

<https://doi.org/10.15388/vu.thesis.454>

<https://orcid.org/0000-0002-9894-3423>

VILNIAUS UNIVERSITETAS

Janina Ličytė

# Oksiduotų 5-metilcitozino formų nustatymas naujais metodais, jų genominis pasiskirstymas ir funkcijos

**DAKTARO DISERTACIJA**

Gamtos mokslai,  
Biochemija (N 004)

VILNIUS 2023

Disertacija rengta 2013–2017 metais Vilniaus universitete, Gyvybės mokslų centre, Biotechnologijos institute.

Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Disertacija ginama eksternu.

**Mokslinė konsultantė** – dr. Edita Kriukienė (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).

Gynimo taryba:

**Pirmininkė** – prof. dr. Edita Sužiedėlienė (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).

**Nariai:**

prof. dr. Skirmantas Kriaučionis (Oksfordo universitetas, gamtos mokslai, biochemija, N 004),

dr. Arūnas Lagunavičius („Thermo Fisher Scientific Baltics“, gamtos mokslai, biochemija, N 004),

prof. dr. Artūras Petronis (Toronto universitetas, gamtos mokslai, biochemija, N 004),

dr. Mindaugas Zaremba (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).

Disertacija ginama viešame Gynimo tarybos posėdyje 2023 m. balandžio mėn. 20 d. 14 val. Vilniaus universiteto Gyvybės mokslų centro R401 auditorijoje ir/ar nuotoliniu būdu. Adresas: Saulėtekio al. 7, LT-10257, Vilnius, Lietuva. el. paštas [janina.licyte@bti.vu.lt](mailto:janina.licyte@bti.vu.lt)

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

<https://doi.org/10.15388/vu.thesis.454>

<https://orcid.org/0000-0002-9894-3423>

VILNIUS UNIVERSITY

Janina Ličytė

# Identification of Oxidized 5-Methylcytosine Derivatives by Novel Technologies, their Distribution and Functions

**DOCTORAL DISSERTATION**

Natural Sciences,  
Biochemistry (N 004)

VILNIUS 2023

The dissertation was prepared between 2013 and 2017 at Vilnius University, Life Sciences Center, Institute of Biotechnology.

The research was supported by the Research Council of Lithuania.

The dissertation is defended on an external basis.

**Academic consultant** – Dr. Edita Kriukienė (Vilnius University, Natural Sciences, Biochemistry, N 004).

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

**Chair** – Prof. Dr. Edita Sužiedėlienė (Vilnius University, Natural Sciences, Biochemistry, N 004).

**Members:**

Prof. Dr. Skirmantas Kriaučionis (University of Oxford, Natural Sciences, Biochemistry, N 004),

Dr. Arūnas Lagunavičius (Thermo Fisher Scientific Baltics, Natural Sciences, Biochemistry, N 004),

Prof. Dr. Artūras Petronis (Toronto University, Natural Sciences, Biochemistry, N 004),

Dr. Mindaugas Zaremba (Vilnius University, Natural Sciences, Biochemistry, N 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 14:00 on 20th of April 2023 in auditorium R401 of the Vilnius University Life Sciences Center and/or online.

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

e-mail: [janina.licyte@bti.vu.lt](mailto:janina.licyte@bti.vu.lt)

The text of this dissertation can be accessed at the library of Vilnius University, as well as on the website of Vilnius University:

[www.vu.lt/lt/naujienos/ivykiu-kalendorius](http://www.vu.lt/lt/naujienos/ivykiu-kalendorius)

## TURINYS

SANTRUMPŲ SĄRAŠAS .....	9
ĮVADAS .....	11
1. LITERATŪROS APŽVALGA .....	16
1.1 DNR modifikacijos.....	16
1.1.1 5mC pasiskirstymas ir funkcijos eukariotuose .....	17
1.1.2 Eukariotų DNR metiltransferazės.....	22
1.1.3 5-metilcitozino demetilinimas žinduoliuose.....	26
1.1.4 TET baltymai.....	30
1.1.5 Oksiduotų 5-metilcitozino formų funkcijos .....	33
1.1.6 Kiti galimi žinduolių aktyvaus demetilinimo mechanizmai .....	39
1.1.7 <i>TET-JPB</i> genų ekspansija modeliniuose grybuose <i>Laccaria</i> <i>bicolor</i> ir <i>Coprinopsis cinerea</i> .....	41
1.2 RNR modifikacijos.....	45
1.2.1 RNR modifikacijų paplitimas ir funkcijos.....	45
1.2.2 5-metilcitozino funkcijos RNR.....	48
1.2.3 Oksiduotų 5-metilcitozino formų funkcijos RNR .....	50
1.3 Citozino modifikacijų nustatymo metodai .....	52
1.4 DNR citozino modifikacijų genomio profiliavimo sekoskaitos metodai .....	55
1.4.1 Afininiu praturtinimu paremti metodai.....	56
1.4.2 Nukleotido skiriamosios gebos metodai, paremti kovalentine fermentine modifikacija ir prijungto oligonukleotido nukreipta sekoskaita .....	61
1.4.3 Restrikcijos endonukleazių savybėmis paremti metodai .....	64
1.4.4 Bisulfitinė sekoskaita ir jos variantai.....	67
1.4.5 Cheminės ar fermentinės konversijos metodai nenaudojant natrio bisulfito.....	72
1.4.6 Trečios kartos sekoskaita.....	74
2. MEDŽIAGOS IR METODAI .....	77
2.1 Medžiagos .....	77

2.1.1 Reagentai .....	77
2.1.2 DNR preparatai.....	78
2.1.3 RNR preparatai.....	82
2.1.4 Baltymų preparatai .....	82
2.1.5 Molekulinės biologijos rinkiniai.....	83
2.1.6 Buferiniai tirpalai.....	84
2.2 Metodai.....	84
2.2.1 Plonasluoksni chromografija (TLC) .....	84
2.2.2 Modelinių DNR fragmentų paruošimas .....	85
2.2.3 DNR analizė naudojant restrikcijos endonukleazes ir kiekybinę PGR .....	85
2.2.4 eM.SssI vykdomo dekarboksilinimo ir žymėjimo azido grupė įvertinimas su modeliniais DNR fragmentais.....	86
2.2.5 eM.SssI vykdomo dekarboksilinimo analizė IMR90 genominiuose regionuose .....	87
2.2.6 5hmC ir nemetilintų CG taikinių apsaugojimas .....	87
2.2.7 wt M.SssI metilinimo efektyvumo ir 5caC lygio įvertinimas bisulfitine sekoskaita lambda bakteriofago DNR.....	88
2.2.8 Lambda DNR paruošimas caCLEAR analizei .....	88
2.2.9 mESC auginimas ir DNR skyrimas .....	89
2.2.10 Aukšto slėgio skysčių chromatografija-tandeminė masių spektrometrija (HPLC-MS/MS) .....	90
2.2.11 caCLEAR bibliotekų paruošimas .....	92
2.2.12 Pelės genominių regionų 5caC analizė caMAB-seq metodu.....	93
2.2.13 Pelės genominių regionų 5caC analizė piridino borano sekoskaita .....	94
2.2.14 Grybų auginimas .....	94
2.2.15 Grybų DNR gryninimas .....	95
2.2.16 Grybų RNR gryninimas.....	95
2.2.17 Kontrolinės RNR paruošimas.....	95
2.2.18 5caC analizė taškiniu imunoblotu.....	96

2.2.19 5caC turinčios RNR cheminė derivatizacija .....	96
2.2.20 Grybų RNR gryninimas į atskiras frakcijas.....	97
2.2.21 5hmC ir 5fC praturtinimo įvertinimas naudojant modelinius DNR fragmentus.....	97
2.2.22 hMe-Seal ir fC-Seal bibliotekų paruošimas.....	98
2.2.23 hmTOP-seq ir foTOP-seq bibliotekų paruošimas .....	99
2.2.24 RNR bibliotekų paruošimas .....	99
2.2.25 Sekoskaitos duomenų analizė.....	99
3. REZULTATAI .....	101
3.1 DNR citozino-5 metiltransferazių atliekama 5caC dekarboksilinimo reakcija .....	101
3.2 5caC nustatymo metodo, caCLEAR, kūrimas ir įvertinimas .....	104
3.2.1 Nemodifikuotų CG ir 5hmC modifikavimo įvertinimas .....	105
3.2.2 eM.SssI atliekamo 5caC dekarboksilinimo ir žymėjimo įvertinimas .....	107
3.2.3 Visos caCLEAR procedūros įvertinimas.....	111
3.3 5caC tyrimas pelės embrioninėse kamieninėse ląstelėse caCLEAR metodu .....	113
3.3.1 Dviejų pelės ESC pluriipotentiško būsenų caCLEAR analizė .....	113
3.3.2 5caC modifikuoti transkripcijos veiksnių ir reguliatorių jungimosi regionai dviejose pluriipotentiško būsenose skiriasi .....	119
3.3.3 5caCG pasiskirsto antiprasminėje aktyvių genų grandinėje.....	124
3.4 Oksiduotų 5-metilcitozino formų tyrimas <i>Laccaria bicolor</i> ir <i>Coprinopsis cinerea</i> papėdgrybiuose .....	130
3.4.1 Oksiduotų 5-metilcitozino formų nustatymas <i>L. bicolor</i> ir <i>C.</i> <i>cinerea</i> .....	131
3.4.2 5caC pasiskirstymas RNR .....	133
3.4.3 5hmC ir 5fC genomis profiliavimas .....	139
3.4.4 5hmC ir 5fC ryšys su genų raiška.....	143
4. REZULTATŲ APTARIMAS .....	150

4.1 Žmogaus DNMT3A/3L ir DNMT3B/3L metiltransferazių atliekama 5caC dekarboksilinimo reakcija .....	150
4.2 Sukurtas eM.SssI 5caC dekarboksilinimo reakcija paremtas 5caC profiliavimo metodas caCLEAR .....	152
4.3 caCLEAR metodas atskleidė 5caC pasiskirstymo skirtumus esant dviems pelės embrioninių ląstelių pliuripotentiškumo būsenoms .....	153
4.4 Oksiduotų 5-metilcitozino formų pasiskirstymas ir jų vaidmuo <i>L. bicolor</i> ir <i>C. cinerea</i> papėdgrybių nukleorūgštyse .....	156
IŠVADOS.....	160
LITERATŪROS SĄRAŠAS.....	161
SUMMARY OF DOCTORAL DISSERTATION .....	187
DISERTACIJOS TEMA PASKELBTI MOKSLINIAI DARBAI.....	218
CURRICULUM VITAE .....	220
PADĖKA.....	223
UŽRAŠAMS .....	225



## SANTRUMPŲ SĄRAŠAS

2i	du MEK ir GSK3 kinazių slopikliai (angl. <i>two inhibitors</i> )
5caC	5-karboksilcitozinas
5fC	5-formilcitozinas
5hmC	5-hidroksimetilcitozinas
5mC	5-metilcitozinas
Ado-6-N <sub>3</sub>	sintetinis SAM kofaktoriaus analogas, turintis šoninę heks-2-inil-azido grandinę
ar.	aminorūgštis(ys)
BER	bazių iškirpimo ir reparacijos sistema (angl. <i>base excision repair</i> )
βGT	T4 bakteriofago β-gliukoziltransferazė
bp	bazių poros
BS-Seq	bisulfitinė sekoskaita (angl. <i>bisulfite sequencing</i> ).
caCLEAR	5caC genomino profiliavimo metodas (angl. <i>5caC clearance</i> ), kuriame naudojamas eM.SssI atliekamas 5caC dekarboksilinimas ir TOP-seq
CG arba CpG	citozino ir guanino dinukleotidas
CGI	CpG sankaupos – CpG salos (angl. <i>CpG island</i> )
DNMT	eukariotų DNR metiltransferazės
eM.SssI	mutantinė (angl. <i>engineered</i> ) CpG metiltransferazė SssI (Gln142 ir Asn370 pakeisti į Ala)
foTOP-seq	5fC genomino profiliavimo metodas, kuriame 5fC redukuojamas į 5hmC ir atliekamas hmTOP-seq
H3K	H3 histono lizinas, nurodoma jo pozicija ir modifikacijos, pvz.: 4me1 – 4 lizino monometilinimas (H3K4me1), 36me3 – 36 lizino trimetilinimas, 27ac – 27 lizino acetilinimas
hmTOP-seq	5hmC profiliavimo metodas, kuriame atliekamas βGT 5hmC žymėjimas ir TOP-seq
HPLC-MS/MS	aukšto slėgio skysčių chromatografija-tandeminė masių spektrometrija (angl. <i>high pressure liquid chromatography-tandem mass spectrometry</i> )
kPGR	kiekybinė polimerazinė grandininė reakcija
M. arba MTazė	metiltransferazė
mESC	pelės embrioninės kamieninės ląstelės (angl. <i>mouse embryonic stem cells</i> )
mTAG	metiltransferazės katalizuojama aktyvių grupių pernaša nuo sintetinio kofaktoriaus analogo ant specifinio substrato

(angl. *methyltransferase-directed transfer of activated groups*)

nt	nukleotidas(ai)
Oksi-mC	oksiduotos 5-metilcitozino formos
R. arba REazė	restrikcijos endonukleazė
rC	RNR citozino bazė
SAH	S-adenozil-L-homocisteinas
SAM	S-adenozil-L-metioninas
TDG	timino DNR glikozilazė
TE	judrieji genomo elementai (angl. <i>transposable elements</i> )
TET	5mC oksidaciją atliekantys baltymai (angl. <i>ten-eleven translocation</i> )
TOP-seq	metodas, kuris yra paremtas eM.SssI nemonifikuotų CG sričių žymėjimu aktyviomis grupėmis ir DNR grandinės pratęsimu nuo prie CG prijungto oligonukleotido (angl. <i>tethered oligonucleotide-primed sequencing</i> )
TSS	transkripcijos pradžios vieta (angl. <i>transcription start site</i> )
UTR	netransliuojamas regionas (angl. <i>untranslated region</i> )
wt	laukinis tipas (angl. <i>wild type</i> )

*Tekste naudojamos standartinės aminorūgščių pavadinimų vienos ir trijų raidžių santrumpos bei nukleorūgščių bazių pavadinimų vienos raidės santrumpos.*

## IVADAS

5-metilcitozinas (5mC) yra paplitusi DNR modifikacija, aptinkama nuo bakterijų iki žinduolių. 5mC susidaro veikiant DNR metiltransferazių fermentams, kurie perkelia metilo grupę nuo S-adenozil-L-metionino (SAM) kofaktoriaus ant citozino C5 pozicijos (Goll ir Bestor, 2005). 5mC yra esminė daugialąsčių eukariotų epigenetinė DNR modifikacija, kuri nepakeičia koduojančios sekos, tačiau yra labai svarbi genų raiškos reguliacijai, pasikartojančių sekų slopinimui, imprintingui, žinduolių X chromosomos nutildymui (Lu ir kt., 2015a).

5mC genominiis pasiskirstymas yra susijęs su ląstelei būdinga transkripcijos būkle, todėl 5mC pokyčiai yra svarbūs ląstelių diferencijacijai vystymosi metu, pliuripotencijos palaikymui, senėjimui ir ligų atsiradimui (Parry ir kt., 2021). DNR demetilinimas gali vykti pasyviu būdu, kai 5mC kiekiai mažėja vykstant DNR replikacijai. Aktyvus demetilinimas vyksta nepriklausomai nuo replikacijos, kai fermentai greitai pašalina 5mC (Delatte ir Fuks, 2013).

Žinduoliuose aktyvus DNR demetilinimas gali vykti veikiant TET baltymams, kurie paeiliui oksiduoja 5mC į 5-hidroksimetilcitoziną (5hmC), 5-formilcitoziną (5fC) ir 5-karboksilcitoziną (5caC) (He ir kt., 2011; Ito ir kt., 2011; Pfaffeneder ir kt., 2011). 5fC ir 5caC yra iškerpami timino DNR glikozilazės, o susidariusi bazės neturinti vieta ištaisoma bazių iškirpimo reparacijos sistemos ir taip grąžinama nemodifikuota citozino būsena (He ir kt., 2011; Maiti ir Drohat, 2011).

Daugėja įrodymų, kad šios oksiduotos 5-metilcitozino formos (oksi-mC) yra ne vien tarpiniai demetilinimo produktai, bet ir atlieka unikalias epigenetines funkcijas žinduoliuose. 5hmC yra gausiausia oksi-mC, svarbi transkripcijos reguliavimui, embriono vystymuisi, nervinių audinių ir atminties formavimuisi (An ir kt., 2017). 5fC ir 5caC yra retesnės modifikacijos ir jų savarankiškos epigenetinės funkcijos iširtos mažai, nors atsiranda tai remiančių duomenų. Pavyzdžiui, yra nustatyti iš dalies skirtingi oksi-mC atpažįstantys baltymai ir oksi-mC vietos genome ne visai sutampa (Spruijt ir kt., 2013; Iurlaro ir kt., 2013; Wu ir kt., 2014).

Kiti galimi žinduolių aktyvaus demetilinimo būdai dar mažai supraci. Vienas iš pasiūlytų alternatyvių mechanizmų – tiesioginis 5caC dekarboksilinimas susidarant nemodifikuotam citozinui (Pfaffeneder ir kt., 2011; Wu ir Zhang, 2010). Be to, buvo pastebėtas tiesioginis 5caC dekarboksilinimas inkubuojant DNR su pelės embrioninių kamieninių ląstelių ekstraktais (Schiesser ir kt., 2012), tačiau reakciją atliekantys fermentai dar nenustatyti. VU BTI DNR modifikacijos tyrimų skyriuje buvo nustatyta, kad

žmogaus DNMT3A ir 3B bei bakterijų M.SssI DNR citozino-5 metiltransferazės geba atlikti 5caC dekarboksilinimą susidarant nemodifikuotam citozinui (Liutkevičiūtė ir kt., 2014). Ši reakcija parodo galimą fermentinį tiesioginės 5caC konversijos į C kelią, nors neaišku, ar metiltransferazės galėtų atlikti šią reakciją esant gausiam SAM *in vivo*.

Žinduolių embrioninio vystymosi metu 5mC lygiai smarkiai keičiasi. Susiformavus zigotai 5mC lygiai krenta ir susidaro oksimC, kurių gausumas po kurio laiko taip pat sumažėja. Vykstant ankstyvojo epiblasto implantacijai į gimdą, 5mC bei 5hmC lygiai vėl pakyla (Yang ir kt., 2020). Auginant pelės embrionines kamienines ląsteles terpėje su dviem kinazių slopikliais (2i), jos yra naivios pliuripotentiškumo būsenos, turi žemus 5mC bei 5hmC lygius ir yra panašios į blastocistos vidinės masės ląsteles prieš implantaciją (Ficz ir kt., 2013; Habibi ir kt., 2013). Tuo tarpu serumo sąlygomis auginamos ląstelės yra panašios į poimplantacinio epiblasto ląsteles ir yra „apmokytos“ diferencijacijai pliuripotentiškumo būsenos. Šiomis dvejomis sąlygomis auginamos ląstelės yra geras modelis demetilinimo ir pliuripotentiškumo būsenų nagrinėjimui, kurių retesnių oksimC pokyčiai dar menkai tyrinėti.

Oksiduotos 5mC formos ir jų funkcijos mažai tirtos kitose filogenetinėse grupėse nei žinduoliai, nors TET baltymų homologai aptinkami įvairiuose organizmuose. Pavyzdžiui, *Agaricales* eilės papėdgrybiai *Laccaria bicolor* ir *Coprinopsis cinerea* turi labai daug TET baltymų homologų genų (atitinkamai 74 ir 47; Iyer ir kt., 2009, 2014). Šie laboratorijose naudojami modeliniai organizmai turi nedidelius (~37 Mb *C. cinerea* ir ~61 Mb *L. bicolor*), žinomas sekos genomus (Martin ir kt., 2008; Stajich ir kt., 2010), todėl yra patogūs tyrimams. *C. cinerea* DNR buvo aptiktos oksimC, nustatytas bendras oksimC bei 5hmC išsidėstymas genome (Chavez ir kt., 2014). Tačiau atskirų oksimC pasiskirstymas ir jų poveikis genų raiškai *C. cinerea* nebuvo išsamiai ištirtas, o *L. bicolor* oksimC dar visai neanalizuotas. Taigi *C. cinerea* ir *L. bicolor* yra įdomūs modeliniai organizmai siekiant suprasti galimas oksimC funkcijas ir TET baltymų reikšmę.

Siekiant išsiaiškinti dar iki galo nesuprastas oksimC funkcijas, reikia jautrių, nukleotido skiriamosios gebos genominių modifikacijų išsidėstymo nustatymo metodų. Dauguma oksimC tyrimams sukurtų didelės skiriamosios gebos metodų naudoja bisulfitinę sekoskaitą, tačiau ji turi trūkumų (Zhao ir kt., 2020). Šiuose metoduose nuskaitomas visas genomus, nors didžioji dalis jo sekų neturi citozino modifikacijų, dažniausiai esančių CG dinukleotide. Kadangi patikimam modifikavimo įvertinimui reikia didelio sekų nuskaitymų skaičiaus, didelių bioinformatinės analizės bei sekoskaitos apimčių, todėl šie

metodai brangūs ir nepraktiški retų modifikacijų nustatymui (Plongthongkum ir kt., 2014).

DNR modifikacijos tyrimų skyriuje genominiam citozino modifikacijų profiliavimui buvo sukurti alternatyvūs metodai, kurie užpildo spragą tarp ekonomiško ir skiriamosios gebos. Prijungto oligonukleotido nukreiptos sekoskaitos metodas (TOP-seq, angl. *tethered oligonucleotide-primed sequencing*; Staševskij ir kt., 2017) analizuoja nemodifikuotus genomo CG. Naudojama mutantinė metiltransferazė eM.SssI (angl. *engineered methyltransferase*) su išplėsta kofaktoriaus prisijungimo kišene, kuri kovalentiškai pažymi CG taikinius, perkeldama aktyvią azido grupę nuo sintetinio SAM kofaktoriaus analogo. Prie azido grupės kovalentiškai prijungiamas oligonukleotidas ir nuo jam komplementaraus pradmens polimerazė atlieka grandinės pratęsimą. Susiformuoja DNR grandinė, kurioje yra prijungta seka, taikinio CG ir gretimas genomo regionas. Atlikus sekoskaitą galima tiesiogiai nustatyti nemodifikuoto CG taikinio vietą nukleotido tikslumu ir neatliekant viso genomo sekoskaitos.

Kadangi eM.SssI atlieka 5caC dekarboksilinimą susidarant nemodifikuotam CG, šią reakciją galima sujungti su eM.SssI vykdomu kovalentiniu CG žymėjimu ir pritaikius TOP-seq būtų galima nustatyti 5caC turėjusius taikinius. 5caC yra rečiausia oksimC, todėl tik 5caC nustatantis metodas būtų daug pigesnis ir efektyvesnis nei metodai, kuriuose naudojama viso genomo sekoskaita.

Remiantis TOP-seq principu, ankstesniuose darbuose buvo sukurtas nukleotido skiriamosios gebos 5hmC genominio profiliavimo metodas – hmTOP-seq (angl. *5hmC-specific tethered oligonucleotide-primed sequencing*; Gibas ir kt., 2020). Šiuo atveju 5hmC pažymimas azido grupę turinčia gliukoze esant β-gliukoziltransferazei. Atlikus cheminę 5fC redukciją į 5hmC, šį metodą būtų galima pritaikyti 5fC nustatymui genome.

Šio darbo **tikslas** buvo ištirti metiltransferazių atliekamą 5caC dekarboksilinimo reakciją ir ja remiantis sukurti 5caC genominio profiliavimo metodą bei pritaikyti TOP-seq paremtus citozino modifikacijų genominio profiliavimo metodus oksimC pasiskirstymo ir funkcijų analizei.

### **Uždaviniai:**

1. Patvirtinti, kad žmogaus metiltransferazės DNMT3A/3L ir DNMT3B/3L atlieka 5caC dekarboksilinimo reakciją bei įvertinti fermentų aktyvumą esant fiziologinėms SAM koncentracijoms.

2. Pasitelkiant eM.SssI vykdomą 5caC dekarboksilinimo reakciją, sukurti ir įvertinti 5caC genominio profiliavimo metodą.
3. Sukurtu 5caC profiliavimo metodu ištirti genominių 5caC pasiskirstymą pelės embrioninėse kamieninėse ląstelėse.
4. Nustatyti *L. bicolor* ir *C. cinerea* papėdgrybių oksiduotų 5mC formų gausumą, pasiskirstymą bei naudojant TOP-seq paremtus genominio profiliavimo metodus ištirti jų galimas funkcijas.

### **Mokslinio darbo naujumas**

Sukurtas naujas CG skiriamosios gebos genominio 5caC profiliavimo metodas caCLEAR, kuris yra jautrus ir specifiškas. Kadangi caCLEAR nustato tik 5caCG taikinius, metodas suteikia ekonomišką 5caC profilių sudarymą – reikia mažesnių sekoskaitos ir bioinformatinės analizės apimčių, palyginus su kitais nukleotido skiriamosios gebos 5caC profiliavimo metodais, kuriuose atliekama viso genomo sekoskaita.

Pirmą kartą atliktas 5caC genominio pasiskirstymo palyginimas dviejose pelės embrioninių kamieninių ląstelių pluriipotentškumo būsenose – naivioje ir „apmokytoje“ (atitinkamai auginant serumo-2i ir serumo sąlygomis). Tarp dviejų būsenų buvo nustatyti žymūs 5caC praturtinimo genomo elementuose bei išsidėstymo svarbių veiksnių jungimosi regionuose skirtumai. 5caC praturtinimas ir dalis 5caC turinčių veiksnių jungimosi vietų (CTCF, c-Myc ir Nanog) tik serumo sąlygomis buvo susiję su „apmokytoje“ būsenoje aktyvesniais vystymosi procesų genais, o 2i sąlygomis sąsajos nustatytos su bendrų ląstelės funkcijų genais. Taip pat buvo pademonstruotas asimetrinis 5caC pasiskirstymas pelės embrioninių kamieninių ląstelių antiprasminėje aktyvių baltymų koduojančių genų grandinėje.

Sukurtas retos RNR modifikacijos, 5-karboksilcitozino (5carC), aptikimą masių spektrometrijos būdu palengvinantis metodas, kuriame naudojama cheminė derivatizacija. *L. bicolor* ir *C. cinerea* papėdgrybių RNR pirmą kartą nustatytos oksimC, ištirtas 5carC bei 5hmC gausumas skirtingose RNR frakcijose.

Sukurtas naujas nukleotido tikslumo 5fC genominio profiliavimo metodas foTOP-seq, kuris nustato tik 5fC taikinius be viso genomo sekoskaitos. *L. bicolor* DNR buvo pirmą kartą nustatytos ir kiekybiškai įvertintos oksimC. Pirmą kartą nustatytas *L. bicolor* 5hmC ir 5fC genominis išsidėstymas nukleotido tikslumu. Taip pat atliktas išsamus *L. bicolor* ir *C. cinerea* 5mC, 5hmC ir 5fC modifikacijų įtakos genų raiškai įvertinimas.

Šie atlikti darbai suteikė daugiau žinių apie oksimC paplitimą, pasiskirstymą ir funkcijas bei buvo sukurti nauji metodai jų tyrimui.

### **Ginamieji teiginiai:**

1. DNMT3A/3L ir DNMT3B/3L žmogaus metiltransferazės atlieka 5caC dekarboksilinimą, kuris yra slopinamas SAM.
2. Sukurtas jautrus CG skiriamosios gebos metodas, specifiskai nustatantis genominius 5caCG taikinius.
3. Serumo-2i ir serumo sąlygomis augintos pelės embrioninės kamieninės ląstelės ženkliai skiriasi 5caC praturtinimu genomo elementuose bei išsidėstymu svarbių veiksmių jungimosi regionuose. Serumo sąlygomis 5caC praturtintas „apmokytoje“ būsenoje aktyvesniuose vystymosi procesų genuose. 5caC susijęs su aktyvia raiška ir yra asimetriškai išsidėstęs antiprasminėje aktyvių baltymų koduojančių genų grandinėje.
4. *L. bicolor* ir *C. cinerea* RNR turi oksimC. Sukurtas 5scarC aptikimą RNR masių spektrometrijos būdu palengvinantis metodas, pagrįstas chemine derivatizacija. 5scarC yra abiejų grybų 3'-poli A sekų neturinčiose ilgesnėse nei 200 nt RNR, *L. bicolor* poli A RNR bei *C. cinerea* trumpesnėse nei 200 nt RNR, tarp kurių 5scarC gausiausias yra už tRNR (~76-90 nt) trumpesnėse RNR.
5. *L. bicolor* ir *C. cinerea* DNR yra visos trys oksimC, iš kurių gausiausias 5hmC ir 5fC. Nustatytas 5hmC ir 5fC genominis išsidėstymas hmTOP-seq ir sukurtu foTOP-seq metodu. 5hmC ir 5fC yra praturtinti judriuosiuose elementuose, šalia kurių esančių genų raiška yra slopinama. *C. cinerea* gausus 5hmCG yra susijęs su aktyviais genais esant mažai 5mCG, o *L. bicolor* gausūs 5fCG/5hmCG susiję su slopinamais genais, nepriklausomai nuo jų 5mCG lygio.

# 1. LITERATŪROS APŽVALGA

## 1.1 DNR modifikacijos

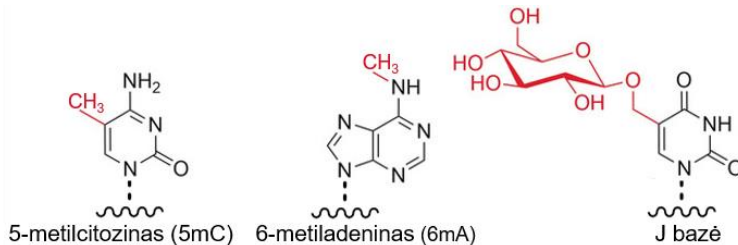
Virusų, bakterijų, archėjų bei eukariotų genomuose be 4 pagrindinių DNR heterociklinių bazių G, A, T ir C, yra aptikta daugiau nei 17 įvairių modifikuotų bazių (Raiber ir kt., 2017). Bakteriofagų genomuose aptiktos retos modifikuotos adenino bazės, kaip 2-aminoadeninas ir N6-karbamoilmetiladeninas. Bakteriofaguose taip pat yra įvairių timino modifikacijų, kurios dažnai susideda iš didelių grupių, prijungtų prie C5 padėties, pvz.:  $\alpha$ -putrescinitiminas, 5-dihidroksipentiluracilas ir  $\alpha$ -glutamiltiminas. Tokios hipermodifikacijos gali pakeisti didelę genomo timino dalį. Manoma, kad jos apsaugo nuo infekuotų bakterijų išskiriamų restrikcijos endonukleazių, o taip pat gali palengvinti DNR pakavimą į viruso apvalkalą (Hofer ir kt., 2019).

Bakterijose ir archėjose susidaro N6-metiladeninas (6mA), N4-metilcitozinas (4mC) arba 5-metilcitozinas (5mC). Pirmiausia šios mikroorganizmų modifikacijos yra svarbus gynybinis mechanizmas, sudarytas iš restrikcijos-modifikacijos sistemų ir apsaugantis nuo virusų. DNR metilinimas vyksta sekai specifinėse vietose ir apsaugo nuo tą pačią seką atpažįstančios restrikcijos endonukleazės kirpimo (de Mendoza ir kt., 2019). Unikalus DNR metilinimo išsidėstymas (metilintų sekų derinys) yra kaip bakterijų rūšies „kodas“, pagal kurį restrikcijos-modifikacijos sistema atskiria svetimą DNR nuo savos (Kriukienė ir kt., 2012). 6mA tyrimai bakterijų DNR atskleidė jo svarbą ir daugeliui kitų procesų. Jis yra būtinas kai kurių bakterijų gyvybingumui, nes yra svarbus DNR replikacijai bei chromosomų segregacijai. Taip pat kai kuriose bakterijose 6mA gali būti svarbus nukleoido išsidėstymui, naujai susintetintų grandinių atpažinimui, reparacijai, transkripcijos reguliacijai (Breiling ir Lyko, 2015; Hofer ir kt., 2019).

Eukariotiniuose organizmuose aptikta keletas įvairių modifikacijų, kai kurios gali būti būdingos tik tam tikrai filogenetinei grupei. Pavyzdžiui, *Trypanosoma brucei* ir kituose kinetoplastiduose aptinkamos dvi timino modifikacijos su reguliacine funkcija – 5-hidroksimetiluracilas (5hmU) ir  $\beta$ -d-gliukozil-hidroksimetiluracilas (J bazė) (1.1 pav.). J bazę jungiantys baltymai JBP1 ir JBP2 (angl. *J binding protein*) oksiduoja timino metilo grupę susidarant 5-hidroksimetiluracilui (5hmU), o kai kurių 5hmU hidroksilo grupė vėliau glikozilinama J-gliukoziltransferazės susidarant J bazei (Pastor ir kt., 2013). Nustatyta, kad bazė J yra praturtinta telomeriniuose pasikartojimuose ir nutildytose genomo srityse bei stabdo policistroninių transkriptų ilgėjimą už transkripcijos sustojimo vietų (Raiber ir kt., 2017).



Neseniai buvo nustatyta, kad N6-metiladeninas (6mA) (1.1 pav.) yra įvairiose eukariotų filogenetinėse grupėse aptinkama DNR modifikacija, galinti reguliuoti genų raišką (Breiling ir Lyko, 2015). Daugėja 6mA buvimo ir funkcijos įrodymų žinduolių genomuose, tačiau norint išsiaiškinti kai kurias nesutapimus ir suprasti pagrindinius mechanizmus, reikia tolesnių tyrimų (Hofer ir kt., 2019).



**1.1 pav.** Kai kurios eukariotiniuose organizmuose aptinkamos modifikuotos DNR bazės.

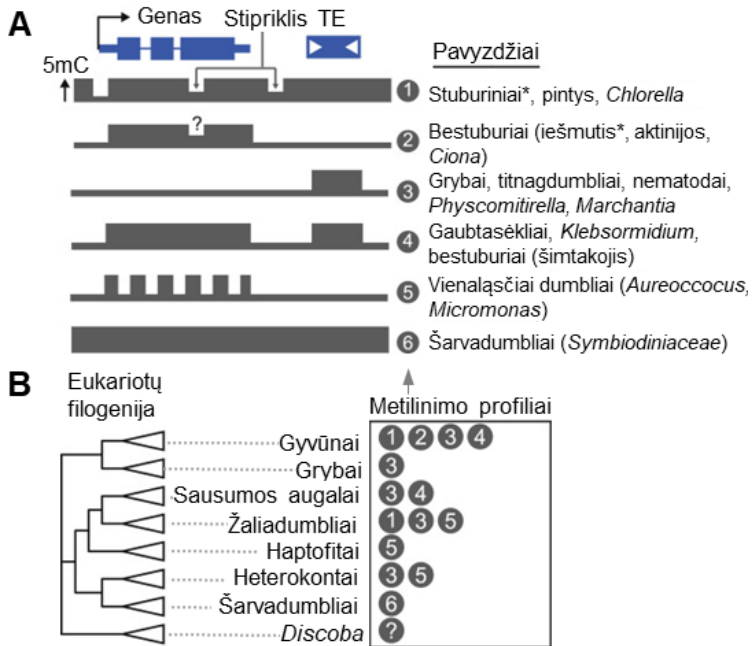
5-metilcitozinas yra vyraujanti eukariotų DNR modifikacija (1.1 pav.), evoliuciškai konservatyvi daugelyje organizmų (Lu ir kt., 2015a). Ji daro didelį poveikį genų raiškai, vystymuisi ir genomo vientisumo palaikymui (Wu ir Zhang, 2014). DNR metilinimas vyksta didžiajame DNR spiralės griovyje ir gali būti atpažįstamas kaip sterinis signalas besijungiantiems baltymams (Kriukienė ir kt., 2012). Eukariotų genomuose cheminės modifikacijos, pridamos prie DNR ar chromatinio baltymų histonų uodegų, veikia kaip pagrindiniai epigenetiniai ženklai, nepakeičiantys koduojančios sekos ir gali būti perduodami po ląstelės pasidalijimo (Suzuki ir Bird, 2008). Išsaugodamos pradinę genetinę informaciją, epigenetinės modifikacijos reguliuoja genų raišką ir suteikia ląstelių tipui specifinę funkciją per sąveikas su modifikuojančiais fermentais, slopikliais, chromatiną pertvarkančiais ir transkripcijos baltymais (Skvortsova ir kt., 2018).

### 1.1.1 5mC pasiskirstymas ir funkcijos eukariotuose

5-metilcitozinas buvo bene pirmoji atrasta ir geriausiai ištirta DNR modifikacija (Raiber ir kt., 2017). 5mC yra plačiai paplitęs ir aptinkamas daugelyje eukariotų genomų, įskaitant grybus, augalus ir gyvūnus (Deniz ir kt., 2019). Daugialąsčiuose eukariotuose 5mC yra bene plačiausiai taikoma genomo judriųjų elementų (TE, angl. *transposable elements*) slopinimo strategija (de Mendoza ir kt., 2019). Manoma, kad poreikis palaikyti TE nutildymą, paskatino raišką slopinančio DNR metilinimo mechanizmo evoliuciją, kuris vėliau buvo nukreiptas kitų funkcijų atlikimui (Deniz ir kt., 2019). Tačiau nemažai eukariotų (tarp jų modeliniai organizmai *Drosophila*

*melanogaster*, *Saccharomyces cerevisiae* ir *Caenorhabditis elegans*) neturi 5mC, kuris nėra būtinas jų vystymuisi bei ląstelių diferenciacijai (Breiling ir Lyko, 2015). Vadinasi, spaudimas išlaikyti 5mC TE slopinimui skirtingose evoliucijos šakose varijavo. Vienas iš galimų šių skirtumų paaiškinimų – didelių populiacijų rūšys gali veiksmingai pašalinti silpnai žalingus TE įsiterpimus ir todėl nėra priklausomos nuo DNR metilinimo (Deniz ir kt., 2019).

Eukariotinis 5mC dažniausiai randamas simetriškų CG dinukleotidų (taip pat vadinamų CpG – citozinas-fosfatas-guaninas) kontekste. 5mC yra mutageniškas, nes gali įvykti spontaniškas deaminavimas ir C pakinta į T. Todėl organizmai, turintys CpG metilinimą, genome taip pat turi mažesnę CpG kiekį (Greenberg ir Bourc’his, 2019). Tuo tarpu metilinimas ne CpG kontekste sudaro mažesnę organizmų genomo metilinimo dalį ir yra aptinkamas daugiausia augaluose bei kai kuriuose grybuose (Zemach ir kt., 2010). Gyvūnuose ne CpG metilinimas buvo pastebėtas tik keliais atvejais – žinduolių pluriipotentinėse ląstelėse ir smegenų audiniuose (Guo ir kt., 2014b; Lister ir kt., 2009).



**1.2 pav.** 5-metilcitozino išsidėstymas įvairių eukariotinių organizmų genomuose. **(A)** Pagrindiniai iki šiol aprašyti DNR metilinimo profiliai su jų tipinėmis rūšimis. Žvaigždutės žymi filogenetines grupes, kuriose buvo aprašytas stipriklių demetilinimas. **(B)** Eukariotų filogenetiniai ryšiai pagal dabartinį sutarimą ir kiekvienoje grupėje aptinkami DNR metilinimo profiliai (nurodyti skaičiais ir pavaizduoti (A)). Pagal de Mendoza ir kt. (2019).

Genominis 5mC išsidėstymas ir gausumas skirtingose eukariotų filogenetinėse grupėse gana įvairus, todėl 5mC funkcijos gali varijuoti (de Mendoza ir kt., 2019). Dažnai sutinkamas mozaikiškas metilinimo išsidėstymas (1.2 pav. A, 2-5 profiliai), kai stipriai metilintos DNR regionai keičiasi su nemetilintais ir yra metilinta mažoji genomo CpG dalis (Suzuki ir Bird, 2008). Mozaikiškas metilinimas būdingas daugeliui bestuburių organizmų ir grybų, taip pat aptinkamas kai kuriuose augaluose (Suzuki ir Bird, 2008). Kai kurių augalų genomuose metilinimas nukreiptas tik į TE (pvz., samanų), o kiti turi ir genų metilinimą (pvz., *Arabidopsis thaliana*) (de Mendoza ir kt., 2019). Tuo tarpu daugelio vabzdžių genomuose tarpgeniniai TE nemetilinimi, o 5mC aptinkamas kai kurių genų egzonuose (Breiling ir Lyko, 2015). Grybų genomuose 5mC metilinimas dažniausiai nukreiptas į TE, o genai metilinti mažai. Nemažai grybų rūšių aptinkamos aukšto metilinimo lygio 5mC sankaupos – ilgi (iki 100 kb) gretimi ruožai, kuriuose gali būti ir genų (Bewick ir kt., 2019; Zemach ir kt., 2010). Kai kuriuose vienaląsčiuose dumbliuose nustatytas 5mC išsidėstymas transkribuojamuose genuose tarp nukleosomų (de Mendoza ir kt., 2019).

Metilinimas taip pat gali būti globalus (1.2 pav. A 1 ir 6 profiliai), kai didžioji dalis genomo CpG yra metilinta (Breiling ir Lyko, 2015). Augaluose didelė dalis genomo (>50%) gali būti metilinama slopinant labai gausius TE (pvz., kukurūzuose), kurie dominuoja tarpgeniniuose regionuose (Suzuki ir Bird, 2008). Globalus 5mC metilinimas aptiktas pintyse ir žaliuosiuose vienaląsčiuose dumbliuose, kuriuose tik kai kurie regionai lieka nemetilinti – dažniausiai promotoriai, o šarvadumbliuose metilinami ir jie (de Mendoza ir kt., 2019). Globalus metilinimas būdingas stuburinių genomams, kuriuose trumpi nemetilinti regionai aptinkami daugiausia aktyvių genų reguliaciniuose elementuose – promotoriuose ir stiprikliuose (angl. *enhancer*, genomo elementai, kurie, pritraukdami baltymų kompleksus į promotorius, skatina transkripciją) (Breiling ir Lyko, 2015).

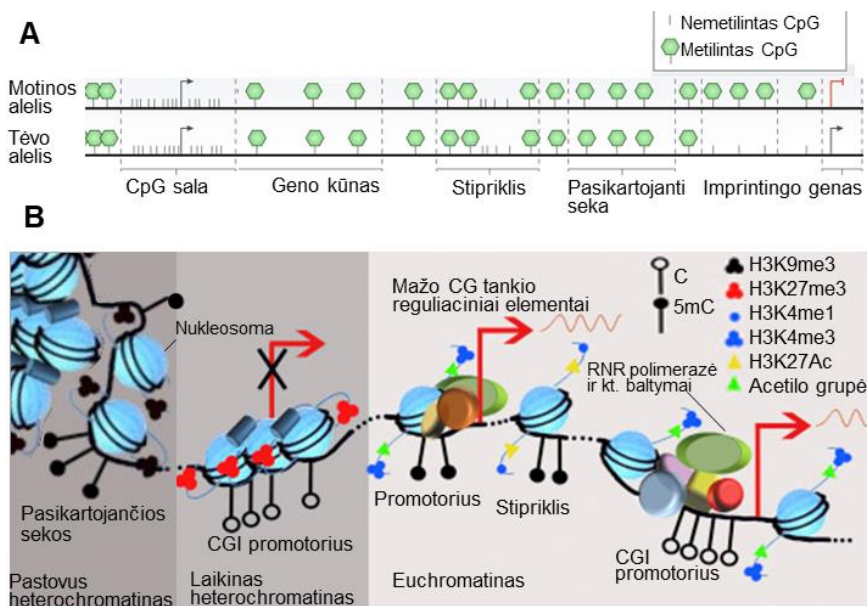
Augaluose ir gyvūnuose 5mC aptinkamas aktyvių genų kūnuose, dažniausiai tuose, kurių raiška yra pastovi. Ne CG metilinimas augalų genuose paprastai yra susijęs su raiškos slopinimu, tai pastebėta ir žinduolių nervinėse ląstelėse (Kumar ir kt., 2018; de Mendoza ir kt., 2019). Didžiojoje dalyje grybų geninis 5mC galbūt nedaro tiesioginio poveikio arba slopina genų raišką – daugumos stipriai metilintų genų raiškos lygis yra žemiausias (Bewick ir kt., 2019). Tačiau aptikta ir teigiama grybų genų metilinimo koreliacija su genų raiška (Bewick ir kt., 2019; Zemach ir kt., 2010).

Daugumos gyvūnų genų promotoriai dažniausiai yra nemetilinti. Nemetilinti promotoriai dažnai yra praturtinti H3 histono 4 lizino trigubo

metilinimo (H3K4me3) modifikacija, kuri susijusi su aktyviais genais (1.3 pav. B). Nemetilintuose promotoriuose dažnai aptinkamos CpG sankaupos (1.3 pav. A) – CpG salos (CGI, angl. *CpG island*), kurios dažniausiai randamos vystymosi ir „namų ruošos“ (pagrindines ląstelės funkcijas užtikrinantys, pastovios raiškos genai, angl. *housekeeping*) genų promotoriuose (Parry ir kt., 2021). Dauguma stuburinių CGI būna nemetilintos, nepriklausomai nuo geno raiškos (de Mendoza ir kt., 2019). Taigi mažai metilintas promotorius ir jo CpG turtingumas gali paskatinti raišką, bet nereiškia aktyvios transkripcijos. Nedidelė dalis CGI (~10%) yra skirtingai metilinos tarp audinių (Lee ir kt., 2014). 5mC kiekis stuburinių stiprikliuose pasižymi neigiama koreliacija su genų aktyvumu. Stiprikliai yra tos genomo vietos, kuriose vyksta didžiausi 5mC pakitimai vykstant embrioniniam vystimuisi ir ląstelių diferenciacijai (de Mendoza ir kt., 2019). Ląstelių tipui būdingi transkripcijos faktoriai prisijungdami gali sukelti šių regionų DNR metilinimo sumažėjimą, taip skatindami susijusių genų raišką (An ir kt., 2017).

Žinduolių genome metilinimas 5-oje citozino padėtyje yra gausiausia DNR modifikacija, kuri taip pat vadinama „penktąja baze“ – ~4-5% visų citozinių yra metilinos. Globalus citozino metilinimo lygis įvairiuose suaugusių žinduolių audiniuose dažniausiai yra labai panašus (Greenberg ir Bourc'his, 2019). Dauguma žinduolių CpG vietų (70-80%) yra stipriai metilintos ir dažniausiai aptinkamos mažo CpG tankio TE, satelitinuose pasikartojimuose, geniniuose ir tarpgeniniuose regionuose (1.3 pav. A). Tuo tarpu beveik nemetilinta nedidelė CpG dalis (<20%) daugiausia aptinkama didelio CpG tankio regionuose, įskaitant promotorių CGI (1.3 pav. B) (An ir kt., 2017; Parry ir kt., 2021). Genų tolimieji reguliaciniai elementai, pvz., stiprikliai, turi vidutinį CpG tankį ir labai skirtingus metilinimo lygius atskiruose ląstelių tipuose (Parry ir kt., 2021).

Citozino metilinimas žinduolių genome yra esminė epigenetinė modifikacija, atliekanti svarbias funkcijas įvairiuose ląstelės procesuose (Wu ir Zhang, 2014). 5mC pasiskirstymas yra susijęs su ląstelei specifine transkripcijos būkle, todėl 5mC pokyčiai yra svarbūs vystymosi procesuose, pluriipotencijos palaikyme, ląstelės atminties susidaryme, senėjime ir ligų atsiradime (Parry ir kt., 2021). Vėžinėse ląstelėse DNR metilinimas labai išreguliuojamas – būdingas visuotinis metilinimo sumažėjimas, kuris tikriausiai pažeidžia genomo vientisumą. Taip pat aptinkamas lokalus CGI promotorių hipermetilinimas, susijęs su nenormalia auglio slopinimo arba reparacijos genų raiška (An ir kt., 2017).



**1.3 pav.** 5-metilcitozinas ir histonų modifikacijos stuburiniuose organizmuose (A) Stuburinių DNR metilinimas aptinkamas daugiausia CpG kontekste ir dauguma jų yra metilinti. Promotoriuose dažnai aptinkami tankiai išsidėstę CpG regionai – CpG salos (CGI), kurios dažniausiai nemetilintos. Aktyvūs stiprikliai (transkripciją skatinantys reguliaciniai elementai) pasižymi mažesniu metiliniu. Imprintingo genai aktyvūs tik viename iš tėvų alelių ir genas reguliuojančiuose regionuose turi skirtingą metilinimą. Pagal Skvortsova ir kt. (2018). (B) DNR metilinimas ir chromatinio struktūra. Kompaktiškas pastoviai neaktyvus chromatinas (heterochromatinas) pasižymi stipriu DNR metiliniu ir pasikartojančiomis sekomis bei į nukleosomas įeinančio 3 histono uodegos 9 lizino trigubu metiliniu (H3K9me3). Laikiniai neaktyvus chromatinas CGI promotoriuose slopinamas H3K27me3 modifikacija. Aktyviame laisvesnės struktūros chromatine (euchromatine) aptinkami mažai metilinti mažo CpG tankio reguliaciniai elementai (promotoriai ir stiprikliai) arba nemetilinti CGI promotoriai, pasižymintys parodytomis genų raišką skatinančiomis histonų modifikacijomis. Ac – acetilinas. Pagal Zampieri ir kt. (2018).

Citozino metilinimas neabejotinai yra susijęs su negrįžtamu daugelio genų slopinimu, kuris dažniausiai vyksta metilinant CGI. 5mC dalyvauja lytinėms ląstelėms specifinių genų slopinime bei labai svarbus TE slopinimui, užtikrinančiam genomo stabilumą (de Mendoza ir kt., 2019). 5mC dalyvauja X chromosomos nutildyme. Veikiant nekoduojančiai XIST RNR, kiekvienoje moteriškos lyties žinduolių ląstelėje atsitiktinai nutildoma po vieną X chromosomą. Šiame procese X chromosomos CGI metilinimas veikia kaip galutinis užraktas, svarbus negrįžtamam genų nutildymui (Greenberg ir Bourc’his, 2019). 5mC svarbus genų imprintingui, kuriame daugiau nei šimtas autosominių chromosomų genų turi skirtingą reguliacinių elementų metilinimą ir raišką priklausomai nuo to, ar genas paveldėtas iš motinos ar iš

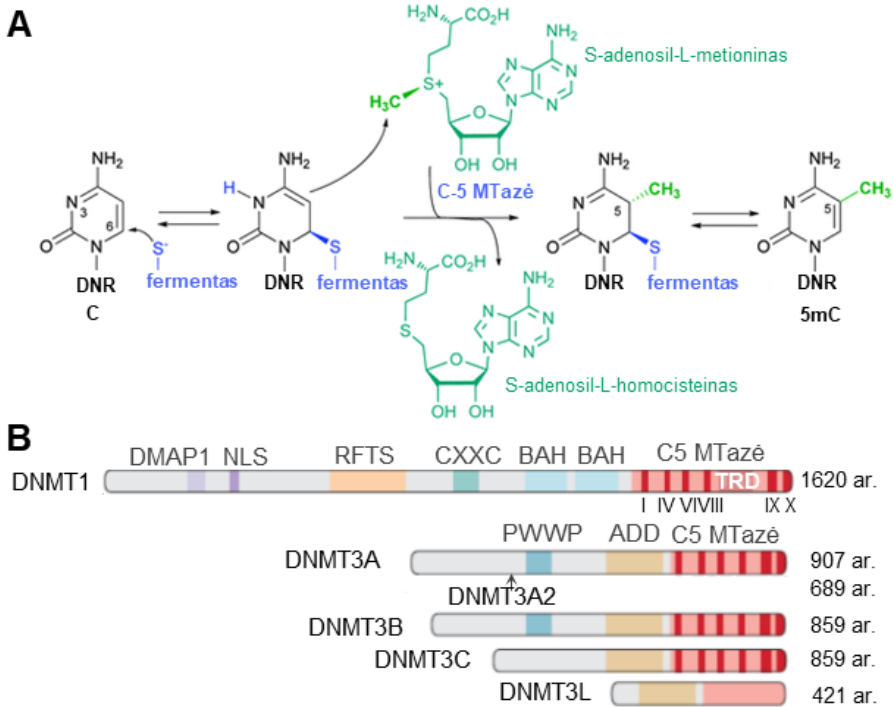
tėvo (Tajima ir kt., 2016). Tuo tarpu dauguma 5mC neturinčių neaktyvių CGI promotorių slopinami PRC2 komplekso vykdomo H3K27 metilinimo (1.3 pav. B), kuris yra lengviau keičiamas, kai reikia aktyvuoti CGI turinčius genus priklausomai nuo aplinkos veiksnių (Greenberg ir Bourc'his, 2019).

5mC yra labai reikšmingas audiniams specifinių genų raiškos ir chromatinos struktūros reguliatorius (Raiber ir kt., 2017). Tačiau vis dar nėra iki galo išaiškinta, kaip promotorių ar stipriklių metilinimas sukelia transkripcijos slopinimą, nes pats savaime 5mC nesukelia nutildymo (Greenberg ir Bourc'his, 2019). DNR ir histonų modifikacijos (1.3 pav. B) sąveikauja su įvairiais chromatinos komponentais ir pertvarkymo baltymais kartu keisdami genų raišką (Raiber ir kt., 2017).

5mC geba atstumti transkripcijos faktorius (TF), sukeldamas sterinius trukdžius jų jungimosi vietose ir taip slopindamas genų raišką (Greenberg ir Bourc'his, 2019). Tačiau didelė dalis TF stipriau jungiasi prie metilintos DNR (de Mendoza ir kt., 2019). 5mC gali paskatinti jį atpažįstančių baltymų prisijungimą, kurie prie tikslinių promotorių gali pritraukti papildomus genų raišką slopinančius baltymų kompleksus (Breiling ir Lyko, 2015). 5mC atpažįsta metil-CpG jungiantį domeną turintys baltymai (MBD, angl. *methyl-CpG-binding domain proteins*). MBD jungimasis prie 5mC siejamas su kompaktiškos, genus slopinančios struktūros susidarymu, vadinamos heterochromatinu (1.3 pav. B) (de Mendoza ir kt., 2019). Visi MBD sąveikauja su nukleosomų pertvarkymo ir histono deacetilazių kompleksais (Sin3, NurD, NCoR/SMRT), kurie skatina transkripcijos slopinimą jungiamuose regionuose (Greenberg ir Bourc'his, 2019).

### 1.1.2 Eukariotų DNR metiltransferazės

DNR metilinimas katalizuojamas veikiant DNR metiltransferazių fermentams (MTazėms), kurie perkelia metilo grupę nuo gausaus S-adenozil-L-metionino (SAM) kofaktoriaus susidarant metilintam produktui ir S-adenozil-L-homocisteinui (SAH) (Goll ir Bestor, 2005). DNR citozino-C5 metiltransferazių vykdomos kovalentinės katalizės metu (1.4 pav. A) bazė išsukama iš DNR duplekso (Klimašauskas ir kt., 1994), susidaro tarpinis kovalentinis junginys tarp fermento katalitinio centro nukleofilinio cisteino liekanos ir citozino žiedo C6 atomo ir tuomet C5 tampa pakankamai nukleofiliškas metilinimo reakcijai įvykti (Kriukienė ir kt., 2012).



**1.4 pav. (A)** DNR metiltransferazių katalizuojamas citozino-5 metilinimas. Fermentiniam metilo grupės pernešimui nuo kofaktoriaus S-adenozil-L-metionino (SAM) reikalingas kovalentinis tarpinio junginio susidarymas tarp fermento cisteino liekanos ir citozino žiedo C6 atomo. Pagal Kriukienė ir kt. (2012). **(B)** Žinduolių citozino-5 metiltransferazių struktūros. DNR metilinimo palaikymo metiltransferazė DNMT1 turi kelis konservatyvius domenus, svarbius jungimuisi prie DNR: DMAP1 baltymą jungiantį domeną, į replikacijos vietas nukreipiančios sekos RFTS domeną ir CpG jungiantį CXXC domeną. NLS – branduolio lokalizacijos signalas. Tiksli bromo homologijos (BAH) domenų funkcija neaiški. Metiltransferazės katalitinis domenas (C5 MTazė) yra baltymo C-gale. TRD – taikinio atpažinimo domenas. *De novo* DNR metiltransferazėse DNMT3A, 3A2 (pradžią pažymėta rodykle) ir D3B yra H3K4 jungiantis ADD domenas ir H3K36me3 jungiantis PWWP domenas bei metiltransferazės katalitinis domenas C-gale. DNMT3C raiška aptinkama pelių patinėlių embrionų lytinėse ląstelėse, šio baltymo genas yra DNMT3B duplikacija, praradusi PWWP domeną. DNMT3L yra neaktyvus baltymas, stimuliuojantis DNMT3A ir DNMT3B katalitinį aktyvumą. Nurodytas baltymų ilgis aminorūgštimis (ar.). Pagal Greenberg ir Bourc’his (2019).

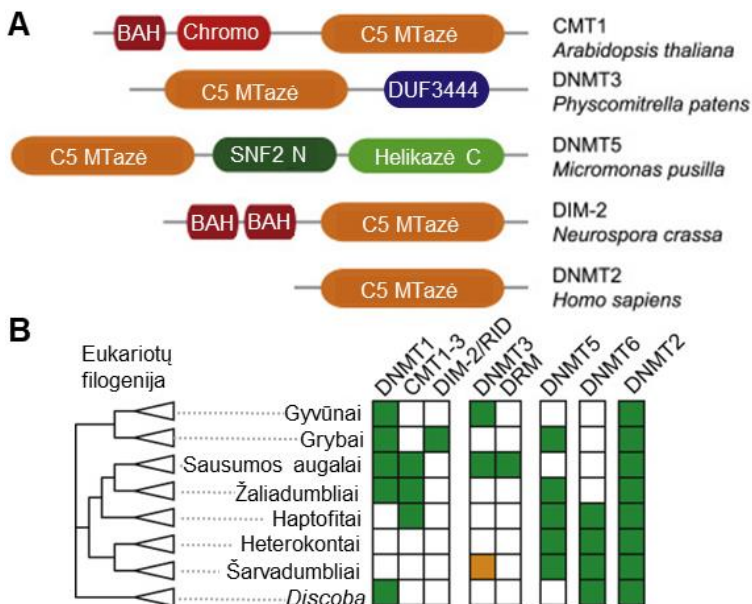
Eukariotinės citozino metiltransferazės yra kilusios iš prokariotų ir pasižymi dideliu katalitinio domeno sekos ir struktūros panašumu į bakterijų metiltransferazes. Šis domenas susideda iš 10-ties sekų motyvų, kuriuose yra SAM jungimosi sritis ir aktyvusis centras (1.4 pav. B). Tarp VIII ir IX motyvų yra mažiau konservatyvus taikinio atpažinimo regionas (TRD), kuris bakterijose yra atsakingas už DNR sekos atpažinimo specifiškumą (Goll ir Bestor, 2005).

Eukariotų MTazės skirstomos į dvi dideles kategorijas – palaikomojo ir *de novo* metilinimo (1.6 pav. A). Žinduolių ankstyvosios embriogenezės metu DNMT3A ir DNMT3B naujai sukuria metilinimo žymes ant nemetilintų CpG, todėl yra klasifikuojamos kaip *de novo* metiltransferazės. O DNMT1, palaikomojo metilinimo metiltransferazė, yra atsakinga už jau esančio citozino metilinimo paveldėjimą po DNR replikacijos dalijantis ląstelėms (An ir kt., 2017). Replikacijos metu UHRF1 baltymas atpažįsta hemimetilintus CpG ir prijungia DNMT1, kad ant naujai besiformuojančios grandinės būtų atkurtas buvęs metilinimo profilis (Greenberg ir Bourc’his, 2019).

Žinduolių DNMT1 (1.4 pav. B) baltymą sudaro keli domenai (1.4 pav. B): DMAP1 jungimosi domenai (dar vadinamas nepriklausomai sulankstytu N-galo domenu, NTD), į replikacijos vietas nukreipiančios sekos (RFTS) domenai, CXXC motyvas, du gretimi bromo homologijos (BAH) domenai ir katalitinis MTazės domenai baltymo C-gale (Greenberg ir Bourc’his, 2019; Tajima ir kt., 2016). DMAP1 baltymas jungiasi prie DNMT1 ir replikacijos vietose slopina transkripciją prijungdamas histonų deacetilazę HDAC2. CXXC (cisteinas-X-X-cisteinas) domenai sąveikauja su nemetilintais CpG. RFTS domenai jungiasi su UHRF1 ir tada atidengiamas katalitinis centras, kurį RFTS užstoja (Tajima ir kt., 2016).

Eukariotų citozino-5 MTazės yra skirstomos į penkias pagrindines grupes (1.5 pav. B), atsižvelgiant į domenų konfigūraciją ir DNR metiltransferazės domeno filogenetinius ryšius (de Mendoza ir kt., 2019). Geriausiai apibūdinti DNMT1 panašūs fermentai, nes jų yra gyvūnuose, augaluose ir grybuose, o jų pagrindinė palaikomojo metilinimo funkcija yra gerai žinoma (Goll ir Bestor, 2005). DNMT1 artimos fermentų formos išsivystė dėl įvykusios duplikacijos ir pakitimų chloroplastus turinčiuose organizmuose. DNMT1 kopija augaluose vadinama MET1, o jos išvestinė kopija davė pradžią CMT grupei (1.5 pav. A), kurioje yra įgytas histonų (H3K9me) atpažinimo chromodomenai. Grybų DNR citozino-5 MTazės RID ir DIM-2 (taip pat žinomos kaip DNMT4) taip pat yra kilę DNMT1 geno duplikacijos, kuris įgijo naujas grupei būdingas domenines struktūras (1.5 pav. A) ir funkcijas (de Mendoza ir kt., 2019).





**1.5 pav.** Eukariotų citozino-5 metiltransferazės. **(A)** Įvairių metiltransferazių klasių domeninė struktūra. BAH – bromo homologijos domenas. C5 MTazė – metiltransferazių katalitinis domenas. SNF2N – su chromatiną pertvarkančiu SNF2 baltymu sąveikaujantis domenas. Chromo – chromodomenas, sąveikaujantis su histonais. DUF – nežinomos funkcijos domenas. **(B)** Kairėje pavaizduoti eukariotų filogenetiniai ryšiai pagal dabartinį sutarimą. Viršuje surašyti pagrindinių eukariotų grupių MTazės. Žali kvadratai žymi fermentus, nustatytus bent viename grupės naryje. Jei fermentų neaptikta, tai pažymėta baltais kvadratais. Oranžinis kvadratas rodo retrotranspozono dalyje koduojamus fermentus. Pagal de Mendoza ir kt. (2019).

Kita pagrindinė eukariotinių C5 MTazių šeima yra DNMT3 fermentai, atsakingi už *de novo* metilinimą. Žinduolių DNMT3 šeimą (1.4 pav. B) sudaro DNMT3A, DNMT3A2, DNMT3B, DNMT3C ir DNMT3L (Pennings ir kt., 2019). DNMT3A, 3A2 (trumpesnė 3A izoforma, aptinkama vyriškų lytinių ląstelių pirmtakuose) ir 3B C-gale yra MTazės domenas, o N-gale yra reguliaciniai domenai – PWWP bei ADD, kurie sąveikauja su chromatinu (Goll ir Bestor, 2005; Greenberg ir Bourc’his, 2019). PWWP (Pro-Trp-Trp-Pro) domenas jungiasi su DNR ir nukreipia į chromatiną regionus per sąveikas su įvairiais baltymais, pvz.: pericentrinio heterochromatino CENP-C baltymu, taip pat sąveikauja su H3K36me3 histonų uodegomis, kurios aptinkamos aktyvių genų kūnuose (Gagliardi ir kt., 2018; Tajima ir kt., 2016). Į augalų homeodomeną (PHD) panašus ADD (ATRX-DNMT3-DNMT3L) cinko pirštų domenas yra turtingas Cys liekanomis ir sąveikauja su nemetilintomis H3K4 histonų uodegomis (per šią sąveiką atsilaivina ADD užstojamas MTazės domenas) bei įvairias su genų slopinimu susijusiais baltymais (Greenberg ir Bourc’his, 2019; Tajima ir kt.,

2016). Neseniai graužikų genomuose aptikta DNR metiltransferazė DNMT3C, kuri metilina evoliuciškai jaunų retrotranspozonų promotorius vyriškų lytinių ląstelių linijoje (Raiber ir kt., 2017). DNMT3C nėra PWWP domeno, tačiau ji turi katalitinį metiltransferazės domeną (Pennings ir kt., 2019). DNMT3L yra PWWP domeno neturintis ir katalitiškai neaktyvus baltymas, kuris sąveikauja su DNMT3A ir DNMT3B, stimuliuoja jų aktyvumą ir yra būtinas retrotranspozonų metilinimui bei slopinimui vystantis lytinėms ląstelėms (Greenberg ir Bourc'his, 2019).

DNMT3 artimų fermentų yra augaluose, gyvūnuose ir keliose kitose eukariotų filogenetinėse grupėse (1.5 pav. B). Sausumos augalai ir jų dumblių giminaičiai (streptofitai) turi pakitusį DNMT3 paralogą DRM, kurios metiltransferazės domenas pertvarkytas (de Mendoza ir kt., 2019). DNMT3 taip pat yra plikasėklių augalų genomuose, o tuo tarpu DRM yra daug plačiau paplitusi sausumos augaluose, vadinasi, evoliucijos metu DNMT3 tapo nereikalinga. DNMT3 neturi bendros domenų architektūros, o struktūra, turinti PWWP ir ADD domenus, iki šiol aptinkama tik gyvūnų karalystėje (de Mendoza ir kt., 2019).

Be DNMT1 ir DNMT3 yra ir kitų eukariotų C5 MTazių šeimų (1.5 pav. B). DNMT2 yra struktūriškai homologiška kitoms eukariotų C5 MTazėms, nors neturi reguliacinių domenų. Tačiau DNMT2 metilina tRNR ir rūšių, kuriose yra tik DNMT2, DNR nėra 5mC (Pennings ir kt., 2019). DNMT5 aptinkama skirtingose eukariotų filogenetinėse grupėse, įskaitant grybus, haptofitus, heterokontus ir chlorofitus (de Mendoza ir kt., 2019). Panašiai kaip ir DNMT1, DNMT5 turi ilgesnę konservatyvią dalį, kurioje yra su chromatiną pertvarkančiu SNF2 baltymu sąveikaujantis domenas. Apibūdinta DNMT6 grupė taip pat yra gana paplitusi (1.5 pav. B) ir turi tik MTazės domeną, kuriame nėra žinomų chromatino sąveikos domenų. Kadangi tirtų DNMT6 turinčių rūšių DNR 5mC neaptiktas, galbūt šis fermentas susijęs su RNR metilinimu. Taip pat grybuose buvo aprašyta DNMTX *de novo* MTazių šeima, kurios ryšys su kitomis šeimomis neaiškus (de Mendoza ir kt., 2019).

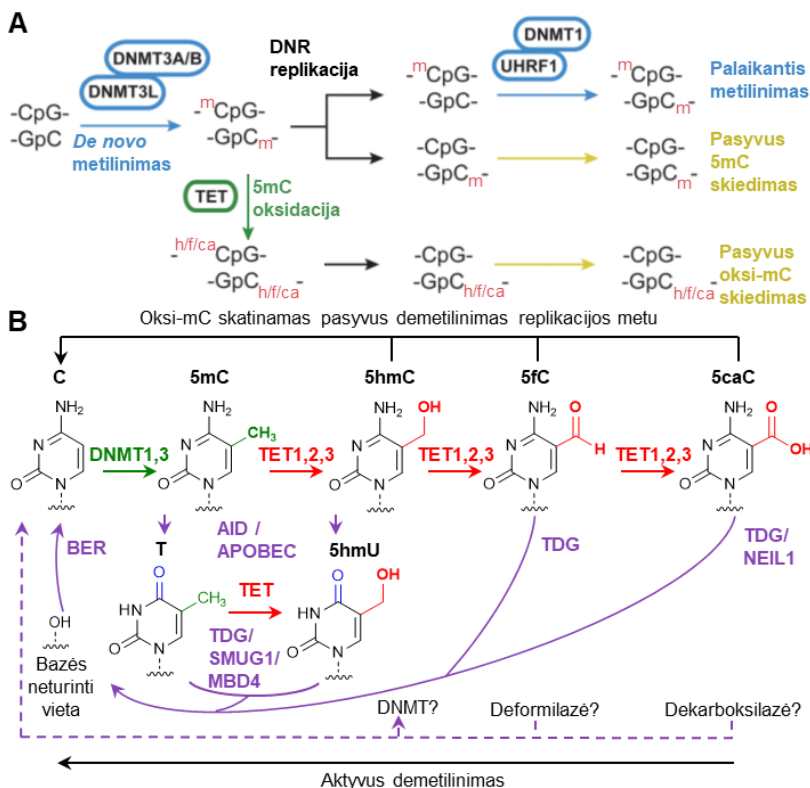
### **1.1.3 5-metilcitozino demetilinimas žinduoliuose**

Ankstyvojo žinduolių vystymosi metu DNR metilinimas keičiasi visame genome, nukreipdamas pliuripotentes kamienines ląsteles diferenciacijai į skirtingas linijas, kurios formuoja įvairius audinius (Raiber ir kt., 2017). 5mC gali būti gražinamas į nemodifikuotą būseną pasyviu arba aktyviu būdu. „Pasyvus“, arba nuo DNR replikacijos priklausomas demetilinimas (1.6 pav. A), reiškia laipsnišką 5-metilcitozino praskiedimą

dalijantis ląstelėms. Šiuo atveju, jei citozino metiltransferazė DNMT1 ir ją jungiantis UHRF1 baltymas yra slopinami arba neištraukiami į replikacijos šakutę, naujai susintetinta grandinė lieka nemetilinta (1.6 pav. A). „Aktyvus“ DNR demetilimas apibūdina greitą, nuo replikacijos nepriklausomą, fermentinį 5mC pašalinimą (Delatte ir Fuks, 2013).

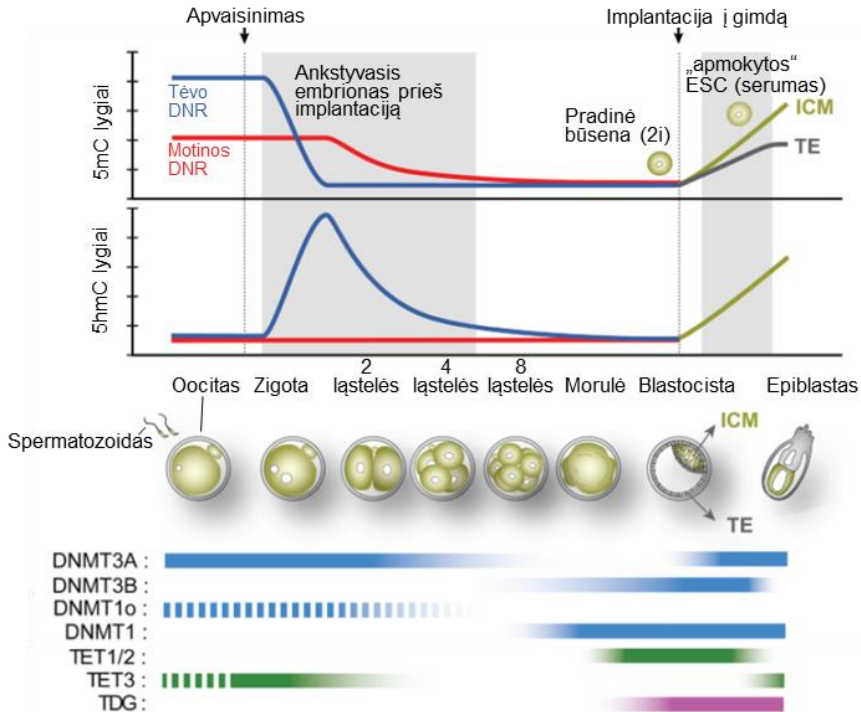
Vienas iš pirmųjų įrodymų, patvirtinančių aktyvų DNR demetilimą žinduoliuose, buvo nustatytas 5mC praradimas pelių zigotose (1.7 pav.). Po apvaisinimo 5mC signalas tėvo DNR greitai sumažėja tiek, kad to negalima visiškai paaiškinti pasyviu demetilimu (Wu ir Zhang, 2017). Aktyvus DNR demetilimas augaluose gali vykti veikiant DNR glikozilazėms ROS1/DME, galinčioms tiesiogiai pašalinti 5mC bazę, tačiau tokių baltymų žinduoliuose neaptikta (Deniz ir kt., 2019).

Molekulinis žinduolių aktyvaus DNR demetilimo mechanizmas atskleistas 2009 m., kai buvo parodyta TET1 fermento katalizuojama 5mC oksidacija į 5hmC (Tahiliani ir kt., 2009) bei pelės smegenų audiniuose aptiktas gausus 5hmC (Kriaučionis ir Heintz, 2009). Vėliau nustatyta, kad TET1, TET2 ir TET3 baltymai paeiliui oksiduoja (1.6 pav. B) 5-hidroksimetilcitoziną (5hmC) į 5-formilcitoziną (5fC) ir 5-karboksilcitoziną (5caC) (He ir kt., 2011; Ito ir kt., 2011; Pfaffeneder ir kt., 2011). 5fC ir 5caC yra atpažįstami ir iškerpami timino DNR glikozilazės (TDG) (He ir kt., 2011; Maiti ir Drohat, 2011), o susidariusi bazės neturinti vieta ištaisoma bazių iškirpimo reparacijos būdu (BER, angl. *base excision repair*) ir taip grąžinama nemodifikuota citozino būseną (1.6 pav. B) (Kohli ir Zhang, 2013). 5caC (bet ne 5fC) taip pat gali būti iškerpamas veikiant NEIL1 glikozilazei, bet jos reikšmė DNR demetilimui dar neaiški (Slyvka ir kt., 2017). Manoma, kad TET atliekama oksidacija skatina nuo replikacijos priklausomą pasyvų DNR demetilimą (1.6 pav. B), nes *in vitro* vienoje CpG taikinio grandinėje esant oksiduotoms 5-metilcitozino formoms (oksi-mC), DNMT1 metilinimą atlieka prasčiau (Seiler ir kt., 2018).



**1.6 pav.** Pasyvaus ir aktyvaus DNR demetilimo mechanizmai žinduoliuose. **(A)** Pasyvaus nuo replikacijos priklausomo 5mC arba oksiduoto 5mC praradimo CpG taikiniuose schema. Nuo replikacijos priklausomas pasyvus 5mC (m) praskiedimas įvyksta sutrikus DNMT1/UHRF1 palaikomojo DNR metilinimo mechanizmui. Oksiduotos 5mC formos (h/f/ca) gali palengvinti pasyvų demetilimą, nes joms esant vienoje grandinėje suprastėja DNMT1 metilinimo efektyvumas. Pagal Wu ir Zhang (2014). **(B)** Epigenetinių DNR citozino modifikacijų susidarymas ir pašalinimas. Citozinas (C) paverčiamas 5-metilcitozinu (5mC) veikiant DNMT1 ir DNMT3 šeimų DNR metiltransferazėms (žaliosis kelias). Pasiūlyti keli DNR demetilimo mechanizmai, kurių metu 5mC paverčiamas atgal į C. Raudonos rodyklės žymi TET baltymų atliekamą oksidaciją: 5mC metilo grupė iš eilės oksiduojama į hidroksimetilo, formilo ir karboksilo grupes, susidarant atitinkamai 5-hidroksimetilcitozinui (5hmC), 5-formilcitozinui (5fC) ir 5-karboksilcitozinui (5caC). TET baltymai taip pat gali oksiduoti T į 5-hidroksimetiluracilą (5hmU). Violetinės iššinės rodyklės rodo deamininimo, bazių iškirpimo ir bazių iškirpimo taisymo (BER) kelius. AID/APOBEC deaminina 5hmC į 5hmU ir 5mC į T. TDG, NEIL1, MBD4 arba SMUG1 glikozilazės iškerpa oksiduotas ar deaminintas bazines, susidarant laikinoms bazių neturinčioms vietoms, kurias ištaiso BER. Brūkšninės rodyklės žymi citozino metiltransferazių *in vitro* atliekamas hidroksimetilinimo ir dehidroksimetilinimo reakcijas bei numanomus fermentus (deformilazę ir dekarboksilazę), kurie galėtų tiesiogiai pašalinti formilo ir karboksilo grupes atitinkamai nuo 5fC ir 5caC. Pagal Kriukienė ir kt. (2012) ir Zampieri ir kt. (2018).

Žinduolių embriono vystymosi metu ląstelėse įvyksta dvi drastiškų 5mC lygio pokyčių bangos. Pirmiausia, po apvaisinimo zigotos motinos ir tėvo genomuose matomas visuotinis 5mC praradimas (1.7 pav.) (Xu ir Wong, 2015). 5mC oksidacija intensyviausiai vyksta vyriškame genome veikiant TET3. Prieš implantaciją į gimdą bendras embriono DNR metilinimas pasiekia gana žemą lygį, o po implantacijos pradžios jis padidėja. Kita DNR demetilinimo banga įvyksta formuojantis poimplantacinio embriono pirminėms lytinėms ląstelėms, o po to vyksta naujas lyčiai specifinis metilinimas (Yang ir kt., 2020).



**1.7 pav.** Dinamiški citozino modifikacijų ir fermentų pokyčiai, kurie vyksta pelės embrione prieš implantaciją į gimdą. Iškart po apvaisinimo 5mC greitai prarandamas pasyvaus DNR demetilinimo būdu ir TET3 vykdam oksidaciją. Oocite (o) aptinkama DNMT1 daugiausia yra ne branduolyje, todėl palaikantis metilinimas neefektyvus. Ląstelėms dalijantis oksiduotų 5mC formų ir 5mC kiekiai mažėja vykstant pasyviu replikaciniam jų skiedimui. Žemiausias 5mC lygis pasiekiamas maždaug blastocistos stadijoje. Po implantacijos į gimdą DNR metilinimas vidinėje ląstelių masėje (ICM), bet ne trofoektoderfoje (TE), DNMT3A/3B metiltransferazių sukuriamas iš naujo. Pradinės būsenos (dviejų slopiklių (2i) sąlygos ląstelių kultūroje) ICM ląstelių genomas metilintas silpnai (~1% 5mC nuo viso C), o po implantacijos (serumo sąlygos ląstelių kultūroje) epiblasto diferenciacijai „apmokytų“ kamieninių embrioninių ląstelių (ESC) DNR yra stipriai metilinta (~4% 5mC). Pagal Wu ir Zhang (2014).

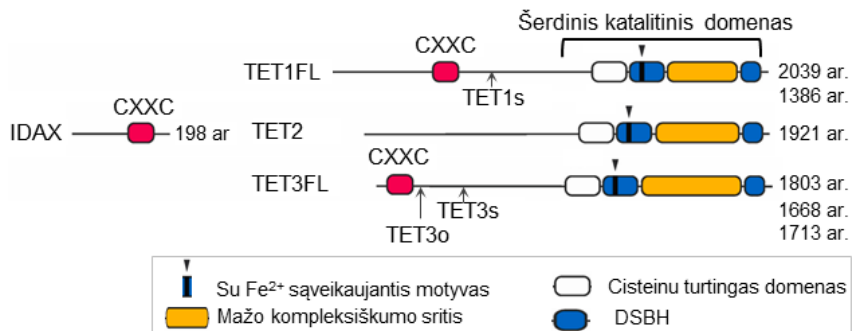
Parodyta, kad visuotinis 5mC demetilinimas žinduoliuose pasiekiamas derinant aktyvų ir pasyvų demetilinimą. TET baltymai daugiausia veikia specifiskai, tam tikriems genams suaktyvinti reikalingose vietose, ypač stiprikluose, ir taip reguliuojama vystymosi genų raiška (Tsiouplis ir kt., 2021). Taip pat sukeliama didelės apimties pasyvus DNR demetilinimas pašalinant UHRF1 nuo chromatino ir taip slopinamas palaikantis DNR metilinimas (Mulholland ir kt., 2020). TET-TDG aktyvaus DNR demetilinimo kelias yra gyvybiškai svarbus žinduolių vystymuisi, nes TET ir TDG fermentų funkcijos praradimas sutrikdo pelių embrioninių kamieninių ląstelių (mESC, angl. *mouse embryonic stem cells*) diferenciaciją ir yra mirtinas pelių embrionams (Raiber ir kt., 2017).

Pradiniuose bandymuose siekiant išgauti iš pelės blastocistos pluriipotentes kamienines embrionines ląsteles, įprastinėmis serumo auginimo sąlygomis jos įgydavo pakitusią hipermetilintą epigenetinę būklę, kuri yra panašesnė į vėlesnės stadijos poimplantacinio epiblasto (Lee ir kt., 2014). Vėliau atrasta, kad naudojant dviejų kinazių slopiklius 2i (angl. *two inhibitors*, slopina MEK ir GSK3) šios ląstelės gali būti grąžintos į pradinę (angl. *ground*) sumažėjusio metilinimo būseną (1.7 pav.), kuri labiau atspindi vidinės blastocistos ląstelių masės epigenetinę būseną (Ficz ir kt., 2013; Habibi ir kt., 2013). Veikiant 2i, yra slopinamas DNR metilinimas ir aktyvinama TET oksidacija (Lee ir kt., 2014). Pradinės būsenos ląstelės (dar vadinamos naiviomis) yra homogeniškos ir gali diferencijuotis į visas ląstelių linijas. Serumo sąlygomis ląstelės yra linkusios į diferenciaciją (vadinamos „apmokytomis“, angl. *primed*), heterogeniškos, pasižymi diferenciacijos faktorių raiška, negali diferencijuotis į lytines ląsteles (Martello ir Smith, 2014). Šiomis dvejomis sąlygomis auginamos ląstelės gali būti naudojamos kaip modelis, skirtas signalų sukeltų metilinimo pokyčių ir ESC epigenetinių programų tyrimams (Pennings ir kt., 2019).

