leidžia išplėsti genetinį kodą (Hoshika *et al.*, 2019; Marx & Betz, 2020).

Ši apžvalga akcentuos įvairias galimybes, atsiveriančias naudojant modifikuotas nukleorūgštis kaip sintetinius molekulinis įrankius. Apžvelgiamų modifikacijų spektras bus apribotas nukleobazių modifikacijomis, kas atitinka šios disertacijos tematiką.



1.1 pav. Galimų DNR cheminių modifikacijų iliustracija. DNR modifikacijos gali varijuoti nuo paprastų atomų pakaitų tam tikrose pozicijose iki labai kompleksinių pakaitų, mažai panašių į natūralią nukleorūgščių struktūrą (pagal Ochoa & Milam, 2020).

1.1 Funkcionalizuotos nukleorūgštys

Sėkmingam taikymui funkcionalizuotų nukleotidų dizainas turi atitikti tam tikrus kriterijus, originaliai suformuluotus sisteminės ligandų evoliucijos naudojant eksponentinį praturtinimą (*angl.* systematic evolution of ligands by exponential enrichment, SELEX; Perrin *et al.*, 1999) procesui, bet taikytinus ir kitoms molekulinės biologijos metodikoms:

- modifikacijos neturėtų trukdyti bazių sąveikai (Watson-Crick ir Hoogsteen);

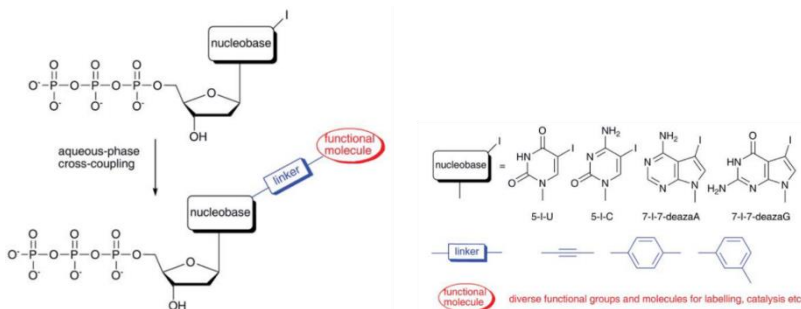
- modifikuoti nukleotidai turėtų būti substratai DNR arba RNR polimerazėms;
- modifikuotų nukleotidų įterpimas turėtų būti efektyvus bet kurioje pozicijoje ar sekos kontekste;
- kai kuriais atvejais funkcionalizuota grandinė turėtų tarnauti kaip matrica polimerazėms.

Bazių modifikacijos paprastai įvedamos į pirimidinų C5 padėtį arba 7-deazapurinų C7 padėtį, nes pakaitai šiose pozicijose orientuojami į didįjį DNR griovį netrikdant spiralinių struktūrų (Jäger *et al.*, 2005). Tyrinėti ir C8 padėtyje modifikuoti purinai, tačiau didelės molekulinės masės pakaitai šioje padėtyje prastai toleruojami polimerazių (Cahová *et al.*, 2008).

Klasikinį funkcionalizuotas nukleobazės turinčių nukleotidų pirmtakų sintezės principą sudaro daugiapakopė aminopropargil, aminopropenil arba aminopropil pakeistų dNTP arba NTP sintezė taikant paladžio katalizuojamą kryžminio sujungimo reakciją (Shaughnessy & DeVasher, 2005; Sonogashira *et al.*, 1975) tarp halogenintų nukleozidų ir CF₃CO- apsauginės grupės turinčių modifikacijų, vėliau sekant trifosforilinizacijai ir apsauginės grupės pašalinimui. Norima funkcinė grupė prijungiama prie amino grupės susidarant amidiniam ryšiui. Šis procesas yra gana sudėtingas ir generuoja mažas tikslinio produkto išeigas. Pavyzdžiui, 5 etapų amidino-dUTP sintezės bendra išeiga sudaro vos apie 8% (Jäger *et al.*, 2005).

Paladžio katalizuojamos kryžminio sujungimo reakcijos plačiai taikomos C-C ryšio formavimui, tačiau jas buvo sunku adaptuoti nukleotidams, kadangi reakcijos terpėse nukleotidų pirmtakai būdavo prastai tirpūs. Atradus vandenyje tirpias katalitines sistemas, vandeninės fazės kryžminio sujungimo reakcijos įgalino nesudėtingą neapsaugotų funkcinių grupių įvedimą tiesiai į dNTP arba NTP (1.2 pav.). Thoresen su kolegomis pademonstravo pirmąsias Sonogashira kryžminio sujungimo reakcijas tarp 5-jodo-dUTP ir fluoresceinu modifikuotų terminalinių acetilenu (Thoresen *et al.*, 2003). Reikalingi halogeninti nukleotidai gali būti gauti chemiškai trifosforilinant

halogenintus nukleozidus (5-I-dU, 5-I-dC, 7-I-7-deaza-dA ir 7-I-7-deaza-dG; Kovacs & Ötvös, 1988).



1.2 pav. Bendrinė funkcionalizuotas nukleobazės turinčių nukleotidų analogų gamybos schema (pagal Hocek & Fojta, 2008).

Dauguma publikuotų tyrimų su bazių modifikacijos turinčiais nukleotidais aprašo C5-pakeistus uracilo analogus, tuo tarpu 7-deazapurinai yra sąlyginai mažiau tyrinėti. Tą galima paaikškinti sudėtinga daugiastadijine tarpinių nukleozidų sinteze (Hocek, 2014).

Ciklizacijos reakcijos yra viena populiariausių strategijų funkcinių grupių konjugavimui, ypač vario katalizuojama azido-alkino grupių ciklizacija (CuAAC), arba „click“ reakcija. „Click“ chemija plačiai taikoma biokonjugavimui ir DNR modifikavimui dėl jos biologinio suderinamumo ir aukšto efektyvumo. Šio tipo reakcijai DNR turi būti modifikuota arba terminaliniu alkinu, arba azido grupe. Alkinai yra suderinami su fosforamiditų sinteze ir fermentinėmis reakcijomis, o azido grupė nėra suderinama su kietafazės fosforamiditų sintezės procesu, tačiau gali būti įterpta į DNR fermentiškai atitinkamai modifikuotų nukleotidų sudėtyje (Ivancová *et al.*, 2019; Panattoni *et al.*, 2018). Svarbu paminėti, kad tipinis DNR funkcionalizavimas yra posintetinis procesas, t.y. substratas „click“ reakcijai yra dgDNR, vgDNR arba oligonukleotidai (ON), turintys vieną, kelis arba daug modifikuotų nukleotidų, įterptų įvairiais fermentiniais metodais, tokiais kaip PGR arba pradmens pratęsimo (PEX) būdu.

1.2 Fermentinis modifikuotų nukleotidų įterpimas

Fermentinėmis reakcijomis pagrįsti metodai modifikuotų nukleorūgščių sintezei yra lankstūs – galima gauti tiek viengrandžius, tiek dvigrandžius produktus, be to, šis sintezės būdas turi mažiau apribojimų tikslinio produkto ilgiui (Whitfield *et al.*, 2018).

Polimerazių sąveikų su modifikuotais substratais tyrimai rodo, kad polimerazių gebėjimas katalizuoti modifikuotų nukleotidų įterpimą priklauso tiek nuo modifikacijos padėties, tiek nuo jos prigimties. Įterpimo efektyvumas taip pat priklauso nuo linkerio tarp funkcinės grupės ir nukleobazės, ypač jei konjuguojamos erdviškai stambios grupės. Paprastai didelės modifikacijos toleruojamos, kai yra prijungtos prie nukleotido per ilgą lankstų linkerį.

Gamtinės polimerazės pasižymi tam tikru substrato atpažinimo plastiškumu. Struktūrinių ypatybių, užtikrinančių nukleotidų analogų įterpimą, analizė padėtų tobulinti fermentus baltymų inžinerijos metodais. Tyrimai vienareikšmiškai rodo, kad B šeimos DNR polimerazės geriau katalizuoja nenatūralių substratų įterpimą už A šeimos polimerazes (Jäger *et al.*, 2005; Kuwahara *et al.*, 2006; Lapa *et al.*, 2016). Modifikuotos RNR sintezė išimtinai remiasi T7 RNR polimerazės panaudojimu, kuri gerai toleruoja bazių modifikacijas, išskyrus modifikuotą GTP +1 padėtyje (Milisavljevič *et al.*, 2018).

1.3 Funkcionalizuotų nukleorūgščių replikacija

Būtina modifikuotų nukleorūgščių pritaikomumo molekuliniėje biologijoje sąlyga yra modifikacijų biologinis suderinamumas – modifikuotos nukleorūgštys turėtų funkciškai mimikuoti jų natūralius analogus (Sanzone *et al.*, 2012). Sąlyginai nedaug tyrimų fokusavosi į dirbtinių matricų replikaciją arba transkripciją. Žinoma, jog nedidelės fosfodiesterinio karkaso modifikacijos toleruojamos polimerazių (Ciafrè *et al.*, 1995), amidinis karkaso variantas buvo netobulai perskaitomas PEX eksperimentuose (Kuwahara *et al.*, 2009), o tam tikri triazolo žiedą turintys DNR karkasai yra

funkcionalūs *in vitro* ir net *in vivo* bakterinėse ir žinduolių ląstelėse (El-Sagheer *et al.*, 2011; Birts *et al.*, 2014).

Shivalingam su bendradarbiais siekė suprasti kokios molekulinės ypatybės įgalintų tikslią dirbtinių DNR karkasų replikaciją (Shivalingam *et al.*, 2017). Mokslininkai susintetino ON su keliais struktūriškai ir elektroniškai skirtingais sintetiniais linkeriais, turinčiais triazolo, amido ir fosforotioato modifikacijas. Linijinio PEX ir eksponentinės PGR amplifikacijos rezultatai parodė, kad net nedidelės fosforotioato modifikacijos gali sutrikdyti DNR kopijavimo procesą. Kita vertus, kai kurie triazolu ar amidu modifikuoti DNR karkasai buvo replikuojami stebėtinai gerai, kas indikuoja, jog fosfato grupė pati savaime nėra būtina. Perskaitymas per sintetinį karkasą identifikuotas kaip reakcijos greitį limituojantis etapas – DNR grandinės kopijavimo greitis koreliavo su dirbtinio karkaso dydžiu. 3'-5' egz nukleazinį aktyvumą turinti Phusion DNR polimerazė kopijavo sintetines matricas geriau nei *Taq* (exo-) polimerazė, kai reakcijos laikas buvo trumpas. Ilginant pratęsimo reakcijos laiką, produkto išėigos reakcijoje su Phusion reikšmingai nedidėjo, skirtingai nei *Taq* polimerazės atveju.

Replikacijos tikslumą autoriai įvertino taikydami NKS. Įdomu tai, kad fosforotioatinis karkasas, nepaisant artimesnio panašumo į natūralų fosfodiesterinį DNR karkasą nei kitos tirtos modifikacijos, buvo kopijuojamas įvedant insercijas, kurių ilgis ir pozicija priklausė nuo naudotos polimerazės. Stipri sąveika tarp sieros atomo ir baltymo greičiausiai trukdė polimerazės perėjimui per modifikacijos sritį ir skatindavo polimerazę pridėti kelis neteisingus dNTP.

Pastebėta, kad polimerazės, turinčios 3'-5' egz nukleazinį aktyvumą, linkusios įvesti delecijas šalia modifikacijos srities. Manoma, kad tokį efektą lemia polimerazės sulėtėjimas ties modifikacijos sritimi ir pratęsimos DNR grandinės perėjimas į egz nukleazinį polimerazės domeną. Polimerizacijos ir hidrolizės procesas tęsiasi tol, kol arba modifikacija yra prašokama, arba modifikacijos regionas „išsukamas“ iš matricinės grandinės ir sintezė tęsiasi toliau. 3'-5' egz nukleazinio aktyvumo neturinčios polimerazės

perskaitymui turi tik vieną bandymą per pratęsimo ciklą, tad delecijų įvedimo tikimybė joms mažesnė. Delecijos replikacijos metu siejamos ir su cukrų konformacijos pokyčiais – delecijos buvo rečiau detektuojamos kopijuojant amidinį karkasą turinčią DNR, kuri reikšmingai nepakeičia cukrų pozicijų ar konformacijų.