#### 1.1.4 TET baltymai

*TET* genų pavadinimai (angl. *ten-eleven translocation*) kilo iš translokacijos, kuri aptinkama retais ūminės mieloidinės leukemijos atvejais, kai žmogaus TET1 genas iš 10 chromosomos susilieja su mišrios linijos leukemijos genu (MLL) 11 chromosomoje (Shen ir kt., 2014). Biocheminė TET baltymų funkcija nebuvo žinoma iki 2009 m., kai buvo parodyta, kad žmogaus DNR esantį 5mC TET1 gali oksiduoti į 5hmC (Iyer ir kt., 2009; Tahiliani ir kt., 2009). Du TET1 paralogai, TET2 ir TET3, buvo nustatyti remiantis sekos homologija (Iyer ir kt., 2009). Hipotezė, kad TET baltymai gali turėti DNR modifikuojantį aktyvumą, buvo pagrįsta J bazę jungiančių

baltymų analize. JBP fermentai priklauso didelei nuo  $\text{Fe}^{2+}$  ir  $\alpha$ -ketoglutarato ( $\alpha$ -KG, taip pat vadinamo 2-oksoglutaratu) priklausomų dioksigenazių superšeimai. AlkB fermentai, oksidaciniu mechanizmu pašalinantys nenormalaus DNR bazių metilinimo pažaidas, yra tos pačios superšeimios nariai (Pastor ir kt., 2013). JBP homologų analizė genomuose nuspėjo nukleorūgštis modifikuojančias dioksigenazes įvairiuose eukariotuose ir bakteriofaguose, įskaitant TET fermentus daugialąsčiuose gyvūnuose (Iyer ir kt., 2009). TET turimas CXXC domenas, jungiantis nemetilintus CpG, JBP homologų aptikimas kartu su DNR metiliniu bei timino panašumas į 5mC lėmė idėją, kad TET baltymai gali oksiduoti 5mC (Pastor ir kt., 2013).



**1.8 pav.** Žinduolių TET baltymų struktūros. TET baltymai turi konservatyvų C-galo šerdies katalitinį domeną, kurį sudaro cisteinu turtingas domenas, veikiantis prisijungimą prie chromatino, dvigrandės  $\beta$ -spiralės (DSBH) domenas, turintis su  $\text{Fe}^{2+}$  sąveikaujantį motyvą (His–bet kuri ar.–Asp) ir mažo kompleksišumo įterpimą. Embrioninėse ląstelėse aptinkami viso ilgio (FL) TET1 ir TET3 baltymai turi N-galinį CXXC domeną, reikalingą jungimuisi prie CpG turtingų DNR sekų. CXXC domeno nėra trumposiose (s) TET1 ir TET3 izoformose (pradžia parodyta rodyklėmis), kurių raiška gausi subrendusiose smegenyse. TET2 neturi CXXC domeno, tačiau sąveikauja su CXXC domeną turinčiu IDAX baltymu. TET3o izoformos raiška yra aptinkama oocituose. Ar. – baltymo ilgis aminorūgštimis. Pagal Dick ir Chen (2021).

TET atlieka oksidaciją naudodami deguonies ir  $\alpha$ -KG substratus bei  $\text{Fe}^{2+}$  kofaktorių, susidarant  $\text{CO}_2$  ir sukcinatui (Bochtler ir kt., 2017). C-gale yra pagrindinis katalitinis TET domenas, sudarytas iš dvigrandės  $\beta$  spiralės (DSBH) domeno ir cisteinu turtingo domeno (1.8 pav.) (Wu ir Zhang, 2017). DSBH domenas kartu jungia  $\text{Fe}^{2+}$ ,  $\alpha$ -KG ir nukleotidą oksidacijai. O cisteinu turtingas domenas, būtinas katalitiniam aktyvumui, apvynioja DSBH šerdį, kad stabilizuotų bendrą TET-DNR struktūros sąveiką (Wu ir Zhang, 2017). DSBH, kuris būdingas visoms nuo  $\text{Fe}^{2+}$  ir  $\alpha$ -KG priklausomoms dioksigenazėms, turi aštuonias konservatyvias antilygiagrečias  $\beta$ -juostas (I–VIII) ir konservatyvų geležį jungiantį motyvą. TET baltymams būdinga unikali ilga nekonservatyvi mažo kompleksišumo sritis, įterpta tarp IV ir V  $\beta$  juostų (Shen ir kt., 2014). Spėjama, kad ši sritis gali atlikti reguliacinę

funkcija, kuri keičiama potransliacinėmis modifikacijomis (pvz., glikoziliniu ir fosforiliniu) (An ir kt., 2017).

Amino galinėje srityje TET1 ir TET3 turi labai konservatyvų DNR jungiantį domeną CXXC (1.8 pav.), kurį sudaro du cisteino tipo cinko pirštų motyvai. Oksidacija gali vykti ir be šio domeno (Shen ir kt., 2014). TET2 neturi CXXC domeno, kuris atsiskyrė chromosomų inversijos metu ir tapo atskiru genu IDAX. Tačiau TET2 sąveikauja su IDAX ir taip iš dalies gali būti lemiamas jo prisijungimas prie chromatino (Wu ir Zhang, 2017). CXXC domenas gali daryti įtaką TET baltymų pasiskirstymui genome: pelių embrioninėse kamieninėse ląstelėse TET1 pirmiausia aptinkamas promotorių CpG salose arba stiprikliuose, o TET2 dažniausiai yra praturtintas genų kūnuose arba stipriklių regionuose (An ir kt., 2017).

Įdomu, kad pelės somatiniuose audiniuose aptinkama N-gale sutrumpinta TET1 forma (TET1s) (1.8 pav.), o viso ilgio forma (vadinama TET1FL) yra ankstyvuosiuose embrionuose, ESC ir pirminėse lytinėse ląstelėse (Wu ir Zhang, 2017). TET1s, kuris neturi CXXC domeno, silpniau jungiasi prie chromatino nei TET1FL ir pasižymi silpnesniu oksidaciniu aktyvumu. TET3 taip pat turi keletą izoformų, įskaitant TET3s ir TET3o, neturinčias CXXC domeno. TET3o yra aptinkama oocituose, o TET3s ir viso ilgio TET3 raiška didėja vykstant neuronų diferenciacijai. TET3s ir TET3o pasižymi stipresniu oksidacijos aktyvumu nei viso ilgio TET3 (Wu ir Zhang, 2017).

Vis labiau akivaizdu, kad TET baltymai ir jų reakcijos produktai yra vieni pagrindinių embriono vystymosi, kamieninių ląstelių funkcijų, diferenciacijos ir lytinių ląstelių vystymosi reguliatoriai (An ir kt., 2017). TET baltymai taip pat gali reguliuoti pluriipotenciją kontroliuodami telomerų ilgį, o jos padeda palaikyti genomo stabilumą ir yra reikalingos ląstelių atsinaujinimui (Yang ir kt., 2020). Taip pat atsiranda duomenų, kad TET fermentai svarbūs TE aktyvumo reguliavimui. Embrione prieš implantaciją stebimas kai kurių TE hidroksimetilinimas ir raiškos padidėjimas (Deniz ir kt., 2019). Tačiau taip pat parodyta, kad ESC TET1 oksiduoja LINE-1 TE elementus ir kartu pritraukia slopinantį SIN3A-histonų deacetilazės kompleksą, užtikrindamas nutildymą (Deniz ir kt., 2019). Nemažai tyrimų rodo TET baltymų svarbą neuronų funkcijoms (Raiber ir kt., 2017). Pvz., TET1 reguliuoja suaugusio organizmo hipokampo neurogenezę ir pažinimo funkcijas, kontroliuodamas nervinių kamieninių ląstelių dauginimąsi (Xu ir Wong, 2015). TET baltymų raiškos ar funkcijos sutrikimas dažnai stebimas įvairiose vėžio formose ir ypač daro įtaką kraujo ląstelių pakitimams (mielocitų ir limfocitų) (An ir kt., 2017). TET2 mutaciją turinčios pelėms



išsivysto mielininiai vėžiniai pakitimai, o TET2 mutacija taip pat stebima leukemijos pacientuose (Xu ir Wong, 2015).

TET baltymai taip pat gali reguliuoti genų raišką nepriklausomai nuo katalitinio aktyvumo, sąveikaudami su kitais baltymais, darančiais įtaką transkripcijai ir chromatinio struktūrai (Tsiouplis ir kt., 2021). Nustatyta, kad tiek TET1, tiek TET2 tiesiogiai sąveikauja su SIN3A ir histono deacetilazių kompleksu, kuris slopina genų raišką, o taip pat su pliuripotencijos faktoriumi Nanog. ESC TET1 taip pat gali pritraukti slopinantį kompleksą PRC2 (per jo EZH2 ir Suz12 komponentus) prie ląstelių linijai specifinių genų promotorių, o vykstant diferenciacijai PRC2 paleidžiamas taip veikiant vystymuisi svarbių genų raišką (Yang ir kt., 2020).

### **1.1.5 Oksiduotų 5-metilcitozino formų funkcijos**

Oksi-mC pasiskirstymas ir funkcijos DNR geriausiai ištirtos žinduoliuose. Oksi-mC aptinkamos daugumoje žinduolių audinių, tačiau jų lygis yra gana žemas ir kitaip nei 5mC (~4% nuo C), labai varijuoja skirtingų tipų ląstelėse (An ir kt., 2017). Suaugusiose pelėse 5hmC gausiausias smegenų Purkinje neuronuose ir su pažintinėmis funkcijomis susijusiose srityse (~0,6-1,3% nuo C), o kituose audiniuose nustatomi mažesni kiekiai (~0,03-0,3% nuo C) (Ito ir kt., 2011; Kriaučionis ir Heintz, 2009; Wagner ir kt., 2015). 5hmC lygiai gali keistis su organizmo amžiumi, pvz., senstant jie didėja smegenų audiniuose (Wagner ir kt., 2015). 5fC yra daug retesnis nei 5hmC, o 5caC kiekiai yra mažiausi. Pavyzdžiui, mESC 5fC (~0,001% nuo C) ir 5caC (~0,0002% nuo C) kiekiai yra viena-dviem eilėmis mažesni nei 5hmC (~0,06%) (Ito ir kt., 2011). Mažesniais kiekiais (~0,00002-0,001%), kurie varijuoja nepriklausomai nuo 5mC ar 5hmC, 5fC taip pat aptinkamas įvairiuose suaugusių žinduolių audiniuose ir ląstelių tipuose (Iurlaro ir kt., 2016). 5caC gausiausias ESC ir embrioniniuose audiniuose (Bachman ir kt., 2015; Ito ir kt., 2011), o suaugusio organizmo ląstelėse dėl mažų kiekių beveik neaptinkamas, nors padidinus aptikimo jautrumą, pvz., nustatytas tiesiosios žarnos vėžiniuose ir jiems gretimuose normaliuose audiniuose (Tang ir kt., 2015).

Oksi-mC lygiai nebūtinai koreliuoja tarpusavyje: pavyzdžiui, nors smegenų žievėje 5hmC gausesnis nei ESC, 5fC ir 5caC atveju yra atvirkščiai (Wu ir Zhang, 2017). Vadinas, oksidacijos etapai yra nevienodai reguliuojami įvairiuose audiniuose. Nevienodas oksi-mC gausumas bent iš dalies gali būti siejamas su TDG vykdomu 5fC ir 5caC (bet ne 5hmC) iškirpimu (He ir kt., 2011). Ir iš tiesų, išveiklinus TDG, 5fC ir 5caC kiekiai mESC padidėja 5-10 kartų (Shen ir Zhang, 2013; Song ir kt., 2013). Be to, gali skirtis oksidacijos

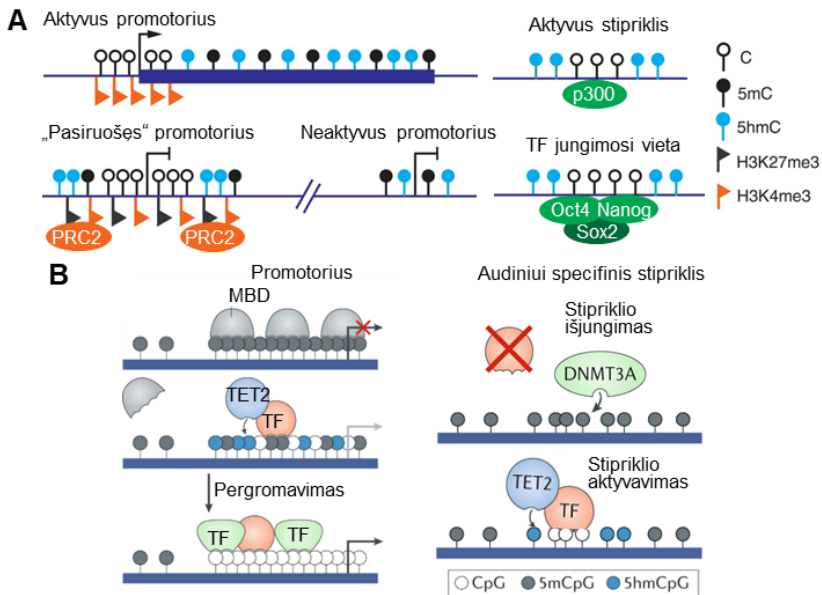
efektyvumas: *in vitro* TET baltymai yra aktyvesni esant 5mC, nei esant 5hmC ir 5fC (Ito ir kt., 2011). Pelių ESC TET fermentai į 5hmC paverčia ~10% 5mC ir tik dalis 5hmC (1-10%) toliau oksiduojamas iki 5fC/5caC. Taigi 5hmC yra mažiau linkęs į tolesnę oksidaciją, todėl yra gausesnis ir dalis jo išlieka kaip stabilus epigenetinis ženklas (An ir kt., 2017). Manoma, kad *in vivo* TET procesyvumas gali priklausyti nuo chromatino aplinkos, nes, pvz., labiau prieinamame chromatine oksidacija dažniau vyksta iki 5fC ir 5caC (Wu ir kt., 2014).

Oksi-mC gali būti svarbios vėžio vystymuisi ir naudingos kaip biožymenys, nes vėžiniuose audiniuose nustatomi jų kiekių pakitimai (An ir kt., 2017). Greitai besidauginančiose ir metastazuojančiose ląstelėse 5hmC kiekiai sumažėja ir tai siejama su prasta prognoze (Klungland ir Robertson, 2017). Pasiūlyta, kad ląstelėms virstant į vėžines gali vykti nuo 5caC priklausoma transkripcijos reguliacija, nes, pvz., 5caC kiekiai padidėja krūties ir įvairiuose nervinių ląstelių vėžiniuose audiniuose (Eleftheriou ir kt., 2015; Ramsawhook ir kt., 2017; Storebjerg ir kt., 2018; Zhou ir kt., 2018).

Oksi-mC kai kuriais atvejais gali būti stabilios DNR modifikacijos. Naudojant žymėjimą stabiliais izotopais pademonstruota, kad 5hmC daugiausia yra stabilus pelės audiniuose ir ląstelių kultūrose, o 5fC stabilus suaugusių pelių audiniuose, smegenyse (Bachman ir kt., 2014, 2015). 5hmC stabilumas, pavyzdžiui, taip pat buvo stebimas tam tikruose eritrocitų genominiuose regionuose, kuriuose 5hmC praturtinimas išliko po dalijimosi ir diferenciacijos (Cao ir kt., 2019). Kitas tyrimas parodė sumažėjusį TDG aktyvumą, kai 5fC buvo jungiamas nukleosomos su H2A.Z ir H3.3 histonų variantais, paprastai praturtintais stipriklių sekose. Vadinas, tam tikruose regionuose 5fC gali būti stabilesnis (Prasad ir kt., 2021). Zigotoje susidarę 5fC ir 5caC praskiedžiami vykstant ląstelių dalijimuisi ir DNR replikacijai, bet nėra greitai iškerpami, todėl yra gana stabilūs ir galėtų būti funkciškai svarbūs vystymuisi iki implantacijos (Inoue ir kt., 2011).

Vis daugiau duomenų rodo, kad oksi-mC veikia ne tik kaip aktyvaus DNR demetilavimo tarpiniai produktai, bet ir kaip epigenetinių funkcijų turinčios modifikacijos. Oksi-mC genomino profiliavimo tyrimai atskleidė jų išsidėstymą reguliaciniuose elementuose, kuriuose vyksta dideli metilavimo pokyčiai (pvz., transkripcijai „pasiruošusiuose“ promotoriuose ir ląstelei specifiniuose stiprikliuose), todėl oksi-mC gali būti svarbios epigenetinėje genų reguliacijoje (Lu ir kt., 2015a). Gausiausios oksi-mC – 5hmC – epigenetinės genų reguliavimo funkcijos vis plačiau pripažįstamos ir labiausiai ištirtos. Tuo tarpu 5fC, o ypač 5caC, epigenetinės funkcijos mažiau ištirtos ir dėl jų dar abejojama, bet atsiranda tai paremiančių duomenų.

5hmC yra svarbus daugeliui procesų: embriono vystymuisi, diferenciacijai, citokinų atsakams, nervinių audinių ir atminties formavimuisi (An ir kt., 2017). 5hmC daugiausia aptinkamas mažo bei vidutinio CpG tankio ir vidutinio 5mC lygio regionuose (1.9 pav. A). 5hmC praturtintas aktyviuose (turinčiuose H3K27Ac ir H3K4me1) ir „pasiruošusiuose“ (turinčiuose tik H3K4me1, angl. *poised*) stiprikluose, promotoriuose, genų kūnuose, taip pat aplink pliuripotencijai svarbių transkripcijos faktorių jungimosi vietas, CTCF regulatoriaus jungiamuose ir atvirą chromatiną žyminčių histonų modifikacijų regionuose (Pastor ir kt., 2011; Song ir kt., 2011; Wu ir kt., 2011; Yu ir kt., 2012). Promotoriuose aptinkamas stiprus hidroksimetilinimas dažniausiai siejamas su maža raiška, o genai su mažu 5hmC lygiu transkripcijos pradžios vietose yra stiprios raiškos (Yang ir kt., 2020). Didelis 5hmC lygis stebimas transkripcijai „pasiruošusiuose“ bivalentiniuose promotoriuose (turinčiuose ir slopinančias H3K27me3 ir aktyvinančias H3K4me3 histonų modifikacijas), kurie yra slopinami ESC, o ląstelėms diferencijuojantis aktyvuojami (Wu ir Zhang, 2017). Genų kūnuose esančio 5hmC įtaka raiškai gali priklausyti nuo ląstelių tipo – pvz., didelis 5hmC lygis ESC ir daugelyje kitų ląstelių koreliuoja su stipria raiška, o neuronų pirmtakuose su žema raiška (Yang ir kt., 2020).



**1.9 pav.** Citozino modifikacijos ir genų raiškos aktyvinimas. (A) Genominis 5hmC pasiskirstymas pelių ESC genome. 5hmC yra praturtintas aktyvių genų kūnuose, transkripcijai „pasiruošusiuose“ bivalentiniuose (turinčiuose slopinančias ir aktyvinančias histonų modifikacijas, pažymėtas vėliavėlėmis) ir neaktyviuose promotoriuose, taip pat aktyviuose stiprikluose (kuriuose aptinkama histonų acetiltransferazė p300) ir dalyje pliuripotencijos transkripcijos faktorių jungiamų

vietų. Pagal Shen ir Zhang (2013). **(B)** Dinamiški ląstelei specifiniai DNR metilinimo pokyčiai dažniausiai vyksta reguliaciniuose elementuose. Kairėje, neaktyvūs metilinti promotoriai gali būti atpažįstami MBD domenais turinčių transkripcijos slopiklių. Tačiau tam tikrose metilintose sekose taip pat gali prisijungti ir transkripcijos aktyvatoriai (pvz., pluriipotencijos faktorius Oct4), kurie skatina ląstelių pakitimus. Pavyzdžiui, vykstant indukuotam ląstelių perprogramavimui, KLF transkripcijos faktorius (TF) pritraukia TET2, kuris sukelia DNR demetilinimą ir genų aktyvavimą. Dešinėje, TF nesijungia prie neaktyvių stipriklių, kurie yra prieinami DNR metiltransferazėms. TET baltymai, veikdami kartu su transkripcijos faktoriais, sukelia DNR demetilinimą ir stipriklių aktyvavimą. Pagal Greenberg ir Bourc'his (2019).

mESC 5fC ir 5caC didžiaja dalimi atspindi 5hmC praturtinimą, tik jie daugiausia išsidėstę nuo transkripcijos pradžios vietos labiau nutolusiuose regionuose (Shen ir kt., 2013). 5fC ir 5caC yra stipriai praturtinti aktyviuose stiprikliuose, pluriipotencijos faktorių, CTCF, p300 (aktyviuose stiprikliuose aptinkamos svarbios histonų acetiltransferazės) jungiamose vietose (Shen ir kt., 2013; Song ir kt., 2013; Wu ir kt., 2014; Lu ir kt., 2015b; Neri ir kt., 2015; Iurlaro ir kt., 2016). 5fC ir 5caC taip pat praturtinti „pasiruošusiuose“ stiprikliuose bei promotoriuose (dauguma tyrimų praturtinimą nustato aktyviuose ir bivalentiniuose), aktyvių genų kūnuose bei tandeminiuose pasikartojimuose (Shen ir kt., 2013; Wu ir kt., 2014; Lu ir kt., 2015b; Neri ir kt., 2015; Song ir kt., 2013).

Pelių ESC 5hmC/5fC/5caC yra labiau praturtinti egzonuose palyginus su intronais, todėl galbūt 5mC oksidacija gali reguliuoti iRNR splaisingą (Wu ir Zhang, 2014). Be to, 5hmC yra labiau praturtintas pastoviuose egzonuose ir ties egzono-introno ribomis (Khare ir kt., 2012). Splaisingo reguliacijoje galbūt dalyvauja CTCF baltymas, kurio jungimasis prie 5caC ar 5hmC alternatyviuose egzonuose sukelia jų įtraukimą (Marina ir kt., 2016).

Oksi-mC praturtintos genomo vietos ne visada sutampa, vadinasi, fermentai gali būti reguliuojami atskirų oksimC susidarymui tam tikruose genomo regionuose (Parry ir kt., 2021). Pavyzdžiui, pelės ESC nustatyta, kad >90% 5hmC modifikuotų CpG nesutampa su 5fC ar 5caC turinčiomis vietomis ir tik ~19% 5fC/5caC vietų sutampa su 5hmC (Wu ir kt., 2014). Tarpusavyje 5fC ir 5caC taip pat persidengia labai mažai (6,5%), todėl atskirų oksimC funkcijos galbūt skirtingos (Lu ir kt., 2015b). Aktyviuose stiprikliuose 5fC praturtinimas yra stipresnis nei 5hmC, todėl 5fC sritys yra aktyvesnės, o tai taip pat patvirtina 5fC srityse (lyginant su 5hmC sritimis) stebimi mažesni 5mC lygiai (Xia ir kt., 2015). Taip pat pastebėta, kad su promotoriais sąveikaujantys transkripciją skatinantys stiprikliai labiausiai persidengia su 5caC ir 5fC (5caC >5fC >5hmC). Todėl manoma, kad 5caC ir 5fC žymi aktyvesnes chromatinio sritis nei 5hmC (Lu ir kt., 2015a).

Oksi-mC funkcijos reguliuojant epigenetinę būklę ir ląstelės lėmtį gali būti skirtingos. Pavyzdžiui, pademonstruota, kad TET oksidacija į 5fC/5caC, bet ne į 5hmC, skatina indukuotų pluriipotentinių ląstelių perprogramavimo efektyvumą (Caldwell ir kt., 2021). mESC nustatyti pluriipotencijos transkripcijos faktorių Nanog, Oct4 ir Sox2 imunoišsodinimo signalai yra stipresni 5fC ir 5caC modifikuotuose regionuose palyginus su tik 5hmC turinčiais regionais, o tai reiškia stipresnį šių veiksmų jungimosi regionų demetilinimą (Wu ir kt., 2014). Kamieninėms ląstelėms diferencijuojantis į neuronus, glijos ląsteles ir hepatocitus 5caC kiekiai laikinai padidėja bei diferenciacijos metu 5caC yra aptinkamas ląstelei specifinių aktyvių genų promotoriuose (Lewis ir kt., 2017b; Wheldon ir kt., 2014). Todėl manoma, kad oksidacija į 5caC gali būti bendras epigenetinio pertvarkymo mechanizmas, vykstant ląstelių linijų formavimuisi.

Biofizikiniai tyrimai rodo, kad oksi-mC gali paveikti fizines DNR savybes. Pavyzdžiui, 5hmC padidina DNR dvigubos spiralės termodinaminį stabilumą (Raiber ir kt., 2015). 5fC sukelia vietinės DNR struktūros pokyčius, padidindamas struktūros lankstumą ir daro įtaką DNR jungiančių baltymų prieinamumui, tikriausiai pakeisdamas DNR superspiralizaciją ir pakavimą į nukleosomas (Raiber ir kt., 2015). 5caC išplečia mažąjį griovį ir atlaisvina dvigrandes dalis (Dai ir kt., 2015), kitaip nei labiau stabilizuojantis 5mC (Harcourt ir kt., 2017; Fu ir kt., 2019). Vadinasi, tiesiogiai paveikdamos DNR struktūrą, oksi-mC gali daryti įtaką baltymų jungimuisi prie DNR (Wu ir Zhang, 2015).

Dėl skirtingų metilo, hidroksimetilo, formilo ir karboksilo grupių cheminių savybių oksi-mC galėtų būti dinamiški epigenetiniai ženklai, kurie skirtingai veikia baltymų prisijungimą ir reguliuoja genų raišką (Lu ir kt., 2015a). Pavyzdžiui, RNR polimerazė II sudaro vandenilinius ryšius su 5caC arba 5fC (karboksilo arba formilo grupėmis) ir dėl to sulėtėja transkripcijos ilginimas genų kūnuose, kuris galbūt gali daryti įtaką transkripcijos reguliavimui (Kellinger ir kt., 2012; Wang ir kt., 2015). Nustatyta, kad 5fC stabilizuoja nukleosomas *in vitro*, o kitų C modifikacijų stabilizavimo efektas yra žymiai mažesnis (Raiber ir kt., 2018). Taip pat pademonstruota, kad *in vivo* 5fC sukelia dažnesnį nukleosomų išsidėstymą sudarydamas grįžtamas kovalentines sąsijas su H3 histono lizinu. Šis padažnėjęs išsidėstymas atsiranda aktyviuose stiprikluose ir koreliuoja su padidėjusia audiniui specifine genų raiška (Raiber ir kt., 2018).

Buvo parodyta, kad oksiduotas 5mC formas atpažįsta iš dalies skirtingi baltymai (Iurlaro ir kt., 2013; Spruijt ir kt., 2013), todėl tikėtina, kad jos gali atlikti skirtingas epigenetines funkcijas. Tarp oksi-mC atpažįstančių

baltymų buvo aptikti ne tik DNR pažaidas taisantys baltymai, bet ir transkripcijos faktoriai, reguliatoriai bei chromatiną modifikuojantys fermentai (Tsiouplis ir kt., 2021). Nepaisant jų retumo, nustatyta daug daugiau 5fC ir 5caC jungiančių baltymų (tarp kurių daug DNR pažaidas taisančių baltymų) nei jungiančių 5hmC, galbūt dėl unikalių formilo ir karboksilo grupių cheminių savybių (Spruijt ir kt., 2013).

Keisdamos citozino modifikavimo būseną, ląstelės galėtų selektyviai kontroliuoti chromatiną baltymų prisijungimą ir disociaciją (1.9 pav. B) (An ir kt., 2017). Oksi-mC gali trukdyti transkripcijos slopiklių prisijungimui ir taip skatinti genų raišką. Pavyzdžiui, MBD1 baltymas specifiškai jungiasi prie 5mC bei prijungia histono H3K9 metiltransferazę ir taip slopinama raiška. Tuo tarpu MBD1 nesijungia prie oksi-mC, todėl jos skatina raišką (Tsiouplis ir kt., 2021).

Oksi-mC gali aktyvinti raišką skatindamos kai kurių transkripcijos faktorių ar reguliatorių prisijungimą. Pavyzdžiui, nervinėse ląstelėse MeCP2 (angl. *methyl-CpG-binding protein 2*) baltymui prisijungus prie 5hmC pasikeičia chromatiną struktūra ir skatinama genų raiška (Tsiouplis ir kt., 2021). 5hmC genų kūnuose taip pat pritraukia baltymus, atliekančius H3K36 trimetilimą, kuris žymi aktyvią transkripciją (Klungland ir Robertson, 2017).

Buvo nustatyti transkripcijos faktoriai, kurie teikia pirmenybę 5caC turinčioms atpažinimo vietoms ir gali būti, jog skatina atvirą aktyvų chromatiną. Pavyzdžiui, onkogeninio transkripcijos faktoriaus c-Myc partneris Max veikia kaip 5caC jutiklis, o 5caC lygiai galbūt gali moduluoti Myc-Max valdomą transkripcijos tinklą ir daryti įtaką vėžio vystymuisi (Wang ir kt., 2017). Kitas 5caC atpažįstantis transkripcijos faktorius, TCF4, yra svarbus smegenų vystymuisi, atminčiai ir pažinimui (Yang ir kt., 2019). Įdomu, kad TET3 CXXC domenas yra labai specifiškas 5caC (Jin ir kt., 2016). Tai galėtų būti svarbu genų reguliavimui, nes TET3 yra praturtintas su lizosomos funkcijomis susijusių aktyvių neuronų genų transkripcijos pradžios vietose, o lizosomos svarbios užkertant kelią neurodegeneracinėms ligoms (Jin ir kt., 2016). Daugiafunkcinis reguliatorius CTCF, kuris, pavyzdžiui, dalyvauja stipriklių-promotorių sąveikose, stipriausiai jungia 5caC modifikuotą DNR (Marina ir kt., 2016). O 5caC kaupimasis TDG neturinčiuose pelių embrionų fibroblastuose sukelia neįprastas CTCF jungimosi vietas visame genome (Nanan ir kt., 2019). Tai gali daryti įtaką genų raiškai, nes CTCF jungimosi pakitimai taip pat gali sukelti DNR metilinimo pokyčius (Prasad ir kt., 2021).

Oksi-mC gali ir slopinti raišką, trukdydami transkripcijos faktorių ir aktyvatorių prisijungimui. Pavyzdžiui, šalia promotoriaus esančioje ciklinio AMP atsako elemento sekoje 5hmC gali sumažinti transkripcijos aktyvatoriaus prisijungimą, todėl gali sumažėti tikslių į šį signalą reaguojančių genų raiška (Tsiouplis ir kt., 2021).

Taigi skirtingos oksimC, sąveikaudamos su ląstelių linijai specifiniu transkripcijos faktorių rinkiniu, gali sukelti unikalius transkripcijos atsakus įvairiose ląstelėse. 5mC ir oksimC gali veikti kaip papildomos genetinio kodo raidės, padidinančios galimų transkripcijos faktorių jungimosi motyvų ir ląstelės transkripcijos sudėtingumą (Parry ir kt., 2021). Tačiau norint tiksliau suprasti šiuos funkcinius ryšius, reikia tolesnių tyrimų.

OksimC aptinkamos ir kituose organizmuose nei žinduoliai, o pvz., 5hmC pirmiausia buvo nustatytas T2, T4 ir T6 bakteriofaguose (Shen ir kt., 2014). 5hmC aptiktas ryžių TE ir heterochromatino regionuose, tačiau 5hmC reikšmė augaluose nėra iki galo aiški (Kumar ir kt., 2018). 5hmC randamas ne tik stuburinių gyvūnų genomuose, bet taip pat buvo aptiktas kempinių, duobagyvių ir bičių genomuose (de Mendoza ir kt., 2019). Manoma, kad kaip ir stuburinių, kitų gyvūnų TET oksidacija taip pat vyksta transkripcijos faktorių jungimosi vietose. Tokių duomenų yra, pvz., pinties ir iešmučio atveju, tačiau kiek tai yra konservatyvu tarp gyvūnų dar reikia patikrinti (de Mendoza ir kt., 2019). 5fC ir 5caC yra retesnės modifikacijos ir nagrinėtos mažiau, tačiau yra duomenų apie jų (kartu su 5hmC) buvimą grybų (Chavez ir kt., 2014) bei augalų genomuose (Tang ir kt., 2014). Taip pat 5caC kartu su 5hmC nustatytas varliagyvio aksolotlio kiaušidžių folikulinių ląstelių branduolių euchromatine (Alioui ir kt., 2012). OksimC epigenetinė reikšmė daugelyje eukariotinių organizmų grupių mažai tirta ir dar gerai nesuprasta.

### **1.1.6 Kiti galimi žinduolių aktyvaus demetilinimo mechanizmai**

TET-TDG vykdomas aktyvaus DNR demetilinimo kelias žinduoliuose yra geriausiai ištirtas ir įrodytas, tačiau pelės zigotose nesant TDG, aktyvus demetilinimas nepaveikiamas (Guo ir kt., 2014a), todėl buvo pasiūlyti alternatyvūs mechanizmai (1.6 pav. B).

AID/APOBEC vykdo citozino deamininimą, o susidariusią klaidingą U:G bazių porą ištaiso glikozilazės ir BER. Šis procesas reikalingas somatinėms hipermutacijoms ir imunoglobulinų genų klasių rekombinacijai B ląstelėse (Klungland ir Robertson, 2017). Pasiūlyta, kad nemodifikuotas citozinas gali būti grąžinamas BER po AID ar APOBEC vykdomo 5mC ar 5hmC deamininimo (1.6 pav. B), tačiau duomenys priešaringi ir norint suprasti deaminazių vaidmenį DNR demetilinime reikia tolesnių tyrimų

(Bochtler ir kt., 2017). Yra duomenų, kad didelė AID/APOBEC raiška ląstelėse sukelia 5hmC deaminavimo produkto 5hmU atsiradimą ir jį gali iškirpti glikozilazės (TDG, MBD4 ar SMUG1). Be to, AID slopinimas sutrikdo pluriipotentškumui svarbių genų promotorių DNR demetilinimą perprogramuojant ląsteles (Raiber ir kt., 2017; Bochtler ir kt., 2017). Tačiau kiti tyrimai nepatvirtino 5hmU susidarymo ir demetilinimo AID/APOBEC vykdomo deaminavimo būdu (Klungland ir Robertson, 2017). *In vitro* AID/APOBEC 5mC deaminavimo aktyvumas silpnesnis lyginant su C, o esant 5hmC dar silpnesnis (Mariani ir kt., 2013). Be to, viengrandė DNR yra daug geresnis AID substratas, o 5hmU gali susidaryti ir TET baltymams oksiduojant T (Pastor ir kt., 2013; Raiber ir kt., 2017). Buvo pasiūlytas ir procesyvaus 5mC demetilinimo kelias, galbūt paaiškinantis nesutampančius duomenis. Šiuo atveju AID galėtų deamininti citoziną, o po to, vykstant plačiai regiono BER, netoli deaminavimo regiono pasroviui esantys 5mC galėtų būti pakeičiami nemodifikuotais citozinais (Bochtler ir kt., 2017).

Kitas galimas mechanizmas yra tiesioginis citozino modifikacijų pašalinimas skeliant C-C ryšį. *In vitro* tyrimai parodė, kad žinduolių DNMT3A ir 3B bei bakterijų DNR metiltransferazės, nesant kofaktoriaus SAM gali pašalinti 5-hidroksimetilo grupę nuo 5hmC susidarant nemodifikuotam citozinui (1.6 pav. B) (Chen ir kt., 2012; Liutkevičiūtė ir kt., 2009, 2014). Tačiau šios reakcijos biologinė svarba yra abejotina, nes SAM yra gausus daugelio svarbių biocheminių reakcijų, vykstančių visų tipų ląstelėse, metilo donoras (Shen ir kt., 2014). *In vitro* DNMT3 gali atlikti ir 5mC demetilinimą, bet tam reikia labai didelių, fiziologinių sąlygų neatitinkančių,  $Ca^{2+}$  koncentracijų (Chen ir kt., 2012). Taip pat parodyta, kad DNMT3 atlieka 5mC deaminimą ir susidariusi T:G pora galėtų būti ištaisoma BER (Métivier ir kt., 2008).

Dar vienas iš pasiūlytų alternatyvių mechanizmų, kad 5caC ir 5fC galbūt gali būti paverčiami atgal į C tiesioginio deformilino arba dekarboksilinimo būdu (1.6 pav. B) (Pfaffeneder ir kt., 2011; Wu ir Zhang, 2010). Buvo parodyta, kad 5caC turinti DNR yra dekarboksilinama esant tiolių, o kamieninių ląstelių ekstraktai taip pat gali atlikti šią reakciją (Schiesser ir kt., 2012). Neseniai atlikti izotopų ar fluoro žymėjimo ir masių spektrometrijos tyrimai patvirtino, kad įvairios pelės ir žmogaus ląstelių kultūros gali vykdyti tiesioginį 5fC deformilinimą į C (Iwan ir kt., 2018; Korytiaková ir kt., 2021; Schön ir kt., 2020), o taip pat parodytas ir tiesioginis 5caC dekarboksilinimas įvairių žinduolių ląstelių genomuose (Feng ir kt., 2021a, 2021b; Kamińska ir kt., 2021). *In vivo* šias reakcijas vykdytys fermentai iki šiol nežinomi, nors *in vitro* buvo parodytas 5caC



dekarboksilinimas esant žmogaus ar bakterijų DNR citozino-5 metiltransferazėms (Liutkevičiūtė ir kt., 2014).

### 1.1.7 *TET-JPB* genų ekspansija modeliniuose grybuose *Laccaria bicolor* ir *Coprinopsis cinerea*

TET baltymų pošeimė priklauso TET-JPB šeimai (Iyer ir kt., 2009). Sekų ir filogenetinė analizė parodė, kad TET-JPB šeima kilo iš bakteriofagų (Iyer ir kt., 2013). TET homologiški baltymai aptinkami įvairiuose organizmuose: viena ar tik kelios *TET-JPB* genų kopijos yra randamos gyvūnuose, *Acanthamoeba*, tam tikruose dumbliuose (pvz, *Coccomyxa*), *Naegleria*, kinetoplastiduose, tam tikrose bakterijose ir bakteriofaguose (Iyer ir kt., 2014). Tuo tarpu grybai, o ypač papėdgrybiai (kepurėtieji grybai, rūdys ir teliomicetai), pasižymi grupei specifiniu daugybės *TET-JPB* genų išplitimu, kurie dažnai yra susiję su unikalia judriųjų elementų klase KDZ (*Kyakuja*, *Dileera*, *Zisupton*; Iyer ir kt., 2009, 2013, 2014).

Papėdgrybių (*Basidiomycota*) *Laccaria bicolor* ir *Coprinopsis cinerea* analizė parodė, kad dauguma *TET-JPB* genų (atitinkamai 47 iš 74 ir 32 iš 47 homologų) galėtų koduoti katalitiškai aktyvius baltymus (Iyer ir kt., 2009, 2014). *L. bicolor* ir *C. cinerea* priklauso *Agaricales* eilei, didelei papėdgrybių grupei. Šie grybai buvo tirti kaip modeliniai organizmai DNR metilinimo analizėje (Zemach ir kt., 2010). *L. bicolor* ir *C. cinerea* grybuose nustatytos DNMT1 ir DNMT5 metiltransferazės, o metilinimas aptinkamas daugiausia pasikartojančiose sekose (Bewick ir kt., 2019; Zemach ir kt., 2010). *In vitro* buvo parodytas *C. cinerea* TET-JPB baltymų aktyvumas su DNR (Zhang ir kt., 2014). Taip pat *C. cinerea* genomineje DNR aptiktos oksimC ir naudojant SMRT sekoskaitą (žr. 1.4.6) nustatytas bendras jų signalas beveik nukleotido tikslumu (Chavez ir kt., 2014). Oksi-mC nustatytos daugiausia genuose be raiškos ir pasikartojančiose sekose, tačiau atskirų oksimC pasiskirstymas ir poveikis genų raiškai nėra aiškus. Tuo tarpu *L. bicolor* oksimC dar visai nebuvo tirtos.

Didelių daugialąsčių grybų vystymosi procesų tyrimuose saprotrofinis *C. cinerea* (pilkasis mėšlagrybis, dar vadinamas *Coprinus cinereus*) yra klasikinis modelinis organizmas, kuris padėjo geriau suprasti grybų vystymąsi, dauginimosi tipų genus bei fotomorfogenezės mechanizmus (Pukkila, 2011; Stajich ir kt., 2010). *C. cinerea* gana lengvai auginamas laboratorijoje (1.10 pav. D, E), kur užbaigia visą savo gyvenimo ciklą (1.10 pav. A) tik per dvi savaites ir gali būti genetiškai manipuluojamas. Tuo tarpu didžiąją daugumą kepurėtųjų grybų auginti laboratorijos sąlygomis daug sunkiau arba nepavyksta pasiekti visų vystymosi ciklo stadijų (Kües, 2000).

Sudugus sporoms formuojasi ilgi ląstelių filamentai (hifai). Susiliejus lytiškai suderinamiems (t. y. su skirtingais poravimosi tipo aleliais) haploidiniams hifams, susidaro diploidinis dikariotinis micelis (kitaiip grybiena), kuriame kiekvienoje ląstelėje yra po haploidinį branduolį iš kiekvieno pirminių hifų. Esant tinkamoms sąlygoms (šviesai, temperatūrai, maisto medžiagoms) dikarionas formuoja vaisiakūnius (1.10 pav. D). Vaisiakūnio ląstelėse, vadinamose papėdėmis (bazidėmis), įvyksta branduolių susijungimas (kariogamija), mejozė ir susidaro haploidinės sporos (Martin ir Selosse, 2008; Pukkila, 2011).

*L. bicolor* (dvispalvė lakabudė) yra naudojama kaip eksperimentinis modelis ektomikorizės simbiozės sąveikoje dalyvaujančių genų ir ekologinės adaptacijos tyrimuose (1.10 pav. B, E). *L. bicolor* yra svarbi kietmedžių ir spygliuočių rūšims, tačiau neturi griežto specifiškumo šeiminkui ir yra gana plačiai geografiškai paplitusi (Martin ir Selosse, 2008). Ektomikoriziniai grybai formuoja mutualistinę (abipusiai naudingą) simbiozę, kuri būdinga daugumai medžių rūšių ir miškų ekosistemų. Šie grybai gali turėti dvejopą gyvenimo būdą: formuoja simbiozę su augalo šaknimis arba yra laikini saprotrofai dirvožemyje (Genre ir kt., 2020). Dikarionas yra dominuojanti *L. bicolor* forma (kaip ir daugumos Agarikomietų), jo vegetatyvinis micelis formuoja ektomikorizę (Martin ir Selosse, 2008). Augalų ir grybų simbiotiniai ryšiai atlieka esminį vaidmenį miško medžių ekologijoje ir biologijoje, nes daro įtaką jų augimui pagerindami vandens ir būtinų maisto medžiagų įsisavinimą, saugo nuo šaknų ligų, palaiko atsparumą stresui ir kenkėjams (Martin ir kt., 2016). Be to, skirtingų augalų ryšiai su ta pačia grybų sancaupa per medžiagų pernašą lemia konkurenciją ar netiesiogines sąveikas tarp augalų ir yra labai svarbūs ekosistemos funkcionavimui. Medžiai maitina grybą angliavandeniais, o grybas gautą energiją panaudoja azoto, fosfato junginių, kitų mineralinių medžiagų, kurias perduoda augalui, įsisavinimui iš dirvožemio (ar net iš tarp hifų įstrigusios mikrofaunos *L. bicolor* atveju) skaidant įvairias medžiagas. Mikorizę būtina suprasti, nes miškai yra labai svarbūs CO<sub>2</sub> fiksavimui, o taip pat siekiant optimizuoti medžių biokuro produktyvumą (Martin ir Selosse, 2008). Ektomikoriziniai grybai gali būti pritaikomi miškų atstatyme po gaisro, ekologiniame kasyklų rekonstravime bei sodinukų produkcijoje (Genre ir kt., 2020).

Ektomikoriziniai grybai sąveikauja su smulkiomis augalo šoninėmis šaknimis ir jų kolonizacija sukelia žymius grybo ir šaknies morfologinius pokyčius suformuojant hibridinį organą – ektomikorizinį šaknies galiuką. Jame grybo hifai sluoksniais apsupa šaknies paviršių, formuodami grybo apvalkalo audinį (vadinamą mantija). Šaknies tarpląstelinėje erdvėje hifai

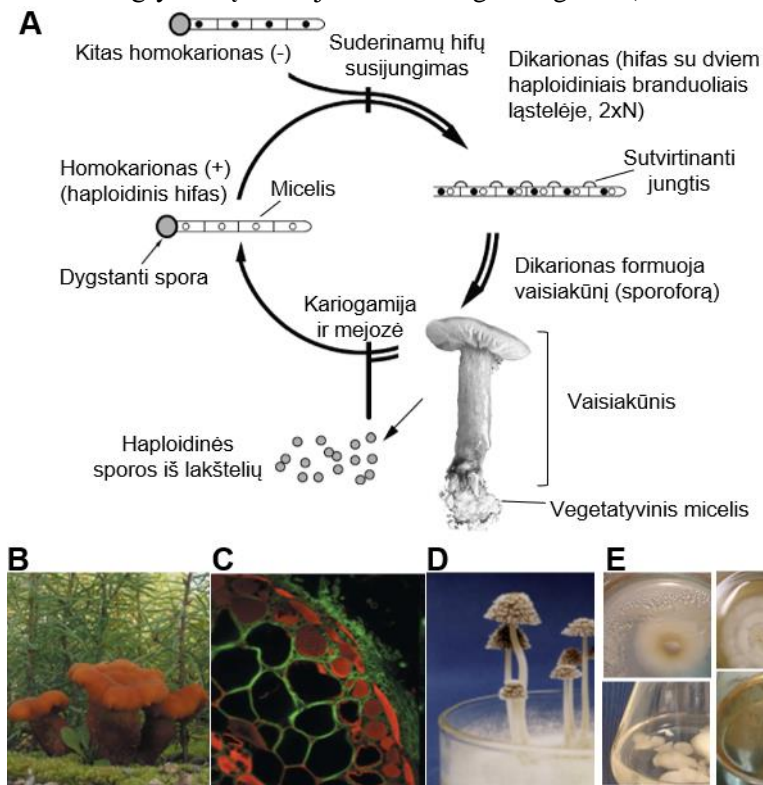
formuoja sudėtingą grybo ir augalo epidermio bei šerdies ląstelių labirintą – *Hartig* tinklą (1.10 pav. C), kuriame maisto medžiagomis apsikeičiama per didelį ląstelių sąveikos paviršių neįsiterpiančią į augalo ląstelę (Martin ir kt., 2016). Užtikrindami ektomikorizinę sąveiką grybai mažina augalo ląstelių sukibimą ardydami pektinus, skatina šaknų ląstelių dauginimąsi didindami augalų augimo hormono auksino kaupimąsi bei švelnina gynybinę atsaką išskirdami mažiau jį aktyvuojančių medžiagų (Genre ir kt., 2020). *L. bicolor*, kaip ir kiti simbiozę sudarantys grybai, pasižymi didele mažų (<300 ar.) sekretuojamų baltymų įvairove. Šie baltymai yra indukuojami simbiozės ir tarnauja kaip efektoriai, skatinantys mutualistinę sąveiką ir paveikiantys augalo atsaką (Martin ir Selosse, 2008). *L. bicolor* taip pat išskiria lipochito-oligosacharidų mišinį, kuris aktyvuoja skirtingiems augalams bendrą simbiotinį signalinį kelią jų šaknyse. Tuo tarpu dauguma grybų genų, kurie yra aktyvuojami simbiozės, būdingi konkrečiai grybų rūšiai (Genre ir kt., 2020).

Lyginant mikorizinių grybų genomus su saprotrofais (pavyzdžiui, *L. bicolor* su *C. cinerea*), galima stebėti skirtumus, kurių daugelis tikriausiai yra atsiradę dėl ekologinių nevienodumų. Saprotrofai turi platų augalų ląstelės sienelės angliavandenius (ligniną, celiuliozę, hemiceliuliozę ir pektinus) ardančių fermentų asortimentą, kuris būtinas organinės paklotės skaidymui. Tuo tarpu simbiotuose stebimas šių fermentų praradimas (Miyachi ir kt., 2020).

Pastebėta, kad mikorizinių grybų genomai yra didesni lyginant su saprotrofiniais. Pagrindinė šių genomų didėjimo varomoji jėga atrodo yra TE kiekis, kuris yra didžiausias ektomikoriziniuose simbiotuose, lyginant su kitais ekologinio gyvenimo būdais. Spekuliuojama, kad TE gausėjimas gali skatinti simbiozei specifinių faktorių evoliuciją (Miyachi ir kt., 2020). *L. bicolor* pasižymi didele judriųjų elementų įvairove ir jie sudaro didelę genomo dalį (iki 37,9 %; Castanera ir kt., 2016), tuo tarpu *C. cinerea* ji daug mažesnė (~6%; Stajich ir kt., 2010). Abiejuose grybuose gausiausi yra retrotranspozonų judrieji elementai (Labbé ir kt., 2012). TE daro didelę įtaką eukariotų genomų evoliucijai, todėl yra svarbus grybų genetinės įvairovės šaltinis. TE gali sukelti įvairius genominius persitvarkymus: duplikacijas (sukeliamos retrotranspozonų), delecijas, įterpimus, translokacijas (Deniz ir kt., 2019). TE sukelti persitvarkymai gali būti neutralūs ar žalingi, tačiau gali lemti ir adaptyvų poveikį populiacijoms prisitaikant prie aplinkos apribojimų (Castanera ir kt., 2017). TE gali skatinti cis-reguliacinių elementų, naujų sulietų genų ar alternatyvių transkriptų (sukėlus egzonų praleidimą) susidarymą (Deniz ir kt., 2019). TE taip pat gali išveiklinti genus ar keisti jų

raišką suteikdami naujus promotorius, įterpdami alternatyvius poliadenilinimo signalus (Muszewska ir kt., 2019) ar keisdami DNR bei chromatino modifikavimą (Seidl ir Thomma, 2017).

Simbioziniuose grybuose stebimas naujų, grupėms specifinių, simbiozės indukuojamų genų įvairovės didėjimas. Daugumos numanomų *L. bicolor* genų funkcijos nežinomos ir net 24% yra būdingi tik šiai rūšiai (Martin ir kt., 2008). Tikėtina, kad genų šeimų ekspansija yra didele dalimi nulemta simbiozės sąveikos ir manoma, kad duplikacijos yra pagrindinė genų įvairovės ir specializavimosi simbiozei varomoji jėga (Martin ir Selosse, 2008). Tę persikėlimai *L. bicolor* genome didėjant jų tankiui tikriausiai sukėlė stebimą neįprastai didelį kai kurių baltymų genų šeimų išplitimą, tokių kaip hidrofobinių baltymai, kurie yra svarbūs simbiozei ir suteikia pranašumą šiam mutualistiniam grybui sąveikaujant su skirtingais augalais (Labbé ir kt., 2012).



**1.10 pav.** (A) Kepurėtiesiems papėdgrybiams būdingą vystymąsi atspindintis *Laccaria bicolor* ciklas (Martin ir Selosse, 2008). (B) *L. bicolor* S238N dikariono vaisiakūnis augantis po eglų (*Pseudotsuga menziesii*) sėjinukais (nuotrauka D. Vairelles; Martin ir kt., 2008). (C) Tuopos (*Populus trichocarpa*) šaknies ir *L. bicolor* S238N ektomikorizės skersinio pjūvio lazerinio skenuojančio konfokalinio mikroskopo nuotrauka (V. Pereda ir V. Legué; Martin ir kt., 2008). Grybo hifų (žalia) mantija apsupa šaknį, į tarpląstelinę erdvę įsiterpę hifai formuoja Hartig tinklą. (D)

*Coprinopsis cinerea* dikariono vaisiakūnis gilioje petri lėkštelėje (nuotrauka Hajime Muraguchi; Pukkila, 2011). (E) *L. bicolor* (kairėje) ir *C. cinerea* (dešinėje) haploidiniai miceliai augantys laboratorijoje, agarizuotoje (viršuje) ir skystoje (apačioje) terpėje (nuotraukos J. Ličytės).

## 1.2 RNR modifikacijos

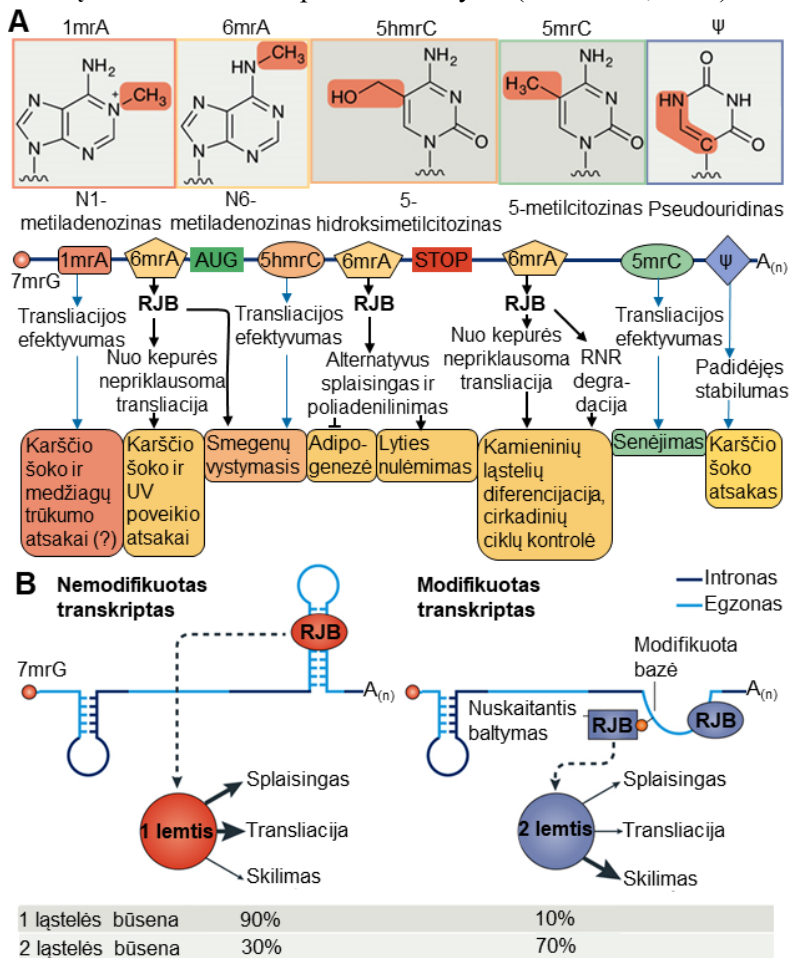
### 1.2.1 RNR modifikacijų paplitimas ir funkcijos

RNR nustatyta bent 160 skirtingų cheminių modifikacijų (<http://mods.rna.albany.edu>; Cantara ir kt., 2011). Vyraujančios žinduolių RNR modifikacijos (1.11 pav. A) yra pseudouridinas ( $\Psi$ ), 2'-O-metilimas, N6-metiladenozinas (6mrA), N1-metiladenozinas (1mrA) ir 5-metilcitozinas (5mrC) (Xue ir kt., 2020). Modifikacijos dalyvauja visuose RNR metabolizmo procesuose veikdamos RNR struktūrą ir sąveikas su baltymais, kurie gali dalyvauti ląstelės signaliniuose keliuose jai reaguojant į aplinkos veiksnius (Nachtergaele ir He, 2017). Daugybė RNR modifikacijų gali daryti įtaką genų reguliacijai, todėl atsirado sritis, vadinama „RNR epigenetika“ arba „epitranskriptomika“, kuri tiria potranskripcinius pokyčius, nepakeičiančius RNR sekos (Zhao ir kt., 2020).

RNR modifikacijos dalyvauja įvairiuose svarbiuose biologiniuose procesuose (1.11 pav. A), tokiuose kaip ląstelių diferenciacija, lyties nulėmimas ir atsakas į temperatūros ar medžiagų trūkumo stresą (Jonkhout ir kt., 2017). Daugelio RNR modifikuojančių fermentų mutacijos ar reguliacijos sutrikimai sukelia įvairias ligas, pvz.: padidėjęs tam tikrų modifikuotų nukleozidų kiekis yra pastebimas vėžiu sergančių pacientų šlapime (Kellner ir kt., 2010), o 6mrA metiltransferazės Mett13 praradimas sukelia pelių ir augalų gemalų žūtį (Dezi ir kt., 2016).

Nekoduojančiose RNR modifikacijos yra kritiškai svarbios jų teisingam funkcionavimui transliacijoje bei splaisinge (Dezi ir kt., 2016). Transportinėse RNR (tRNR) nustatomas didžiausias modifikacijų gausumas ir įvairovė. tRNR modifikacijų visuma užtikrina transliaciją, veikdama tRNR susilankstymą ir informacinės RNR (iRNR) iškodavimą (Helm ir Motorin, 2017). Labai gausiose ilgose ribosominėse RNR (rRNR) taip pat aptinkama didelė modifikacijų įvairovė ir gausumas. Įvairių rRNR modifikacijų visuma užtikrina ribosomų veikimo efektyvumą. Modifikacijos yra linkusios pasiskirstyti labai konservatyviuose, svarbiausiose rRNR funkcijai srityse, kurios yra būtinos ribosomų biogenezei ir tikslios bei veiksmingos baltymų sintezės palaikymui (Nachtergaele ir He, 2017). Tuo tarpu gana nedidelė modifikacijų įvairovė aptikta ne tokiose gausiose nekoduojančiose RNR, pvz., mažosiose branduolio ar branduolėlio RNR bei mikroRNR (miRNR) (Helm ir Motorin, 2017).

Anksčiau manyta, kad modifikacijos yra stabilios ir daugiausia reguliuoja chemostruktūrinės nekoduojančių RNR savybes. Dabar RNR modifikacijos suvokiamos kaip dinamiškai reguliuojami, kintantys ir grįžtami potranskripciniai reguliatoriai, greitai reaguojantys į ląstelių signalus. Kad RNR modifikacijos gali būti grįžtamos ir dinamiškos buvo suprasta atradus, kad žinduolių informacinėse RNR 6mrA gali būti tiesiogiai demetilinamas oksidacijos būdu veikiant nuo  $Fe^{2+}/\alpha$ -KG priklausomai dioksigenazei FTO ir jos homologui ALKBH5 (Shen ir kt., 2014). Šios reakcijos metu susidaro N6-hidroksimetiladenozino ir N6-formiladenozino produktai, kurie suskyla atgal į adenoziną dėl nestabiliu tampačio C–N ryšio (Shen ir kt., 2014).



**1.11 pav.** RNR modifikacijos ir jų funkcijos informacinėje RNR. (A) Modifikuotų iRNR bazių cheminės struktūros (viršuje) ir jų funkcijos (apačioje). Parodyti žinomų modifikaciją turinčią RNR jungiančių baltymų (RJB) paveikiami iRNR procesai ir biologinė reikšmė organizmui. Mėlynos rodyklės rodo kelius, kuriems nėra žinomų modifikaciją nuskaitančių baltymų. Pavyzdžiui, 5mrC dalyvauja su replikaciniu

senėjimu susijusių genų transliacinėje kontrolėje. 5-hidroksimetilcitozinas iRNR padidina transliacijos efektyvumą *D. melanogaster*, jo lygis didžiausias smegenyse ir 5hrmC daro įtaką jų vystymuisi. A(n) – poliadenilinizimas; 7mrG – 7-metilguanozinas; AUG – transliacijos pradžios kodonas; STOP – pabaigos kodonas. **(B)** RNR modifikavimo stechiometrijos dinamika. Tam tikro geno RNR yra modifikuojama tik dalyje transkriptų. Modifikavimas tam tikromis ląstelių sąlygomis gali būti labai žemas (1 ląstelės būseną). Modifikacijoms sukėlus RNR struktūrinius pertvarkymus ir pritraukus ją atpažįstančius baltymus, modifikuotos iRNR įgauna kitokią lemtį, dėl kurios, pavyzdžiui, gali padidėti skilimas. Keičiantis ląstelių sąlygoms, gali pasikeisti ir modifikacijų stechiometrija, o tai gali sukelti didesnę (2 ląstelės būseną) modifikuotų transkriptų procentą dėl nuskaitančių, modifikuojančių ir (arba) modifikaciją panaikinančių baltymų aktyvumo ar raiškos pokyčių. Brūkšninės rodyklės rodo netiesioginį poveikį. Pagal Lewis ir kt. (2017a).

Nors dauguma modifikacijų yra nekoduojančiose RNR, tačiau kai kurios modifikacijos taip pat aptinkamos iRNR. Daugėja duomenų, kad jos gali veikti visus genų raiškos aspektus, kisti reaguojant į aplinkos veiksnius ir reguliuoti ląstelės procesus (1.11 pav. A) (Dezi ir kt., 2016). Pavyzdžiui, 1mrA skatina nuo iRNR kepurės priklausomą transliaciją, o 1mrA lygius keičia maistinių medžiagų trūkumas arba karščio šokas, galbūt per 1mrA nukreipiant ląstelių atsakus šiomis sąlygomis. Tuo tarpu 6mrA palengvina nuo kepurės nepriklausomą transliaciją stiprindamas sąveiką su eukariotų transliacijos iniciacijos faktoriaus 3 (eIF3) kompleksu ląstelei reaguojant į šilumos šoką arba UV poveikį. Be to, iRNR stabilumo reguliavimas per 6mrA yra labai svarbus kamieninių ląstelių diferenciacijai ir cirkadinio ciklo valdymui. O iRNR 3' netransliuojamuose regionuose (UTR, angl. *untranslated region*) esantis pseudouridinas padidina transkriptų stabilumą šilumos šoko metu. Daugiau modifikacijų poveikio iRNR procesams ir funkcijoms pavyzdžių pateikta (1.11 pav. A).

Daugėja duomenų, kad potranskripcinė genų reguliacija priklauso ne tik nuo iRNR sekos, bet ir nuo aptinkamo susilankstymo į antrines struktūras, kurios daro įtaką baltymų prisijungimui prie iRNR, splaisingui, transliacijai, lokalizacijai ląstelėje ir stabilumui (Harcourt ir kt., 2017). Pavyzdžiui, nestruktūruotos iRNR pasižymi didesniu transliacijos efektyvumu, o kai kurie poli A „uodegos“ neturintys iRNR transkriptai sudaro 3'UTR struktūras, didinančias stabilumą ir transliaciją (Lewis ir kt., 2017a). Seniai žinoma, kad modifikacijos keičia tRNR ir rRNR struktūras ir taip veikia jų funkcijas, todėl veidamos iRNR struktūras modifikacijos taip pat galėtų plačiai paveikti minėtus jų procesus. Net mažų grupių, tokių kaip metilo, pridėjimas veikia RNR struktūrą: bazių gebėjimą poruotis, sąveiką su gretimomis bazėmis, vienos konformacijos ar susilankstymo įgijimą prieš kitą (Harcourt ir kt., 2017). Pavyzdžiui, 6mrA atveju parodyta, kad jis destabilizuoja RNR

dupleksus *in vitro*, o *in vivo* taip pat nustatyta 6mrA modifikuotų iRNR vietų tendencija nesudaryti struktūrų. Dinamiškos ir grįžtamos RNR modifikacijos gali stabilizuoti arba sutrikdyti struktūrinius elementus bei moduluoti prie RNR besijungiančių baltymų sąveiką. Taip gali būti paveikiama iRNR transliacija ar skilimas greitai reaguojant į aplinkos pokyčius (1.11 pav. B) (Lewis ir kt., 2017a). RNR modifikacijos gali būti viena iš pagrindinių transkripto struktūrinio plastiškumo varomųjų jėgų. Keičiant tam tikrą modifikaciją turinčių transkriptų dalį, galėtų būti dinamiškai sukuriamas to paties RNR transkripto funkcinė įvairovė (1.11 pav. B). Dalies transkriptų modifikavimas sukurtų dvi skirtingas iRNR rūšis, kurios galėtų skirtis savo struktūromis arba prie jų prisijungiančiais baltymais (Lewis ir kt., 2017a).

### 1.2.2 5-metilcitozino funkcijos RNR

Nors ankstesniuose RNR metilinimo tyrimuose daugiausia dėmesio buvo skiriama 6mrA, palaipsniui buvo atskleistas ir 5mrC vaidmuo potranskripciniame reguliavime (Xue ir kt., 2020). 5mrC yra plačiai paplitęs tiek koduojančiose, tiek nekoduojančiose įvairių organizmų RNR iš visų gyvybės domenu (Edelheit ir kt., 2013; Squires ir kt., 2012). Jis nustatytas ne tik rRNR, iRNR, tRNR, bet ir virusinėse RNR, skliautinėse RNR (angl. *vault*), mažose branduolio ir branduolėlio RNR, pseudogenuose, natūraliuose antiprasminiuose transkriptuose, stipriklių RNR (eRNR), ilgose nekoduojančiose RNR ir mikro RNR (miRNR). Parodyta, kad 5mrC ir jo reguliatorių pakitimai veikia RNR stabilumą, genų raišką bei baltymų sintezę, todėl daro didelį poveikį įvairiems ląstelių procesams (Xue ir kt., 2020).

Ribosominėse RNR 5mrC yra svarbus struktūros palaikymui, transliacijos tikslumui, tRNR atpažinimui, ribosominių baltymų prisijungimui ir yra būtinas ribosomų biogenezei. Pavyzdžiui, 25S RNR 5mrC ir ribozės metilinimo praradimas (atitinkamai C2278 ir G2288 pozicijoje) lemia struktūrinius pokyčius ir dalies 60S didžiojo subvieneto ribosominių baltymų praradimą, dėl to sutrinka ribosomų biogeneze (Xue ir kt., 2020).

tRNR 5mrC yra išsidėstęs aplink kintamąjį regioną ir antikodono kilpą. Jis stabilizuoja antrinę tRNR struktūrą, veikia aminoacilinimą, kodono atpažinimą, saugo nuo degradacijos ir reguliuoja ląstelių metabolizmą (Xue ir kt., 2020). Pavyzdžiui, 5mrC *S. cerevisiae* tRNR-Phe 40 padėtyje svarbus Mg<sup>2+</sup> jungimuisi, kuris sukelia visos antikodono kilpos konformacinius pakitimus (Motorin ir kt., 2010). Žmogaus, pelės ir augalų 5mrC susidarymas veikiant NSUN2 (TRM4) ir DNMT2 metiltransferazėms apsaugo tRNR nuo oksidacinio streso sukkelto skilimo (Zhao ir kt., 2017). Metilinimo perprogramavimas yra svarbus atsakas į stresą: paveikus vandenilio



peroksidu, mielių tRNA-Leu-CCA padaugėja 5mC ir tai skatina streso atsako baltymų transliaciją (Helm ir Alfonzo, 2014). Išveiklinus DNMT2 ir NSUN2 metiltransferazes, tRNR praranda 5mC ir tai sukelia jos nestabilumą bei tRNR fragmentų kaupimąsi (Liu ir Jia, 2014). Taip yra slopinama baltymų sintezė bei palaikoma kamieninė ląstelių būseną. Praradus NSUN2 tokiu pat būdu sustiprinamas ir auglio kamieninių ląstelių atsinaujinimas. Tačiau tokios ląstelės yra jautrios stresui, nes nesugeba aktyvinti streso atsako kelių, kuriems reikalingas tRNR metilinimas (Nachtergaele ir He, 2017). Suskilusios tRNR fragmentai gali reguliuoti genų raišką ir šių fragmentų kaupimasis siejamas su neuronų praradimu, vėžio progresavimu (Yu ir kt., 2021; Nachtergaele ir He, 2017).

5mC paplitimas iRNR gali smarkiai skirtis skirtinguose organizmuose: žmogaus ląstelėse nustatyta ~8000 metilintų vietų tarp visų iRNR, o tuo tarpu mielių ląstelėse nustatyta viena metilinta pozicija tarp visų iRNR (Gilbert ir kt., 2016). 5mC praturtintas ląstelei specifiniuose iRNR transkriptuose, daugiausia 5' netransliuojamuose regionuose netoli transliacijos pradžios kodono ir kai kurių transkriptų 3'UTR (Amort ir kt., 2017). Kadangi 5mC yra praturtintas iRNR 3'UTR netoli numanomų struktūras sudarančių reguliacinių sekų bei Argonauto 1-4 baltymų (pagrindiniai genų raišką slopinančio miRNR/RISC komplekso komponentai) jungiamose iRNR sekose, todėl 5mC galbūt daro įtaką iRNR transliacijai ir stabilumui (Squires ir kt., 2012).

5mC gali veikti įvairius iRNR procesus. Pavyzdžiui, parodyta, kad 5mC įgalina iRNR transportą iš branduolio į citoplazmą žmogaus HeLa ląstelėse, o taip pat suaktyvina iRNR transportavimą į tolimas *Arabidopsis thaliana* augalo dalis, taip reguliuodamas vystymąsi ir skatindamas šaknų augimą (Xue ir kt., 2020). Yra šiek tiek duomenų, kad 5mC gali paveikti genų funkcijas vykstant reguliuojamam genetinio kodo perjungimui. Į iRNR įvedus 5mC, 4% prolino *E. coli* buvo perkoduojami kaip leucinas. Tokiu atveju, kai iRNR modifikavimas sukelia alternatyvius baltymų variantus, net ir retai išsidėsčiusios modifikacijos gali daryti didelį biologinį poveikį (Gilbert ir kt., 2016). Taip pat gali kisti transliacijos efektyvumas: pavienių 5mC (o taip pat 6mC arba Ψ) įvedimas į tam tikrus kodonus sumažina baltymų gamybą *E. coli* nuo 20 iki 70 %. 5mC veikia tRNR susilankstymą, todėl tikėtina veikia ir galinčias susidaryti iRNR struktūras bei nuo jų priklausančius iRNR procesus (Gilbert ir kt., 2016). Nors metilinimas citozino 5 padėtyje labai mažai veikia bazių poravimąsi, bet padidina didžiojo RNR griovio hidrofobiškumą ir tikriausiai sustiprina stekingo sąveikas (Harcourt ir kt., 2017).

5mC aptinkamas įvairiose reguliacinėse RNR. Pavyzdžiui, jis buvo aptiktas XIST ir HOTAIR ilgose nekoduojančiuose RNR, kurios dalyvauja histonų modifikavimo reguliavime (Amort ir kt., 2013). XIST yra reikalinga žinduolių moteriškų ląstelių ilgalaikiam vienos X chromosomos išveiklinimui, o HOTAIR dalyvauja organizmo vystymuisi svarbių genų reguliacijoje. 5mC buvo nustatytas HOTAIR ir XIST RNR regionuose (arba arti jo), kurie sąveikauja su chromatiną modifikuojančiais kompleksais. 5mC paveikė XIST funkciją, trukdydamas prisijungti baltymų kompleksui PRC2, kuris slopina genų raišką atlikdamas H3K27 metilinimą (Amort ir kt., 2013). Taip pat 5mC yra dažnai aptinkamas miRNR. Pavyzdžiui, buvo nustatyta, kad subrendusios miR-181a-5p metilinimas slopina duplekso su iRNR formavimąsi ir tai panaikina naviko slopinimo poveikį bei siejasi su prasta daugiaformės glioblastomos prognoze (Cheray ir kt., 2020). O ląstelės signalinių kelių reguliacijoje dalyvaujančiuose skliautinėse RNR aptinkamas 5mC skatina iš jų susidaryti trumpas RNR, kurios veidamos kaip miRNR slopina transliaciją ir skatina diferenciaciją į epidermį (Xue ir kt., 2020).

5mC gali jungti metabolinės būklės informaciją su transliacija ir daryti įtaką genų reguliavimo tinklams (Helm ir Alfonzo, 2014). 5mC modifikuotos eRNR sąveikauja su specifinių metabolinių genų stiprikliais ir tikriausiai sustiprina jų sąveikas su promotoriais. 5mC pagausėja šiose specifinėse eRNR rūšyse po metabolinio badavimo streso poveikio kepenyse. 5mC metiltransferazės NSun7 slopinimas lemia šių eRNR sumažėjimą, vadinasi, 5mC veikia eRNR stabilumą (Aguilo ir kt., 2016).

Taigi RNR 5mC metilinimas yra būtinas normalioms ląstelių ir organizmų fiziologinėms funkcijoms palaikyti, o nenormalus jo pasiskirstymas ir gausa yra susiję su įvairiomis ligomis. 5mC metiltransferazės NSUN2 raiška yra didesnė smegenyse, o jos mutacijos sukelia protinę negalią (Dezi ir kt., 2016). Buvo parodyta, kad RNR 5mC modifikacija ir jos reguliatorių pokyčiai susiję su šlapimo pūslės vėžio, kepenų karcinomos, daugiaformės glioblastomos ir leukemijos patogenezė. Todėl 5mC gali būti perspektyvus vėžio diagnostikoje, prognozėje ir gydyme (Xue ir kt., 2020). Įdomu, kad virusų RNR taip pat gali turėti 5mC, kuris slopina arba skatina replikaciją ir transliaciją (Xue ir kt., 2020).

### **1.2.3 Oksiduotų 5-metilcitozino formų funkcijos RNR**

Buvo parodyta, kad *in vitro* TET baltymai gali oksiduoti RNR 5-metilcitoziną iki visų trijų oksimC (5hmC, 5fC, 5caC) (Basanta-Sanchez ir kt., 2017; Fu ir kt., 2014). Taip pat 5hmC gali susidaryti iš 5mC (Huber ir

kt., 2015) *in vivo*, dalyvaujant TET oksigenazėms (Fu ir kt., 2014). Vadinas, RNR gali turėti tą patį demetilinimo mechanizmą kaip ir DNR.

5m<sup>r</sup>C poveikis RNR funkcijoms yra labiau ištirtas, nei iš jo galinčių susidaryti oksim<sup>r</sup>C. Manoma, kad 5m<sup>r</sup>C gali būti dinamiška modifikacija ir galbūt 5m<sup>r</sup>C įtaka tos RNR funkcijai galėtų būti keičiama vykstant oksidacijai. Oksim<sup>r</sup>C buvo aptiktos koduojančiose ir nekoduojančiose RNR, įvairiuose organizmuose: žinduoliuose, augaluose, vabzdžiuose (Delatte ir kt., 2016; Fu ir kt., 2014; Huang ir kt., 2016; Huber ir kt., 2015; Zhang ir kt., 2016). Dėl C-C ryšio su baze fiziologinėmis sąlygomis oksim<sup>r</sup>C yra chemiškai stabilios (skirtingai nei 6m<sup>r</sup>A oksidacijos produktai), todėl galėtų būti funkciškai svarbūs (Shen ir kt., 2014).

Nors 5m<sup>r</sup>C oksiduotų formų funkcijos RNR vis dar mažai žinomos, manoma, kad jos gali daryti įtaką RNR stabilumui ir transliacijai (1.11 pav. A). Pavyzdžiui, buvo parodyta, kad 5hm<sup>r</sup>C gali skatinti iRNR transliaciją vaisinėje muselėje *D. melanogaster* (Delatte ir kt., 2016). 5hm<sup>r</sup>C yra gausiausias *Drosophila* smegenyse ir yra funkciškai svarbus, nes išveiklinus TET ir dėl to sumažėjus 5hm<sup>r</sup>C, smegenų vystymasis sutrinka, muselės žūsta lėliukės stadijoje (Delatte ir kt., 2016). 5hm<sup>r</sup>C aptiktas įvairiuose pelės audiniuose, žmogaus smegenyse ir vėžinių ląstelių kultūrose (Fu ir kt., 2014). 5hm<sup>r</sup>C kiekiai sumažėja žmogaus vėžiniuose audiniuose (Huang ir kt., 2016) ir Parkinsono ligos modelinių pelių smegenyse (Miao ir kt., 2016). Žinduoliuose 5hm<sup>r</sup>C gausiausias iRNR (Huang ir kt., 2016; Huber ir kt., 2015), todėl jis gali daryti įtaką genų reguliacijai. Pavyzdžiui, žinduolių Socs3 iRNR įvykus TET2 vykdomai 5m<sup>r</sup>C oksidacijai, susidaręs 5hm<sup>r</sup>C skatina prisijungti Adar1 baltymą, kuris destabilizuoja iRNR. Oksidacija skatina infekcijos sužadintą mielopoezę, nes Socs3 slopina JAK-STAT signalinį kelią, kuris yra būtinas citokinų skatinamam mieloidinių leukocitų susidarymui (Shen ir kt., 2018). Įdomu, jog pelių endogeninių L retrovirusų transkriptai, kurie yra aktyvuojami dviejų ląstelių stadijos embrionuose ir koreliuoja su totipotentiškumu, gali būti reguliuojami veikiant TET2 – 5hm<sup>r</sup>C susidarymas šiuose transkriptuose sukelia jų skilimą (Amort ir kt., 2013; Guallar ir kt., 2018).

5f<sup>r</sup>C pirmiausia buvo nustatytas jaučio ir askaridės mitochondrinėje tRNR-Met. Vėliau 5f<sup>r</sup>C nustatytas ir įvairių kitų rūšių (žmogaus, vištos, varlės, vaisinės muselės, kalmaro) mitochondrinėje tRNR-Met, todėl manoma, kad tai gali būti universali šios RNR modifikacija eukariotuose (Van Haute ir kt., 2017). 5f<sup>r</sup>C yra būtinas AUA (neuniversalus kodonas mitochondrijose, įprastai koduoja izoleuciną) teisingam atpažinimui ir transliacijai. 5f<sup>r</sup>C taip pat aptiktas citoplazminėje tRNR-Leu (Kawarada ir kt.,

2017). 5frC nustatytas žmogaus ląstelių kultūrų ir audinių, įvairių graužikų audinių visuminėje RNR (Huang ir kt., 2016; Huber ir kt., 2015; Jiang ir kt., 2017). 5frC taip pat buvo aptiktas pelės kepenų ir mielių iRNR, tačiau galima tokio modifikavimo funkcija nežinoma (Huang ir kt., 2016; Tardu ir kt., 2019).

Neseniai žmogaus bei pelės ląstelių ir audinių RNR buvo aptiktas 5carC, pelės kepenyse jis gausiausias iRNR (Huang ir kt., 2016; Xiong ir kt., 2017). Tačiau dar nėra daugiau žinių apie jo paplitimą ir funkcijas, o 5hmrC arba 5frC jau yra ištirti daugiau (Huber ir kt., 2015; Roundtree ir kt., 2017; Wang ir kt., 2021; Zhang ir kt., 2016).

Reikėtų pažymėti, kad pelės embrioninėse kamieninėse ląstelėse, kuriose išveiklinti visi TET baltymai, vis dar aptinkamas nedidelis 5hmrC kiekis (Fu ir kt., 2014). Vadinasi, 5mrC oksidaciją gali katalizuoti ir kiti fermentai. Be to, 5hmrC modifikacija taip pat aptinkama *C. elegans* ir *A. thaliana*, kurių genomuose nėra TET ortologų (Huber ir kt., 2015). Naujausi tyrimai parodė, kad mitochondrinėje tRNR-Met ALKBH1 gali oksiduoti 5mrC į 5hmrC ir tuomet į 5frC, o 5carC neaptiktas (Kawarada ir kt., 2017). ALKBH1 pašalinimas stipriai sumažina transliaciją mitochondrijose bei daro neigiamą poveikį jų funkcijai (Kawarada ir kt., 2017).

### 1.3 Citozino modifikacijų nustatymo metodai

Siekiant ištirti ir suprasti nukleorūgščių modifikacijų funkcijas, dažnai pirmaisia yra įvertinami globalūs jų lygiai prieš atliekant sudėtingesnius ir brangesnius eksperimentus, pavyzdžiui, sekoskaitą (Song ir kt., 2012).

Radioaktyviai pažymėtų nukleotidų **plonasluoksnė chromatografija** (TLC, angl. *thin layer chromatography*) – bene seniausias metodas, anksčiau plačiai naudotas RNR ir DNR modifikacijų analizei, kuriuo pirmaisia buvo nustatytos kai kurios modifikacijos (Song ir kt., 2012). Jame nukleotidai žymimi fosforilinant  $^{32}\text{P}$  arba  $^{33}\text{P}$  radionuklidais ir aptinkami autoradiografiniu būdu pagal skirtingą migraciją TLC celiuliozės plokštelėse. Metodą galima atlikti standartinėje laboratorijoje ir jis yra pigus (Kriukienė ir kt., 2012; Yuan ir Feng, 2014). Vadinamoje „artimiausio kaimyno“ žymėjimo strategijoje vienoje DNR grandinėje DNazė I atsitiktinai įveda trūkius, tada DNR polimerazė įjungia radioaktyviai pažymėtą  $[\alpha\text{-}^{33}\text{P}]\text{-dGTP}$  į trūkius ten, kur C yra priešingoje grandinėje. Tuomet DNR fermentiškai hidrolizuojama iki nukleotidų ir radioaktyvi fosfato grupė atsiduria ant 5' kaimyninio nukleotido. CG dinukleotiduose esantį modifikuotą citoziną taip pat galima praturtinti pritaikant R.MspI arba R.TaqI restrikcijos endonukleazes, kurios kerpa tarp

pirimidinų atitinkamai C<sup>3</sup>CGG ir T<sup>3</sup>CGA taikiniuose, nepriklausomai nuo C modifikavimo antroje pozicijoje. Susidarę DNR fragmentai su CG 5' gale pažymimi <sup>33</sup>P ir DNR suskaidoma iki 5'-NMP. Šio metodo jautrumas yra didesnis dėl selektyvesnio <sup>33</sup>P žymėjimo (matomi tik citozinai CCGG arba TCGA kontekste), tačiau tai apriboja analizę iki dalies CG taikinių (Kriukienė ir kt., 2012). TLC metodas gali būti varginantis, dažniausiai kokybinis, jautrumas priklauso nuo radioaktyvumo ir nukleotidų atskyrimo, tikslumas nėra didelis, todėl šis metodas retai naudojamas (Yuan ir Feng, 2014).

Visi nukleozidai gali būti analizuojami **skysčių chromatografija** (LC, angl. *liquid chromatography*). Šiame metode DNR fermentiškai hidrolizuojama iki nukleozidų, kurie dažniausiai išskirstomi didelio slėgio ir efektyvumo (HP, angl. *high-performance (pressure)*) atvirkščios fazės skysčių chromatografijos kolonėle ir gali būti kiekybiškai nustatomi UV detektoriumi (Yuan ir Feng, 2014). Analizei būtinas geras nukleozidų atskyrimas, tačiau, pavyzdžiui, 5hmC 2'-deoksিনukleozidas išplaunamas šalia 2'-deoksicitidino ir iš dalies yra užstojamas jo signalo, todėl nustatymas naudojant UV sugertį yra ne visuomet tinkamas rutininei analizei (Kriukienė ir kt., 2012). Be to, šis metodas nėra labai jautrus, reikia didesnių nukleorūgščių kiekių (Yuan ir Feng, 2014).

Didesnį tikslumą ir jautrumą galima pasiekti kartu su HPLC naudojant **masių spektrometrijos** (MS) detektorių ir nustatant nukleozidus pagal jų masę. Modernūs metodai pasitelkia mikro, nano kapiliarų ar itin didelio (angl. *ultra*) slėgio chromatografijos sistemas. Jos sumažina reikalingos medžiagos kiekį, trumpina išskirstymo laiką, palaiko didesnę medžiagų koncentraciją bei siauras ir aiškias signalo smailes, o tai padidina aptikimo jautrumą (Schwartz ir Motorin, 2017; Traube ir kt., 2019).

Masių spektrometrijos jautrumas taip pat priklauso nuo naudojamo prietaiso ir jo parametrų. Elektroįpurškimo jonizacijos (angl. *electrospray ionization*) šaltinis būtinas geresnei jonizacijai ir jautrumui (Traube ir kt., 2019). Tandemine masių spektrometrija (MS/MS) atskiriami tą patį masės ir krūvio santykį turintys nukleozidai, nes jie fragmentuojami (tipiškai per nukleozidinę jungtį) ir nustatomi iš jų susidarę skirtingi fragmentai (Kellner ir kt., 2010). Nustatymo jautrumą dar labiau padidina MS/MS naudojamas trigubo kvadrupolio masių spektrometras, dėl jo kelių reakcijų stebėjimo režimo (angl. *multiple reaction monitoring*), kuriame stebimas pirminis jonas ir jo fragmentas. Taip vietoje viso masių spektro atrenkamos ir stebimos tik pasirinktos dominančios molekulės. Gaunamas tam tikram nukleozidui unikalus signalas, kuris integruojamas ir kiekybiniam nustatymui lyginamas su sintetinių standartų kalibravimo kreivių tiesine lygtimi (Raiber ir kt., 2017;

Traube ir kt., 2019). „Auksiniu standartu“ laikomas kiekybinis įvertinimas, kai į mėginius įmaišomas vidinis, stabiliais izotopais pažymėtas standartas (atskiriamas nuo mėginio dėl kitokios masės). Dėl tokios pat jonizacijos ir sutampančio išėjimo laiko, tokie standartai suteikia papildomą patikimumą ir tikslumą nustatytam modifikacijos signalui (koreguojami tūrio, DNR kiekio, kartu išplaunamų medžiagų efektai) (Raiber ir kt., 2017; Yuan ir Feng, 2014).

Izotopinis žymėjimas ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$ ) LC-MS/MS taip pat pasitelkiamas norint sekti DNR modifikacijų gyvavimo trukmę, kilmę ir dinamiką. Pavyzdžiui, teikiant ląstelėms  $^{13}\text{C}^2\text{H}_3$  pažymėtą metioniną, kuris paverčiamas metilo donoru SAM kofaktoriuje (Hofer ir kt., 2019; Raiber ir kt., 2017).

HPLC-MS/MS analizėje kartu išeinančios medžiagos gali sukelti trukdžius – pavyzdžiui, gausūs nemodifikuoti nukleozidai silpnina retų oksimC signalą (Yuan ir Feng, 2014). Siekiant pašalinti šiuos trukdžius, 5fC ir 5caC nustatymui genomineje DNR pirmiausia gali būti reikalingas papildomas jų gryninimas HPLC frakcionavimu prieš tikslią HPLC-MS/MS analizę (Ito ir kt., 2011). Masės signalus dar galima sustiprinti prieš MS/MS atliekant modifikuotos DNR praturtinimą arba modifikuotų bazių cheminį derivatizavimą, kuris pagerina jonizaciją ir pakeičia išėjimo laiką (Li ir kt., 2018). Toks signalo stiprinimas pritaikytas retoms modifikacijoms nustatyti (Huang ir kt., 2016; Tang ir kt., 2014, 2015; Zhang ir kt., 2016).

Šiuolaikiniai masių spektrometrai gali tiksliai įvertinti DNR modifikacijas plačiame koncentracijų diapazone ir yra jautrūs – gali aptikti retų modifikuotų nukleozidų femtomolių eilės ar net mažesnius kiekius (Raiber ir kt., 2017). Globaliam genominei DNR oksimC lygių įvertinimui MS metodai tinka geriausiai, tačiau jiems reikia specialios įrangos, todėl jų negali atlikti kiekviena laboratorija (Kriukienė ir kt., 2012). HPLC-MS/MS yra svarbi ir plačiai taikoma technologija epigenetikos tyrimuose, nes jautrus kiekybinis modifikuotų DNR bazių aptikimas yra svarbus jų atradimui ir analizei organizmų genomuose. HPLC-MS/MS naudojamas modifikacijų skirtumų analizei tarp skirtingų organizmų, vėžio stadijų, audinių ir jų amžiaus. Modifikuotų bazių kiekybinį nustatymą galima pritaikyti padidintos tam tikrų modifikuojančių fermentų raiškos ar išveiklinimo poveikio įvertinimui (Raiber ir kt., 2017).

Dar vienas metodas – **kapiliarinė elektroforezė** – atskiria nukleozidus ar nukleotidus elektriniame lauke. Po to juos galima nustatyti UV ar masių spektrometrija (Kellner ir kt., 2010). Prie nukleotidų chemiškai prijungus fluorescuojančius dažus galima pasiekti jautrų nustatymą lazeriu indukuota fluorescencija (Cadet ir Wagner, 2014).

Citozino modifikacijų aptikimui taip pat naudojami įvairūs **antikūnais pagrįsti metodai**, kurie yra dažnai pasirenkami dėl jų paprastumo. Šie metodai visada priklauso nuo antikūno specifškumo ir jungimosi afiniškumo (Traube ir kt., 2019). Pagrindinis antikūnų trūkumas yra dažnai nustatomas kryžminis reagavimas, ypač tarp RNR ir DNR bazių, todėl į gautus duomenis gali būti įvedamos klaidos (Wu ir Zhang, 2014).

Imuninis chromogeninis dažymas ir imunofluorescencija yra taikomi 5mC ir jo oksiduotų formų nustatymui vizualizuojant jas įvairiose ląstelėse ir audiniuose (pvz., oocituose, zigotoje, kamieninėse ląstelėse, vėžiniuose audiniuose) (Cadet ir Wagner, 2014; Yuan ir Feng, 2014). Informatyvūs gali būti taškiniai imunoblotai arba imunofermentinės analizės metodas ELISA (angl. *enzyme-linked immunosorbent assay*), kuriame naudojamas chemiliuminescencinis aptikimas (Cadet ir Wagner, 2014). Tačiau dėl nelinijško signalo šie metodai suteikia tik pusiau kiekybinę informaciją (Song ir kt., 2012).

Taip pat modifikacijų aptikimui naudojami **cheminio ar fermentinio žymėjimo būdai**. Siekiant padidinti aptikimo jautrumą, 5fC galima chemiškai žymėti su reagentais, formuojančiais fluorescuojančius junginius (Yuan ir Feng, 2014). 5hmC aptikimui pritaikytas 5-hidroksimetilo grupės glikozilimas, atliekamas T4 fago β-gliukoziltransferazės (βGT) esant uridino 5'-difosfo-d-gliukozės (UDP-gliukozės) kofaktoriui. Naudojant [<sup>3</sup>H]-žymėtą gliukozę, nustatymas gali būti kiekybinis, jautresnis ir tikslesnis nei naudojant antikūnus (Cadet ir Wagner, 2014; Song ir kt., 2012). Taip pat pritaikoma gliukozė su azido grupe, prie kurios aptikimui galima chemiškai prijungti biotiną (Song ir kt., 2011).

#### **1.4 DNR citozino modifikacijų genomio profiliavimo sekoskaitos metodai**

Siekiant suprasti DNR citozino modifikacijų funkcijas, reikia nustatyti jų išsidėstymą genome, geriausia nukleotido tikslumu (Wu ir Zhang, 2017). DNR modifikacijų genomio profiliavimo sekoskaitos metodų sukūrimas smarkiai išplėtė epigenomikos tyrimus, atskleidė ryšius tarp modifikacijų kitimo ir ląstelės fenotipo pokyčių (Wu ir Zhang, 2014). 5hmC, 5fC ir 5caC atradimas paskatino sukurti naujus genomio profiliavimo sekoskaitos metodus jų biologinių funkcijų ir susijusių molekulinų mechanizmų apibūdinimui (Zhao ir kt., 2020). Genominio modifikacijų išsidėstymo analizavimas naudojant naujas technologijas suteikė žinių apie aktyvaus DNR demetilavimo procesus ir TET baltymų funkcijas (Wu ir Zhang, 2017). Tačiau dar reikia tolimesnių tyrimų, siekiant nustatyti tikslias

oksiduotų 5mC formų funkcijas, o profiliavimą apsunkina jų retumas genome (Plongthongkum ir kt., 2014).

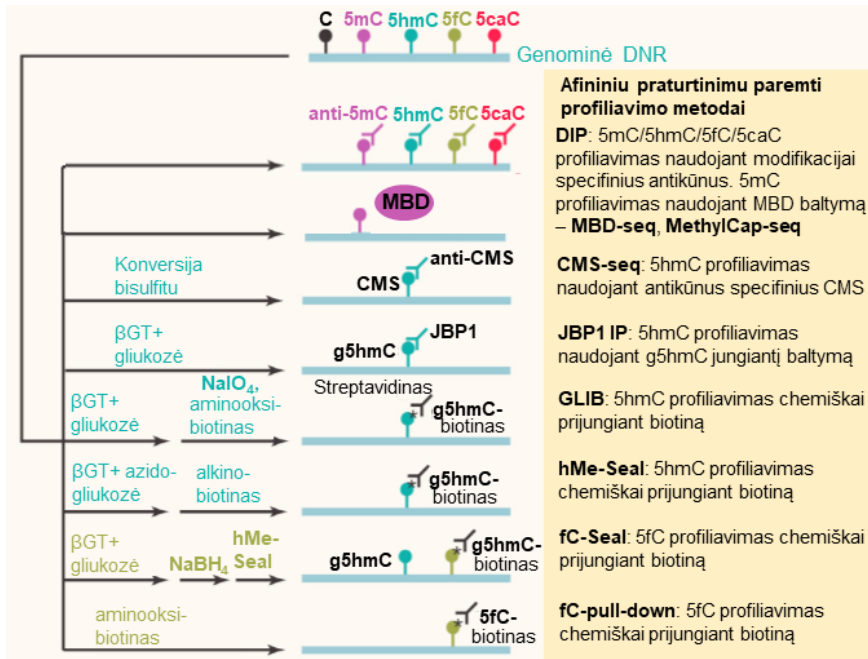
Modifikacijų genomino profiliavimo metodus galima skirstyti į afininio modifikuotų sekų praturtinimo (nedidelės skiriamosios gebos) ir nukleotido tikslumo metodus (Shen ir kt., 2014). Afininiam praturtinimui gali būti naudojami antikūnai, modifikacijas atpažįstantys baltymai, cheminis ir/ar fermentinis modifikavimas. Nukleotido tikslumas gali būti pasiekiamas naudojant restrikcijos endonukleazes, pakeičiant sekos nuskaitymą fermentine arba chemine konversija (bisulfitine ar kitokia) ar atliekant tiesioginį modifikacijų nuskaitymą trečios kartos sekoskaita. Taip pat sukurti nukleotido tikslumo metodai, kuriuose pasitelkiamas chemofermentinis žymėjimas ir prie žymės prijungto oligonukleotido pratęsimas.

Skirtingi metodai turi tam tikrų privalumų ir trūkumų (aprašyta žemiau). Vystant DNR modifikacijų genomino profiliavimo metodus, siekiama optimizuoti jų kainą, minimalų reikalingą DNR kiekį, tikslumą ir metodo našumą (Plongthongkum ir kt., 2014). Taip pat renkantis metodą yra svarbūs jautrumo, specifiskumo reikalavimai eksperimentui, paprastumas, prieinami reagentai ir analizės galimybės (Kurdyukov ir Bullock, 2016). Dėl skirtingų naudojamų nustatymo metodų ar duomenų analizės strategijų tarp modifikacijų profiliavimo duomenų gali atsirasti nesutapimai (Wu ir Zhang, 2014). Analizę galima nukreipti į dominančias genomo sritis (angl. *targeted*), tai sumažina kainą ir yra efektyviau esant daug mėginių. Nukreiptai analizei dažnai yra taikoma mikrogardelių hibridizacija, bet atliekama ir sekoskaita. Tuo tarpu viso genomo analizė yra svarbi tiriant įvairių ląstelių nežinomus globalius epigenomo pokyčius (Plongthongkum ir kt., 2014).

#### **1.4.1 Afininiu praturtinimu paremti metodai**

Afininio praturtinimo metoduose (1.12 pav.) atliekamas modifikacijas turinčių sekų praturtinimas, pasitelkiant fermentinį ar/ir cheminį poveikį arba modifikacijas atpažįstančius antikūnus (Wu ir Zhang, 2017). Šie metodai pasižymi žema sekoskaitos kaina, nes nuskaitomos tik praturtintos, modifikacijas turinčios sekos. Jie gana paprastai atliekami, tačiau nustato tik santykinę modifikacijų gausumą, priklauso nuo jų tankio, yra nedidelės skiriamosios gebos (keli šimtai bazių porų) ir nesuteikia grandinės informacijos (Wu ir Zhang, 2015). Praturtinimu paremti metodai gali būti jautresni nustatant tankiai išsidėsčiusias mažai modifikuotas vietas, kai tuo tarpu retai išsidėsčiusias modifikacijas geriau nustato nukleotido tikslumo metodai (Plongthongkum ir kt., 2014; Wu ir Zhang, 2017).





**1.12 pav.** Afininių praturtinimų pagrįsti citozino modifikacijų genomio profiliavimo metodai. DNR imunoišsodinimo (DIP) metuose 5mC, 5hmC, 5fC ar 5caC modifikacijas galima praturtinti pasitelkiant jas tiesiogiai atpažįstančius antikūnus. 5mC taip pat gali būti praturtinamas naudojant jį atpažįstantį baltymą MBD. Kituose metuose oksiduotos 5-metilcitozino formos prieš praturtinimo stadiją yra chemiškai ar fermentiniu būdu modifikuojamos (5hmC nustatymui skirtas JBP1 baltymą pasitelkiantis imunoišsodinimas (JBP1 IP)) bei gali būti prijungiamas biotinas ir praturtinimui naudojama sąveika su streptavidinu: glikozilino, perjudato oksidacijos ir biotininimo sekoskaita (GLIB); 5-metilenosulfonato sekoskaita (CMS-seq); selektyvus chemofermentinis žymėjimas (hMe-Seal 5hmC ir fC-Seal 5fC nustatymui); 5fC praturtinimas (fC-pull-down). Nustatomi modifikuoti DNR fragmentai, tačiau nenustatoma kur tiksliai juose yra modifikacija. Pagal Wu ir Zhang (2014).

### Antikūnų ar baltymų atpažinimu pagrįsti afininio praturtinimo metodai

DNR imunoišsodinimo (DIP, angl. *DNA immunoprecipitation*) būdas plačiai taikomas dėl paprastos procedūros – 5mC, 5hmC, 5fC ar 5caC turinčios sekos praturtinamos naudojant jas atpažįstančius antikūnus (1.12 pav.). Šių metodų praturtinimo efektyvumas yra didesnis esant didesniam modifikacijų tankiui, todėl tokios sekos aptinkamos geriau (Shen ir kt., 2014). Antikūnų specifiškumas priklauso nuo jų kokybės, galimas kryžminis reagavimas, todėl rezultatai atskirose laboratorijose gali skirtis (Pastor ir kt., 2013).

Atliekant papildomą 5hmC modifikavimą, gaunamas didesnis IP specifiškumas ir mažesnis foninis signalas. **CMS-seq** (angl. *5-methylenesulfonate sequencing*) metode 5hmC veikiamas natrio bisulfitu ir paverčiamas 5-metilenosulfonatu (CMS), kuris praturtinamas didelio specifiškumo antikūnais prieš CMS. Tačiau didelė cheminė grupė gali sukelti PGR pokyčius. JBP1 imunoišsodinimo metode (**JBP1 IP**) 5hmC glikozilinamas veikiant  $\beta$ GT ir praturtinamas pasitelkiant JBP1 baltymą, kuris natūraliai atpažįsta susidariusią modifikaciją. Tačiau šio metodo praturtinimas yra silpnesnis nei naudojant antikūnus prieš 5hmC (Wu ir Zhang, 2014).

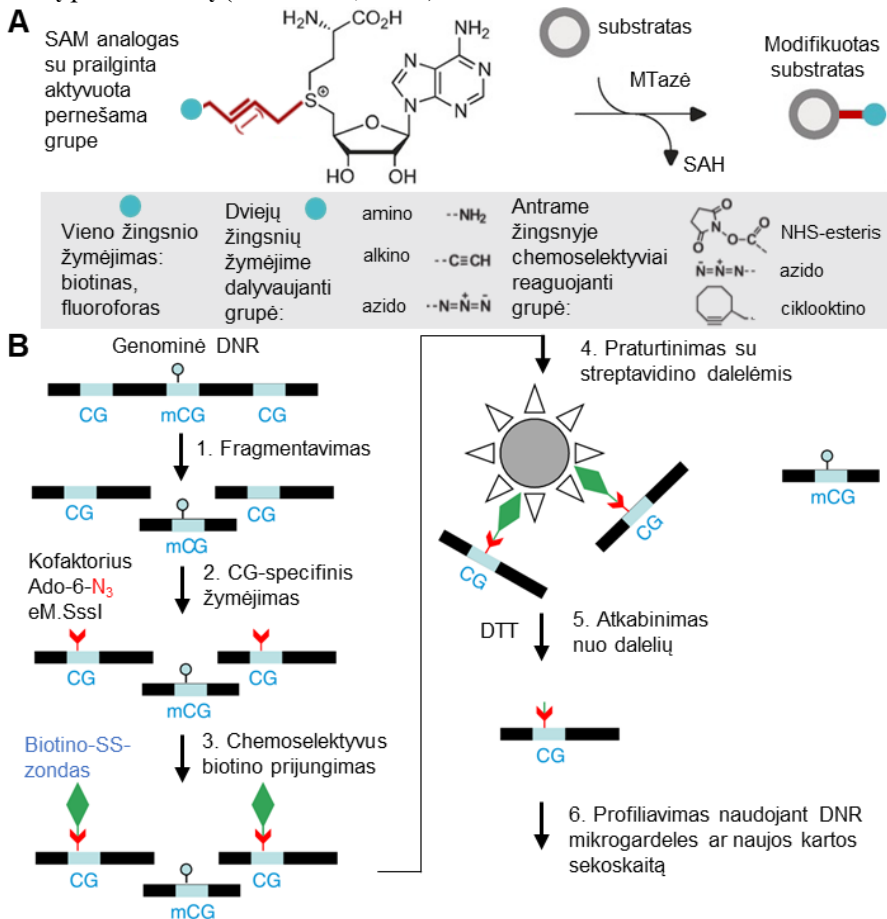
Taip pat gali būti pritaikomi modifikacijas atpažįstantys baltymai. 5mC nustatymo metodui **MBD-seq** (angl. *methyl-binding protein sequencing*) pritaikytas jį atpažįstantis baltymas MBD. MBD-seq protokolas buvo pakeistas (**MethylCap-seq**) atliekant išplovimą nuo žemos iki didelės NaCl koncentracijos. Taip pasiekiamas vienodesnis skirtingo CpG tankio regionų nuskaitymų skaičius (t.y. kiek kartų regionas buvo nuskaitytas, toliau tekste – padengimas) bei didesnis bendras CpG padengimas (Plongthongkum ir kt., 2014).

### **Kovalentiniu modifikavimu pagrįsti afininio praturtinimo metodai**

Kita didelė DNR modifikacijų afininio praturtinimo profiliavimo metodų grupė yra paremta fermentiniu ar/ir cheminiu žymėjimu prijungiant biotiną (1.12 pav.) bei praturtinimu pasitelkiant avidino/streptavidino daleles. Dėl stiprios ir selektyvios biotino sąveikos su avidinu/streptavidinu šie metodai yra jautresni ir turi mažesnę foninę signalą nei naudojant antikūnus, tačiau išlieka tam tikra praturtinimo priklausomybė nuo modifikacijų tankio.

$\beta$ GT yra efektyvi ir nepriklauso nuo DNR sekos, todėl patogi 5hmC žymėjimui (Shen ir kt., 2014; Wu ir Zhang, 2014). **hMe-Seal** (angl. *5hmC-selective chemical labeling*) metode (1.12 pav.)  $\beta$ GT prijungia azido grupę turinčią gliukozę, tuomet „click“ chemijos būdu prie azido grupės prijungiamas ciklooktino-biotinas ir atliekamas praturtinimas (Song ir kt., 2011, 2013). **GLIB** (angl. *glucosylation, periodate oxidation, biotinylation*) metode  $\beta$ GT prijungia paprastą gliukozę, kurios hidroksilo grupės natrio perjodatu ( $\text{NaIO}_4$ ) oksiduojamos į aldehido grupes, o šios reaguoja su aminooksi-biotinu (Shen ir kt., 2014). hMe-Seal pagrindu buvo sukurtas **fC-Seal** metodas 5fC praturtinimui. Natūralus 5hmC pirmiausia apsaugomas glikozilinant paprastą gliukozę, tuomet 5fC redukuojamas į 5hmC natrio borohidridu ( $\text{NaBH}_4$ ), o susidaręs 5hmC žymimas ir praturtinamas atliekant hMe-Seal žingsnius (Song ir kt., 2011, 2013). 5fC galima praturtinti po

tiesioginio žymėjimo biotinu (**fc-pull-down**) aminorooksialdehido kondensacijos būdu (Raiber ir kt., 2012), o 5caC – prie karboksigrupės prijungus aminą esant EDC (1-etil-3-(3-dimetilaminopropil)karbodiimidis) (Lu ir kt., 2013). Tačiau abu būdai turi pašalinių reakcijų su kitais modifikuotais citoziniais arba DNR karkasu, o tai lemia didelį nespecifinį foninį praturtinimą (Shen ir kt., 2014).



**1.13 pav.** Žymėjimas metiltransferazės atliekant aktyvuotų grupių pernašą (mTAG). (A) mTAG žymėjimas naudojant įvairius SAM kofaktoriaus analogus (pagal Deen ir kt., 2017; Lukinavičius ir kt., 2013; Tomkuvienė ir kt., 2016). Mėlynas rutuliukas žymi įvairias grupes, galinčias būti pernešamos grandinės gale – jų pavyzdžiai parašyti apačioje. Dviejų žingsnių žymėjime gali būti prijungta norima molekulė naudojant chemoselektyviai reaguojančią grupę. NHS – N-hidroksisukcinimidas. Gali būti naudojamos įvairios MTazės, modifikuojančios DNR, RNR, baltymų ar mažų molekulių substratus. (B) mTAG žymėjimu paremto nemetilintų genomo CpG taikinių profiliavimo metodo schema (pagal Kriukienė ir kt., 2013). Genominė DNR ardoma į trumpus fragmentus (1 žingsnis) ir veikiant mutantine metiltransferaze (eM.SssI) bei sintetiniu kofaktoriaus analogu (Ado-6-azidu arba Ado-6-aminu), prie nemodifikuotų CpG taikinių prijungiamos reaktyvios grupės (2 žingsnis). Prie

modifikuotų taikinių prijungiamas biotinas (3 žingsnis, biotino disulfido NHS-esteris arba dibenzilciklooktino-SS-PEG-biotinas atitinkamai amino ir azido grupei) ir pažymėti taikiniai praturtinami naudojant streptavidinu padengtas magnetines daleles (4 žingsnis). Fragmentai atkabunami nuo dalelių ditiotreitolium (DTT) skeliant disulfidinių ryšių biotino zonde (5 žingsnis). Prie praturtintų fragmentų prijungiami adapteriai, DNR padauginama PGR būdu ir analizuojama (6 žingsnis) naudojant mikrogardes (mTAG-chip) ar sekoskaitą (mTAG-seq).

Bakterijose aptinkamos DNR metiltransferazės vykdo specifinį daugybės įvairių sekų (2-8 nt) kovalentinį modifikavimą, kuris yra patrauklus tikslingo žymėjimo pritaikymui. Buvo sukurti sintetiniai SAM kofaktoriaus analogai (1.13 pav. A), kuriuose metilo grupė pakeista įvairiomis prailgintomis grupėmis, efektyviam modifikavimui „aktyvuotomis“ nesočiuoju ryšiu šalia reaguojančio anglies atomo (Dalhoff ir kt., 2006). mTAG (angl. *methyltransferase-directed Transfer of Activated Groups*) reakcijoje MTazės perneša aktyvuotas grupes nuo sintetinių SAM kofaktoriaus analogų specifiskai pažymėdamos substratą (1.13 pav. A) (Lukinavičius ir kt., 2007). Atliekant dviejų žingsnių žymėjimą, pernešamos grandinės gale yra reaktyvi pirminio amino, alkino ar azido grupė, prie kurios galima chemoselektyviai prijungti norimą žymę. Vieno žingsnio žymėjimui mTAG būdu iškart prijungiamos, pvz., biotino ar fluorescuojančios žymės (Tomkuvienė ir kt., 2016). Parodytas įvairių bakterijų MTazių, kurios priklauso visoms trimis klasėms (adenino-N6, citozino-N4 ir citozino-C5), vykdomas žymėjimas su kofaktorių analogais (Tomkuvienė ir kt., 2019). Laukinio tipo MTazės dažnai neefektyviai atlieka mTAG, todėl ši problema buvo išspręsta išplečiant C5 MTazių kofaktoriaus prijungimo kišenę (Lukinavičius ir kt., 2012; Kriukienė ir kt., 2013). mTAG žymėjimas gali būti pritaikomas DNR praturtinimui, imobilizavimui ant paviršių, transfekcijos stebėjimui, DNR nanostruktūrų kūrimui, fluorescenciniam DNR „barkodavimui“, *in vivo* žymėjimui. Naudojant įvairias kitas MTazes mTAG reakcija taip pat buvo pritaikyta RNR, baltymų ir mažų molekulių žymėjimui (Tomkuvienė ir kt., 2016).

Remiantis mTAG žymėjimu, buvo sukurtas genomo modifikavimo būsenos nustatymo sekoskaitos metodas **mTAG-seq**, kuriame atliekamas kovalentinis chemofermentinis nmodifikuotų CpG žymėjimas (Kriukienė ir kt., 2013). mTAG žymėjimui pasitelkta bakterijų CpG-spezifinė mutantinė metiltransferazė eM.SssI (angl. *engineered methyltransferase, Spiroplasma sp.* MQ1), kurioje yra išplėsta kofaktoriaus prisijungimo kišenė (Gln142 ir Asn370 pakeisti į Ala). Nuo sintetinio kofaktoriaus analogo eM.SssI perneša azido arba amino grupę turinčią molekulę. Tuomet chemoselektyviai prijungiamas biotinas, pažymėta DNR praturtinama naudojant streptavidino

daleles ir atliekama analizė mikrogardelėmis (mTAG-chip) ar naujos kartos sekoskaita (mTAG-seq) (1.13 pav. B). Analizė lengvai pritaikoma efektyviems didelės apimties epigenominiams tyrimams (Kriukienė ir kt., 2013).

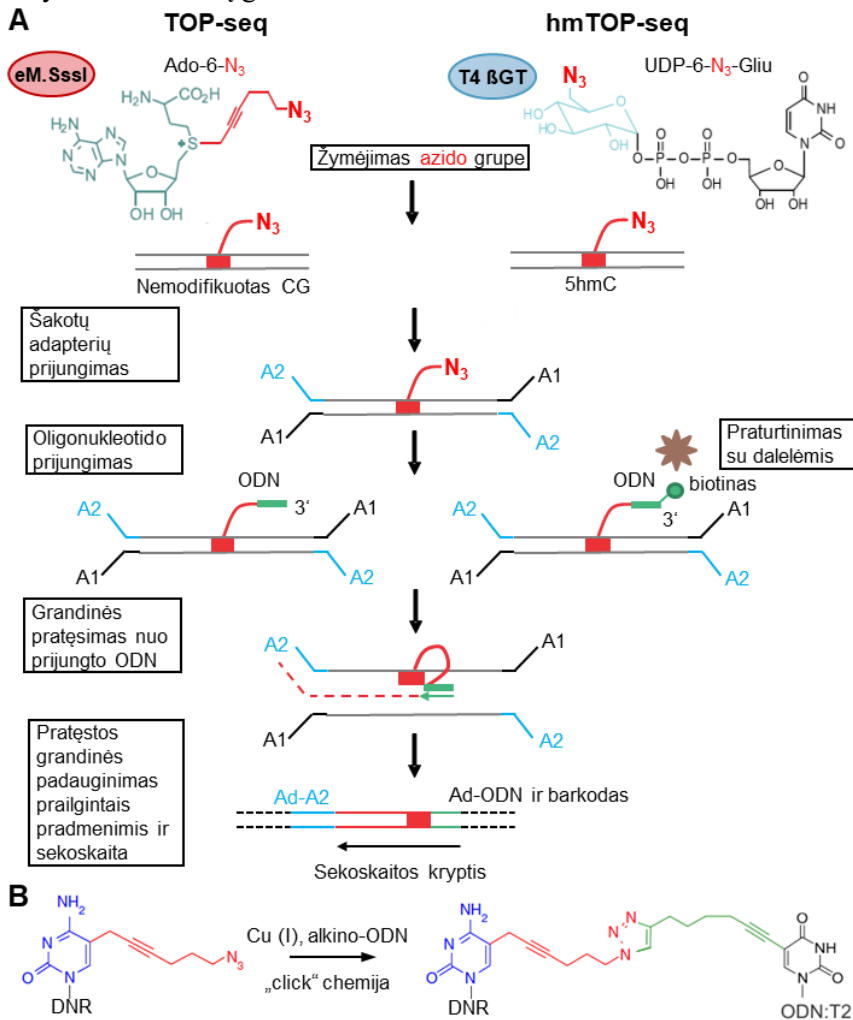
Eksperimentai su žmogaus DNR parodė, kad mTAG-seq metodas suteikia didesnę tikslumą ir apima daugiau nei dauguma afininio praturtinimo metodų. Metodas suteikia unikalų nemodifikuotų regionų nustatymą įvairiose geno vietose, kaip pasikartojančios sekos ir dalinai metilinti regionai, kurie dažnai nematomi, pvz., MRE-seq (1.15 pav. A). Dėl kovalentinio žymėjimo, griežto afininio praturtinimo ir švelnaus cheminio DNR išlaisvinimo nuo dalelių, mTAG-seq genominei analizei reikalingas pradinis DNR kiekis yra ties apatine 5mC praturtinimo metodų riba (tokių kaip MethylCap-seq, MBD-seq). Be to, lyginant su MBD-seq, mTAG-seq metodas parodė geresnę vidutinio ir žemo CpG tankio regionų nustatymą, kurie sudaro >50% visų CpG ir yra nepakankamai padengti kitų praturtinimo metodų (Kriukienė ir kt., 2013). O šių regionų metilinimas smarkiai keičiasi vykstant ląstelių diferenciacijai ir jie gali būti svarbūs ligų vystymuisi (Parry ir kt., 2021).

Nemodifikuoti citozinai sudaro mažesnę CpG dalį lyginant su metilintais (žmogaus genome, priklausomai nuo audinio, ~60-70% CpG būna metilinti; Lister ir kt., 2009). Todėl nemetilintos DNR dalies mTAG-seq analizė gali sumažinti sekoskaitos kainą, statistinių palyginimų skaičių ir gali būti jautresnė subtilių epigenetinių pokyčių nustatymui. Citozino modifikacijos gali būti sunkiai atskiriamos viena nuo kitos, tad papildydamas 5mC nustatymo metodus, mTAG-seq suteikia unikalų būdą sudėtingo epigenomo analizei.

#### **1.4.2 Nukleotido skiriamosios gebos metodai, paremti kovalentine fermentine modifikacija ir prijungto oligonukleotido nukreipta sekoskaita**

Nors sekoskaitos kaina mažėja, ji vis dar pernelyg aukšta didelės apimties epigenetinių ligų tyrimams viso geno bisulfitine sekoskaita (BS-Seq, angl. *bisulfite sequencing*, žr. 1.4.4). Pasiūlytas alternatyvus būdas DNR modifikacijų išsėdystymo analizei, kuris užpildo spragą tarp ekonomiško ir skiriamosios gebos – prijungto oligonukleotido nukreiptos sekoskaitos metodas **TOP-seq** (angl. *Tethered Oligonucleotide-Primed sequencing*; Staševskij ir kt., 2017). Pasitelktas prieš tai sukurtas mTAG-seq metodas (Kriukienė ir kt., 2013), kuriame nemodifikuotų ar hemimetilintų CG žymėjimui naudojama eM.SssI metiltransferazė perneša aktyvią azido grupę nuo sintetinio SAM kofaktoriaus analogo. Buvo parodyta, kad prie perneštos

azido grupės kovalentiškai prijungus 5' gale alkino grupę turintį oligodeoksirobonukleotidą (ODN) (1.14 pav. B), polimerazė nuo jam komplementaraus pradmens geba atlikti grandinės pratęsimą ir yra nukreipiama į taikinio CpG. Taip susiformuoja DNR grandinė, kurioje po prijungto ODN sekos yra taikinio CG ir gretimas genomo regionas (1.14 pav. A). Po padauginimo atlikus sekoskaitą, galima nukleotido tikslumu tiesiogiai nustatyti taikinio vietą genome.



**1.14 pav.** Prijungto oligonukleotido inicijuotos sekoskaitos metodai. (A) Nemodifikuotų CG (kairėje, TOP-seq; Staševskij ir kt., 2017) ir 5hmC (dešinėje, hmTOP-seq; Gibas ir kt., 2020) analizės metodų schema. Fragmentuota DNR TOP-seq atveju pažymima veikiant mutantinei metiltransferazei eM.SssI ir esant sintetiniam SAM kofaktoriaus analogui su azido grupe (Ado-6-N<sub>3</sub>). hmTOP-seq atveju DNR pažymima esant T4 β-gliukoziltransferazei (T4 βGT) ir UDP-6-azido-6-deoksi-D-gliukozei (UDP-6-N<sub>3</sub>-gliu). Tuomet prijungiami šakoti adapteriai (Ad,

A1/A2). Prie azido grupę žymėtų taikinių „click“ chemijos būdu prijungiamas oligodeoksiribonukleotidas (ODN). hmTOP-seq metodui naudojamas 3' gale biotiną turintis ODN ir papildomai atliekamas pažymėtos DNR praturtinimas pasitelkiant magnetines daleles. Vykdomas grandinės pratęsimas nuo prijungtam ODN komplementaraus pradmens. Prijungtą ODN turinti DNR specifiskai padauginama naudojant prailgintus (punktyrinė linija) pradmenis su sekoskaitos platformai reikalingomis 5' adapterių sekomis (Ad-A2 ir Ad-ODN). Sekoskaita inicijuojama nuo adapterio sekos, esančios Ad-ODN barkoduoto padauginimo pradmens 5' dalyje. **(B)** ODN su alkino grupę (antrame nukleotide, T2) prijungimas prie azido grupę modifikuoto citozino vario (I) jonų katalizuojamos azidų-alkinų ciklo prijungimo „click“ reakcijos metu. Prie azido grupę turinčia gliukoze modifikuoto 5hmC ODN prijungiamas analogiskai. Pagal Staševskij ir kt. (2017) ir Gibas ir kt. (2020).

Bandomieji TOP-seq eksperimentai parodė, kad metodas geba nustatyti epigenetinius bruožus, įprastai matomus tik BS-Seq. Priešingai nei BS-Seq, šiam metodui nereikia viso genomo sekoskaitos, todėl jis yra ekonomiška genomino DNR metilinimo profiliavimo alternatyva. Skirtingai nei bisulfitine sekoskaita paremti metodai, TOP-seq metodas negali nustatyti absoliutaus metilinimo lygio, tačiau galima nustatyti santykinę tam tikro regiono nemodifikuotų CG profilį.

TOP-seq apima įvairaus CG tankio regionus bei pasižymi didesniu tikslumu nustatant įvairias genomo vietas ir trumpus genominius elementus nei afininio praturtinimo metodai ar MRE-seq. MBD-seq ir MRE-seq reikia ~30-50 mln. genomui prilyginamų nuskaitymų, o su 30 mln. TOP-seq nuskaitymų pasiekiamas sekoskaitos išsotinimas ir geras sutapimas su BS-Seq, kuriam reikia bent eile didesnio nuskaitymų skaičiaus. Be to, TOP-seq nuskaitymai pasižymi lengvu duomenų apdorojimu ir informatyvumu, nes >90% sekų turi informaciją apie modifikavimą. Tuo tarpu viso genomo bisulfitinė sekoskaita sukuria didelius duomenų kiekius, tačiau didžioji dalis nuskaitymų (50-80%) duoda mažai arba jokios informacijos apie CG metilinimą (Ziller ir kt., 2013). TOP-seq metodas yra prieinamos kainos, CG skiriamosios gebos, apima visą genomą ir yra tinkamas didelės apimties epigenetiniams tyrimams.

Remiantis TOP-seq principu, buvo sukurtas vieno nukleotido skiriamosios gebos 5hmC genomino profiliavimo metodas – **hmTOP-seq** (angl. *5hmC-specific tethered oligonucleotide-primed sequencing*, 5hmC-specifinė prijungto oligonukleotido sekoskaita; Gibas ir kt., 2020). 5hmC pažymimas azido grupę turinčia gliukoze esant βGT ir UDP-6-azido-6-deoksi-D-gliukozei (tradiciškai vadinama UDP-6-azido-gliukoze, toliau tekste), tuomet prijungiamas biotinilintas oligodeoksiribonukleotidas, naudojant magnetines daleles praturtinama DNR su prijungtu ODN ir tolimesni žingsniai atliekami analogiskai kaip TOP-seq (1.14 pav. A).

Taikant šį metodą, 5hmC profiliavimui reikia mažiau DNR nei bisulfitine sekoskaita paremtiems metodams. Be to, pademonstruotas viso genomo 5hmC nustatymas pelės DNR esant tik ~20 mln. apdorotų nuskaitymų. Kitiems nuo natrio bisulfito nepriklausomiems vieno nukleotido skiriamosios gebos metodams (AMD ar ACE-Seq, žr. 1.4.5) reikia viso genomo sekoskaitos, todėl išauga kaina, ypač didelės apimties klinikinių mėginių analizėje. Dėl tikslingos, į 5hmC nukreiptos sekoskaitos hmTOP-seq metodas pasižymi dideliu ekonomišku ir lengvesniu duomenų apdorojimu. Nors hmTOP-seq negali nustatyti absoliutaus modifikacijos lygio, tačiau pagal padengimą galima nustatyti santykinę 5hmC lygį ir skaičiuoti regionų h-tankio profilį palyginimui tarp pavyzdžių (Gibas ir kt., 2020).

**Jump-seq** (Hu ir kt., 2019) – dar vienas panašus ekonomiškas metodas pusiau kiekybinei genominei 5hmC sekoskaitai. 5hmC nustatomas beveik nukleotido tikslumu, naudojant mažiau nei 50 ng DNR ir nedideles sekoskaitos apimtis (1/30 TAB-Seq, žr. 1.4.4). Pasitelkiant transpozazę, DNR yra sufragmentuojama ir tuo pačiu prijungiami biotiną turintys adapteriai, o tai sumažina reikalingą pradinės DNR kiekį ir galima atplauti nespecifinį oligonukleotidų prikibimą. Šiame metode taip pat pasitelkiamas 5hmC žymėjimas 6-azido-gliukoze, prijungiamas oligonukleotidas ir vykdomas grandinės pratęsimas. Po padauginimo 5hmC nustatomas pagal sekas, esančias iškart po prijungto oligonukleotido.

### 1.4.3 Restrikcijos endonukleazių savybėmis paremti metodai

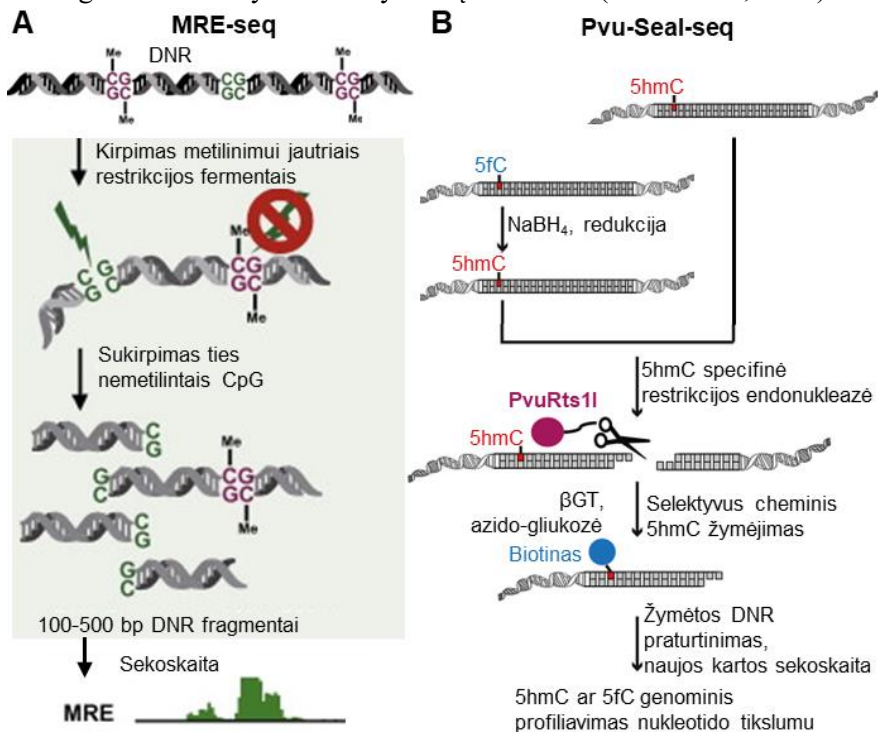
Citozino modifikacijų nustatymui genome pritaikytos restrikcijos endonuklezės (REazės), kurių gebėjimas perkirpti atpažįstamą DNR taikinį priklauso nuo DNR modifikavimo būsenos. Restrikcijos endonuklezėmis paremtais metodais galima analizuoti tiek nemodifikuotą, tiek modifikuotą genomines DNR dalį. Pagrindinis tokių metodų privalumas – nedideli naudojamos DNR kiekiai bei viso genomo sekoskaitos nereikalaujanti analizė. Tačiau jų analizuojama genomo dalis ir tikslumas yra priklausomi nuo fermentų atpažįstamos modifikacijos ir sekos. Kai kurie metodai gali netiksliai nustatyti modifikacijos poziciją arba nesuteikti kiekybinio modifikavimo įvertinimo (Wu ir Zhang, 2017). Dauguma REazių aiškiai neatskiria 5mC ir 5hmC, nors vėliau atrastos ir 5hmC-specifinės restrikcijos endonuklezės. 5fC ir 5caC modifikacijos dar labiau apsunkina tokią analizę (Kriukienė ir kt., 2012).

5mC nustatymui sukurtas **MRE-seq** (angl. *methylation-sensitive restriction-enzyme digestion and sequencing*) metodas (1.15 pav. A), praturtinantis nemetilintą DNR (Li ir kt., 2015). Lygiagrečiai naudojamos



metilinimui jautrios REazės, dažniausiai R.HpaII (C<sup>^</sup>CGG), R.HinPII (G<sup>^</sup>CGC) ir R.AciI (C<sup>^</sup>CGC). Po kirpimo išgryninami 50-300 bp fragmentai. DNR metilinimo lygis REazių taikiniuose yra nustatomas pagal atvirkštinį ryšį tarp nuskaitymų ir metilinimo. Įvertinamas CpG skiriamosios gebos santykinis metilinimo lygis, kurį galima lyginti tarp mėginių, tačiau nustatomos vietos yra apribotos REazių atpažinamų sekų. Aptinkamos ir mažai metilintos sekos, kurios sunkiau nustatomos kitais metodais (Li ir kt., 2015).

5mC nustatymui dažnai taikomas **Methyl-Seq** metodas (Brunner ir kt., 2009). Genominė DNR kerpama atskirai R.HpaII ir R.MspI REaze CCGG sekoje ir prie DNR sukirptų galų prijungiami adapteriai. Atlikus sekoskaitą, atliekamas abiejų REazių bibliotekų palyginimas. Kadangi R.HpaII kerpa tik nemodifikuotus taikinius, jos bibliotekos sekų galuose aptinkami CG nustatomi kaip nemetilinti, nes tik įvykus kirpimui prie jų gali būti prijungiami adapteriai. R.MspI kerpa esant ir metilintoms ir nemetilintoms sekoms, todėl tik R.MspI bibliotekos sekų galuose šalia adapterių aptinkami CG (ir neaptinkami R.HpaII bibliotekoje) yra nustatomi kaip metilinti. Naudojant skirtingai fluorescuojančius pradmenis, metodą galima taikyti mikrogardelėms (**HELP**, angl. *HpaII fragment enrichment by ligation-mediated PCR*) ir tuomet gaunamas santykinis kiekybinis įvertinimas (Booth ir kt., 2015).



**1.15 pav.** Citozino modifikacijų genomino profiliavimo metodai pasitelkiant restrikcijos endonukleazes. **(A)** MRE-seq (angl. *methylation-sensitive restriction-enzyme digestion and sequencing*) metodas. Genominė DNR kerpama naudojant kelias 5-metilcitozinui jautrias REazes, išgryninami susidarę tam tikro ilgio fragmentai ir atliekama jų sekoskaita. Nuskaitytos sekos parodo nemetilintų CpG, kurie yra naudotų fermentų taikiniuose, vietas genome. Pagal Li ir kt. (2015) **(B)** Pvu-Seal-seq metode pritaikyta PvuRtsII restrikcijos endonukleazė, kerpanti 11-12 bp už 5hmC. Kad būtų padidintas metodo specifiškumas, 5hmC pažymimas pasitelkiant  $\beta$ -gliukoziltransferazę ir azido grupę turinčią gliukozę. Po selektyvaus fragmentų praturtinimo prijungus biotiną, atliekama jų sekoskaita. Metodas taip pat gali būti pritaikomas 5fC nustatymui, su  $\text{NaBH}_4$  atlikus 5fC redukciją į 5hmC. Pagal Sun ir kt. (2015).

Taip pat 5mC nustatymui pritaikyta REazė McrBC, atpažįstanti 5mC ir optimaliai kerpanti tarp dviejų metilintų citozinų esant 55-100 bp. **Methyl-MAPS** (angl. *methylation mapping analysis by paired-end sequencing*) metode, naudojant McrBC bei penkias metiliniui jautrias REazes, analizuojama metilinta ir nemetilinta genomo dalis. Atlikus sekoskaitą ir lyginant mėginius metilinimo statusas nustatomas pagal tai, kurie CpG buvo atsparūs ar jautrūs kirpimui (Edwards ir kt., 2010).

Pasitelkiant  $\beta$ GT gebėjimą glikozilinti 5hmC, yra sukurti keli nustatymo metodai. **Aba-seq** metodui naudojama restrikcijos endonukleazė AbaSI specifiškai kerpa DNR, turinčią glikozilintą 5hmC. Kirpimas vyksta siaurame intervale už 5hmC, bet nustatoma modifikacijos pozicija yra apytikslė. AbaSI geriau kerpa simetriškus taikinius, taip pat yra nedidelis aktyvumas kitų citozino formų atžvilgiu (Sun ir kt., 2013a).

**Pvu-Seal-seq** metode (Sun ir kt., 2015) DNR kerpama PvuRts11, kuri atpažįsta 5hmC ir kerpa 11-12 bp už jo (1.15 pav. B). Kadangi REazė taip pat šiek tiek kerpa esant 5mC ar C, siekiant padidinti metodo specifiškumą 5hmC žymimas biotinu kaip hMe-Seal metode ir tuomet praturtinami 5hmC turintys fragmentai. Atlikus sekoskaitą, 5hmC turinčios vietos nustatomos ekonomiškai ir nukleotido tikslumu. Nors metodas negali įvertinti absoliutaus modifikavimo lygio, jo privalumas (kaip ir Aba-seq atveju) yra jautrus silpnai modifikuotų vietų aptikimas be didelių sekoskaitos apimčių, reikalingų tokių vietų nustatymui, pvz., bisulfitinės sekoskaitos atveju. Atliekant 5fC redukciją (į 5hmC naudojant  $\text{NaBH}_4$ ), Pvu-Seal-seq metodu galima nustatyti ir šią modifikaciją (Wu ir Zhang, 2015).

Sukurtas sumažintos 5hmC reprezentacijos metodas **RRHP** (angl. *reduced representation 5hmC profiling*) geba nustatyti 5hmC nukleotido tikslumu. DNR sukerpama R.MspI ir prijungiami adapteriai. 5hmC glikozilinamas ir taip apsaugomas nuo atliekamo antro kirpimo R.MspI. Tuomet, atliekant PGR, padauginami tik nepažeistą adapterį ir 5hmC turėję

fragmentai, o po sekoskaitos galima nustatyti 5hmC poziciją (Plongthongkum ir kt., 2014). Restrikcijos endonukleazių analizei alternatyviame **SCL-exo** (angl. *selective chemical labelling*) metode 5hmC pažymimas biotinu (analogiškai kaip hMe-Seal). Tuomet DNR praturtinama ir kerpama egzonukleaze, kuri sustoja ties biotinilintu 5hmC, todėl modifikacija nustatoma nukleotido tikslumu (Wu ir Zhang, 2017).

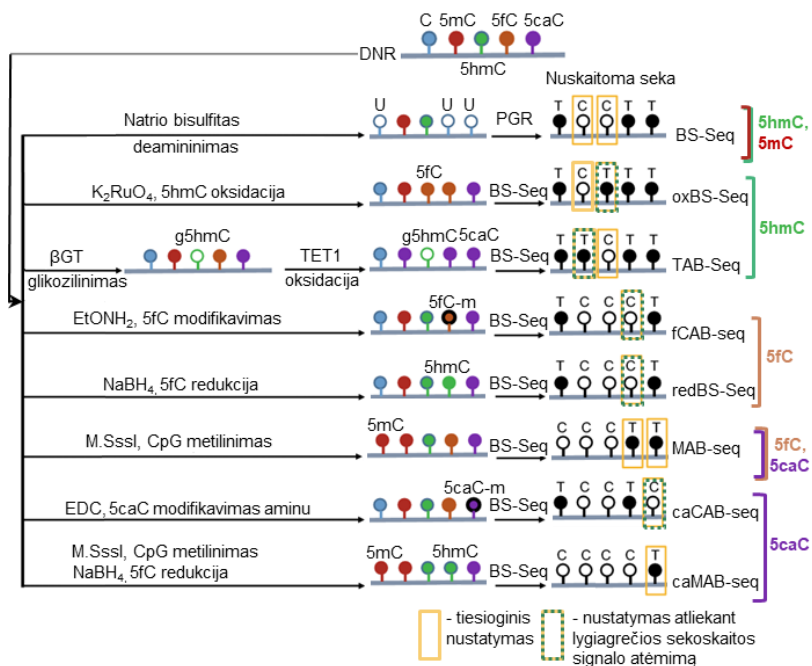
Taip pat, pritaikant modifikacijoms jautrias REazes ir specifinius pradmenis, tikro laiko PGR galima kiekybiškai įvertinti modifikavimo lygį specifiniame geno regione, kuriame yra vienas kirpimo taikynys. 5mC įvertinimui pasitelkiamos įvairios metilinizacijos jautrios restrikcijos endonuklezės, kartu su tą pačią seką atpažįstančia modifikavimui nejautria REaze. Dažniausiai naudojamos modifikavimui jautri R.HpaII ir nejautri R.MspI. Absolutus modifikavimo lygis įvertinamas pagal tikro laiko PGR signalo skirtumą tarp abiejų REazių. 5hmC nustatymui jis apsaugomas nuo R.MspI kirpimo glikozilinant βGT ir tuomet 5hmC lygiai įvertinami lyginant nekirptos, kirptos ir glikozilintos kirptos DNR kiekius (Booth ir kt., 2015).

#### **1.4.4 Bisulfitinė sekoskaita ir jos variantai**

Dauguma metodų, nustatančių modifikacijas nukleotido tikslumu (1.16 pav.), yra paremti bisulfitine sekoskaita. DNR konversija natrio bisulfitu sujungta su viso geno naujos kartos sekoskaita yra laikoma „auksiniu“ standartu 5mC profiliavimui ir yra plačiai taikoma didelės skiriamosios gebos metilinizacijos pokyčių tyrimui (Peng ir kt., 2016). BS-Seq paremta skirtingu citozino ir modifikuotų jo formų reagavimu su natrio bisulfitu (NaHSO<sub>3</sub>). Esant natrio bisulfitui, nemodifikuotas citozinas formuoja 5,6-dihidro-6-sulfonil junginį, įvyksta jo deaminavimas į uracilą, kuris po PGR padauginimo virsta T. Tuo tarpu 5mC yra labai atsparus natrio bisulfito sukeliama deaminavimui ir yra nuskaitomas kaip C. BS-Seq yra tiksli (nes konversija siekia 99%) ir suteikia kiekybinę modifikavimo informaciją visame genome – skaičiuojant C ir T nuskaitymų skaičių taikinyje, galima nustatyti metilinizacijos procentus. Absolutus modifikavimo lygio įvertinimas yra svarbus siekiant gauti tikslią biologinę informaciją. Kadangi po poveikio natrio bisulfitu susidaro nekomplementarios grandinės, taip pat galima nustatyti kiekvienos grandinės modifikavimą, o to negalima padaryti praturtinimo metodais (Zhao ir kt., 2020).

Atradus 5hmC pasirodė, kad BS-Seq negali jo atskirti nuo 5mC. Natrio bisulfitas atakuoja 5-hidroksimetil grupę susidarant stabiliam citozino 5-metilenosulfoniatui. Tuo tarpu 5fC ir 5caC elgiasi panašiai kaip nemodifikuotas citozinas ir sekoskaitos metu yra nuskaitomi kaip T. Todėl

standartinė BS-Seq atspindi bendrą 5mC ir 5hmC signalą, o tai apsunkina rezultatų interpretavimą ir reikia iš naujo įvertinti iki 5hmC atradimo gautus duomenis (Kriukienė ir kt., 2012). Prieš poveikį natrio bisulfitu atlikus fermentinį ar cheminį modifikavimą, standartinė BS-Seq gali būti pritaikyta oksiduotų 5mC formų atskyrimui viena nuo kitos (1.16 pav.) (tačiau nėra metodo, atskiriančio visas citozino modifikacijas vienu metu). Oksi-mC nustatomos tiesiogiai arba atliekamas palyginimas su įprastos BS-Seq duomenimis (Pastor ir kt., 2013).



**1.16 pav.** Nukleotido skiriamosios gebos DNR citozino modifikacijų profiliavimo metodai, paremti bisulfitinės konversijos sekoskaita (BS-Seq). Modifikacijos nustatomos tiesiogiai iš nuskaitomos sekos arba lygiagrečiai atliekant kitą sekoskaitą (nustatančią kitas modifikacijas) ir atimant jos signalą. 5mC profiliavimui kartu su 5hmC gali būti atliekama bisulfitinė sekoskaita. 5hmC atskirai nuo 5mC nustato oksidacinės bisulfitinės sekoskaitos metodas (oxBS-Seq) bei bisulfitinė sekoskaita naudojant TET (TAB-Seq). 5fC profiliavimui sukurta bisulfitinė sekoskaita su cheminiu žymėjimu (fCAB-seq) ar redukcija (redBS-Seq). Bisulfitinė sekoskaita veikiant M.SssI metiltransferaze (MAB-seq) nustato 5caC ir 5fC, o atliekant 5fC redukciją nustatomas tik 5caC (caMAB-seq). Bisulfitinė sekoskaita su cheminiu modifikavimu (CAB-seq) nustato 5caC. g5hmC – 5hmC su prijungta gliukoze, 5caC-m – aminu modifikuotas 5caC, 5fC-m – modifikuotas 5fC. Pagal Zhao ir kt. (2020).

5hmC nustatymui sukurta **oxBS-Seq** (angl. *oxidative bisulfite sequencing*; Booth ir kt., 2012) metodas. 5hmC kalio perrutenatu (KRuO<sub>4</sub>) yra oksiduojamas į 5fC, kuris virsta U ir nuskaitomas kaip T, kai tuo tarpu 5mC

nuskaitomas kaip C. Absoliutus 5hmC lygis ir pozicija nustatoma oxBS-Seq signalą atimant iš BS-Seq (5mC+5hmC signalo), todėl patikimam 5hmC nustatymui reikia giles (gylis – sekoskaitos sukuriamas sekų ar bazių skaičius, lemiantis padengimą) ir dvigubos sekoskaitos (Zhao ir kt., 2020). Metodo trūkumas yra oksidacijos metu sukeliama DNR fragmentacija, todėl reikia gana didelio pradinio DNR kiekio. Oksidacijos žingsnyje sumažinus fragmentaciją, metodą galima taikyti ~100 ng DNR (Peng ir kt., 2016). Kitame 5hmC nustatymo metode, **TAB-Seq** (angl. *TET-assisted bisulfite sequencing*; Yu ir kt., 2012), 5hmC pirmiausia glikozilinamas pasitelkiant βGT, o 5mC oksiduojamas iki 5caC naudojant pelės TET1. Paveikus natrio bisulfitu ir atlikus sekoskaitą, 5hmC tiesiogiai nuskaitomas kaip C (βGT modifikavimas >90%), o tuo tarpu visos kitos citozino formos nuskaitomos kaip T (Yu ir kt., 2012). Atimant TAB-Seq signalą iš BS-Seq galima nustatyti 5mC. Dėl 5hmC retumo, esant vidutiniam sekoskaitos gyliui (17-26× padengimas citozinui) TAB-Seq gali nustatyti didesnę nei 20% 5hmC modifikavimą, o mažesnio lygio aptikimui TAB-Seq ar oxBS-Seq metodu sekoskaitoje gali būti reikalingas >100× genomo padengimas (Wu ir Zhang, 2015).

5fC nustatymui buvo sukurtas **fCAB-seq** (angl. *5fC chemical modification–assisted bisulfite sequencing*; Song ir kt., 2013) metodas. Cheminis 5fC modifikavimas O-etilhidroksilaminu (EtONH<sub>2</sub>) apsaugo nuo deaminavimo veikiant natrio bisulfitu ir 5fC nustatomas lyginant mėginį su standartinės BS-Seq rezultatais. 5fC galima nustatyti netgi esant tik kelių procentų modifikacijos lygiui. Kitame metode, **redBS-Seq** (angl. *reduced BS-Seq*; Booth ir kt., 2014), 5fC nustatomas jį selektyviai redukuojant į 5hmC natrio borohidridu (NaBH<sub>4</sub>) ir paveikus natrio bisulfitu rezultatai lyginami su standartinė BS-Seq. fCAB-seq metode 5fC modifikuojamas 50-60%, tuo tarpu redBS-Seq beveik 97% efektyvumu (Zhao ir kt., 2020).

5caC galima nustatyti **CAB-seq** (angl. *Chemical modification-assisted bisulfite sequencing*; Lu ir kt., 2013, 2015b) metodu. 5caC nuo deaminavimo natrio bisulfitu apsaugomas EDC katalizuojamoje reakcijoje prie karboksilo grupės prijungiant aminą ir tuomet atliekamas palyginimas su paprasta BS-Seq.

Kadangi 5fC ir 5caC yra ypač retos modifikacijos, naudingiau jas nustatyti be papildomo, sekoskaitos kaštus didinančio, lyginimo su lygiagrečiu paprastos BS-Seq eksperimentu. Tam sukurtas **MAB-seq** (angl. *methylase-assisted bisulfite sequencing*; Wu ir kt., 2014; Neri ir kt., 2015) metodas, nustatantis 5fC ir 5caC kartu, kaip vieną bendrą signalą. Nemodifikuoti citozinai yra metilinami naudojant CpG metiltransferazę M.SssI ir taip apsaugomi, o po poveikio natrio bisulfitu, CpG taikiniuose tik

5fC ir 5caC yra nuskaitomi kaip T (Peng ir kt., 2016). MAB-seq metode 99,5% 5caC ir 84,7% 5fC yra efektyviai konvertuojami (Zhao ir kt., 2020). Siekiant kaip T nuskaityti tik 5caC ir jį nustatyti atskirai, metodas pakeistas (**caMAB-seq**; Wu ir kt., 2016) pridėjus NaBH<sub>4</sub> vykdomą 5fC redukciją į 5hmC. Kadangi citozino modifikacijos dažniausiai aptinkamos CpG kontekste, MAB-seq metodas tinkamas daugumai audinių (Peng ir kt., 2016).

Nors BS-Seq paremtais metodais galima nustatyti modifikavimo lygį nukleotido tikslumu, jie turi tam tikrų trūkumų. Dauguma citozinių (~95%) virsta T, todėl sekos kompleksiskumas sumažėja beveik tik iki trijų bazių ir tai lemia prastą sekoskaitos kokybę, sumažėja nuskaitymų skaičius, kuriuos galima prilyginti prie geno. Sekoskaitos analizė yra gana sudėtinga, nes reikalingas nuskaitymų atrinkimas ir statistiniai testai bei skirtingos vietos padengiamos nevienodai (Zhao ir kt., 2020). Taip pat gali susidaryti chimeriniai PGR produktai, atsirasti nuo sekos priklausomi tendencingumai. Todėl metodo patikimumas yra geresnis atliekant mažiau PGR ciklų ir priklauso nuo naudojamos polimerazės tikslumo bei DNR kiekio (Kumar ir kt., 2018). Nomodifikuoti citozinai gali būti ne iki galo paverčiami timiniais (1,3% ir mažiau), taip pat gali įvykti netinkamas 5mC konvertavimas į timiną (gali siekti 2,7%; Holmes ir kt., 2014). Todėl BS-Seq paremtuose metoduose į mėginius turi būti įmaišoma kontrolinė DNR, kad būtų galima įvertinti konversijos efektyvumą ar neteisingą konversiją, kuri gali skirtis dėl techninės ar biologinės variacijos.

Bene didžiausias BS-Seq trūkumas yra kaina, kuri išlieka aukšta daugumai laboratorijų, nes nuskaitomas visas genomai ir patikimam modifikavimo įvertinimui reikia didelio sekų nuskaitymų skaičiaus. Atlikus nepakankamai gilią sekoskaitą, kai kurios geno vietos gali turėti mažą padengimą, įvykti atsitiktinės atrankos klaidos. Pavyzdžiui, kad padengiant 99,5% žmogaus geno būtų nustatytas mažiausiai 10% metilinimas, reikia atlikti 60 Gb sekoskaitą esant vidutiniškai 20× padengimui (Plongthongkum ir kt., 2014). Siekiant sumažinti sekoskaitos apimtį ir kainą kartu su BS-Seq taikomos įvairios praturtinimo strategijos, leidžiančios atlikti tik dalies geno sekoskaitą. CpG turtingos sekos prieš ar po BS-Seq gali būti praturtinamos naudojant hibridizaciją su imobilizuotais oligonukleotidais (nukreipta BS-Seq, angl. *targeted*) ir taip analizuojama dominanti maža geno dalis (Kurdyukov ir Bullock, 2016). Taip pat BS-Seq ir jos variantus oksiduotoms 5mC formoms galima atlikti konkrečiam lokusui su jam specifiniais pradmenimis po konversijos atliekant PGR (Plongthongkum ir kt., 2014).

Yra žinoma, kad DNR metilinimas dažniausiai vyksta CpG kontekste, tačiau jų dažnis stuburinių genomuose yra mažesnis nei kitų dinukleotidų. Pavyzdžiui, ~53% nepersidengiančių žmogaus genomo regionų po 100 bp neturi CpG. Todėl nukreipiant sekoskaitą į CpG turtingus regionus, kaip CpG salos ir promotoriai, galima sumažinti reikalingų nuskaitymų skaičių bent 10 kartų (Plongthongkum ir kt., 2014). Tam pasitelkiamas sumažintos reprezentacijos metodas **RRBS** (angl. *reduced representation bisulfite sequencing*), kuriame CpG turtingų sekų atrinkimui pritaikoma metilinimui nejautri REazė (pvz., R.MspI (C<sup>^</sup>CGG) ar R.ApeKI (G<sup>^</sup>CWGC)), po kirpimo pašalinama nesukirpta ar ilga DNR, prijungiami adapteriai ir atliekama BS-Seq. Pasitelkiant R.MspI žmogaus genome galima išanalizuoti 10% CpG ir matoma ~85% CpG salų, o kerpant R.ApeKI ir R.MspI analizuojama 20% CpG (Kurdyukov ir Bullock, 2016). Tačiau RRBS neapima mažo CpG tankio regionų, galinčių persidengti su introniniais regionais ar stiprikliais, kuriuose pavyzdžiui, stebimi metilinimo pokyčiai žmogaus indukuotose pluriipotentinėse ląstelėse (Plongthongkum ir kt., 2014). RRBS taip pat gali būti atliekama kartu su oxBS-Seq 5hmC nustatymui ar su redBS-Seq 5fC nustatymui. Svarbu paminėti, kad R.MspI, naudojama standartiniame RRBS, tik iš dalies kerpa 5fC modifikuotas sekas ir visiškai nekerpa esant 5caC. Todėl analizuojant šias modifikacijas, 5fC turi būti redukuojamas į 5hmC arba pritaikoma modifikacijoms nejautri restrikcijos endonukleazė, kaip TaqαI, (T<sup>^</sup>CGA) (Wu ir Zhang, 2015).

Dėl 5fC ir 5caC retumo genome patikimam šių modifikacijų nustatymui reikia ypač didelio sekoskaitos gylio. Pavyzdžiui, fCAB-seq reikia ~1000× ar didesnio konkretaus lokuso padengimo. Taigi fCAB-seq ar caCAB-seq pritaikymas viso genomo sekoskaitai yra nepraktiškas (Shen ir kt., 2014). Todėl šis metodas taikomas kartu su DNR imunoišsodinimu (**DIP-CAB-seq**, angl. *DNA immunoprecipitation-coupled CAB-seq*; Lu ir kt., 2015b) – 5fC ar 5caC prieš CAB-seq yra praturtinami naudojant antikūnus ir taip nuskaitytos tik modifikacijas turinčios genomo vietos. Tuomet nustatomas santykinis modifikacijų praturtinimas, o ne absoliutus lygis. Taip pat galima jungti BS-Seq metodų variantus su chromatino imunoišsodinimu (ChIP, angl. *chromatin immunoprecipitation*), nustatant dalį modifikacijomis praturtintų vietų (Shen ir kt., 2014). Pavyzdžiui, galima atlikti praturtinimą pasitelkiant antikūnus prieš H3 histono 4 lizino monometilinimą (H3K4me1), tuomet atlikti fCAB-seq ir taip analizuoti modifikaciją aktyviuose/pasiruošusiuose stiprikluose.

Veikiant natrio bisulfitu prarandamos pirimidinų bazės, tai palengvina grandinės trūkius tam tikrose sekose ir didelė DNR dalis fragmentuojasi. Todėl BS-Seq paremtiems metodams dažnai reikia nemažai

DNR (Raiber ir kt., 2017). Tačiau naudojant patobulintus bibliotekų ruošimo metodus, BS-Seq galima atlikti su ~30 ng DNR (Kurdyukov ir Bullock, 2016).

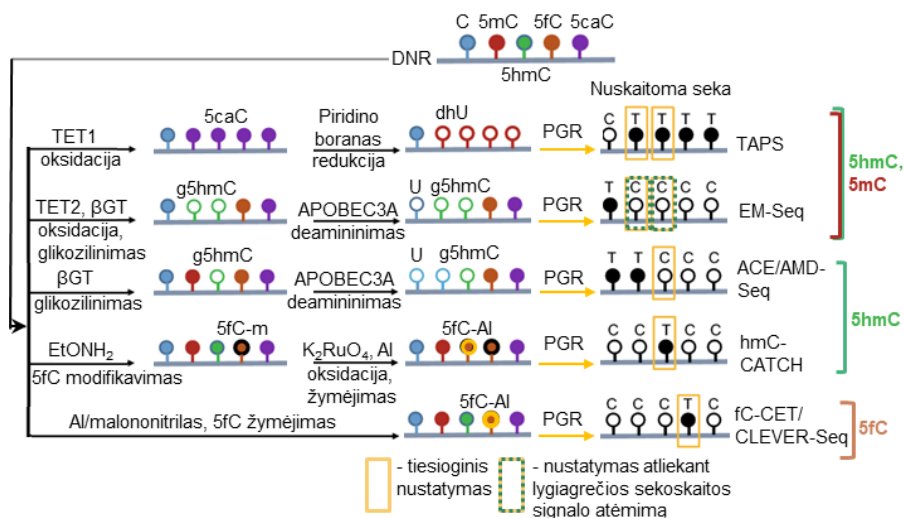
#### 1.4.5 Cheminės ar fermentinės konversijos metodai nenaudojant natrio bisulfito

Siekiant išspręsti bisulfitinės sekoskaitos problemas, buvo ieškoma švelnesnių ir/ar nemodifikuoto citozino neveikiančių konversijos būdų. Tam buvo sukurti metodai, kuriuose citozino bazės paveikiamos cheminiu ar/ir fermentiniu būdu nenaudojant natrio bisulfito, po PGR bazė toje vietoje pakinta ir po viso genomo sekoskaitos modifikacijos nustatomos nukleotido tikslumu (1.17 pav.). Šiais metodais galima įvertinti ir modifikavimo lygį. Dauguma metodų buvo sukurti dar palyginus neseniai, todėl dar nėra plačiai naudojami.

5mC ir 5hmC nustatymui sukurta piridino borano sekoskaita (1.17 pav.) pasitelkiant TET (**TAPS**, angl. *TET-assisted pyridine borane sequencing*; Liu ir kt., 2019). 5mC ir 5hmC yra oksiduojami TET iki 5caC, o šis piridino boranu efektyviai (98%) redukuojamas į dihidrouracilą (dhU), kuris po PGR nuskaitymas kaip T.  $\beta$ GT glikozilinus 5hmC arba TET pakeitus kalio perrutenatu atitinkamai galima atskirai nustatyti 5mC (**TAPS $\beta$** ) ir 5hmC (**CAPS**, angl. *chemical-assisted pyridine borane sequencing*). Taip pat, neatliekant TET oksidacijos ir chemiškai modifikuojant 5caC arba 5fc, galima atskirai nustatyti ir šias modifikacijas. Lyginant su BS-Seq, piridino borano sekoskaita nesukelia DNR fragmentavimo, turi mažesnę neteisingai nustatomo metilavimo dažnį (vidutiniškai 0,23% lyginant su 0,6%). Taip pat stebima geresnė sekų kokybė, kompleksškumas ir prilyginimas prie tiriamo genomo, todėl reikia mažesnio nuskaitymų skaičiaus bei DNR kiekio (Liu ir kt., 2019).

Kitas natrio bisulfito nenaudojantis 5mC ir 5hmC nustatymo metodas – **EM-Seq** (angl. *Enzymatic Methyl-Seq*). Jame APOBEC3A deaminina nemodifikuotus citozinius, paversdama juos į U. Prieš tai 5mC apsaugomas nuo deaminavimo atliekant oksidaciją su TET2 ir susidariusio 5hmC glikozilinimą veikiant  $\beta$ GT. Neatliekant TET oksidacijos, galima nustatyti 5hmC **ACE-Seq** (angl. *APOBEC-coupled epigenetic sequencing*) ar **AMD-Seq** (angl. *APOBEC3A-mediated deamination sequencing*) metodu. Nors APOBEC naudojančių metodų (1.17 pav.) sekų prilyginimas geresnis nei BS-Seq, bet kompleksškumas vis tiek yra mažas ir mažinant DNR kiekį didėja nepanaudojamų nuskaitymų dalis (Zhao ir kt., 2020).





**1.17 pav.** Nukleotido skiriamosios gebos DNR citozino modifikacijų profiliavimo metodai, kuriuose konversijai nenaudojamas natrio bisulfitas. Atliekamas įvairus cheminis ar fermentinis modifikavimas pakeičia bazių nuskaitymą po PGR. Modifikacijos nustatomos tiesiogiai iš nuskaitytos sekos arba lygiagrečiai atliekant kitą sekoskaitą (nustatančią kitas modifikacijas) ir atimant jos signalą. 5mC profiliavimui kartu su 5hmC gali būti atliekama piridino borano sekoskaita naudojant TET (TAPS) arba fermentinė 5mC sekoskaita (EM-Seq). 5hmC atskirai nuo 5mC nustato APOBEC pasitelkianti sekoskaita (ACE-Seq) bei cheminė 5hmC konversija į T (hmC-CATCH). 5fC profiliavimui sukurta ciklizacijos (fC-CET) ar žymėjimo malononitrilu (CLEVER-Seq) nulemta 5fC į T konversija. g5hmC – 5hmC su prijungta gliukoze, dhU – dihidrouracilas, 5fC-m – modifikuotas 5fC, AI – 1,3-indandionas. Pagal Zhao ir kt. (2020).

5fC nustatymui (1.17 pav.) sukurtas **fC-CET** (angl. *5fC cyclization-enabled C-to-T transition*) metodas (Xia ir kt., 2015). 5fC selektyviai pažymimas ir praturtinamas pritaikant azido grupę turintį 1,3-indandioną, o po PGR tik pažymėtas 5fC pavirsta į T. Konversija nesukelia ryškios DNR fragmentacijos ir išlaikomas didelis sekų kompleksiškumas (Peng ir kt., 2016). Remiantis šiuo principu, 5hmC nustatymui sukurtas **hmC-CATCH** (angl. *Chemical-assistant C-to-T conversion of 5hmC sequencing*) metodas, paremtas 5hmC oksidacija į 5fC su kalio rutenatu ir tolimesniu cheminiu 5fC žymėjimu, kuris po PGR nuskaitymas kaip T. Kalio rutenatas (K<sub>2</sub>RuO<sub>4</sub>) sukelia nedidelį DNR praradimą, todėl galima naudoti gana mažus pradinės DNR kiekius. Metodas buvo pritaikytas vėžio DNR tyrimams (Zhao ir kt., 2020). **CLEVER-Seq** (angl. *chemical-labeling-enabled C-to-T conversion sequencing*) metode vyksta 5fC į T konversija po malononitrilo žymėjimo ir šis metodas buvo pritaikytas vienos ląstelės sekoskaitai (Zhu ir kt., 2017).

### 1.4.6 Trečios kartos sekoskaita

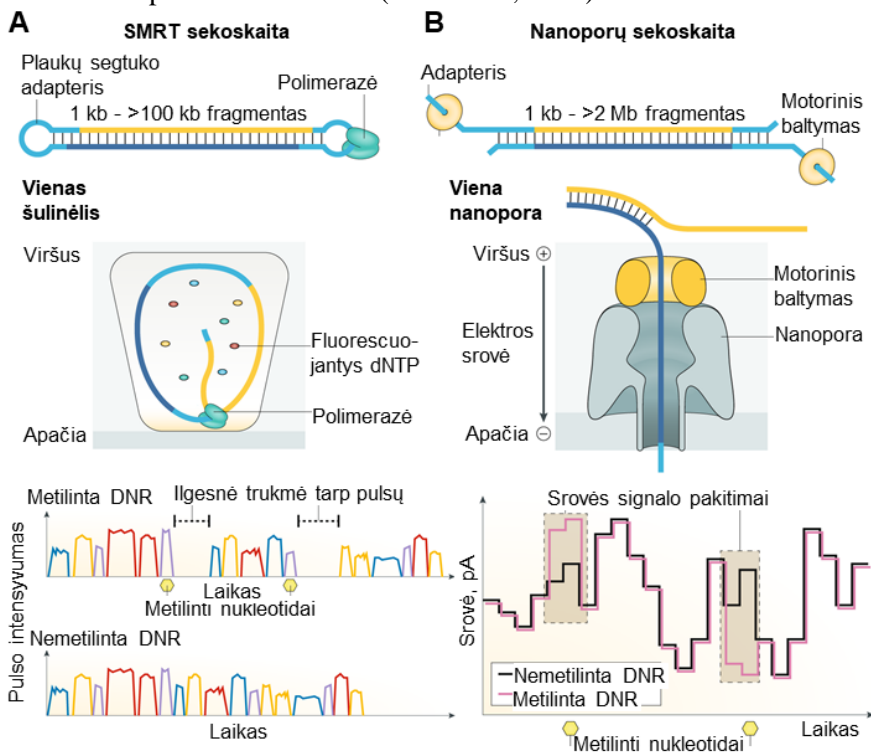
Dauguma minėtų DNR modifikacijų profiliavimo metodų yra paremti naujos kartos sekoskaita, kuri turi gana trumpą nuskaitymų sekų ilgį. Tuo tarpu trečios kartos sekoskaita galima gauti daug ilgesnius vienos molekulės nuskaitymus, ji pasižymi didesniu greičiu, pavyzdžių paruošimui nereikia PGR padauginimo. Taip pat ilgų sekų nuskaitymas pagerina pasikartojančių sekų vietos nustatymą genome (Plongthongkum ir kt., 2014). Nors trečios kartos sekoskaitai reikia daugiau DNR, ji yra patraukli dėl galimo tiesioginio DNR modifikacijų nustatymo (nukleotido tikslumu) be papildomo modifikavimo (Zhao ir kt., 2020).

Vienos molekulės realaus laiko sekoskaitos (**SMRT**, angl. *Single-Molecule Real-Time*, *Pacific Biosciences*) metu polimerazė įjungia skirtingai fluorescuojančius žymėtus nukleotidus, realiu laiku stebima fluorescencija ir individualios polimerazės kinetika (1.18 pav. A) (Zhao ir kt., 2020). Polimerazės replikacijos greitis ties ir aplink modifikuotas bazes yra lėtesnis lyginant su nemodifikuotomis (Wu ir Zhang, 2014). Stebint šį kinetinį parašą ir įjungimo laiką galima nustatyti 5mC, 5hmC, 5fC, 5caC ir 6mA (Shen ir kt., 2014). Tačiau parašo skirtumai gali būti gana menki ir taip lemiantys nustatymo netikslumus, todėl geresniam modifikacijų atskyrimui galima atlikti papildomą bazių modifikavimą, didinantį kinetinius skirtumus. 5fC ir 5caC sukelia žymų replikacijos uždelsimą be papildomo modifikavimo, tačiau sunku juos atskirti vieną nuo kito (Pastor ir kt., 2013). Silpno 5mC signalo sustiprinimui jis oksiduojamas iki 5caC, o 5hmC nustatymui pritaikytas hMe-Seal metodas, padedantis jį atskirti nuo 5mC. Tokiu būdu, pasitelkiant cheminį ar fermentinį žymėjimą, galima kiekvienoje grandinėje ir gana dideliu patikimumu nustatyti 5mC ir jo oksiduotas bazes (Plongthongkum ir kt., 2014; Shen ir kt., 2014; Wu ir Zhang, 2014).

Kita vienos molekulės sekoskaitos technologija, kuri turi ilgiausią nuskaitymų sekų ilgį ir geba atskirti DNR modifikacijas – **nanoporų sekoskaita** (*Oxford Nanopore Technologies*). Jos metu, realiu laiku ir be PGR, yra stebimas srovės signalas, kai individualios DNR molekulės praveriamos pro nanometrines baltymų (arba kietos būsenos sintetines) poras (1.18 pav. B). Skirtingos molekulės sukuria joms būdingą signalą, kuris pasitelkiamas atskiriant nukleozidus vieną nuo kito (Zhao ir kt., 2020). Su modeliniais DNR substratais parodyta, kad nanoporų sekoskaita galima atskirti skirtingas citozino modifikacijas vieną nuo kitos. Modifikacijų nustatymui reikia sudėtingų skaičiavimų, kontrolinės DNR, su kuria vykdomas algoritmų apmokymas. Nustatymo tikslumas priklauso nuo sekos

konteksto, naudojamos DNR ir skaičiavimų (pvz., viename tyrime parodytas ~92-98% citozino modifikacijų nustatymo tikslumas (Wescoe ir kt., 2014), kitame – ~80% (Rand ir kt., 2017)). Neseniai nanoporų sekoskaita buvo pritaikyta žmogaus metilomos kiekybiniam įvertinimui (Simpson ir kt., 2017). Tačiau ši technologija dar nėra pritaikyta oksi-mC sekoskaitai natūraliame genome be papildomo modifikavimo.

Jungiant įvairius modifikacijų profiliavimo metodus su trečios kartos sekoskaita galima tikėtis gana tikslios ilgų sekų epigenetinės sekoskaitos. Pavyzdžiui, ilgiems nuskaitymams SMRT ir nanoporų sekoskaita buvo pritaikytas TAPS metodas (Liu ir kt., 2020). Tačiau platesniam DNR modifikacijų nustatymo pritaikymui reikia tolimesnio šių technologijų tikslumo bei apimties tobulinimo (Zhao ir kt., 2020).



**1.18 pav.** Trečios kartos sekoskaitos metodai. **(A)** *Pacific Biosciences* vienos molekulės realaus laiko sekoskaitoje (angl. *Single-Molecule Real-Time*, SMRT) prie fragmentuotos DNR prijungiami plaukų segtuko struktūros adapteriai (šviesiai mėlyna) ir suformuojama žiedinė molekulė, kuri su prisijungusia polimeraze patalpinama į pikolitrų talpos šulinėlio dugną. Fluorescuojančią žymę turintis dNTP trumpai prisijungia polimerazės aktyviame centre ir tuomet iš apačios sužadinta fluorescencija (skirtingo bangos ilgio kiekvienai bazei) yra detekuojama kameros. Šviesos pulsų seka parodo nuskaitytą seką, o tarpai tarp pulsų – DNR matricoje esančias kovalentines modifikacijas. Įjungiant nukleotidą fluorescuojanti žymė nuskyta, o sužadėjimas vyksta tik pačiame dugne. **(B)** *Oxford Nanopore Technologies*

sekoskaitoje prie ilgos DNR prijungiami adapteriai su prisijungusiu motoriniu baltymu. Sekoskaitos kasetės sintetinėje lipidų membranoje yra įterptos baltyminės nanoporos. Sekoskaitos adapteris įsiterpia į nanoporą, o motorinis baltymas išvynioja DNR grandinę. Elektros srovė kartu su motoriniu baltymu traukia neigiamai įkrautą DNR grandinę pro porą. Judėdama DNR sukuria charakteringus srovės tekėjimo trukdžius priklausomai nuo nukleotidų. Metilinti nukleotidai įveda išsiskiriančius signalus ir modifikacijos nustatomos lyginant su nemetilintos sekos signalais. Pagal Beaulaurier ir kt. (2018), Logsdon ir kt. (2020).