Šis tyrimas pasiūlė bendrinius dirbtinių DNR karkasų biologinio suderinamumo principus ir atskleidė, jog kai kurie triazolo žiedą turintys karkasai yra gerai replikuojami, kas leidžia „click“ reakciją taikyti sintetinės modifikuotos DNR gamybai.

1.4 Didelio našumo nukleorūgščių analizė

Mažėjant naujos kartos sekoskaitos (NKS) kainai, tikėtina, kad didelio našumo nukleorūgščių analizės metodai taps dar labiau paplitę ir bus įdiegti į klinikinę praktiką (D'Argenio, 2018).

Visos šiulaikinės sekoskaitos platformos reikalauja tam tikro nukleorūgščių paruošimo prieš sekoskaitą. Paprastai tai apima DNR arba RNR fragmentavimą bei galų modifikavimą tam, kad jie būtų tinkami ligavimo reakcijai. Sintetiniai adapteriai vėliau prijungiami naudojant ligazės fermentą. Sekoskaitai paruošta nukleorūgščių fragmentų biblioteka turėtų būti pažymėta sekoskaitos platformai specifinių adapterių sekomis abiejuose fragmentų galuose (Buermans & den Dunnen, 2014).

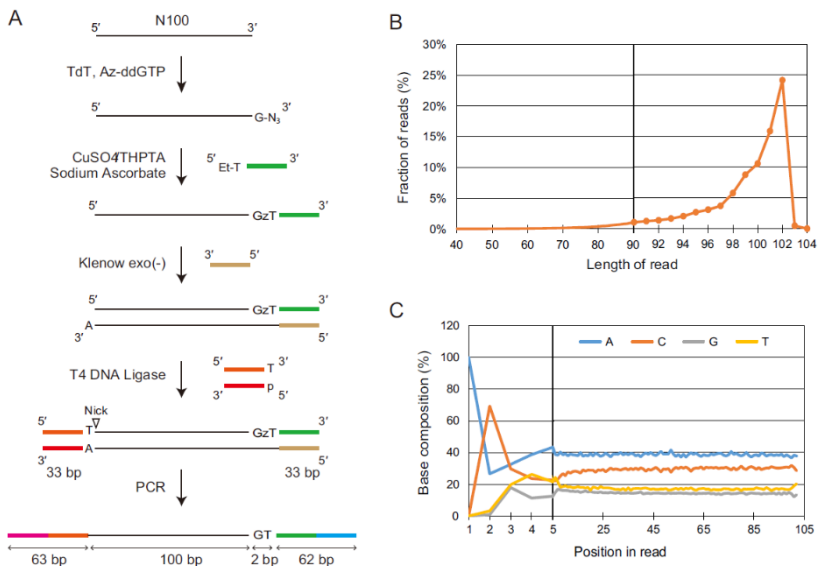
Disertacijoje apžvelgiami DNR ir kDNR fragmentų bibliotekų paruošimo technologijos, besiremiančios modifikuotų nukleorūgščių ir cheminio ligavimo taikymu.

1.4.1 DNR sekoskaita

Efektyvus vgDNR žymėjimas adapteriais įdomus daugeliui DNR sekoskaitos taikymų: metodai originaliai skirti aukšto degradacijos laipsnio senovinės DNR analizei (Gansauge & Meyer, 2013) vėliau buvo pritaikyti kitų fragmentuotos DNR mėginių – cirkuliuojančios DNR bei DNR išgrynintos iš formalinu fiksuotų į parafiną įlietu audinių – analizei. Tarp šių metodų privalumų yra galimybė analizuoti

natūraliai susidariusius 5' ir 3' galus, kas įgalina biologinio fragmentavimo dėsningumą analizę (Troll *et al.*, 2019).

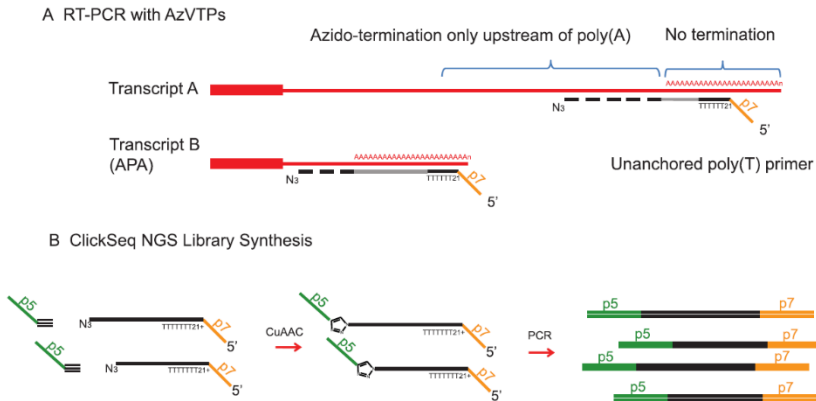
Miura su kolegomis sukūrė chemofermentinį NKS bibliotekų paruošimo iš vgDNR metodą, paremtą terminalinės nukleotidil transferazės (TdT) žymėjimu bei CuAAC grįstu cheminiu ligavimu (Miura *et al.*, 2018). TdT buvo naudota vieno 3'-azidu modifikuoto dideoksinukleotido įterpimui į vgDNR fragmentus. Vėliau prie azidu modifikuotų fragmentų buvo prijungiami 5'-etnil modifikaciją turintys sintetiniai adapteriai naudojant „click“ reakciją. Po komplementarios grandinės sintezės antras sekoskaitos adapteris buvo liguojamas prie susidariusių dgDNR fragmentų naudojant T4 DNR ligazę (1.3 pav.). Autoriai optimizavo cheminio ligavimo efektyvumą iki 20-30%. Komplementarios grandinės sintezei per triazolo žiedą turinčią modifikaciją buvo ieškoma tinkamos DNR polimerazės. Iš 16 tikrintų fermentų, 14 DNR polimerazių nesugeneravo produktų PEX eksperimentuose, o dvi polimerazės (Klenow fragmentas ir Klenow exo- fragmentas) gebėjo replikuoti triazolinę modifikaciją turinčias matricas. Principiniai eksperimentai su vgDNR ON mišiniu sėkmingai sugeneravo NKS bibliotekas, tačiau 12 žingsnių protokolo bendra išeiga buvo vos 0,6%. Sekoskaitos duomenų analizė parodė, kad protokolas turi tam tikrų šalutinių efektų – 57,8% nuskaitymų buvo trumpesni nei tikėtasi greičiausiai dėl vario sąlygotos DNR degradacijos „click“ reakcijos metu.



1.3 pav. NKS bibliotekos paruošimas iš vgDNR. **(A)** – principinių eksperimentų, naudojant 100 nt ilgio vgDNR oligonukleotidus, schema. **(B)** – sekoskaitos nuskaitymų ilgių pasiskirstymas. **(C)** – bazių pasiskirstymas kiekvienoje nuskaitytoje pozicijoje. Pirmą nuskaitymo poziciją atitinka cheminio ligavimo regioną (pagal Miura *et al.*, 2018).

1.4.2 RNR sekoskaita

Routh su bendradarbiais ištyrė galimybę taikyti „click“ chemiją mėginių paruošimui didelio našumo RNR sekoskaitai (Routh *et al.*, 2015; Routh *et al.*, 2017). Jų grupės sukurtas metodas ClickSeq ir jo variantas poly(A)-ClickSeq remiasi atvirkštine transkripcija reakcijos mišinyje esant azido-2',3'-dideoksinukleotidams, kurie atsitiktinai terminuoja kDNR sintezę. Tokie kDNR fragmentai gryninami ir vėliau CuAAC reakcijos metu sujungiami su 5'-etinil modifikaciją turinčiais adapteriais (1.4 pav.).



1.4 pav. NKS bibliotekų paruošimas naudojant poly(A)-ClickSeq metodiką. **(A)** – AT-PGR reakcija vykdoma su azido modifikaciją turinčiais ddATP, ddGTP ir ddCTP. **(B)** – kDNR fragmentai, pažymėti azido grupėmis, sujungiami su 5'-heksinil modifikaciją turinčiais adapteriniais oligonukleotidais (pagal Routh *et al.*, 2017).

Autoriams pavyko gauti ~10% cheminio ligavimo efektyvumą ir paruošti geros kokybės fragmentų bibliotekas iš virusinės RNR. Nepaisant to, kad *Taq* DNR polimerazė gebėjo PGR metu padauginti tikslinius fragmentus, vieno pradmens pratęsimo ciklo efektyvumas buvo labai žemas (<4%), kas, tikėtina, reikšmingai sumažintų konversijos efektyvumą, jei sekoskaitai būtų ruošiamas kompleksiškesnis RNR mėginys. Visgi, autoriai teigia, jog ClickSeq metodas sumažina nepageidaujamos rekombinacijos dažnį, kadangi azido-ddNTP įterpimas nebeleidžia fragmentams dalyvauti grandinės pakeitimo ar kitose šalutinėse reakcijose.

Aptarti pavyzdžiai rodo, kad chemofermentiniai nukleorūgščių modifikavimo metodai gali būti taikomi didelio našumo nukleorūgščių analizei. Suvienijus chemijos ir molekulinės biologijos technologines galimybes įmanoma ne tik supaprastinti analizės procesus, bet ir kurti visiškai naujus sekoskaita paremtus metodus.

2. MEDŽIAGOS IR METODAI

Šiame skyriuje aprašomi ddN^{ON}TP savybių tyrimo eksperimentai ir pagrindiniai DNR ir RNR sekoskaitos metodai, sukurti ddN^{ON}TP pagrindu. Visi atlikti darbai aprašyti disertacijoje.

2.1 Oligonukleotidu modifikuotų dideoksinukleotidų sintezė

Visi komponentai buvo pridėti į reakcijų mišinius kaip vandeniniai tirpalai, jei nenurodyta kitaip. Modifikuoti ON naudoti konjugacijai prie dideoksinukleotidų pateikti lentelėje 2.1. Visi ON buvo pagaminti Metabion GmbH ir išgryninti efektyvios skysčių chromatografijos (HPLC) metodu.

Lentelė 2.1. Šiame darbe naudoti modifikuoti oligonukleotidai

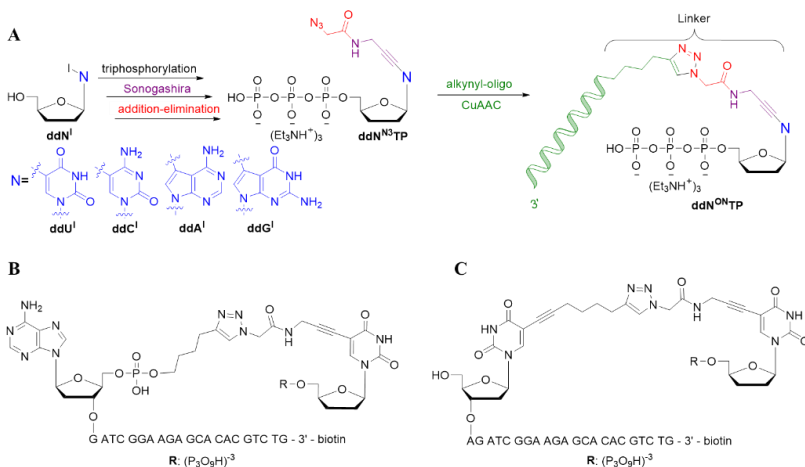
| Nr. | Oligonukleotido seka |
|------|--|
| ON1 | 5'-(AldU)-AGATCGGAAGAGCACACGTCTG-biotinas-3' |
| ON2 | 5'-heksinil-AGATCGGAAGAGCACACGTCTG-biotinas-3' |
| ON3 | 5'-heksinil-AGATCGGAAGAGCACACGTCTG-fosfatas-3' |
| ON4 | 5'-heksinil-AGATCGGAAGAGCACACGT*C*T*G-fosfatas-3' |
| ON5 | 5'-heksinil-NNNNNNNAGATCGGAAGAGCACACGTCTG-biotinas-3' |
| ON6 | 5'-heksinil-NNNNNNNAGATCGGAAGAGCACACGT*C*T*G-fosfatas-3' |
| ON7 | 5'-(AldU)-NNNNNNNAGATCGGAAGAGCGTCTGTGA-biotinas-3' |
| ON8 | 5'-heksinil-NNNNNNNAGATCGGAAGAGCGTCTGTGA-biotinas-3' |
| ON9 | 5'-heksinil-NNNNNNNAGATCGGAAGAGCGTCTGTAGGAAAGAG-fosfatas-3' |
| ON10 | 5'-(AldU)-AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATGCCTAAA TCTCGTATGCCGTCTTCTGCTTG-biotinas-3' |

„AldU“ reiškia 5-(okta-1,7-diinil)-dUTP, * - fosforotioatinės jungtys.

Oligonukleotidai buvo konjuguoti su azido modifikaciją turinčiais dideoksinukleotidais taikant „click“ chemiją. Modifikuotų dideoksinukleotidų pirmtakais buvo 5-(3-(2-azidoacetamido)prop-1-inil)-2',3'-dideoksipirimidinų-5'-trifosfatai arba 7-deaza-7-(3-(2-azidoacetamido)prop-1-inil)-2',3'-dideoksipurinų-5'-trifosfatai.

ddC^{N3}TP, ddU^{N3}TP, ddG^{N3}TP arba ddA^{N3}TP (3 ekv.) tirpalas buvo pridėtas į 5'-etinil modifikaciją turinčio ON (200-210 nmol) tirpalą natrio fosfato buferyje (1 mL, 100 mM, pH 7). Tuomet buvo pridėtas iš anksto sumaišytas CuSO₄ (100 mM, 12 ekv.) ir THPTA (250 mM, 5 CuSO₄ ekv.) tirpalas, o vėliau – natrio askorbato tirpalas (1 M, 50 CuSO₄ ekv.). Reakcijos mišinys buvo inkubuotas 20 min. 42°C, reakcija sustabdyta pridėjus 0,5 M EDTA-Na₂ tirpalo (1 mL, pH 8). Konjugatai buvo išgryninti taikant C18 atvirkštinių fazių chromatografiją naudojant 100 mM TEAAc/ACN (10-30%, priklausomai nuo azido-ddNTP prigimties ir ON modifikacijos) kaip eliuentą ir nudruskinti naudojant vandenį/ACN (0-100%) kaip eliuentą.

Oligonukleotidu modifikuoti dideksinukleotidai (ddN^{ON}TP) buvo gaunami su >20% išeiga ir >95% grynumu. Grynumas buvo vertinamas HPLC, o molekulinės masės patvirtintos skysčių chromatografijos-masių spektrometrijos (LC-MS) metodu. Bendrinis sintezės principas ir pavyzdinės ddN^{ON}TP struktūros pateiktos 2.1 paveiksle.



2.1 pav. (A) – bendrinė ddN^{ON}TP sintezės schema. Mėlynai pažymėta „N⁴“ - nukleobazė. **(B)** – ddN^{ON}TP su ON modifikacija konjuguota susidarant fosfatas-nukleobazė jungčiai. Kaip pavyzdys parodyta ddU^{ON2}TP junginio

struktūra. (C) – ddN^{ON}TP su ON modifikacija konjuguota susidarant nukleobazė-nukleobazė jungčiai. Kaip pavyzdys parodyta ddU^{ON}TP junginio struktūra.

2.2 Įterpimo ir replikacijos tyrimai

Matricų paruošimas ddN^{ON}TP įterpimo tyrimams. Oligonukleotidų dupleksai, turintys skirtingos sekos 5' išsikišančius galus, buvo paruošti sulydžius dalinai komplementarius oligonukleotidus 1× sulydymo buferyje (10 mM Tris-HCl pH 8, 1 mM EDTA, 50 mM NaCl) iki galutinės 2 μM duplekso koncentracijos. Pradmens ON buvo pažymėti fluorescenciniu dažu 5' gale patogiai reakcijų produktų detekcijai gelyje.

Pradmenų pratęsimo (PEX) reakcijų sąlygos. PEX reakcijos buvo vykdytos optimaliuose buferiuose ir optimaliose arba artimose optimalioms temperatūrose kiekvienam tirtam fermentui. Kontrolinės reakcijos atliktos su natūraliais dNTP, siekiant įsitikinti, jog kiekviena polimerazė yra aktyvi duotomis sąlygomis.

Visais atvejais PEX reakcijai naudoti 2 pmol ON duplekso (arba viengrandžio pradmens TdT testavimo atveju) ir 20 pmol ddN^{ON}TP. Polimerazės, pasižymėjusios geriausiu ddN^{ON}TP įterpimo efektyvumu, vėliau buvo tikrintos konkurencinio įterpimo eksperimentuose, kur modifikuoti terminatoriai buvo sumaišyti su atitinkamais dNTP santykiais 1:1, 1:2 arba 1:3.

PEX produktai išfrakcionuoti 15% TBE-Urea PAAG. Prieš užnešimą į gelį mėginiai sumaišyti santykiu 1:1 su 2× DNR užnešimo dažu (98% formamido, 10 mg/mL mėlynojo dekstrano, 10 mM EDTA), inkubuoti 95°C 5 min. ir atšaldyti lede. Elektroforezė vykdyta 1× TBE buferyje esant 400 V įtampai 1 val. 55°C. Vizualizavimas vykdytas naudojant Typhoon™ FLA 9500 (GE Healthcare) sistemą.

RNR polimerazių tyrimai. RNR polimerazės tirtos *in vitro* transkripcijos arba 3' galo žymėjimo būdu pagal atitinkamus gamintojų protokolus („TranscriptAid™ T7 High Yield Transcription Kit“, Thermo Fisher Scientific, ir Poly(U) Polymerase, Cat. No.

M0337S, NEB), išskyrus tai, kad į reakcijos mišinį pridėta ddN^{ON}TP. Reakcijų produktai išgryninti ir analizuoti Agilent 2100 Bioanalyzer sistema naudojant „RNA 6000 Pico“ arba „Small RNA“ rinkinius (Agilent Technologies).

Matricos paruošimas replikacijos tyrimams. Siekiant išsiaiškinti, kurios polimerazės gali replikuoti DNR, turinčią nenatūralų ddN^{ON}TP linkerį, trumpas ddN^{ON}TP įterpimo produktas buvo naudotas kaip PEX matrica. Pradmuo (5'-TGCAGACATGGGTAGGCATCCTTGGCGTA-3') buvo suldydytas su RNR ON (5'- auacgccaaggaugccuacccaugucugca-3') 1× suldydymo buferyje. Dupleksas turėjo 5'-A iškyšą, kuri tarnavo kaip matrica ddU^{ON2}TP įterpimui reakcijoje su SuperScript IV (30 min. 50°C). RNR ON buvo pašalintas paveikus reakcijos mišinį RNaze H, o tiksliniai produktai bei laisvas ddU^{ON2}TP buvo išgryninti naudojant streptavidinu dengtas magnetines daleles „Dynabeads™ M-270 Streptavidin“. Tiksliniai produktai buvo atskirti nuo laisvo ddU^{ON2}TP frakcionuojant 4% agarozės gelyje „E-Gel EX“, o tikslinis produktas išgrynintas naudojant „PureLink™ Quick Gel Extraction Kit“. ~100 fmol išgryninto produkto buvo naudota tolimesnėse PEX reakcijose.

PEX reakcijų sąlygos replikacijos tyrimams. Nenatūralų linkerį turinčios DNR replikacija buvo tirta naudojant termostabilias B šeimos DNR polimerazes bei kelis A ir AT šeimų atstovus. Visais atvejais naudota 1 pmol pradmens komplementaraus ON2 modifikacijai (5'-CAGACGTGTGCTCTTCC-3'). PEX reakcijos vykdytos optimaliuose buferiuose su optimaliais dNTP kiekiais kiekvienai tirtai polimerazei.

Laisvas pradmuo ir kiti viengrandžiai ON buvo pašalinti paveikus reakcijos mišinį Exo I. Dvigrandžiai replikacijos produktai buvo išgryninti taikant precipitaciją etanolium. PEX produktai analizuoti Agilent 2100 Bioanalyzer sistema naudojant „Small RNA“ rinkinį.

2.3 Fragmentų bibliotekų paruošimas taikant *in vitro* transkripciją

T7 RNR polimerazės inžinerija. Siekiant palengvinti linijinio amplifikavimo produktų sekoskaitą, T7 RNRP buvo siekiama pakeisti taip, kad polimerazė gebėtų sintetinti vgdNR – tai leistų transkriptus tiesiogiai naudoti kaip PGR matricas, o ddN^{ON}TP žymėjimas įgalintų nesudėtingą fragmentų bibliotekų paruošimą. Be to, fermentinė vgdNR sintezė yra patrauklus būdas gaminti ON molekulinės biologijos taikymams.

In vitro evoliucijos metodas T7 RNRP variantų su mažesniu selektyvumu substratui atrankai rėmėsi fluorescencijos aktyvuotu lašelių rūšiavimu, kaip aprašyta Kapustina *et al.*, 2021b. Iš 1463 T7 RNRP mutantų bibliotekos buvo atrinkti fermentai, gebantys katalizuoti 2'-deoksipirimidinų įterpimą. Šie mutantiniai fermentai patikrinti ir su kitomis NTP/dNTP substratų kombinacijomis. V783M mutantas pademonstravo geriausią *in vitro* transkripcijos (IVT) efektyvumą su dTTP, dCTP, dATP ir 2'-F-dGTP substratų mišiniu.

Fragmentų bibliotekų paruošimas taikant IVT. Siekiant įvertinti galimybę linijinio amplifikavimo produktus žymėti ddN^{ON}TP ir tiesiogiai amplifikuoti PGR, IVT buvo vykdyta reakcijos mišinyje su 250 ng žiedinės pTZ19R plazmidinės DNR su Illumina P7 adapterio fragmentu, klonuotu už T7 promotoriaus sekos, 2 mM 2'-F-dGTP arba GTP, 2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dTTP, 0.5 μM ddU^{ON8}TP arba ddC^{ON8}TP, 0.1 U neorganinės pirofosfatazės, 8% DMSO, 200 U T7 RNRP V783M ir 1× TranscriptAidTM reakcijos buferio. Reakcijos mišinys inkubuotas 37°C 3 val. Biotinilintais ddN^{ON}TP žymėti transkriptų fragmentai išgryninti naudojant „DynabeadsTM M-270 Streptavidin“ magnetines daleles ir amplifikuoti reakcijos mišinyje su 1× InvitrogenTM CollibriTM Library Amplification Master Mix, 20 U Phusion exo- fermento, 1 μM indeksinių pradmenų iš „InvitrogenTM CollibriTM Stranded RNA Library Prep Kit“ ir benukleazinio vandens iki 50 μL galutinio tūrio. Ciklinimo temperatūrinės sąlygos atitiko gamintojų rekomendacijas. Fragmentų

bibliotekos išgrynintos naudojant „Dynabeads™ Cleanup Beads“ magnetines daleles pagal protokolą iš „Invitrogen™ Collibri™ Stranded RNA Library Prep Kit“ (versija C.0, “Purify the amplified cDNA”).

2.4 Pusiau taikininė 16S rRNR geno sekoskaita

Mėginiai. Principiniams eksperimentams naudota iš *Escherichia coli* BL21 ląstelių išgryninta genomine DNR. Naujas sekoskaitos metodas taip pat įvertintas naudojant komercinius DNR standartus: „ZymoBIOMICS™ Microbial Community DNA Standard“ ir „ATCC 20 Strain Even Mix Genomic Material“. Mikrobiomo sekoskaita buvo atlikta ir su gamtiniais metagenominiais mėginiais. DNR iš 250 mg dirvožemio buvo išgryninta naudojant „ZymoBIOMICS™ DNA Miniprep Kit“.

Pusiau taikininė 16S rRNR geno sekoskaita. PEX reakcijai naudoti pradmenys komplementarūs konservatyviai 16S rRNR geno sričiai tarp V2 ir V3 variabilių regionų.

Bibliotekų paruošimui naudota 30 ng metagenominės DNR. Ekvimoliarinio pradmenų mišinio pratęsimas vykdytas reakcijos mišinyje su Thermo Sequenase fermentu ir ddN^{ONT}P su ON2 modifikacija kaip aprašyta Kapustina *et al.*, 2021a. Žymėti pratęsimo produktai išgryninti „Dynabeads™ M-270 Streptavidin“ magnetinėmis dalelėmis ir amplifikuoti.

Viso metagenomo sekoskaita. 20 ng DNR tirpintos 50 µL 10 mM Tris-HCl (pH 8) buferio sufragmentuota iki ~300 bp naudojant Covaris™ E220 Evolution Focused-ultrasonicator instrumentą. 25 µL fragmentuotos DNR naudoti bibliotekų paruošimui su „Invitrogen™ Collibri™ PS DNA Library Prep Kit“ pagal gamintojo nurodymus.