## 2. MEDŽIAGOS IR METODAI

### 2.1 Medžiagos

#### 2.1.1 Reagentai

*BD - Difco*: agaras, mielių ekstraktas, salyklo ekstraktas.

*BroadPharm*: DBCO-S-S-PEG3-biotinas.

*Carl Roth*: akrilamido-bisakrilamido tirpalas *Rotiphorese gel* 40% (19:1), chloroformas, fenolis (pH 8), etidžio bromidas, amonio persulfatas, tetrametiletilendiaminas, NaOH, NaCl, KCl, LiCl, HCl, CH<sub>3</sub>COONa, NaH<sub>2</sub>PO<sub>4</sub>, HEPES, EDTA, Tris, izopropanolis, Tween 20, glicerolis, vitaminas C, alfa-ketoglutaratas, FeSO<sub>4</sub>, acetonitrilas, metanolis, skruzdžių rūgštis, natrio dodecilsulfatas (NDS), 2-merkaptioetanolis, urėja, gliukozė, metileno mėlis.

*Jena Bioscience*: UDP-6-azido-gliukozė, DBCO-S-S-PEG3-biotinas.

*MERCK*: CaCl<sub>2</sub>, formamidas.

*Molecular Research Center, Inc*: RNazol RT.

*PerkinElmer*: [ $\gamma$ -<sup>33</sup>P]-ATP, >2500 Ci/mmol.

*Reachim*: CH<sub>3</sub>COOH, H<sub>3</sub>BO<sub>3</sub>.

*Sigma-Aldrich*: DMSO, DTT, CH<sub>3</sub>COONH<sub>4</sub>, SAH, N-hidroksisukcinimidas, CuBr (99,999%), THPTA, 1-etil-3-(3-dimetilaminopropil)-karbodiimido hidrochloridas, piridino boranas, 3,3',5,5'-tetrametilbenzidinas (TMB), NaBH<sub>4</sub>.

*Thermo Fisher Scientific*: agarozė, MgCl<sub>2</sub>, ATP, dATP, dNTP, 2×RNR mėginių dažas, 6×DNR mėginių dažas, glikogenas, SAM, PEG 4000, SYBR™ Gold.

*Trilink Biotechnologies*: citozino modifikacijas turintys nukleotidai.

*Vilniaus degtinė, AB*: 96% etanolis.

## 2.1.2 DNR preparatai

### DNR ilgio standartai

DNR markeriai *GeneRuler™ Ladder Mix*, *MassRuler™ Low Range DNA*, *MassRuler™ High Range DNA Ladder* – *Thermo Fisher Scientific*.

### Genominė DNR

λ bakteriofago DNR (dam–, dcm–), žmogaus kraujo DNR – *Thermo Fisher Scientific*. Žmogaus IMR90 genominė DNR – ECACC, žmogaus smegenų genome DNR – gauta iš prof. A Petronio.

**Sintetiniai oligonukleotidai** (*Metabion*, jei nenurodyta kitaip), \* žymi fosforotioato jungtis.

Oligonukleotidai TLC eksperimentams:

C-14 5′-CGCATCATACTCTA

5caC-14 5′-5caCGCATCATACTCTA

C-17 5′-CGGATGTTGTGGGTCAG

5mC-31 3′-GCCTACAACACCCAGTCG5mCGTAGTATGAGAT

C-31 3′-GCCTACAACACCCAGTCGCGTAGTATGAGAT

Pradmenys DNR fragmentų paruošimui PGR būdu:

BMX-dir 5′-TGTGTTACTGTGTGGAAAAGACC

BMX-rev 5′-CCACTCCTTATAGTTTGGCTG

Sp-PCR-dir 5′-GTGTTGGGGTGACTATTATG

Sp-PCR-rev 5′-CCTATACTCAGCGCATCC

188-rev 5′-GCATCCTGGAGATTGTGGGCA

Mod-PCR-rev 5′-GCATCCTGGAGATTGTGGGCAACATCMGG

(M=5hmC (*IBA GmbH*), 5fC (*TriLink Biotechnologies*) arba 5caC)

c-fos-gu-dir 5′-TTACACAGGATGTCCATATTAGG

c-fos-gu-rev 5′-CTGTGGAGCAGAGCTGGGTA

Pradmenys 188 bp fragmento praturtinimo analizei kiekybine PGR (kPGR):

Su-test-dir 5′-ATGTGTTGGAGTGTGCCTGA

Su-test-rev 5′-GTGGCTCTGATTGATGGCTC

Oligonukleotidas taškiniam imunoblotui:

pTAA-5caC 5′-TAATAATAAA5caCCGTAATAATAATAAT

Oligonukleotidai TOP-seq bibliotekų paruošimui (X – keičiamos barkodų sekos nukleotidai):

A1 adapteris 5'-P-GATTGGAAGAGTGGTTCAGCAGGAATGCTGAG, P = 5'-fosfatas

A2 adapteris 5'-ACACTCTTTCCTACATGACACTCTTCCAATCT ODN („click“ būdu prijungiamas oligonukleotidas, su arba be biotino): 5'-TXXXXTGTGTGGTTTGGAGACTGACTACCAGATGTAACA-biotinas, (X=C8-alkin-dT), *Base-click*

ODN komplementarus pradmuo 5'-

TGTTACATCTGGTAGTCAGTCTCCAAACCACA\*C\*A\*A, *Exiqon*

A-BC-S (PGR pradmuo su barkodu) 5'-

CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXXXTGTTCATCTGGTAGTCAGT\*C\*T\*C

trP1 (PGR pradmuo) 5'-

CCTCTCTATGGGCAGTCGGTGATCACTCTTTCCTACATGAC\*A\*C\*T

*Ion Torrent* platformos adapteriai bisulfitinei sekoskaitai (visi citozinai sekoje buvo pakeisti 5mC):

A-BS-se 5'-

CCATCTCATCCCT\*G\*CGTGTCTCCGACTCAGXXXXXXXXXXXXGAT

A-BS-as 5'-ATCXXXXXXXXXXCTGAGTCGGAGACACGC

P1-BS-se 5'-

CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT

P1-BS-as 5'-

ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGG\*T\*T

*Ion Torrent*<sup>TM</sup> platformai skirtų adapterių rinkinys (ne TOP-seq bibliotekoms) su barkodais ligavimui iš *Kapa Biosystems*.

Pradmenys *Ion Torrent* bibliotekų (ne TOP-seq) padauginimui:

Ion-PCR-Fw 5'-CCACTACGCCTCCGCTTTCCTCTCT\*A\*T\*G

Ion-PCR-Rv 5'-CCATCTCATCCCTGCGT\*G\*T\*C

Pradmenys IMR90 DNR genominių regionų analizei:

Regio- no nume- ris	Genominis elementas	kPGR pradmenys	Vieta chromoso- moje (hg19 genomas)
#1	Tarpge- nis regionas	Inter1-dir 5'- TGGTGTCCCCTAAAGCATAAGT Inter1-rev 5'- TCAAGCCACATTTCCCATCG	chr7: 117340531- 117340636
#2	<i>CDH13</i> genas	CDH1-dir 5'- CAGAAATGCAGTGATGGGTGA CDH1-rev 5'- GGGCAGTCCTTAATCGGGA	chr16: 82662108- 82662222
#3	<i>AGTR1</i> genas	AGTR -dir 5'- TCTTCTTACATGGGCCTATGTGA AGTR -rev 5'- GGACTAGGAGAAGCTGAGGG	chr3: 148426487- 148426616
#4	<i>TOP3A</i> genas	TOP3A-dir 5'- CAGTACATGACAGAGAGCCCT TOP3A-rev 5'- TGTGTCATGCAGAGGGTCAT	chr17: 18210312- 18210435
#5	<i>IGF1R</i> genas	IGFR-dir 5'- ATGCTCCTTATGTACCATGTGC IGFR-rev 5'- TGCTGTCCAAATTATCCATCAAG	chr15: 99324967- 99325066
#6	<i>CDH13</i> genas-2	CDH2-dir 5'- AAGTGGGTTCCCTGTCTCAG CDH2-rev 5'- GAGGGATTGCCTCTGACCTT	chr16: 83191684- 83191805

Pradmenys skirti pelės genominių regionų bisulfitinei caMAB-seq ir piridino borano sekoskaitai:

Geno pavadini- mas, grandinė	caMAB-seq pradmenys	Vieta chromo- somoje



		(mm10 genomas)
Nr. 1, <i>Neol</i> , +	Fw 5'- GGAAGAAAAATATAGAGGGTAAAATG Rv 5'- ATTCCTAACCAAATCACCATAAAAA	chr9: 58979144- 58979315
Nr. 1, <i>Neol</i> , -	Fw 5'- GGAAATTTTTAGTTGGTGGTATTATTT Rv 5'-CAATCTAAAACCAACTTCACCACA	chr9: 58979094- 58979293
Nr. 2, <i>Crhr1</i> , +	Fw 5'-GTAGGTTTGGTGTAGGTTGGTGATA Rv 5'-CTAACCCACATCCTCAATAAAAA	chr11: 104170118- 104170327
Nr. 2, <i>Crhr1</i> , -	Fw 5'- TTGGTTTATATTTTTAGTGAGGAGAGG Rv 5'- CAAACCTAATACAAACTAATAACAACC	chr11: 104170119- 104170327
Nr. 3, <i>Nudcd3</i> , +	Fw 5'-GAAAAGTTATTGTTGGTGGGAGAG Rv 5'- CCAATACACCTCAAACATTTAAACTC	chr11: 6163918- 6164101
Nr. 3, <i>Nudcd3</i> , -	Fw 5'- ATTTTATTGTTTGAGATAGGATTGTT Rv 5'- AACTAAATACACAAAACCAAACACACA	chr11: 6163941- 6164136
Nr. 4, <i>4933411K16</i> , +	Fw 5'- GGAGGTTTTATGTTGTGGGTTTATAG Rv 5'- CTTCCCATACCTACTTTCAAATAAA	chr19: 42052423- 42052618
Nr. 4, <i>4933411K16</i> , -	Fw 5'- GTTTTTAGGTGAGTGGTTATTGTTGT Rv 5'- AAAAACCTCATACTATAAACTCACAAA	chr19: 42052423- 42052605
Genominis elementas	Pradmenys piridino borano sekoskaitai	Vieta chromo- somoje (mm10 genomas)
Nr. 1, <i>Neol</i> genas	Fw 5'-CTATTCCTAGCCAGGTCACCATAG Rv 5'- CCATGATGAGATTACTAACAGGAAGA	chr9: 58979124- 58979317

Nr. 2, <i>Crhr1</i> genas	Fw 5'-CTGGTGCAGGCTGGTGAC Rv 5'-CTGGCCCACATCCTCAGT	chr11: 104170124- 104170327
Nr. 3, <i>Nudcd3</i> genas	Fw 5'-AGTCACTGCTGGTGGGAGAG Rv 5'-GCCAGTACACCTCAGGCATT	chr11: 6163922- 6164102
Nr. 4, <i>4933411K16</i> genas	Fw 5'-GTGGGCTCACAGGAAGAAAA Rv 5'-ATGCCTGCTTTCAGGTGAGT	chr19: 42052437- 42052611
Nr. 5, <i>Cercam</i> genas	Fw 5'-ATGCTTTGGGAGCTACAGGA Rv 5'-TCAAGTCCCATGACAGCTTCT	chr2: 29880548- 29880739
Nr. 6, tarpgeninis regionas	Fw 5'-AGGGTTGAGATCAGCCTCCT Rv 5'-TCTGGAATTTCCAAGCTTCA	chr9: 69748564- 69748756

### 2.1.3 RNR preparatai

*Thermo Fisher Scientific*: RNR ilgio standartas – *RiboRuler™ Low Range RNA Ladder*.

### 2.1.4 Baltymų preparatai

*Active Motif*: NgTet1 (*Naegleria gruberi*), triušio antikūnai prieš 5caC (Cat# 61225), su krienų peroksidaze HRP sujungti ožkos antikūnai prieš triušio IgG, (Cat# 15015).

*BPS Bioscience*: rekombinantinės žmogaus metiltransferazės DNMT3A/DNMT3L (Lot#100824, 1,8 mg/ml, ≥95% grynumas), DNMT3B/DNMT3L (Lot#120911, 0,13 mg/ml, ≥85% grynumas, arba Lot#121003, 0,19 mg/ml, ≥90% grynumas).

*Pierce*: Jaučio serumo albuminas (BSA, 2 mg/ml).

*Sigma-Aldrich*: Nukleazė P1.

*Thermo Fisher Scientific*: ribonukleazė A, A/T1 ribonukleazių mišinys, Proteinazė K, T4 polinukleotidkinazė (PNK), T4 DNR ligazė, DNR polimerazės (Taq, DreamTaq, Klenowo exo-, Pfu, *Phusion U HS*), PGR mišinys *Platinum SuperFi PCR Master Mix (2X)*, šarminė fosfatazė FastAP,

restrikcijos endonukleazės (HpaII, Hin6I, MspI, TaiI, Eco47III, Crf10I, EheI, NotI), DNazė I, T4 β-gliukoziltransferazė (βGT), agarazė, laukinio tipo (wt, angl. *wild type*) M.SssI ir fermentų buferiniai tirpalai.

Mutantinė eM.SssI išgryninta Giedrės Urbanavičiūtės (Kriukienė ir kt.; 2013, 3.2, 3.5 A, B pav. eksperimentai) arba *Thermo Fisher Scientific* (visi kiti eksperimentai). M.HhaI ir M.HpaII MTazės buvo išgrynintos VU BTI DNR modifikacijos tyrimų skyriuje pagal aprašytą metodiką (Vilkaitis ir kt., 2000; Lukinavičius ir kt., 2012).

### 2.1.5 Molekulinės biologijos rinkiniai

*Agilent Technologies*: kapiliarinės elektroforezės rinkiniai bioanalizatoriui *Agilent HS DNA kit*, *Agilent RNA 600 Nano Kit*, *Agilent RNA 600 Pico Kit*.

*Beckman Coulter*: magnetinės dalelės *Ampure XP*.

*Bio-Rad*: chromatografijos kolonėlės *Micro Bio-Spin*.

*Illumina*: rRNR pašalinimo rinkinys *Ribo-Zero Gold (Yeast)*.

*Thermo Fisher Scientific*: didelės išeigos transkripcijos *in vitro* rinkinys *TranscriptAid™ T7 High Yield Transcription Kit*, rinkinys gryninimui iš agarozės gelio *GeneJet Gel Extraction Kit*, kiekybinės PGR mišinys *Maxima SYBR Green/ROX qPCR Master Mix (2X)*, DNR gryninimo rinkiniai *GeneJet PCR purification kit*, *NGS cleanup kit*, *MagJET NGS Cleanup and Size Selection Kit*, DNR galų sutvarkymo rinkinys *Fast DNA End Repair Kit*, bisulfitinės konversijos rinkinys *EpiJET Bisulfite Conversion Kit*, nukleorūgščių koncentracijos nustatymo rinkiniai *Qubit RNA HS Assay Kit*, *Qubit DNA HS Assay Kit*, iRNR gryninimo rinkinys *Dynabeads™ mRNA Purification Kit*, streptavidino magnetinės dalelės *Dynabeads MyOne Streptavidin C1*.

*WiseGene*: 5-mC oksidacijos rinkinys *5mC Tet1 Oxidation Kit*.

*Zymo Research*: RNR gryninimo rinkinys *RNA Clean & Concentrator Kit*, DNR gryninimo rinkiniai *DNA Clean and Concentrator kit*, *Genomic DNA Clean and Concentrator kit*, bisulfitinės konversijos rinkinys *EZ DNA Methylation-Gold™ Kit*, oligonukleotidų gryninimo rinkinys *Oligo Clean and Concentrator kit*.

## 2.1.6 Buferiniai tirpalai

Eliucijos buferis (EB): 10 mM Tris-HCl pH 8,5 (~25°C).

Eliucijos buferis nukleorūgščių skyrimui iš poliakrilamidinio gelio: 0,5 M CH<sub>3</sub>COONH<sub>4</sub>, 0,1 mM EDTA, 0,1% NDS.

P1 buferis: 10 mM NaOAc pH 5,2, 1 mM ZnOAc.

eM.SssI dekarboksilinimo buferis (1×): 10 mM Tris-HCl pH 8 (~22°C), 50 mM NaCl, 0,5 mM EDTA.

STE buferis (1×): 50 mM NaCl, 10 mM TrisHCl pH 8 (~22°C), 0,5 mM EDTA (pH 8).

M.HhaI buferis (1×): 10 mM Tris-HCl pH 7,4 (~22°C), 50 mM NaCl, 0,5 mM EDTA.

Etidžio bromido tirpalas: 5 µg/ml etidžio bromido vandenyje.

## 2.2 Metodai

### 2.2.1 Plonasluoksnė chromatografija (TLC)

31 bp DNR duplexų paruošimui sintetinis C-14 arba 5caC-14 DNR oligonukleotidas pažymėtas 5'-<sup>33</sup>P naudojant [ $\gamma$ -<sup>33</sup>P]-ATP ir T4 polinukleotidkinazę pagal gamintojo rekomendacijas. Po išsodinimo etanoliu, oligonukleotidas sulydytas su C-17 ir 5mC-31 arba C-31 oligonukleotidais (pakaitinta 5 min 85°C ir temperatūra buvo mažinama po 0,01°C/s iki 4°C) ir tarpas sujungtas naudojant T4 DNR ligazę pagal gamintojo rekomendacijas. Siekiant gauti abiejose grandinėse 5caC turinčią DNR, 5caC/5mC duplexas buvo oksiduotas naudojant TET1 (žr. 2.2.2).

Dekarboksilinimo reakcijose ~90 nM DNR buvo inkubuojama su 500 nM eM.SssI (3 val. 22°C, esant 50 mM Tris-HCl pH 8,0 (~22°C), 0,25 mM EDTA, 5 mM NaCl) arba 160-400 nM žmogaus MTazių (37°C temperatūroje 4-6 val., esant 10 mM Tris-HCl pH 7,4 (~22°C), 50 mM NaCl, 0,1 mg/ml BSA, 2 mM DTT).

Reakcijai su cisteinu, MTazės inkubuotos su 20 mM *N*-etilmaleimidu 1 val. 25°C be DNR. DNR buvo išsodinta etanoliu ir sukurpta į nukleotidus su P1 nukleaze (0,02-0,01 U/µl) 1 val. 55°C. Mėginiai buvo užnešti ant TLC plokštelių (*PEI Cellulose F*, *MERCK*) ir išskirstyti buferyje A (izosviesto rūgštis/vanduo/konc. amoniakas, tūriu santykiu 66:17:4) arba B (izopropanolis/37% HCl/vanduo, tūriu santykiu 4,7:1:1). Plokštelės buvo išdžiovintos, vizualizuotos autoradiografijos būdu ir rezultatai analizuoti naudojant FLA-5100 skaitytuvą ir *MultiGauge* programinę įrangą (*Fujifilm*).

## 2.2.2 Modelinių DNR fragmentų paruošimas

188 bp (1 CG taikiny) ir 200 bp modeliniai DNR fragmentai buvo gauti atliekant PGR su Sp-PCR-dir ir Sp-PCR-rev pradmenimis 200 bp fragmentui, Sp-PCR-dir ir 188-rev 188 bp fragmentui nuo C57BL/6 pelės genominės DNR (išskirtos iš uodegos, kuri gauta iš VU BTI Imunologijos ir ląstelės biologijos skyriaus). Modifikacijos (5hmC, 5fC arba 5caC) į 188 bp fragmento vieną CG taikinio grandinę įvestos su Mod-PCR-rev pradmeniu. 155 bp fragmentas (BMX geno dalis, 1 CG taikiny) buvo padaugintas nuo žmogaus DNR su BMX-dir ir BMX-rev pradmenimis. 428 bp fragmentas (37 CG) buvo padaugintas nuo žmogaus *c-fos* geno promotoriaus regiono (įterpto pUC19 plazmidėje, kuri gauta iš Giedrės Urbanavičiūtės) su c-fos-gu-dir ir c-fos-gu-rev pradmenimis. Po PGR fragmentai buvo išpjauti iš agarozės gelio ir išgryninti *GeneJET Gel Extraction* rinkiniu.

Siekiant įvesti 5hmC į GCGC taikinius, DNR buvo inkubuojama esant 5-ių kartų M.HhaI pertekliui prieš GCGC taikinius jos buferyje su 13 μM formaldehidu kambario temperatūroje 1 val. Tada mėginiai buvo veikiami 0,2 mg/ml proteinaze K (ProtK) ir 0,1% NDS 1 val. 50°C, o po to 20 min. 65°C. DNR buvo išgryninta *DNA Clean and Concentrator* rinkiniu. Šiame ir tolesniuose kolonėlių gryninimo etapuose DNR eliuavimui buvo naudojamas 10 mM Tris-HCl pH 8,5 (~25°C) tirpalas, jei nenurodyta kitaip.

Siekiant vienoje grandinėje modifikuotą citoziną turinčių modelinių DNR fragmentų kitoje grandinėje įvesti 5mC, nemodifikuoti CG buvo metilinti naudojant wt M.SssI pagal gamintojo protokolą. Į abi DNR fragmentų grandines 5caC buvo įvestas po nemodifikuoto DNR fragmento metilinimo su TET1 atlikus 5mC oksidaciją *5mC Tet1 Oxidation* rinkiniu pagal gamintojo rekomendacijas.

caCLEAR eksperimentui su dviem modeliniais DNR fragmentais, 5caC 188 bp fragmente buvo įvestas per PGR, naudojant atvirkštinį pradmenį su 5caC CG taikinyje. Šis fragmentas buvo naudojamas kartu su 155 bp fragmentu, turinčiu M.HhaI įvestą 5hmC GCGC taikinyje.

## 2.2.3 DNR analizė naudojant restrikcijos endonukleazes ir kiekybinę PGR

10-20 ng DNR buvo inkubuojama su 3-10 U atitinkamos REazės (MspI, HpaII, Hin6I, TaiI, Eco47III, Crf10I, EheI, NotI) gamintojo buferyje 37°C 3 val. ir atliktas rekomenduojamas išveiklinimas karščiu.

Kirpimas buvo analizuojamas kPGR būdu su pradmenimis, skirtais pasirinktoms genominės DNR sritims arba fragmentams. Nesukirptos DNR

kiekiai buvo apskaičiuoti naudojant žinomų DNR koncentracijų kalibravimo kreivę. Analizei buvo apskaičiuota karpymui atsparios DNR dalis (%) nuo DNR kiekio, nustatyto REaze neveiktame kontroliniame mėginyje. Kiekybinės PGR reakcijos buvo atliktos su *Rotor-Gene Q* tikro laiko PGR sistema (*Qiagen*), naudojant *Maxima SYBRGreen qPCR Master Mix* PGR mišinį. Reakcijose buvo 0,3 μM kiekvieno pradmens. kPGR reakcija vykdyta taip: 95°C 10 min., 40 ciklų 95°C 15 s, 60°C 1 min. 188 bp DNR fragmento praturtinimo analizei buvo naudojami CG taikinio neapimantys pradmenys Su-test-dir ir Su-test-rev.

#### **2.2.4 eM.SssI vykdomo dekarboksilinimo ir žymėjimo azido grupe įvertinimas su modeliniais DNR fragmentais**

3.5 pav. A parodytam dekarboksilinimo eksperimentui su skirtinga pH, buvo naudota 100 ng 188 bp fragmento su 5caC vienoje grandinėje. Reakcija atlikta 20 μl tūryje, dekarboksilinimo buferyje 30 min. 22°C esant 1,31 μM eM.SssI, fermentas išveiklintas 20 min. 65°C. 3.5 pav. B eksperimente 188 bp fragmentas buvo sumaišytas su ultragarsu fragmentuota žmogaus smegenų genomine DNR (1:100 ng, kad būtų imituojamas mažas 5caC gausumas) ir dekarboksilinimas vykdytas 3 val. (ir tolesniuose eksperimentuose). Skirtingo eM.SssI (1,13 μM) pertekliaus prieš CG taikinius eksperimentui (3.5 pav. C) 188 bp fragmentas sumaišytas su ultragarsu fragmentuota lambda bakteriofago DNR santykiu 1:150 ng. Tada mėginiai buvo paveikti ProtK ir išgryninti (žr. 2.2.2). Likęs 5caC lygis buvo įvertintas kPGR būdu pagal R.MspI kirpimui atsparios DNR kieki.

Žymėjimui azido grupe, dekarboksilinimo reakcijos mišinys (be gryninimo ir išveiklinimo) buvo papildytas Ado-6-azido kofaktoriumi iki 0,2 mM koncentracijos (Lukinavičius ir kt., 2013; Masevičius ir kt., 2016; 3a junginys, susintetintas prof. Viktoro Masevičiaus laboratorijoje) ir inkubuotas 30°C 1 val., pakaitintas 65°C 20 min., paveiktas ProtK ir išgrynintas (žr. 2.2.2).

Žymėjimo įvertinimui praturtinimo būdu (3.5 pav. E) 188 bp modelinis fragmentas sumaišytas su ultragarsu fragmentuota žmogaus kraujo DNR (1:300 ng). Pažymėta DNR buvo biotinilinta inkubuojant 10 mM Tris-HCl pH 8,5 (~22°C) tirpale 37°C 2 val iki 0,4 mM koncentracijos pridėjus DBCO-S-S-PEG3-biotino (*BroadPharm*). Po to DNR buvo išgryninta *DNA Clean and Concentrator* rinkiniu. Tada pažymėta DNR buvo praturtinta: 0,1 mg *Dynabeads MyOne Streptavidin C1* buvo inkubuota su DNR 10 mM Tris-HCl pH 8,5 (~22°C), 1 M NaCl 3 val. kambario temperatūroje, vartant. Po to

dalelės buvo tris kartus plaunamos su 15 mM Tris-HCl pH 7,5 (~22°C), 1,5 mM EDTA, 3 M NaCl ir du kartus su 5 mM Tris-HCl pH 7,5 (~22°C), 0,5 mM EDTA, 1 M NaCl. Dalelių ir supernatanto frakcijose esantis DNR kiekis buvo įvertintas kPGR.

### **2.2.5 eM.SssI vykdomo dekarboksilinimo analizė IMR90 genominiuose regionuose**

IMR90 DNR buvo fragmentuota ultragarsu naudojant *Bioruptor UCD-200 (Diagenode)* iki ~500 bp fragmentų. 5mC oksidacija iki 5caC buvo atlikta naudojant TET1 (žr. 2.2.2). 75 µl reakcijoje 65 ng DNR su 1,13 µM eM.SssI buvo inkubuojama dekarboksilinimo buferyje 22°C 3 val. ir išgryninta kaip aprašyta 2.2.2. Tuomet 9 ng genominės DNR buvo kerpama su 10 U atitinkamos restrikcijos endonukleazės. DNR buvo analizuojama kPGR (žr. 2.2.3) būdu naudojant pasirinktų genominių regionų pradmenis (žr. 2.1.2). 5mC ir 5caC lygiai buvo įvertinti pagal nesukirptos DNR kiekį (%) atitinkamai naudojant R.HpaII ir R.MspI. Dekarboksilinimo efektyvumas buvo apskaičiuotas kaip R.MspI nesukirptos DNR kiekio skirtumas prieš ir po dekarboksilinimo.

### **2.2.6 5hmC ir nemetilintų CG taikinių apsaugojimas**

5hmC buvo glikozilinamas 37°C 2 val. naudojant T4 beta-gliukoziltransferazę (βGT) pagal gamintojo rekomendacijas (5 U βGT ir 1 µg DNR 50 µl reakcijoje), o po to βGT buvo išveiklinta 65°C 20 min. Mėginiai buvo paveikti ProtK ir išgryninti (žr. 2.2.2). Antrasis βGT glikozilinimo etapas buvo atliktas 16 valandų 37°C (be veikimo ProtK) ir DNR vėl išgryninta.

Nemodifikuoti CG buvo metilinami naudojant 10-ties kartų laukinio tipo M.SssI fermento perteklių (1/10 reakcijos tūrio) prieš CG taikinius su 0,6 mM SAM gamintojo buferyje 37°C 4 val. Po fermento išveiklinimo mėginiai buvo paveikti ProtK ir išgryninti (žr. 2.2.2). Tuomet taip pat buvo atliktas antrasis metilinimo ir gryninimo etapas, tik metilinimas buvo atliktas 37°C 16 valandų. Metilinimo įvertinimui su modeliniais fragmentais (3.4 pav. B) buvo naudojamas 188 bp modelinis fragmentas sumaišytas su *E. coli* genome DNR (1:100ng).

### **2.2.7 wt M.SssI metilinimo efektyvumo ir 5caC lygio įvertinimas bisulfitine sekoskaita lambda bakteriofago DNR**

nCG metilinimo įvertinimui ultragarsu fragmentuotos lambda DNR (1,4 µg, žr. 2.2.8) galai buvo sutvarkyti *DNA End Repair* rinkiniu 15 min. 20°C pagal gamintojo rekomendacijas ir DNR išgryninta naudojant *GeneJet Purification Kit*. Tuomet 60 µl reakcijoje esant 15 U T4 ligazės, jos buferio, 0,5 mM ATP, 10 mM MgCl<sub>2</sub> ir 5% PEG 4000 buvo prijungti metilinti *Ion Torrent* platformos adapteriai (A-BS ir P1-BS, žr. 2.1.2, 42 pmol, suardyti STE buferyje) inkubuojant per naktį 22°C. DNR išgryninta *GeneJET NGS Cleanup* rinkiniu. Ligazės buferyje 30 min. 37°C atliktas fosforilinimas su 5 U T4 PNK ir likęs tarpas sujungtas inkubuojant su 5 U T4 ligazės 1 val. 20°C, išveiklinta 1 min. 60°C. DNR vėl išgryninta *GeneJET NGS Cleanup* rinkiniu. Tada buvo pritaikytas metilinimo protokolas (žr. 2.2.6) ir atlikta bisulfitinė konversija naudojant *EpiJET Bisulfite Conversion Kit*. Tada naudojant *Phusion U HS* polimerazę ir Ion-PCR-Fw bei Ion-PCR-Rv pradmenis DNR buvo padauginta 15 ciklų. Bibliotekų koncentracijos buvo įvertintos naudojant *Agilent Bioanalyzer*.

5caC lygio įvertinimui prie lambda DNR (su įvestu 20% 5caC, žr. 2.2.8) buvo prijungti adapteriai, kaip aprašyta aukščiau. Tuomet po TET1 oksidacijos galintis likti nedidelis 5fC kiekis buvo redukuotas į 5hmC naudojant NaBH<sub>4</sub> (žr. 2.2.12). Po to 5caC lygio įvertinimui atlikta ta pati metilinimo, bisulfitinės konversijos ir padauginimo procedūra kaip aprašyta aukščiau. Siekiant įvertinti netinkamą bisulfito konversiją, PGR būdu su 5mCTP ir Pfu polimeraze buvo paruoštas kontrolinis 200 bp fragmentas, kuriame visi citozinai yra metilinti. Šis fragmentas sumaišytas su lambda DNR mėginiais prieš natrio bisulfito konversiją. Skaičiuojant konversiją buvo naudojamas tik regionas, esantis už pradmenų jungimosi vietų.

### **2.2.8 Lambda DNR paruošimas caCLEAR analizei**

Naudojant *Covaris M220* prietaisą DNR buvo fragmentuota ultragarsu iki vidutiniškai 200-250 bp fragmentų. Pirmajame eksperimente (3.6 pav.) 5mC buvo įvestas į CG naudojant 2,5 kartų wt M.SssI perteklių prieš jos taikinius, reakciją vykdant 4 valandas. GCGC sekos buvo metilintos esant 10-ties kartų M.HhaI pertekliui jos buferyje su 300 µM SAM 37°C temperatūroje 2 val. Fermentai buvo termiškai išveiklinti 20 min. 65°C, paveikti ProtK ir DNR išgryninta (žr. 2.2.2). M.SssI ir M.HhaI metilinta DNR buvo sumaišyta su nemonifikuota lambda DNR, kad CG ir GCGC taikiniai būtų modifikuoti atitinkamai ~30% ir ~60%. Tuomet su TET1 atlikta 5mC



oksidacija į 5caC (žr. 2.2.2). Kontroliniams mėginiams lygiagrečiai buvo paruošta lambda DNR, turinti 30% nemetilintų CG (ir 70% metilintų).

3.7 pav. B eksperimentui 5mC buvo įvestas į GCGC taikinius su M.HhaI metilinant 30 min., naudojant 0,125:1 fermento ir taikinių santykį. Tada 5mC buvo oksiduojamas 10-ties kartų NgTet1 pertekliumi esant 50 mM Bis-tris pH 6 (~22°C), 50 mM NaCl, 2 mM vitaminui C, 1 mM alfa-ketoglutaratui ir 0,1 mM FeSO<sub>4</sub> 1,5 val. 34°C temperatūroje. Tada reakcijos mišinys buvo papildytas 50 mM EDTA, paveiktas ProtK 30 min. 50°C ir išgrynintas. 5caC lygis buvo įvertintas naudojant masių spektrometriją (žr. 2.2.10). 5hmC buvo įvestas naudojant 10-ties kartų M.HpaII perteklių prieš CCGG taikinius inkubuojant su 13 μM formaldehidu M.HhaI buferyje 37°C 3,5 val. ir 20 min. 65°C. Tada mėginiai buvo paveikti ProtK ir išgryninti (žr. 2.2.2). Hidroksimetilinta ir karboksilinta DNR buvo metilinta wt M.SssI ir sumaišyta su 30% nemetilintos 5caC turinčios DNR, kad būtų gautas galutinis mėginys, kuriame yra 70% metilintų CG, 10% G5caCGC ir 10% C5hmCGG vietų. Taip pat buvo paruošta tokia pat kontrolinė lambda DNR be 5caC.

3.7 pav. C eksperimentui 5mC buvo įvestas naudojant M.HhaI (M.HhaI:GCGC santykis 0,125:1), o tada naudojant NgTet1 buvo atlikta 5mC oksidacija iki 5caC. Tuomet, sumaišius karboksilintą DNR su M.SssI metilinta lambda DNR, buvo paruoštas DNR mėginys su 20% 5caC ir 70% 5mC atitinkamai GCGC ir CG taikiniuose.

### 2.2.9 mESC auginimas ir DNR skyrimas

E14TG2a wt, mutantines *Tdg* neturinčias (*Tdg*<sup>-/-</sup>) ir visų TET1/2/3 neturinčias (TET TKO) pelių ESC linijas padovanojo prof. Guo-Liang Xu. Ląstelės buvo auginamos dr. Vaidoto Stankevičiaus serumo terpėje su leukemiją slopinančiu veiksmu (LIF) pagal Wu ir kt. (2014) arba serumo terpėje su 2i ir LIF pagal Yu ir kt. (2012).

mESC genomine DNR buvo išgryninta inkubuojant ląsteles skyrimo buferyje (200 mM Tris HCl pH 8,0 (~22°C), 250 mM NaCl, 25 mM EDTA ir 0,5-2% NDS, priklausomai nuo lizės) su ProtK 55°C 4-6 val. DNR buvo išgryninta lygiu tūriu fenolio, du kartus lygiu tūriu chloroformo ir išsodinta etanoliu. DNR ištirpinta EB ir paveikta 0,13 μg/μl ribonukleaze A 37°C 2 val. Tuomet DNR taip pat dar kartą išgryninta fenoliu-chloroformu ir jos koncentracija nustatyta *Qubit*<sup>TM</sup> *dsDNA HS Assay Kit* rinkiniu.

### 2.2.10 Aukšto slėgio skysčių chromatografija-tandeminė masių spektrometrija (HPLC-MS/MS)

Po fragmentavimo ultragarsu (iki ~1000 bp fragmentų), DNR buvo išgryninta naudojant *DNA Clean and Concentrator kit* ir eliuojama vandeniu. Tada DNR denatūruota 10 min. 80°C ir sukarpyta nukleaze P1 (naudojant ~0,33 U 1 µg DNR) 50°C temperatūroje 4 val. P1 buferyje ir 37°C per naktį defosforilinta pridėdant FastAP fosfatazės (~1 U/5 µg DNR). Reakcijos buvo sustabdytos kaitinant 75°C 10 min., o baltymų nuosėdos nusuktos 30 min. 14 000 × g 4°C. Nukleozidų standartai buvo paruošti defosforilinus nukleotidus. Mėginiai buvo analizuojami naudojant integruotą HPLC/ESI-MS/MS sistemą (*Agilent 1290 Infinity/6410B triple Quadruple*), turinčią *Supelco Discovery® HS C18* kolonėlę (7,5 cm × 2,1 mm, 3 µm), per kurią tirpalai buvo leidžiami 0,3 ml/min. greičiu 30°C temperatūroje. Dalį HPLC-MS/MS atliko ir sąlygas visur optimizavo Audronė Rukšėnaitė.

mESC 5caC ir 5fC analizei naudota ~22 µg karpytos DNR, kuri pirmiausia buvo per 4 kartus frakcionuojama HPLC naudojant linijinį A (10 mM amonio formiatas pH 5,5) ir B (80% metanolis) tirpalų gradientą: 0 min., 0% B; 9-23 min., 0-10% B; 23-27 min., 10-100% B; 27-31 min., 100% B; 31,5-37 min., 0% B. 5caC ir 5fC frakcijos buvo surinktos atitinkamai ~1,1-3 ir ~13,6-17,4 min., išdžiovinintos vakuume ir ištirpintos vandenyje. Kalibravimo kreivėms žinomų 5caC ir 5fC kiekių standartai buvo frakcionuojami iš mišinio su 5 µg dam<sup>-2</sup>cm<sup>-1</sup> lambda bakteriofago DNR. Grybų DNR oksimC frakcionavimui naudota ~9-23 µg DNR, taip pat frakcionuotas ir 5hmC (~4,3-5,8 min).

Masių spektrometrijos analizei buvo naudojamas A (0,0075 % skruzdžių rūgšties vandenyje) ir B (0,0075% skruzdžių rūgšties acetonitrile) tirpalų gradientas. Gradientas 5caC analizei: 0-6 min., 0% B; 6-10 min., 0-100% B; 10-14 min., 100 % B; 14,5-19,5 min., 0% B. Masių spektrometras veikė teigiamos jonizacijos MRM (angl. *multiple reaction monitoring*) režimu ir buvo stebimas nukleozidui specifinės m/z 272,1→156 jonų tranzicijos signalo intensyvumas. Jonizacijos kapiliaro įtampa 1800 V, džiovinimo dujų temperatūra 150°C, srautas 10 l/min., kolizijos energija 15V. 5fC analizei gradientas buvo: 0-24 min., 0% B; 24-25 min., 0-100% B; 25-29 min., 100 % B; 29,5-35 min., 0% B. Stebėta 256,1→140 m/z jonų tranzicija, naudota 300°C džiovinimo dujų temperatūra.

5mC analizei naudota 20-50 ng DNR, kalibravimui buvo sumaišyti žinomi nukleozidų standartų kiekiai. mESC 5hmC analizei naudota 0,5-2 µg DNR, 5hmC standartai sumaišyti su tiriamus mėginius atitinkančiais karpytos

lambda DNR kiekiais. 5hmC ir 5mC analizei naudotas gradientas: 0-6 min., 0% B; 6-18 min., 10 % B; 20-24 min., 100 % B; 25-33 min., 0% B, dujų temperatūra buvo 300°C. Registruotos jonų m/z tranzicijos: 5mC 242,1→126,1; 5hmC 258,1→142,1; G 268,1→152,1, C 228,1→112,1. 5fC fragmentacija buvo nustatyta 80 V, dhU ir kitiems nukleozidams 100 V.

5fC redukcijos į 5hmC įvertinimui naudota 2,5 µg serumo-2i *Tdg*-/- DNR. Grybų 5fC ir 5hmC po vitC poveikio įvertinimui analizuota ~2 µg DNR. Genominio 5hmC glikozilinimo analizei naudota 0,87 µg M.HhaI-hidroksimetilintos lambda DNR prieš ir po glikozilinimo. Piridino borano konversijos vertinimui registruota dhU m/z tranzicija 231,1→115,1, džiovinimo dujų temperatūra 150°C. Visi šie mėginiai buvo analizuoti be papildomo nukleozidų gryninimo frakcionuojant, analizei naudotas gradientas: 0-24 min, 0% B; 24-27 min., 0-100% B; 27-32 min., 100 % B; 33-38 min., 0% B.

RNR oksimC analizei, ~20-24 µg *C. cinerea* RNR ir 5 µg nemodifikuotos RNR sumaišytos su standartais kalibravimui, pirmiausia buvo išskirstytos į atskiras 5carC, 5hmrC ir 5frC nukleozidų frakcijas, kaip ir DNR atveju (tik pakoreguoti surinkimo laikai). 5mrC įvertinimui naudota nefrakcionuota RNR (20-50 ng). Registruotos nukleozidų m/z tranzicijos: rC 244,1→112, rG 284,1→152,1, 5mrC 258,1→126,1, 5carC 288,1→156, 5hmrC 274,1→142, 5frC 272,1→140. Kapiliaro jonizacijos įtampa 1800 V, džiovinimo dujų temperatūra 150°C, srauto greitis 10 l/min, kolizijos energija 15 V, celės greitintuvas 7 V. Fragmentacija 80 V 5frC ir 100 V kitiems nukleozidams, tirpalų gradientas buvo toks pat kaip atitinkamų DNR nukleozidų.

5carC derivatizacijos analizės HPLC gradientas: 0-6 min., 0% B; 6-22 min., 0-100% B; 22-26 min., 100% B; 26-27 min., 100-0% B; 27-35 min., 0% B; arba 0-6 min., 0% B; 6-13,5 min., 0-6% B; 13,5-21,5 min., 6-50% B; 22-26 min., 100% B; 26-27 min., 100-0% B; 27-35 min., 0% B. 5carC darinio m/z tranzicija 432,2→300,1, fragmentacija 80 V, kolizijos energija 8 V, celės greitintuvas 5V. Kiekviename eksperimente lygiagrečiai su mėginiais taip pat buvo atliekama derivatizacija su kontroline 5carC turinčia RNR, kad būtų atsižvelgta į galimą reakcijos efektyvumo variaciją. Tada naudojant MS/MS šioje RNR buvo įvertintas pradinis 5carC kiekis ir derivatizacijos efektyvumas. Ši RNR buvo sumaišyta su 5 µg nemodifikuotos RNR ir naudojama 5carC darinio kiekio kalibravimui. Apskaičiuota 5carC darinio MS/MS signalo ploto priklausomybė nuo pradinio 5carC kiekio. Signalai buvo kiekybiškai įvertinti jiems esant tiesiniame kalibravimo kreivės

diapazone ir tik aptikti kai signalo ir triukšmo santykis buvo  $\geq 3$ , bet signalas nepateko į kalibracinę kreivę.

Didelės skiriamosios gebos 5carC derivatizacijos produkto masės spektrai buvo gauti naudojant Q-TOF 6250 masės spektrometrą (*Agilent*), turintį *Dual-ESI* jonizacijos šaltinį.

### 2.2.11 caCLEAR bibliotekų paruošimas

Genominė DNR buvo fragmentuota naudojant *Covaris M220* iki vidutiniškai 200-250 bp fragmentų. mESC DNR po to dar buvo išgryninta naudojant *DNA Clean and Concentrator* rinkinį.

**1 žingsnis.** 5hmC taikiniai buvo apsaugoti glikozilinant  $\beta$ GT (žr. 2.2.6).

**2 žingsnis.** Nemodifikuoti genomo CG buvo apsaugomi atliekant wt M.SssI metilinimą (žr. 2.2.6). Po antrosios metilinimo reakcijos DNR buvo išgryninta *Oligo Clean and Concentrator* rinkiniu.

**3 žingsnis.** 5caC dekarboksilinimas buvo atliktas dekarboksilinimo buferyje 22°C 3 val. naudojant 10-ties kartų eM.SssI perteklių (skiedžiant iki 1,13  $\mu$ M koncentracijos) prieš CG taikinius.

**4 žingsnis.** Dekarboksilinimo reakcija buvo papildyta Ado-6-azido kofaktoriumi iki 0,2 mM ir inkubuojama 30°C 1 val., tada fermentas buvo išveiklintas, paveiktas ProtK ir DNR išgryninta (žr. 2.2.4).

**5 žingsnis.** TOP-seq.

**5a žingsnis.** DNR galai buvo sutvarkyti *DNA End Repair Kit* rinkiniu 15 min. 20°C pagal gamintojo rekomendacijas ir DNR išgryninta naudojant *GeneJet Purification Kit*. Tada DNR ( $\sim 1 \mu$ g į 50  $\mu$ l reakcijos) galuose buvo prijungtas A nukleotidas inkubuojant 37°C 45 min. su 0,5 mM dATP, 0,25 U/ $\mu$ l Klenow exo- ir jos buferiu. Po to fermentas inaktyvuotas 75°C 15 min. ir DNR išgryninta *GeneJet Purification Kit*. Iš dalies komplementarūs STE buferyje sulydyti A1/A2 adapteriai (2,25  $\mu$ M) buvo prijungti inkubuojant DNR ( $\sim 0,7 \mu$ g į 50  $\mu$ l) su 0,25 U/ $\mu$ l T4 DNR ligaze ir jos buferiu 22°C per naktį, po to termiškai išveiklinant 65°C 10 min. DNR išgryninta *DNA Clean and Concentrator* rinkiniu.

**5b žingsnis.** Išgryninta DNR buvo papildyta 20  $\mu$ M biotinilintu DNR alkino oligonukleotidu (ODN),  $\frac{1}{5}$  reakcijos tūrio CuBr (8 mM) ir THPTA (24 mM) DMSO ištirpintu mišiniu ir iki galutinės 50% koncentracijos buvo pridėta DMSO. Tada reakcijos mišinys buvo inkubuojamas 45°C 20 min. ir po to praskiedžiamas iki  $\sim 1\%$  DMSO prieš gryninimą per *GeneJET NGS Cleanup Kit* kolonėles. Tada DNR su prijungtu ODN buvo praturtinta: 0,1 mg *Dynabeads MyOne Streptavidin C1* buvo inkubuojama su DNR 10 mM Tris-

HCl pH 8,5 (~22°C), 1 M NaCl ir 0,2 % Tween 20 kambario temperatūroje 3 val., vartant. Dalelės buvo plaunamos kaip aprašyta (Song ir kt., 2017), naudojant 150 µl buferio. DNR buvo atkabinta daleles inkubuojant vandenyje 5 min. 95°C, 5 min. 4°C. 3.6 B, 3.7 A pav. pateiktiems eksperimentams buvo naudojamas oligonukleotidas be biotino ir nebuvo atliktas praturtinimas. Praturtinimui naudota ~50 ng lambda ir 827 ng mESC DNR.

**5c žingsnis.** Visa atkabinta DNR buvo papildyta 0,5 µM ODN komplementariu pradmeniu, 0,2 mM dNTP, 0,05 u/µl Pfu polimeraze, Pfu MgSO<sub>4</sub> buferiu, papildomu 1 mM MgSO<sub>4</sub> ir 20 µl reakcijoje buvo inkubuojama 5 ciklus: 95°C 1 min., 65°C 10 min., 72°C 10 min.

**5d žingsnis.** Visa DNR 100 µl reakcijoje padauginta naudojant *Platinum™ SuperFi™ PCR Master Mix* 94°C 4 min., 15 ciklų 95°C 1 min., 60°C 1 min., 72°C 1 min. Tuomet *MagJet NGS Cleanup and Size-selection* rinkiniu pagal gamintojo rekomendacijas išgryninti ~300 bp fragmentai ir atlikta *Ion Proton (Thermo Fisher Scientific, TFS)* sekoskaita. Ši ir kitos *Ion Proton* DNR bibliotekų sekoskaitos atliktos A. Rukšėnaitės.

### 2.2.12 Pelės genominių regionų 5caC analizė caMAB-seq metodu

Buvo paruošti du mėginiai: serumo-2i TET TKO DNR neigiama kontrolė ir 5caC turinti serumo-2i *Tdg*<sup>-/-</sup> DNR. Genominė DNR buvo išgryninta naudojant *Genomic DNA Clean and Concentrator* rinkinį, metilinta (žr. 2.2.6), išgryninta tuo pačiu rinkiniu ir eliuota vandeniu. 5fC redukcija atlikta remiantis Booth ir kt. (2014). 10 ng/µl DNR vandenyje buvo inkubuojama esant 250 mM NaBH<sub>4</sub> 1 val. kambario temperatūroje, dažnai maišant ir nusukant. Reakcija sustabdyta 0,5 tūrio 750 mM NaAc (pH 5). Kai nebesiskyrė dujos, buferis buvo pakeistas vandeniu naudojant *Microcon®* centrifuginius filtrus (*MERCK*). Kiekvienoje reakcijoje 500 ng DNR buvo veikiami natrio bisulfitu naudojant standartinį *EZ DNA Methylation-Gold™* rinkinio protokolą, išgryninta DNR buvo eliuojama 10 µl tūryje. Tuomet naudojant kiekvienai grandinei specifinius pradmenis (2.1.2) pasirinkti regionai 50 µl reakcijoje buvo padauginti 25 ciklus naudojant *Phusion U HS* polimerazę ir 1,3 µl DNR matricos. Reakcijos buvo išgrynintos naudojant *DNA Clean and Concentrator* rinkinį. Fragmentai buvo išgryninti iš poliakrilamidinio gelio naudojant *SYBR™ Gold* dažymą, susmulkintas gelis buvo inkubuojamas buferyje skyrimui iš gelio 2 val. 37°C, purtant 600 aps./min. Gelio dalelės nufiltruotos per *Ultrafree-MC* centrifuginį filtrą (*Millipore*) ir DNR buvo išgryninta naudojant *Ampure XP* magnetines daleles pagal gamintojo protokolą. Tuomet buvo sutvarkyti DNR galai (žr. 2.2.7). Atliktas *Ion Torrent™* platformai skirtų adapterių su barkodais (naudotas 10-

20× perteklius prieš DNR galus) prijungimas (žr. 2.2.7, be fosforilavimo etapo) ir DNR išgryninta *GeneJET NGS Cleanup* rinkiniu. Po 4 PGR ciklų naudojant Ion-PCR-Fw bei Ion-PCR-Rv pradmenis ir *Platinum SuperFi PCR Master Mix* PGR mišinį, DNR buvo išgryninta su *GeneJET NGS Cleanup* rinkiniu.

### 2.2.13 Pelės genominių regionų 5caC analizė piridino borano sekoskaita

5caC nustatymui buvo atliktas TAPS metodo variantas (pagal Liu ir kt., 2019), praleidžiant TET oksidacijos žingsnį (kai kurie etapai atlikti K. Skardžiūtės). Serumo-2i TET TKO (neigiama kontrolė) ir serumo-2i *Tdg*-/*mESC* genominė DNR buvo išgryninta naudojant *Genomic DNA Clean and Concentrator* rinkinį ir tuomet atlikta 5fC redukcija, kaip aprašyta caMAB-seq. Tada DNR (150 ng 50 µl tūriui) buvo inkubuojama su 1 M piridino boranu (*Sigma-Aldrich*) ir 0,6 M natrio acetatu (pH 4,3) 37°C 16 valandų, purtant 850 aps./min. Išgryninus DNR *Genomic DNA Clean ir Concentrator* rinkiniu, pasirinkti regionai buvo padauginti 16 ciklų su kiekvienam regionui specifiniais pradmenimis (2.1.2) naudojant *DreamTaq* DNR polimerazę. PGR produktai buvo išgryninti poliakrilamidinio gelio, kaip aprašyta 2.2.12. Po eliacijos į tirpalą buvo pridėta 0,05 µg/µL glikogeno ir DNR išsodinta etanoliu. DNR fragmentai buvo paruošti *Ion Torrent™* sekoskaitai, kaip aprašyta caMAB-seq 2.2.12.

### 2.2.14 Grybų auginimas

*Laccaria bicolor* (Maire) Orton S238N-H82 haploidinė monokariotinė padermė (NCBI:txid486041) buvo padovanota prof. Francis M. Martin. *Coprinopsis cinerea* haploidinė monokariotinė padermė Okayama 7 (NCBI: txid240176) įsigyta iš FGSC, #9003.

*C. cinerea* micelis buvo auginamas ant agarizuotos (1%) YMG terpės (Rao ir Niederpruem, 1969). Nukleorūgščių skyrimui į 300 ml skystos terpės 1 L kolboje buvo įdėtas nedidelis kiekis agarizuotoje terpėje auginto micelio ir auginta tamsoje be purtymo 30°C 14 dienų. TET aktyvumo sustiprinimui, micelis buvo augintas 7 dienas ir tada paveiktas 5 mM vitaminu C (vitC) 2 dienas (nebent nurodyta kitaip), neveikta kontrolė auginta tokį pat laiką tarpą.

*L. bicolor* micelis buvo auginamas 25°C tamsoje ant agarizuotos (2%) pakeistos Pachlewski terpės: 10 g/L gliukozės, 3 g/L salyklo ekstraktas, 0,25 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0,5 g/L KH<sub>2</sub>PO<sub>4</sub>, 0,15 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 0,05 g/L CaCl<sub>2</sub>, 0,025 g/L NaCl, 0,0012% FeCl<sub>3</sub>, 100 mg/L tiamino, 2,3 mg/L ZnSO<sub>4</sub>·7H<sub>2</sub>O, 8,5 mg/L H<sub>3</sub>BO<sub>3</sub>, 0,5 mg MnCl<sub>2</sub>·4H<sub>2</sub>O, 0,6 mg/L CuSO<sub>4</sub>·5H<sub>2</sub>O, 0,4 mg/L

$(\text{NH}_4)_2\text{MoO}_4$  (pagal Robin ir kt., 2011). Nukleorūgščių skyrimui micelis buvo auginamas 25°C skystoje terpėje įdėjus nedidelį kiekį micelio nuo agarizuotos terpės ir surinktas po 8 savaitių. vitC (5 mM) buvo veikta 2 paskutines dienas.

### 2.2.15 Grybų DNR gryninimas

Grybų micelis buvo susmulkintas iki miltelių su žiupsneliu polivinilpirolidono (PVP, K25), naudojant grūstuvę ir skystą azotą. Milteliai (~1 ml) buvo inkubuojami 55°C 90 min. 4,5 ml lizės buferio (100 mM Tris-HCl pH 8,0 (~22°C), 20 mM EDTA, 2 M NaCl, 2 % heksadeciltrimetilamonio bromidas (CTAB), 2 % PVP) su Proteinaze K. Tada DNR buvo 2 kartus išgryninta chloroformu ir išsodinta 1,5 tūrio izopropanolio ir 0,4 tūrio 3 M natrio acetato (pH 5,2) 30 min. 4°C. DNR buvo ištirpinta 10 mM Tris-HCl ir paveikta ribonukleazėmis A/T1 1,5 val. 37°C, o po to išgryninta chloroformu ir išsodinta etanoliu. DNR kokybė buvo patikrinta agarozės elektroforeze, koncentracija nustatyta *Qubit™ dsDNA HS Assay Kit* rinkiniu.

### 2.2.16 Grybų RNR gryninimas

Micelis pirmiausia buvo susmulkintas kaip ir DNR skyrimui. *C. cinerea* RNR taškiniam imunoblotui ir *L. bicolor* RNR buvo išskirta naudojant *RNAzol® RT* reagentą pagal gamintojo rekomendacijas.

Kitiems eksperimentams visuminė *C. cinerea* RNR buvo išskirta modifikuotu CTAB metodu (Peter ir kt., 2006). Sutrintas micelis (~100-500 µl) 10 min. inkubuotas 65°C temperatūroje 1 ml lizės buferio (2 % CTAB, 2 % PVP K25, 100 mM Tris-HCl pH 8,0 (~22°C), 25 mM EDTA pH 8, 2 M NaCl) purtant 550 aps./min. Tada RNR buvo išgryninta chloroformu ir išsodinta etanoliu su natrio acetatu. Taškiniam imunoblotui RNR paveikta DNaze I pagal gamintojo rekomendacijas. RNR kokybė buvo patikrinta agarozės arba kapiliarine elektroforeze (*Agilent RNA 600 Nano Kit*), koncentracija nustatyta *Qubit® RNA HS Assay Kit* rinkiniu.

### 2.2.17 Kontrolinės RNR paruošimas

Kontrolinė RNR buvo susintetinta *in vitro* transkripcijos būdu pagal gamintojo rekomendacijas, naudojant *TranscriptAid T7 High Yield Transcription Kit* rinkinį. DNR matricai buvo naudojama sukirpta pUC18 plazmidė (gauta iš dr. Miglės Tomkuvienės), turinti sR47 geną (*P. Abyssis*; Nolivos ir kt., 2005). 5carC turinčios sR47 RNR sintezei reakcijoje buvo 25% 5caCTP nuo viso CTP. Po poveikio DNaze I pagal gamintojo rekomendacijas,

RNR buvo išgryninta naudojant *RNA clean and concentrator* rinkinį. 5caC kiekiai kontrolinėje RNR buvo įvertinti naudojant HPLC-MS/MS.

### 2.2.18 5caC analizė taškiniu imunoblotu

Kontrolei buvo naudojama nemodifikuota ir 5caC turinti sr47 RNR. Teigiamai DNR kontrolei buvo naudojamas 5caC turintis oligonukleotidas pTAA-5caC, kuris buvo sumaišytas su  $\lambda$  fago DNR iki reikiamo galutinio kiekio (ng), kaip ir mėginiuose. Lambda DNR taip pat buvo naudojama kaip neigiama kontrolė. RNR buvo denatūruota vandenyje 70°C 10 min., DNR – 95°C 10 min. Mėginiai buvo užnešti ant *Hybond<sup>TM</sup>-NX* membranos (*GE Healthcare Life Sciences Amersham<sup>TM</sup>*), išdžiovinti ir pritvirtinti prie membranos veikiant UV. Membrana buvo nudažyta 0,04 % metileno mėliu 0,3 M natrio acetate (pH 5,2) 5 min., praskalauta vandeniui, nufotografuota ir 5 min. nuplauta PBS buferiu (137 mM NaCl, 2,7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>) su 0,1 % Tween 20.

Membrana buvo užblokuota inkubuoiant TBS buferyje (50 mM Tris-HCl pH 7,4 (~22°C), 150 mM NaCl) su 2% lieso pieno 1 val. kambario temperatūroje. Po to buvo inkubuojama per naktį ~6°C su 4 ml TBS buferio, 2% pieno, 0,1% Tween 20 ir 1,5 µg antikūnų prieš 5caC. Tuomet membrana buvo plaunama TBS buferiu su 0,1 % Tween 20 3 kartus po 5 min. Tada į 4 ml buferį su TBS, 2 % pieno ir 0,1 % Tween 20 buvo pridėta 1,6 µl antrinių ožkos antikūnų su HRP (prieš triušio IgG) ir inkubuota 1 val. kambario temperatūroje. Tuomet 3 kartus po 5 min. buvo atliktas plovimas TBS buferiu su 0,1% Tween 20, 2 kartus plauta su vandeniui ir membrana buvo nudažyta chromogeniniu substratu TMB.

### 2.2.19 5caC turinčios RNR cheminė derivatizacija

5caC turinčios RNR derivatizavimas buvo pagrįstas Lu ir kt. (2013). (4-aminometil)benzilazido hidrochloridas susintetintas prof. Viktoro Masevičiaus ir dr. Gražinos Petraitytės (žr. Ličytė ir kt., 2022).

Prieš reakciją RNR buvo denatūruojama vandenyje 70°C 10 min. ir atšaldoma ant ledo. Derivatizavimas buvo atliktas dviem etapais. Siekiant pirmajame etape optimizuoti reakciją (atliko Eglė Godliauskaitė), 5caC turinti sr47 RNR buvo inkubuojama 50 µl reakcijoje su 10 mM MES pH 6 (~22°C), 20 mM NHS (N-hidroksisukcinimidas) su 4 skirtingais kitų reagentų deriniais: a) 2 mM EDC (1-etil-3-(3-dimetilaminopropil)-karbodiimido hidrochloridas); b) 8 mM EDC; c) 2 mM EDC, 45 % DMSO; d) 8 mM EDC, 45 % DMSO. NHS ir EDC buvo šviežiai ištirpinami, sumaišomi tarpusavyje



ir į reakciją pridedami paskutiniai. Pirmajame etape reakcija buvo inkubuojama 30 min. 37°C temperatūroje, purtant 550 aps./min. Antrame etape į galutinį 150 µl tūrį buvo pridėta 10 mM (4-aminometil)benzilazido hidroklorido, 100 mM natrio fosfato pH 7,5, 150 mM NaCl ir inkubuojama 1 val. 37°C, purtant 550 aps./min. Reakcijos buvo 2,5 karto praskiestos vandeniui, o RNR išgrynintos *RNA clean and concentrator* rinkiniu.

Tolesniam grybų RNR derivatizavimui pirmajame etape buvo naudojamos d) sąlygos. Pirmajame etape 0,7-25 µg RNR buvo inkubuota 60 µl tūryje, o antrajame etape galutiniame 180 µl tūryje.

### 2.2.20 Grybų RNR gryninimas į atskiras frakcijas

RNR buvo išgryninta į ilgų (>200 nt) ir trumpų (<200 nt) RNR frakcijas naudojant *RNA clean and concentrator* kit rinkinį. Poli A RNR buvo išgryninta iš ilgų RNR frakcijos (100-106 µg), naudojant iRNR skyrimo rinkinį *Dynabeads™ mRNA Purification Kit*. Ilgosios (28S, 18S) ribosominės RNR buvo išgrynintos iš žemos lydymosi temperatūros agarozės gelio, veikiant agaraze pagal gamintojo rekomendacijas. Trumpų RNR frakcijos buvo išpjautos po denatūruojančio poliakrilamidinio gelio elektroforezės, nudažius etidžio bromidu. RNR išplautos iš gelio kaip aprašyta 2.2.12. Mėginiai buvo filtruojami per *Ultrafree-MC* centrifuginį filtrą (*Millipore*) ir RNR išsodinta naudojant etanolį.

### 2.2.21 5hmC ir 5fC praturtinimo įvertinimas naudojant modelinius DNR fragmentus

188 bp modeliniai DNR fragmentai, turintys vieną 5hmCG arba 5fCG taikinį vienoje grandinėje, buvo paruošti PGR būdu kaip nurodyta 2.2.2. Kitoje grandinėje esantis CG buvo metilintas naudojant wt M.SssI pagal gamintojo protokolą. Modeliniai DNR fragmentai buvo sumaišyti su ultragarsu fragmentuota *L. bicolor* DNR santykiu 13:1287 ng, imituojant nedidelį modifikacijų gausumą.

5hmC ir 5fC praturtinimui naudojome modifikuotus hMe-Seal ir fC-Seal protokolus (Song ir kt., 2011, 2013). Norint nustatyti 5fC, buvo atlikta jo redukcija į 5hmC naudojant NaBH<sub>4</sub> (žr. 2.2.12). Siekiant natūralioje genominėje DNR 5fC atskirti nuo 5hmC, 5hmC buvo apsaugotas glikozilinant paprasta gliukoze. 5hmC apsaugojimo įvertinimui prieš žymėjimą buvo atliktas 5hmC/5mC fragmento glikozilinimas su nemodifikuota UDP-gliukoze (žr. 2.2.6). 5hmC, redukuotą 5fC arba apsaugotą 5hmC turintys DNR fragmentai buvo pažymėti naudojant 50 µM UDP-6-azido-gliukozę ir T4

$\beta$ GT, 50  $\mu$ l reakcijoje naudojant 10 U fermento ir 1  $\mu$ g DNR. Reakcija buvo inkubuojama 2 val. 37°C, po to 65°C 20 min. atliktas fermento išveiklinimas ir DNR išgryninta naudojant *DNA Clean and Concentrator* rinkinį.

Tada iki 0,4 mM koncentracijos pridėta DBCO-S-S-PEG3-biotino (*Jena Bioscience*) ir buvo inkubuojama 10 mM Tris-HCl pH 8,5 (~22°C) tirpale 37°C 2 val. DNR buvo išgryninta *DNA Clean and Concentrator* rinkiniu. Tada DNR buvo praturtinta kaip aprašyta 2.2.4, tik inkubavimo ir plovimo etapuose papildomai dedant 0,01 % Tween 20. Dalelių ir supernatanto frakcijose esantis DNR kiekis buvo įvertintas kPGR.

### 2.2.22 hMe-Seal ir fC-Seal bibliotekų paruošimas

Genominė DNR buvo fragmentuota ultragarsu *Bioruptor UCD-200* (*Diagenode*) instrumentu 10 mM Tris-HCl pH 8,5 (~22°C) tirpale iki vidutiniškai 200-250 bp fragmentų ir išgryninta naudojant *DNA Clean and Concentrator kit*.

5hmC apsaugojimas ir 5fC redukcija į 5hmC buvo atlikta kaip aprašyta 2.2.6 ir 2.2.12. Norint pasiekti panašų 5fC ir 5hmC mėginių paruošimą, abiem mėginiam buvo atlikti tie patys gryninimo etapai, tik 5hmC mėginiam nebuvo atlikta glikozilinimo su nemodifikuota UDP-gliukoze ir redukcijos NaBH<sub>4</sub>. Naudojant UDP-6-azido gliukozę ir T4  $\beta$ GT (žr. 2.2.21) buvo pažymėta 2,6-2,7  $\mu$ g DNR. Taip pat buvo paruošti du kontroliniai DNR mėginiai: 5hmC kontroliniams mėginiam žymėjimo etape buvo nedėta  $\beta$ GT; 5hmC apsaugojimo įvertinimui 5fC mėginiuose, buvo naudojama glikozilinta DNR be redukcijos žingsnio.

Biotino prijungimo etapas buvo atliktas 130  $\mu$ l reakcijoje (kaip aprašyta 2.2.21), o po to DNR galai sutvarkyti naudojant *DNA End Repair Kit* (žr. 2.2.11 5a) ir DNR išgryninta naudojant *DNA Clean and Concentrator* rinkinį. Tada biotinita DNR buvo praturtinta kaip aprašyta 2.2.21, 40  $\mu$ l tūryje su 0,1 mg dalelių *Dynabeads MyOne Streptavidin C1*. Tuomet, po inkubacijos su 20  $\mu$ l 50 mM DTT, 60 mM Tris pH 7,8 (~22°C) 1 val. variant, DNR buvo atkabinta ir atskirta nuo dalelių.

Prie DNR buvo prijungta 4,2 pmol *Ion Torrent*<sup>TM</sup> platformos adapterių inkubuojant per naktį 22°C 30  $\mu$ l reakcijoje su 0,5 mM ATP, 10 mM MgCl<sub>2</sub>, 5% PEG 4000 ir 7,5 U T4 ligazės. Po to adapteriai buvo pašalinti naudojant *MagJet NGS Cleanup and Size-selection* rinkinį. Tada DNR buvo inkubuojama su 10 mM 2-merkoptoetanoliu 10 min. kambario temperatūroje (kad nesusidarytų disulfidiniai kryžminiai ryšiai). Po to 50  $\mu$ l reakcijoje su Taq (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> buferiu, 0,2 mM dNTP, 2 mM MgCl<sub>2</sub>, 1  $\mu$ M Ion-PCR-Fw ir Ion-PCR-Rv pradmenų ir 2,5 U Taq DNR polimerazės atliktas PGR. PGR

sąlygos buvo: 1 min. 50°C, 5 min. 72°C, 4 min. 94°C, 1 min. 65°C, 5 min. 72°C, 15 ciklų po 1 min. 94°C, 1 min. 65°C, 1 min. 72°C ir 72°C 5 min. *MagJet NGS Cleanup and Size-selection* rinkiniu buvo išgryninti ~300 bp fragmentai ir atlikta *Ion Proton* (TFS) sekoskaita.

### 2.2.23 hmTOP-seq ir foTOP-seq bibliotekų paruošimas

DNR buvo sufragmentuota ultragarsu iki vidutiniškai 200-250 bp fragmentų su *Covaris M220* prietaisu ir išgryninta naudojant *DNA Clean and Concentrator* rinkinį. 5fC redukcija į 5hmC, 5hmC apsaugojimas ir 5hmC žymėjimas azido grupe (5,8 µg DNR šiame etape) buvo atlikti kaip aprašyta hMe-Seal ir fC-Seal bibliotekoms (2.2.22). Tada buvo sutvarkyti DNR galai, prijungtas dA ir adapteriai kaip aprašyta 2.2.11 5a. Taip pat buvo paruošti atitinkami kontroliniai mėginiai kaip aprašyta 2.2.22. Buvo atliktas TOP-seq bibliotekų praturtinimas ir paruošimas kaip nurodyta 2.2.11 5b-d ir atlikta *Ion Proton* (TFS) sekoskaita.

### 2.2.24 RNR bibliotekų paruošimas

*L. bicolor* RNR buvo išskirta naudojant *RNAzol® RT* reagentą ir paveikta DNaze I pagal gamintojo rekomendacijas. Tuomet *RNA clean and concentrator* rinkiniu išgryninta >200 nt frakcija ir pagal gamintojo rekomendacijas buvo pašalintos ribosominės RNR naudojant *Ribo-Zero Gold* rinkinį. *RNA clean and concentrator* rinkiniu dar kartą išgryninta >200 nt frakcija. Tuomet naudojant *Collibri Stranded RNA Library Prep Kit* (TFS) rinkinį skirtą *Illumina* platformai *Thermo Fisher Scientific Baltics* buvo paruoštos RNR bibliotekos ir atlikta sekoskaita (~200 mln. nuskaitymų iš abiejų galų vienam mėginiui).

### 2.2.25 Sekoskaitos duomenų analizė

Visus su caCLEAR susijusius sekoskaitos duomenis išanalizavo dr. Povilas Gibas, kaip aprašyta Ličytė ir kt. (2020). Didelio patikimumo mESC 5caCG taikinių nustatymui visi TET TKO bibliotekose aptikti CG buvo pašalinti iš tikslinių bibliotekų. Tikslinėse bibliotekose buvo naudojami tik tie CG, kurie turėjo didesnę nei vidutinę arba vidutinį padengimą (5× serumo-2i *Tdg*<sup>-/-</sup>, 6× serumo *Tdg*<sup>-/-</sup>, 4× serumo-2i wt ir 2× serumo wt bibliotekoms). Wt mėginiuose buvo palikti tik tie CG, kurie taip pat buvo nustatyti atitinkamų sąlygų *Tdg*<sup>-/-</sup> mėginiuose. Taip pat pašalinta 10% CG taikinių, kurie turėjo didžiausią padengimo skirtumą tarp techninių replikų.

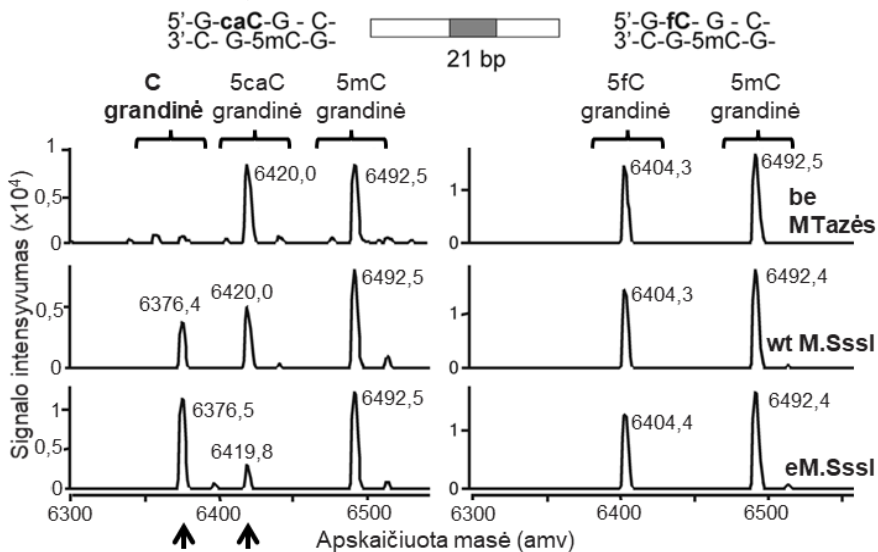
Dr. Povilas Gibas ir Vita Tomkutė atliko pradinę grybų hMe-Seal ir fC-Seal sekoskaitos bioinformatinę analizę, kurią pabaigė Kotryna Kvederavičiūtė. Visą tolesnę grybų sekoskaitos duomenų analizę atliko K. Kvederavičiūtė, kaip aprašyta Ličytė ir kt. (2022). hMe-Seal ir fC-Seal analizei genomas buvo padalintas į nepersidengiančius 200 bp regionus, kuriuose mėginių signalas buvo skaičiuotas jei padengimas buvo 1,2 karto didesnis nei kontrolinių mėginių. hmTOP-seq/foTOP-seq duomenims apjungus technines replikas naudoti bent 2× padengimą turintys CG, o ir kontroliniuose mėginiuose aptikti CG nebuvo naudojami.

Šio darbo sekoskaitos duomenys pasiekiami internete NCBI GEO GSE142319 ir GSE185551.

### 3. REZULTATAI

#### 3.1 DNR citozino-5 metiltransferazių atliekama 5caC dekarboksilinimo reakcija

Fermentinis citozino metilinimas vyksta metiltransferazėms atliekant metilo grupės perkėlimą nuo SAM kofaktoriaus ant C5 pozicijos dalyvaujant kataliniam cisteinui (Kriukienė ir kt., 2012). Aktyvus DNR demetilinimas žinduoliuose vyksta TET fermentams oksiduojant 5mC ir susidarant 5-hidroksimetilcitozinui (5hmC), 5-formilcitozinui (5fC) ir 5-karboksilcitozinui (5caC), o tuomet 5fC bei 5caC bazės gali būti fermentiškai pašalinamos timino DNR glikozilazės (TDG) ir bazių iškirpimo reparacijos sistema (BER) grąžina nemodifikuotą citozino būseną (He ir kt., 2011; Ito ir kt., 2011; Maiti ir Drohat, 2011; Pfaffeneder ir kt., 2011; Tahiliani ir kt., 2009). Iškelta hipotezė, kad citozino modifikacijos galėtų būti tiesiogiai pašalinamos specialiais fermentais skeliant C-C ryšį (Pfaffeneder ir kt., 2011; Wu ir Zhang, 2010), o inkubuojant DNR su pelės embrioninių kamieninių ląstelių ekstraktais buvo pastebėtas tiesioginis 5caC dekarboksilinimas į nemodifikuotą citoziną (Schiesser ir kt., 2012).

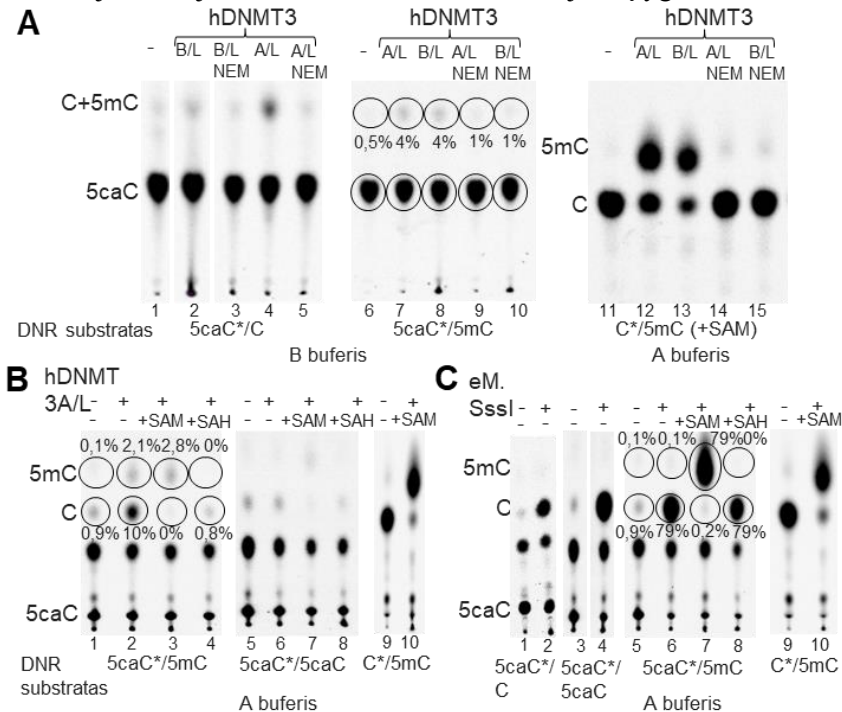


**3.1 pav.** Bakterinės metiltransferazės SssI 5caC dekarboksilinimo aktyvumas. 21 bp ilgio hemimetilintas DNR dupleksas kitoje taikinio grandinėje turintis 5caC arba 5fC buvo veiktas mutantine (e, ji turi išplėstą kofaktoriaus prijungimo kišenę) arba laukinio tipo (wt) M.SssI. DNR grandinės analizuotos Q-TOF masių spektrometru neigiamoje jonizacijoje, apskaičiuotos jų masės atominiais masės vienetais (amv). Po poveikio SssI sumažėja 5caC grandinės signalas ir atsiranda (pažymėta rodyklėmis) nemodifikuotos grandinės signalas (C grandinė), o tai rodo metiltransferazės dekarboksilinimo aktyvumą. Pagal Liutkevičiūtė ir kt. (2014).

Buvo parodyta, kad bakterinės (Liutkevičiūtė ir kt., 2009) ir žmogaus (Chen ir kt., 2012) metiltransferazės geba pašalinti 5-hidroksimetil grupę susidarant nemodifikuotam citozinui. Siekiant geriau suprasti fermentinį C-C ryšio skėlimą ir galimus aktyvaus demetilavimo kelius, DNR modifikacijos tyrimų skyriuje buvo tiriamas metiltransferazių gebėjimas pašalinti DNR citozino-C5 modifikacijas. Buvo nustatyta, kad nesant kofaktoriaus SAM, bakterinės metiltransferazės M.HhaI ir M.SssI (3.1 pav.) geba atlikti ne tik 5hmC dehidroksimetilinimą, bet ir 5caC dekarboksilinimą susidarant nemodifikuotam citozinui, o tuo tarpu 5fC ir 5mC išlieka nepakitę (Liutkevičiūtė ir kt., 2014).

Žmogaus metiltransferazių DNMT3A/DNMT3L kompleksas taip pat rodė aiškų 5caC dekarboksilinimo aktyvumą, DNMT3B/DNMT3L aktyvumas buvo silpnesnis, bet aptinkamas (Liutkevičiūtė ir kt., 2014). Šių tyrimų kontekste siekėme įsitikinti, kad žmogaus DNMT3 baltymų preparate 5caC dekarboksilinimo reakciją atlieka būtent DNR metiltransferazės. Tyrėme metiltransferazių aktyvumą naudojant 31 bp DNR, kurios vienoje grandinėje buvo <sup>33</sup>P žymėtas citidino ar 5-karboksilcitidino nukleotidas CG taikinyje (žr. 2.2.1, 2.1.2), o tai leido plonasluoksnės chromatografijos būdu selektyviai stebėti taikinio nukleotido kovalentinius pokyčius. Veikėme žmogaus metiltransferazės *N*-etilmaleimidu, kuris reaguoja su tioliais, todėl gali reaguoti su katalitiniu cisteinu. Kadangi po šio poveikio stebimas žmogaus DNMT3A/3L ir DNMT3B/3L 5caC dekarboksilinimo aktyvumas buvo panaikintas (3.2 pav. A), todėl dekarboksilinimo reakcijai yra svarbus metiltransferazių katalitinis cisteinas ir labiausiai tikėtina, kad reakciją DNMT3 baltymų preparate atlieka metiltransferazė. Atitinkamai metiltransferazinis aktyvumas buvo prarastas ir kontrolinėje metilinimo reakcijoje. Eksperimentai su bakterijų metiltransferazėmis, kuriose pakeistas katalitinis cisteinas taip pat rodė 5caC dekarboksilinimo aktyvumo praradimą (Liutkevičiūtė ir kt., 2014). Taip pat siekėme nustatyti, ar 5caC dekarboksilinimo reakcija galėtų vykti esant fiziologinėms SAM koncentracijoms (30 μM; Clarke ir Banfield, 2001). Esant reakcijoje SAM, žmogaus DNMT3A/3L 5caC dekarboksilinimo aktyvumas žymiai sumažėjo (3.2 pav. B), todėl mažai tikėtina, jog šis kompleksas vykdo dekarboksilinimą įprastinėmis sąlygomis *in vivo*. 5caC dekarboksilinimo reakcija taip pat buvo žymiai slopinama ir SAM produkto S-adenozil-homocisteino (SAH), o toks specifinis reakcijos slopinimas kofaktoriumi ir jo produktu dar kartą parodo, kad 5caC dekarboksilinimą vykdo metiltransferazė (3.2 pav. B). Tuo tarpu bakterinės metiltransferazės eM.SssI vykdomos 5caC dekarboksilinimo reakcijos SAM ar SAH neslopino bei ji dekarboksilinimą efektyviai vykdė

esant įvairiems substratams: 5caC vienoje grandinėje, o nemodifikuotam ar metilintam citozinui kitoje, ar esant 5caC abiejose grandinėse (3.2 pav. C). Stebima netipinė metiltransferazių vykdoma C-C ryšio skėlimo reakcija parodo įmanomą fermentinį tiesioginės 5caC konversijos į C kelią veikiant DNMT, jų kompleksams su kitais baltymais ar dar nežinomiems specialioms fermentams *in vivo*. Taip pat metiltransferazių vykdomas 5caC vertimas į nemodifikuotą citoziną siūlo unikalią galimą 5caC nustatymo būdą genominės DNR sekoje naudojant švelnias fermentines reakcijos sąlygas.