Palyginimas su tradiciniais 16S rRNR sekoskaitos metodais. Siekiant palyginti pusiau taikininės sekoskaitos generuojamus rezultatus su tradiciniais metodais, bibliotekos iš komercinių DNR standartų ir dirvožemio DNR buvo paruoštos naudojant rinkinius išvardintus lentelėje 2.2. Kiekvienu atveju buvo naudotas

rekomenduojamas DNR kiekis (1-10 ng). Visos bibliotekos ruoštos atliekant du techninius pakartojimus.

Lentelė 2.2. 16S rRNR geno sekoskaitos NKS bibliotekų paruošimo rinkiniai naudoti palyginamajai studijai

| Rinkinys | Kat. Nr. | Gamintojas | Analizuojamas regionas |
|---------------------------------------|-----------------|-------------------|-------------------------------|
| QIAseq™ 16S/ITS Screening Panel | 333812 | Qiagen | Visas genas |
| Swift™ Amplicon 16S+ITS Panel | AL-51648 | Swift Biosciences | Visas genas |
| NEXTFLEX™ 16S V1-V3 Amplicon-Seq Kit | NOVA-4202-02 | PerkinElmer | V1-V3 |
| NEXTFLEX™ 16S V4 Amplicon-Seq Kit 2.0 | NOVA-4203-02 | PerkinElmer | V4 |
| Quick-16S™ NGS Library Prep Kit | D6410 | Zymo Research | V1-V2 |
| 16S Illumina amplicon protocol | N/A | EMP | V4 |

2.5 Genų raiškos analizė

iRNR 3' galų praturtinimas. Principiniai eksperimentai atlikti naudojant gerai charakterizuotą žmogaus RNR – UHRR ir HeLa. Bibliotekos ruoštos iš 100 pg – 1 µg suminės RNR pridendant „Invitrogen™ ERCC ExFold Spike-In Mixes“ kontrolių. Taip pat ištirtas tiesioginis bibliotekų paruošimas iš eukariotinių ląstelių lizatų naudojant 10 – 10000 HEK-293 arba BALB/3T3 ląstelių suspenduotų 1× PBS, pH 7,4. Norint palyginti sekoskaitos duomenis, gaunamus ruošiant bibliotekas iš ląstelių lizatų ir grynos RNR, suminė RNR išgryninta iš 1 mln. HEK-293 ląstelių naudojant „Invitrogen™ PureLink™ RNA Mini Kit“ pagal gamintojo rekomendacijas.

Atvirkštinė transkripcija (AT) vykdyta 20 µL tūrio reakcijos mišinyje su 200 U SuperScript IV, 50 pmol AT pradmens (5'-CTGGAGTTCAGACGTGTGCTCTTCCGATCT(T)₃₀-3'), 20 pmol dNTP mišinio, 40 U RiboLock RNazių inhibitoriaus, 5 mM DTT, 2 pmol ddU^{ON9}TP, 0,4 pmol ddC^{ON9}TP ir 1× SuperScript IV AT buferio. Ruošiant bibliotekas iš ląstelių lizatų, AT reakcija buvo papildyta 0,3% IGEPAL™ CA-630, kuris užtikrina ląstelių lizę. Reakcija vykdyta 30 min. 50°C, sustabdyta inkubuojant 80°C 10 min. Po AT reakcijos mišinys tiesiogiai naudotas kDNR amplifikacijai naudojant „Invitrogen™ Collibri™ Library Amplification Master Mix“ ir Phusion exo-fermentą. Bibliotekos išgrynintos naudojant „Dynabeads™ Cleanup Beads“ magnetines daleles.

Ruošiant bibliotekas iš labai mažų RNR kiekių (100 pg – 500 pg) buvo reikalingas papildomas amplifikavimo žingsnis. Reamplifikavimas vykdytas naudojant „Invitrogen™ Collibri™ Library Amplification Master Mix with Primer mix“ pagal rekomenduojamą protokolą, ciklą skaičius buvo 6-12. Galutinės bibliotekos išgrynintos naudojant „Dynabeads™ Cleanup Beads“ magnetines daleles.

Palyginimas su tradiciniais iRNR 3' galo sekoskaitos metodais. Naujas genų raiškos analizės metodas palygintas su tradiciniu – bibliotekos iš 500 pg, 10 ng, 100 ng ir 500 ng UHRR ir HeLa suminės RNR paruoštos naudojant „Collibri™ 3' mRNA Library Prep Kit for Illumina™ Systems“ pagal gamintojo rekomendacijas. Bibliotekos ruoštos atliekant tris techninius pakartojimus.

2.6 NKS bibliotekų kokybės kontrolė ir sekoskaita

Visų sekoskaitai paruoštų bibliotekų kokybė įvertinta kapiliarine elektroforeze Agilent 2100 Bioanalyzer sistemoje naudojant „High Sensitivity DNA“ rinkinį arba Fragment Analyzer sistemoje naudojant „NGS High Sensitivity“ rinkinį.

Bibliotekų koncentracija matuota kPGR metodu naudojant „Invitrogen™ Collibri™ Library Quantification Kit“ pagal gamintojo rekomendacijas.

Sekoskaitos atliktos Illumina™ MiSeq™ instrumentu naudojant žemiau išvardintus rinkinius:

- „MiSeq Reagent Nano/Micro Kit v2“ (300 ciklų) pilotiniams ir kokybės kontrolės eksperimentams arba žemo kompleksiško mėginiai;
- „MiSeq Reagent Kit v2“ (300 ciklų) DNR ir RNR sekoskaitoms (viso genomo arba viso transkriptomo sekoskaitos);
- „MiSeq Reagent Kit v3“ (150 ciklų) genų raiškos analizės eksperimentams;
- „MiSeq Reagent Kit v3“ (600 ciklų) 16S rRNR geno sekoskaitai.

Gilesnės sekoskaitos atliktos su Illumina™ NovaSeq™ 6000 sistema UAB „Novogene Europe“ (Kembrižas, Jungtinė Karalystė). Sekoskaitos atliktos arba sekvenuojant abu fragmentų galus (2×150 bp, 2×300 bp, 21+100 bp), arba tik vieną galą (1×150 bp), priklausomai nuo bibliotekos struktūros.

Žemo kompleksiško mėginiai prieš sekoskaitą buvo maišomi su 5-20% PhiX kontrolės (Illumina).

2.7 Duomenų analizė

NKS duomenų analizę sudarė standartiniai išankstinio duomenų apdorojimo etapai, tokie kaip filtravimas pagal kokybę, adapterių sekų pašalinimas, normalizavimas pagal sekoskaitos gylį, nuskaitymų prilyginimas prie referentinio genomo. Tolimesni etapai priklausė nuo tirtos bibliotekų paruošimo metodo. Duomenų analizę atliko dr. Gediminas Alzbutas, dr. Varvara Dubovskaja, Karolis Matjošaitis ir Gytis Mackevičius.

3. PAGRINDINIAI REZULTATAI

3.1 ddN^{ON}TP panaudojimas fermentinėse reakcijose

Didžiausią praktinę reikšmę įterpiant ddN^{ON}TP turėtų AT tipo ir A arba B šeimos termostabilios DNR polimerazės. Suderinamas AT fermentas įgalintų kDNR žymėjimą, o termostabilūs fermentai leistų vystyti protokolus naudojant ir vgDNR, ir dgDNR matricas. Paprastai ddN^{ON}TP pažymėti DNR fragmentai toliau amplifikuojami naudojant PGR, tad nenatūralų linkerį turinčios terminuotos kDNR replikacijai labiausiai tiktų termostabilios DNR polimerazės.

3.1.1 ddN^{ON}TP kaip fermentų substratas

ddN^{ON}TP įterpimo eksperimentais buvo tikrinamas ne tik fermentų gebėjimas toleruoti stambias ON modifikacijas, bet ir dideoksiribozės toleravimas bei diskriminacija tarp modifikuotų ir nemodifikuotų nukleotidų. Kaip buvo galima tikėtis, efektyviausios DNR polimerazės iš A ir B šeimų neturėjo egzozonukleazinio aktyvumo bei gebėjo kaip substratus naudoti nukleotidų analogus. Dauguma atvirkštinių transkriptazių gebėjo katalizuoti ddN^{ON}TP įterpimą. Be to, pastebėta, kad TdT, T7 RNRP mutantas ir PUP irgi toleravo modifikuotą substratą.

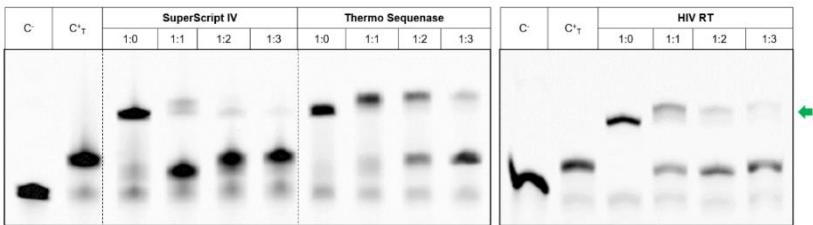
Visų fermentų galimybės katalizuoti modifikuotų terminatorių įterpimą buvo vertintos naudojant ddU^{ON}TP analogus, o pasirinkti kandidatai (Thermo Sequenase ir SuperScript IV AT) buvo papildomai patikrinti su visų rūšių nukleotidų analogais. Tyrimų rezultatai apibendrinti lentelėje 3.1.

Lentelė 3.1. Oligonukleotidu modifikuotų dideoksinukleotidų įterpimas naudojant skirtingas DNR arba RNR polimerazes

| Polimerazė | ddN ^{ON} TP | | | |
|--------------------------|----------------------|-------------------|-------------------|-------------------|
| | ddU ^{ON} | ddC ^{ON} | ddG ^{ON} | ddA ^{ON} |
| <i>Taq</i> | - | nt | nt | nt |
| <i>Taq</i> (exo-) | - | nt | nt | nt |
| <i>Tth</i> | - | nt | nt | nt |
| Platinum™ II <i>Taq</i> | - | nt | nt | nt |
| DyNAzyme™ II | - | nt | nt | nt |
| DyNAmo™ IV | - | nt | nt | nt |
| Thermo Sequenase | ++ | ++ | ++ | + |
| CycleSeq™ | ++ | nt | nt | nt |
| KlenTaq | - | nt | nt | nt |
| Sequenase™ V2.0 | ++ | nt | nt | nt |
| T7 DNR polimerazė | + | nt | nt | nt |
| Klenow fragmentas (exo-) | - | nt | nt | nt |
| Bsm DNR polimerazė | - | nt | nt | nt |
| DNR polimerazė I | - | nt | nt | nt |
| Platinum™ SuperFi™ | - | nt | nt | nt |
| Phusion™ (exo-) | + | nt | nt | nt |
| <i>Pfu</i> (exo-) | - | nt | nt | nt |
| Phusion™ U | - | nt | nt | nt |
| T4 DNR polimerazė | - | nt | nt | nt |
| Phi29 polimerazė | - | nt | nt | nt |
| TdT | ++ | nt | nt | nt |
| Maxima™ AT | ++ | nt | nt | nt |
| SuperScript™ IV | ++ | ++ | + | + |
| SuperScript™ IV Q190N | ++ | nt | nt | nt |
| SuperScript™ IV Q190F | - | nt | nt | nt |
| SuperScript™ IV K103A | + | nt | nt | nt |
| SuperScript™ II | ++ | nt | nt | nt |
| RevertAid™ AT | ++ | nt | nt | nt |
| AMV AT | - | nt | nt | nt |
| MarathonRT | - | nt | nt | nt |
| ŽIV AT | ++ | nt | nt | nt |
| T7 RNRP V783M | ++ | ++ | nt | nt |
| PUP | ++ | nt | nt | nt |

- neįterpia, + įterpia žemu efektyvumu, ++ efektyvus įterpimas, nt – netikrinta

Konkurencinio įterpimo tyrimai atskleidė polimerazių diskriminacijos tarp natūralių deoks nukleotidų ir modifikuotų terminatorių laipsnį (3.1 pav.). Struktūriškai skirtingos (Das & Georgiadis, 2004) M-MLV (SuperScript IV) ir ŽIV AT pademonstravo skirtingus rezultatus: ŽIV AT vienodu efektyvumu įterpė tiek dTTP, tiek ddU^{ON2}TP, tuo tarpu SuperScript IV akivaizdžiai teikė pirmenybę dTTP. Šis rezultatas atitinka anksčiau skelbtus duomenis, rodančius, kad ŽIV AT nuo DNR priklausoma sintezė yra mažiau tiksli nei M-MLV ir AMV AT, tačiau autoriai pastebėjo, kad šie skirtumai mažiau išreikšti su RNR matricomis (Sebastián-Martín *et al.*, 2018).



3.1 pav. Konkurencinio įterpimo eksperimentai. PEX reakcijos atliktos su ddU^{ON2}TP (1:0) arba su ddU^{ON2}TP ir dTTP mišiniais (1:1, 1:2, 1:3). C⁻ - neigiama kontrolė, C⁺ – teigiama kontrolė, žalia rodyklė nurodo produkto su įterptu ddU^{ON2}TP poziciją.

Įdomu tai, kad Thermo Sequenase ir T7 RNRP mutantas V783M teikė pirmenybę ddN^{ON}TP lyginant su dNTP: Thermo Sequenase įterpė ddN^{ON}TP ~2 kartus efektyviau, o T7 RNRP V783M net >1000 kartų efektyviau už nemodifikuotus dNTP. Thermo Sequenase fermentas buvo identifikuotas kaip puikus kandidatas DNR fragmentų bibliotekų paruošimui atliekant PEX ir žymėjimą ddN^{ON}TP naudojant DNR matricas.

3.1.2 ddN^{ON}TP žymėtų nukleorūgščių replikacija

Šiame tyrime išbandytos dvi ddN^{ON}TP konfigūracijos, kuomet dideoksinukleotidas buvo konjuguotas prie ON susidarant bazė-fosfatas jungčiai (2.1 pav. B; modifikacijos ON2, ON3, ON4, ON5, ON6, ON8, ON9) arba bazė-bazė jungčiai (2.1 pav. C; modifikacijos ON1, ON7, ON10). Pastarasis variantas galiausiai buvo atmestas, kadangi alkino grupė, prijungta prie nukleobazės modifikuoto ON sudėtyje, buvo prieinama tik su pirimidiniais, kas priverstų įterpti papildomą nukleotidą į Illumina adapterių sekas. Bazė-fosfatas jungtis nepriklausė nuo sekos ir įgalino molekulių žymėjimą be nepageidaujamų insercijų. Ši konfigūracija buvo išbandyta replikacijos eksperimentuose, naudojant ON modifikaciją kaip pradmens hibridizacijos sritį. Buvo identifikuotos egzozonukleazinio aktyvumo neturinčios polimerazės, gebančios atlikti sintezę per nenatūralų linkerį (lentelė 3.2).

Lentelė 3.2. Įvairių DNR polimerazių replikacijos per ddN^{ON}TP linkerį efektyvumas

| Polimerazė | Sąlygos | Efektyvumas |
|--|--|--------------------|
| Klenow fragmentas (exo-) | 30°C 30 min | + |
| Thermo Sequenase | | + |
| Platinum TM SuperFi TM | 95°C 1 min, 60°C 1 min, | - |
| Phusion U | 72°C 15 min | - |
| Phusion (exo-) | | ++ |
| Phusion (exo-) | 95°C 1 min, 60°C 1 min, 72°C 10 min | ++ |
| Phusion (exo-) | 95°C 1 min, 60°C 1 min, 72°C 5 min | + |
| SuperScript TM IV | 50°C 30 min | + |

- linkerio perskaitymas nedetektuotas, + detektuotas, tačiau neefektyvus perskaitymas (efektyvumas iki 10%), ++ geras perskaitymo efektyvumas (>20% efektyvumas)

Phusion exo- fermentas pasižymėjo geriausiu perskaitymo per nenatūralų linkerį efektyvumu naudojant vieną sintezės ciklą (25-

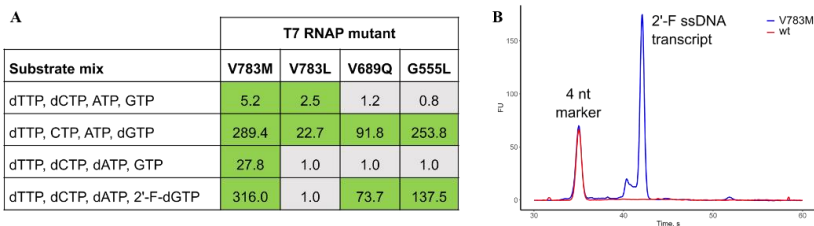
30%). Efektyvumas priklausė nuo reakcijos laiko, Phusion exo-pasiekė didžiausią efektyvumą 10 min. reakcijoje.

3.1.3 Linijinio amplifikavimo produktų žymėjimas

DNR ir RNR polimerazės turi savitus mechanizmus, skirtus atpažinti nekanoninių cukrų turinčius nukleotidų analogus. Substrato atpažinimo veiksmų supratimas įgalintų fermentų katalizinių savybių išplėtimą. Ypatingai įdomi galimybė būtų RNR polimerazės pavertimas DNR polimeraze, tai leistų izoterminėse sąlygose sintetinti vgDNR.

Siekiant sukurti naują linijinės DNR amplifikacijos metodą su integruota galimybe žymėti amplifikavimo produktus tolimesnei didelio našumo sekoskaitai, *in vitro* evoliucijos metodai buvo pasitelkti T7 RNRP variantų, gebančių katalizuoti dNTP įterpimą, atrinkimui.

Mutantas V783M parodė geriausius rezultatus su dTTP, dCTP, dATP ir 2'-F-dGTP substratų mišiniu (3.2 pav.), todėl buvo naudotas tolimesniuose principiniuose eksperimentuose. Nė vienas iš atrastų variantų negebėjo sintetinti nemodifikuotos vgDNR, tačiau GTP pakeitimas į 2'-F-dGTP bei naudojimas kartu su kitais nemodifikuotais deoksinukleotidais buvo sėkmingas. 2'-fluoro modifikacija reikšmingai nekeičia transkripto konformacijos lyginant su RNR (Zhu *et al.*, 2015), kadangi elektroneigiamas fluoro atomas užtikrina į RNR panašią C3'-endo cukraus konformaciją.

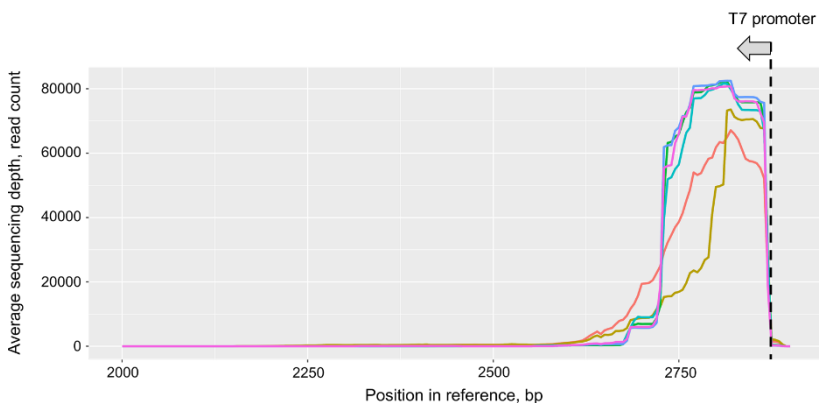


3.2 pav. Mažiau selektyvių substratui T7 RNRP variantų atranka. **(A)** – T7 RNRP mutantai, gebantys naudoti įvairius NTP ir dNTP substratų mišinius. **(B)** – transkripto iš dTTP, dCTP, dATP ir 2'-F-dGTP, susintetinto T7 RNRP V783M mutanto, elektroforegrama.

Izoterminis žiedinės DNR regiono padauginimas IVT reakcijoje su dNTP ir ddN^{ON}TP įgalino gauti atsitiktinai terminuotų, žymėtų transkripcijos produktų biblioteką. Įdomu tai, kad empirinis optimalaus ddN^{ON}TP ir dNTP santykio nustatymas atskleidė, kad T7 RNRP mutantas teikia pirmenybę ddN^{ON}TP – vidutinio ~300 bp dydžio bibliotekos gautos naudojant 5000× modifikuotų terminatorių deficitą. Sekoskaitos duomenų analizės rezultatai parodė, kad vidutiniškai 93,5% nuskaitymų prisilygino prie referentinės plazmidės sekos ir dengė tikėtiną regioną iškart už T7 promotoriaus sekos (3.3 pav.).

Nuskaitymų struktūra atitiko tikėtiną bazių kompoziciją: pirmos 8 bazės pasiskirstė atsitiktinai, kas atitinka randomizuotą regioną ddN^{ON}TP sudėtyje, o G arba A bazių dominavimas 9-toje padėtyje atitiko ddCTP ir ddUTP įsistatymo pozicijas.

Šis metodas įgalina patogią linijinę izoterminę DNR amplifikaciją ir paprastą stabilių transkripcijos produktų procesavimą.



3.3 pav. IVT produktų, susintetintų naudojant dTTP, dCTP, dATP, 2'-F-dGTP ir ddN^{ON}TP substratų mišinį, sekoskaita. Grafikas vaizduoja gautą pTZ19R plazmidės padengimą.

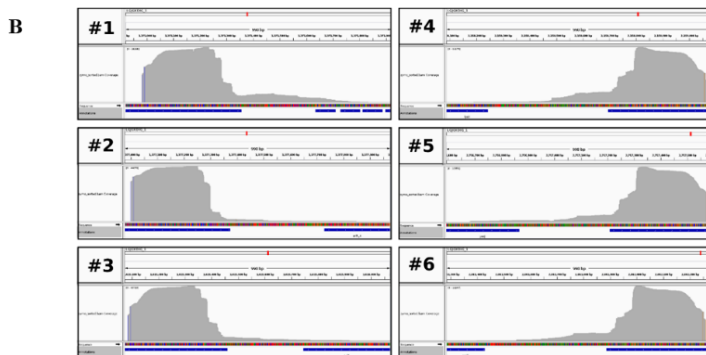
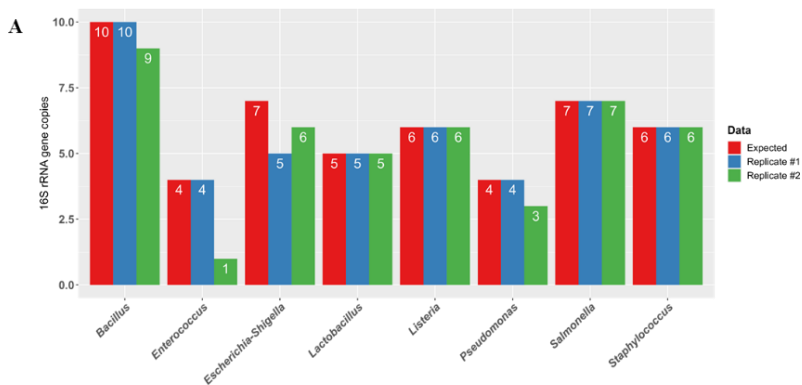
3.2 Pusiau taikininė 16S rRNR geno sekoskaita

Šiame darbe sukurtas naujas pusiau taikininės 16S rRNR geno sekoskaitos metodas (st16S-seq), tiesiogiai siejantis kiekvienos 16S rRNR geno kopijos seką su šalia esančiu genominiu regionu. Tai leidžia tiksliau klasifikuoti ir tiksliau kiekybiškai charakterizuoti bakterijų populiacijas (Kapustina *et al.*, 2021a).

Naujo metodo diagnostinis potencialas iš pradžių buvo įvertintas *in silico* analizuojant 13570 deponuotų duomenų bazėje bakterinių genomų. Pastebėta, kad prieš 16S rRNR geną esančių regionų įtraukimas į analizę iš tiesų pagerina klasifikacijos tikslumą rūšių lygmenyje, be to, gerėjimo tendencija stebima įtraukiant vis ilgesnį genominių fragmentą iki 400 bp. Vidutinis klasifikacijos tikslumas, gaunamas naudojant viso ilgio 16S rRNR geno seką, buvo panašus į V1-V2 regionų kartu su prieš srovę esančiomis genominėmis sekomis. Individualių variabilių regionų klasifikacijos tikslumas buvo ženkliai mažesnis lyginant tiek su viso ilgio 16S rRNR genu, tiek su V1-V2 regionais, sujungtais su genominais lokusais.