**3.2 pav.** Žmogaus (h) DNMT3A ir 3B bei bakterijų eM.SssI DNR citozino-5 metiltransferazių 5caC dekarboksilinimo aktyvumas. 31 bp ilgio DNR dupleksas turintis <sup>33</sup>P žymėtą (pažymėta žvaigždute) 5caC arba C nukleotidą vienoje CG taikinio grandinėje ir C, 5mC arba 5caC kitoje (parodyta kaip 5caC\*/C, 5caC\*/5mC, C\*/5mC, 5caC\*/5caC) buvo veiktas bakterijų metiltransferaze eM.SssI arba žmogaus metiltransferazėmis DNMT3 A arba B komplekse su L. DNR buvo sukarpyta iki nukleotidų ir analizuota plonasluoksne chromatografija A arba B buferyje (žr. 2.2.1), <sup>33</sup>P žymėtas taikinio nukleotidas ir jo produktai vizualizuoti autoradiografija. Densitometrinis taškų intensyvumo įvertinimas parodytas apskritimais (% nuo signalų sumos takelyje). (A) DNMT3A/L ir DNMT3B/L rodė aptinkamą 5caC dekarboksilinimo aktyvumą, kuris buvo panaikintas prieš tai fermentus paveikus su cisteinu reaguojančia medžiaga (20 mM *N*-etilmaleimidu (NEM), 1 h, 25 °C). Atitinkamai beveik nebevyko ir nuo SAM priklausomas citozino metilinimas. (B) hDNMT3A/L 5caC dekarboksilinimo aktyvumas buvo žymiai slopinamas esant kofaktoriui SAM ar jo produktui SAH (abu 30 μM). Tuo tarpu toks slopinimas nebuvo stebimas eM.SssI vykdomoje 5caC dekarboksilinimo reakcijoje (C).

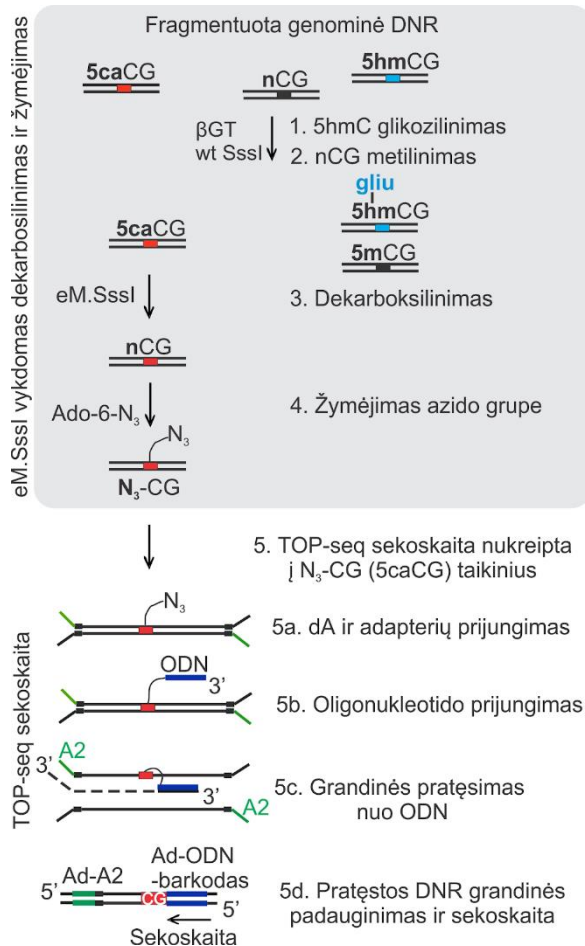
### 3.2 5caC nustatymo metodo, caCLEAR, kūrimas ir įvertinimas

Oksiduotų 5-metilcitozino formų (oksi-mC) biologinės funkcijos nėra iki galo suprastos, tačiau yra įrodymų, kad jos gali būti ne vien demetilinimo kelio tarpiniai produktai, bet ir turėti epigenetines funkcijas. Oksi-mC biologinei svarbai ir DNR demetilinimui geriau iširti reikalingi jautrūs didelės skiriamosios gebos modifikacijų genominio profiliavimo metodai. Žinduolių genome 5caC yra rečiausia modifikacija iš oksii-mC, o tai apsunkina jos funkcijų tyrimus. 5caC tyrimams sukurti ir patvirtinti didelės skiriamosios gebos metodai naudoja bisulfitinę sekoskaitą (Lu ir kt., 2015b; Wu ir kt., 2014; Neri ir kt., 2015) kartu su papildomu fermentiniu ar cheminiu veikimu. Tačiau kai kurie metodai nuskaito 5caC ir 5fC kaip vieną signalą arba atliekamas modifikuoto 5caC ir paprastos BS-Seq mėginio palyginimas (kad 5caC būtų atskirtas nuo kitų citozino formų), o tai reikalauja dvigubos sekoskaitos, reiškančios techninį ir ekonominį iššūkį genomo analizei. Be to, dėl 5caC retumo tokiuose sekoskaitos metoduose didžioji dalis gaunamų sekų yra neinformatyvios.

Mūsų tikslas buvo sukurti vieno nukleotido skiriamosios gebos metodą 5caC nustatymui genome, išvengiant bisulfitinės sekoskaitos, keliančios technines ir bioinformatinės analizės kliūtis. Metiltransferazių vykdoma 5caC dekarboksilinimo reakcija vyksta švelniomis sąlygomis, yra specifiskumas nukleotido modifikacijai: nors jos panašiai aktyvios su 5hmC, bet ne su 5fC ir 5mC (3.1 pav.; Liutkevičiūtė ir kt., 2014). Todėl nusprendėme pabandyti pritaikyti šią metiltransferazių reakciją 5caC nustatymui genome. Mutantinė bakterinė metiltransferazė eM.SssI su išplėsta kofaktoriaus prijungimo kišene patogi šiam pritaikymui, nes atpažįsta CG dinukleotidą, kuriame citozino modifikacijos daugelyje organizmų yra aptinkamos dažniausiai (Zemach ir kt., 2010) bei 5caC dekarboksilinimo reakciją vykdo efektyviau nei laukinio tipo M.SssI (wt, angl. *wild type*) (3.1 pav.), o tuo tarpu 5hmC dehidrosimetilinimą vykdo silpniau (Liutkevičiūtė ir kt., 2014). Be to, eM.SssI geba perkelti azido grupę nuo sintetinio kofaktoriaus Ado-6-N<sub>3</sub> ant nmodifikuotų CG (Kriukienė ir kt., 2013). Apjungus šias dvi eM.SssI vykdomas reakcijas – 5caC dekarboksilinimą sukuriantį nmodifikuotus CG (nCG) ir nCG žymėjimą, buvusius 5caC taikinius galima nustatyti prie azido grupės prijungto pradmens nukreipta sekoskaita – TOP-seq (Staševskij ir kt., 2017; žr. 1.4.2). Tuo tarpu siekiant apsaugoti 5hmC ir natūralius nCG nuo žymėjimo, prieš tai atitinkamai galima atlikti jų glikozilinimą ir metilinimą. Toks sekoskaitos būdas leistų nustatyti 5caC nukleotido tikslumu ir nuskaitant tik 5caC turinčias sekas. Kuriamą metodą pavadino caCLEAR (angl. *5caC*



*Clearance*), jo schema pateikta 3.3 pav.. Pirmiausia siekėme įvertinti atskirų žingsnių ir viso siūlomo metodo efektyvumą ir specifiškumą modelinėse sistemose.



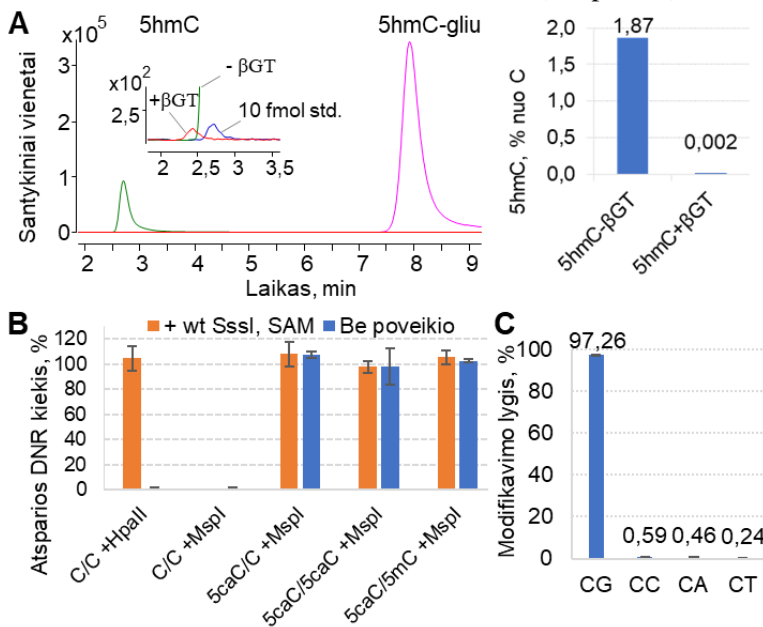
**3.3 pav.** Kuriamo 5caC nustatymo metodo caCLEAR schema. 1. 5hmC modifikavimas gliukoze (gliu) naudojant T4 bakteriofago  $\beta$ -gliukoziltransferazę ( $\beta$ GT) ir kofaktorių UDP-gliukozę. 2. Nemodifikuotų CG (nCG) metilinimas naudojant CG-specifinę laukinio tipo (wt) metiltransferazę SssI ir SAM. 3. 5caC dekarboksilinimas ir naujų nemodifikuotų CG sukūrimas naudojant mutantinę metiltransferazę eM.SssI. 4. Kovalentinis naujai sukurtų nCG žymėjimas azido grupe esant eM.SssI ir Ado-6-N<sub>3</sub>. 5. Azido grupe žymėtų CG taikinių profiliavimas didelio našumo prijungto pradmens nukreipta sekoskaita TOP-seq. ODN – oligodeoksiribonukleotidas, A2 – viena dalinai komplementaraus adapterio grandinė.

### 3.2.1 Nemodifikuotų CG ir 5hmC modifikavimo įvertinimas

Kadangi eM.SssI taip pat geba pašalinti 5-hidroksimetil grupę nuo 5hmC, šią modifikaciją nusprendėme apsaugoti nuo žymėjimo (3.3. pav., 1

žingsnis) naudojant  $\beta$ -gliukoziltransferazės ( $\beta$ GT) atliekamą 5hmC modifikavimą gliukoze (Song ir kt., 2011). O kad atskirtume genominėje DNR natūraliai esančius nCG nuo naujų nCG, įvestų po 5caC dekarboksilinimo, genominius nCG metilnometilino laukinio tipo M.SssI esant SAM (3.3. pav., 2 žingsnis). Kad būtų pasiektas visiškas 5hmC ir nCG apsaugojimas, fermentais DNR veikėme po du kartus (žr. 2.2.6), optimizavome reakcijų sąlygas modelinėse DNR sistemose bei genominėje DNR ir vertinome jų efektyvumą įvairiais metodais.

Naudojant lambda bakteriofago genomine DNR, kurios GCGC taikiniuose įvedėme 5hmC (naudojant M.HhaI hidroksimetilinimą, žr. 2.2.2), vertinome glikozilinimo efektyvumą kiekybine aukšto slėgio skysčių chromatografija-tandemine masių spektrometrija (dalį HPLC-MS/MS atliko ir sąlygas visur parinko Audronė Rukšėnaitė). Stebėjome efektyvų glikozilinimą (~99,89%), nes lyginant su pradiniu kiekiu po reakcijos aptikome tik ~0,11 % likusio nmodifikuoto 5hmC (3.4 pav. A).



**3.4 pav.** caCLEAR metodo 5hmC ir nmodifikuotų CG apsaugojimo žingsnių įvertinimas su modeliniais DNR fragmentais ir modeline genomine DNR. (A) Lambda bakteriofago genomine DNR su 5hmC GCGC taikiniuose buvo inkubuota su T4 bakteriofago  $\beta$ -gliukoziltransferaze ( $\beta$ GT) ir UDP-gliukoze optimizuotomis sąlygomis (žr. 2.2.6). Kairėje,  $\beta$ GT glikozilinimo efektyvumas įvertintas analizuojant reakcijos produktus HPLC-MS/MS. Nukleozidų masių chromatogramos rodo efektyvų 5hmC (žalia spalva rodo lygį prieš reakciją, raudona – po) pavertimą į glikozilintą produktą (5hmC-gliu, rožinė). 5hmC standartas parodytas mėlyna spalva. Dešinėje, 5hmC santykis su citozinu prieš glikozilinimą ir po jo. (B) Laukinio tipo (wt) M.SssI metilinimo įvertinimas kPGR naudojant modelinius DNR fragmentus,

turinčius vieną R.MspI/HpaII taikinį (CCGG). C/C ir pan. – nurodyta CG taikinio citozino modifikavimo būseną DNR grandinėse. Buvo stebimas įsotintas nCG metilinimas, kadangi DNR fragmentai buvo atsparūs R.HpaII kirpimui. Metilinimo metu karboksilo grupės pašalinimas C5caCGG taikiniuose nebuvo stebimas, nes įvairiai CCGG taikiniuose modifikuoti DNR fragmentai liko atsparūs R.MspI kirpimui ir po metilinimo. (C) caCLEAR metilinimo žingsnio įvertinimas (įvairiuose C dinukleotiduose) atliekant bisulfitinę lambda bakteriofago genomo sekoskaitą po metilinimo wt SssI.

DNR modifikavimo būseną analizavome naudojant restrikcijos endonukleazes, jautrias jų keramos sekos CG taikinio modifikacijoms: R.HpaII kerpa tik esant nemodifikuotam CG, o R.MspI CCGG taikinyje nekerpa tik esant 5caCG (Liutkevičiūtė ir kt., 2014). Šiomis REazėmis po metilinimo kirpome modelinius DNR fragmentus, kurių vienoje arba abiejose CCGG taikinio grandinėse buvo įvestas 5caC (5caC/C, 5caC/5mC, 5caC/5caC). kPGR būdu nustatėme, kad analizuotuose taikiniuose wt SssI MTazė efektyviai metilina nCG (nemodifikuota DNR tampa atspari R.HpaII kirpimui) ir reakcijos metu nepašalina karboksilo grupės nuo 5caC (5caC DNR išlieka atspari R.MspI, 3.4. pav. B). Taip pat metilinimo efektyvumą įvertinome atliekant bisulfitinę lambda bakteriofago genomo sekoskaitą (šią ir kitas su caCLEAR metodu susijusias sekoskaitas analizavo dr. Povilas Gibas, visas DNR bibliotekų sekoskaitas atliko Audronė Rukšėnaitė). Analizė parodė beveik visišką (~97,3%) genome esančių nCG metilinimą (3.4. pav. C), atitinkantį paskelbtus duomenis (Wu ir kt., 2014). Atsižvelgiant į žinomą „neteisingą“ metilintų citozinų konversiją į timinus veikiant natrio bisulfitu (2,7% su kai kuriais bisulfitinės konversijos rinkiniais (Holmes ir kt., 2014) ir 3,5% nustatytą mūsų eksperimentuose), pasiektas metilinimo lygis gali būti vertinamas kaip įsotintas.

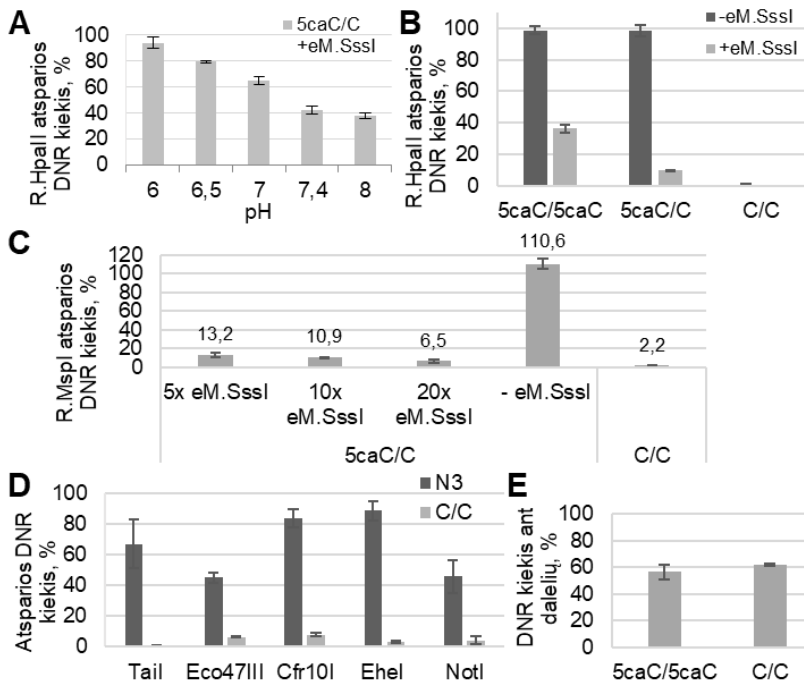
### 3.2.2 eM.SssI atliekamo 5caC dekarboksilinimo ir žymėjimo įvertinimas

Siekėme įvertinti, kaip efektyviai eM.SssI geba pašalinti karboksilo grupę nuo 5caC (3 žingsnis) ir pažymėti susidariusį nemodifikuotą CG (3 žingsnis, 3.3 pav.).

5caC dekarboksilinimo efektyvumą (3 žingsnis) vertinome kPGR būdu naudojant modelinius DNR fragmentus su CCGG taikinyje įvestu 5caC, kurie po eM.SssI dekarboksilinimo tampa nebeatsparūs R.MspI ar R.HpaII kirpimui. Įvertinome reakcijos priklausomybę nuo tirpalo pH ir stebėjome efektyviausią 5caC dekarboksilinimą esant pH 8 (liko mažiausias kiekis nesukirpto fragmento 3.5 pav. A), todėl toliau reakcijos buvo vykdomos tokiomis sąlygomis. Taip pat patikrinome 5caC dekarboksilinimo efektyvumą

modeliniams fragmentams (kai 5caC yra vienoje arba abiejose grandinėse) esant genominei DNR mišinyje (santykiu 1 ng frag. : 100 ng žmogaus genominei DNR), imituojant į natūralias panašias sąlygas su reta 5caC modifikacija. eM.SssI gebėjo efektyviai atlikti 5caC dekarboksilinimą mišinyje bei esant įvairiems substratams (3.5 pav. B, tai taip pat matyti 3.2 pav. C) – 5caC esant abiejose ar vienoje CG taikinio grandinėje. Siekiant nustatyti pastovias efektyvios reakcijos sąlygas, vertinome eM.SssI santykį su taikiniaus, reikalingą efektyviai reakcijai. 5caC dekarboksilinimo įvertinimas su modeliniu 5caC turinčiu DNR fragmentu parodė, kad reakcija yra efektyvi esant įvairiam fermento pertekliui prieš CG taikinius (5×-20×). Kadangi efektyvumas imant didesnę nei 10× perteklių padidėjo nedaug (3.5 pav. C, eksperimentą atliko Kotryna Skardžiūtė), tolesniuose eksperimentuose naudojome šias sąlygas, jei nenurodyta kitaip.

Nors eM.SssI geba atlikti 5caC dekarboksilinimą esant kofaktoriui (3.2 pav. C), atliekant caCLEAR procedūrą dekarboksilinimo (3 žingsnis) ir žymėjimo azido grupėje (4 žingsnis) reakcijas nusprendėme atlikti vieną po kitos siekiant didesnio abiejų reakcijų efektyvumo. Šios reakcijos gali būti patogiai vykdomos viena po kitos viename mėgintuvėlyje be papildomo DNR gryninimo žingsnio – atlikus 5caC dekarboksilinimą yra įmaišomas Ado-6-N<sub>3</sub> kofaktorius. Siekdami tinkamai suderinti šias reakcijas, vertinome žymėjimo žingsnio efektyvumą 5caC dekarboksilinimo reakcijos sąlygomis nemodifikuotame *c-fos* geno promotoriaus regiono 428 bp PGR fragmente, turinčiame 37 CG taikinius. Naudojant restrikcijos endonukleazes, jautrias citozino modifikacijoms (TatI, Eco47III, Crf10I, turinčias po vieną CG taikinį fragmente, bei EheI, NotI, turinčias po 2 taikinius), kPGR būdu įvertinome 9 CG taikinius ir stebėjome vidutiniškai 66% žymėjimo azido grupėje efektyvumą (3.5 pav. D). Abiejų dekarboksilinimo ir žymėjimo žingsnių efektyvumą kartu įvertinome kPGR būdu, atlikus biotino prijungimą prie žymėtos DNR ir praturtinimą streptavidino magnetinėmis dalelėmis. Modelinis DNR fragmentas buvo sumaišytas su žmogaus DNR (1:300 ng). Fragmento, turėjusio 5caC abiejose grandinėse, praturtinimas buvo panašus kaip ir nemodifikuoto fragmento, o tai rodo efektyviai įvykusį dekarboksilinimą ir žymėjimą (3.5 pav. E).

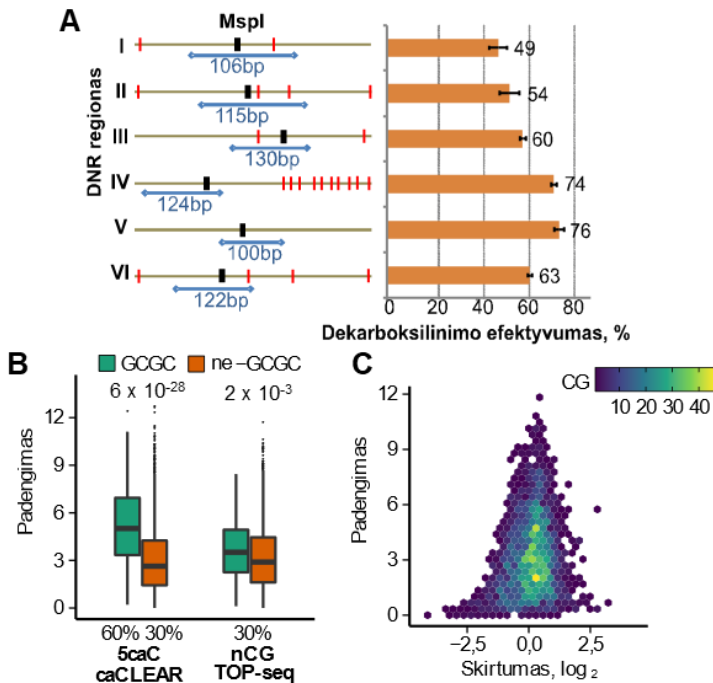


**3.5 pav.** eM.SssI metiltransferazės 5caC dekarboksilinimo ir žymėjimo azido grupe tyrimas modeliniuose fragmentuose turinčiuose 5caC vienoje (5caC/C) ar abiejose (5caC/5caC) CG taikinio grandinėse. C/C – nemodifikuotas fragmentas. 5caC dekarboksilinimas vykdytas 30 min (A) arba 3 val. (B-E). CG taikinių modifikavimas vertintas kerpant tam jautriomis restrikcijos endonukleazėmis (A-D). DNR kiekiai įvertinti kPGR. (A) 5caC dekarboksilinimo aktyvumas (kurį rodo mažėjantis R.HpaII atsparios DNR kiekis) esant skirtingam reakcijos buferio pH. (B) 5caC dekarboksilinimo efektyvumas modeliniam fragmentui (1 ng) esant mišinyje su genomine DNR (100 ng). (C) Modelinis DNR fragmentas buvo inkubuotas esant skirtingam eM.SssI pertekliui (5, 10, 20) ir tuomet kirptas su R.MspI. Buvo stebimas ~94% C5caCGG dekarboksilinimas esant 20× fermento pertekliui prieš CG taikinius (~6% DNR buvo atspari R.MspI kirpimui). (D) Nemodifikuotų CG žymėjimo vertinimas azido grupe naudojant eM.SssI ir sintetinį kofaktorių Ado-6-N<sub>3</sub> *c-fos* geno promotoriaus regiono DNR fragmente, turinčiame 37 CG taikinius. Azido grupės prijungimo (N3, tamsiai pilka) efektyvumas buvo vertintas naudojant restrikcijos endonukleazes, kurios yra jautrios citozino modifikacijoms jų taikiniuose. Naudotos restrikcijos endonukleazės, turinčios vieną (Tail, Eco47III, Crf10I) arba du (EheI, NotI) taikinius DNR fragmente. Kontrolinė nemodifikuota DNR (C/C) parodyta šviesiai pilkai. (E) Dekarboksilinimo ir žymėjimo azido grupe efektyvumo įvertinimas atliekant biotino prijungimą prie žymėto modelinio DNR fragmento ir praturtinimą streptavidino magnetinėmis dalelėmis.

Tuomet tyrėme eM.SssI vykdomą 5caC dekarboksilinimą genomo lygyje. Su žmogaus fibroblastų IMR90 DNR atlikus 5mC oksidaciją su TET1, paruošėme 5caC turinčią modelinę genomine DNR. Tuomet kPGR būdu vertinome eM.SssI dekarboksilinimo efektyvumą pasirinktuose 5caCG

taikiniuose pasitelkiant skirtingą R.MspI restrikcijos endonukleazės jautrumą modifikacijoms jos CCGG taikinyje: R.MspI kerpa esant visoms citozino modifikacijoms, išskyrus 5caC. Pasirinkome šešis įvairių CG taikinių išsidėstymą turinčius genomo regionus. DNR atsparumo R.MspI kirpimui skirtumo prieš ir po veikimo eM.SssI analizė parodė ~49-76% dekarboksilinimo efektyvumą skirtinguose regionuose (3.6 pav. A). Darant prielaidą, kad eksperimentui buvo naudota daug 5caC turinti DNR, (paprastai 60-70% CG IMR90 DNR yra metilinti (Lister ir kt., 2009), kurie po TET1 oksidacijos tikėtina buvo efektyviai paversti į 5caCG, nes DNR buvo 90-100% atspari R.MspI), o 5caC paprastai genome išsidėsto retai ir turi žemą modifikavimo lygį (pelės kamieninėse embrioninėse ląstelėse 5caC/5fC modifikavimo procentas yra apie 8-10%, su ekstremumais iki 80%; Neri ir kt., 2015), tikimės efektyvesnio karboksilo grupės pašalinimo nuo 5caCG natūralioje genomineje DNR.

Tuomet, siekdami įvertinti eM.SssI vykdomą dekarboksilinimą ir žymėjimą azido grupė viso genomo lygyje, į lambda bakteriofago DNR CG bei GCGC taikinius įvedėme atitinkamai 30% ir 60% 5caC (naudojant M.HhaI bei M.SssI metilintų taikinių oksidaciją TET1, žr. 2.2.8). Atlikome 2-5 caCLEAR metodo žingsnius, paruošėme bibliotekas ir atlikome jų sekoskaitą. caCLEAR signalo įvertinimui CG taikinyje skaičiavome kiek kartų jis buvo nuskaitytas (t.y. kokį sekų skaičių turėjo tas regionas), toliau tekste tai vadiname padengimu (po sekoskaitos gavome vidutiniškai 50× padengimą CG taikiniui; lambda genome yra 6226 CG). Analizė parodė, kad caCLEAR CG padengimas atspindėjo 5caC modifikacijos lygį (3.6 pav. B) – padengimas buvo didesnis 60% 5caC turėjusiuose GCGC taikiniuose lyginant su 30% 5caC turėjusiais kitais CG taikiniais. Vadinasi, caCLEAR metodu galima palyginti 5caC modifikavimo stiprumą. Taip pat paruošėme kontrolinę TOP-seq biblioteką (tik 4 ir 5 caCLEAR žingsniai, yra nustatomi nCG) su DNR, kurioje nmodifikuotų CG lygis buvo 30% (o CG metilinimas 70%, 5caC nebuvo). Siekdami įvertinti 5caC dekarboksilinimo išbaigtumą, palyginome CG padengimą tarp 30% nCG turėjusio TOP-seq mėginio ir taip pat 30% 5caCG lygį turėjusio caCLEAR mėginio (dėl kitokio modifikavimo lygio, GCGC taikiniai analizėje nenaudoti). Daugumoje CG nebuvo padengimo skirtumo tarp caCLEAR ir TOP-seq mėginių (3.6 pav. C). Vadinasi, eM.SssI beveik visiškai pašalina karboksilo grupes nuo 5caC ir todėl galėjo atlikti 5caC turėjusių CG taikinių žymėjimą panašiu efektyvumu kaip ir nmodifikuotuose CG.



**3.6 pav.** caCLEAR metodo dekarboksilinimo ir žymėjimo žingsnių efektyvumo patikrinimas modelinėje genomineje DNR. **(A)** Genominių regionų 5caC taikinių dekarboksilinimas. Kairėje, šešių regionų, naudotų 5caC dekarboksilinimo įvertinimui kPGR, schematinis vaizdas (R.MspI taikinio CG parodytas juodai, kiti gretimi CG raudonai, PGR produktas parodytas mėlynai). Dešinėje, 5caC dekarboksilinimo efektyvumas, parodytas kaip R.MspI atsparios DNR kiekio skirtumas taikinyje prieš ir po eM.SssI dekarboksilinimo. **(B)** Modelinio lambda bakteriofago genomo caCLEAR bibliotekų padengimas (normalizuotas pagal visą nuskaitymų skaičių), atspindintis buvusio 5caC lygį GCGC ar kituose CG (neskaičiuojant GCGC) taikiniuose, kurie atitinkamai turėjo vidutiniškai 60% ir 30% 5caC. Kontroliniai TOP-seq mėginiai rodo maksimalų padengimą CG taikiniuose (30% nemodifikuotų CG, nėra 5caC). Skaičiai virš grafiko rodo p vertes, apskaičiuotas dvipusiu poriniu t-testu. **(C)** Normalizuoto padengimo skirtumas CG taikiniuose (neskaičiuojant GCGC) tarp caCLEAR ir kontrolinių TOP-seq bibliotekų esant įvairiems padengimo slenksčiams. Spalvų kodas nurodo CG taikinių skaičių.

### 3.2.3 Visos caCLEAR procedūros įvertinimas

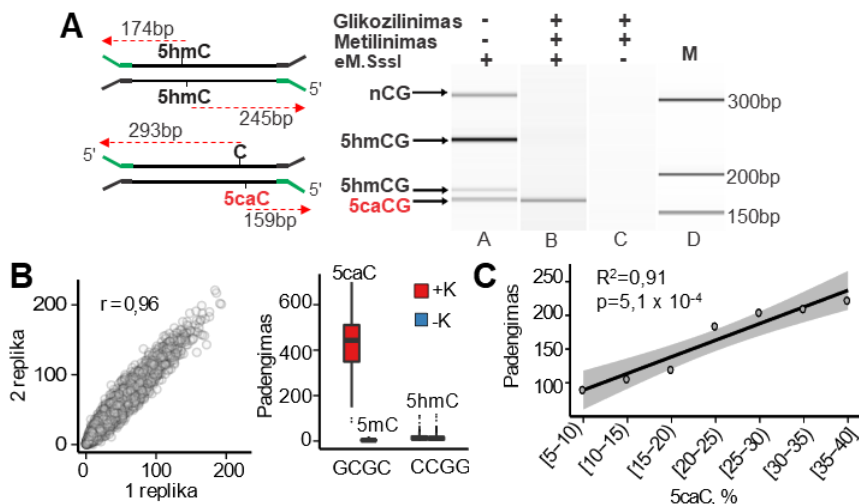
Tuomet patikrinome visą caCLEAR procedūrą naudojant modelinius DNR fragmentus (eksperimentą atliko K. Skardžiūtė). Modelinę sistemą sudarė du lygiomis dalimis sumaišyti DNR fragmentai, kurių kiekviename buvo po vieną skirtingai modifikuotą CG taikinį – vieno fragmento abiejose grandinėse buvo įvestas 5hmC, kitas fragmentas vienoje grandinėje turėjo 5caC, o kitoje nCG (3.7 pav. A). Kadangi atstumai nuo DNR grandinės 5' galo iki kiekvieno CG yra skirtingi, analizuojant elektroforetiškai galima

pagal ilgį atskirti, nuo kurio CG taikinio susidarė galutinis caCLEAR PGR produktas (3.7 pav. A). Neatlikus glikozilino ir metilino žingsnių bei eM.SssI vykdant 5hmC dehidroksimetilinimą, 5caC dekarboksilinimą ir nCG žymėjimą susidarė 4 skirtingi PGR produktai nuo kiekvieno CG (3.7 pav. A, A takelis). Atlikus visus caCLEAR žingsnius (1-5, 3.3 pav.), susidarė vienas PGR produktas, kilęs nuo 5caC turinčios DNR grandinės (3.7 pav. A, B takelis), tuo tarpu 5hmC ar nemodifikuotą C turėję CG PGR produktų nesudarė. Kontrolinėje reakcijoje be eM.SssI PGR produktai nesudarė (3.7 pav. A, C takelis). Vadinasi, 5hmC ir nCG buvo efektyviai apsaugoti nuo žymėjimo ir caCLEAR procedūra specifiskai nustatė 5caCG.

Taip pat visą caCLEAR procedūrą analizavome modeliniame lambda bakteriofago genome. Pasitelkiant M.HpaII hidroksimetilinimą ir M.HhaI metilintų taikinių oksidaciją (žr. 2.2.8), į DNR buvo įvesta 10% 5hmC CCGG taikiniuose (328 CG genome) ir 10% 5caC GCGC taikiniuose (215 CG genome), o 70% CG buvo metilinti M.SssI (atspindint žinduolių metilino lygį). Atlikus visus caCLEAR žingsnius, buvo paruoštos bibliotekos ir atlikta sekoskaita. Geresniam specifiskumui buvo naudojamas biotininintas ODN ir aliekamas DNR su prijungtu ODN praturtinimas (kaip ir hmTOP-seq, žr. 1.14 pav.). Taip pat atliktas kontrolinis caCLEAR eksperimentas su analogiskai paruošta DNR, neturinčia 5caC. Sekoskaitos analizė parodė gerą koreliaciją tarp bibliotekų techninių replikų (*Pearson*  $r = 0,96$ , 3.7 pav. B, kairėje). Kaip ir tikėtasi, dauguma nuskaitymų kilo iš 5caC turėjusių GCGC taikinių, tuo tarpu nuo kitų CG taikinių, įskaitant CCGG 5hmC taikinius, susidarė tik foninio lygio nuskaitymų skaičiai tikslinėje ir kontrolinėje bibliotekose (3.7 pav. B, dešinėje). Tai dar kartą parodo metodo specifiskumą 5caC, o taip pat ir jautrumą nedideliems natūraliai aptinkamiems 5caC modifikacijos lygiams (pelės kamieninėse embrioninėse ląstelėse 5caC/5fC modifikavimas yra vidutiniškai 8-10%; Neri ir kt., 2015). Papildomam 5caC genomio profiliavimo įvertinimui, į lambda DNR GCGC taikinius įvedėme vidutiniškai 20% 5caC, 70% CG buvo metilinti (žr. 2.2.8). Įvertinome 5caC modifikacijos lygį CG taikiniuose caMAB-seq (žr. 1.4.4, 2.2.7) bisulfitinės sekoskaitos metodu (5caC lygis varijavo tarp 4-40%) ir taip pat analizavome šią DNR caCLEAR metodu. Įvertinome caCLEAR padengimo ryšį su 5caC lygiu (5% 5caC lygio intervaluose). Tarp bisulfitinės sekoskaitos ir caCLEAR buvo stebima linijinė priklausomybė – didėjant 5caC lygiui, caCLEAR padengimas didėjo (3.7 pav. C). Tai dar kartą patvirtina, kad caCLEAR metodas gali suteikti palyginamąją kiekybinę 5caC modifikacijos lygio informaciją.

Visi šie išsamūs eksperimentai patvirtino caCLEAR metodo tinkamumą 5caCG taikinių nustatymui genome.





**3.7 pav.** caCLEAR procedūros patikrinimas modelinėse sistemose. **(A)** Dviejų modelinių DNR fragmentų caCLEAR analizė. Kiekvienas fragmentas (kairėje) turi skirtingai modifikuotus citozinas abiejose CG taikinio grandinėse. Raudonomis rodyklėmis pažymėti galintys susidaryti galutinių bibliotekų PGR produktai. Dešinėje, elektroforetinė caCLEAR produktų analizė bioanalizatoriumi (*Agilent*) parodė vieną nuo 5caC kilusį PGR produktą (spendžiant pagal ilgį, B takelis) iš keturių galimų (A takelis). PGR produktai nesusidarė, kai reakcijoje nebuvo žymėjimą atliekančios eM.SssI (C takelis). **(B)** caCLEAR metodo tikrinimas su modeline lambda bakteriofago genomine DNR atliekant sekoskaitą. Kairėje, caCLEAR bibliotekų techninių replikų padengimo sklaidos grafikas. Dešinėje, lambda DNR techninių replikų vidurkio caCLEAR padengimo pasiskirstymas mėginiuose, kurie turėjo ~10% 5hmC ir 5caC, atitinkamai CCGG ir GCGC taikiniuose (raudona, +K). Kontrolinėse bibliotekose nebuvo 5caC (mėlyna, -K) ir jų GCGC taikiniai buvo metilinti (5mC). **(C)** caCLEAR padengimo priklausomybė nuo 5caC lygio GCGC taikiniuose, nustatyto bisulfitine sekoskaita modelinėje lambda DNR. Linijinis modelis buvo pritaikytas naudojant vidutinį 5caC procentų grupės padengimą (parodyta *p*-vertė ir pakoreguotas R kvadratas). Laužtiniai skliaustai rodo intervale įskaičiuojamą reikšmę, paprasti skliaustai – neįskaičiuojamą.

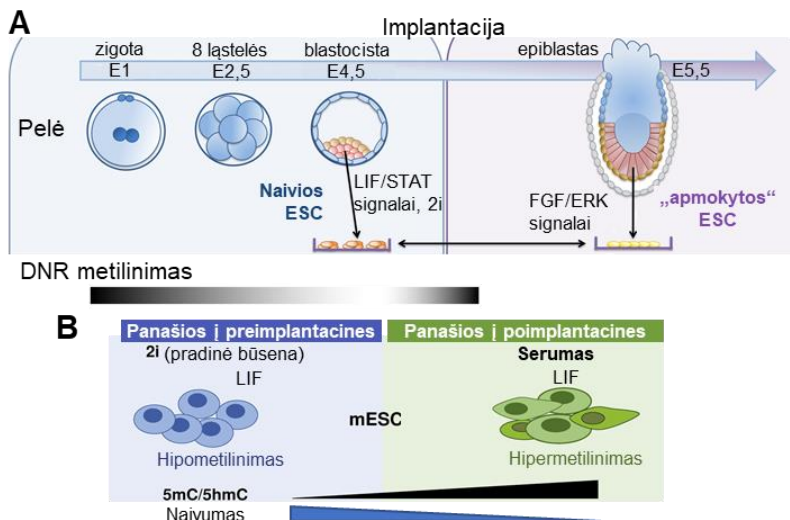
### 3.3 5caC tyrimas pelės embrioninėse kamieninėse ląstelėse caCLEAR metodu

#### 3.3.1 Dviejų pelės ESC pluripotensiškumo būsenų caCLEAR analizė

Patvirtinus caCLEAR modelinėse DNR sistemose, toliau juo analizavome 5caC natūralioje genomine DNR. Nusprendėme 5caC tirti pelės embrioninėse kamieninėse ląstelėse, nes yra atlikta jų 5caC profiliavimo tyrimų (Lu ir kt., 2015b; Shen ir kt., 2013), su kuriais būtų galima palyginti ir taip patvirtinti caCLEAR gaunamus duomenis. Buvo tiriamos laukinio tipo (wt) ląstelės, o *Tdg* geno neturinčios (*Tdg*<sup>-/-</sup>) buvo naudojamos kaip teigiama kontrolė – TDG nevykdant 5caC iškirpimo, jo kiekiai ląstelėse padidėja (He

ir kt., 2011). Ląstelės su išveiklintais TET1/2/3 (TET TKO, angl. *triple knockout*) genais buvo naudojamos kaip neigiama kontrolė – nesant TET nesusidaro 5mC oksiduotos formos.

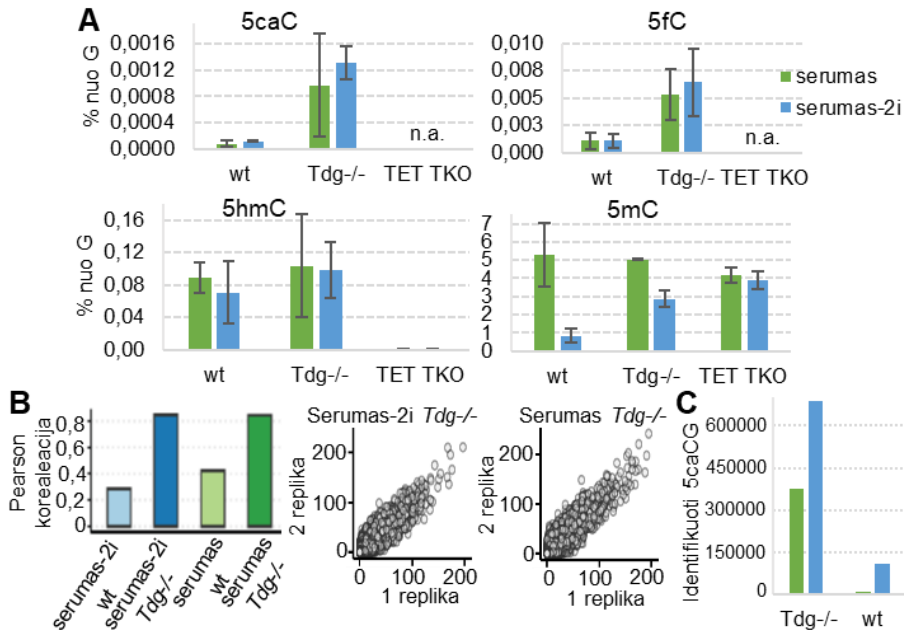
Priklausomai nuo auginimo sąlygų, mESC gali įgyti dvi būsenas (3.8 pav.), kurios gali keistis viena į kitą ir kurios atspindi skirtingas vystymosi stadijas (Habibi ir kt., 2013; von Meyenn ir kt., 2016). mESC auginant terpėje, turinčioje serumo ir leukemiją slopinančio veiksnio (LIF, angl. *leukemia inhibitory factor*), ląstelės yra panašios į ankstyvą epiblastą po implantacijos į gimdą ir yra vadinamos „apmokytomis“ (angl. *primed*) diferenciacijai. Tuo tarpu mESC auginant beseruminėje terpėje esant dviems mažų molekulių slopikliams (o taip pat ir LIF), ląstelės labai panašios į esančias blastocistos vidinėje ląstelių masėje prieš implantaciją (ankstesnės stadijos) ir yra vadinamos naivios arba pradinės (angl. *ground*) pluriipotentišku būsenos (Marks ir kt., 2012; Martello ir Smith, 2014). Ši slopikliai (PD0325901 bei CHIR99021, atitinkamai slopina mitogenų aktyvuojamą baltymų kinazę (MEK) ir glikogeno sintazės kinazę 3 (GSK3)) sukelia didelius pokyčius kamieninėse ląstelėse – stebimas pakitęs histonų modifikavimas ir DNR demetilimas (Habibi ir kt., 2013; Marks ir kt., 2012; von Meyenn ir kt., 2016). Nusprendėme tirti 5caC pasiskirstymą auginant ląsteles šiomis skirtingomis sąlygomis (esant 2i ir be 2i), kadangi jos atspindi skirtingas pluriipotentišku būsenas ir yra įdomus modelis demetilimo dinamikos tyrimams.



**3.8 pav.** Naivi ir „apmokyta“ pelės embrioninių kamieninių ląstelių (mESC) būseną. (A) Besivystančio pelės embriono ląstelių naivi ir „apmokyta“ pluriipotentišku būseną ir *in vitro* sąlygos, reikalingos palaikyti/paversti vieną būseną į kitą (adaptuota pagal Eguizabal ir kt., 2019). E – embriono vystymosi diena; LIF – leukemiją slopinantis veiksnys; STAT – signalo perdavėjas ir transkripcijos aktyvatorius 3 (angl.

*signal transducer and activator of transcription 3*); FGF – fibroblastų augimo veiksnys; ERK – ekstraląstelinė reguliuojama kinazė; 2i – du kinazių slopikliai. Apačioje parodytas embrione vykstantis DNR metilinimo kitimas (pagal Kim ir kt., 2014). **(B)** Auginant mESC su dviem kinazių slopikliais (2i), ląstelės yra pradinės pliuripotentiškumo būsenos, panašios į esančias vidinėje blastocistos ląstelių masėje, o DNR hipometilinta. Tuo tarpu serumo sąlygomis ląstelių DNR yra hipermetilinta ir metilinimo išsidėstymas atitinka poimplantacinėse epiblasto ląstelėse nustatomą metilinimą (adaptuota pagal Habibi ir kt., 2013).

Siekiant palaikyti labiau vienalytę mESC ląstelių populiaciją, nemažai publikuotų DNR modifikacijų tyrimų (Booth ir kt., 2012; Liu ir kt., 2019; Song ir kt., 2013; Yu ir kt., 2012) yra naudojamas serumo-2i-LIF mišinys, o šios sąlygos atitinka naivią pliuripotentiškumo būseną (Martinez-Val ir kt., 2021). Todėl auginome minėtas ląstelių linijas serumo ir serumo-2i sąlygomis esant LIF ir lygindami jų genominių 5caC pasiskirstymą analizavome pliuripotentiškumo būsenas. Pirmiausia įvertinome bendrą 5mC ir oksimC lygį atliekant kiekybinę HPLC-MS/MS (3.9 pav. A). Analizė parodė hipometilintą laukinio tipo serumo-2i ląstelių būseną (~1% 5mC), o tai atitinka 2i sąlygomis auginamų mESC publikuotus duomenis (Leitch ir kt., 2013). Tuo tarpu serumo wt mESC ląstelėse nustatėme ~5% citozino metilinimą (Leitch ir kt. (2013) bei Stadler ir kt. (2011) nustatė ~4% 5mC). Nors 5hmC kiekiai buvo panašūs abejomis auginimo sąlygomis, serumo-2i mESC ląstelėse buvo galima stebėti nedidelę kitų oksimC didėjimo tendenciją (Wheldon ir kt. (2014) taip pat buvo stebimas 5caC signalo sustiprėjimas 2i sąlygomis). Nustatėme ~0,00008% 5caC (skaičiuojant nuo guanino) serumo wt mESC (o tai atitinka anksčiau paskelbtus duomenis; Ito ir kt., 2011; Wheldon ir kt., 2014) ir ~0,00012% 5caC serumo-2i wt ląstelėse. Kaip ir tikėtasi, *Tdg*<sup>-/-</sup> ląstelėse buvo nustatyti didesni 5caC kiekiai (~0,00097% ir ~0,0013% nuo guanino atitinkamai serumo ir serumo-2i sąlygų *Tdg*<sup>-/-</sup> ląstelėse), lyginant su laukinio tipo ląstelėmis. 5caC ir 5fC nebuvo aptikti abejų sąlygų TET TKO ląstelėse, nors buvo aptikti labai nedideli 5hmC kiekiai.



**3.9 pav.** mESC analizė. (A) Citozino modifikacijų HPLC-MS/MS analizė. Laukinio tipo (wt), *Tdg* geno neturinčios (*Tdg*<sup>-/-</sup>) ir TET1/2/3 išveiklintus genus turinčios mESC (TET TKO) buvo auginamos serumo arba serumo-2i sąlygomis. 5caC, 5fC, 5hmC ir 5mC % nuo viso G buvo apskaičiuoti naudojant atitinkamo nukleozido standarto kalibracinę kreivę. Buvo naudoti bent du biologiniai pakartojimai su papildomais dvejais techniniais pakartojimais kai kuriuose mėginiuose. n.a. – signalas nebuvo aptiktas. (B) mESC caCLEAR bibliotekų korelacija tarp techninių replikų. Korelacijos sklaidos grafikai parodyti *Tdg* neturinčiai mESC linijai. (C) Mėginiuose identifikuotų 5caCG skaičius po griežto jų atrinkimo.

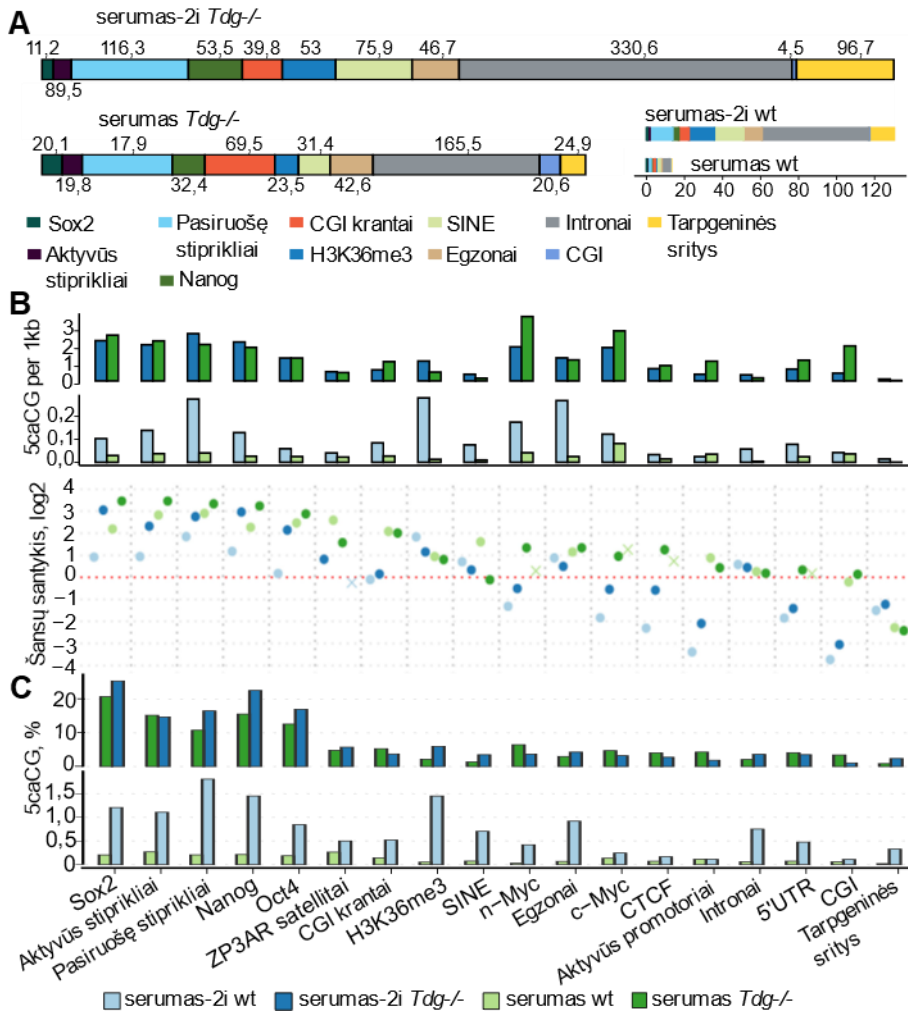
Tuomet esant abejoms auginimo sąlygoms visose ląstelių linijose buvo nustatyti 5caC išsidėstymo profiliai genominiu mastu. Kadangi 5caC kiekiai tarp ląstelių linijų gana skirtingi, galutiniai bibliotekų dydžiai ir nustatyti 5caCG skaičius atspindi santykinę 5caC gausumą. Buvo gauta ~1,7 mln. apdorotų (atrinktų pagal ilgį, sekoskaitos kokybę, atstumą iki CG bei pašalinus adapterius žr. Staševskij ir kt., 2017; Ličytė ir kt., 2020) vienos krypties nuskaitymų kiekvienam techniniam pakartojimui serumo wt mESC ir ~15 mln. serumo *Tdg*<sup>-/-</sup> mESC. Atitinkamai buvo gauta ~9 mln. ir ~25 mln. nuskaitymų serumo-2i wt ir *Tdg*<sup>-/-</sup> mESC. Techniniai *Tdg*<sup>-/-</sup> mESC caCLEAR bibliotekų pakartojimai turėjo aukštą korelaciją (*Pearson*  $r=0,85$ ), tuo tarpu laukinio tipo mESC bibliotekų korelacija buvo žemesnė (0,29-0,42; 3.9 pav. B). TET TKO caCLEAR bibliotekose buvo labai nedaug nuskaitymų (~0,4 mln.) ir buvo nustatyta ~196 000 CG, kurie beveik nesidengė tarp replikų (5 501 besidengiančių CG, 2,8% nuo visų nustatytų CG). Vadinasi, šie CG yra atsitiktinis triukšmas, klaidingai teigiami taikiniai. Siekiant patikimai

nustatyti visus 5caCG taikinius wt ir *Tdg*<sup>-/-</sup> mESC, naudojome didelio griežtumo duomenų atrinkimą (žr. 2.2.25), po kurio buvo nustatyta iš viso 8 685 (vidutiniškai 5× padengimas) ir 374 488 (vidutiniškai 20× padengimas) 5caCG atitinkamai serumo wt ir *Tdg*<sup>-/-</sup> ląstelėse; 110 753 (~8× padengimas) ir 689 303 (~18× padengimas) 5caCG atitinkamai serumo-2i wt ir *Tdg*<sup>-/-</sup> mESC (3.9 pav. C). TET TKO ląstelėse po atrinkimo buvo nustatyti tik 1 259 CG.

Toliau palyginome caCLEAR su kitais 5caC profiliavimo metodais. Serumo *Tdg*<sup>-/-</sup> mESC 5caCG rinkinys gerai persidengė su 5caC praturtintais regionais nustatytais toje pačioje ląstelių linijoje antikūnais paremtu DNR imunoišsodinimo metodu (5caC-DIP; Shen ir kt., 2013): 58% 5caCG persidengė su 5caC-DIP praturtintais regionais (šansų santykis =18;  $p < 2,2 \times 10^{-16}$ , Fišerio tikslusis testas). Be to, 62% ir 80% 5caC-DIP regionų persidengė su bent vienu 5caCG atitinkamai serumo ir serumo-2i *Tdg*<sup>-/-</sup> ląstelių duomenų rinkiniuose. Po to serumo *Tdg*<sup>-/-</sup> mESC caCLEAR metodu nustatytus 5caCG lyginome su bisulfitiniu modifikavimu paremta 5caC sekoskaita CAB-seq (Lu ir kt., 2013, 2015b). Savo griežto atrinkimo duomenų rinkiniuose aptikome 46% CAB-seq taikinių (CAB-seq buvo nustatyti 4 806 5caCG taikiniai), o 78% 5caCG taikinių nustatėme naudojant ne tokį griežtą atrinkimą (kai naudojami CG taikiniai su 1× ir didesniu padengimu). M.SssI metiltransferaze paremta bisulfitinė sekoskaita (MAB-seq; Neri ir kt., 2015; Wu ir kt., 2014) vienu metu nustato abi 5caC ir 5fC modifikacijas, kurių genominės padėtyys mažai persidengia (Lu ir kt., 2015b), todėl negalėjome su juo palyginti savo nustatytų 5caCG.

Nors nustatytų 5caCG pasiskirstymas genų reguliacijos ir genominiuose elementuose varijavo skirtingomis sąlygomis augintose mESC (3.10 pav. A), dauguma 5caCG taikinių buvo praturtinti pasiruošusiuose ir aktyviuose stiprikliuose (turinčiuose atitinkamai H3K4me1 ir H3K27ac/H3K4me1), įvairių su pluriipotentišku susijusių veiksnių prisijungimo vietose (tokių kaip Sox2, Nanog, Oct4 (Pou5FI)) bei kai kurių tipų satelitiniuose pasikartojimuose ir SINE judriuosiuose elementuose (3.10 pav. B). Dvi pluriipotentišku būsenos skyrėsi 5caCG praturtinimu įvairiuose svarbiuose genomo elementuose: serumo mESC 5caCG buvo praturtinti CG salose ir jų krantuose, 5' netransliuojamuose regionuose, aktyvių stipriklių regionuose, aktyvių promotorių regionuose (pažymėtų H3K9ac) bei izoliuojančio baltymo CTCF jungimosi srityse, o serumo-2i mESC 5caCG šiuose regionuose buvo nepraturtinti ar praturtinti mažiau. Lyginant su serumo mESC, abiejose wt ir *Tdg*<sup>-/-</sup> serumo-2i ląstelių linijose 5caCG buvo santykinai labiau praturtinti tarpgeniniuose regionuose,

intronuose ir aktyviai transkribuojamuose regionuose, pažymėtuose H3K36me3. Kadangi šie elementai sudaro didelę geno dalį, tai atitinka labiau demetilintą serumo-2i ląstelių geno būseną. Nepaisant didesnio 5caCG gausumo daugumoje wt ir *Tdg*<sup>-/-</sup> serumo-2i mESC geno elementų (3.10 pav. C), šiose ląstelėse 5caCG yra pasiskirstę mažiau specifiskai – tai rodo mažesnės praturtinimo šansų santykio reikšmės daugumoje tirtų elementų, lyginant su serumo mESC (3.10 pav. B).

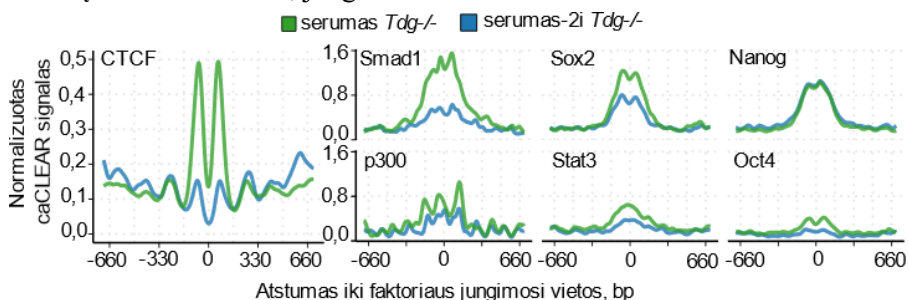


**3.10 pav.** 5caCG pasiskirstymas naivios ir „apmokytos“ pluriopotentiškumo būsenų pelės ląstelėse. (A) Nustatytų 5caCG pasiskirstymas (tūkstančiais) genominiuose elementuose ir regionuose. 5'UTR – 5' netransliuojamas regionas. (B) 5caCG skaičiai (1 kb dydžio regionams) ir 5caCG taikinių praturtinimo šansų santykiai (log<sub>2</sub>) nustatyti Fišerio tiksluoju testu įvairiose wt ir *Tdg*<sup>-/-</sup> mESC ląstelių geno srityse. Pasiruošę stiprikliai – regionai su H3K4me1 histonų žyme; aktyvūs stiprikliai – regionai su H3K4me1 ir H3K27ac žymėmis; aktyvūs promotoriai – 2 kb dydžio

regionai prieš srovę nuo genų, kurie persidengia su H3K9ac histonų žyme. Statistiškai nereikšmingi ( $p \geq 0,05$ ) įverčiai pažymėti „x“. (C) 5caCG procentas nuo visų CG įvairiuose genomo elementuose ir transkripcijos faktorių jungimosi regionuose.

### 3.3.2 5caC modifikuoti transkripcijos veiksnių ir reguliatorių jungimosi regionai dviejose pluriipotentiško būsenose skiriasi

Kitame etape tyrėme 5caCG pasiskirstymą 12 transkripcijos veiksnių ir reguliatorių (Nanog, Oct4, Sox2, STAT3, Smad1, Zfx, c-Myc, n-Myc, E2f1, CTCF, p300 ir Suz12), kurie yra svarbūs embrioninių kamieninių ląstelių pluriipotentiško būsenai palaikyti (Chen ir kt., 2008; Martello ir Smith, 2014), jungimosi vietose. Analizavome 5caCG pasiskirstymą ir viešai prieinamus mESC ChIP-seq baltymų jungimosi regionų duomenis (Atlasi ir kt., 2019; Chen ir kt., 2008; Galonska ir kt., 2015). Centrinės veiksnių jungimosi regionų (VJR, vidutinis regionų ilgis 16-50 bp; Atlasi ir kt., 2019; Chen ir kt., 2008) srityse buvo vidutiniškai 1-2 5caCG. Abejomis auginimo sąlygomis Nanog, Sox2 ir Oct4 jungimosi regionai (~600-1000 bp; Galonska ir kt., 2015) buvo praturtinti 5caCG ~300 bp atstumu nuo centrinės regiono srities ir turėjo vidutiniškai 3-4 5caCG. caCLEAR profiliai aplink VJR parodė 5caCG praturtinimą ties Sox2, Nanog, Oct4, Smad1, STAT3, CTCF ir p300 jungimosi vietomis (3.11 pav.). Įdomu, kad nustatėme periodinį 5caCG pasiskirstymą aplink CTCF jungimosi regionus: atstumas tarp 5caCG signalų maksimumų buvo 165-bp, o tokie atstumai yra būdingi DNR jungtukams tarp nukleosomų. Panašus pasiskirstymo sutapimas su nukleosomų išsidėstymu CTCF jungimosi vietose taip pat buvo parodytas 5hmC ir 5fC atveju (Sun ir kt., 2013a, 2015). Mūsų rezultatai dar kartą patvirtina pasiūlytą chromatinio struktūros įtaką TET baltymų aktyvumui (Sun ir kt., 2013b). Periodišką 5caCG pasiskirstymą, nors mažiau išreikštą, taip pat stebėjome aplink p300 koaktyvatoriaus, kuris sąveikauja su įvairiais transkripcijos veiksniais ir yra histonų acetiltransferazė, jungimosi vietas.

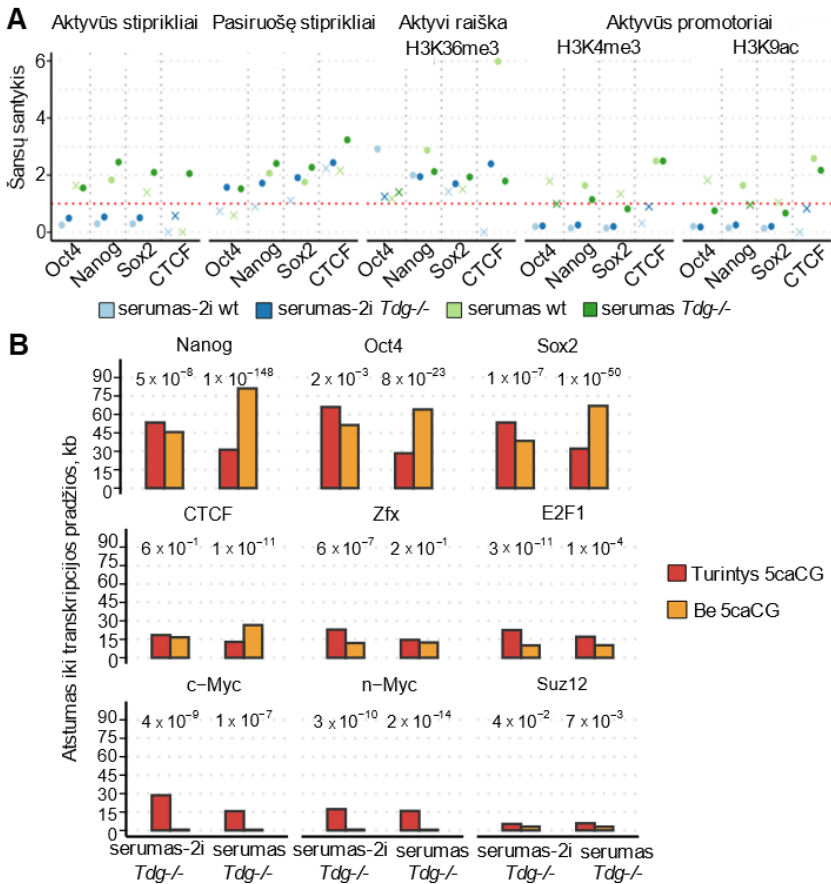


**3.11 pav.** 5caCG pasiskirstymas aplink įvairių veiksnių jungimosi regionus (normalizuota pagal CG tankį). Veiksnių jungimosi regionams apskaičiuoti naudoti

chromatino imunoišsodinimo sekoskaitos duomenys (Atlasi ir kt., 2019; Chen ir kt., 2008).

Abejų auginimo sąlygų ląstelių 5caC modifikuoti VJR (5caC-VJR) persidengė su H3K36me3 chromatino modifikaciją turinčiais (žyminčiais aktyvią raišką) ir pasiruošusių stipriklių regionais (3.12 pav. A). Priešingai nei serumo-2i mESC, serumo sąlygomis 5caC-VJR buvo praturtinti aktyviuose stiprikliuose ir aktyviems promotoriams būdinguose regionuose su H3K4me3/H3K9ac chromatino žymėmis. Yra žinoma, kad pliuripotentiškumo veiksniai gali persikelti nuo genams artimų į tolimus regionus vykstant ląstelių virsmui pereinant iš serumo į 2i auginimo sąlygas (Galonska ir kt., 2015; Tesar ir kt., 2007). Kadangi stiprikliai pritraukia veiksnis, paskaičiavome 5caC-VJR atstumą iki artimiausios anuotos transkripcijos pradžios vietos (TSS, angl. *transcription start site*). Nustatėme, kad daugumos tirtų veiksnų 5caC-VJR buvo išsidėstę toliau nuo TSS lyginant su 5caC nepraturtintais VJR, ir šie atstumai buvo žymiai ilgesni serumo-2i mESC (3.12 pav. B). Pažymėtina, kad serumo sąlygomis 5caCG modifikuotos pagrindinių pliuripotentiškumo veiksnų, Nanog, Oct4 ir Sox2 jungimosi vietos (bei CTCF) buvo arčiau TSS lyginant su 5caC nepraturtintais VJR.

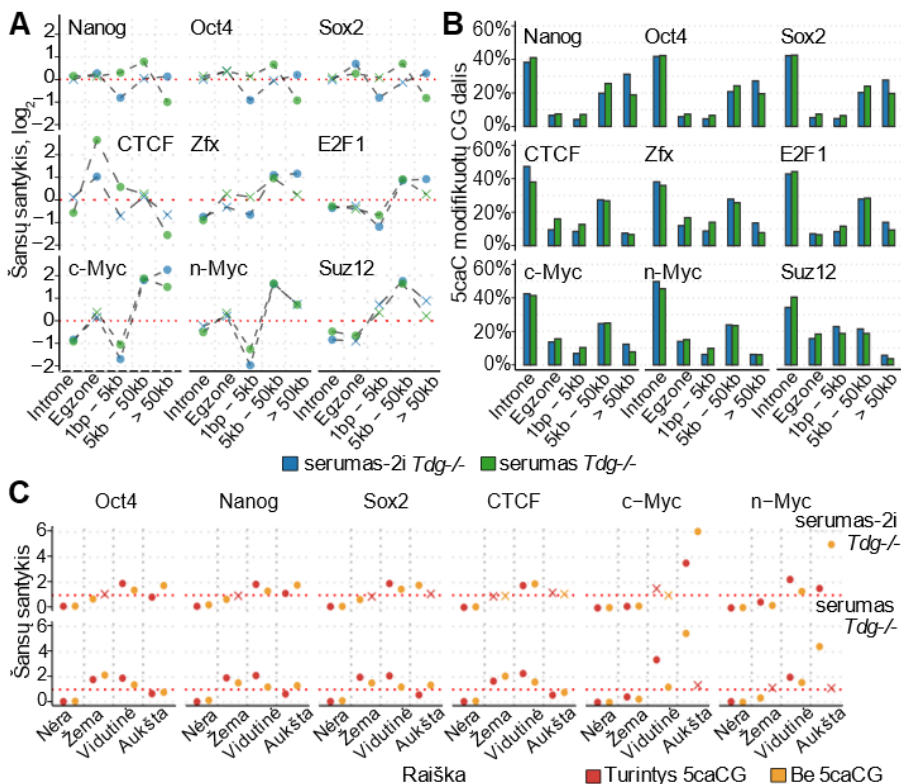




**3.12 pav.** 5caCG modifikuoti veiksmų jungimosi regionai (5caCG-VJR) dviejose mESC pluriopotentiškumo būsenose pasiskirsto skirtingai. **(A)** 5caCG-VJR praturtinimo šansų santykiai nustatyti Fišerio tiksluoju testu pasiruošusiuose ar aktyviuose stiprikluose, atitinkamai pažymėtuose H3K4me1 ar H3K4me1 ir H3K27ac, aktyviuose promotoriuose, pažymėtuose H3K4me3 ir H3K9ac, bei aktyvios raiškos regionuose, pažymėtuose H3K36me3. Statistiškai nereikšmingi ( $p \geq 0,05$ ) įverčiai pažymėti „×“. **(B)** 5caCG-VJR atstumų iki transkripcijos pradžios mediana lyginant su 5caG neturinčiais VJR. Reikšmės virš stulpelių rodo p vertes, paskaičiuotas naudojant Mann-Whitney testą.

Tuomet analizavome 5caCG-VJR praturtinimą ir skaičiavome 5caCG dalį (frakciją) nuo visų elemente esančių CG promotoriuose, genuose ir tarpgeninėse srityse, esant įvairiems atstumams iki TSS (3.13 pav. A, B). Abiejų būsenų mESC c-Myc, n-Myc, E2F1, Suz12 ir Zfx 5caCG-VJR buvo linkę pasiskirstyti tarpgeninėse srityse, tuo tarpu 5caCG-CTCF sritys buvo išskirtinai stipriai praturtintos egzonuose (3.13 pav. A). Kadangi veiksmų jungimosi vietose intronai ir egzonai (3.13 pav. B) buvo su žymia 5caCG frakcija, įvertinome VJR turinčių genų raiškos lygį (3.13 pav. C). Visi tirti 5caCG-VJR buvo praturtinti vidutinės raiškos genų grupėse, išskyrus c-Myc

jungimosi regionus serumo-2i mESC, kurie buvo praturtinti stiprios raiškos genuose. Vyraujanti c-Myc funkcija ESC yra pluriipotentiškumo palaikymas (Cartwright ir kt., 2005; Singh ir Dalton, 2009). Neseniai buvo pasiūlyta, kad c-Myc partneris Max veikia kaip 5caC jutiklis, o jo lygiai gali moduluoti Myc-Max valdomą transkripcijos tinklą (Wang ir kt., 2017). c-Myc jungiamų genų, kurių raiška padidėja serumo mESC, genų ontologijos (GO, angl. *gene ontology*) klasifikacija nenustatė kategorijų, susijusių su vystymosi procesais (Marks ir kt., 2012). Tačiau mūsų funkcinė analizė visgi parodė, kad genai su 5caCG-c-Myc jungimosi regionais turi sąsajas su vystymosi procesais serumo sąlygomis (GO terminai susiję su embriono vystymusi,  $p=3 \times 10^{-5}$ ). Todėl keliamo hipotezę, kad 5caCG modifikuotos sritys pritraukia c-Myc, kuris dalyvauja mESC keičiantis į labiau „apmokytą“ būseną, paruošiant ląsteles vystymosi procesų aktyvavimui.

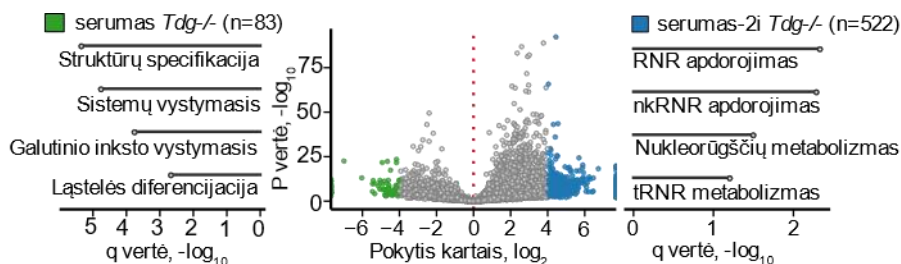


**3.13 pav.** 5caCG modifikuotų veiksmų jungimosi regionų (VJR) analizė. **(A)** Fišerio tiksliojo testo šansų santykiai ( $\log_2$ ) 5caCG-VJR praturtinimui geno elementuose bei genui artimuose ir tolimuose regionuose. **(B)** Veiksmų jungimosi regionų 5caCG frakcija genuose ir tarpgeniniuose regionuose esant skirtingam atstumui iki transkripcijos pradžios vietos. **(C)** 5caCG-VJR ir 5caCG neturinčių VJR praturtinimo pasiskirstymas skirtingose genų raiškos grupėse. Statistiškai nereikšmingi įverčiai ( $p \geq 0,05$ ) pažymėti „x“.

Prieš tai minėti duomenys parodė, kad 5caCG-CTCF regionai buvo praturtinti genuose. CTCF genome atlieka daugialypes funkcijas: skatina ilgo atstumo sąveikas tarp stipriklių ir promotorių bei izoluoja aktyvios raiškos sritis (Ong ir Corces, 2014; Phillips ir Corces, 2009). Daug pliuripotentiškumui svarbių genų-stipriklių sąveikų mESC, kurios paprastai pasikeičia vykstant diferenciacijai yra užtikrinamos jungiantis CTCF (Beagan ir kt., 2017). Nors 5caC modifikacijos lygiai aplink CTCF jungimosi regionus buvo didesni serumo mESC, abi pliuripotentiškumo būsenos rodė panašų 5caCG išsidėstymą – sritys šalia jungimosi regiono buvo labiau praturtintos 5caCG nei pats regionas (3.11 pav.). Be to, 5caCG-CTCF sritys buvo stipriai praturtintos aktyviems promotoriams būdingas histonų modifikacijas turinčiose vietose „apmokytoje“ mESC būsenoje (3.12 pav. A). Kad patikrintume genų, persidengiančių su 5caCG-CTCF jungimosi sritimis, biologines sąsajas abiejose ląstelių būsenose, atlikome GO funkcinės anotacijos analizę. Serumo mESC specifinėje genų grupėje buvo baltymai, dalyvaujantys vystymosi procesuose ( $p=5,5 \times 10^{-8}$ ) ir signalinių kelių reguliacijoje ( $p=6 \times 10^{-7}$ ), tuo tarpu serumo-2i mESC nustatėme baltymus, dalyvaujančius nukreipime į membranas ( $p=3,5 \times 10^{-5}$ ). Tai gali reikšti, jog galbūt pagal jungimosi regionų modifikavimo būseną CTCF gali moduluoti genų raišką, skatinant kitimą iš naivių mESC į paruoštas vystymuisi. Iš 5caCG modifikuotų pliuripotentiškumo veiksnių jungimosi regionų, tik Nanog regionus turintys genai rodė sąsajas su vystymosi procesais (teigiama ląstelės procesų reguliacija,  $p=3,5 \times 10^{-7}$ ; vystymosi procesas,  $p=4,7 \times 10^{-5}$ ).

Siekiant pažiūrėti, ar 5caC apibūdina pliuripotentiškumo būsenai specifinę genų raišką, lyginome 5caC praturtintus genus, specifinius naivioms ir „apmokytooms“ mESC. Kaip tikėtasi, daugiau 5caCG praturtintų genų nustatėme serumo-2i mESC (522 lyginant su 83 genais, ANOVA q-reikšmė  $< 0,05$ ; absoliutus  $\log_2$  pokytis kartais  $>4$ ; 3.14 pav.). Tačiau šiame dideliame genų rinkinyje GO funkcinė analizė neparodė stiprių sąsajų su jokiais kategorijomis išskyrus silpną ryšį su nukleorūgščių metaboliniais procesais. Tuo tarpu tarp serumo sąlygomis augintoms mESC specifinių 5caC praturtintų genų buvo DNR jungiantys transkripcijos faktoriai, daugialąsčio organizmo vystymąsi ir ląstelės lemties specifikaciją reguliuojančiuose procesuose dalyvaujančios baltymų grupės (3.14 pav.). Genai, kurių raiška 2i sąlygomis mESC yra didesnė nei serumo sąlygomis, yra susiję su metaboliniais procesais, o serumo mESC padidėjusios raiškos genai (kurių raiškos nėra arba ji mažesnė 2i sąlygomis) yra susiję su vystymosi procesais (Marks ir kt., 2012). Kadangi aptikome 5caC ryšį su aktyviais ir mESC būsenai būdingais

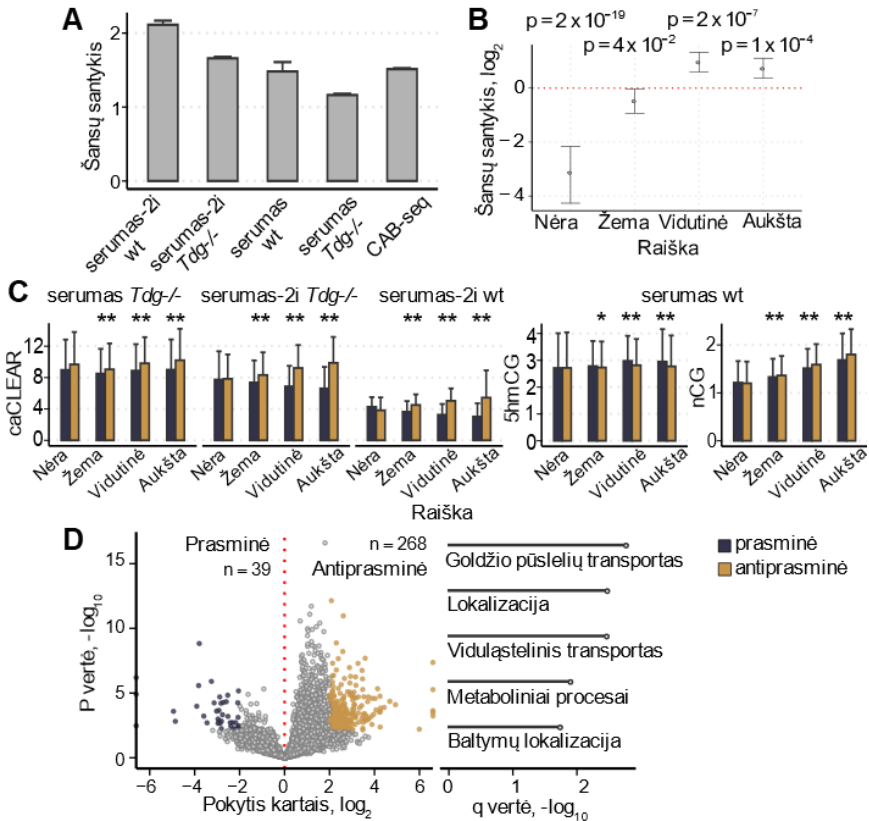
genais, tai rodo aktyvaus DNR demetilavimo dalyvavimą pliuripotentiškumo būsenų sąlygojime.



**3.14 pav.** Vulcano grafikas, rodantis 5caCG praturtintų genų pasiskirstymą dviejų pliuripotentiškumo būsenų mESC. Funkcinė anotacijos analizė parodė reikšmingiems genų rinkiniams: 83 ir 522 genai parodyti žaliai ir mėlynai atitinkamai serumo ir serumo-2i *Tdg*<sup>-/-</sup> ląstelėms (ANOVA F testo  $q$  reikšmės  $<0,05$ ; absoliutus  $\log_2$  pokytis kartais  $>4$ ). nkRNR – nekoduojančios RNR.

### 3.3.3 5caCG pasiskirsto antiprasminėje aktyvių genų grandinėje

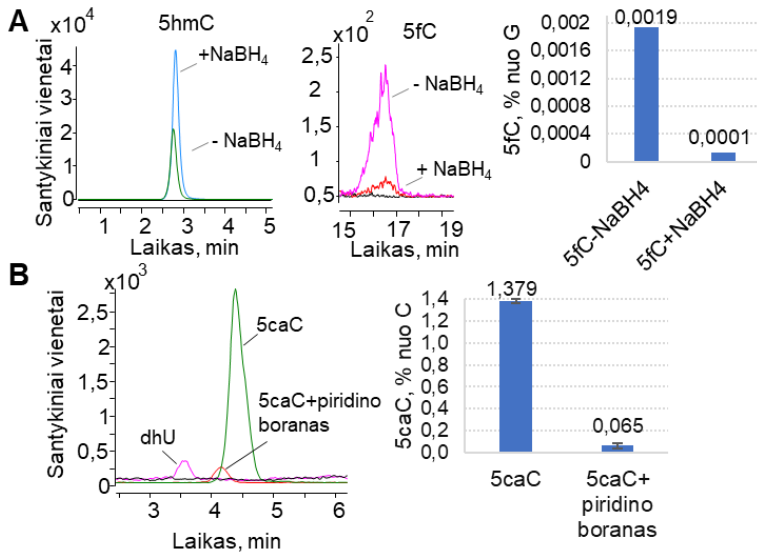
Yra žinoma, kad 5mC DNR grandinių CG taikiniuose pasiskirsto simetriškai dėl DNR metilavimo palaikymo mechanizmo, tuo tarpu didelė dalis 5hmC ir 5fC CG taikiniuose pasiskirsto asimetriškai (Booth ir kt., 2014; Wu ir kt., 2014; Yu ir kt., 2012). MAB-seq metodu 5fCG/5caCG taikiniuose buvo nustatyta stipri modifikavimo asimetrija: tik ~5% nustatytų CG taikinių (22 590 iš 454 400 CG diadų) buvo simetriškai modifikuoti 5fC/5caC (Wu ir kt., 2014). Kadangi 5caC ir 5fC padėtys genome persidengia mažai (Lu ir kt., 2015b), nusprendėme ištirti 5caCG pasiskirstymą genomines DNR grandinėse abejomis ląstelių auginimo sąlygomis. Pirmiausia, naudojant lambda genomines DNR eksperimentą (parodytą 3.7 pav. B ir C) patikrinome, ar nėra galimo nuo grandinės priklausančio asimetrinio eM.SssI vykdomo CG žymėjimo pasiskirstymo – neaptikome jokių grandinei specifinių caCLEAR signalo skirtumų CG taikiniuose ( $p=0,18$ , dvipusis porinis t testas).



**3.15 pav.** 5caCG pasiskirsto antiprasminėje aktyvios raiškos genų grandinėje. (A) caCLEAR 5caCG praturtinimo šansų santykis baltymus koduojančių genų antiprasminėje grandinėje wt ir *Tdg*<sup>-/-</sup> ląstelių linijose. Viesiems parodytiems įverčiams  $p < 2,2 \times 10^{-16}$ . 5caCG praturtinimas antiprasminėje genų grandinėje taip pat aptiktas 2 803 5caCG, nustatytiems bisulfitiniu modifikavimu paremtu metodu CAB-seq (Lu ir kt., 2015b). (B) Genų su 5caCG grandinių asimetrija praturtinimas skirtingose RNR raiškos grupėse serumo-2i *Tdg*<sup>-/-</sup> mESC. Reikšmės viršuje nurodo p vertes, paskaičiuotas Fišerio tiksluoju testu. (C) 5caCG, 5hmCG ir nemodifikuotų CG asimetrinis pasiskirstymas tarp grandinių skirtingose genų raiškos grupėse. 5hmCG ir nCG duomenys parodyti serumo wt mESC. Dvipusio porinio t testo p reikšmės nurodytos žvaigždutėmis (\* –  $p < 0,05$ ; \*\* –  $p < 1 \times 10^{-7}$ ). (D) Kairėje, vulkano grafikas, rodantis genų su 5caCG asimetrija pasiskirstymą prasminėje ir antiprasminėje (statistiškai reikšmingi genai atitinkamai pažymėti tamsiai mėlyna ir oranžine spalva) grandinėje serumo-2i mESC. Dešinėje, GO funkcinė anotacijos analizė, parodyta genams su 5caCG asimetrija antiprasminėje grandinėje (ANOVA F testo q reikšmės  $< 0,05$ ; absoliutus log<sub>2</sub> pokytis kartais  $> 2$ ).