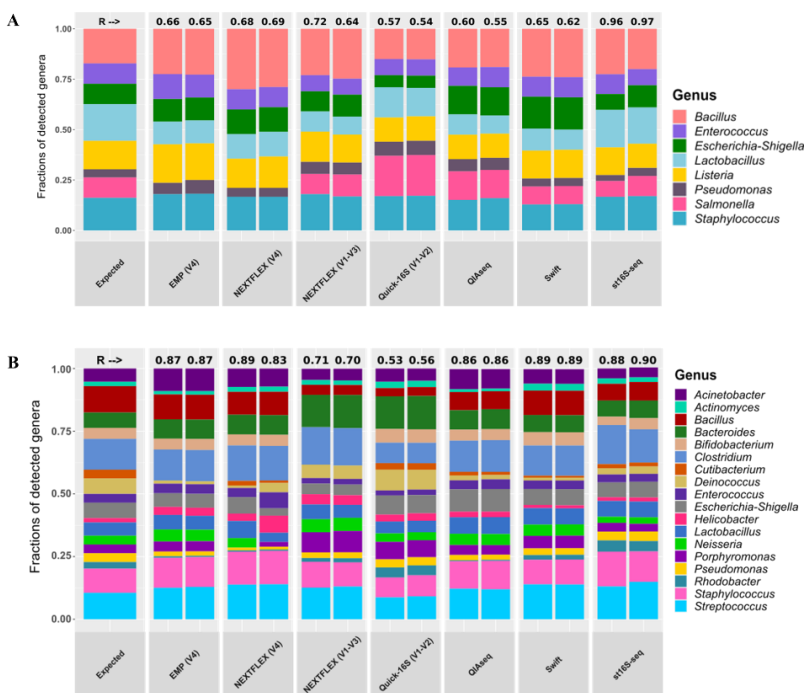
Galimybė iš sekoskaitos duomenų tiksliai identifikuoti 16S rRNR geno kopijų skaičių reikšmingai pagerintų kiekybinę taksonų analizę. 16S rRNR geno hipervariabilios sritys gebėjo patikimai atskirti ~50% geno kopijų individualiuose genomuose. Viso ilgio geno seka padidino identifikuojamų geno kopijų dalį iki 60%, o prieš srovę esančių genominių regionų sujungimas su V1-V2 arba V1-V9 padidino diskriminacijos efektyvumą iki 99,7% geno kopijų. Mikrobiomų mėginius dažnai sudaro artimai giminingų bakterijų mišiniai. Tai apsunkina kiekybinę analizę, nes skirtingų kamienų genomuose yra identiškų 16S rRNR geno kopijų. Siekiant įvertinti 16S rRNR geno kopijų diferenciaciją tarp kamienų, į analizę buvo įtrauktos rūšys su mažiausiai dviem kamienais. Nors nė vienas analizės būdas nepasiekė absoliutaus kopijų atskyrimo, genominių regionų įtraukimas leido identifikuoti >80% 16S rRNR geno kopijų, tuo tarpu viduginės sekos atskyrė $\leq 70\%$ kopijų. Aukštesniame taksoniniame lygmenyje tarprūšinė sekų variacija šiek tiek mažina viso ilgio 16S rRNR geno ir atskirų hipervariabilių regionų sekų diskriminacinę galią, tačiau neturi įtakos šalia esančių genominių sekų gebėjimui atskirti genų kopijas.

Praktinis st16S-seq metodo išbandymas buvo pradėtas nuo gerai charakterizuoto *E. coli* genomo, turinčio 7 rRNR operono kopijas (Maeda *et al.*, 2015), analizės. Iš sekoskaitos duomenų surinkus kontigus, įtraukiant šalia 16S rRNR geno esančius genominius regionus, buvo identifikuotos visos 7 16S rRNR geno kopijos. Taip pat beveik visos 16S rRNR genų kopijos buvo identifikuotos 8-iuose prokariotiniuose organizmuose, sudarančiuose ZymoBIOMICS™ mikrobiominį standartą (3.4 pav.). Pažymėtina, kad intragenominiai sekų skirtumai atskiruose genominiuose lokusuose yra stebimi ribiniame atstume nuo 16S rRNR geno, kurį st16S-seq metodas dar gali patikimai detektuoti naudojant trumpus nuskaitymus, todėl kai kuriais atvejais ne visos rRNR geno kopijos buvo atskirtos.



3.4 pav. 16S rRNR geno kopijų identifikavimas naudojant st16S-seq. **(A)** – 16S rRNR geno kopijų skaičiai nustatyti ZymoBIOMICS™ standarto mėginiuose. **(B)** – šešių 16S rRNR geno kopijų padengimas *Listeria monocytogenes* genome.

16S rRNR geno kopijų nustatymas leidžia normalizuoti nuskaitymų skaičių tiksliam kiekybiniam taksonų pasiskirstymo įvertinimui. st16S-seq duomenys parodė stiprią koreliaciją su tikėtinu pasiskirstymu abiejuose mikrobiominiuose standartuose: ZymoBIOMICS mėginiuose gauti Pearson koreliacijos koeficientai 0,96-0,97 (3.5 pav. A), o ATCC mėginiuose – 0,88-0,90 (3.5 pav. B). Palyginimui, PGR grįstos metodikos pademonstravo prastesnius (Quick-16S ir NEXTFLEX V1-V3) arba nenuoseklius (EMP, NEXTFLEX V4, QIAseq ir Swift) rezultatus, kas gali būti paaiškinta PGR pradmenų dizaino apspręstais apribojimais ir nevienoda individualių 16S rRNR hipervariabilių regionų diskriminacine galia (Klindworth *et al.*, 2013; Winand *et al.*, 2019; Soriano-Lerma *et al.*, 2020).

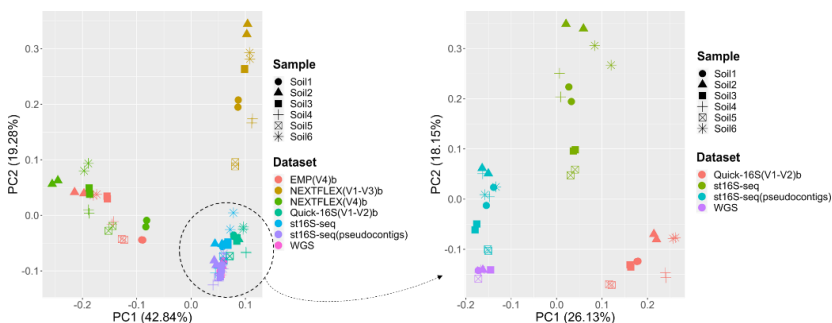


3.5 pav. st16S-seq metodo validavimas su mikrobiominiais standartais ir palyginimas su tradicinėmis technologijomis. (A) – nuskaitymų

pasiskirstymas tarp bakterinių genčių bibliotekose, paruošose iš ZymoBIOMICS™ DNR. Skaičiai virš stulpelių nurodo Pearson koreliacijos koeficientus tarp tikėtino ir gauto pasiskirstymo. Kiekvienam mėginiui parodyti 2 techniniai pakartojimai. **(B)** – nuskaitymų pasiskirstymas tarp bakterinių genčių bibliotekose, paruošose iš ATCC MSA-1002™ DNR. Skaičiai virš stulpelių nurodo Pearson koreliacijos koeficientus tarp tikėtino ir gauto pasiskirstymo. Kiekvienam mėginiui parodyti 2 techniniai pakartojimai.

Prieš 16S rRNR geną esančių sekų informacija pagerino klasifikacijos tikslumą rūšių lygmenyje. Buvo įvertintas bakterinių rūšių klasifikacijos tikslumas mikrobiominiuose standartuose naudojant tiek sujungtas (*angl.* merged), tiek nesujungtas (*angl.* unmerged) nuskaitymų poras. Pastarasis analizės variantas st16S-seq atveju sugeneravo, kaip ir galima buvo tikėtis, palyginamus su V1-V2 amplikonais rezultatus. Priešingai, sujungtų nuskaitymų porų naudojimas ir kontigų surinkimas st16S-seq duomenyse leido teisingai identifikuoti visas bakterines rūšis, išskyrus *Streptococcus agalactiae*, o PGR metodų tikslumas pakeitus analizės strategiją nepagerėjo.

Siekiant ištirti st16S-seq galimybes charakterizuoti kompleksiškas bakterijų bendruomenes, buvo nusekvenuotos iš dirvožemio DNR paruoštos bibliotekos. Nuskaitymų pasiskirstymo principinių komponentų analizė parodė, kad duomenys, gauti V4 amplikonams ir su NEXTFLEX V1-V3 rinkiniu formuoja atskirus klasterius, o st16S-seq ir Quick-16S duomenys aproksimuoja nuskaitymų pasiskirstymą, gaunamą sekvenuojant visą metagenomą. Be to, kai st16S-seq duomenys buvo išanalizuoti kaip pseudokontigai, t.y. visi nuskaitymai, susiję su individualiais taksonominiais vienetais, buvo analizuoti kaip viena visuma, st16S-seq klasterizavosi su viso metagenomo sekoskaitos duomenimis dar geriau (3.6 pav.).



3.6 pav. Kompleksinių bakterijų bendruomenių charakterizavimas naudojant st16S-seq ir tradicines technologijas. Grafikai rodo nuskaitymų pasiskirstymą tarp bakterinių rūšių principinių komponentžių analizės rezultatus. Grafikas dešinėje rodo duomenis, kurie klasterizuojasi šalia viso metagenomo sekoskaitos duomenų.

Apibendrinant, buvo pademonstruota, kad tiesioginis 16S rRNR geno V1-V2 sričių sujungimas su prieš geną esančiais genomiais regionais reikšmingai pagerina mikrobiomų charakterizavimą.

3.3 Genų raiškos analizė

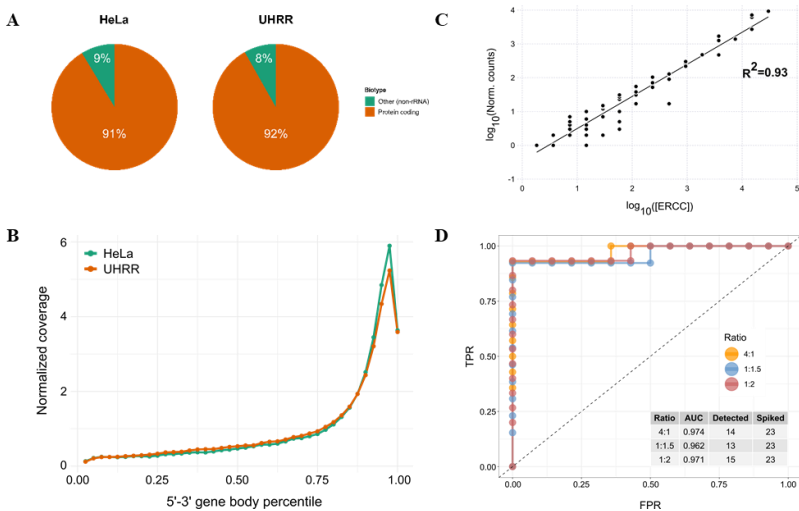
Atliekant genų raiškos analizę su trumpų nuskaitymų sekoskaitos instrumentais, NKS bibliotekos dažnai ruošiamos taip, kad nuo kiekvieno transkripto būtų sugeneruotas vienas kDNR fragmentas, dažniausiai atitinkantis sekas šalia 5' arba 3' transkripto galo.

3' netransliuojamų regionų (3' UTR) praturtinimas įgalina 3' UTR izoformų, kurios prisideda prie individualių ląstelių fenotipinės įvairovės (Velten *et al.*, 2015), variabilumo tyrimus. Šiame darbe buvo sukurtas naujas didelio našumo genų raiškos analizės metodas MTAS-seq, generuojantis NKS bibliotekas iš 3' galinių transkriptų regionų pasitelkiant paprastą vieno mėgintuvėlio protokolą.

Metodas pirmiausia buvo patikrintas analizuojant bibliotekas paruoštas iš gerai charakterizuotų HeLa ir UHRR suminių RNR mėginių, pridėdamas į mėginius kontrolinių ERCC transkriptų mišinį. 99,4 – 99,8% sekoskaitos nuskaitymų prisilygino prie referentinio

žmogaus genomo ir ERCC transkriptų sekų su specifiškumu grandinei >99%. Turint vos 2 mln. nuskaitymų, UHRR mėginiuose buvo detekuota daugiau kaip 19000 genų, o HeLa – beveik 15000. Atvirkštinės transkripcijos sąlygos parodė aukštą specifiškumą iRNR nepaisant to, kad pradinis mėginys buvo suminė RNR: sekoskaitos duomenyse praktiškai nebuvo detekuota rRNR sekų, be to, kaip ir buvo galima tikėtis, transkriptų padengimas koncentravosi šalia 3' galinio regiono (3.7 pav. A-B).