Siekiant mESC kiekybiškai įvertinti 5caCG pasiskirstymą grandinėse, pirmiausia analizavome baltymus koduojančius genus, kurie yra praturtinti 5caCG. Abiejų wt ir *Tdg*<sup>-/-</sup> ląstelių linijų analizė parodė asimetrinį 5caCG taikinių pasiskirstymą ( $p < 2,2 \times 10^{-16}$ , Fišerio tikslusis testas; 3.15 pav. A)

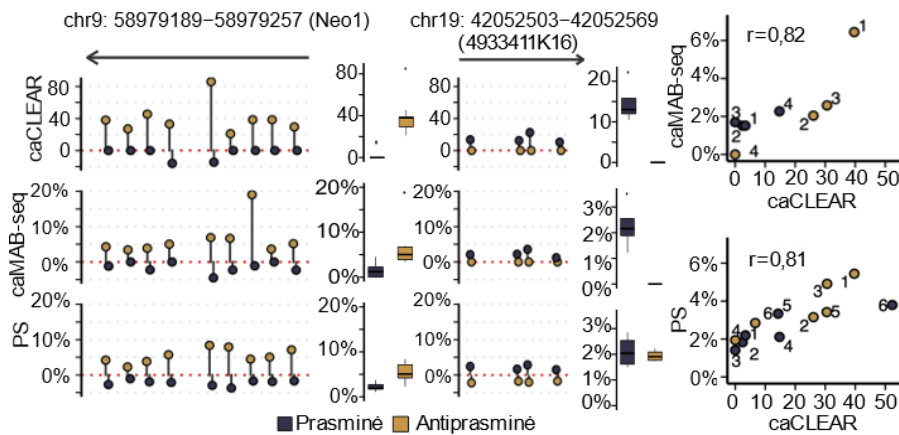
antiprasminėje (nekoduojančioje) transkribuojamų genų grandinėje (3.15 pav. C). 5caCG grandinių asimetrija buvo ryškesnė serumo-2i mESC stiprios raiškos genuose (3.15 pav. A, C). Be to, įvertinome vidutinius nemetilintų CG ir 5hmCG lygius tuose pačiuose serumo wt mESC genuose, naudojant mūsų laboratorijos TOP-seq metodo (Staševskij ir kt., 2017) ir 5hmC analizei sukurto hmTOP-seq metodo duomenis (Gibas ir kt., 2020). Panašiai kaip ir 5caCG, didesnės raiškos genuose nCG gausumas buvo didesnis antiprasminėje grandinėje, bet 5hmCG buvo linkę pasiskirstyti prasminėje tų pačių genų rinkinio grandinėje (3.15 pav. C). Tai sutampa su paskelbtais pelės ir žmogaus audinių 5hmC duomenimis (Booth ir kt., 2012; Gibas ir kt., 2020; Wen ir kt., 2014) ir parodo, kad 5mC demetilimas yra skirtingai reguliuojamas abiejose geno grandinėse. Stebėtą 5caC pasiskirstymo asimetriją baltymus koduojančiuose genuose taip pat nustatėme analizuojant 2 803 5caCG, identifikuotus CAB-seq metodu (Lu ir kt., 2015b) serumo *Tdg*<sup>-/-</sup> mESC (3.15 pav. A). Pažymėtina, jog reikšmingų grandinei specifinių 5caCG lygio skirtumų nenustatėme kituose geno elementuose (CG salose, pasikartojančiose sekose, antiprasminėse RNR, ilgose nekoduojančiose RNR, pseudogenuose). Tai rodo mechanizmų, palaikančių 5caC pasiskirstymo asimetriją išskirtinai baltymus koduojančiuose genuose, egzistavimą. Naudojant serumo-2i *Tdg*<sup>-/-</sup> mESC duomenis, suskaičiavome genus su reikšmingu grandinei specifiniu 5caC lygiu: vulkano grafikas parodė poslinkį link antiprasminės grandinės (3.15 pav. D). Nustatėme 268 ir 39 genus su didesniu 5caCG gausumu atitinkamai antiprasminėje ir prasminėje grandinėse (absoliutus log<sub>2</sub> pokytis kartais >2; ANOVA q reikšmė <0,05; 3.15 pav. D). Genai praturtinti 5caCG antiprasminėje grandinėje sutapo su esančiais stiprios raiškos grupėje (3.15 pav. B) ir buvo daugiausia susiję su metabolinėmis ir viduląstelinio transporto kategorijomis GO funkcinėje analizėje (3.15 pav. D). Prasminės grandinės 5caC asimetriškų genų rinkinyje nebuvo rasta reikšmingų GO terminų.



**3.16 pav.** 5caC ir 5fC konversijos įvertinimas caMAB-seq ir piridino borano sekoskaitai atliekant HPLC-MS/MS analizę. **(A)** 5fC redukcijos į 5hmC naudojant NaBH<sub>4</sub> įvertinimas serumo-2i *Tdg*<sup>-/-</sup> mESC DNR. Kairėje, nukleozidų masių signalų chromatogramos rodo efektyvią 5fC konversiją (rožinė rodo lygį prieš poveikį, raudona – po), o natūraliai esančio 5hmC kiekiai padidėjo. Juoda spalva rodo kontrolinę nemodifikuotą DNR. Dešinėje grafikas rodo 5fC % nuo viso guanino prieš ir po redukcijos NaBH<sub>4</sub>. Aptikti likutiniai 5fC kiekiai rodo efektyvią 5fC konversiją į 5hmC genomineje DNR. **(B)** Lambda DNR 5caC konversijos į dihidrouracilą (dhU) piridino boranu įvertinimas. Kairėje, nukleozidų masių chromatogramos rodo efektyvią 5caC konversiją (žalia rodo 5caC signalą prieš poveikį, raudona – po) į dhU (rožinė, juoda rodo signalą prieš poveikį). Dešinėje grafikas rodo 5caC santykį su visu citoziniu prieš ir po piridino borano reakcijos. Likutiniai 5caC kiekiai rodo efektyvią 5caC konversiją genomineje DNR.

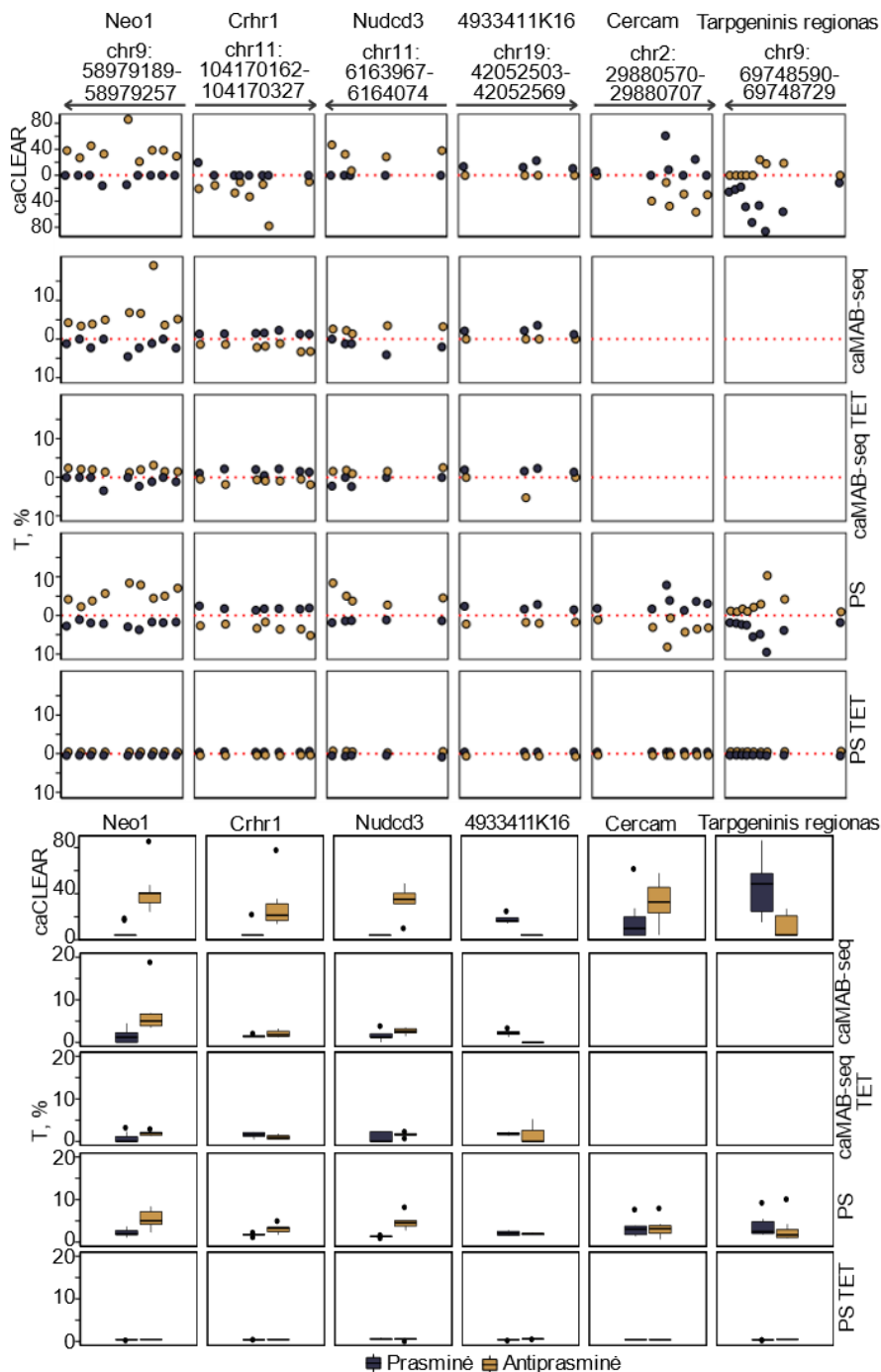
Siekiant patvirtinti caCLEAR stebėtą 5caCG pasiskirstymo asimetriškumą, įvertinome 5caC lygius keturiuose pasirinktuose geno lokusuose naudodami 5caC nustatymui pritaikytą regionui specifinį MAB-seq metodą (caMAB-seq, žr. 1.4.4, 2.2.12; Wu ir kt., 2016). Kad šioje bisulfitinėje sekoskaitoje atskirtume 5caC nuo 5fC, 5fC buvo paverstas į 5hmC naudojant natrio borohidrido redukciją (Song ir kt., 2013). Tuo tarpu nemodifikuoti CG taikiniai buvo metilinti wt *SssI* kaip caCLEAR metilinimo protokole. Po redukcijos HPLC-MS/MS analize nustatyti tik labai maži 5fC kiekiai, lyginant su pradiniu, rodė efektyvų (~95%) 5fC pavertimą į 5hmC genomineje DNR (3.16 pav. A). Po metilinimo ir redukcijos žingsnių bei natrio bisulfito konversijos 5caC buvo nuskaitomas kaip T, tuo tarpu visos kitos citozino būsenos CG kontekste liko atsparios natrio bisulfito poveikiui ir buvo nuskaitomos kaip C. Be to, palyginimui su caCLEAR pasitelkėme natrio bisulfito nenaudojančio piridino borano sekoskaitos metodo (PS, žr. 1.4.5,

2.2.13; Liu ir kt., 2019) variantą, kuris leidžia nustatyti 5caC (tuose pačiuose 4 ir dar dviejuose papildomuose regionuose, 3.18 pav.). 5fC redukavus į 5hmC, visos citozino būsenos atpažįstamos kaip C, išskyrus 5caC, kuris paverčiamas į dihidrouracilą ir nuskaitomas kaip T. HPLC-MS/MS analizė parodė efektyvią 5caC konversiją (~95%, 3.15 pav. B). Grandinei specifinė caMAB-seq ir PS analizė patvirtino caCLEAR matomą 5caCG asimetriją pasirinktuose lokusuose (3.17, 3.18 pav.). Visi trys metodai rodė gerą sutapimą šioje pavienių CG analizėje, caCLEAR rodė aukštą koreliaciją su kitais metodais (0,81-0,82; 3.16 dešinėje, caMAB-seq ir PS tarpusavio koreliacija buvo 0,79), o tai patvirtina caCLEAR metodo veiksmingumą. Be to, caCLEAR gebėjo nustatyti itin žemus vos kelių % 5caC lygius, nustatomus kitais metodais (3.16, 3.17 pav.), o tai parodo didelį metodo jautrumą.



**3.17 pav.** caCLEAR metodu matomos asimetrijos patvirtinimas kitais metodais. Kairėje, *Neo1* ir *4933411K16* genų lokusų (kurie buvo nustatyti kaip turintys 5caCG asimetrinį pasiskirstymą atitinkamai antiprasminėje ir prasminėje grandinėse) caCLEAR padengimas ir 5caC procentai abiejų grandinių CG taikiniuose, nustatyti caMAB-seq ir piridino borano sekoskaita (PS). Parodytos lokusų koordinatės genome bei geno kryptis. Stačiakampės diagramos šalia rodo signalo pasiskirstymą kiekvienoje regiono grandinėje, iš kurio matyti tarp grandinių esantis skirtumas. Dešinėje, visų analizuotų lokusų (parodytų skaičiais 1-6, žr. 3.17 pav.) vidutinio grandinės 5caC signalo koreliacija tarp caCLEAR ir caMAB-seq arba PS (koreliacijų p vertės <0,014). Taip pat žr. 3.18 pav.





**3.18 pav.** Lokusui specifinė caCLEAR identifikuotų 5caCG analizė caMAB-seq (5caC%, nuskaitomas kaip T) bisulfitinė sekoskaita ir piridino borano sekoskaita (PS) serumo-2i *Tdg*<sup>-/-</sup> pelės ESC. caMAB-seq buvo analizuoti 4 genų lokusai: 4933411K16, *Crhr1*, *Neo1* ir *Nudcd3*; PS papildomai buvo analizuotas tarpgenis

regionas ir *Cercam* genas. Taip pat parodyta caMAB-seq ir PS analizė TET TKO embrioninėse kamieninėse ląstelėse – caCLEAR šių ląstelių regionuose neidentifikuoja 5caC turinčių CG. Stačiakampės diagramos apačioje rodo signalo pasiskirstymą kiekvienoje regiono grandinėje, iš kurio matyti tarp grandinių esantis skirtumas. Taip pat žr. 3.17 pav.

### 3.4 Oksiduotų 5-metilcitozino formų tyrimas *Laccaria bicolor* ir *Coprinopsis cinerea* papėdgybiuose

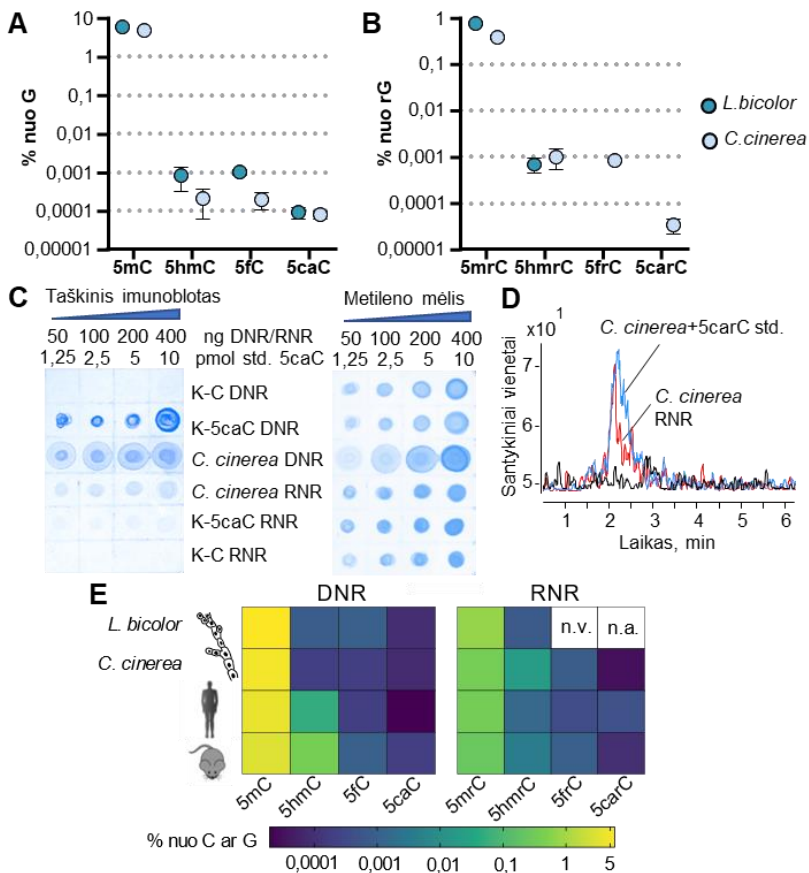
Oksiduotos 5-metilcitozino formos plačiai tiriamos žinduoliuose, tuo tarpu kitos filogenetinės grupės dar mažai ištirtos. *Agaricales* eilės papėdgybiai *Laccaria bicolor* ir *Coprinopsis cinerea* turi daug *TET-JPB* šeimos genų (Iyer ir kt., 2009, 2013, 2014), spėjama, kad dauguma jų galėtų koduoti katalitiškai aktyvius baltymus (47 iš 74 ir 32 iš 47 homologų atitinkamai *L. bicolor* ir *C. cinerea*). Šie grybai turi nedidelius genomus, kurių seka yra nuskaityta (~37 Mb *C. cinerea* ir ~61 Mb *L. bicolor*; Martin ir kt., 2008; Stajich ir kt., 2010) ir surinkta (*L. bicolor* iš dalies), todėl jie yra patogūs tyrimams. Be to, yra publikuoti viso genomo DNR metilinimo duomenys (Zemach ir kt., 2010). Saprotrofinis *C. cinerea* yra naudojamas tyrimams kaip daugialąsčių grybų modelinis organizmas, jį galima lengvai auginti laboratorinėmis sąlygomis (Pukkila, 2011; Stajich ir kt., 2010). *L. bicolor* yra svarbus kietmedžio ir spygliuočių rūšių ektomikorizinis simbiotas, kuris yra naudojamas simbiozės sąveikos ir ekologinės adaptacijos genams tirti (Daguerre ir kt., 2017; Martin ir kt., 2008). Dėl esminio grybų vaidmens medžių augimui ir pagrindinių maistinių medžiagų apykaitai, ektomikorizinė simbiozė turi visuotinę ekologinę ir ekonominę reikšmę (Martin ir Selosse, 2008).

*C. cinerea* DNR buvo aptiktos oksimC, gausiausia iš jų – 5caC modifikacija (Chavez ir kt., 2014). Bendras visų *C. cinerea* oksimC išsidėstymas genome buvo nustatytas naudojant SMRT sekoskaitą ir papildomai, atliekant citozino 5-metilensulfonato imunoišsodinimą, buvo nustatytas 5hmC (Chavez ir kt., 2014). Tačiau atskirų oksimC pasiskirstymas genome ir jų poveikis genų raiškai *C. cinerea* nebuvo išsamiai ištirtas, o *L. bicolor* oksimC dar visai netirtos ir nežinoma, kaip jų pasiskirstymas veikiamas didesnio skaičiaus *TET-JPB* homologų genų. Siekėme ištirti oksimC pasiskirstymą ir funkcijas šiuose grybuose, tam pritaikant laboratorijoje kuriamus TOP-seq paremtus sekoskaitos metodus.

### 3.4.1 Oksiduotų 5-metilcitozino formų nustatymas

#### *L. bicolor* ir *C. cinerea*

Pirmiausia, HPLC-MS/MS ir taškinio imunobloto metodais įvertinome oksiduotų 5-metilcitozino formų kiekius genominiame DNR, išskirtoje iš *L. bicolor* ir *C. cinerea* grybų haploidinio micelio. HPLC-MS/MS metodu *L. bicolor* nustatėme 4-5 kartus didesnius 5hmC ir 5fC kiekius (~0,0008 ir ~0,001% nuo G), nei *C. cinerea* (~0,0002% 5hmC ir 5fC; 3.19 pav. A). Abiejuose grybuose 5hmC ir 5fC kiekiai buvo panašūs tarpusavyje, skirtingai nuo žinduolių (3.19 pav. A ir E), kuriuose 5hmC yra pati gausiausia oksiduota 5mC forma (pavyzdžiui, ~0,06% 5hmC, ~0,0007% 5fC, ~0,0002% 5caC nuo C pelės embrioninėse kamieninėse ląstelėse; Ito ir kt., 2011). Tuo tarpu 5caC lygiai buvo žemiausi iš visų oksimC (~0,0001%, 3.19 pav. A). Šie rezultatai nesutapo su publikuotais duomenimis, kad 5caC yra gausiausia oksimC (Chavez ir kt., 2014), bei taškiniu imunoblotu, kuriuo stebėjome stiprų 5caC signalą *C. cinerea* DNR naudojant antikūnus prieš 5caC (3.19 pav. C). Todėl iškėlėme hipotezę, kad galbūt 5caC signalas, stebimas taškiniu imunoblotu, atsiranda dėl 5caC-antikūnų reagavimo su priemaišomis iš grybų ekstraktų arba priemaišinės likusios RNR, kuri taip pat gali turėti 5caC. Iš tiesų, taškiniu imunoblotu stebėjome 5caC signalą *C. cinerea* RNR (5scarC) (3.19 pav. C) ir patvirtinome jo buvimą HPLC-MS/MS (3.19 pav. B, D). HPLC-MS/MS taip pat nustatėme 5mrC, 5hmrC ir 5frC (3.19 pav. B), o stebėti lygiai buvo panašūs į aptinkamus žinduolių RNR (% nuo G: ~0,25% 5mrC; ~0,004% 5hmrC; ~0,0001% 5scarC; Huang ir kt., 2016; taip pat žr. Huber ir kt., 2015; 3.19 pav. E). HPLC-MS/MS DNR ir RNR nustatyti 5caC kiekiai buvo daug mažesni, nei būtų galima tikėtis iš taškinio imunobloto, o tai rodo galimą kryžminį antikūnų reaktyvumą su nežinomais grybų ekstraktų komponentais. Be to, yra žinoma, kad IgG antikūnai gali atpažinti trumpus DNR sekų pasikartojimus (Lentini ir kt., 2018).



**3.19 pav.** Oksiduotos 5-metilcitozino formos (oksi-mC) *L. bicolor* ir *C. cinerea* papėdgybiuose. **(A, B)** Citozino modifikacijų, analizuotų kiekybine HPLC-MS/MS, paplitimas *L. bicolor* ir *C. cinerea* **(A)** DNR ir **(B)** RNR. 5caC, 5fC, 5hmC ir 5mC % nuo viso G buvo apskaičiuoti naudojant atitinkamo nukleozido kalibracinę kreivę. Buvo naudojami bent du biologiniai pakartojimai, išskyrus 5fC RNR (5frC), kuris buvo kiekybiškai įvertintas tik vienoje biologinėje replikoje. Oksi-mC nukleozidai buvo papildomai išgryninti HPLC frakcionavimu (žr. 2.2.10) prieš tikslinę HPLC-MS/MS analizę (išskyrus 5hmrC *L. bicolor* RNR). *L. bicolor* įprastomis sąlygomis 5carC signalas RNR nebuvo aptiktas, 5frC nebuvo vertintas. **(C)** Kairėje, 5caC nustatymas taškiniu imunoblotu nukleorūgštyse, išskirtose iš *C. cinerea* micelio. Nurodyti naudotos DNR arba RNR kiekiai (ng) bei 5caC kiekiai (pmol) kontrolinėse nukleorūgštyse. K-C – kontrolinės nmodifikuotos nukleorūgštys, K-5caC – modifikuotos kontrolinės nukleorūgštys. Dešinėje, dažymas metileno mėliu, siekiant įvertinti užneštus nukleorūgščių kiekius (dažas gali nudažyti ir kitas neigiamai įkrautas molekules, kurios, pavyzdžiui, gali atsirasti iš grybų ląstelių sienelių priemaišų). **(D)** 5caC MS/MS signalo chromatograma *C. cinerea* RNR (raudona). Signalas padidėja į *C. cinerea* mėginį pridėjus 5caC standarto (25 fmol, mėlynas), o tai patvirtina jo specifiskumą. Signalas nmodifikuotoje RNR parodytas juoda spalva. **(E)** grybų *L. bicolor* ir *C. cinerea* bei žinduolių modifikuotų citozinių gausumo DNR ir visuminėje RNR palyginimas, išreikštas % nuo viso G arba C (grafiką sugeneravo P. Gibas). Pelės vidutiniai apytiksliai įvairių audinių DNR duomenys buvo paimti iš

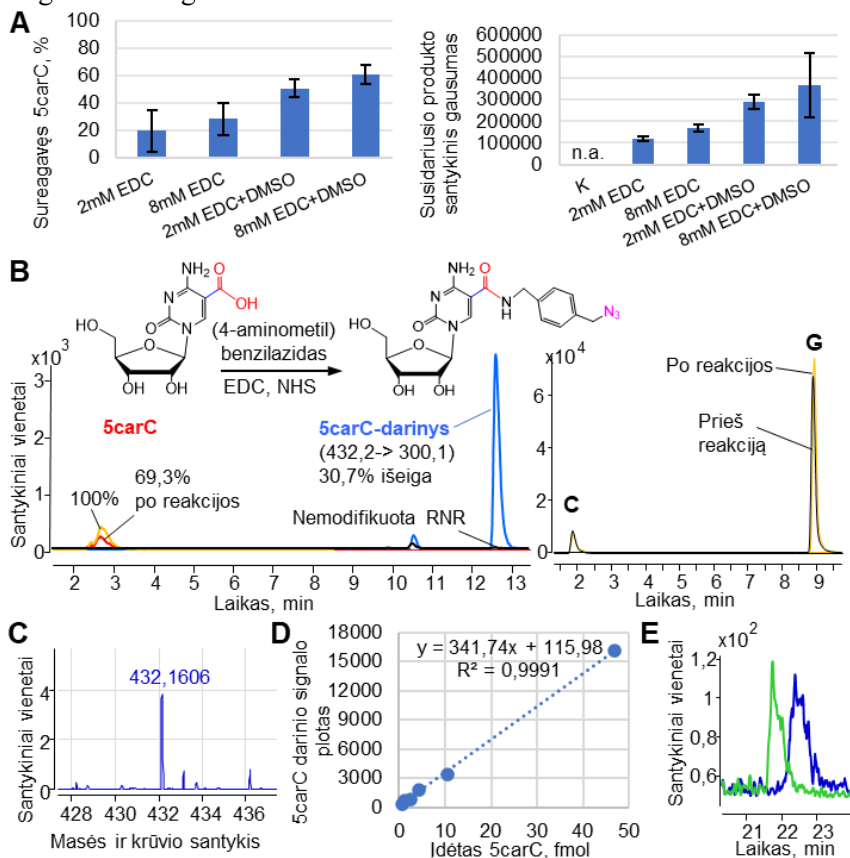
Ito ir kt. (2011), žmogaus iš Tang ir kt. (2015), RNR duomenys paimti iš Huang ir kt. (2016). Grybų DNR ir RNR buvo naudoti neveiktų mėginių (be vitamino C ar derivatizacijos) duomenys (5frc *L. bicolor* nebuvo vertintas (n.v.), o 5scarC neaptiktas (n.a.)).

### 3.4.2 5caC pasiskirstymas RNR

Aptikus 5scarC signalą *C. cinerea* RNR, nusprendėme ištirti šios modifikacijos pasiskirstymą ir gausumą abiejų grybų RNR. Ši modifikacija dar labai mažai tirta, tik neseniai nustatyta žinduolių RNR (Huang ir kt., 2016; Tardu ir kt., 2019) ir jos funkcijos nežinomos, kai tuo tarpu apie kitas RNR oksiduotas 5-metilcitozino formas duomenų yra daugiau (Huber ir kt., 2015; Roundtree ir kt., 2017; Wang ir kt., 2021; Zhang ir kt., 2016).

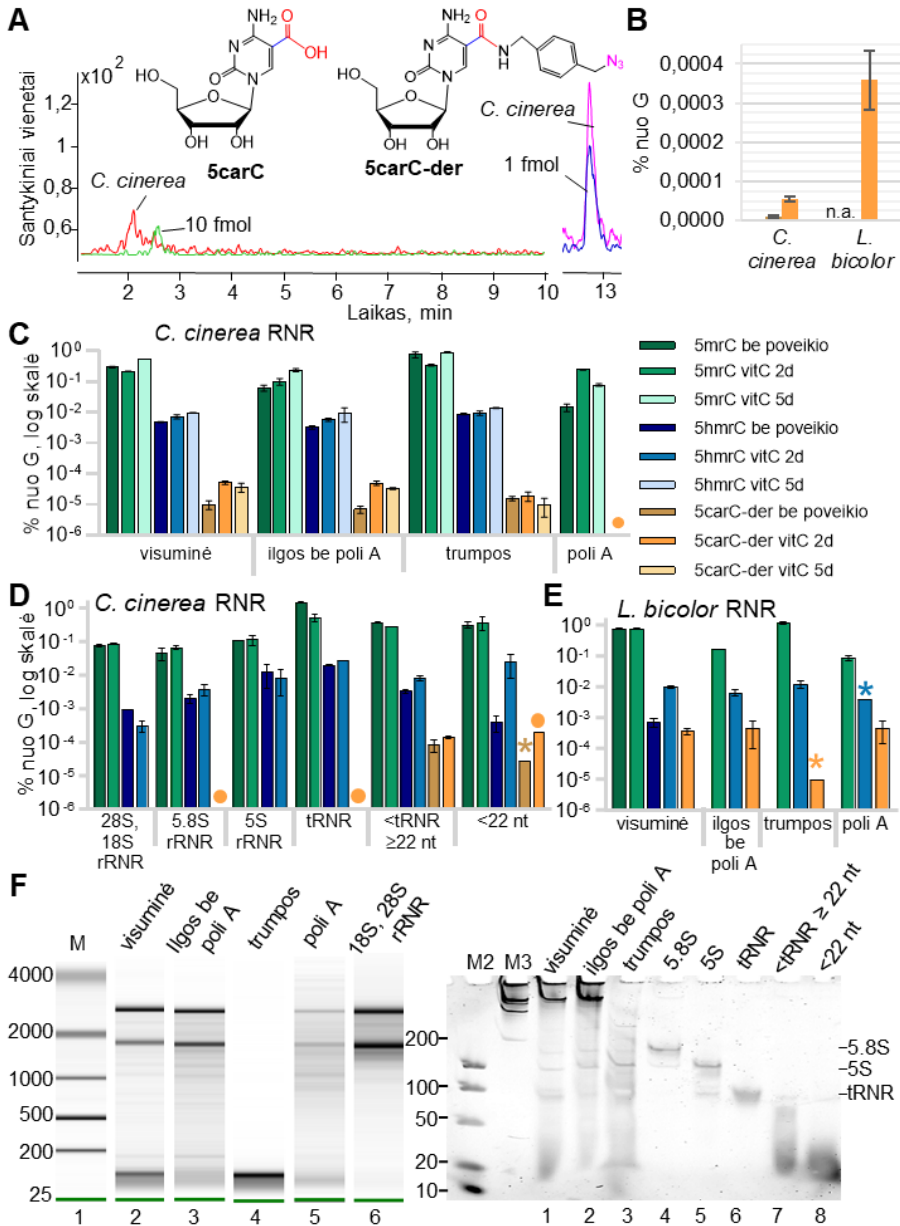
5scarC tyrimui RNR pasitelkėme HPLC-MS/MS metodą. 5scarC kiekybinis įvertinimas yra sudėtinga užduotis: reikalingi dideli RNR kiekiai (~20 µg) bei norint 5scarC atskirti nuo gausaus citozino, turinčio panašų HPLC išplovimo laiką, prieš tikslią HPLC-MS/MS analizę turi būti atliekamas papildomas nukleozidų frakcionavimas ir gryninimas HPLC (žr. 1.3, 2.2.10). Yra žinoma, kad įvairių retų modifikuotų nukleozidų HPLC-MS/MS analizę galima palengvinti atlikus jų cheminį modifikavimą, po kurio chromatografijoje įvykus atsiskyrimui nuo kitų gausių nukleozidų ar pagerėjus junginio jonizacijai gali smarkiai sustiprėti produkto MS/MS signalas (Li ir kt., 2018). Siekiant palengvinti 5scarC aptikimą, pasitelkėme 5caC DNR sekoskaitai naudojamą cheminę derivatizaciją (Lu ir kt., 2013, 2015b), kuriame prie karboksilo grupės prijungiamas (4-aminometil)benzilazidas (sintezę atliko prof. Viktoras Masevičius ir dr. Gražina Petraitytė). Optimizavus sąlygas (optimizavimą atliko Eglė Godliauskaitė), sintetinėje RNR sureagavo ~61% 5scarC (3.20 pav. A), panašiu efektyvumu kaip ir 5caC DNR (Lu ir kt., 2013) bei susidarė reikalingas produktas (3.20 pav. B, C). Pažymėtina, kad 5scarC darinio MS signalo intensyvumas padidėjo ~20 kartų, lyginant su paprastu 5scarC (3.20 pav. B), o dėl skirtingo išėjimo laiko nuo kitų nukleozidų, papildomas nukleozidų frakcionavimas nebebuvo reikalingas. 5scarC darinio MS/MS signalas rodė tiesinę priklausomybę nuo pradinio 5scarC kiekio mėginiuose (3.20 pav. D). 5scarC darinio signalas buvo selektyvus ir reakcijos efektyvumas nepriklausė nuo RNR kiekio, nes panašaus intensyvumo signalas buvo gerai aptinkamas mėginiuose, sumaišytuose su skirtingais nemodifikuotos RNR kiekiais (3.20 pav. E). Kadangi vidutiniškai buvo galima aptikti net 0,6 fmol derivatizuoto 5scarC, sukurtas jautrus nustatymo

metodas gali būti tinkamas 5carC matavimams grybuose ir kituose įvairiuose biologiniuose mėginiuose.



**3.20 pav.** Cheminė derivatizacija su (4-aminometil)benzilazidu 5carC nustatymui RNR. **(A)** 5carC derivatizacijos optimizavimas sintetinėje 5carC turinčioje RNR įvertintas naudojant nukleozidų HPLC-MS/MS. Kairėje, sureagavusio 5carC procentas skirtingomis reakcijos sąlygomis (žr. 2.2.19). K – kontrolė be derivatizacijos, n.a. – signalas neaptiktas. Dešinėje, 5carC darinio signalo intensyvumas. **(B)** Kairėje, 5carC (oranžinė prieš ir raudona, po reakcijos) ir jo derivatizacijos produkto (mėlynas) MS/MS signalai. Nomodifikuota RNR nerodo derivatizacijos produkto signalo (juoda). Po reakcijos liko 69,3% nesureagavusio 5carC; derivatizacija žymiai padidina 5carC signalą, kaip matyti iš stipresnio produkto signalo, atitinkančio 30,7% pradinio 5carC, sureagavusio su (4-aminometil)benzilazidu. Dešinėje, C ir G MS/MS signalai prieš (juoda) ir po reakcijos (geltona) rodo vienodą mėginio užnešimą. **(C)** 5carC darinio didelės skiriamosios gebos masės spektras. Nustatytas tikslus 5carC nukleozido darinio masės ir krūvio santykis (teorinis m/z 432,1626). **(D)** 5carC darinio MS/MS signalo plotas parodė tiesinį padidėjimą ir didelę koreliaciją priklausomai nuo įdėtų 5carC kiekių mėginiuose prieš reakciją. **(E)** 5carC derivatizacijos efektyvumas nepriklauso nuo nemodifikuotos RNR kiekio mėginyje, nes 5carC darinio MS/MS signalas yra labai panašus, kai 0,8 μg (mėlyna) arba 8 μg (žalia) nemodifikuotos RNR buvo sumaišyta su RNR, turinčia 20 fmol 5carC.

Esant standartinėms abiejų grybų auginimo sąlygoms, derivatizacija leido aptikti 5carC tik *C. cinerea* visuminėje RNR (3.21 pav. A, B). Norėdami padidinti 5carC kiekį, abu grybus paveikėme vitaminu C (vitC), kuris sustiprina TET fermentų oksidacinį aktyvumą žinduolių ląstelėse nekeičiant jų RNR lygio (Minor ir kt., 2013; Yin ir kt., 2013). Kaip ir tikėtasi, poveikis vitC sustiprino signalą *C. cinerea* (3.21 pav. A, C) bei sukėlė labai stiprų 5carC darinio signalą *L. bicolor* (3.21 pav. A, E). Tai patvirtino 5carC buvimą *C. cinerea* (3.21 pav. A) ir parodė, kad 5carC gali atsirasti *L. bicolor* tam tikromis biologinėmis aplinkybėmis.



**3.21 pav.** Kiekybinis 5carC nustatymas HPLC-MS/MS naudojant cheminę derivatizaciją su (4-aminometil)benzilazidu *L. bicolor* ir *C. cinerea* grybų RNR. **(A)** 5carC (raudona, ~20 µg RNR) ir 5carC darinio (rožinė, ~5 µg RNR) MS/MS signalas *C. cinerea* RNR. 5carC standartas parodytas žaliai (10 fmol), po derivatizacijos susidaręs 5carC produktas (5carC-der) mėlynai (1 fmol). **(B)** 5carC aptikimas RNR naudojant HPLC-MS/MS ir cheminę derivatizaciją. 5carC darinio kiekiai buvo įvertinti skaičiuojant jo % nuo G prieš ir po veikimo vitaminu C (vitC) 2 dienas. Naudoti du biologiniai pakartojimai, signalas neaptiktas (n.a.) *L. bicolor* be vitC. Kalibravimui buvo naudojama sintetinė 5carC turinti RNR. Chemiškai derivatizuoto 5carC HPLC-MS/MS analizė skirtingose *C. cinerea* (**C**, **D**) ir *L. bicolor* (**E**) RNR



frakcijose. Grybai buvo veikiami vitaminu C (vitC) 2 (C, D, E) arba 5 dienas (C). Neveiktos *L. bicolor* RNR nebuvo frakcionuota, nes nerodė 5carC signalo. Buvo naudojami du biologiniai pakartojimai. Žvaigždutės žymi mėginius su aptiktu ir kiekybiškai įvertintu signalu (tiesiniame kalibracinės kreivės diapazone) tik viename iš dviejų pakartojimų; apskritimai žymi mėginius, kur signalas buvo aptinkamas, bet negalėjo būti kiekybiškai įvertintas (nebuvo kalibracinės kreivės linijiniame diapazone) viename iš dviejų pakartojimų; mėginiai, kurie nerodė jokio signalo, nebuvo įtraukti (taip pat žr. 3.1 lentelę). (F) 5carC HPLC-MS/MS nustatymui išskirtos (žr. 2.2.20) *C. cinerea* RNR frakcijos (be poveikio vitC): kairėje, analizuotos *Bioanalyzer* (Agilent); dešinėje, analizuotos denatūruojančio poliakrilamido gelio elektroforeze. M – DNR ilgio standartas; M2 – *FastRuler Ultra Low Range DNA Ladder*; M3 – *RiboRuler High Range RNA Ladder* (visi iš *Thermo Scientific*). Visuminė – visuminė RNR; trumpos – <200 nt RNR frakcija, praturtinta naudojant kolonėles; ilgos be poli A – >200 nt praturtintos RNR iš kurių buvo pašalintos 3'-poli A sekas turinčios RNR (poli A); <tRNA<sub>≥22</sub> nt – RNR trumpesnės nei tRNR ir ilgesnės nei 22 nt; <22 nt – RNR trumpesnės nei 22 nt. Trumpųjų (<200 nt RNR) frakcija apima 5.8S rRNR, 5S rRNR, tRNR ir RNR trumpesnės nei tRNR; ilgų be poli A RNR frakcija apima 28S, 18S RNR ir kitas >200 nt RNR.

Ištyrėme 5carC pasiskirstymą tarp įvairių tipų RNR: 3'-poli A sekas turinčių RNR, ilgų RNR (>200 nt), iš kurių pašalintos poli A RNR, ir trumpų (<200 nt) RNR. *C. cinerea* atveju mes papildomai atskyrėme ilgas ribosomines RNR ir išfrakcionavome trumpąsias RNR į 5,8S rRNR, 5S rRNR, tRNR (~76-90 nt), <tRNR<sub>≥22</sub> nt frakciją ir RNR, trumpesnės nei 22 nt (tipišką miRNR ilgį) (3.21 pav. F). Po poveikio vitC *L. bicolor* 5carC buvo gana daug ilgose RNR, iš kurių pašalintos poli A, o signalas buvo sunkiai aptinkamas trumpų RNR (<200 nt) mišinyje (3.21 pav. E). Neveiktame *C. cinerea* didžiausias 5carC lygis buvo nustatytas trumpose RNR, tačiau veikimas vitC padidino 5carC kiekį ilgose RNR po poli A pašalinimo (3.21 pav. C). Poli A RNR, 5.8S rRNR ir tRNR rodė silpną signalą po veikimo vitC tik viename biologiniame pakartojime, o 28S, 18S ir 5S rRNR frakcijose nebuvo aptinkamo 5carC abejomis sąlygomis (3.21 pav. D, F, 3.1 lentelė). Po poveikio vitC 5carC lygis taip pat padidėjo <tRNR<sub>≥22</sub> nt ir <22 RNR frakcijose (3.21 pav. D, 3.1 lentelė). Santykinai stiprus 5carC signalas šiose frakcijose rodo, kad jos sudaro didžiąją dalį 5carC signalo trumpų <200 nt RNR mišinyje. Stebėjome 5mrC sumažėjimą ir 5hmrC padidėjimą visuminėje ir trumpose RNR po 2 dienų veikimo vitC, o tai potencialiai liudija, jog oksimC susidaro TET baltymams nuosekliai oksiduojant 5mrC. 5 dienų veikimas vitC dar padidino 5mrC ir 5hmrC kiekius, bet ne 5carC, galbūt dėl 5mrC atkūrimo proceso (3.21 pav. C). Įdomu tai, kad 5hmrC nebuvo aptiktas *C. cinerea* poli A RNR visomis sąlygomis.

**3.1 lentelė.** Kai kurių RNR frakcijų 5carC kiekiai, analizuoti atlikus cheminę derivatizaciją ir HPLC-MS/MS, 3.11 pav. papildantys duomenys. 5carC kiekiai, pateikti kaip % nuo viso G; (+) aptikti silpni, bet kiekybiškai neįvertinami signalai (kai signalo ir triukšmo santykis buvo  $\geq 3$ , bet signalas buvo ne kalibravimo kreivės tiesiniame diapazone); (-) aiškaus signalo neaptikta. Bio – biologinė replika, vitC – vitamino C poveikis 2 arba 5 dienas.

RNR	Poveikis	5carC-darinys	
		Bio 1	Bio 2
<i>C. cinerea</i>			
poli A	be poveikio	-	-
	vitC 2d.	+	-
	vitC 5d.	-	-
28S, 18S rRNR	be poveikio	-	-
	vitC 2d.	-	-
5.8S	be poveikio	-	-
	vitC 2d.	+	-
5S	be poveikio	-	-
	vitC 2d.	-	-
tRNR	be poveikio	-	-
	vitC 2d.	+	-
<tRNR $\geq$ 22 nt	be poveikio	0,0001%	0,00006%
	vitC 2d.	0,00013%	0,00015%
<22 nt	be poveikio	0,00005%	-
	vitC 2d.	0,00019%	+
<i>L. bicolor</i>			
<200 nt	vitC 2d.	0,00001%	-

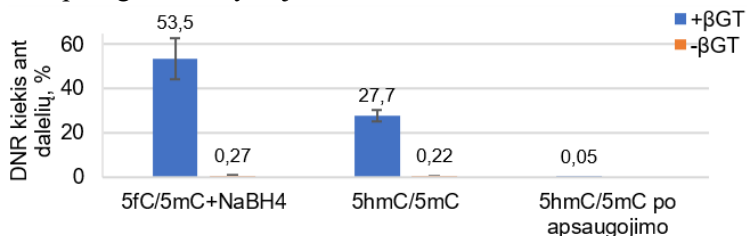
Apibendrinant, 5carC buvimas ilgose RNR, bet ne gausiausiose 28S ir 18S rRNR, *C. cinerea* rodo modifikacijos ryšį su iRNR, kurios yra su trumpomis 3'-poli A sekomis arba be jų, arba su nekoduojančiomis RNR. Atsižvelgiant į tai, kad poli A RNR *L. bicolor* įgavo 5carC po veikimo vitC, ši modifikacija gali būti įgyta tam tikromis biologinėmis sąlygomis. Reikėtų pažymėti, kad 5carC kiekiai liko beveik nepaveikti vitC trumpų *C. cinerea* ir *L. bicolor* RNR mišinyje, vadinasi, ši modifikacija nesukelia ar nerodo RNR degradacijos. Nepaisant to, negalime visiškai atmesti galimybės, kad silpnas 5carC signalas bendrame mažų RNR mišinyje yra dalies ilgų RNR skilimo pasekmė. Galbūt signalas ateina iš labai trumpų reguliacinių RNR, tokių kaip

miRNR, kurios buvo identifikuotos grybuose (į mikroRNR panašios RNR buvo aptiktos *C. cinerea*; Lau ir kt., 2020). Tačiau dar reikia iširti, kokios RNR lemia šį signalą.

### 3.4.3 5hmC ir 5fC genominis profiliavimas

Toliau tyrėme gausiausių oksiduotų 5-metilcitozino formų, 5hmC ir 5fC, pasiskirstymą *L. bicolor* ir *C. cinerea* genomuose. Naudojome hMe-Seal metoda ir jo modifikuotą versiją fC-Seal (žr. 1.4.1), kurie yra pagrįsti atitinkamai 5hmC arba 5fC afininiu praturtinimu, atliekant T4 bakteriofago  $\beta$ -gliukoziltransferazės vykdomą 5hmC modifikavimą azido grupę turinčia gliukoze (Song ir kt., 2011, 2013). 5fC nustatymui natūralus 5hmC pirmiausia glikozilinamas paprasta gliukoze, tuomet naudojant  $\text{NaBH}_4$  5fC redukuojamas į 5hmC, kuris pažymimas azido grupę turinčia gliukoze.

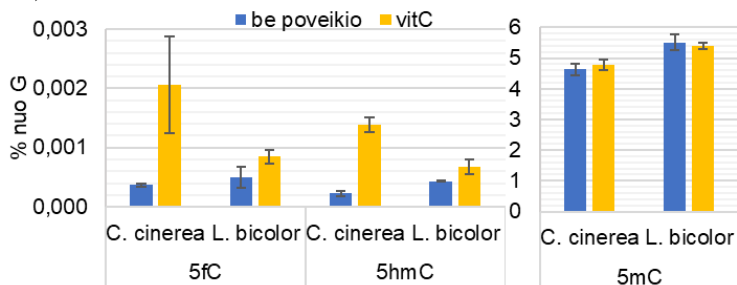
Grybų DNR analizei pirmiausia optimizavome hMe-Seal ir fC-Seal protokolus (žr. 2.2.21), remdamiesi efektyviomis glikozilinimo ir redukcijos sąlygomis, nustatytomis atitinkamai caCLEAR ir caMAB-seq (žr. pav. 3.4 A, 3.16 A). Patikrinome hMe-Seal ir fC-Seal protokolus su modeliniais DNR fragmentais, kurių vieno CG taikinio vienoje grandinėje buvo 5hmC arba 5fC (o kitoje grandinėje 5mC). Siekiant imituoti modifikacijų retumą, fragmentai buvo sumaišyti su *L. bicolor* genomine DNR (13:1287 ng) ir buvo atlikta žymėjimas azido grupę turinčia gliukoze (5fC turinčio fragmento atveju prieš tai atlikta redukcija į 5hmC), biotino prijungimas ir praturtinimas naudojant streptavidino magnetines daleles. Kiekybinės PGR būdu stebėjome pakankamai efektyvų (~28-54%) abi modifikacijas turinčių fragmentų praturtinimą (3.22 pav. A). Tuo tarpu prieš žymėjimą azido grupę turinčia gliukoze atlikus 5hmC modifikavimą paprasta gliukoze, praturtinimas buvo beveik neaptinkamas, kaip ir atliktose kontrolinėse reakcijose be  $\beta$ GT esant UDP-6-azido-gliukozei (3.22 pav. A). Taigi metodas buvo specifiškas 5hmC ir 5fC turinčiai tikslinei DNR, o siekiant 5fC atskirti nuo 5hmC, 5hmC galima efektyviai apsaugoti nuo žymėjimo.



**3.22 pav.** hMe-Seal ir fC-Seal (+NaBH<sub>4</sub>) metodų įvertinimas atitinkamai atliekant 5hmC ir 5fC praturtinimą. Modeliniai DNR fragmentai, turintys vieną CG taikinį, kuriame yra 5hmC arba 5fC (o kitoje grandinėje 5mC – 5hmC/5mC, 5fC/5mC), buvo

sumaišyti su *L. bicolor* DNR ir žymėti esant UDP-6-azido-gliukozei ir T4 bakteriofago  $\beta$ -gliukoziltransferazei (+ $\beta$ GT, mėlyna) arba be jos (- $\beta$ GT, oranžinė). Taip pat atlikta kontrolinė žymėjimo reakcija naudojant 5hmC/5mC DNR fragmentą, kuris prieš tai buvo glikozilintas esant nemodifikuotai UDP-gliukozei (apsaugotas nuo  $\beta$ GT žymėjimo azido grupę turinčia gliukoze). Po biotininimo ir afininio praturtinimo, kurio efektyvumas įvertintas kPGR (žr. 2.2.21), matomas geras specifiskumas 5hmC ir 5fC turinčiai tikslinei DNR.

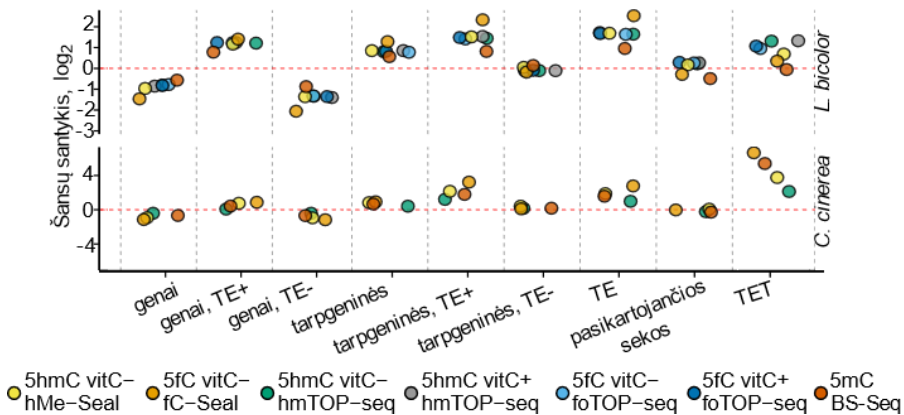
Siekiant sustiprinti oksimC signalus ir taip patikimiau nustatyti oksimC išsidėstymą genomineje DNR, paveikėme abu grybus vitaminu C. Kaip ir tikėtasi, HPLC-MS/MS analizė parodė oksimC kiekių padidėjimą: nuo 0,0004% iki 0,0007% 5hmC ir nuo 0,0005 iki 0,0009% 5fC *L. bicolor*; *C. cinerea* 5hmC didėjo nuo 0,0002 iki 0,0014% ir 5fC nuo 0,0004 iki 0,002% (3.23 pav.).



**3.23 pav.** *L. bicolor* ir *C. cinerea* citozino modifikacijų HPLC-MS/MS analizė prieš (mėlyna) ir po poveikio vitaminu C (vitC, geltona). 5fC, 5hmC ir 5mC procentai nuo viso G buvo apskaičiuoti naudojant atitinkamų nukleozidų kalibracines kreives. Naudotos 2-3 biologinės replikos. Šis eksperimentas buvo atliktas be papildomo nukleozidų gryninimo HPLC frakcionavimu skirtingai nuo 3.19 pav. A eksperimento (žr. 2.2.10).

Siekdami palyginti 5hmC ir 5fC pasiskirstymą *L. bicolor* ir *C. cinerea* genomineje DNR, šių modifikacijų praturtinimui pritaikėme optimizuotus hMe-Seal ir fC-Seal protokolus. Buvo paruoštos įprastinėmis sąlygomis augintų ir vitC veiktų *L. bicolor* ir *C. cinerea* DNR hMe-Seal ir fC-Seal bibliotekos. Atlikus sekoskaitą, buvo gautas vidutinis padengimas ( $25\times$  *L. bicolor* ir  $45\times$  *C. cinerea*, beveik visa grybų sekoskaitos analizė atlikta Kotrynos Kvederavičiūtės, kai kurie pav. dr. P. Gibo). Analizuojant 200 bp ilgio nepersidengiančius genomo regionus, abiejuose grybuose 5hmC ir 5fC signalai tarpusavyje persidengė vidutiniškai (*Jaccard* koeficientai 0,51 ir 0,33 atitinkamai *L. bicolor* ir *C. cinerea*). 5hmC ir 5fC gerai dengėsi su viso genomo bisulfitinės sekoskaitos signalu (duomenys iš Zemach ir kt., 2010), kuris rodo bendrą 5mC (jis sudaro didžiąją signalo dalį) ir 5hmC signalą (analizuojant 200-bp nepersidengiančius genomo regionus, 56,65 % ir 69,7 % BS-Seq modifikavimo signalo persidengia su hMe-Seal, 50,97 % ir 79,63 %

signalo sutampa su fC-Seal genomo regionais, atitinkamai *L. bicolor* ir *C. cinerea*). 5hmC bei 5fC turinčių regionų praturtinimo analizė įvairiuose genominiuose elementuose parodė, kad abi modifikacijos (panašiai kaip ir 5mC) stipriausiai praturtintos su genais susijusiuose ir tarpgeniniuose judriuosiuose genomo elementuose (TE) bei tarpgeninėse srityse (3.24 pav.). Tai sutampa su *C. cinerea* 5hmC rezultatais naudojant kitus profiliavimo metodus – 5hmC buvo nustatytas TE ir juos supančiuose regionuose, centromerose, genuose be raiškos, homologų neturinčiuose genuose ir daugiakopijiniuose genų paraloguose (Chavez ir kt., 2014). Taip pat stebėjome 5hmC/5fC praturtinimą abiejų grybų TET genuose, kuris buvo stipresnis *C. cinerea*. Mėginiuose po vitC poveikio stebėjome analogišką praturtinimą genomo elementuose kaip ir mėginiuose be poveikio (3.24 pav.).

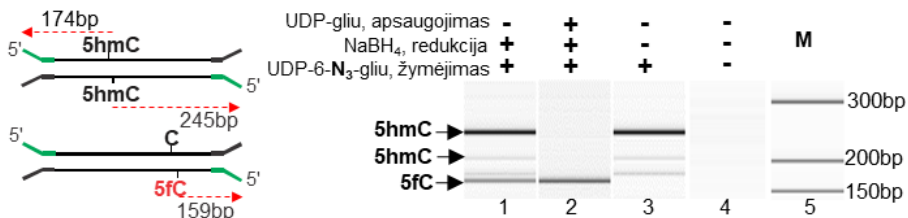


**3.24 pav.** *L. bicolor* ir *C. cinerea* 5hmC ir 5fC signalų praturtinimas (šansų santykis,  $\log_2$ ) įvairiuose genominiuose elementuose, nustatytas skirtingais metodais: hMe-Seal ir fC-Seal (200 bp dydžio regionų analizė); hmTOP-seq ir foTOP-seq (pavienių CG analizė); 5mC signalas nustatytas pagal BS-Seq duomenis (naudoti >50% metilinti CG; Zemach ir kt., 2010). Patikimumas nustatytas naudojant tikslų Fišerio testą. TE – judrieji genomo elementai, pasikartojančios sekos – kitos pasikartojančios sekos neįskaitant TE (tandeminiai pasikartojimai, į kuriuos įeina mini ir mikrosatelitinės sekos). Mėginiai po veikimo vitaminu C pažymėti vitC+, be poveikio – vitC-. TE+, TE- atitinkamai genominiai elementai turintys TE ir ne.

Kadangi genai tiriamuose grybuose yra gana trumpi (vidutiniškai 1,5 kb *L. bicolor*; Martin ir kt., 2008; ir 1,8 kb *C. cinerea*; Stajich ir kt., 2010), vidutinė afiniu praturtinimu paremtų metodų skiriamoji geba (~150-400, remiantis mūsų eksperimentuose gaunamu DNR fragmentų ilgiu) gali sukurti persidengiančius oksimC pasiskirstymo rezultatus genominiuose elementuose, ypač TE, kurių vidutinis ilgis 256 bp ir 233 bp atitinkamai *L. bicolor* ir *C. cinerea*. Todėl toliau siekėme patikslinti 5hmC ir 5fC pasiskirstymą nukleotido skiriamosios gebos profiliavimo metodais, kurie yra

pagrįsti TOP-seq (Staševskij ir kt., 2017; žr. 1.4.2) labiau akcentuojant dar menkai ištirtą *L. bicolor*.

Didelės skiriamosios gebos 5hmC profiliavimui siekėme pritaikyti laboratorijoje sukurtą hmTOP-seq (žr. 1.4.2) metodą (Gibas ir kt., 2020), kuriame veikiant βGT 5hmC žymimas azido grupę turinčia gliukoze ir prie jos prijungiamas pradmuo leidžia po sekoskaitos nustatyti 5hmC nukleotido tikslumu. Remiantis fC-Seal protokolu, 5fC profiliavimui *L. bicolor* sukūrėme modifikuotą hmTOP-seq versiją – foTOP-seq, kuriame prieš atliekant hmTOP-seq žingsnius natūralus 5hmC modifikuojamas, o 5fC redukuojamas į 5hmC ir taip galima nustatyti 5fC nukleotido tikslumu. Rėmėmės fC-Seal naudotomis redukcijos ir glikozilinimo sąlygomis (žr. 2.2.21). Išbandėme foTOP-seq su dviem vieną CG taikinių turinčiais modeliniais DNR fragmentais, kurių viename buvo 5hmC (abiejose grandinėse), o kitame 5fC (vienoje grandinėje). Kadangi atstumas nuo 5' galo iki kiekvieno šių DNR fragmentų 5hmC/5fC taikinių skirtingas, analizuojant elektroforetiškai pagal ilgį galima atskirti, nuo kurio iš taikinių susidaro galutinis PGR produktas. Atlikus visus foTOP-seq žingsnius, stebėjome vieną nuo 5fC taikinio susidariusį fragmentą (3.25 pav. 2 takelis, sprendžiant pagal fragmento ilgį) vietoje 3, kai nebuvo atliktas 5hmC glikozilinimas paprasta gliukoze (1 takelis). Neatlikus 5fC redukcijos ir glikozilinimo paprasta gliukoze, susidarė tik 5hmC PGR produktai, o 5fC PGR produktas nesusidarė. Kontrolinėje reakcijoje be βGT PGR produktai visai nesusidarė. Vadinas, 5hmC buvo efektyviai apsaugotas nuo žymėjimo ir foTOP-seq metodas gali specifiskai nustatyti 5fC, o taip pat 5hmC galima atskirai nustatyti hmTOP-seq metodu.



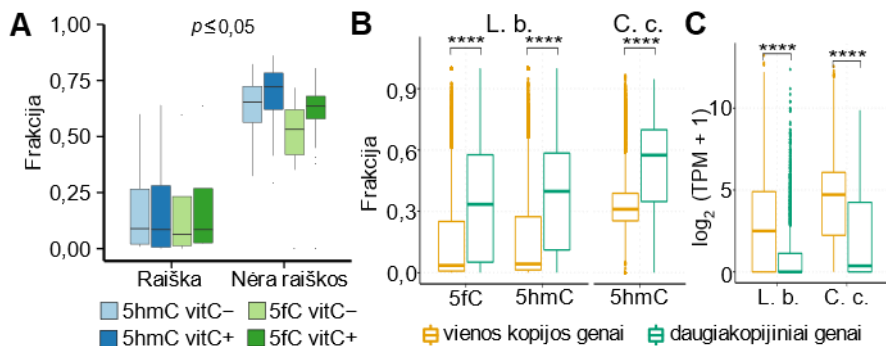
**3.25 pav.** foTOP-seq metodo įvertinimas naudojant du modelinius DNR fragmentus turinčius 5hmC arba 5fC (kairėje). Raudonomis rodyklėmis pažymėti galintys susidaryti galutinių bibliotekų PGR produktai. Dešinėje, elektroforetinė PGR produktų analizė bioanalizatoriumi (*Agilent*). Kai 5hmC buvo apsaugotas βGT glikozilinant paprasta gliukoze (gliu), 5fC redukuotas NaBH<sub>4</sub> ir susidaręs 5hmC pažymėtas naudojant UDP-6-azido-gliukozę bei βGT susidaro vienas 5fC specifinis PGR produktas (2 takelis).

Buvo paruoštos įprastai augintos ir vitC paveiktos *L. bicolor* DNR hmTOP-seq ir foTOP-seq, bei įprastai auginto *C. cinerea* DNR hmTOP-seq bibliotekos, atlikta jų sekoskaita. Analizavome modifikacijas tik CG

dinukleotiduose, kadangi 5mC *L. bicolor* ir *C. cinerea* genomuose daugiausia yra CG kontekste (Zemach ir kt., 2010). *L. bicolor* DNR nustatėme 275 922 5hmCG ir 271 761 5fCG atitinkamai dviejose hmTOP-seq ir foTOP-seq replikose (vidutiniškai 10× padengimas), kurių skaičius, kaip ir tikėtasi, padidėjo po poveikio vitC į 315 170 5hmCG ir 349 758 5fCG. Auginant standartinėmis sąlygomis *C. cinerea* genome identifikavome 663 841 5hmCG dviejose hmTOP-seq bibliotekų replikose (vidutiniškai 10× padengimas). Gana didelė koreliacija tarp 5hmCG ir 5fCG turinčių regionų genominiuose elementuose (Pearson  $r=0,85$  genams ir 0,49 pasikartojantiems elementams) rodė didelį 5hmC ir 5fC persidengimą *L. bicolor*. Oksi-mC modifikuotų CG praturtinimo analizė genominiuose elementuose gerai sutapo su prieš tai minėtais hMe-Seal ir fC-Seal rezultatais ir rodė panašias tendencijas vitC veiktuose ir neveiktuose mėginiuose (3.24 pav.): 5hmCG ir 5fCG buvo praturtinti TE, tarpgeninėse srityse ir TE turinčiuose genuose.

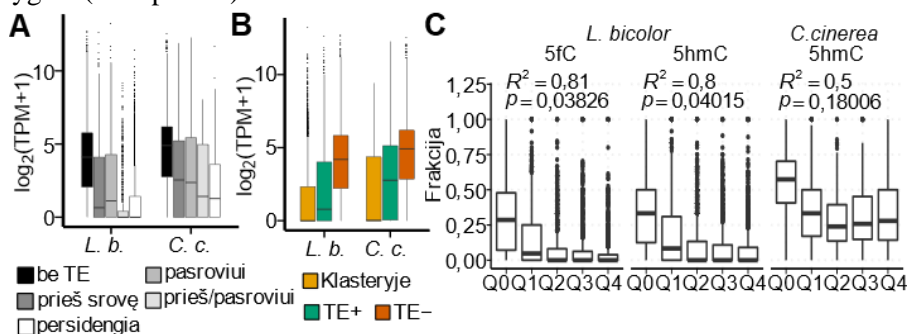
#### 3.4.4 5hmC ir 5fC ryšys su genų raiška

Siekiant įvertinti oksimC įtaką genų raiškai grybuose, atlikome standartinėmis sąlygomis ir su vitC auginto *L. bicolor* micelio visos transkriptomos sekoskaitą (bibliotekos paruoštos ir sekoskaita atlikta *Thermo Fisher Scientific Baltics*). *C. cinerea* atveju naudojome viešai prieinamus standartinėmis sąlygomis auginto micelio poli A RNR sekoskaitos duomenis (Plaza ir kt., 2014). 5fC/5hmC modifikavimo vertinimui skaičiavome modifikuotų CG dalį (frakciją) nuo visų tame geno elemente esančių CG (žr. Gordevičius ir kt., 2020; Narmontė ir kt., 2021). Nors dauguma iš mūsų identifikuotų 46 *L. bicolor* *TET* genų turi 5fC/5hmC modifikacijas, tačiau tik 4 iš jų su nedidele oksimCG frakcija (3.26 pav. A) buvo su raiška (TPM>0,05). Šie aktyvūs *TET* genai potencialiai galėtų dalyvauti oksimC susidaryme. Kiti daugiakopijiniai genai taip pat turėjo daugiau 5fCG/5hmCG ir buvo mažesnės raiškos lyginant su vienos kopijos genais (3.26 pav. B, C), o tai sutampa su *C. cinerea* 5hmC duomenimis (Chavez ir kt., 2014).



**3.26 pav.** 5hmC ir 5fC modifikavimas *L. bicolor* ir *C. cinerea* genuose. **(A)** 5hmCG ir 5fCG (hmTOP-seq ir foTOP-seq) frakcija (dalis nuo visų elemento CG) *L. bicolor* *TET* genuose su raiška (TPM>0,05) arba be jos. Visos porinio Wilcox testo p vertės buvo mažesnės ar lygios nurodytai reikšmei. Mėginiai veikti vitaminu C pažymėti vitC+, mėginiai be poveikio – vitC-. **(B)** 5hmCG ir 5fCG frakcijos daugiakopijiniuose ar turinčiuose vieną kopiją genuose bei jų raiškos lygis **(C)**. Žvaigždutės žymi p vertes, paskaičiuotas naudojant Wilcox (frakcijoms) ir t-testą (raiškai). \*\*\*\* – p-vertė  $\leq 0,0001$ . L. b. – *L. bicolor*; C. c. – *C. cinerea*. TPM – geno transkriptai milijonui visų transkriptų.

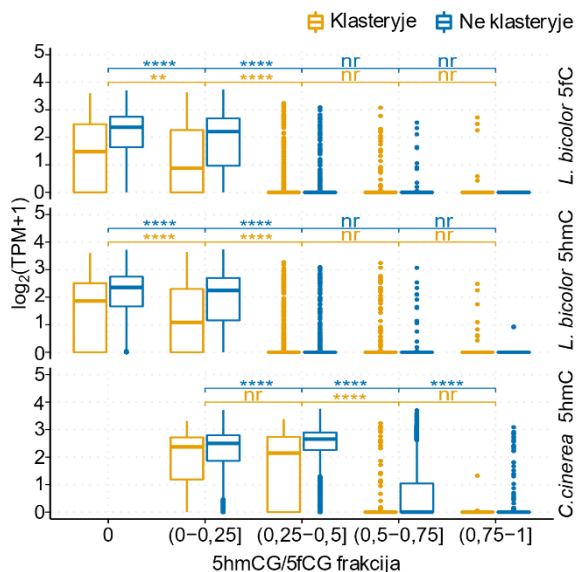
Kadangi genai ir tarpgeniniai regionai yra labiau praturtinti 5hmCG/5fCG, kai persidengia su TE (3.24 pav.), toliau analizavome raišką priklausomai nuo TE buvimo. Yra žinoma, kad grybuose TE įsiterpimas ar net buvimas arti genų gali juos slopinti (Castanera ir kt., 2016). Nors yra paskelbtų duomenų, kad grybuose TE gali būti išveiklinami DNR metilinimo (Borgognone ir kt., 2018) ir heterochromatino formavimosi (Seidl ir Thomma, 2017; Soyer ir kt., 2014), 5mC oksidacijos įtaka TE veiklumui anksčiau nebuvo tirta. Nustatėme, kad abiejuose grybuose su TE susiję genai (atstumas iki TE <1 kb, kai genas yra apsuptas TE, TE yra prieš arba už geno arba persidengia su genu) buvo mažesnės raiškos nei TE neturintys genai ir skirtumas buvo ryškesnis *L. bicolor* (3.27 pav. A, B). *L. bicolor* genai, kurie persidengė arba iš abiejų pusių buvo apsupti TE, rodė mažiausias raiškos lygius (3.27 pav. A).





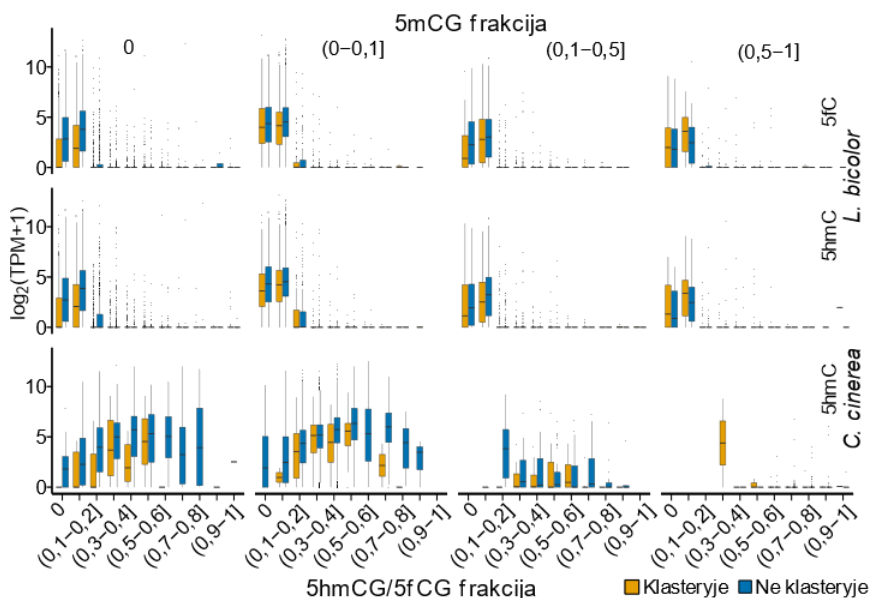
**3.27 pav.** *L. bicolor* ir *C. cinerea* judriųjų elementų (TE) apsupties ir jų 5hmC/5fC modifikavimo ryšys su genų raiška. **(A)** Genų raiška pagal TE padėtį jų atžvilgiu (be TE – TE nėra 1 kb atstumu, prieš srovę arba pasroviui – TE aptiktas 1 kb atstumu atitinkamai prieš srovę arba pasroviui nuo geno, prieš/pasroviui – TE aptiktas iš abiejų pusių prieš srovę ir pasroviui nuo geno, persidengia – TE persidengia su genu). Skirtumams grupėse įvertinti ( $p$  vertė  $\leq 0,05$ ) buvo naudota vienpusė ANOVA. **(B)** Genų asocijuotų su TE (TE+, atstumas  $< 1$  kb), jų neturinčių (TE-) arba esančių TE klasteriuose raiška ( $\log_2(\text{TPM}+1)$ ). Skirtumams grupėse įvertinti ( $p$  vertė  $\leq 0,05$ ) buvo naudota vienpusė ANOVA. **(C)** TE 5hmCG ir 5fCG modifikavimo frakcija (dalis nuo visų elemento CG, įvertinta hmTOP-seq ir foTOP-seq) ir susijusių genų raiška (atstumas iki TE  $< 1$  kb, raiškos kvartiliai Q1-4, Q0 – nėra raiškos). Duomenims pritaikytas linijinis modelis. *L. b.* – *L. bicolor*; *C. c.* – *C. cinerea*. TPM – geno transkriptai milijonui visų transkriptų.

Kai kuriuose grybuose buvo parodyta, kad labiau metilinti ir mažiau aktyvūs genai aptinkami TE sankaupose – klasteriuose (Borgognone ir kt., 2018; Castanera ir kt., 2016). *L. bicolor*, priešingai nei *C. cinerea*, TE nesikaupia tik subtelomeriniuose ar centromeriniuose regionuose, bet formuodami TE klasterius pasiskirsto po visą genomą (Martin ir kt., 2008; Stajich ir kt., 2010). Palyginamajai *L. bicolor* ir *C. cinerea* analizei *de novo* nustatėme TE klasterius (žr. Ličytė ir kt., 2022). Kaip ir tikėtasi, gerokai didesnis TE klasterių skaičius aptiktas *L. bicolor* (207 lyginant su 82 *C. cinerea*). Stebėjome, jog TE klasteriuose esantys genai buvo mažiausios raiškos, lyginant su genais, esančiais ne klasteriuose ar neturinčiais TE (3.27 pav. B). Dauguma TE klasteriuose esančių genų (4 312 iš 7 095) buvo be raiškos (vidutiniškai 0,28 5hmCG/5fCG ir 0,11 5mCG frakcijos), o tuo tarpu tik 3 618 iš 15 815 genų buvo nutildyti ne klasteriuose (vidutiniškai 0,24 5hmCG/5fCG ir 0,13 5mCG frakcijos) ir tai patvirtina slopinančių TE poveikį. Genų su raiška modifikuotų CG frakcijos klasteriuose buvo šiek tiek didesnės nei ne klasteriuose: vidutiniškai 0,06 5hmCG, 0,06 5fCG ir 0,07 5mCG klasteriuose; 0,03 5hmCG, 0,02 5fCG ir 0,04 5mCG ne klasteriuose. Tuomet analizavome pačių TE modifikavimą esant skirtingai su jais susijusių (atstumas iki TE  $< 1$  kb) genų raiškai. Nustatėme, kad labiausiai 5hmCG/5fCG modifikuoti TE buvo būdingi mažiausios raiškos asocijuotiems genams (3.27 pav. C), todėl stiprus TE 5hmCG/5fCG modifikavimas susijęs su raiškos slopinimu. Taip pat nustatėme, kad labiau modifikuoti prieš srovę esantys 1 kb regionai (galintys turėti promotorius) buvo susiję su mažesne raiška, kuri buvo dar mažesnė TE klasteriuose (3.28 pav.). Taigi, kadangi genominiai elementai su TE turi didesnę 5hmCG/5fCG praturtinimą (3.24 pav.), o su jais susijusių genų raiška yra mažesnė, TE daro įtaką didesniai 5fC/5hmC modifikavimui ir tuo pačiu daro neigiamą poveikį raiškai.



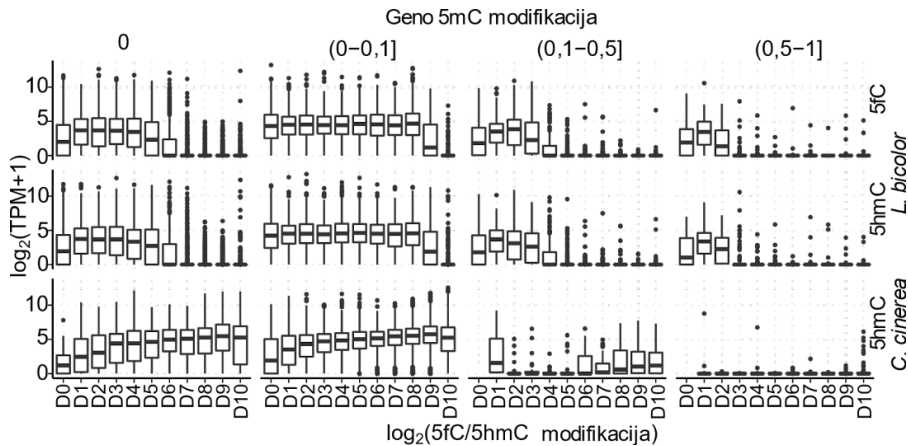
**3.28 pav.** *L. bicolor* ir *C. cinerea* regionų, esančių prieš geną, 5hmC ir 5fC modifikavimo ryšys su raiška. Genų, išsidėsčiusių TE klasteriuose ir už jų, raiška ( $\log_2(\text{TPM}+1)$ ) pagal prieš srovę nuo genų esančių regionų (apibrėžtų kaip 1 kb regionas prieš srovę nuo geno pradžios) 5hmCG ir 5fCG frakcijas (didėjant po 0,25). Žvaigždutės žymi p vertes, paskaičiuotas naudojant *t* testą. nr (nereikšmingas) – p vertė > 0,05; \*\* –  $p \leq 0,01$ ; \*\*\*\* –  $p \leq 0,0001$ . Laužtiniai skliaustai rodo intervale įskaičiuojamą reikšmę, paprasti skliaustai – neįskaičiuojamą. TPM – geno transkriptai milijonui visų transkriptų.

Kadangi genai ir TE turi visas tris citozino modifikacijas – 5mC, 5hmC, 5fC – toliau siekėme tirti jų tarpusavio sąveiką genų raiškos reguliacijoje. Siekiant palyginti oksidini-5mCG ir jų pirmtako 5mCG pasiskirstymą genuose ir TE, skaičiavome modifikuotų CG frakcijas. Sugrupavome visus genus pagal jų 5mCG frakcijas ir įvertinome jų raišką priklausomai nuo didėjančių oksidini-5mC frakcijų. Stebėtina, tačiau ryšys tarp raiškos ir 5hmCG lygio skyrėsi *L. bicolor* ir *C. cinerea* (3.29 pav.). Mažai metilintuose genuose (0-0,1 5mCG frakcija) *C. cinerea* buvo stebimas teigiamas raiškos ryšys su didėjančiu 5hmCG lygiu, o tai sutampa su stebima teigiama 5hmC įtaka raiškai žinduoliuose (Wu ir Zhang, 2014), tuo tarpu raiška buvo silpna arba jos nebuvo stipriai metilintuose ir hidroksimetilintuose genuose (>0,5 frakcijos). *L. bicolor* teigiamas ryšys tarp raiškos ir 5hmCG/5fCG frakcijų buvo pastebimas tik lyginant dvi genų grupes – neturinčias oksidini-5mC ir turinčias mažas oksidini-5mCG frakcijas (iki 0,1), tuo tarpu genai su didesnėmis nei 0,1 5hmCG/5fCG frakcijomis dažniausiai buvo nutildyti nepriklausomai nuo jų 5mCG lygio. Ta pati tendencija buvo stebima genams, esantiems TE klasteriuose ir jų išorėje, tik raiška buvo silpnesnė genams esant klasteriuose.



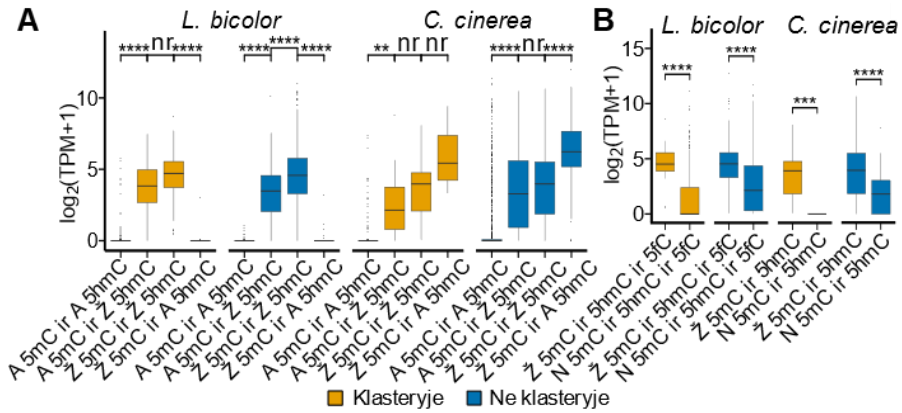
**3.29 pav.** 5hmC, 5fC ir 5mC modifikacijų ryšys su genų raiška *L. bicolor* ir *C. cinerea*. Raiška ( $\log_2(\text{TPM}+1)$ ) išskirstyta pagal genų 5fCG ir 5hmCG frakcijas skirtingai metilintuose genuose (5mCG frakcijos parodytos virš grafiko; BS-Seq duomenys iš Zemach ir kt., 2010) TE klasteriuose ir jų išorėje. Laužtiniai skliaustai rodo intervale įskaičiuojamą reikšmę, paprasti skliaustai – neįskaičiuojamą. TPM – geno transkriptai milijonui visų transkriptų.

Siekiant įvertinti naudojamo CG frakcijų matavimo tinkamumą genų modifikavimo analizei, papildomai paskaičiavome CG padengimo vertes tose pačiose 5mCG modifikavimo grupėse. Apskritai rezultatai sutapo, išskyrus didelį 5hmCG modifikavimą turinčias grupes, kurios nerodė stipraus slopinimo *C. cinerea* ir rodė mažesnę slopinimą mažai metilintuose genuose *L. bicolor* (3.30 pav.). Vadinas, yra genų, kurie turi nedidelį kiekį stipriai modifikuotų (t.y. stipriai padengtų) 5hmCG taikinių ir demonstruoja, kad naudojant abu parametrus TOP-seq duomenų analizėje galima gauti papildančios informacijos.



**3.30 pav.** *L. bicolor* ir *C. cinerea* genų raiškos ( $\log_2(\text{TPM}+1)$ ) priklausomybė nuo normalizuoto nustatytų 5hmCG ir 5fCG padengimo (decilės D1-10, D0 – nėra 5hmCG/5fCG) genų grupėse, išskirstytose pagal 5mC modifikacijos lygius (šimtiosios dalys, viršuje, BS-Seq; Zemach ir kt., 2010). Laužtiniai skliaustai rodo intervale įskaičiuojamą reikšmę, paprasti skliaustai – neįskaičiuojamą. TPM – geno transkriptai milijonui visų transkriptų.

Tuomet tarpusavyje lyginome genus su didelėmis ir mažomis (atitinkamai paskutinė ir pirma decilė) skirtingų citozino modifikacijų frakcijomis. Vėlgi stebėjome, kad *L. bicolor* abi oksiduotos 5-metilcitozino formos (duomenys parodyti tik 5hmC) stipriai slopina genus TE klasteriuose ir už jų: abiejose stipriai arba silpnai metilintų genų grupėse tik genai turintys žemą 5hmCG frakciją rodė raišką, o tuo tarpu daug 5hmCG turintys genai buvo nutildyti (3.31 pav. A). Tuo tarpu *C. cinerea* mažai metilinti genai su didele 5hmCG frakcija žymėjo aktyvius genus (kurie buvo patikimai aktyvesni nei genai su mažai 5hmCG ne klasteriuose), o daug 5hmCG ir 5mCG kartu buvo stipriai slopinamuose genuose ir slopinimas buvo mažesnis esant mažai 5hmCG (bei mažai 5hmCG ir daug 5mCG turintys genai linkę būti labiau slopinami, nei esant mažai 5mCG). Iš kitos pusės, nepaisant stebimo slopinančio oksimC vaidmens abiejuose grybuose ir 5mC *C. cinerea*, genai turintys žemus visų citozino modifikacijų lygius buvo didesnės raiškos nei genai, visai neturintys modifikacijų ( $p$  vertės  $<0,001$ , 3.31 pav. B). Stebimi skirtumai tarp grybų galbūt rodo esančius skirtingus TE reguliuojančius mechanizmus dėl *L. bicolor* būdingo didesnio TE gausumo lyginant su *C. cinerea*. Nors negalime visiškai atmesti, kad skirtumai galbūt iš dalies susidaro dėl prastos TE ir genų anotacijos *L. bicolor* genome (Castanera ir kt., 2016; Martin ir kt., 2008).