Siekiant iširti kiekybinį MTAS-seq tikslumą, buvo palyginti detektuoti ERCC transkriptų skaičiai su tikėtiniais ir pastebėta, kad detektavus bendrai bent 50 skirtingų ERCC transkriptų, koreliacija (R^2) tarp tikėtino ir gauto kiekybinio pasiskirstymo yra 0,91-0,94 (3.7 pav. C). Diskriminacinė diferencinės raiškos identifikavimo galia buvo įvertinta analizuojant ERCC santykių (tarp „Mix 1“ ir „Mix 2“ kontrolių) detekciją. Kiekvienam santykiui detektavus bent 13 ERCC transkriptų, ploto po kreive (AUC) analizė parodė gerą MTAS-seq diagnostinę galią – tikimybė teisingai detektuoti diferencinę genų raišką buvo >0,96 visiems kiekybiniais santykiams (3.7 pav. D).

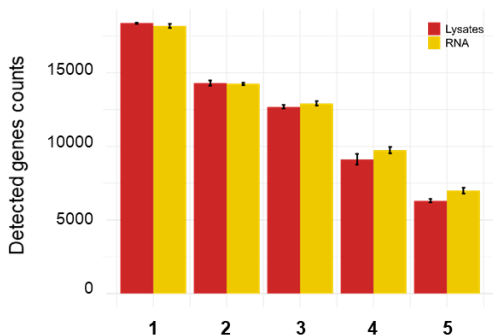


3.7 pav. Duomenų, gautų naudojant MTAS-seq metodą, kokybiniai parametrai. (A) – RNR tipai detektuoti MTAS-seq bibliotekose, paruošose iš gerai charakterizuotos RNR. Į kitų („Other“) transkriptų kategoriją patenka

ERCC RNR kontrolės, kurios buvo detektuotos dėl turimos poliA uodegos. **(B)** – tipinis transkripto padengimas. **(C)** – koreliacijos koeficientas tarp detektuotų ir tikėtinų ERCC transkriptų skaičių MTAS-seq bibliotekose, paruoštose iš 500 ng UHRR su ~2% ERCC kontrolės, buvo 0,93. Iš viso detektuoti 55 skirtingi ERCC transkriptai. **(D)** – *erccdashboard* analizė, vertinanti diferencinės raiškos detekcijos tikslumą.

Norint patikrinti ar aukštos kokybės NKS bibliotekos gali būti paruoštos tiesiogiai eukariotinių ląstelių lizatuose, taip leidžiant nedaryti RNR gryninimo žingsnio, iš HEK-293 ląstelių pirmiausiai buvo išskirta suminė RNR ir pamatuotas apytikslis RNR kiekis vienoje ląstelėje – 12 pg. Tuomet buvo paruoštos MTAS-seq bibliotekos iš įvairių HEK-293 ląstelių kiekių ir lygiagrečiai – iš atitinkamo kiekio grynos suminės RNR. Kiekvienam mėginiui daryti 4 techniniai pakartojimai.

Vidutiniškai 99,1% sekoskaitos nuskaitymų prisilygino prie žmogaus genomo lizatų mėginiuose ir 99,2% - grynos RNR mėginiuose. Gautas labai geras genų detekcijos galimybių atitikimas tarp RNR ir lizatų mėginių (3.8 pav.) ir aukštas techninis atsikartojamumas ruošiant NKS bibliotekas ląstelių lizatuose.



3.8 pav. Detektuotų genų skaičiai MTAS-seq bibliotekose, paruoštose iš ląstelių lizatų ir grynos RNR. 1 – 120 ng RNR arba 10000 ląstelių, 2 – 12 ng RNR arba 1000 ląstelių, 3 – 6 ng RNR arba 500 ląstelių, 4 – 1,2 ng RNR arba 100 ląstelių, 5 – 0,12 ng RNR arba 10 ląstelių.

Tiesioginio bibliotekų paruošimo ląstelių lizatuose metodas buvo pritaikytas kitos rūšies ląstelėms – pelės BALB/3T3 fibroblastams. Gauti aukštos kokybės sekoskaitos duomenys patvirtino MTAS-seq metodo patikimumą bei galimybę analizuoti mažus RNR kiekius.

MTAS-seq duomenų kokybė buvo palyginta su komerciniu rinkiniu, generuojančiu panašios struktūros bibliotekas. MTAS-seq pasirodė geriau specifiškumo iRNR prasme, tačiau kai pradinis RNR kiekis buvo labai mažas, MTAS-seq detektavo mažiau genų palyginus su tradiciniu metodu (lentelė 3.3). Mažesnę jautrumą galėjo nulemti ribotas atvirkštinės transkripcijos su modifikuotais terminatoriais efektyvumas esant mažam RNR matricos kiekiui.

Lentelė 3.3. Techninių parametrų tarp MTAS-seq („ddN^{ON}TP“) ir tradicinio metodo („Trad.“) palyginimas. Visi mėginiai ruošti atliekant 3 techninius pakartojimus; sekoskaitos gylis buvo 2 mln. nuskaitymų mėginiui. Nurodytos reikšmės yra vidurkiai ± standartiniai nuokrypiai.

| Parametras | UHRR | | HeLa | |
|-------------------------------|----------------------|------------|----------------------|------------|
| | ddN ^{ON} TP | Trad. | ddN ^{ON} TP | Trad. |
| Prilyginti nuskaitymai, % | 99.45±0.03 | 99.05±0.39 | 99.68±0.05 | 99.01±0.77 |
| Specifiškumas grandinei, % | 99.21±0.06 | 98.58±0.30 | 99.50±0.04 | 99.14±0.10 |
| Baltymus koduojantys genai, % | 91.71±0.31 | 74.07±4.74 | 91.38±0.18 | 73.90±2.16 |
| rRNR tarša, % | 0.06±0.02 | 7.51±3.94 | 0.03±0.01 | 9.51±3.73 |
| Genų skaičius (500 ng) | 19046±23 | 19124±54 | 14874±20 | 15032±10 |
| Genų skaičius (100 ng) | 18745±38 | 18656±22 | 14443±22 | 14794±59 |
| Genų skaičius (10 ng) | 15633±163 | 15134±926 | 11877±15 | 11554±334 |
| Genų skaičius (0.5 ng) | 7273±73 | 9938±393 | 5055±55 | 7388±602 |

IŠVADOS

1. 5-pakeisti 2',3'-dideoksipirimidinių ir 7-deaza-7-pakeisti-2',3'-dideoksipurinų nukleotidų analogai, turintys 22-67 nt oligonukleotidines modifikacijas, konjuguotas per triazolo žiedą turinčią jungtį, yra substratai Thermo Sequenase, CycleSeq, Sequenase V2.0, T7 DNR, Phusion exo-, TdT polimerazėms, Maxima, SuperScript IV, SuperScript II, RevertAid, ŽIV atvirkštinėms transkriptazėms bei poliU RNR polimerazei.
2. *In vitro* evoliucijos būdu identifiikuotas T7 RNR polimerazės mutantas V783M, gebantis sintetinti transkriptus iš dTTP, dCTP, dATP ir 2'-F-dGTP bei įterpinėti ddN^{ONT}TP, generuojant atsitiktinai terminuotus transkripcijos produktus, pažymėtus oligonukleotidu 3' gale.
3. Klenow fragmentas (exo-), Thermo Sequenase, Phusion exo- ir SuperScript IV geba sintetinti komplementarią DNR grandinę per nenatūralią ddN^{ONT}TP jungtį.
4. Thermo Sequenase fermentas pažymi sintetinamas DNR grandines ddN^{ONT}TP pradmenų pratęsimo reakcijose esant pradmenims ir vgDNR arba dgDNR matricoms. Gaunami fragmentai sudaro charakteringos struktūros sekoskaitai paruoštą biblioteką: fragmentų 5' galai atitinka pratęsimui naudotų pradmenų sekas, o 3' galų pozicijos pasiskirsčiusios atsitiktinai dėl stochastinio modifikuotų dideoksinukleotidų įterpimo pobūdžio.
5. Naujas pusiau taikininės sekoskaitos metodas, sujungęs 16S rRNR geno V1-V2 regionų sekas su genominėmis sekomis, esančiomis prieš 16S rRNR geną, reikšmingai pagerino mikrobiomo charakterizavimą. Naujas metodas įgalina nustatyti 16S rRNR geno kopijų skaičių bei reikšmingai pagerina klasifikacijos tikslumą rūšių lygmenyje.
6. SuperScript IV pažymi sintetinamas kDNR grandines ddN^{ONT}TP atvirkštinės transkripcijos reakcijoje esant pradmenims ir RNR matricoms. Gaunami kDNR fragmentai sudaro sekoskaitai tinkamą

biblioteką, praturtintą 3' UTR/poliA regionais arba dengiančią visą transkripto ilgį, priklausomai nuo naudotų pradmenų sekų.

7. Naujas 3' iRNR sekoskaitos metodas paremtas ddN^{ON}TP žymėjimu tinka tiek tradicinei RNR sekoskaitai, tiek pavienių ląstelių analizei. Naujas metodas lenkia tradicinius bibliotekų ruošimo būdus specifiškumu iRNR, tačiau nusileidžia jautrumu, kuomet mėginyje RNR kiekis labai mažas (<10 ng suminės RNR).

PUBLIKACIJŲ SĄRAŠAS¹

Patentinės paraiškos

- A. Lubys, I. Čikotienė, **Ž. Kapustina**, A. Berezniakovas, J. Medžiūnė, S. Žeimytė. Tethered Oligos And Uses Thereof. Tarptautinė paraiška Nr. PCT/US2020/039009.
- A. Lubys, **Ž. Kapustina**, A. Jasponė. Mutant Polymerases And Methods Of Using The Same. USPTO paraiška Nr. 63/023,026.

Publikacijos

- **Kapustina Ž**, Medžiūnė J, Alzbutas G, Rokaitis I, Matjošaitis K, Mackevičius G, Žeimytė S, Karpus L, Lubys A. High-resolution microbiome analysis enabled by linking of 16S rRNA gene sequences with adjacent genomic contexts. *Microb Genom.* 2021a;7(9). doi: 10.1099/mgen.0.000624.
- **Kapustina Ž**, Jasponė A, Dubovskaja V, Mackevičius G, Lubys A. Enzymatic Synthesis of Chimeric DNA Oligonucleotides by *in Vitro* Transcription with dTTP, dCTP, dATP, and 2'-Fluoro Modified dGTP. *ACS Synth Biol.* 2021b;10(7):1625-1632.
- Medžiūnė J, **Kapustina Ž**, Žeimytė S, Jakubovska J, Sindikevičienė R, Čikotienė I, Lubys A. Advanced preparation of fragment libraries enabled by oligonucleotide-modified 2',3'-dideoxynucleotides. Rankraštis recenzuojamas.

¹ Tik tiesiogiai su disertacijos tema susijusios publikacijos. Visas publikacijų sąrašas pateiktas disertacijoje.

Stendiniai pranešimai mokslinėse konferencijose

- **Kapustina Ž**, Šulgaitė J, Žeimytė S, Lubys A. Innovative approach for digital gene expression and alternative polyadenylation profiling. European Human Genetics Virtual Conference, ESHG 2020.
- **Kapustina Ž**, Medžiūnė J, Žeimytė S, Liucvaikytė A, Lubys A. Innovative approach for digital gene expression and alternative polyadenylation profiling. American Society of Human Genetics Virtual Meeting, ASHG 2020.
- **Kapustina Ž**, Alzbutas G, Šulgaitė J, Žeimytė S, Lubys A. High-precision characterization of microbial communities by the analysis of 16S rRNA gene genomic context. Cold Spring Harbor Laboratory Microbiome Virtual Meeting, 2020.
- **Kapustina Ž**, Medžiūnė J, Drazdauskienė U, Sindikevičienė R, Dubovskaja V, Lubys A. Exploratory analysis of genetic rearrangements. 2021 Advances in Genome Biology and Technology Virtual General Meeting, AGBT 2021.
- Jakubovska J, Sindikevičienė R, Medžiūnė J, Kerzhner M, Kuhn P, **Kapustina Ž**, Lubys A. Oligonucleotide-modified terminators for high-throughput single-stranded DNA sequencing. Revolutionizing Next-Generation Sequencing (Virtual 4th edition), VIB Tools and Technologies conference 2021.
- Drazdauskienė U, **Kapustina Ž**, Medžiūnė J, Gasiūnienė M, Sindikevičienė R, Dubovskaja V, Sabaliauskaitė R, Lubys A. Semi-targeted sequencing of fusion transcripts in prostate cancer. European Human Genetics Conference, ESHG 2021.