**3.31 pav.** *L. bicolor* ir *C. cinerea* 5hmC, 5fC ir 5mC gausumo ryšys su genų raiška. (A) Skirtingai modifikuotų genų raiškos ( $\log_2(\text{TPM}+1)$ ) palyginimas, sugrupuotų į turinčius aukštą (A, paskutinė decilė) ir žemą (Ž, pirmą decilė) genų 5mCG ir 5hmCG frakcijas, TE klasteriuose ir jų išorėje. (B) Genų raiška, turinčių žemą visų trijų 5mC, 5fC ir 5hmC modifikuotų CG frakcijas (Ž) *L. bicolor*, o *C. cinerea* 5mCG ir 5hmCG ir 5hmCG, arba visai neturinčių šių modifikacijų (N). Žvaigždutės žymi p vertes, paskaičiuotas naudojant *t* testą. nr (nereikšmingas) – p vertė >0,05; \*\* – p ≤0,01; \*\*\* – p ≤0,001; \*\*\*\* – p ≤0,0001. TPM – geno transkriptai milijonui visų transkriptų.

## 4. REZULTATŲ APTARIMAS

DNR demetilimas yra svarbus genų raiškos reguliacijos keitimo mechanizmas vykstant ląstelių diferenciacijai. Žinduoliuose geriausiai apibūdintas aktyvaus demetilimo kelias, kurio metu TET baltymai oksiduoja 5mC į 5hmC, 5fC ir 5caC, o tuomet 5fC ir 5caC gali būti pašalinami veikiant TDG glikozilazei ir BER. Daugėja įrodymų, kad šios oksiduotos 5-metilcitozino formos gali būti ne vien 5mC demetilimo tarpiniai produktai, bet ir turėti epigenetines funkcijas, kurios šiuo metu dar nėra iki galo suprastos. Alternatyvūs aktyvaus 5mC demetilimo būdai žinduoliuose dar mažai ištirti. Buvo nustatyta, kad DNR metiltransferazės geba atlikti 5caC dekarboksilinimą tiesiogiai susidarant nemodifikuotam citozinui (Liutkevičiūtė ir kt., 2014). Tyrėme šią reakciją kaip galimą nemetilintos citozino būsenos grąžinimo būdą bei siekėme ją pritaikyti veiksmingo genomino 5caC profiliavimo metodo kūrimui, kuris padėtų geriau suprasti galimas šios retos ir todėl sunkiai nustatomos modifikacijos funkcijas.

Oksiduotos 5mC formos mažai nagrinėtos kitų nei žinduoliai organizmų genomuose. Todėl analizavome daug *TET-JPB* genų kopijų turinčių papėdgybių *L. bicolor* ir *C. cinerea* oksi-mC pasiskirstymą ir jų galimas funkcijas, pritaikant laboratorijoje kuriamus DNR modifikacijų genomino profiliavimo metodus. Šie atlikti darbai suteikė daugiau žinių apie 5mC demetilimą, oksi-mC funkcijas ir pasiskirstymą bei buvo sukurti nauji metodai jų tyrimui.

### 4.1 Žmogaus DNMT3A/3L ir DNMT3B/3L metiltransferazių atliekama 5caC dekarboksilinimo reakcija

Patvirtinome, kad nustatytoje karboksilo grupės pašalinimo nuo 5caC reakcijoje esant žmogaus DNMT3A/3L ar DNMT3B/3L (Liutkevičiūtė ir kt., 2014), dekarboksilinimą atlieka metiltransferazės, kadangi stebėjome jo slopinimą po poveikio su metiltransferazių katalitiniu cisteinu reaguojančiu *N*-etilmaleimidu. Bakterinių metiltransferazių katalitinį cisteiną pakeitus serinu, 5caC dekarboksilinimas nebevyksta (Liutkevičiūtė ir kt., 2014) ir tai sutampa su mūsų rezultatu, kad DNMT3A/3L ir 3B/3L atliekamai dekarboksilinimo reakcijai reikalingas šis katalitinis cisteinas. Taip pat parodytas 5hmC dehidroksimetilimo slopinimas, kai minėta mutacija buvo įvesta į pelės DNMT3A ir 3B (Chen ir kt., 2012) bei bakterijų (Liutkevičiūtė ir kt., 2009) metiltransferazes. Mechanistinė *in vitro* analizė taip pat parodė, kad 5caC gali būti dekarboksilinamas tiolio nukleofilui laikinai jungiantis C6 pozicijoje (Schiesser ir kt., 2013), analogiškai citozino metiltransferazių

metilinio mechanizmui, kuriame fermento cisteinas laikinai jungiasi C6-toje pozicijoje (Kriukienė ir kt., 2012).

Manoma, kad egzistuoja nuo TDG nepriklausomas aktyvus DNR demetilinio kelias, nes išveiklinus TDG aktyvus 5mC demetilimas pelės zigotose nepakinta (Guo ir kt., 2014a). Be to, TDG-BER vykdomas oksi-mC pašalinimas, esant dažnam ar simetriškam modifikacijų išsidėstymui, galėtų sukelti tankius ar net dvigrandžius trūkius, o tai galėtų pažeisti tokio lokuso genetinį integralumą. Šis sudėtingas procesas turi daug žingsnių, jame dalyvauja daugiau nei 10 skirtingų fermentų, todėl jam reikia daug energijos (Feng ir kt., 2021a). Taigi nuo TDG-BER nepriklausomas tiesioginis nemetiltos citozino būsenos grąžinimas būtų saugesnė ir mažiau energijos naudojanti, paprastesnė, iš mažiau žingsnių susidedanti alternatyva. Naudojant su bazių neturinčiomis vietomis reaguojančią medžiagą, buvo parodyta, kad 5fC ir 5caC pašalinimas nepadidina potencialiai kenksmingų BER tarpinių produktų lygio (Carell ir kt., 2018). Šį rezultatą galima aiškinti dviem būdais: arba TDG/BER veikimas yra ne toks gausus, nei manyta iki šiol, arba TDG pagrįstas pašalinimas vyksta glaudžiam komplekse kartu su kitais baltymais ir bazės neturinti vieta ištaisoma labai greitai, todėl bazės iškirpimo taisymo tarpiniai produktai negali kauptis iki didelių lygių (Carell ir kt., 2018).

Neseniai buvo paskelbti nauji rezultatai, rodantys *in vivo* vykstantį C-C ryšio skilimą. Nustatytas tiesioginis 5caC dekarboksilinimas į C įvairių kultivuojamų žmogaus ir pelės ląstelių genomuose bei transfekavus į jas 5caC turinčią DNR (Feng ir kt., 2021a, 2021b; Kamińska ir kt., 2021). Masių spektrometrija pagrįstas proteominis tyrimas pelės embrioninėse kamieninėse ląstelėse atskleidė DNMT1 sąveiką su 5caC (Spruijt ir kt., 2013), tačiau slopinant žmogaus *DNMT1* raišką nepakinta transfekuotos 5caC turinčios DNR dekarboksilinimas (Feng ir kt., 2021b). *In vitro* grybų izoorotato dekarboksilazė rodo silpną 5caC dekarboksilinimą (Xu ir kt., 2013). Žmogaus ląstelėse nėra izoorotato dekarboksilazei homologiško baltymo, o aptinkama struktūriškai panaši orotidino 5'-monofosfato dekarboksilazė, tačiau jos raiškos slopinimas nepakeičia transfekuotos DNR dekarboksilinimo (Feng ir kt., 2021b). Taigi *in vivo* tiesioginę 5caC dekarboksilinimo reakciją vykdytys fermentai iki šiol nenustatyti.

Savo darbuose stebėjome žmogaus DNMT3A/3L vykdomos 5caC dekarboksilinimo reakcijos specifinį slopinimą SAM kofaktoriumi ir jo produktu SAH, o tai dar kartą patvirtina metiltransferazių veikimą šioje reakcijoje. Atitinkamai, SAM taip pat turi nebūti, kad žmogaus DNMT3A ir DNMT3B metiltransferazės galėtų vykdyti 5hmC dehidroksimetilinimą

(Chen ir kt., 2012). Kadangi DNMT3A/3L vykdoma 5caC dekarboksilinimo reakcija buvo slopinama esant fiziologinėms SAM koncentracijoms, mažai tikėtina, kad ji vyksta įprastinėmis sąlygomis *in vivo*. Tačiau gali egzistuoti klinikinės būsenos, mažinančios SAM lygius ir galbūt sudarančios sąlygas DNMT vykdyti 5caC dekarboksilinimą: ląstelinis SAM lygis sumažėja Dauno sindrome; MAT fermentų, katalizuojančių metionino pavertimą SAM, sumažėja sergant ciroze; žiurkių, maitinamų folio rūgšties stokojančiu maistu (dėl kurio sumažėja SAM kiekis), kepenyse pastebėti priešvėžiniai pokyčiai, susiję su globaliu hipometiliniu (van der Wijst ir kt., 2015). Intensyvi ląstelių proliferacija sukelia vietinę hipoksiją naviko viduje, kuri sumažina SAM lygį hepatomos ląstelėse ir sukelia genomines DNR demetilinimą CpG salose (Liu ir kt., 2011). Taip pat pasiūlyta, kad veikiant įvairiems baltymams, chromatino mikroaplinkoje gali susidaryti neįprastos ribinės medžiagų koncentracijos, kurios galbūt sudaro sąlygas DNMT veikti DNR dehidrosimetilinu ar dekarboksilinu lokaliame lygyje (van der Wijst ir kt., 2015). Pavyzdžiui, vykstant histonų demetilinimui išskiriamas H<sub>2</sub>O<sub>2</sub> (van der Wijst ir kt., 2015), kuris galbūt galėtų padėti DNMT3 efektyviau vykdyti 5hmC dehidrosimetilinu (Chen ir kt., 2012). Jei vyktų tokie lokalūs SAM koncentracijų pokyčiai, svyravimai galėtų paveikti DNMT3A ar 3B funkciją (van der Wijst ir kt., 2015). *In vivo* įrodymų, kad DNMT iš tikrųjų dalyvauja 5caC dekarboksilinu, dar nėra, tačiau *in vitro* nustatytas aktyvumas parodo chemiškai ir fermentiškai įmanomą tiesioginę 5caC pavertimo į C reakciją. Taigi ląstelėse 5caC dekarboksilinimo reakciją galbūt galėtų atlikti specialus fermentas, naudojantis MTazėms analogišką reakcijos mechanizmą.

#### **4.2 Sukurtas eM.SssI 5caC dekarboksilinimo reakcija paremtas 5caC profiliavimo metodas caCLEAR**

Siekiant geriau suprasti aktyvų žinduolių DNR demetilinimą, reikalingi jautrūs, didelės skiriamosios gebos metodai viso genomo oksiduotų citozino formų profiliavimui. Mūsų tikslas buvo sukurti jautraus viso genomo 5caCG nustatymo strategiją išvengiant bisulfito konversijos, dėl kurios kyla daug techninių ir bioinformatinės analizės iššūkių. eM.SssI gebėjimas efektyviai vykdyti 5caC dekarboksilinimą įvairiai modifikuotuose (kai 5caC yra vienoje grandinėje, o nmodifikuotas ar metilintas citozinas kitoje, arba kai 5caC yra abiejose grandinėse) CG taikiniuose ir perkelti aktyvią azido grupę nuo sintetinio SAM kofaktoriaus analogo sudaro sąlygas kovalentiniam 5caCG taikinių žymėjimui. Remiantis šiomis švelniomis fermentinėmis reakcijomis, sukūrėme 5caCG pozicijas genome identifikuojantį caCLEAR



metodą: po eM.SssI dekarboksilavimo susidarę nemodifikuoti CG taikiniai nustatomi TOP-seq profiliavimu, o natūralūs genominiai nCG ir 5hmCG taikiniai neįtraukiami į analizę atitinkamai atliekant jų wt M.SssI metilinimą ir βGT glikozilinimą. Atskirų žingsnių veiksmingumą bei viso metodo specifiskumą, jautrumą ir efektyvumą patvirtinome modelinėse DNR sistemose, įvairiuose taikinių kontekstuose. Kadangi caCLEAR specifiskai nustato tik 5caCG taikinius, reikia mažesnių sekoskaitos ir bioinformatinės analizės apimčių, palyginus su kitais viso genomo sekoskaitos reikalaujančiais 5caC profiliavimo metodais, paremtais bisulfito (Wu ir kt., 2016) ar piridino borano konversija (Liu ir kt., 2019, 2021). Taigi caCLEAR suteikia ekonomišką, o kartu ir didelės skiriamosios gebos 5caC profilių sudarymą.

### **4.3 caCLEAR metodas atskleidė 5caC pasiskirstymo skirtumus esant dviems pelės embrioninių ląstelių pliuripotentiškumo būsenoms**

Pelės ESC caCLEAR analizė parodė, kad 5caC gali būti gausesnis, nei buvo manyta anksčiau. Nustatėme daugiau 5caC taikinių nei aptikta bisulfito konversija paremtu CAB-seq metodu (Lu ir kt., 2015b). CAB-seq metode 5caC regionų praturtinimui dar papildomai naudojami antikūnai, tai mažina reikalingų nuskaitymų skaičių, bet nustatomi ne visi taikiniai. caCLEAR metodo jautrumas taip pat leido aptikti 5caCG laukinio tipo mESC, todėl norint aptikti karboksilintas CG vietas nebėra būtinas TDG išveiklinimas. Be to, analizuojant mESC genominius regionus, stebėjome, jog caCLEAR geba aptikti vos kelių procentų 5caC modifikavimą, matomą bisulfito ar piridino borano konversija paremta sekoskaita. Stebėjome didelį nustatytų 5caCG taikinių persidengimą su 5caC vietomis, parodytomis kitais metodais bei publikuotus duomenis atitinkantį 5caCG praturtinimą genominiuose genomo elementuose, o tai dar kartą patvirtina caCLEAR metodo specifiskumą.

Dviejų pliuripotentiinių mESC būsenų 5caCG genomis pasiskirstymas atskleidė jų vyraujančią kaupimąsi stiprikluose ir kituose atviruose genomo regionuose, kurie paprastai yra sujungti transkripcijos faktorių tinklu, o tai rodo dinamišką 5mC kaitą šiuose lokusuose. Nustatėme, kad globalus 5caCG pasiskirstymas yra veikiamas pliuripotencijos būsenos. Keletas tyrimų pabrėžė transkripcijos ir DNR modifikacijų pokyčius ląstelėms pereinant iš vienos pliuripotentiškumo būsenos į kitą (Marks ir kt., 2012; Ficiz ir kt., 2013). Stebėjome, kad daugumos svarbių veiksnių ir transkripcijos faktorių jungimosi regionai, kuriuose yra 5caCG (5caCG-VJR), serumo-2i mESC buvo toliau nuo genų nei serumo sąlygomis. Vadinas, 2i sąlygomis

daugiau 5caC susidarė tolimųjų stipriklių reguliacinių elementų VJR ir tai sutampa su 2i sąlygomis nustatytu aktyviu DNR demetiliniu stiprikliuose bei labiau demetilinta būseną (Habibi ir kt., 2013; von Meyenn ir kt., 2016; Galonska ir kt., 2015). O tuomet serumo sąlygomis didesnė dalis 5caC susidarė VJR, kurie yra TSS artimesniuose stiprikliuose ar promotoriuose, tai atitinka tik serumo sąlygomis aktyviuose promotoriuose mūsų matomą 5caC praturtinimą bei vykstant kamieninių ląstelių diferenciacijai nustatytą 5caC susidarymą aktyvių genų promotoriuose (Lewis ir kt., 2017b; Wheldon ir kt., 2014). Be to, 5caCG-VJR abejomis sąlygomis buvo toliau nuo genų nei 5caC neturintys regionai (išskyrus Nanog, Oct4, Sox2 ir CTCF regionus serumo sąlygomis). Tai sutampa su duomenimis, kad didžioji 5caC signalo dalis yra stipriklių tolimuosiuose reguliaciniuose genomo elementuose, kuriuose ir yra didelė dalis VJR (Shen ir kt., 2013; Chen ir kt., 2008; Galonska ir kt., 2015). Serumo sąlygomis matomi genams tolimesni 5caC neturintys Nanog, Oct4 ir Sox2 VJR tikriausiai atspindi parodytą šių faktorių jungimąsi prie jau mažai metilintų serumo sąlygoms specifinių mESC stipriklių, kuriuose pereinant iš 2i sąlygų 5mC lygis beveik nekinta ir todėl DNR demetilinimas mažas (Galonska ir kt., 2015). Nors serumo mESC 5caCG yra mažiau, palyginus su serumo-2i mESC, „apmokytose“ mESC 5caCG praturtintas genuose, kurie svarbūs ląstelių vystymosi procesams. Be to, 5caCG-VJR kaupiasi aktyviuose stiprikliuose ir promotoriuose „apmokytose“ mESC, tuo tarpu tokie VJR apsiriboja pasiruošusiais stiprikliais naivioje būsenoje.

Kadangi 5caCG praturtintas įvairių pagrindinių veiksnių ir transkripcijos faktorių surišimo srityse, iškėlėme hipotezę, kad 5caC dalyvauja veiksnių valdomame genų raiškos reguliavime. Nustatėme tendenciją, kad 5caCG-VJR yra praturtinti vidutinės raiškos genuose ir praturtinimas yra stipresnis palyginus su 5caCG neturinčiais VJR. Galbūt tai gali būti susiję su transkripcijos ilgėjimo trikdymu gene esant 5caC (Kellinger ir kt., 2012). Įdomu, kad dalis 5caCG modifikuotų CTCF, c-Myc ir Nanog jungimosi vietų buvo linkusios pasiskirstyti genuose, susijusiuose su ląstelių vystymosi procesais, išskirtinai „apmokytose“ serumo sąlygų mESC. Todėl galbūt pritraukdamos šiuos faktorius 5caCG modifikuotos sritys dalyvauja genų reguliavime mESC keičiantis į labiau „apmokytą“ būseną ir paruošia ląsteles vystymosi procesų aktyvavimui. 5caCG-CTCF sritys buvo išskirtinai stipriai praturtintos egzonuose, o 5caC taip pat labiausiai praturtintas egzonuose lyginant su intronais ir tarpgeninėmis sritimis. Tai sutampa su duomenimis, kad CTCF stipriausiai jungiasi prie 5caC modifikuotos DNR (Marina ir kt., 2016), o 5caC taip pat sukelia neoptimalių CTCF sekų atpažinimą ir gali būti šio baltymo diferencinio jungimosi mechanizmas

ląstelėse (Nanan ir kt., 2019). Įdomu, jog CTCF jungimasis prie 5caC/5hmC alternatyviuose egzonuose sukelia jų įtraukimą (Marina ir kt., 2016), o tai parodo dar vieną galimą 5caC vaidmenį genų reguliacijoje.

Apibendrinant, nustatėme 5caCG ir 5caCG-VJR ryšį su aktyvia genų raiška. Be to, mes nustatėme asimetrinį 5caCG pasiskirstymą stiprios raiškos baltymus koduojančių genų antiprasminėje grandinėje, kuris buvo ryškesnis serumo-2i mESC. Reikia tolesnių tyrimų, kad būtų nustatyta, ar ši tendencija būdinga diferencijuotiems pelių audiniams ar kitiems organizmams. Genai, kuriuose yra ryškus grandinei specifinis 5caCG pasiskirstymas serumo-2i mESC, daugiausia yra susiję su bendromis pagrindinėmis ląstelės funkcijomis, tokiomis kaip medžiagų apykaita, signalų perdavimas ir ląstelės ciklo kontrolė. Asimetrinį oksi-mC išsidėstymą atitinka nustatyta TET kristalinė struktūra su simetriškai metilintu DNR substratu, prie kurio prisijungęs vienas baltymas išsuka metilintą bazę oksidacijai tik iš vienos grandinės (Hu ir kt., 2013). Asimetrinį 5caCG pasiskirstymą turinčių genų antiprasminėje grandinėje taip pat nustatėme didesnį nCG kiekį, o tuo tarpu 5hmC linkęs pasiskirstyti prasminėje genų grandinėje (Booth ir kt., 2012; Gibas ir kt., 2020; Wen ir kt., 2014). Tai rodo skirtingą 5mC oksidacijos ir demetilinimo reguliaciją tarp geno grandinių, galbūt svarbią palaikant genų raišką ESC. Dar turi būti ištirta, kaip susidaro toks asimetrinis oksi-mC pasiskirstymas, galbūt tam turi įtakos kažkokių būdu skirtingai reguliuojami TET/TDG, jų procesyvumas ar dalyvaujantys kiti DNR demetilinimo mechanizmai. Asimetriją galėtų lemti skirtingas grandinių prieinamumas, pavyzdžiui, transkripcijos metu susidarant R-kilpos struktūrai: koduojanti prasminė DNR grandinė atsiskiria, o matricinė antiprasminė grandinė sudaro hibridą su nurašoma RNR. Iš tiesų yra parodyta, kad TET baltymai yra nukreipiami į R-kilpas oksidacijai bei jose aptinkamas oksi-mC praturtinimas (Arab ir kt., 2019). Taip pat nustatyta, kad R-kilpos yra būdingos aktyviems ESC genams ir jų išsidėstymas genome koreliuoja su 5hmC genominiu pasiskirstymu, o 5hmC taip pat skatina R-kilpų susidarymą (Sabino ir kt., 2022). TET baltymai gali veikti su įvairiais substratais – viengrande DNR, RNR/DNR hibridais, nors efektyvumas geriausias dvigrandėje DNR (Xue ir kt., 2020). Tad įmanoma, kad susidarius 5hmC ir susiformavus R-kilpai, tolimesnė oksidacija nukreipiama į priešingą grandinę, sudarančią hibridą su RNR.

Apibendrinant, caCLEAR parodė, kad 5caC yra aktyvios genų raiškos ženklas, kuris, pritraukdamas veiksnius į TET oksiduotus genomo taikinius, dalyvauja genų reguliacijoje keičiantis mESC pliuripotencijos būsenoms.

Todėl 5caC gali būti geras dinamiško ląstelės DNR modifikavimo ir chromatinio pertvarkymo indikatorius.

#### **4.4 Oksiduotų 5-metilcitozino formų pasiskirstymas ir jų vaidmuo *L. bicolor* ir *C. cinerea* papėdėgrybių nukleorūgštyse**

*L. bicolor* ir *C. cinerea* grybuose siekėme iširti oksi-mC gausą, pasiskirstymą, galimas funkcijas bei tuo pačiu suprasti šiuose grybuose gausių *TET-JPB* genų reikšmę. Abiejų grybų micelio DNR aptikome visas tris oksi-mC, o 5hmC ir 5fC lygiai buvo panašūs, skirtingai nei žinduolių DNR, kurioje 5hmC daug gausnis palyginus su 5fC (3.19 pav. E). Tai gali reikšti kitokią 5hmC ir 5fC kaitą arba biologinius vaidmenis grybuose. Priešingai nei žinduoliai, kurie turi tik tris *TET* genus, *L. bicolor* ir *C. cinerea* yra daug *TET-JPB* genų kopijų (47-74 genus) (Iyer ir kt., 2009; Wu ir Zhang, 2014). Nepaisant daugybės *TET-JBP* genų, tik keli iš jų pasižymėjo silpna raiška (šis tyrimas ir Chavez ir kt., 2014), todėl *L. bicolor* ir *C. cinerea* stebėtas santykinai mažas oksi-mC kiekis nėra netikėtas. 5caC kiekis *C. cinerea* DNR buvo mažiausias iš trijų oksi-mC modifikacijų (3.19 pav. A, E), priešingai nei Chavez ir kt. (2014) nustatytas gausiausias 5caC. Aptiktą 5caC gausumo neatitikimą tarp HPLC-MS/MS ir taškinio imunobloto galima paaiškinti naudojamų 5caC antikūnų kryžminiu reaktyvumu, o tai kelia abejonę dėl antikūnais pagrįstų grybų mėginių analizės metodų patikimumo. Dėl šios priežasties išsamiam *L. bicolor* ir *C. cinerea* DNR ir RNR oksi-mC apibūdinimui mes naudojome HPLC-MS/MS metodą.

Atradome 5carC *C. cinerea* RNR ir sukūrėme cheminės derivatizacijos metodą, kuris palengvintų šios rečiausios ir mažiausiai iširtos RNR oksi-mC aptikimą HPLC-MS/MS. Tarp skirtingų RNR tipų 5carC stebėjome ilgose RNR po poli A pašalinimo, o tai rodo galimą 5carC ryšį su nekoduojančiomis RNR arba iRNR, kurios neturi 3'-poli A sekų ar jos yra trumpos. Parodyta, kad kai kuriuose grybuose yra nemaža dalis iRNR, neturinčių 3'-poli A sekų (Griffin, 1996), o trumpos poli A „uodegos“ yra paplitusios stiprios raiškos eukariotų genuose (Lima ir kt., 2017). 5carC signalas taip pat buvo matomas trumpose *C. cinerea* RNR, o tarp jų didžiausi 5carC kiekiai buvo RNR, trumpesnėse nei tRNR. Poveikis vitaminu C grybuose sukėlė oksi-mC padidėjimą RNR ir DNR, o tai paremia galimą TET baltymų dalyvavimą grybų 5mC oksidacijoje. Standartinėmis auginimo sąlygomis mūsų metodu *L. bicolor* 5carC neaptikome, tačiau vitC paskatino didelių 5carC kiekių susidarymą ilgose RNR. Vadinasi, ši modifikacija gali susidaryti esant padidėjusiai vitC koncentracijai ląstelėse (grybai natūraliai

sintetina vitC arba jo analogus; Loewus, 1999) arba, pavyzdžiui, esant jų sąveikai su augalu. Įdomu, jog *C. cinerea* poli A RNR frakcijoje neaptikome 5hmrC ir jį nustatėme po veikimo vitC tik *L. bicolor*, o tai rodo kitokį oksimC vaidmenį grybuose lyginant su žinduoliais, kur didžiausias 5scarC ir 5hmrC kiekis buvo nustatytas poli A RNR (Huang ir kt., 2016). Nors 5scarC lygiai bendroje *C. cinerea* (~0,00001-0,00005% nuo G) ir *L. bicolor* (~0,0004%) RNR yra žemi, jie patenka į panašų diapazoną, kaip ir skirtinguose žinduolių audiniuose (0,00004-0,0005%; Huang ir kt., 2016) (3.19 pav. E). *In vitro* parodytas santykinai silpnas *C. cinerea* TET baltymų aktyvumas DNR generuojant 5caC (Zhang ir kt., 2014) galbūt lemia mažiausius 5scarC arba 5caC kiekius tarp visų oksimC. Žinduoliuose 5caC gali būti aktyviai pašalinamas bazių iškirpimo mechanizmais, pavyzdžiui, veikiant timino DNR glikozilazei arba vykstant tiesioginiam 5caC dekarboksilinimui į citoziną (Feng ir kt., 2021c). Tačiau iki šiol grybuose timino DNR glikozilazių (*C. cinerea* nustatytas galimas genas; Chavez ir kt., 2014) aktyvumas ar kiti 5caC pašalinimo mechanizmai dar nenustatyti.

Mūsų 5scarC tyrimas grybų RNR prideda vertingos informacijos apie šią menkai ištirtą modifikaciją, nors jos biologinio veikimo mechanizmai dar turi būti išaiškinti. Galbūt 5scarC gali paveikti RNR struktūrą, panašiai kaip 5caC DNR, išplėsdamas mažąjį griovį ir atlaisvindamas dvigrandes dalis (Dai ir kt., 2015), priešingai nei labiau stabilizuojantis 5mC (Harcourt ir kt., 2017; Fu ir kt., 2019). Įvairiuose organizmuose parodyta, jog 5-citozino modifikacijos daro įtaką RNR savybėms, stabilumui ir funkcijoms. Pavyzdžiui, 5mrC veikia tRNR, rRNR susilankstymą ir stabilizuoja antrinę struktūrą, todėl tikėtina veikia ir aptinkamas iRNR struktūras, kurios gali daryti įtaką iRNR brendimui bei transliacijai (Gilbert ir kt., 2016). iRNR esant 5mrC jos lengviau pernešamos iš žmogaus HeLa ląstelių branduolio, o *A. thaliana* – į tolimes augalo dalis (Xue ir kt., 2020). Priešingai nei 5mrC, kuris sumažina iRNR transliaciją, 5hmrC gali ją skatinti *Drosophila* (Delatte ir kt., 2016). Kai kurie eksperimentai rodo galimą 5mrC įtaką iRNR stabilumui (Xue ir kt., 2020), o 5hmrC destabilizuoja pelių endogeninių retrovirusų transkriptus (Amort ir kt., 2013; Guallar ir kt., 2018). 5mrC aptinkamas įvairiose reguliacinėse RNR (miRNR, stipriklių, skliautinėse, ilgose nekoduojančiose RNR) ir yra duomenų apie 5mrC poveikį jų stabilumui, funkcijoms bei dalyvavimui genų reguliacijoje (Xue ir kt., 2020). Pavyzdžiui, 5mrC daro įtaką ilgų nekoduojančių reguliacinių RNR prisijungimui prie chromatiną modifikuojančių kompleksų (Amort ir kt., 2013). Dinamiškai kintančios oksimC modifikacijos galbūt galėtų būti greitas būdas struktūrinių

5mC modifikuotos RNR savybių pakeitimui ar surišančių baltymų sąveikos moduliavimui taip keičiant RNR funkcijas ir atsakant į aplinkos veiksmus.

TE sudaro didesnę ektomikorizinės *L. bicolor* genomo dalį (iki 37,9%, Castanera ir kt., 2016), lyginant su saprotofiniu *C. cinerea* (~6%, Stajich ir kt., 2010). Manoma, kad TE gausėjimas grybuose yra susijęs su abipusiai naudinga arba parazitine sąveika su augalais, taip prisidedant prie genetinės variacijos, naudingos prisitaikymui prie augalo-šeimininko (Castanera ir kt., 2017). Ektomikoriziniuose simbiozuose nustatytas žymiai didesnis TE kiekis, palyginus su kitais grybų ekologinio prisitaikymo būdais (Miyachi ir kt., 2020). Mūsų atliktas 5hmCG ir 5fCG dinukleotidų profiliavimas hmTOP-seq (Gibas ir kt., 2020) ir sukurtu foTOP-seq metodu abiejų grybų DNR parodė, kad oksimC kaupiasi TE, o tai sutampa su paskelbtu abiejų grybų 5mC (Zemach ir kt., 2010) ir *C. cinerea* 5hmC pasiskirstymu (Chavez ir kt., 2014). Nustatėme, kad dauguma šių grybų genų turi oksimC. Genų buvimas arti oksimC turtingų TE buvo susijęs su genų slopinimu ir didesniu 5hmC/5fC modifikacijų praturtinimu, o tai atitinka kitų grybų 5mC duomenis (Borgognone ir kt., 2018; Castanera ir kt., 2016). Taip pat abiejuose grybuose nustatėme didesnę genų slopinimą jiems esant TE klasteriuose, palyginus su genais, esančiais TE klasterių išorėje ar neturinčiais TE.

Analizuodami modifikuotų CG frakcijas, galėjome tiesiogiai palyginti trijų citozino būsenų – 5mC, 5hmC ir 5fC – kiekius ir jų ryšį su genų raiškos reguliavimu. Atskirų citozino modifikacijų analizė genuose atskleidė, kad didelis geno 5hmC arba 5fC, bet ne 5mC modifikavimas daro stiprų slopinamąjį poveikį raiškai *L. bicolor*. Tuo tarpu *C. cinerea* genai esant silpnam metilinizavimui parodė palaiptumui didėjančią raišką daugėjant 5hmC ir tik esant stipriam 5mCG modifikavimui aukštas geno 5hmCG lygis neigiamai veikė raišką. Nepaisant stipraus genų slopinančio oksimCG vaidmens, stebėto abiejuose grybuose, aktyviai raiškai yra reikalingas nedidelis 5mC/5hmC/5fC modifikavimo lygis.

Mūsų žiniomis, tai yra pirmoji tokia išsami 5mC ir jo oksiduotų formų 5hmC/5fC išsidėstymo ir jų poveikio raiškai analizė grybuose. Stebėjome genų, kurių aukštą oksimC lygį lemia stipriai 5hmCG/5fCG modifikuoti TE, nutildymą. Todėl galima manyti, jog 5mC oksidacija gali veikti kaip papildomas TE reguliavimo/slopinimo mechanizmas, kuris grybuose veikia, pavyzdžiui, į TE nukreipiant represinius baltymus, kaip buvo parodyta TET baltymams žinduoliuose (Deniz ir kt., 2019). Šis mechanizmas taip pat gali būti svarbus autoreguliuojant su TE susijusius (Iyer ir kt., 2013, 2014) ir 5hmC/5fC modifikuotus gausius *TET-JPB* genus. Evoliucijos metu

simbiozėje su augalais *L. bicolor* galėjo atsirasti keletas mechanizmų, kurie kartu su TE kaupimu, prisidedančiu prie adaptacijos, suaktyvina stiprią slopinančią kontrolę prieš žalingą gausių TE poveikį. Skirtingas oksimC ir 5mC poveikis genų raiškai tarp *L. bicolor* ir *C. cinerea*, dar kartą patvirtina, kad citozino modifikacijos gali atlikti įvairius vaidmenis skirtingų rūšių grybuose (Bewick ir kt., 2019; Zemach ir kt., 2010).

Nustatytas 5mC ir oksimC modifikacijų pasiskirstymas kartu su dviejų grybų genų raiškos profiliais yra unikalus šaltinis tiriant epigenetinius mechanizmus, kuris pabrėžia individualių citozino modifikacijų svarbą. Norint gauti gilesnių įžvalgų apie oksimC vaidmenų įvairovę ir pasiskirstymą kitose *L. bicolor* ir *C. cinerea* vystymosi stadijose ir daugelyje skirtingų grybų, reikia atlikti tolimesnius tyrimus.

## IŠVADOS

1. 5caC dekarboksilinimui esant DNMT3A/3L ar DNMT3B/3L žmogaus metiltransferazėms reikalingas jų katalitiniame centre esantis cisteinas, bet reakcija yra slopinama SAM. Todėl nors metiltransferazės geba atlikti dekarboksilinimą, mažai tikėtina, kad įprastinėmis sąlygomis jos vykdo reakciją *in vivo*.
2. Remiantis eM.SssI vykdoma 5caC dekarboksilinimo reakcija, sukurtas bei patvirtintas naujas, jautrus ir ekonomiškai CG skiriamosios gebos caCLEAR metodas specifiniam 5caCG taikinių profiliavimui genominiame DNR.
3. caCLEAR metodu buvo nustatyti genominio 5caC pasiskirstymo skirtumai tarp serumo-2i ir serumo sąlygomis augintų pelės embrioninių kamieninių ląstelių, atitinkančių dvi pluriipotentiško būsenas – „naivia“ ir „apmokyta“. Parodytas skirtingas 5caC praturtinimas genomo elementuose, išsidėstymas svarbių veiksnių jungimosi regionuose bei išskirtinai serumo sąlygomis nustatytas 5caC praturtinimas „apmokytoje“ būsenoje aktyvesniuose vystymosi procesų genuose. 5caC yra susijęs su aktyvia raiška ir pasižymi asimetriniu išsidėstymu antiprasminėje aktyvių baltymų koduojančių genų grandinėje.
4. *C. cinerea* ir *L. bicolor* papėdgrybių RNR aptiktos oksiduotos 5mC formos. Pritaikius cheminę derivatizaciją sukurtas jautrus masių spektrometrijos metodas 5scarC nustatymui RNR. Šiuo metodu nustatyta, kad 5scarC daugiausia yra abiejų grybų 3'-poli A sekų neturintiose ilgesnėse nei 200 nt RNR, *L. bicolor* poli A RNR bei *C. cinerea* trumpesnėse nei 200 nt RNR, tarp kurių 5scarC gausiausias RNR, trumpesnėse nei tRNR.
5. *L. bicolor* ir *C. cinerea* DNR nustatytos visos trys oksiduotos 5mC formos, iš kurių gausiausios 5hmC ir 5fC. Naudojant hmTOP-seq ir sukūrus modifikuotą jo versiją foTOP-seq, atliktas 5hmC ir 5fC genominis profiliavimas *L. bicolor* ir *C. cinerea*. Parodytas 5hmC ir 5fC praturtinimas judriuosiuose elementuose ir arti jų esančių genų raiškos slopinimas. *C. cinerea* nustatyta stipraus 5hmCG modifikavimo sąsaja su aktyviais genais esant mažai 5mCG, o *L. bicolor* didelis 5fCG/5hmCG modifikavimas susijęs su slopinamais genais, nepriklausomai nuo jų 5mCG gausumo.



## LITERATŪROS SĄRAŠAS

1. Aguilo, F., Li, S., Balasubramanian, N., Sancho, A., Benko, S., Zhang, F., Vashisht, A., Rengasamy, M., Andino, B., Chen, C.-H., Zhou, F., Qian, C., Zhou, M.-M., Wohlschlegel, J. A., Zhang, W., Suchy, F. J., ir Walsh, M. J. (2016). Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1 $\alpha$ . *Cell Reports*, 14(3), 479–492.
2. Alioui, A., Wheldon, L. M., Abakir, A., Ferjentsik, Z., Johnson, A. D., ir Ruzov, A. (2012). 5-Carboxylcytosine is localized to euchromatic regions in the nuclei of follicular cells in axolotl ovary. *Nucleus*, 3(6), 565–569.
3. Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., Jia, X.-Y., Micura, R., ir Lusser, A. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biology*, 18(1), 1.
4. Amort, T., Soulière, M. F., Wille, A., Jia, X.-Y., Fiegl, H., Wörle, H., Micura, R., ir Lusser, A. (2013). Long non-coding RNAs as targets for cytosine methylation. *RNA Biology*, 10(6), 1003–1008.
5. An, J., Rao, A., ir Ko, M. (2017). TET family dioxygenases and DNA demethylation in stem cells and cancers. *Experimental and Molecular Medicine*, 49(4), e323.
6. Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., ir Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nature Genetics*, 51(2), 217–223.
7. Atlasi, Y., Megchelenbrink, W., Peng, T., Habibi, E., Joshi, O., Wang, S.-Y., Wang, C., Logie, C., Poser, I., Marks, H., ir Stunnenberg, H. G. (2019). Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nature Cell Biology*, 21(5), 568–578.
8. Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H. E., Iurlaro, M., Reik, W., Murrell, A., ir Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nature Chemical Biology*, 11(8), 555–557.
9. Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., ir Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature Chemistry*, 6(12), 1049–1055.
10. Basanta-Sanchez, M., Wang, R., Liu, Z., Ye, X., Li, M., Shi, X., Agris, P. F., Zhou, Y., Huang, Y., ir Sheng, J. (2017). TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. *Chembiochem: A European Journal of Chemical Biology*, 18(1), 72–76.

11. Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V., Gillis, D. R., ir Phillips-Cremins, J. E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Research*, 27(7), 1139–1152.
12. Beaulaurier, J., Schadt, E. E., ir Fang, G. (2018). Deciphering bacterial epigenomes using modern sequencing technologies. *Nature Reviews. Genetics*, 20(3), 157–172.
13. Bewick, A. J., Hofmeister, B. T., Powers, R. A., Mondo, S. J., Grigoriev, I. V., James, T. Y., Stajich, J. E., ir Schmitz, R. J. (2019). Diversity of cytosine methylation across the fungal tree of life. *Nature Ecology and Evolution*, 3(3), 479–490.
14. Bochtler, M., Kolano, A., ir Xu, G.-L. (2017). DNA demethylation pathways: Additional players and regulators. *Bioessays*, 39(1):1-13.
15. Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., ir Balasubramanian, S. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science*, 336(6083), 934–937.
16. Booth, M. J., Marsico, G., Bachman, M., Beraldi, D., ir Balasubramanian, S. (2014). Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nature Chemistry*, 6(5), 435–440.
17. Booth, M. J., Raiber, E.-A., ir Balasubramanian, S. (2015). Chemical methods for decoding cytosine modifications in DNA. *Chemical Reviews*, 115(6), 2240–2254.
18. Borgognone, A., Castanera, R., Morselli, M., López-Varas, L., Rubbi, L., Pisabarro, A. G., Pellegrini, M., ir Ramírez, L. (2018). Transposon-associated epigenetic silencing during *Pleurotus ostreatus* life cycle. *DNA Research*, 25(5), 451–464.
19. Breiling, A., ir Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics and Chromatin*, 8(1), 1–9.
20. Brunner, A. L., Johnson, D. S., Kim, S. W., Valouev, A., Reddy, T. E., Neff, N. F., Anton, E., Medina, C., Nguyen, L., Chiao, E., Oyolu, C. B., Schroth, G. P., Absher, D. M., Baker, J. C., ir Myers, R. M. (2009). Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Research*, 19(6), 1044–1056.
21. Cadet, J., ir Wagner, J. R. (2014). TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine.

*Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 764-765, 18–35.

22. Caldwell, B. A., Liu, M. Y., Prasasya, R. D., Wang, T., DeNizio, J. E., Leu, N. A., Amoh, N. Y. A., Krapp, C., Lan, Y., Shields, E. J., Bonasio, R., Lengner, C. J., Kohli, R. M., ir Bartolomei, M. S. (2021). Functionally distinct roles for TET-oxidized 5-methylcytosine bases in somatic reprogramming to pluripotency. *Molecular cell*, 81(4), 859–869.e8.
23. Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A. P., Fabris, D., ir Agris, P. F. (2011). The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Research*, 39(Database issue), D195–D201.
24. Cao, J. Z., Hains, A. E., ir Godley, L. A. (2019). Regulation of 5-Hydroxymethylcytosine Distribution by the TET Enzymes. *RNA Technologies*, 111(48):E5149-58.
25. Carell, T., Kurz, M. Q., Müller, M., Rossa, M., ir Spada, F. (2018). Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angewandte Chemie*, 57(16), 4296–4312.
26. Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., ir Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development*, 132(5), 885–896.
27. Castanera, R., Borgognone, A., Pisabarro, A. G., ir Ramírez, L. (2017). Biology, dynamics, and applications of transposable elements in basidiomycete fungi. *Applied Microbiology and Biotechnology*, 101(4), 1337–1350.
28. Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., Stajich, J. E., ir Ramírez, L. (2016). Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLoS Genetics*, 12(6), e1006108.
29. Chavez, L., Huang, Y., Luong, K., Agarwal, S., Iyer, L. M., Pastor, W. A., Hench, V. K., Frazier-Bowers, S. A., Korol, E., Liu, S., Tahiliani, M., Wang, Y., Clark, T. A., Korlach, J., Pukkila, P. J., Aravind, L., ir Rao, A. (2014). Simultaneous sequencing of oxidized methylcytosines produced by TET/JBP dioxygenases in *Coprinopsis cinerea*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(48), E5149–E5158.
30. Chen, C.-C., Wang, K.-Y., ir Shen, C.-K. J. (2012). The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-

- hydroxymethylcytosine dehydroxymethylases. *The Journal of Biological Chemistry*, 287(40), 33116–33121.
31. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y.-H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., ... Ng, H.-H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6), 1106–1117.
  32. Cheray, M., Etcheverry, A., Jacques, C., Pacaud, R., Bougras-Cartron, G., Aubry, M., Denoual, F., Peterlongo, P., Nadaradjane, A., Briand, J., Akcha, F., Heymann, D., Vallette, F. M., Mosser, J., Ory, B., ir Cartron, P.-F. (2020). Cytosine methylation of mature microRNAs inhibits their functions and is associated with poor prognosis in glioblastoma multiforme. *Molecular Cancer*, 19(1), 36.
  33. Clarke, S., ir Banfield, K. (2001). S-adenosylmethionine-dependent methyltransferases. *Homocysteine in Health and Disease* (p. 62–77). Cambridge University Press.
  34. Daguerre, Y., Levati, E., Ruytinx, J., Tisserant, E., Morin, E., Kohler, A., Montanini, B., Ottonello, S., Brun, A., Veneault-Fourrey, C., ir Martin, F. (2017). Regulatory networks underlying mycorrhizal development delineated by genome-wide expression profiling and functional analysis of the transcription factor repertoire of the plant symbiotic fungus *Laccaria bicolor*. *BMC Genomics*, 18(1):737.
  35. Dai, Q., Sanstead, P. J., Peng, C. S., Han, D., He, C., ir Tokmakoff, A. (2015). Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chemical Biology*, 11(2), 470–477.
  36. Dalhoff, C., Lukinavičius, G., Klimašauskas, S., ir Weinhold, E. (2006). Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nature Chemical Biology*, 2(1), 31–32.
  37. Deen, J., Vranken, C., Leen, V., Neely, R. K., Janssen, K. P. F., ir Hofkens, J. (2017). Methyltransferase-Directed Labeling of Biomolecules and its Applications. *Angewandte Chemie*, 56(19), 5182–5200.
  38. Delatte, B., ir Fuks, F. (2013). TET proteins: on the frenetic hunt for new cytosine modifications. *Briefings in Functional Genomics*, 12(3), 191–204.
  39. Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., Wetzel, C., Kreher, J., Soin, R., Creppe, C., Limbach, P. A., Gueydan, C., Kruijs, V., Brehm, A.,

- Minakhina, S., ... Fuks, F. (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science*, 351(6270), 282–285.
40. de Mendoza, A., Lister, R., ir Bogdanovic, O. (2019). Evolution of DNA Methylome Diversity in Eukaryotes. *Journal of Molecular Biology*, S0022-2836(19)30659-X.
  41. Deniz, Ö., Frost, J. M., ir Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nature Reviews. Genetics*, 20(7), 417–431.
  42. Dezi, V., Ivanov, C., Haussmann, I. U., ir Soller, M. (2016). Nucleotide modifications in messenger RNA and their role in development and disease. *Biochemical Society Transactions*, 44(5), 1385–1393.
  43. Dick, A., ir Chen, A. (2021). The role of TET proteins in stress-induced neuroepigenetic and behavioural adaptations. *Neurobiology of Stress*, 15, 100352.
  44. Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O., ir Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m<sup>5</sup>C within archaeal mRNAs. *PLoS Genetics*, 9(6), e1003602.
  45. Edwards, J. R., O'Donnell, A. H., Rollins, R. A., Peckham, H. E., Lee, C., Milekic, M. H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., Gingrich, J. A., Haghghi, F., Nutter, R., ir Bestor, T. H. (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Research*, 20(7), 972–980.
  46. Eguizabal, C., Aran, B., Chuva de Sousa Lopes, S. M., Geens, M., Heindryckx, B., Panula, S., Popovic, M., Vassena, R., ir Veiga, A. (2019). Two decades of embryonic stem cells: a historical overview. *Human Reproduction Open*, 2019(1):hoy024.
  47. Eleftheriou, M., Pascual, A. J., Wheldon, L. M., Perry, C., Abakir, A., Arora, A., Johnson, A. D., Auer, D. T., Ellis, I. O., Madhusudan, S., ir Ruzov, A. (2015). 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clinical Epigenetics*, 7(1):88.
  48. Feng, Y., Chen, J.-J., Xie, N.-B., Ding, J.-H., You, X.-J., Tao, W.-B., Zhang, X., Yi, C., Zhou, X., Yuan, B.-F., ir Feng, Y.-Q. (2021a). Direct decarboxylation of ten-eleven translocation-produced 5-carboxylcytosine in mammalian genomes forms a new mechanism for active DNA demethylation. *Chemical Science*, 12(34), 11322–11329.
  49. Feng, Y., Xie, N.-B., Tao, W.-B., Ding, J.-H., You, X.-J., Ma, C.-J., Zhang, X., Yi, C., Zhou, X., Yuan, B.-F., ir Feng, Y.-Q. (2021b).

- Transformation of 5-Carboxylcytosine to Cytosine Through C–C Bond Cleavage in Human Cells Constitutes a Novel Pathway for DNA Demethylation. *CCS Chemistry*, 3(4), 994–1008.
50. Ficzc, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R., ir Reik, W. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell*, 13(3), 351–359.
  51. Fu, L., Guerrero, C. R., Zhong, N., Amato, N. J., Liu, Y., Liu, S., Cai, Q., Ji, D., Jin, S.-G., Niedernhofer, L. J., Pfeifer, G. P., Xu, G.-L., ir Wang, Y. (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *Journal of the American Chemical Society*, 136(33), 11582–11585.
  52. Fu, T., Liu, L., Yang, Q.-L., Wang, Y., Xu, P., Zhang, L., Liu, S., Dai, Q., Ji, Q., Xu, G.-L., He, C., Luo, C., ir Zhang, L. (2019). Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chemical Science*, 10(31), 7407–7417.
  53. Gagliardi, M., Strazzullo, M., ir Matarazzo, M. R. (2018). DNMT3B Functions: Novel Insights from Human Disease. *Frontiers in Cell and Developmental Biology*, 6:140.
  54. Galonska, C., Ziller, M. J., Karnik, R., ir Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell*, 17(4), 462–470.
  55. Genre, A., Lanfranco, L., Perotto, S., ir Bonfante, P. (2020). Unique and common traits in mycorrhizal symbioses. *Nature Reviews. Microbiology*, 18(11), 649–660.
  56. Gibas, P., Narmontė, M., Staševskij, Z., Gordevičius, J., Klimašauskas, S., ir Kriukienė, E. (2020). Precise genomic mapping of 5-hydroxymethylcytosine via covalent tether-directed sequencing. *PLoS Biology*, 18(4), e3000684.
  57. Gilbert, W. V., Bell, T. A., ir Schaening, C. (2016). Messenger RNA modifications: Form, distribution, and function. *Science*, 352(6292), 1408–1412.
  58. Goll, M. G., ir Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, 74, 481–514.
  59. Gordevičius, J., Narmontė, M., Gibas, P., Kvederavičiūtė, K., Tomkutė, V., Paluoja, P., Krjutškov, K., Salumets, A., ir Kriukienė, E. (2020). Identification of fetal unmodified and 5-hydroxymethylated CG sites in

- maternal cell-free DNA for non-invasive prenatal testing. *Clinical Epigenetics*, 12(1), 153.
60. Greenberg, M. V. C., ir Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews. Molecular Cell Biology*, 20(10), 590–607.
  61. Griffin, D. H. (1996). *Fungal Physiology*. John Wiley and Sons.
  62. Guallar, D., Bi, X., Pardavila, J. A., Huang, X., Saenz, C., Shi, X., Zhou, H., Faiola, F., Ding, J., Haruehanroengra, P., Yang, F., Li, D., Sanchez-Priego, C., Saunders, A., Pan, F., Valdes, V. J., Kelley, K., Blanco, M. G., Chen, L., ... Wang, J. (2018). RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nature Genetics*, 50(3), 443–451.
  63. Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., Walsh, C. P., Li, J., Tang, F., ir Xu, G.-L. (2014a). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell*, 15(4), 447–459.
  64. Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G.-L., ir Song, H. (2014b). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, 17(2), 215–222.
  65. Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., Benedetti, R., Altucci, L., Jansen, J. H., Walter, J., Gut, I. G., Marks, H., ir Stunnenberg, H. G. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell*, 13(3), 360–369.
  66. Harcourt, E. M., Kietrys, A. M., ir Kool, E. T. (2017). Chemical and structural effects of base modifications in messenger RNA. *Nature*, 541(7637), 339–346.
  67. He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., ir Xu, G.-L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 333(6047), 1303–1307.
  68. Helm, M., ir Alfonzo, J. D. (2014). Posttranscriptional RNA Modifications: Playing Metabolic Games in a Cell's Chemical Legoland. *Chemistry and Biology*, 21(2), 174–185.
  69. Helm, M., ir Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. *Nature Reviews. Genetics*, 18(5), 275–291.

70. Hofer, A., Liu, Z. J., ir Balasubramanian, S. (2019). Detection, Structure and Function of Modified DNA Bases. *Journal of the American Chemical Society*, *141*(16), 6420–6429.
71. Holmes, E. E., Jung, M., Meller, S., Lisse, A., Sailer, V., Zech, J., Mengdehl, M., Garbe, L.-A., Uhl, B., Kristiansen, G., ir Dietrich, D. (2014). Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One*, *9*(4), e93933.
72. Hu, L., Liu, Y., Han, S., Yang, L., Cui, X., Gao, Y., Dai, Q., Lu, X., Kou, X., Zhao, Y., Sheng, W., Gao, S., He, X., ir He, C. (2019). Jump-seq: Genome-Wide Capture and Amplification of 5-Hydroxymethylcytosine Sites. *Journal of the American Chemical Society*, *141*(22), 8694–8697.
73. Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., Wang, P., Wang, Y., Fang, D., Cui, Q., ... Xu, Y. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature*, *527*(7576):118-22.
74. Huang, W., Lan, M.-D., Qi, C.-B., Zheng, S.-J., Wei, S.-Z., Yuan, B.-F., ir Feng, Y.-Q. (2016). Formation and determination of the oxidation products of 5-methylcytosine in RNA. *Chemical Science*, *7*(8), 5495–5502.
75. Huber, S. M., van Delft, P., Mendil, L., Bachman, M., Smollett, K., Werner, F., Miska, E. A., ir Balasubramanian, S. (2015). Formation and abundance of 5-hydroxymethylcytosine in RNA. *ChemBioChem*, *16*(5):752-5.
76. Inoue, A., Shen, L., Dai, Q., He, C., ir Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Research*, *21*(12), 1670–1676.
77. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., ir Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300–1303.
78. Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.-A., Bachman, M., Booth, M. J., Andrews, S., Balasubramanian, S., ir Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, *14*(10), R119.
79. Iurlaro, M., McInroy, G. R., Burgess, H. E., Dean, W., Raiber, E.-A., Bachman, M., Beraldi, D., Balasubramanian, S., ir Reik, W. (2016). In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biology*, *17*(1), 141.



80. Iwan, K., Rahimoff, R., Kirchner, A., Spada, F., Schröder, A. S., Kosmatchev, O., Ferizaj, S., Steinbacher, J., Parsa, E., Müller, M., ir Carell, T. (2018). 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nature Chemical Biology*, *14*(1):72-78.
81. Iyer, L. M., Tahiliani, M., Rao, A., ir Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, *8*(11), 1698–1710.
82. Iyer, L. M., Zhang, D., Burroughs, A. M., ir Aravind, L. (2013). Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Research*, *41*(16), 7635–7655.
83. Iyer, L. M., Zhang, D., de Souza, R. F., Pukkila, P. J., Rao, A., ir Aravind, L. (2014). Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proceedings of the National Academy of Sciences*, *111*(5):1676-83.
84. Jiang, H. P., Liu, T., Guo, N., Yu, L., Yuan, B. F., ir Feng, Y. Q. (2017). Determination of formylated DNA and RNA by chemical labeling combined with mass spectrometry analysis. *Analytica chimica acta*, *981*, 1–10.
85. Jin, S.-G., Zhang, Z.-M., Dunwell, T. L., Harter, M. R., Wu, X., Johnson, J., Li, Z., Liu, J., Szabó, P. E., Lu, Q., Xu, G.-L., Song, J., ir Pfeifer, G. P. (2016). Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Reports*, *14*(3), 493–505.
86. Jonkhout, N., Tran, J., Smith, M. A., Schonrock, N., Mattick, J. S., ir Novoa, E. M. (2017). The RNA modification landscape in human disease. *RNA*, *23*(12), 1754–1769.
87. Kamińska, E., Korytiaková, E., Reichl, A., Müller, M., ir Carell, T. (2021). Intragenomic Decarboxylation of 5-Carboxy-2'-deoxycytidine. *Angewandte Chemie International Edition*, *60*(43), 23207–23211.
88. Kawarada, L., Suzuki, T., Ohira, T., Hirata, S., Miyauchi, K., ir Suzuki, T. (2017). ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Research*, *45*(12), 7401–7415.
89. Kellinger, M. W., Song, C.-X., Chong, J., Lu, X.-Y., He, C., ir Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature Structural and Molecular Biology*, *19*(8), 831–833.

90. Kellner, S., Burhenne, J., ir Helm, M. (2010). Detection of RNA modifications. *RNA Biology*, 7(2), 237–247.
91. Khare, T., Pai, S., Koncevicus, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., Tice, R., Tochigi, M., Moréra, S., Nazarians, A., Belsham, D., Wong, A. H. C., Blencowe, B. J., Wang, S. C., Kapranov, P., ... Petronis, A. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature Structural and Molecular Biology*, 19(10), 1037–1043.
92. Kim, S., Günesdogan, U., Zylicz, J. J., Hackett, J. A., Cougot, D., Bao, S., Lee, C., Dietmann, S., Allen, G. E., Sengupta, R., ir Ma, S. (2014). PRMT5 Protects Genomic Integrity during Global DNA Demethylation in Primordial Germ Cells and Preimplantation Embryos. *Molecular cell*, 56(4), 564–579.
93. Klimašauskas, S., Kumar, S., Roberts, R. J., ir Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, 76(2), 357–369.
94. Klungland, A., ir Robertson, A. B. (2017). Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radical Biology and Medicine*, 107, 62–68.
95. Kohli, R. M., ir Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502(7472), 472–479.
96. Korytiaková, E., Kamińska, E., Müller, M., ir Carell, T. (2021). Deformylation of 5-Formylcytidine in Different Cell Types. *Angewandte Chemie*, 60(31), 16869–16873.
97. Kriaucionis, S., ir Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 324(5929), 929–930.
98. Kriukienė, E., Labrie, V., Khare, T., Urbanavičiūtė, G., Lapinaitė, A., Koncevičius, K., Li, D., Wang, T., Pai, S., Ptak, C., Gordevičius, J., Wang, S.-C., Petronis, A., ir Klimašauskas, S. (2013). DNA unmethylome profiling by covalent capture of CpG sites. *Nature Communications*, 4, 2190.
99. Kriukienė, E., Liutkeviciūtė, Z., ir Klimašauskas, S. (2012). 5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA. *Chemical Society Reviews*, 41(21), 6916–6930.
100. Kües, U. (2000). Life history and developmental processes in the basidiomycete *Coprinus cinereus*. *Microbiology and Molecular Biology Reviews: MMBR*, 64(2), 316–353.

101. Kumar, S., Chinnusamy, V., ir Mohapatra, T. (2018). Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Frontiers in Genetics*, 9:640.
102. Kurdyukov, S., ir Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology*, 5(1), 3.
103. Labbé, J., Murat, C., Morin, E., Tuskan, G. A., Le Tacon, F., ir Martin, F. (2012). Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*. *PLoS One*, 7(8), e40197.
104. Lau, A. Y. T., Xie, Y., Cheung, M. K., Cheung, P. C. K., ir Kwan, H. S. (2020). Genome-wide mRNA and miRNA analysis in the early stages of germ tube outgrowth in *Coprinopsis cinerea*. *Fungal Genetics and Biology*, 142, 103416.
105. Lee, H. J., Hore, T. A., ir Reik, W. (2014). Reprogramming the Methylome: Erasing Memory and Creating Diversity. *Cell Stem Cell*, 14(6), 710-719.
106. Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A., Surani, M. A., ir Hajkova, P. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural and Molecular Biology*, 20(3), 311–316.
107. Lentini, A., Lagerwall, C., Vikingsson, S., Mjoseng, H. K., Douvlataniotis, K., Vogt, H., Green, H., Meehan, R. R., Benson, M., ir Nestor, C. E. (2018). A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nature Methods*, 15(7), 499–504.
108. Lewis, C. J. T., Pan, T., ir Kalsotra, A. (2017a). RNA modifications and structures cooperate to guide RNA-protein interactions. *Nature Reviews. Molecular Cell Biology*, 18(3), 202–210.
109. Lewis, L. C., Lo, P. C. K., Foster, J. M., Dai, N., Corrêa, I. R., Durczak, P. M., Duncan, G., Ramsawhook, A., Aithal, G. P., Denning, C., Hannan, N. R. F., ir Ruzov, A. (2017b). Dynamics of 5-carboxylcytosine during hepatic differentiation: Potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics*, 12(4), 277–286.
110. Li, D., Zhang, B., Xing, X., ir Wang, T. (2015). Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods*, 72, 29–40.
111. Li, Q.-Y., Yuan, B.-F., ir Feng, Y.-Q. (2018). Mass spectrometry-based nucleic acid modifications analysis. *Chemistry Letters*, 47(12), 1453-1459.

112. Ličytė, J., Gibas, P., Skardžiūtė, K., Stankevičius, V., Rukšėnaitė, A., ir Kriukienė, E. (2020). A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell Reports*, 32(11), 108155.
113. Ličytė, J., Kvederavičiūtė, K., Rukšėnaitė, A., Godliauskaitė, E., Gibas, P., Tomkutė, V., Petraitytė, G., Masevičius, V., Klimašauskas, S., ir Kriukienė, E. (2022). Distribution and regulatory roles of oxidized 5-methylcytosines in DNA and RNA of the basidiomycete fungi and. *Open Biology*, 12(3), 210302.
114. Lima, S. A., Chipman, L. B., Nicholson, A. L., Chen, Y.-H., Yee, B. A., Yeo, G. W., Coller, J., ir Pasquinelli, A. E. (2017). Short poly(A) tails are a conserved feature of highly expressed genes. *Nature Structural and Molecular Biology*, 24(12), 1057–1063.
115. Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., ir Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315–322.
116. Liu, J., ir Jia, G. (2014). Methylation Modifications in Eukaryotic Messenger RNA. *Journal of Genetics and Genomics*, 41(1), 21–33.
117. Liu, Q., Liu, L., Zhao, Y., Zhang, J., Wang, D., Chen, J., He, Y., Wu, J., Zhang, Z., ir Liu, Z. (2011). Hypoxia Induces Genomic DNA Demethylation through the Activation of HIF-1 $\alpha$  and Transcriptional Upregulation of MAT2A in Hepatoma Cells. *Molecular Cancer Therapeutics*, 10(6), 1113–1123.
118. Liu, Y., Cheng, J., Siejka-Zielińska, P., Weldon, C., Roberts, H., Lopopolo, M., Magri, A., D'Arienzo, V., Harris, J. M., McKeating, J. A., ir Song, C.-X. (2020). Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome Biology*, 21(1), 54.
119. Liu, Y., Hu, Z., Cheng, J., Siejka-Zielińska, P., Chen, J., Inoue, M., Ahmed, A. A., ir Song, C.-X. (2021). Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. *Nature Communications*, 12(1), 618.
120. Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., ir Song, C.-X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology*, 37(4), 424–429.

121. Liutkevičiūtė, Z., Kriukienė, E., Ličytė, J., Rudytė, M., Urbanavičiūtė, G., ir Klimašauskas, S. (2014). Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *Journal of the American Chemical Society*, 136(16), 5884–5887.
122. Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D., ir Klimašauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chemical Biology*, 5(6), 400–402.
123. Loewus, F. (1999). Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi. *Phytochemistry*, 52(2), 193–210.
124. Logsdon, G. A., Vollger, M. R., ir Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews. Genetics*, 21(10), 597–614.
125. Lu, X., Song, C.-X., Szulwach, K., Wang, Z., Weidenbacher, P., Jin, P., ir He, C. (2013). Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *Journal of the American Chemical Society*, 135(25), 9315–9317.
126. Lu, X., Zhao, B. S., ir He, C. (2015a). TET family proteins: oxidation activity, interacting molecules, and functions in diseases. *Chemical Reviews*, 115(6), 2225–2239.
127. Lu, X., Han, D., Zhao, B. S., Song, C.-X., Zhang, L.-S., Doré, L. C., ir He, C. (2015b). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Research*, 25(3), 386–389.
128. Lukinavičius, G., Lapienė, V., Staševskij, Z., Dalhoff, C., Weinhold, E., ir Klimašauskas, S. (2007). Targeted Labeling of DNA by Methyltransferase-Directed Transfer of Activated Groups (mTAG). *Journal of the American Chemical Society*, 129(10), 2758–2759.
129. Lukinavičius, G., Lapinaitė, A., Urbanavičiūtė, G., Gerasimaitė, R., ir Klimašauskas, S. (2012). Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. *Nucleic Acids Research*, 40(22), 11594–11602.
130. Lukinavičius, G., Tomkuvienė, M., Masevičius, V., ir Klimašauskas, S. (2013). Enhanced chemical stability of adomet analogues for improved methyltransferase-directed labeling of DNA. *ACS Chemical Biology*, 8(6), 1134–1139.
131. Maiti, A., ir Drohat, A. C. (2011). Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *Journal of Biological Chemistry*, 286(41), 35334–35338.

132. Mariani, C., Madzo, J., Moen, E., Yesilkanal, A., ir Godley, L. (2013). Alterations of 5-Hydroxymethylcytosine in Human Cancers. *Cancers*, 5(4), 786–814.
133. Marina, R. J., Sturgill, D., Bailly, M. A., Thenoz, M., Varma, G., Prigge, M. F., Nanan, K. K., Shukla, S., Haque, N., ir Oberdoerffer, S. (2016). TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *The EMBO Journal*, 35(3), 335–355.
134. Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A., ir Stunnenberg, H. G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, 149(3), 590–604.
135. Martello, G., ir Smith, A. (2014). The nature of embryonic stem cells. *Annual Review of Cell and Developmental Biology*, 30, 647–675.
136. Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E. G. J., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., Salamov, A., Shapiro, H. J., Wuyts, J., Blaudez, D., Buée, M., Brokstein, P., Canbäck, B., Cohen, D., Courty, P. E., ... Grigoriev, I. V. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, 452(7183), 88–92.
137. Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C., ir Hibbett, D. S. (2016). Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews. Microbiology*, 14(12), 760–773.
138. Martin, F., ir Selosse, M.-A. (2008). The *Laccaria* genome: a symbiont blueprint decoded. *The New Phytologist*, 180(2), 296–310.
139. Martinez-Val, A., Lynch, C. J., Calvo, I., Ximénez-Embún, P., Garcia, F., Zarzuela, E., Serrano, M., ir Munoz, J. (2021). Dissection of two routes to naïve pluripotency using different kinase inhibitors. *Nature Communications*, 12(1), 1863.
140. Masevičius, V., Nainytė, M., ir Klimašauskas, S. (2016). Synthesis of S-Adenosyl-L-Methionine Analogs with Extended Transferable Groups for Methyltransferase-Directed Labeling of DNA and RNA. *Current Protocols in Nucleic Acid Chemistry*, 64, 1.36.1–1.36.13.
141. Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., ir Salbert, G. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, 452(7183), 45–50.

142. Miao, Z., Xin, N., Wei, B., Hua, X., Zhang, G., Leng, C., Zhao, C., Wu, D., Li, J., Ge, W., Sun, M., ir Xu, X. (2016). 5-hydroxymethylcytosine is detected in RNA from mouse brain tissues. *Brain Research*, 1642, 546–552.
143. Minor, E. A., Court, B. L., Young, J. I., ir Wang, G. (2013). Ascorbate Induces Ten-Eleven Translocation (Tet) Methylcytosine Dioxygenase-mediated Generation of 5-Hydroxymethylcytosine. *Journal of Biological Chemistry*, 288(19), 13669–13674.
144. Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E., Andreopoulos, B., Barry, K. W., Bonito, G., Buée, M., Carver, A., Chen, C., Cichocki, N., Clum, A., Culley, D., Crous, P. W., Fauchery, L., Girlanda, M., ... Martin, F. M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nature Communications*, 11(1), 5125.
145. Motorin, Y., Lyko, F., ir Helm, M. (2010). 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Research*, 38(5), 1415–1430.
146. Mulholland, C. B., Nishiyama, A., Ryan, J., Nakamura, R., Yiğit, M., Glück, I. M., Trummer, C., Qin, W., Bartoschek, M. D., Traube, F. R., Parsa, E., Ugur, E., Modic, M., Acharya, A., Stolz, P., Ziegenhain, C., Wierer, M., Enard, W., Carell, T., ... Leonhardt, H. (2020). Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nature Communications*, 11(1), 1–24.
147. Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., ir Ginalska, K. (2019). Transposable elements contribute to fungal genes and impact fungal lifestyle. *Scientific Reports*, 9(1), 1–10.
148. Nachtergaele, S., ir He, C. (2017). The emerging biology of RNA post-transcriptional modifications. *RNA Biology*, 14(2), 156–163.
149. Nanan, K. K., Sturgill, D. M., Prigge, M. F., Thenoz, M., Dillman, A. A., Mandler, M. D., ir Oberdoerffer, S. (2019). TET-Catalyzed 5-Carboxylcytosine Promotes CTCF Binding to Suboptimal Sequences Genome-wide. *iScience*, 19, 326–339.
150. Narmontė, M., Gibas, P., Daniūnaitė, K., Gordevičius, J., ir Kriukienė, E. (2021). Multiomics Analysis of Neuroblastoma Cells Reveals a Diversity of Malignant Transformations. *Frontiers in Cell and Developmental Biology*, 9, 727353.
151. Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F., ir Oliviero, S. (2015). Single-Base

- Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Reports*, 10(5), 674–683.
152. Nolivos, S., Carpousis, J.C., ir Clouet-d'Orval, B. (2005). The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Research*, 33(20), 6507–6514.
  153. Ong, C.-T., ir Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nature Reviews. Genetics*, 15(4), 234–246.
  154. Parry, A., Rulands, S., ir Reik, W. (2021). Active turnover of DNA methylation during cell fate decisions. *Nature Reviews. Genetics*, 22(1), 59–66.
  155. Pastor, W. A., Aravind, L., ir Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Reviews. Molecular Cell Biology*, 14(6), 341–356.
  156. Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., Brudno, Y., Mahapatra, S., Kapranov, P., Tahiliani, M., Daley, G. Q., Liu, X. S., Ecker, J. R., Milos, P. M., Agarwal, S., ir Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473(7347), 394–397.
  157. Peng, J., Xia, B., ir Yi, C. (2016). Single-base resolution analysis of DNA epigenome via high-throughput sequencing. *Science China. Life Sciences*, 59(3), 219–226.
  158. Pennings, S., Revuelta, A., McLaughlin, K. A., Hadi, N. A. A., Petchreing, P., Ottaviano, R., ir Meehan, R. R. (2019). Dynamics and Mechanisms of DNA Methylation Reprogramming. *Epigenetics and Regeneration* (p. 19–45). Elsevier.
  159. Peter, M., Kohler, A., Duplessis, S. ir Martin, F. (2006). RNA extraction from ectomycorrhizas. *Handbook of Methods Used in Rhizosphere Research* (p. 458–459). Swiss Federal Research Institute WSL, Birmensdorf.
  160. Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeyer, C., ir Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie*, 50(31), 7008–7012.
  161. Phillips, J. E., ir Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, 137(7), 1194–1211.
  162. Plaza, D. F., Lin, C.-W., van der Velden, N. S. J., Aebi, M., ir Künzler, M. (2014). Comparative transcriptomics of the model mushroom



- Coprinopsis cinerea reveals tissue-specific armories and a conserved circuitry for sexual development. *BMC Genomics*, 15(1), 1–17.
163. Plongthongkum, N., Diep, D. H., ir Zhang, K. (2014). Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nature Reviews. Genetics*, 15(10), 647–661.
164. Prasad, R., Yen, T. J., ir Bellacosa, A. (2021). Active DNA demethylation—The epigenetic gatekeeper of development, immunity, and cancer. *Advanced Genetics*, 2, e10033.
165. Pukkila, P. J. (2011). *Coprinopsis cinerea*. *Current Biology: CB*, 21(16), R616–R617.
166. Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H. E., Branco, M. R., Murat, P., Oxley, D., Booth, M. J., Reik, W., ir Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biology*, 13(8), R69.
167. Raiber, E.-A., Hardisty, R., van Delft, P., ir Balasubramanian, S. (2017). Erratum: Mapping and elucidating the function of modified bases in DNA. *Nature Reviews Chemistry*, 1–13.
168. Raiber, E.-A., Murat, P., Chirgadze, D. Y., Beraldi, D., Luisi, B. F., ir Balasubramanian, S. (2015). 5-Formylcytosine alters the structure of the DNA double helix. *Nature Structural and Molecular Biology*, 22(1), 44–49.
169. Raiber, E.-A., Portella, G., Martínez Cuesta, S., Hardisty, R., Murat, P., Li, Z., Iurlaro, M., Dean, W., Spindel, J., Beraldi, D., Liu, Z., Dawson, M. A., Reik, W., ir Balasubramanian, S. (2018). 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nature Chemistry*, 10(12), 1258–1266.
170. Ramsawhook, A., Lewis, L., Coyle, B., ir Ruzov, A. (2017). Medulloblastoma and ependymoma cells display increased levels of 5-carboxylcytosine and elevated expression. *Clinical Epigenetics*, 9, 18.
171. Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., ir Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, 14(4), 411–413.
172. Rao, P. S., ir Niederpruem, D. J. (1969). Carbohydrate metabolism during morphogenesis of *Coprinus lagopus* (sensu Buller). *Journal of Bacteriology*, 100(3), 1222–1228.
173. Robin, P. D., Amadou, B., Daniel, M., Antoine, G., Ezékiel, B., Bernard, D., ir Yves, P. (2011). Biotechnological Processes Used in

- Controlled Ectomycorrhizal Practices. *Diversity and Biotechnology of Ectomycorrhizae*, (p. 143–155). Springer Berlin Heidelberg.
174. Roundtree, I. A., Evans, M. E., Pan, T., ir He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, *169*(7), 1187–1200.
175. Sabino, J. C., de Almeida, M. R., Abreu, P. L., Ferreira, A. M., Caldas, P., Domingues, M. M., Santos, N. C., Azzalin, C. M., Grosso, A. R., ir de Almeida, S. F. (2022). Epigenetic reprogramming by TET enzymes impacts co-transcriptional R-loops. *eLife*, *11*, e69476.
176. Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeyer, C., Truss, M., ir Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angewandte Chemie*, *51*(26), 6516–6520.
177. Schiesser, S., Pfaffeneder, T., Sadeghian, K., Hackner, B., Steigenberger, B., Schröder, A. S., Steinbacher, J., Kashiwazaki, G., Höfner, G., Wanner, K. T., Ochsenfeld, C., ir Carell, T. (2013). Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *Journal of the American Chemical Society*, *135*(39), 14593–14599.
178. Schön, A., Kaminska, E., Schelter, F., Ponkkonen, E., Korytiaková, E., Schiffers, S., ir Carell, T. (2020). Analysis of an Active Deformylation Mechanism of 5-Formyl-deoxycytidine (fdC) in Stem Cells. *Angewandte Chemie*, *59*(14), 5591–5594.
179. Schwartz, S., ir Motorin, Y. (2017). Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA Biology*, *14*(9), 1124–1137.
180. Seidl, M. F., ir Thomma, B. P. H. J. (2017). Transposable Elements Direct the Coevolution between Plants and Microbes. *Trends in Genetics: TIG*, *33*(11), 842–851.
181. Seiler, C. L., Fernandez, J., Koerperich, Z., Andersen, M. P., Kotandeniya, D., Nguyen, M. E., Sham, Y. Y., ir Tretyakova, N. Y. (2018). Maintenance DNA Methyltransferase Activity in the Presence of Oxidized Forms of 5-Methylcytosine: Structural Basis for Ten Eleven Translocation-Mediated DNA Demethylation. *Biochemistry*, *57*(42), 6061–6069.
182. Shen, L., Song, C.-X., He, C., ir Zhang, Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annual Review of Biochemistry*, *83*, 585–614.

183. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H.-L., Zhang, K., ir Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, *153*(3), 692–706.
184. Shen, L., ir Zhang, Y. (2013). 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Current Opinion in Cell Biology*, *25*(3), 289–296.
185. Shen, Q., Zhang, Q., Shi, Y., Shi, Q., Jiang, Y., Gu, Y., Li, Z., Li, X., Zhao, K., Wang, C., Li, N., ir Cao, X. (2018). Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature*, *554*(7690), 123–127.
186. Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., ir Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, *14*(4), 407–410.
187. Singh, A. M., ir Dalton, S. (2009). The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming. *Cell Stem Cell*, *5*(2), 141–149.
188. Skvortsova, K., Iovino, N., ir Bogdanović, O. (2018). Functions and mechanisms of epigenetic inheritance in animals. *Nature Reviews. Molecular Cell Biology*, *19*(12), 774–790.
189. Slyvka, A., Mierzejewska, K., ir Bochtler, M. (2017). Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision. *Scientific Reports*, *7*(1), 9001.
190. Song, C.-X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., Gao, J., Liu, P., Li, L., Xu, G.-L., Jin, P., ir He, C. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, *153*(3), 678–691.
191. Song, C.-X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P., ir He, C. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, *29*(1), 68–72.
192. Song, C.-X., Yi, C., ir He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. *Nature Biotechnology*, *30*(11), 1107–1116.
193. Song, C.-X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., Liu, B., Xiong, J., Zhang, W., Hu, J., Zhou, Z., Dong, B., Tian, Z., Jeffrey, S. S., Chua, M.-S., So, S., Li, W., Wei, Y., Diao, J., ... Quake, S. R. (2017).

- 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Research*, 27(10), 1231–1242.
194. Soyer, J. L., El Ghalid, M., Glaser, N., Ollivier, B., Linglin, J., Grandaubert, J., Balesdent, M.-H., Connolly, L. R., Freitag, M., Rouxel, T., ir Fudal, I. (2014). Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genetics*, 10(3), e1004227.
195. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., ... Vermeulen, M. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, 152(5), 1146–1159.
196. Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., Suter, C. M., ir Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Research*, 40(11), 5023–5033.
197. Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., ir Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378), 490–495.
198. Stajich, J. E., Wilke, S. K., Ahrén, D., Au, C. H., Birren, B. W., Borodovsky, M., Burns, C., Canbäck, B., Casselton, L. A., Cheng, C. K., Deng, J., Dietrich, F. S., Fargo, D. C., Farman, M. L., Gathman, A. C., Goldberg, J., Guigó, R., Hoegger, P. J., Hooker, J. B., ... Pukkila, P. J. (2010). Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proceedings of the National Academy of Sciences of the United States of America*, 107(26), 11889–11894.
199. Staševskij, Z., Gibas, P., Gordevičius, J., Kriukienė, E., ir Klimašauskas, S. (2017). Tethered Oligonucleotide-Primed Sequencing, TOP-Seq: A High-Resolution Economical Approach for DNA Epigenome Profiling. *Molecular Cell*, 65(3), 554–564.e6.
200. Storebjerg, T. M., Strand, S. H., Høyer, S., Lynnerup, A.-S., Borre, M., Ørntoft, T. F., ir Sørensen, K. D. (2018). Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC),

- 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer. *Clinical Epigenetics*, 10(1):105.
201. Sun, Z., Terragni, J., Borgaro, J. G., Liu, Y., Yu, L., Guan, S., Wang, H., Sun, D., Cheng, X., Zhu, Z., Pradhan, S., ir Zheng, Y. (2013a). High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Reports*, 3(2), 567–576.
202. Sun, M., Song, C.-X., Huang, H., Frankenberger, C. A., Sankarasharma, D., Gomes, S., Chen, P., Chen, J., Chada, K. K., He, C., ir Rosner, M. R. (2013b). HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(24), 9920–9925.
203. Sun, Z., Dai, N., Borgaro, J. G., Quimby, A., Sun, D., Corrêa, I. R., Zheng, Y., Zhu, Z., ir Guan, S. (2015). A Sensitive Approach to Map Genome-wide 5-Hydroxymethylcytosine and 5-Formylcytosine at Single-Base Resolution. *Molecular cell*, 57(4), 750–761.
204. Suzuki, M. M., ir Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews. Genetics*, 9(6), 465–476.
205. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., ir Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929), 930–935.
206. Tajima, S., Suetake, I., Takeshita, K., Nakagawa, A., ir Kimura, H. (2016). Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases. *Advances in Experimental Medicine and Biology*, 945, 63–86.
207. Tang, Y., Xiong, J., Jiang, H.-P., Zheng, S.-J., Feng, Y.-Q., ir Yuan, B.-F. (2014). Determination of Oxidation Products of 5-Methylcytosine in Plants by Chemical Derivatization Coupled with Liquid Chromatography/Tandem Mass Spectrometry Analysis. *Analytical Chemistry*, 86(15), 7764–7772.
208. Tang, Y., Zheng, S.-J., Qi, C.-B., Feng, Y.-Q., ir Yuan, B.-F. (2015). Sensitive and simultaneous determination of 5-methylcytosine and its oxidation products in genomic DNA by chemical derivatization coupled with liquid chromatography-tandem mass spectrometry analysis. *Analytical Chemistry*, 87(6), 3445–3452.