CURRICULUM VITAE

| | |
|---------------------------|---|
| Vardas, Pavardė | Žana Kapustina |
| Bendra informacija | |
| Gimimo data | 1991 m. gruodžio mėn. 20 d. |
| Pagrindinė darbovietė | UAB „Thermo Fisher Scientific Baltics“ V.A.Graičiūno g. 8, Vilnius 02241 |
| Asmeniniai kontaktai | Tel. nr: +37068371332 El. paštas: zana.kapustina@thermofisher.com zana@olimpiados.lt |
| Išsilavinimas | |
| 2016 – 2020 | Doktorantūros studijos (Biologija, N 010); Biomokslų institutas, Gyvybės mokslų centras, Vilniaus universitetas. |
| 2014 – 2016 | Genetikos magistras (<i>Magna Cum Laude</i>), Gyvybės mokslų centras, Vilniaus universitetas. |
| 2010 – 2014 | Genetikos bakalauras (<i>Cum Laude</i>), buvęs Gamtos mokslų fakultetas, Vilniaus universitetas. |
| 2008 – 2010 | Klaipėdos Ažuolyno gimnazija. |
| 1998 – 2008 | Klaipėdos Maksimo Gorkio mokykla. |
| Mokymo kursai | |
| 2020 | „Vadovo startas“; ISM Vadovų mokykla, Vilnius, Lietuva. |
| 2018 | „Single-cell analysis“; TATAA Biocenter, Single Cell Europe 2018 konferencija, Praha, Čekija. |
| Darbo patirtis | |
| Nuo 2020 | MTEPC grupės vadovė; UAB „Thermo Fisher Scientific Baltics“. |
| 2015 – 2020 | Jaunesnioji mokslo darbuotoja ir mokslo darbuotoja; UAB „Thermo Fisher Scientific Baltics“. |
| 2012 – 2014 | Laborantė; buvęs Gamtos mokslų fakultetas, Vilniaus universitetas. |
| Dėstymo patirtis | |
| 2018 – 2021 | Molekulinės genetikos metodai (VU GMC Genetikos magistrantūros programa); Transkriptomika (VU MF Sistemų biologijos magistrantūros programa). |

LITERATŪROS SĄRAŠAS

1. Bilyard MK, Becker S, Balasubramanian S. Natural, modified DNA bases. *Curr Opin Chem Biol.* 2020;57:1-7.
2. Birts CN, Sanzone AP, El-Sagheer AH, Blaydes JP, Brown T, Tavassoli A. Transcription of click-linked DNA in human cells. *Angew Chem Int Ed Engl.* 2014;53(9):2362-5.
3. Buermans HP, den Dunnen JT. Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta.* 2014;1842(10):1932-1941.
4. Cahová H, Pohl R, Bednářová L, Nováková K, Cvacka J, Hocek M. Synthesis of 8-bromo-, 8-methyl- and 8-phenyl-dATP and their polymerase incorporation into DNA. *Org Biomol Chem.* 2008;6(20):3657-60.
5. Carell T, Brandmayr C, Hienzsch A, Müller M, Pearson D, Reiter V, Thoma I, Thumbs P, Wagner M. Structure and function of noncanonical nucleobases. *Angew Chem Int Ed Engl.* 2012;51(29):7110-31.
6. Ciafrè SA, Rinaldi M, Gasparini P, Seripa D, Bisceglia L, Zelante L, Farace MG, Fazio VM. Stability and functional effectiveness of phosphorothioate modified duplex DNA and synthetic 'mini-genes'. *Nucleic Acids Res.* 1995;23(20):4134-42.
7. D'Argenio V. The High-Throughput Analyses Era: Are We Ready for the Data Struggle? *High Throughput.* 2018;7(1):8.
8. Das D, Georgiadis MM. The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. *Structure.* 2004;12(5):819-29.
9. Dhuri K, Bechtold C, Quijano E, Pham H, Gupta A, Vikram A, Bahal R. Antisense Oligonucleotides: An Emerging Area in Drug Discovery and Development. *J Clin Med.* 2020;9(6):2004.
10. El-Sagheer AH, Sanzone AP, Gao R, Tavassoli A, Brown T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2011;108(28):11338-43.
11. Gansauge MT, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc.* 2013;8(4):737-48.

12. Hocek M, Fojta M. Cross-coupling reactions of nucleoside triphosphates followed by polymerase incorporation. Construction and applications of base-functionalized nucleic acids. *Org Biomol Chem*. 2008;6(13):2233-41.
13. Hocek M. Synthesis of base-modified 2'-deoxyribonucleoside triphosphates and their use in enzymatic synthesis of modified DNA for applications in bioanalysis and chemical biology. *J Org Chem*. 2014;79(21):9914-21.
14. Hollenstein M, Hipolito C, Lam C, Dietrich D, Perrin DM. A highly selective DNazyme sensor for mercuric ions. *Angew Chem Int Ed Engl*. 2008;47(23):4346-50.
15. Hollenstein M. DNA Catalysis: The Chemical Repertoire of DNazymes. *Molecules*. 2015;20(11):20777-804.
16. Hoshika S, Leal NA, Kim MJ, Kim MS, Karalkar NB, Kim HJ, Bates AM, Watkins NE Jr, SantaLucia HA, Meyer AJ, DasGupta S, Piccirilli JA, Ellington AD, SantaLucia J Jr, Georgiadis MM, Benner SA. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science*. 2019;363(6429):884-887.
17. Ivancová I, Leone DL, Hocek M. Reactive modifications of DNA nucleobases for labelling, bioconjugations, and cross-linking. *Curr Opin Chem Biol*. 2019;52:136-144.
18. Jäger S, Rasched G, Kornreich-Leshem H, Engeser M, Thum O, Famulok M. A versatile toolbox for variable DNA functionalization at high density. *J Am Chem Soc*. 2005;127(43):15071-82.
19. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013;41(1):e1.
20. Koh HR, Roy R, Sorokina M, Tang GQ, Nandakumar D, Patel SS, Ha T. Correlating Transcription Initiation and Conformational Changes by a Single-Subunit RNA Polymerase with Near Base-Pair Resolution. *Mol Cell*. 2018;70(4):695-706.e5.
21. Kovacs T, Otvos L. Simple synthesis of 5-vinyl- and 5-ethynyl-2'-deoxyuridine-5'-triphosphates. *Tetrahedron Lett*. 1988;29:4525-28.
22. Kuwahara M, Nagashima J, Hasegawa M, Tamura T, Kitagata R, Hanawa K, Hososhima S, Kasamatsu T, Ozaki H, Sawai H. Systematic characterization of 2'-deoxynucleoside-5'-triphosphate analogs as

- substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. *Nucleic Acids Res.* 2006;34(19):5383-94.
23. Kuwahara M, Takeshima H, Nagashima J, Minezaki S, Ozaki H, Sawai H. Transcription and reverse transcription of artificial nucleic acids involving backbone modification by template-directed DNA polymerase reactions. *Bioorg Med Chem.* 2009;17(11):3782-8.
 24. Lapa SA, Chudinov AV, Timofeev EN. The Toolbox for Modified Aptamers. *Mol Biotechnol.* 2016;58(2):79-92.
 25. Maeda M, Shimada T, Ishihama A. Strength and Regulation of Seven rRNA Promoters in *Escherichia coli*. *PLoS One.* 2015;10(12):e0144697.
 26. Marx A, Betz K. The Structural Basis for Processing of Unnatural Base Pairs by DNA Polymerases. *Chemistry.* 2020;26(16):3446-3463.
 27. Milisavljević N, Perlíková P, Pohl R, Hocek M. Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org Biomol Chem.* 2018;16(32):5800-5807.
 28. Miura F, Fujino T, Kogashi K, Shibata Y, Miura M, Isobe H, Ito T. Triazole linking for preparation of a next-generation sequencing library from single-stranded DNA. *Nucleic Acids Res.* 2018;46(16):e95.
 29. Nakatani K., Tor Y. *Modified Nucleic Acids, Nucleic Acids and Molecular Biology Series*; Springer, 2016; Vol. 31:1-276.
 30. Ochoa S, Milam VT. Modified Nucleic Acids: Expanding the Capabilities of Functional Oligonucleotides. *Molecules.* 2020;25(20):4659.
 31. Panattoni A, Pohl R, Hocek M. Flexible Alkyne-Linked Thymidine Phosphoramidites and Triphosphates for Chemical or Polymerase Synthesis and Fast Postsynthetic DNA Functionalization through Copper-Catalyzed Alkyne-Azide 1,3-Dipolar Cycloaddition. *Org Lett.* 2018;20(13):3962-3965.
 32. Perrin DM, Garestier T, Hélène C. Expanding the catalytic repertoire of nucleic acid catalysts: simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. *Nucleosides Nucleotides.* 1999;18(3):377-91.
 33. Routh A, Head SR, Ordoukhanian P, Johnson JE. ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of

- Adaptors to Stochastically Terminated 3'-Azido cDNAs. *J Mol Biol.* 2015;427(16):2610-6.
34. Routh A, Ji P, Jaworski E, Xia Z, Li W, Wagner EJ. Poly(A)-ClickSeq: click-chemistry for next-generation 3'-end sequencing without RNA enrichment or fragmentation. *Nucleic Acids Res.* 2017;45(12):e112.
 35. Sanzone AP, El-Sagheer AH, Brown T, Tavassoli A. Assessing the biocompatibility of click-linked DNA in *Escherichia coli*. *Nucleic Acids Res.* 2012;40(20):10567-75.
 36. Sebastián-Martín A, Barrioluengo V, Menéndez-Arias L. Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases. *Sci Rep.* 2018;8(1):627.
 37. Shaughnessy KH, DeVasher RB. Palladium-Catalyzed Cross-Coupling in Aqueous Media: Recent Progress and Current Applications. *Curr Org Chem.* 2005;9:585-604.
 38. Shivalingam A, Tyburn AE, El-Sagheer AH, Brown T. Molecular Requirements of High-Fidelity Replication-Competent DNA Backbones for Orthogonal Chemical Ligation. *J Am Chem Soc.* 2017;139(4):1575-1583.
 39. Sonogashira K, Tohda Y, Hagihara N. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Letters.* 1975;16(50):4467-70.
 40. Soriano-Lerma A, Pérez-Carrasco V, Sánchez-Marañón M, Ortiz-González M, Sánchez-Martín V, Gijón J, Navarro-Mari JM, Garcia-Salcedo JA, Soriano M. Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples. *Sci Rep.* 2020;10(1):13637.
 41. Thoresen LH, Jiao GS, Haaland WC, Metzker ML, Burgess K. Rigid, conjugated, fluoresceinated thymidine triphosphates: syntheses and polymerase mediated incorporation into DNA analogues. *Chemistry.* 2003;9(19):4603-10.
 42. Troll CJ, Kapp J, Rao V, Harkins KM, Cole C, Naughton C, Morgan JM, Shapiro B, Green RE. A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genomics.* 2019;20(1):1023.

43. Velten L, Anders S, Pekowska A, Järvelin AI, Huber W, Pelechano V, Steinmetz LM. Single-cell polyadenylation site mapping reveals 3' isoform choice variability. *Mol Syst Biol.* 2015;11(6):812.
44. Whitfield CJ, Little RC, Khan K, Ijro K, Connolly BA, Tuite EM, Pike AR. Self-Priming Enzymatic Fabrication of Multiply Modified DNA. *Chemistry.* 2018;24(57):15267-15274.
45. Winand R, Bogaerts B, Hoffman S, Lefevre L, Delvoye M, Braekel JV, Fu Q, Roosens NH, Keersmaecker SC, Vanneste K. Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. *Int J Mol Sci.* 2019;21(1):298.
46. Xu W, Chan KM, Kool ET. Fluorescent nucleobases as tools for studying DNA and RNA. *Nat Chem.* 2017;9(11):1043-1055.
47. Zhou J, Rossi J. Aptamers as targeted therapeutics: current potential and challenges. *Nat Rev Drug Discov.* 2017;16(3):181-202.
48. Zhu B, Hernandez A, Tan M, Wollenhaupt J, Tabor S, Richardson CC. Synthesis of 2'-Fluoro RNA by Syn5 RNA polymerase. *Nucleic Acids Res.* 2015;43(14):e94.

UŽRAŠAMS

UŽRAŠAMS

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p.: info@leidykla.vu.lt, www.leidykla.vu.lt
Tiražas 15 egz.