209. Tardu, M., Jones, J. D., Kennedy, R. T., Lin, Q., ir Koutmou, K. S. (2019). Identification and Quantification of Modified Nucleosides in mRNAs. *ACS Chemical Biology*, *14*(7), 1403–1409.
210. Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L., ir McKay, R. D. G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, *448*(7150), 196–199.
211. Tomkuvienė, M., Kriukienė, E., ir Klimašauskas, S. (2016). DNA Labeling Using DNA Methyltransferases. *Advances in Experimental Medicine and Biology*, *945*, 511–535.
212. Tomkuvienė, M., Mickutė, M., Vilkaitis, G., ir Klimašauskas, S. (2019). Repurposing enzymatic transferase reactions for targeted labeling and analysis of DNA and RNA. *Current Opinion in Biotechnology*, *55*, 114–123.
213. Traube, F. R., Schiffers, S., Iwan, K., Kellner, S., Spada, F., Müller, M., ir Carell, T. (2019). Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nature Protocols*, *14*(1), 283–312.
214. Tsiouplis, N. J., Bailey, D. W., Chiou, L. F., Wissink, F. J., ir Tsagaratou, A. (2021). TET-Mediated Epigenetic Regulation in Immune Cell Development and Disease. *Frontiers in Cell and Developmental Biology*, *8*:623948.
215. van der Wijst, M. G. P., Venkiteswaran, M., Chen, H., Xu, G.-L., Plösch, T., ir Rots, M. G. (2015). Local chromatin microenvironment determines DNMT activity: from DNA methyltransferase to DNA demethylase or DNA dehydroxymethylase. *Epigenetics*, *10*(8), 671–676.
216. Van Haute, L., Powell, C. A., ir Minczuk, M. (2017). Dealing with an Unconventional Genetic Code in Mitochondria: The Biogenesis and Pathogenic Defects of the 5-Formylcytosine Modification in Mitochondrial tRNAMet. *Biomolecules*, *7*(1), 24.
217. Vilkaitis, G., Dong, A., Weinhold, E., Cheng, X., ir Klimašauskas, S. (2000). Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. *The Journal of Biological Chemistry*, *275*(49), 38722–38730.
218. von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., Kroeze, L. I., Nesterova, T. B., Jansen, J. H., Xie, H., He, C., Reik, W., ir Stunnenberg, H. G. (2016). Impairment of DNA Methylation

- Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell*, 62(6), 848-861.
219. Wagner, M., Steinbacher, J., Kraus, T. F. J., Michalakis, S., Hackner, B., Pfaffeneder, T., Perera, A., Müller, M., Giese, A., Kretzschmar, H. A., ir Carell, T. (2015). Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angewandte Chemie*, 54(42), 12511–12514.
220. Wang, D., Hashimoto, H., Zhang, X., Barwick, B. G., Lonial, S., Boise, L. H., Vertino, P. M., ir Cheng, X. (2017). MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Research*, 45(5), 2396–2407.
221. Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.-D., ir Wang, D. (2015). Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, 523(7562), 621–625.
222. Wang, Y., Chen, Z., Zhang, X., Weng, X., Deng, J., Yang, W., Wu, F., Han, S., Xia, C., Zhou, Y., Chen, Y., ir Zhou, X. (2021). Single-Base Resolution Mapping Reveals Distinct 5-Formylcytidine in mRNAs. *ACS Chemical Biology*, 17(1):77-84.
223. Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., Yu, M., Liu, X., Zhu, P., Li, X., Hou, Y., Guo, H., Wu, X., He, C., Li, R., ... Qiao, J. (2014). Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biology*, 15(3), 1–17.
224. Wescoe, Z. L., Schreiber, J., ir Akeson, M. (2014). Nanopores discriminate among five C5-cytosine variants in DNA. *Journal of the American Chemical Society*, 136(47), 16582–16587.
225. Wheldon, L. M., Abakir, A., Ferjentsik, Z., Dudnakova, T., Strohbuecker, S., Christie, D., Dai, N., Guan, S., Foster, J. M., Corrêa, I. R., Jr, Loose, M., Dixon, J. E., Sottile, V., Johnson, A. D., ir Ruzov, A. (2014). Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Reports*, 7(5), 1353–1361.
226. Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., ir Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes and Development*, 25(7), 679–684.

227. Wu, H., Wu, X., Shen, L., ir Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nature Biotechnology*, 32(12), 1231–1240.
228. Wu, H., Wu, X., ir Zhang, Y. (2016). Base-resolution profiling of active DNA demethylation using MAB-seq and caMAB-seq. *Nature Protocols*, 11(6), 1081–1100.
229. Wu, H., ir Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*, 156(1-2), 45–68.
230. Wu, H., ir Zhang, Y. (2015). Charting oxidized methylcytosines at base resolution. *Nature Structural ir Molecular Biology*, 22(9), 656–661.
231. Wu, S. C., ir Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nature Reviews. Molecular Cell Biology*, 11(9), 607–620.
232. Wu, X., ir Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews. Genetics*, 18(9), 517–534.
233. Xia, B., Han, D., Lu, X., Sun, Z., Zhou, A., Yin, Q., Zeng, H., Liu, M., Jiang, X., Xie, W., He, C., ir Yi, C. (2015). Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nature Methods*, 12(11), 1047–1050.
234. Xiong, J., Liu, X., Cheng, Q.-Y., Xiao, S., Xia, L.-X., Yuan, B.-F., ir Feng, Y.-Q. (2017). Heavy Metals Induce Decline of Derivatives of 5-Methylcytosine in Both DNA and RNA of Stem Cells. *ACS Chemical Biology*, 12(6), 1636–1643.
235. Xu, G.-L., ir Wong, J. (2015). Oxidative DNA demethylation mediated by Tet enzymes. *National Science Review*, 2(3), 318–328.
236. Xu, S., Li, W., Zhu, J., Wang, R., Li, Z., Xu, G.-L., ir Ding, J. (2013). Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Research*, 23(11), 1296–1309.
237. Xue, C., Zhao, Y., ir Li, L. (2020). Advances in RNA cytosine-5 methylation: detection, regulatory mechanisms, biological functions and links to cancer. *Biomarker Research*, 8, 43.
238. Yang, J., Bashkenova, N., Zang, R., Huang, X., ir Wang, J. (2020). The roles of TET family proteins in development and stem cells. *Development*, 147(2):dev183129.
239. Yang, J., Horton, J. R., Li, J., Huang, Y., Zhang, X., Blumenthal, R. M., ir Cheng, X. (2019). Structural basis for preferential binding of human



- TCF4 to DNA containing 5-carboxylcytosine. *Nucleic Acids Research*, 47(16), 8375–8387.
240. Yin, R., Mao, S.-Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., Jiao, Y., Li, C., Liu, S., Wu, D., Gu, W., Yang, Y.-G., Xu, G.-L., ir Wang, H. (2013). Ascorbic Acid Enhances Tet-Mediated 5-Methylcytosine Oxidation and Promotes DNA Demethylation in Mammals. *Journal of the American Chemical Society*, 135(28), 10396–10403.
241. Yu, M., Hon, G. C., Szulwach, K. E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J.-H., Jin, P., Ren, B., ir He, C. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, 149(6), 1368–1380.
242. Yu, X., Xie, Y., Zhang, S., Song, X., Xiao, B., ir Yan, Z. (2021). tRNA-derived fragments: Mechanisms underlying their regulation of gene expression and potential applications as therapeutic targets in cancers and virus infections. *Theranostics*, 11(1), 461–469.
243. Yuan, B.-F., ir Feng, Y.-Q. (2014). Recent advances in the analysis of 5-methylcytosine and its oxidation products. *TrAC Trends in Analytical Chemistry*, 54, 24–35.
244. Zampieri, M., Ciccarone, F., ir Caiafa, P. (2018). Defective DNA Methylation/Demethylation Processes Define Aging-Dependent Methylation Patterns. *Epigenetics of Aging and Longevity* (p. 33–58). Elsevier.
245. Zemach, A., McDaniel, I. E., Silva, P., ir Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328(5980), 916–919.
246. Zhang, H.-Y., Xiong, J., Qi, B.-L., Feng, Y.-Q., ir Yuan, B.-F. (2016). The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. *Chemical Communications*, 52(4), 737–740.
247. Zhang, L., Chen, W., Iyer, L. M., Hu, J., Wang, G., Fu, Y., Yu, M., Dai, Q., Aravind, L., ir He, C. (2014). A TET Homologue Protein from *Coprinopsis cinerea* (CcTET) That Biochemically Converts 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine. *Journal of the American Chemical Society*, 136(13), 4801–4804.
248. Zhao, B. S., Roundtree, I. A., ir He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nature Reviews. Molecular Cell Biology*, 18(1), 31–42.

249. Zhao, L.-Y., Song, J., Liu, Y., Song, C.-X., ir Yi, C. (2020). Mapping the epigenetic modifications of DNA and RNA. *Protein and Cell*, *11*(11), 792–808.
250. Zhou, D., Alver, B. M., Li, S., Hlady, R. A., Thompson, J. J., Schroeder, M. A., Lee, J.-H., Qiu, J., Schwartz, P. H., Sarkaria, J. N., ir Robertson, K. D. (2018). Distinctive epigenomes characterize glioma stem cells and their response to differentiation cues. *Genome Biology*, *19*(1), 43.
251. Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., ir Yi, C. (2017). Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell*, *20*(5), 720–731.e5.
252. Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., ir Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*(7463), 477–481.

## SUMMARY OF DOCTORAL DISSERTATION

### List of abbreviations

2i	two small-molecule kinase inhibitors
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Ado-6-N <sub>3</sub>	synthetic analog of the SAM cofactor, with an extended 6-azidohex-2-ynyl group
βGT	T4 phage β-glucosyltransferase
BS-Seq	bisulfite sequencing
caCLEAR	“5caC <u>C</u> learance”, a method for genomic mapping of 5caC
eM.SssI	engineered methyltransferase SssI (Q142A/N370A)
foTOP-seq	5fC-specific tethered oligonucleotide-primed sequencing
H3K	modifications in lysine of histone 3, for example: H3K4me1 – monomethylation of 4th lysine, H3K36me3 – trimethylation of 36th lysine, H3K27ac – acetylation of 27th lysine.
hmTOP-seq	5hmC-specific tethered oligonucleotide-primed sequencing
HPLC-MS/MS	high-performance liquid chromatography–tandem mass spectrometry
M. or MTase	methyltransferase
mESCs	mouse embryonic stem cells
oxi-mCs	oxidized forms of 5-methylcytosine
R. or REase	restriction endonuclease
rC	cytosine base in RNA
SAM	S-adenosyl-L-methionine
TDG	thymine DNA glycosylase
TEs	transposable elements
TET	ten-eleven translocation oxygenase
TOP-seq	tethered oligonucleotide-primed sequencing, a method for mapping of uCGs
uCG	unmodified CG dinucleotide
wt	wild type

## INTRODUCTION

5-methylcytosine (5mC) is a widespread DNA modification found from bacteria to mammals and is produced by DNA methyltransferases (MTases), which transfer a methyl group from the cofactor S-adenosyl-L-methionine (SAM) to the C5 position of cytosine [1]. In multicellular eukaryotes 5mC is a crucial epigenetic mark, which does not change the coding sequence, but is very important for suppression of repetitive elements, regulation of gene expression, imprinting, mammalian X-chromosome inactivation [2].

Changes in genomic distribution of 5mC are linked to the cell-specific transcriptional status, important in development and disease [3]. DNA can be demethylated passively by replication or actively, when enzymes remove 5mC independently of replication [4].

Mammalian active DNA demethylation can occur via successive oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by TET proteins [5–7]. 5fC and 5caC are removed by thymine DNA glycosylase (TDG) and the base excision repair restores the unmodified C [6,8].

Increasing evidence suggests that these oxidized forms of 5-methylcytosine (oxi-mCs) are not only demethylation intermediates but also have unique epigenetic functions in mammals. 5hmC is the most abundant oxi-mC, important for transcriptional regulation, development, formation of neural tissue and memory [9]. 5fC and 5caC are more rare and their independent epigenetic functions are underexplored, although supporting data is emerging – oxi-mCs have partially distinct reader proteins and the genomic sites of oxi-mCs do not always overlap [10,11,12].

Alternative mammalian active demethylation pathways are still poorly understood. One of the proposed mechanisms is a direct decarboxylation of 5caC to cytosine [7,13], which has been observed in cell extracts by an yet unknown enzyme [14]. It was found that human DNMT3A, 3B and bacterial M.SssI DNA cytosine-5 MTases decarboxylate 5caC [15]. This reaction provides a potential enzymatic pathway for a direct reversal of 5caC to the unmodified C, although it is unclear whether MTases could perform this reaction in the presence of abundant SAM *in vivo*.

5mC levels change dramatically during mammalian development. After the zygote formation, 5mC levels drop and oxi-mCs are produced, which abundance later decreases. As the early epiblast implants into the uterus, levels of 5mC and 5hmC increase again [16]. In culture medium with two kinase

inhibitors (2i) mouse embryonic stem cells (mESCs) have a naive state of pluripotency, low levels of 5mC and 5hmC and resemble the inner cell mass of the blastocyst before implantation [17,18], whereas cells grown in serum conditions are similar to the post-implantation epiblast cells and are primed for differentiation. Cells cultivated under these two conditions provide a model to study demethylation and pluripotency states in which knowledge about changes of rare oxi-mCs is limited.

Functions of oxi-mCs in phylogenetic groups other than mammals are still underexplored. The basidiomycete fungi *Laccaria bicolor* and *Coprinopsis cinerea* have many genes encoding TET homologs (74 and 47, respectively; [19,20]). These laboratory model organisms are convenient for research, since they have small genomes of known sequence (~37 Mb for *C. cinerea* and ~61 Mb for *L. bicolor* [21,22]). Oxi-mCs were detected in *C. cinerea* DNA, 5hmC sites and the combined genomic profile of oxi-mCs was determined [23]. However, the distribution of individual oxi-mCs and their impact on gene expression have not been studied in detail in *C. cinerea* and have not yet been explored in *L. bicolor*. Therefore, *C. cinerea* and *L. bicolor* are interesting models to study the possible functions of oxi-mCs and the significance of TET proteins.

Sensitive, nucleotide-resolution methods for genome-wide mapping of oxi-mCs are needed to investigate their functions in detail. Most of the high-resolution methods for oxi-mCs use bisulfite sequencing (BS-Seq), which has many drawbacks [24]. In these methods the whole genome is sequenced, although most of the sequences do not contain cytosine modifications, usually found in the CG dinucleotide. Therefore, sensitive mapping of modifications demands high sequencing coverage, cost and analytical efforts, which is impractical for rare oxi-mCs [25].

Alternative genomic profiling methods which bridge the existing economy-versus-resolution gap have been developed. The tethered oligonucleotide-primed sequencing method (TOP-seq; [26]) analyzes unmodified genomic CGs (uCGs). An engineered MTase eM.SssI with an expanded cofactor binding pocket is used, which covalently tags uCGs with an azide group using a synthetic SAM cofactor analog. An oligonucleotide is covalently attached to the azide group and the polymerase performs strand extension from a complementary primer. Newly formed DNA strand contains the attached sequence, target CG and adjacent genomic region. Sequencing directly determines the location of the uCG site at nucleotide-resolution without the whole genome sequencing.

As eM.SssI decarboxylates 5caC yielding uCG, this reaction can be combined with eM.SssI-mediated covalent labeling of uCGs, and then previously carboxylated CG sites could be identified by TOP-seq. Because 5caC is the rarest oxi-mC, method determining only 5caC sites would be much more cost-efficient compared with whole-genome sequencing methods.

Based on TOP-seq, a nucleotide-resolution profiling method of genomic 5hmC, hmTOP-seq (5hmC-specific TOP-seq [27]), was developed in which 5hmC is labeled with an azide-glucose by  $\beta$ -glucosyltransferase. After chemical reduction of 5fC to 5hmC, this method could be applied for genomic mapping of 5fC.

The **aim** of this work was to study the 5caC decarboxylation reaction performed by methyltransferases, use it for the development of a genomic profiling method of 5caC and to apply TOP-seq-based genome-wide mapping methods of cytosine modifications for the analysis of oxi-mC distribution and functions.

### **The main tasks:**

1. To confirm that the human methyltransferases DNMT3A/3L and DNMT3B/3L perform the 5caC decarboxylation reaction and to evaluate their enzymatic activity at physiological SAM concentrations.
2. To develop and evaluate a 5caC genomic profiling method using the eM.SssI-mediated 5caC decarboxylation reaction.
3. To investigate the genomic distribution of 5caC in mouse embryonic stem cells using the developed 5caC profiling method.
4. To determine the abundance and distribution of oxidized 5mC forms and to investigate their possible functions using TOP-seq-based genomic profiling methods in the basidiomycete fungi *L. bicolor* and *C. cinerea*.

### **Scientific novelty of the study**

A novel sensitive CG-resolution 5caC genomic profiling method, caCLEAR, has been developed. As caCLEAR specifically identifies only 5caCG sites, it provides a cost-effective 5caC profiling, requiring less sequencing and analytical effort compared to other nucleotide-resolution 5caC mapping methods which use whole-genome sequencing.

For the first time, the genomic distribution of 5caC in two pluripotency states of mESCs (cultured in serum-2i and serum conditions) was compared. Considerable differences in the 5caC enrichment in genomic elements and distribution in binding regions of important factors was

determined between the two states. 5caC enrichment and a fraction of 5caC-modified binding sites of some factors (CTCF, c-Myc and Nanog) exclusively in serum conditions were associated with genes important for developmental processes which are more active in the primed state, while in 2i conditions associations were determined with genes of general cell functions. Asymmetric distribution of 5caC in the antisense strand of active protein-coding genes in mESCs was also demonstrated.

A method to facilitate the detection of a rare RNA modification, 5-carboxylcytosine (5carC), by mass spectrometry was developed using chemical derivatization. For the first time, oxi-mCs were determined in RNA of basidiomycete fungi *L. bicolor* and *C. cinerea*, and the abundance of 5carC and 5hmrC was analyzed in different RNA fractions.

A new nucleotide-resolution 5fC genomic profiling method, foTOP-seq, has been developed for identification of 5fC sites without whole-genome sequencing. Oxi-mCs were detected and quantified in *L. bicolor* DNA for the first time. For the first time, the nucleotide-resolution genomic distribution of 5hmC and 5fC in *L. bicolor* was determined, and such a detailed assessment of 5mC, 5hmC and 5fC influence on gene expression was performed in *L. bicolor* and *C. cinerea*.

Overall, new methods have been developed to study oxi-mCs in this work, which provided more information about the prevalence, distribution and functions of oxi-mCs.

## MATERIALS AND METHODS

Enzymes, unless indicated otherwise, were obtained from Thermo Fisher Scientific (TS).

### Thin layer chromatography (TLC)

5'-5caCGCATCATACTCTA (or without 5caC) oligonucleotide was 5'-<sup>33</sup>P-labeled using T4 polynucleotide kinase, ethanol-precipitated, annealed with 5'-TAGAGTATGATG5mCGCTGACCCACAACATCCG, 3'-GACTGGGTGTTGTAGGC and ligated using T4 DNA ligase, yielding 31 bp DNA duplexes. ~90 nM DNA was incubated with 160-400 nM DNMT3A/3L or 3B/3L complex (BPS Bioscience) in 10 mM Tris-HCl pH 7.4 (~22°C), 50 mM NaCl, 0.1 mg/ml BSA, 2 mM DTT for 4-6h at 37°C. DNA was ethanol-precipitated and digested with P1 nuclease (Sigma-Aldrich). After elution on PEI Cellulose F plates (MERCK) with solvent A (isobutyric acid/H<sub>2</sub>O/conc. ammonia, 66:17:4 vol.) or B (propanol-2/37% HCl/H<sub>2</sub>O, 4.7:1:1 vol.), samples analyzed by autoradiography.

### Preparation of model DNA

Oxi-mCs in one strand of the CG site of the 188 bp DNA fragment were introduced with a modified primer by PCR as in [15].

5hmC was introduced to GCGC (using M.HhaI [28] at room temperature (RT) for 1h) or CCGG (using M.HpaII [29] at 37°C for 3.5h) sites in 10 mM Tris-HCl pH 7.4 (~22°C), 50 mM NaCl, 0.5 mM EDTA and 13 μM formaldehyde. Enzymes were inactivated for 20 min at 65°C and treated with 0.2 mg/ml Proteinase K (ProtK) and 0.1% of SDS for 1h at 50°C, 20 min at 65°C. DNA was purified with DNA clean and concentrator (DCC) kit (Zymo research, ZR) and eluted with 10 mM Tris-HCl pH 8.5 (~25°C) in this and further column-purifications.

Lambda bacteriophage (λ) DNA was methylated at CGs or GCGCs by wild-type (wt) SssI and M.HhaI. 5caC was introduced using the 5mC TET1 Oxidation kit (WiseGene). Methylated, unmodified, 5caC and 5hmC containing λ DNA was mixed to produce the final desired modification level.

### Cultivation

E14TG2a wt, *Tdg*-depleted and Tet1/2/3 triple knockout (negative control) mESC lines were kindly provided by Prof. Guo-Liang Xu and cultured by Dr. Vaidotas Stankevičius [30].

*Coprinopsis cinerea* haploid monokaryotic strain Okayama 7 (FGSC, #9003) was grown in liquid YMG medium [31] without shaking at 30°C for



14 d. For oxidation enhancement, the mycelium was grown for 7 days and then treated with 5 mM vitC for 2 or 5d.

*Laccaria bicolor* (Maire) Orton S238N-H82 haploid monokaryotic strain (gift from Prof. Francis M. Martin) was collected after 8 weeks from a liquid modified Pachlewski medium (10 g/L glucose, 3 g/L malt extract, 0.25 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.5 g/L KH<sub>2</sub>PO<sub>4</sub>, 0.15 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.05 g/L CaCl<sub>2</sub>, 0.025 g/L NaCl, 0.0012% FeCl<sub>3</sub>, 100 mg/L thiamine, 2.3 mg/L ZnSO<sub>4</sub>·7H<sub>2</sub>O, 8.5 mg/L H<sub>3</sub>BO<sub>3</sub>, 0.5 mg MnCl<sub>2</sub>·4H<sub>2</sub>O, 0.6 mg/L CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.4 mg/L (NH<sub>4</sub>)<sub>2</sub>MoO<sub>4</sub>). VitC treatment was done for 2d.

### **High-performance liquid chromatography – tandem mass spectrometry (HPLC-MS/MS)**

Nuclease P1 and FastAP phosphatase-digested nucleosides were analyzed on a HPLC/ESI-MS/MS system (Agilent 1290 Infinity/6410B triple quadrupole) with a Supelco Discovery® HS C18 column at 30°C, using a gradient of solvents A (0.0075% HCOOH in H<sub>2</sub>O) and B (0.0075% HCOOH in acetonitrile) at 0.3 ml/min. Nucleoside-specific ion m/z transitions were recorded in the positive MRM mode and nucleoside-standards used for calibration. HPLC-MS/MS was partly performed and all conditions (see [30,32]) optimized by Audronė Rukšėnaitė. 5caC and 5fC (and 5hmC for fungi) were first HPLC-fractionated from ~9-24 μg DNA/RNA in 4 runs with a gradient of 10 mM NH<sub>4</sub>HCO<sub>2</sub> (pH 5.5) and 80% methanol. Fractions were vacuum-dried and dissolved in H<sub>2</sub>O.

### **caCLEAR procedure**

**Step 1.** 5hmC sites in genomic DNA (sonicated to ~200-250 bp) were glycosylated using T4 beta-glucosyltransferase (βGT) at 37°C for 2h according to the vendor's protocol. After inactivation at 65°C for 20 min, ProtK-treatment and purification, a second round of glycosylation was done for 16h and DNA was purified.

**Step 2.** uCGs were blocked by methylation using 10-fold excess of wt M.SssI (1/10 volume) over CGs with 0.6 mM SAM in the vendor's buffer at 37°C for 4h. After inactivation for 20 min at 65°C, ProtK-treatment and purification, a second methylation for 16h was performed accordingly and DNA purified with Oligo Clean and Concentrator (ZR) kit.

**Step 3.** Decarboxylation was performed using 10-fold excess of eM.SssI [33] (1.13 μM) over CGs in 10 mM Tris-HCl pH 8 (~22°C), 50 mM NaCl, 0.5 mM EDTA at 22°C for 3h, 20 min at 65°C. Then, DNA was ProtK-treated and purified.

**Step 4.** Decarboxylation reaction was supplemented with 0.2 mM Ado-6-azide [34,35] and incubated at 30°C for 1h, 65°C for 20 min, then ProtK-treated and purified.

**Step 5.** TOP-seq.

**Step 5a.** Adaptors were ligated as in [26].

**Step 5b.** DNA was incubated at 45°C for 20 min with 20  $\mu$ M 5'-TXXXXTGTGTGGTTTGGAGACTGACTACCAGATGTAACA-biotin (X=C8-alkyne-dU, Base-click) oligonucleotide,  $\frac{1}{5}$  volume of CuBr (8 mM): THPTA (24 mM) mixture (Sigma-Aldrich) in DMSO (final 50%). After dilution to ~1% DMSO, DNA was purified with GeneJET NGS Cleanup Kit (TS). ~50 ng of  $\lambda$  or 827 ng of mESC DNA was used for enrichment with 0.1 mg of Dynabeads MyOne Streptavidin C1 (TS) in 10 mM Tris-HCl pH 8.5 (~22°C), 1 M NaCl and 0.2 % of Tween 20 at RT for 3h on a roller. Beads were washed as in [36]. DNA was eluted in H<sub>2</sub>O at 95°C for 5 min, 5 min at 4°C. Fig. 4A experiment was performed without enrichment.

**Step 5c.** DNA, 0.5  $\mu$ M priming strand 5'-TGTTACATCTGGTAGTCAGTCTCCAAACCACACAA, 0.2 mM dNTP and 0.05 u/ $\mu$ l Pfu polymerase in its MgSO<sub>4</sub> buffer with additional 1 mM MgSO<sub>4</sub>, in 20  $\mu$ l was incubated for 5 cycles at: 95°C 1 min, 65°C 10 min, 72°C 10 min.

**Step 5d.** Primed DNA was amplified with Platinum™ SuperFi™ PCR Master Mix (TS) in 100  $\mu$ l and size-selected for 300 bp fragments as in [26]. All DNA libraries were subjected to Ion Proton (TS) sequencing, performed by A. Rukšėnaitė. All caCLEAR-related sequencing data was analyzed by Dr. Povilas Gibas (see [30]).

### **Derivatization of 5carC-RNA**

RNA (heat-denatured in H<sub>2</sub>O) was incubated with 10 mM MES pH 6 (~22°C), 20 mM N-hydroxysuccinimide, 8 mM 1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride and 45% DMSO in 60  $\mu$ l for 30 min at 37°C, shaking 550 rpm. Then 10 mM (4-aminomethyl)benzylazide hydrochloride (synthesis by Prof. Viktoras Masevičius and Dr. Gražina Petraitytė, see [32]), 100 mM sodium phosphate pH 7.5 and 150 mM NaCl were added to 180  $\mu$ l and incubated for 1h at 37°C, shaking 550 rpm. After 2.5x dilution with H<sub>2</sub>O, RNA was purified with RNA clean and concentrator kit (ZR).

Control sR47 5carC-RNA [37] was prepared using 5caCTP (TriLink BioTechnologies) and TranscriptAid T7 High Yield Transcription kit (TS).

After 5carC derivatization efficiency evaluation, this RNA was used for MS/MS calibration.

### **Fungal sequencing libraries**

Sonicated ~200-250 bp DNA was purified using DCC (ZR). Modified hMe/fC-Seal [38,39] protocols were used. For detection of 5fC, 5hmC was blocked with unmodified glucose as for caCLEAR and 5fC was reduced according to [40]. ~2.65 µg DNA was labeled using 50 µM UDP-6-azide-glucose (full name UDP-6-azido-6-deoxy-D-glucose, Jena Bioscience) and 26.5 U of βGT, incubating for 2h at 37°C, 65°C for 20 min and purified using DCC kit (ZR). For the 5hmC control sample, βGT was omitted in the labeling step; for the 5fC control, glycosylated DNA without the reduction was used.

DNA was then incubated with 0.4 mM DBCO-S-S-PEG3-biotin (Jena Bioscience) in 10 mM Tris-HCl pH 8.5 (~22°C) at 37°C for 2h and purified with DCC kit (ZR). DNA was enriched using Dynabeads MyOne Streptavidin C1 (TS) in 10 mM Tris-HCl pH 8.5 (~22°C), 1 M NaCl and 0.01 % of Tween 20 at RT for 3h on a roller. The beads were washed 3× with 15 mM Tris-HCl (pH 7.5), 1.5 mM EDTA, 3 M NaCl, 0.01 % of Tween 20 and 2× with 5 mM Tris-HCl pH 7.5 (~22°C), 0.5 mM EDTA, 1 M NaCl, 0.01 % of Tween 20. DNA was separated from the beads after incubation with 50 mM DTT, 60 mM Tris pH 7.8 (~22°C) for 1h on a roller.

DNA was then end-filled using a DNA End Repair Kit (TS) and purified with the DCC kit (ZR). DNA was ligated to Ion Torrent™ adapters (Kapa Biosystems), which were then removed using MagJet NGS Cleanup and Size-selection kit (TS). DNA was incubated with 10 mM 2-mercaptoethanol for 10 min at RT, followed by 50 µl PCR containing 0.2 mM dNTP, 2 mM MgCl<sub>2</sub>, 1 µM primers, 2.5 U of Taq DNA Polymerase and its (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> buffer (TS). PCR conditions: 1 min 50°C, 5 min 72°C, 4 min 94°C, 1 min 65°C, 5 min 72°C, 15 cycles of 1 min 94°C, 1 min 65°C, 1 min 72°C and the final extension at 72°C for 5 min. The libraries were size-selected for 300 bp fragments (MagJet NGS Cleanup and Size-selection kit, TS).

For fo/hmTOP-seq, 5.8 µg DNA was labeled as in hMe/fC-Seal and TOP-seq was then done as in caCLEAR.

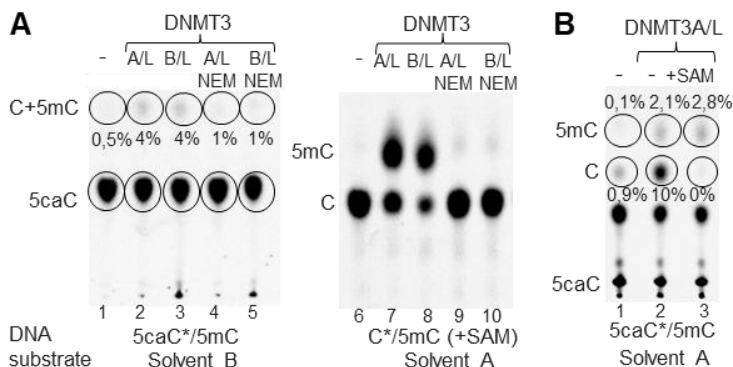
*L. bicolor* whole-transcriptome RNA libraries were prepared and sequenced at Thermo Fisher Scientific Baltics, using Collibri Stranded RNA Library Prep Kit (TS) for Illumina.

Dr. P. Gibas and Vita Tomkutė performed initial analysis of hMe-Seal and fC-Seal, Kotryna Kvederavičiūtė analyzed all further fungal sequencing data (see [32]).

## RESULTS

### 5caC decarboxylation by DNA methyltransferases

Besides the well-established mammalian TET-TDG active 5mC demethylation pathway, an alternative direct decarboxylation of 5caC to C mechanism was suggested [7,13] and observed in cell extracts by an unknown enzyme [14]. It was found that 5caC is decarboxylated in the presence of human MTase DNMT3A/L or 3B/L complexes, also bacterial MTases M.HhaI and M.SssI [15].

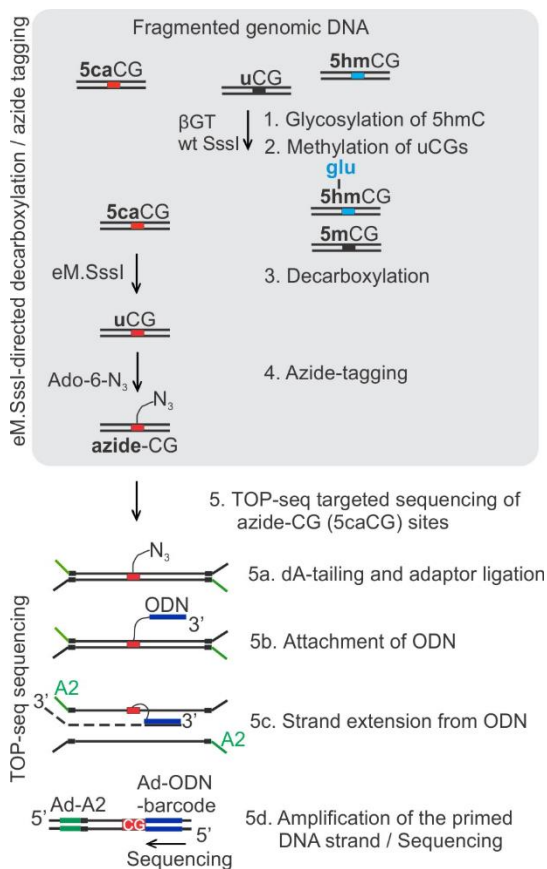


**Fig. 1.** 5caC decarboxylation by human MTases. 31-bp DNA substrate containing 5caCG or CG on one strand and 5mCG on the other was used. After digestion,  $^{33}\text{P}$ -labeled (\*) target nucleotide was analyzed by TLC and autoradiography. Densitometric quantitation of spot intensities is shown in circles (% of total counts in the lane). (A) Decarboxylation of 5caC and methylation by DNMT3A/L and B/L was abolished by pre-treatment with a cysteine-reactive substance *N*-ethylmaleimide (NEM, 20 mM, 1h 25°C). (B) DNMT3A/L 5caC decarboxylation activity was considerably inhibited by the SAM cofactor.

We sought to confirm that the 5caC decarboxylation in the human DNMT3 protein preparation is performed by MTases. We observed covalent changes of cytidine or 5-carboxylcytidine nucleotide by its  $^{33}\text{P}$ -labeling and TLC. Treatment of human MTases with a cysteine-reactive probe *N*-ethylmaleimide abolished 5caC decarboxylation (as well as methylation, Fig. 1A), showing that catalytic cysteine (essential for methylation by MTases [41]) is important for the reaction which is most likely performed by a MTase. At physiological SAM concentrations (30  $\mu\text{M}$ , [42]) we observed a significant suppression of 5caC decarboxylation by DNMT3A/L (Fig. 1B), suggesting that it is unlikely to occur in normal conditions *in vivo* but again confirming MTase activity. This atypical MTase activity indicates a plausible direct enzymatic 5caC conversion to C *in vivo* and also offers a potential approach to detect 5caC in genomic DNA sequence.

## Development and evaluation of caCLEAR method for the genomic mapping of 5caC

The biological functions of oxi-mCs are not fully understood, thus sensitive high-resolution genomic profiling methods are needed. We aimed to develop a single-nucleotide resolution method for genomic mapping of 5caC, avoiding BS-Seq (used in 5caC profiling methods [12,43,44]) and its drawbacks.



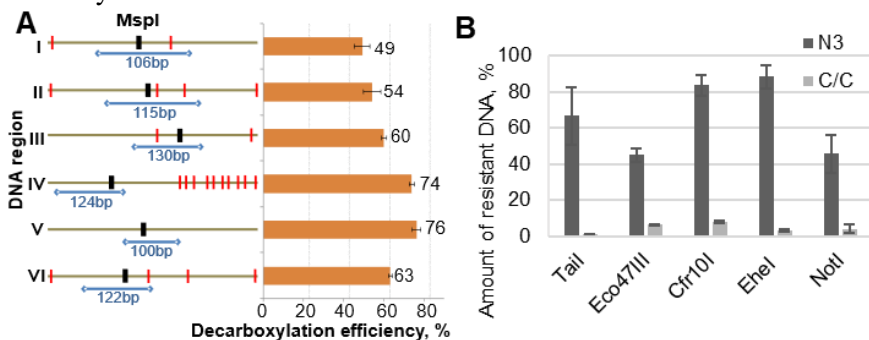
**Fig. 2.** Steps of the caCLEAR method for 5caC profiling. 1: modification of 5hmC with glucose (glu) using  $\beta$ GT and UDP-glucose. 2: protection of uCGs by methylation using the wt SssI MTase and SAM. 3: decarboxylation of 5caC using the engineered eM.SssI which introduces new uCGs. 4: covalent tagging of new uCGs with an azide group in the presence of eM.SssI and Ado-6- $N_3$ . 5: profiling of the azide-labeled CG sites by TOP-seq. ODN, oligodeoxyribonucleotide; A2, one strand of a partially complementary adaptor (Ad).

MTase eM.SssI is able to decarboxylate 5caC [15] and transfer an azide group from the synthetic cofactor Ado-6- $N_3$  to uCGs [33]. By combining these mild enzymatic reactions, former 5caCG targets can be determined by

the TOP-seq method [26]. As eM.SssI also dehydroxymethylates 5hmC [15], it can be protected by glycosylation using phage T4  $\beta$ -glucosyltransferase ( $\beta$ GT) and UDP-glucose [38], while native uCGs can be methylated using the wild-type (wt) M.SssI. The developed method, named caCLEAR (“5caC Clearance”, Fig. 2), would allow a cost-efficient nucleotide-resolution mapping of 5caCG sites only.

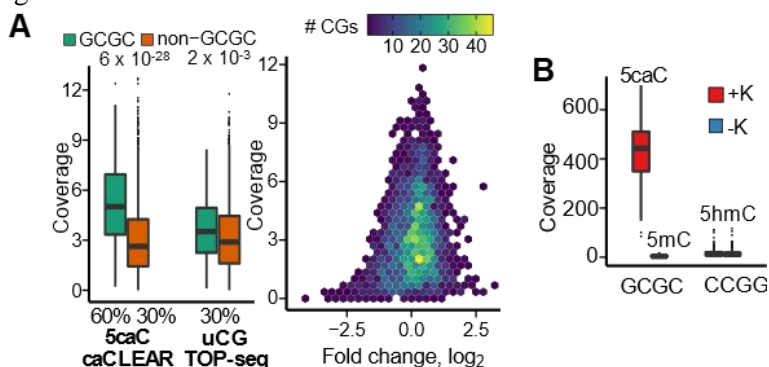
First, we evaluated the efficiency and specificity of the caCLEAR method in model DNA systems. We optimized protection of 5hmC and uCGs. HPLC-MS/MS analysis of hydroxymethylated (at GCGC, ~1.87% of C) lambda phage ( $\lambda$ ) genomic DNA showed an efficient ~99,9% glycosylation. BS-Seq of  $\lambda$  DNA demonstrated 97.3% methylation of all CGs, which can be interpreted as complete, considering the inappropriate bisulfite 5mC conversion to thymine (2.7% with some kits [45]).

Next, we explored the 5caC decarboxylation by eM.SssI in 6 genomic regions of the model human IMR90 DNA (carboxylated by TET1-oxidation) using R.MspI (it can cleave after the 5-carboxyl group is removed). Analysis revealed 49-76% decarboxylation efficiency (Fig. 3A). In the caCLEAR protocol azide-labeling can be conveniently performed after the 5caC decarboxylation in one tube, adding Ado-6-N<sub>3</sub> to the decarboxylation mixture. In a DNA fragment containing 37 uCGs (using 5 modification-sensitive restriction endonucleases (REases), recognizing 1-2 non-overlapping CGs) we observed an efficient (~66%, Fig. 3B) eM.SssI azide-labeling under decarboxylation conditions.



**Fig. 3.** Assessment of 5caC decarboxylation and CG labeling by eM.SssI using modification-sensitive REases and quantitative PCR (qPCR). **(A)** Left: 6 genomic regions were assessed (target CG sites are indicated in black, other CGs in red and PCR products in blue). Right: efficiency of decarboxylation is indicated as a difference in the DNA amount resistant to the R.MspI cleavage before and after decarboxylation. **(B)** eM.SssI azide-labeling (N3) using the cofactor Ado-6-N<sub>3</sub> in a model *c-fos* gene promoter DNA fragment containing 37 CGs was evaluated using REases, which had 1 (TaiI, Eco47III, Crf10I) or 2 (EheI, NotI) recognition sites. C/C – unmodified DNA.

Further, to evaluate the eM.SssI decarboxylation and azide-labeling we applied steps 2-5 of caCLEAR on  $\lambda$  genomic DNA (6,226 CGs) containing 30% and 60% 5caC at CG and GCGC sites, respectively. caCLEAR signal reflected carboxylation – sites with a higher 5caC level had a higher coverage (Fig. 4A). To test the completeness of decarboxylation at CGs (30% 5caC), we compared their caCLEAR coverage with the control TOP-seq sample unmethylated to the same extent (30% uCGs and 70% 5mCGs). No difference in coverage was observed between the libraries for most of the CGs (excluding GCGC), indicating that 5caC decarboxylation was complete and azide-labeling was similar as in uCGs.

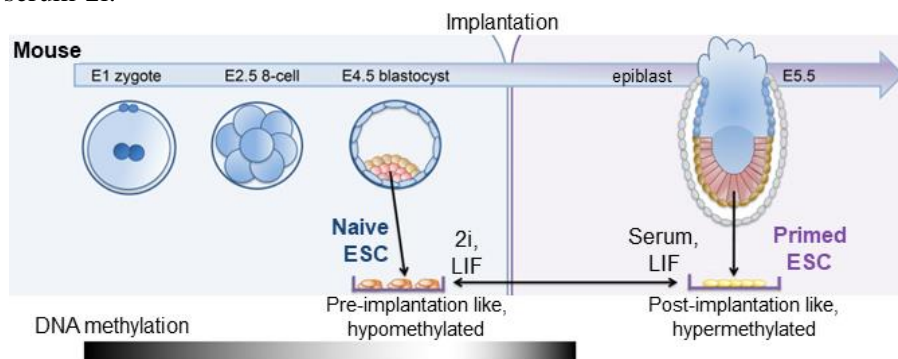


**Fig. 4.** Assessment of caCLEAR on a model genomic  $\lambda$  DNA. **(A)** Left, caCLEAR coverage (normalized by total read count) in GCGC or CG sites (excluding GCGC), containing ~60% and 30% of 5caC, respectively. Control TOP-seq samples contained 30% of uCGs and no 5caC. p values (above) were calculated with a two-sided paired t test. Right: fold change of normalized coverage at CG sites (excluding GCGC) between the caCLEAR and TOP-seq libraries at various coverage cutoffs. Color code denotes numbers of CGs. **(B)** Distribution of caCLEAR coverage in averaged technical replicates of samples that contained ~10% of 5hmc and 5caC (red, +K) at CCGG and GCGC sites, respectively. Control libraries had 5mC instead of 5caC (blue, -K).

Further, we analyzed caCLEAR libraries of  $\lambda$  DNA containing 10% of 5hmc and 5caC at CCGG (total 328) and GCGC (215) sites, respectively (and total 70% 5mCGs, reflecting mammalian DNA). Control sample without 5caC was also prepared. Importantly, most sequencing reads originated from the carboxylated GCGC sites, while other CGs and 5-hydroxymethylated CCGGs, had only background reads (Fig. 4B). This confirms caCLEAR specificity and also shows sensitivity for low levels of 5caC. Altogether, these results confirmed the validity of caCLEAR in detection of genomic 5caCG sites.

## caCLEAR analysis of two pluripotency states of mouse ESCs

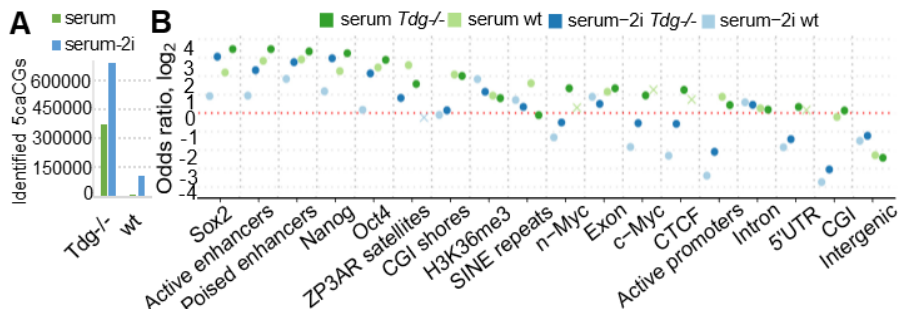
Next, we aimed to map genomic 5caC using caCLEAR in two pluripotency states of mESCs: naive (when grown with 2i), resembling the hypomethylated cells of the blastocyst inner cell mass and primed for development (when in serum conditions), similar to hypermethylated cells of the early post-implantation epiblast (Fig. 5). Many papers on DNA modifications [39,46–48] used a mixture of serum-2i to get more homogeneous mESCs, which corresponds to naive state [49]. Therefore, we cultivated wt and *Tdg*-depleted (*Tdg*<sup>-/-</sup>, positive control) mESCs in serum or serum-2i.



**Fig. 5.** Pluripotency states of cells in the developing mouse embryo: naive (before implantation) and primed for development (after implantation). Bottom – *in vitro* conditions required to maintain/convert each state to the other. LIF, leukemia inhibitory factor; E, embryonic day; 2i – two small inhibitors of kinases. The bottom shows the changes in DNA methylation. Adapted from [50,51].

More 5caCGs were identified by caCLEAR in *Tdg*<sup>-/-</sup> and serum-2i mESCs (Fig. 6A). Majority of 5caCGs were enriched in poised and active enhancers and binding regions of pluripotency-related transcription factors (TFs) (Fig. 6B). Compared to serum-2i, in serum mESCs 5caCGs were more enriched at CG islands and shores, 5'UTRs, active promoters (marked by H3K9ac) and enhancers, the insulator protein CTCF-binding sites. In line with the more demethylated state of serum-2i cells, 5caCGs were relatively more enriched in intergenic regions, introns and actively transcribed regions marked by H3K36me3, as compared to serum mESCs. Despite the higher 5caCG abundance in serum-2i mESCs, their distribution is less specific (mostly lower enrichment values).

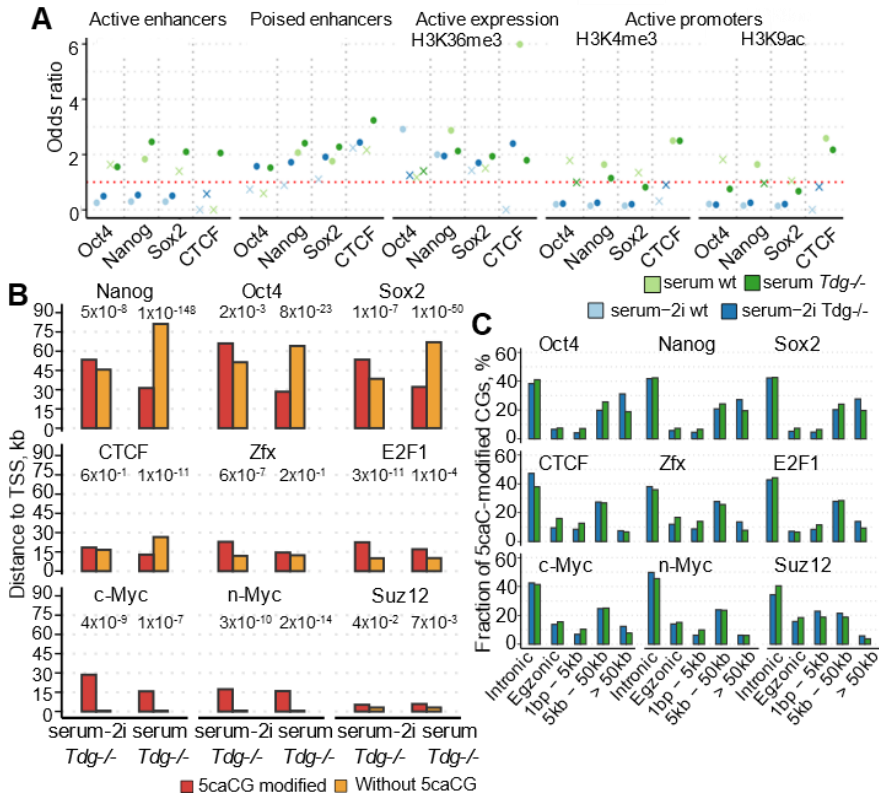




**Fig. 6. (A)** 5caCGs identified by caCLEAR in serum (green) or serum-2i mESCs (blue). **(B)** Enrichment of 5caCGs (odds ratio ( $\log_2$ ), Fisher’s exact test) in various genomic features. Poised enhancers, regions with H3K4me1 histone marks; active enhancers, marked by H3K4me1 and H3K27ac; active promoters, 2-kb regions upstream of genes marked with H3K9ac. “×” – non-significant estimates,  $p \geq 0.05$ .

### 5caC-modified factor binding sites differs between the two pluripotency states

We then investigated the distribution of 5caCGs at the binding sites (data from [52–54]) of TFs and regulators (Nanog, Oct4, Sox2, Zfx, c-Myc, n-Myc, E2F1, CTCF and Suz12) that play important roles in ESCs [52,55]. In both conditions, the 5caC-modified factor binding regions (5caC-FBRs) overlapped with active expression and poised enhancers (Fig. 7A). In contrast to serum-2i, in serum mESCs 5caCG-FBRs were enriched at active enhancers and promoters. We also determined that the majority of 5caCG-FBRs were at longer distances to the nearest transcription start site (TSS) compared to the 5caCG-depleted FBRs and that these distances were longer in serum-2i mESCs (Fig. 7B).



**Fig. 7.** 5caCG-modified factor binding regions (5caCG-FBRs) distribute differently in two pluripotency states of mESCs **(A)** Enrichment of 5caCG-FBRs (odds ratio, Fisher's exact test) in poised/active enhancers marked by H3K4me1 and H3K27ac/H3K4me1, respectively; active promoters (H3K4me3 and H3K9ac); and regions of active transcription (H3K36me3). **(B)** Median distance of 5caG-depleted or 5caCG-modified FBRs to transcription start sites (TSSs). Above the bar plots – p values calculated with the Mann-Whitney test. **(C)** Fractional distribution of 5caCG of FBRs in genes and intergenic regions at different distances from TSSs.

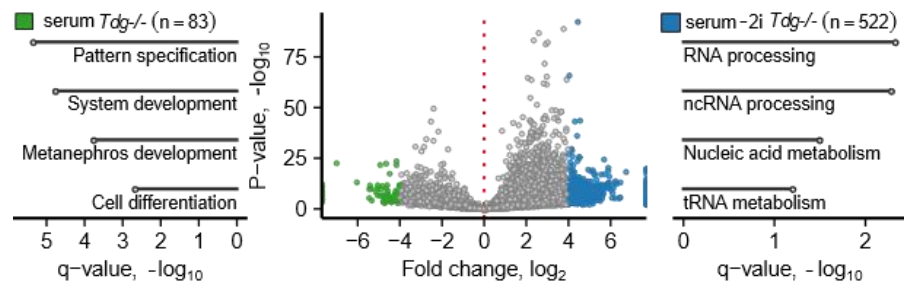
As a considerable fraction of 5caCGs from all CGs at FBRs was in genes (Fig. 7C), we evaluated their expression (Fig. 8). All investigated 5caCG-FBRs were enriched in the groups of moderate expression, except for the c-Myc 5caCG-FBRs, which were enriched in highly expressed genes in serum-2i mESCs. Gene Ontology (GO) functional analysis of genes with 5caCG-FBRs for c-Myc showed links to developmental processes in serum conditions (GO terms associated with embryo development,  $p=3 \times 10^{-5}$ ). Similarly, genes with CTCF 5caCG-FBRs were associated with developmental processes ( $p=5,5 \times 10^{-8}$ ) and regulation of signaling ( $p=6 \times 10^{-7}$ ) in the serum mESCs, while in the serum-2i mESCs we determined links to protein targeting to membrane categories ( $p=3,5 \times 10^{-5}$ ). This suggests that,

through sensing the modification state of its binding sites, CTCF and c-Myc might modulate expression of genes, promoting the switch between the naive and primed mESCs. Of pluripotency TFs, only Nanog 5caC-FBRs containing genes were linked with developmental processes (positive regulation of cellular process,  $p=3,5\times 10^{-7}$ ; developmental process,  $p=4,7\times 10^{-5}$ ).



**Fig. 8.** 5caCG-modified FBRs are associated with active expression. Enrichment of 5caCG-depleted and 5caCG-modified FBRs in different gene expression groups. “x” – non-significant estimates,  $p \geq 0.05$ .

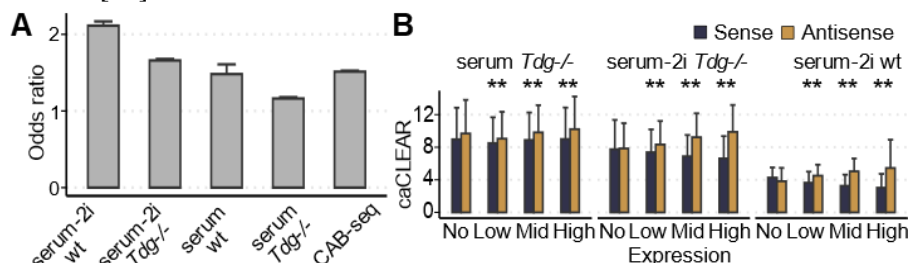
All 5caCG-enriched genes specific to serum-2i *Tdg*<sup>-/-</sup> mESCs (522) had no strong enrichment in GO analysis, except for a weak association with nucleic acid metabolic processes. In contrast, 5caCG-enriched genes specific to serum mESCs (83) revealed functional links with DNA-binding TF activity, multicellular organism development and cell-fate specification (Fig. 9). This is in line with the reported relationship of genes upregulated in 2i mESCs (which expression is lower or undetected in serum mESCs) with metabolic processes, while upregulated genes in serum mESCs are linked to developmental processes [56], pointing to an involvement of active demethylation in the establishment of the pluripotency states.



**Fig. 9.** Distribution of 5caCG-enriched genes, a volcano plot. Functional annotation analysis is shown for the significant genes specific to serum (n=83) and serum-2i *Tdg*<sup>-/-</sup> mESCs (n=522) (ANOVA and F test q values  $< 0.05$ ; absolute  $\log_2$  FC  $> 4$ ). ncRNA – non-coding RNA.

## 5caCGs distribute in the antisense strand of active genes

It is known that 5mC levels are symmetric in CG sites, but a major part of 5hmC and 5fC [12,40,47] and the aggregate 5fC/5caC signal [43] is asymmetric. We set out to investigate the strand-specific distribution of 5caCGs. First, we confirmed that there is no eM.SssI labeling-mediated strand bias in the  $\lambda$  DNA caCLEAR signal across CGs. mESC analysis strikingly showed an asymmetric distribution of 5caCG in the antisense (non-coding) strand of active protein-coding genes (Fig. 10, not visible in other genomic elements). Such asymmetry was also visible in the bisulfite-based CAB-seq method [43].

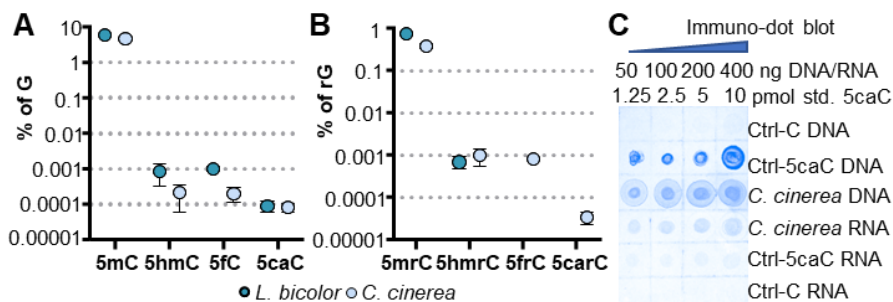


**Fig. 10.** 5caCGs distribute in the antisense strand of actively transcribed genes in mESCs. **(A)** Odds ratio for enrichment of 5caC in caCLEAR and bisulfite-based CAB-seq method [43] at the antisense strand of protein-coding genes. For all estimates,  $p < 2.2 \times 10^{-16}$ , Fisher's exact test. **(B)** The strand bias of 5caCGs in different gene expression groups. \*\* –  $p < 1 \times 10^{-7}$ , two-sided paired t test.

## Analysis of oxi-mCs in *Laccaria bicolor* and *Coprinopsis cinerea* basidiomycete fungi

Saprotrophic *C. cinerea* and important ectomycorrhizal symbiont *L. bicolor* have abundant homologues of *TET* genes, therefore are interesting organisms to study active DNA demethylation. We aimed to investigate the distribution and functions of oxi-mCs in these fungi.

In both fungi HPLC-MS/MS analysis of DNA revealed similar amounts between 5fC and 5hmC (both higher in *L. bicolor*) and levels of 5caC were the lowest (Fig. 11A). In contrast, immuno-dot blot demonstrated a high 5caC signal in *C. cinerea* DNA (Fig. 11C), which, we hypothesized, might appear from the residual RNA, potentially containing 5caC. Indeed, by immuno-dot blot we detected 5caC in *C. cinerea* RNA (5carC), which amounts were much lower in HPLC-MS/MS, indicating a possible antibody cross-reactivity. We also detected 5mrC, 5hmrC and 5frC (Fig. 11B) by HPLC-MS/MS.



**Fig. 11.** HPLC-MS/MS quantification of cytosine modifications in (A) DNA and (B) RNA of *L. bicolor* and *C. cinerea* grown in conventional conditions. At least 2 biological replicates were used, except for 5frC. Oxi-mCs were purified by HPLC prior to HPLC-MS/MS (except for 5hmrC in *L. bicolor*). Under standard *L. bicolor* cultivation 5scarC was not detected, 5frC was not assessed. (C) Detection of 5caC by immuno-dot blot in nucleic acids of *C. cinerea*. Ctrl – control unmodified (C) or modified (containing 5caC, shown in pmol) nucleic acids.

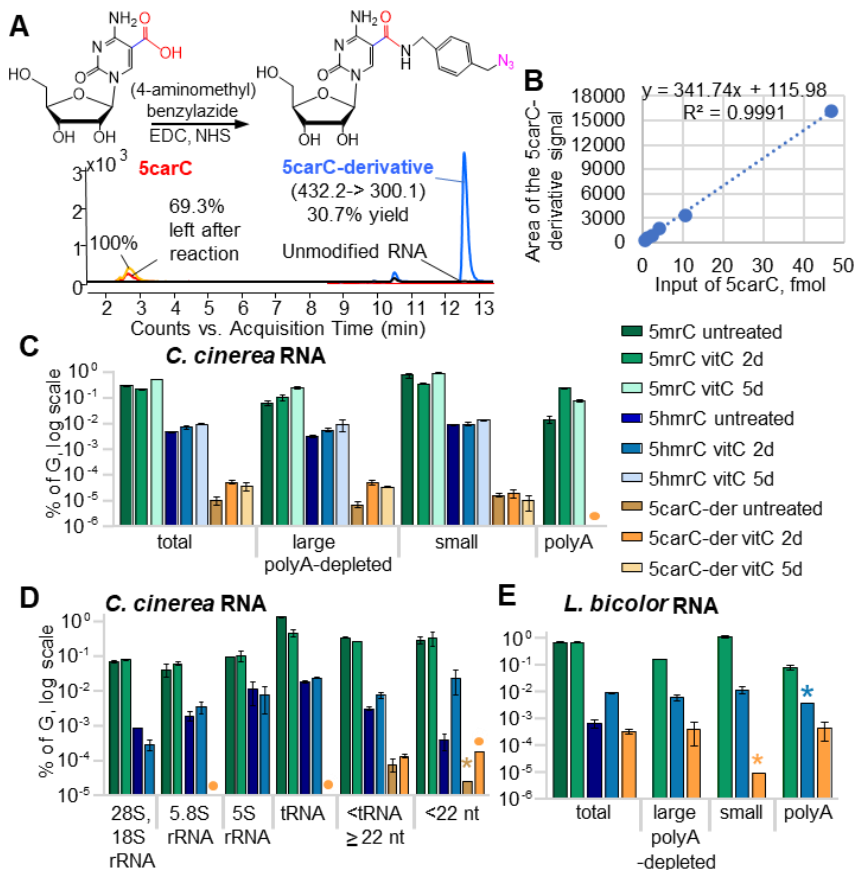
### Distribution of 5caC in RNA

We sought to explore distribution of the discovered 5scarC in *C. cinerea* RNAs by HPLC-MS/MS as this least investigated oxi-mC of RNA was only recently discovered in mammals [57].

To facilitate a difficult detection of 5scarC, which requires high amounts of RNA and additional HPLC nucleoside purification prior to final HPLC-MS/MS, we employed its chemical derivatization with (4-aminomethyl)benzylazide previously used for 5caC CAB-seq in DNA [43,58]. We achieved ~61% derivatization (optimized by Eglé Godliauskaitė) of 5scarC in synthetic RNA. The MS/MS signal of the 5scarC derivative had a linear dependence on 5scarC amounts and its intensity increased ~20× compared to underivatized 5scarC (Fig. 12A, B). Moreover, its new retention time away from other nucleosides eliminated the need for prior fractionation. Since as little as ~0.6 fmol of derivatized 5scarC could be detected, the developed sensitive method is well suited for 5scarC analysis in various biological samples.

We then investigated the distribution of 5scarC in various RNAs. Under standard cultivation 5scarC was not detected in *L. bicolor*, but vitamin C (vitC, known to enhance TET activity [59,60]) induced a strong signal in large (>200 nt) and polyA RNAs, while it was hardly detectable in the small <200 nt RNAs (Fig. 12E). In untreated *C. cinerea*, the highest 5scarC levels were in small RNAs; however, vitC increased 5scarC in the large polyA-depleted RNAs, while polyA RNAs showed a weak signal in one replicate (Fig. 12C). More detailed analysis of *C. cinerea* RNA (Fig. 12D) demonstrated no 5scarC signal in 28S and 18S large rRNAs, indicating its

association with mRNAs that lack or have short polyA tails or non-coding RNAs. In the pool of <200 nt RNAs of *C. cinerea*, 5.8S rRNA and tRNA showed a weak 5carC signal after the vitC treatment in only one biological replicate; 5S rRNA had no signal; while RNAs shorter than tRNA (<RNA<sub>≥22</sub> nt and <22 nt fractions) had a relatively strong signal, suggesting that they contribute the most to the 5carC in the <200 nt RNAs.



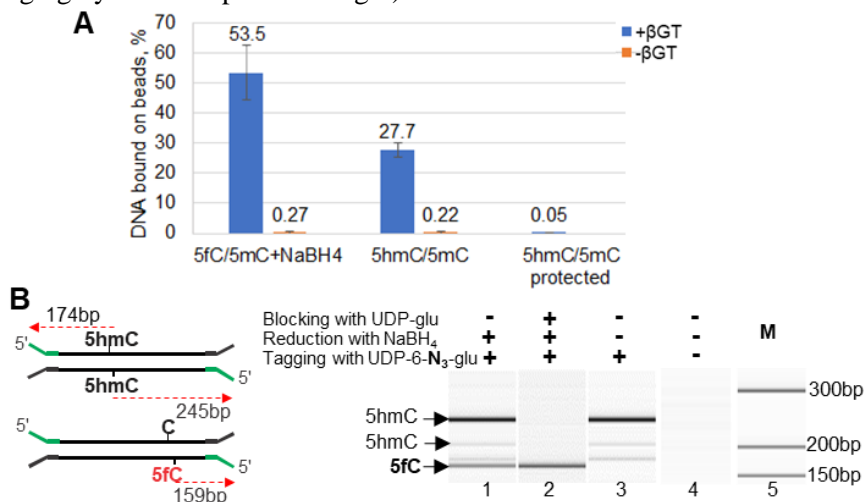
**Fig. 12.** HPLC-MS/MS detection of 5carC in RNA using chemical derivatization with (4-aminomethyl)benzylazide. **(A)** MS/MS signals of 5carC (yellow, before and red, after the reaction) and its derivatization product (blue). Unmodified RNA shows no signal (black). **(B)** MS/MS signal area of the 5carC derivative depending on the input amounts of 5carC before derivatization. **(C, D)** Quantification of 5carC (% of G) using chemical derivatization in different RNA fractions of *C. cinerea* and **(E)** *L. bicolor*. Fungi were treated with vitC for 2 (C, D, E) or 5 days (C). Untreated *L. bicolor* RNA did not show 5carC signal and was not fractionated. 2 biological replicates were used. Stars denote the samples with a detected and quantified signal (within a linear range of the calibration curve) in only one replicate; circles denote where the signal was detectable but could not be quantified (was outside a linear range of the calibration curve and signal-to-noise ratio was  $\geq 3$ ) in one replicate; samples without signal were not included. Small, <200 nt RNAs; large polyA-depleted, >200 nt RNAs, from which polyA tail containing RNAs were depleted; The small RNA fraction includes 5.8S

rRNA, 5S rRNA, tRNA and RNAs shorter than tRNA (~76-90 nt); large polyA-depleted RNA fraction includes 28S, 18S RNAs and other >200 nt RNAs.

### Profiling of 5hmC and 5fC genome-wide

We next investigated the genomic distribution of the *L. bicolor* and *C. cinerea* most abundant oxi-mCs, 5hmC and 5fC, using affinity enrichment methods hMe-Seal and fC-Seal, respectively [38,39]. 5hmC is labeled using  $\beta$ GT and UDP-6-azide-glucose. For 5fC enrichment, 5hmC is blocked with an unmodified glucose and 5fC is reduced to 5hmC using NaBH<sub>4</sub>. We used caCLEAR glycosylation conditions and also achieved an efficient 5fC reduction (~95%). Assessment hMe/fC-Seal protocols on model DNA fragments showed a fairly efficient (~28-54%) enrichment, specific for 5hmC or 5fC DNA (Fig. 13A).

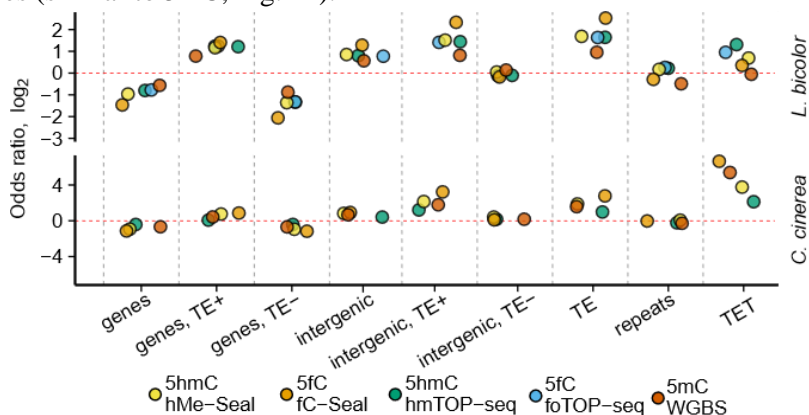
Low hMe/fC-Seal resolution (~280 bp) may generate overlapping oxi-mC distribution in relatively short genomic elements in fungi [22,21]. Therefore, we used our nucleotide resolution 5hmC profiling method hmTOP-seq [27], which uses 5hmC labeling by  $\beta$ GT. For 5fC mapping we additionally developed the foTOP-seq method, using 5hmC protection and 5fC reduction as in fC-Seal before the hmTOP-seq steps. Assessment of foTOP-seq on model DNA fragments showed 5fC-specific detection (Fig. 13B, lane 2, judging by the PCR product length).



**Fig. 13. (A)** qPCR evaluation of the 5hmC and 5fC affinity enrichment in hMe-Seal and fC-Seal (+NaBH<sub>4</sub>) methods, respectively. Model DNA fragments with one CG site modified by 5hmC/5mC or 5fC/5mC (mixed with *L. bicolor* DNA, 13:1287 ng) were labeled with (blue) or without (orange)  $\beta$ GT and UDP-6-azide-glucose. Control labeling reaction was performed on a 5hmC/5mC fragment that was previously glycosylated using unmodified UDP-glucose (protected). **(B)** Evaluation of the

foTOP-seq method using two model DNA fragments containing 5hmC or 5fC (left). Right, PCR products analyzed by Bioanalyzer (Agilent).

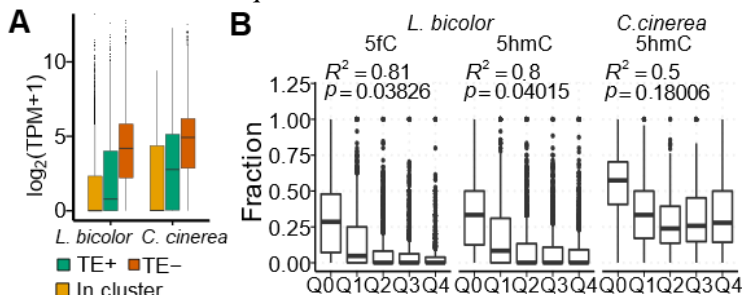
We prepared and sequenced the hMe-Seal, fC-Seal, hmTOP-seq of both fungi and foTOP-seq libraries of *L. bicolor*. As 5mC in both fungi occurs mainly in a CG context [61], we analyzed fo/hmTOP-seq in CGs. Results generally coincided between the methods: 5hmC and 5fC was enriched at genic and intergenic transposable elements (TEs), intergenic areas, also TET genes (similar to 5mC, Fig. 14).



**Fig. 14.** Enrichment (odds ratio ( $\log_2$ ), Fisher's exact test) of the 5hmC and 5fC in genomic features of *L. bicolor* and *C. cinerea* for fC/hMe-Seal (200-bp bins), fo/hmTOP-seq (single CGs) and the 5mC from BS-Seq data [61] (>50% methylated CGs were used). TE – transposable elements; repeats – tandem repeats (mini and microsatellites) excluding TE.

### 5hmC/5fC and gene expression

To evaluate the impact of oxi-mCs on gene expression we performed RNA sequencing of *L. bicolor* and used publicly available *C. cinerea* RNA-seq data [62]. We calculated the fraction of modified CGs from all CGs in the element in the fo/hmTOP-seq data.

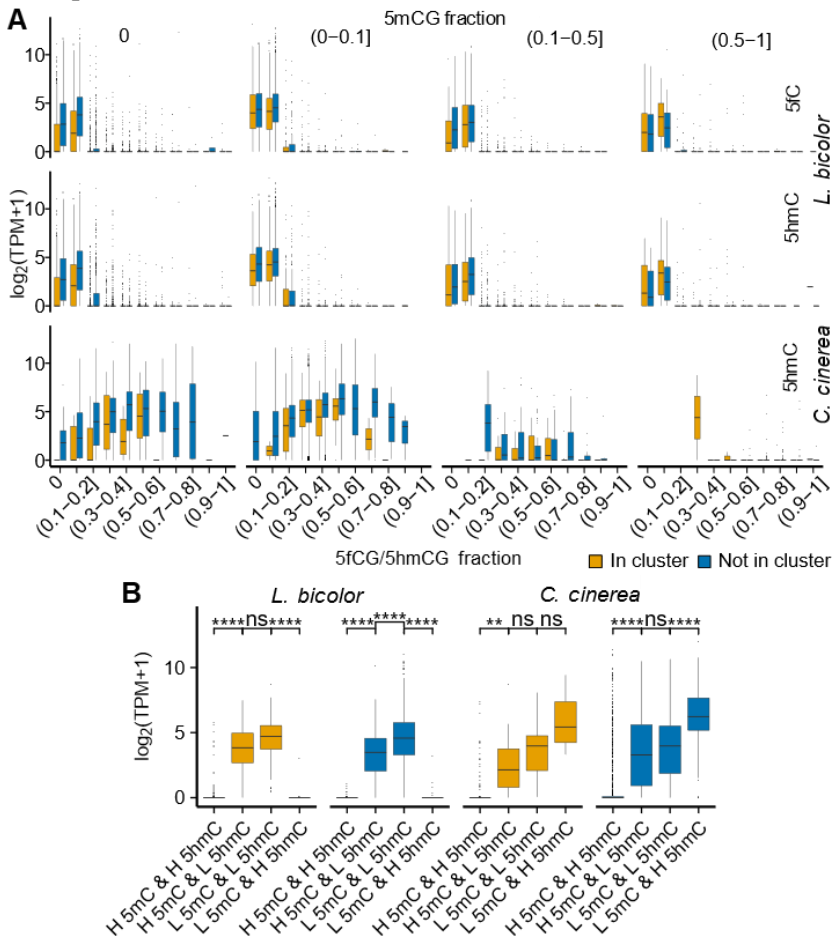


**Fig. 15.** (A) Expression of genes which are associated with TEs (TE+, distance <1 kb) or not (TE-), or present in TE clusters (In cluster). One-way ANOVA was used to test differences within groups ( $p$ -value  $\leq 0.05$ ). (B) 5hmCG and 5fCG fractions in TEs in



relation to expression quartiles of associated genes (distance to TE <1 kb) (Q1-4, Q0 – no expression). Linear model was used to fit the plotted data.

As TE-containing genes were enriched in 5hmCGs/5fCGs, we explored gene expression depending on the presence of TEs. In some fungi proximity to TE and their clusters is related to downregulated genes of higher methylation [63,64]. We observed that genes in TE clusters were least expressed compared to genes not in clusters or without TE (Fig. 15A). Moreover, highly 5hmCG/5fCG-modified TEs were associated (distance to TE <1 kb) with genes of low expression (Fig. 15B). These results show that TEs have an influence on higher gene 5fC/5hmC modification and negatively impact expression.



**Fig. 16.** 5hmC, 5fC and 5mC abundance and gene expression in *L. bicolor* and *C. cinerea* in TE clusters and outside them. **(A)** Expression ( $\log_2(\text{TPM}+1)$ ) stratified according to gene 5fCG and 5hmCG and 5mCG (BS-Seq data from [61]) fractions. **(B)** Expression of genes grouped according to high (H, last decile) and low (L, first decile) gene 5mCG and 5hmCG fractions. p-values were calculated using the t-test.

ns. (non-significant) –  $p > 0.05$ ; \*\* –  $p \leq 0.01$ ; \*\*\*\* –  $p \leq 0.0001$ . TPM – transcripts per million of all transcripts.

We further evaluated relationship between gene expression and 5mCG and oxi-mC fractions. In *C. cinerea*, expression was increasing with higher 5hmCG amounts in genes containing low 5mCG, while it was low in genes with high 5mCG and 5hmCG fractions (Fig. 16A). In contrast, in *L. bicolor* genes with high 5hmCG/5fCG fractions were repressed irrespective of their 5mCG fraction. The same trend was observed in TE clusters and outside them, except that expression was lower inside clusters.

Analysis of genes with high or low fractions of cytosine modifications again revealed that high gene 5hmCG or 5fCG (not shown) but not 5mCG amounts exerts a strong suppressive effect in *L. bicolor* (Fig. 16B). In contrast, high 5hmCG levels mark active genes when 5mCG is low in *C. cinerea*, while both abundant 5hmCGs and 5mCGs were in repressed genes. The observed differences between the fungi might suggest different mechanisms for regulation of more abundant TEs in *L. bicolor* than in *C. cinerea* (up to ~38% compared to ~6% [63,21]).

## CONCLUSIONS

1. Decarboxylation of 5caC in the presence of human methyltransferases DNMT3A/3L or DNMT3B/3L requires a cysteine residue in their catalytic center, but the reaction is inhibited by SAM. Therefore, although methyltransferases are capable of decarboxylation, it is unlikely that they perform the reaction under normal conditions *in vivo*.
2. Based on the 5caC decarboxylation reaction performed by eM.SssI we developed and validated a new, sensitive and cost-effective CG-resolution caCLEAR method for specific profiling of 5caCG sites in genomic DNA.
3. The caCLEAR method revealed differences in the genomic 5caC distribution between serum-2i and serum-cultured mouse embryonic stem cells corresponding to two pluripotency states – naive and primed. We showed a different enrichment of 5caC in genomic elements, a distinct distribution in the binding regions of important factors and also demonstrated that exclusively in serum conditions 5caC is enriched in genes of developmental processes which are more active in the primed state. 5caC is associated with active expression and is asymmetrically distributed towards the antisense strand of active protein-coding genes.
4. Oxidized forms of 5mrC were detected in RNA of basidiomycete fungi *C. cinerea* and *L. bicolor*. A sensitive mass spectrometry method for the detection of 5carC in RNA was developed using chemical derivatization. Using this method, it was determined that 5carC is mainly present in longer than 200 nt RNAs lacking polyA tails in both fungi, in polyA RNAs in *L. bicolor* and in shorter than 200 nt RNAs in *C. cinerea*, among which 5carC is most abundant in RNAs shorter than tRNA.
5. All three oxidized forms of 5mC were detected in DNA of *L. bicolor* and *C. cinerea*, of which 5hmC and 5fC are the most abundant. Genomic profiling of 5hmC and 5fC in *L. bicolor* and *C. cinerea* was performed, for which we used hmTOP-seq and developed its modified version foTOP-seq. Enrichment of 5hmC and 5fC in transposable elements and suppression of gene expression near them was shown. In *C. cinerea*, strong 5hmCG modification is associated with active genes when 5mCG is low, while in *L. bicolor*, high 5fCG/5hmCG modification is associated with repressed genes, independent of their 5mCG abundance.

## REFERENCES

1. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem.* 2005;74: 481–514.
2. Lu X, Zhao BS, He C. TET family proteins: oxidation activity, interacting molecules, and functions in diseases. *Chem Rev.* 2015;115: 2225–2239.
3. Parry A, Rulands S, Reik W. Active turnover of DNA methylation during cell fate decisions. *Nat Rev Genet.* 2021;22: 59–66.
4. Delatte B, Fuks F. TET proteins: on the frenetic hunt for new cytosine modifications. *Brief Funct Genomics.* 2013;12: 191–204.
5. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science.* 2011;333: 1300–1303.
6. He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science.* 2011;333: 1303–1307.
7. Pfaffeneder T, Hackner B, Truss M, Münzel M, Müller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl.* 2011;50: 7008–7012.
8. Maiti A, Drohat AC. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J Biol Chem.* 2011;286: 35334–35338.
9. An J, Rao A, Ko M. TET family dioxygenases and DNA demethylation in stem cells and cancers. *Exp Mol Med.* 2017;49: e323.
10. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell.* 2013;152: 1146–1159.
11. Iurlaro M, Ficiz G, Oxley D, Raiber E-A, Bachman M, Booth MJ, et al. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* 2013;14: R119.
12. Wu H, Wu X, Shen L, Zhang Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnol.* 2014;32: 1231–1240.
13. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol.* 2010;11: 607–620.
14. Schiesser S, Hackner B, Pfaffeneder T, Müller M, Hagemeyer C, Truss M, et al. Mechanism and stem-cell activity of 5-carboxylcytosine

- decarboxylation determined by isotope tracing. *Angew Chem Int Ed Engl.* 2012;51: 6516–6520.
15. Liutkevičiūtė Z, Kriukienė E, Ličytė J, Rudytė M, Urbanavičiūtė G, Klimašauskas S. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J Am Chem Soc.* 2014;136: 5884–5887.
  16. Yang J, Bashkenova N, Zang R, Huang X, Wang J. The roles of TET family proteins in development and stem cells. *Development.* 2020;147.
  17. Ficiz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, et al. FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell.* 2013;13: 351–359.
  18. Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HHD, Matarese F, et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell.* 2013;13: 360–369.
  19. Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle.* 2009;8: 1698–1710.
  20. Iyer LM, Zhang D, de Souza RF, Pukkila PJ, Rao A, Aravind L. Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proc Natl Acad Sci USA.* 2014;11: 1676–1683.
  21. Stajich JE, Wilke SK, Ahrén D, Au CH, Birren BW, Borodovsky M, et al. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc Natl Acad Sci USA.* 2010;107: 11889–11894.
  22. Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, et al. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature.* 2008;452: 88–92.
  23. Chavez L, Huang Y, Luong K, Agarwal S, Iyer LM, Pastor WA, et al. Simultaneous sequencing of oxidized methylcytosines produced by TET/JBP dioxygenases in *Coprinopsis cinerea*. *Proc Natl Acad Sci USA.* 2014;111: E5149–58.
  24. Zhao L-Y, Song J, Liu Y, Song C-X, Yi C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell.* 2020;11: 792–808.
  25. Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet.* 2014;15: 647–661.

26. Staševskij Z, Gibas P, Gordevičius J, Kriukienė E, Klimašauskas S. Tethered Oligonucleotide-Primed Sequencing, TOP-Seq: A High-Resolution Economical Approach for DNA Epigenome Profiling. *Mol Cell*. 2017;65: 554–564.e6.
27. Gibas P, Narmontė M, Staševskij Z, Gordevičius J, Klimašauskas S, Kriukienė E. Precise genomic mapping of 5-hydroxymethylcytosine via covalent tether-directed sequencing. *PLoS Biol*. 2020;18: e3000684.
28. Vilkaitis G, Dong A, Weinhold E, Cheng X, Klimašauskas S. Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. *J Biol Chem*. 2000;275: 38722–38730.
29. Lukinavičius G, Lapinaitė A, Urbanavičiūtė G, Gerasimaitė R, Klimašauskas S. Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. *Nucleic Acids Res*. 2012;40: 11594–11602.
30. Ličytė J, Gibas P, Skardžiūtė K, Stankevičius V, Rukšėnaitė A, Kriukienė E. A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell Rep*. 2020;32: 108155.
31. Rao PS, Niederpruem DJ. Carbohydrate metabolism during morphogenesis of *Coprinus lagopus* (sensu Buller). *J Bacteriol*. 1969;100: 1222–1228.
32. Ličytė J, Kvederavičiūtė K, Rukšėnaitė A, Godliauskaitė E, Gibas P, Tomkutė V, et al. Distribution and regulatory roles of oxidized 5-methylcytosines in DNA and RNA of the basidiomycete fungi and. *Open Biol*. 2022;12: 210302.
33. Kriukienė E, Labrie V, Khare T, Urbanavičiūtė G, Lapinaitė A, Koncevičius K, et al. DNA unmethylome profiling by covalent capture of CpG sites. *Nat Commun*. 2013;4: 2190.
34. Lukinavičius G, Tomkuvienė M, Masevičius V, Klimašauskas S. Enhanced chemical stability of adomet analogues for improved methyltransferase-directed labeling of DNA. *ACS Chem Biol*. 2013;8: 1134–1139.
35. Masevičius V, Nainytė M, Klimašauskas S. Synthesis of S-Adenosyl-L-Methionine Analogs with Extended Transferable Groups for Methyltransferase-Directed Labeling of DNA and RNA. *Curr Protoc Nucleic Acid Chem*. 2016;64: 1.36.1–1.36.13.
36. Song C-X, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res*. 2017;27: 1231–1242.

37. Nolivos S, Carpousis JC and Clouet-d'Orval B. The K-loop, a general feature of the *Pyrococcus* C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res.* 2005; 33: 6507–6514.
38. Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol.* 2011;29: 68–72.
39. Song C-X, Szulwach KE, Dai Q, Fu Y, Mao S-Q, Lin L, et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell.* 2013;153: 678–691.
40. Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem.* 2014;6: 435–440.
41. Kriukienė E, Liutkevičiūtė Z, Klimašauskas S. 5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA. *Chem Soc Rev.* 2012;41: 6916.
42. Clarke S, Banfield K. S-adenosylmethionine-dependent methyltransferases. In: Carmel R, Jacobsen DW, editors. *Homocysteine in Health and Disease.* Cambridge University Press; 2001. pp. 62–77.
43. Lu X, Han D, Zhao BS, Song C-X, Zhang L-S, Doré LC, et al. Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Res.* 2015;25: 386–389.
44. Neri F, Incarnato D, Krepelova A, Rapelli S, Anselmi F, Parlato C, et al. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* 2015;10: 674–683.
45. Holmes EE, Jung M, Meller S, Lisse A, Sailer V, Zech J, et al. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One.* 2014;9: e93933.
46. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science.* 2012;336: 934–937.
47. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012;149: 1368–1380.
48. Liu Y, Siejka-Zielińska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-

- hydroxymethylcytosine at base resolution. *Nat Biotechnol.* 2019;37: 424–429.
49. Martinez-Val A, Lynch CJ, Calvo I, Ximénez-Embún P, Garcia F, Zarzuela E, et al. Dissection of two routes to naïve pluripotency using different kinase inhibitors. *Nat Commun.* 2021;12: 1863.
  50. Kim S, Günesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, et al. PRMT5 Protects Genomic Integrity during Global DNA Demethylation in Primordial Germ Cells and Preimplantation Embryos. *Mol Cell.* 2014;56: 564–579.
  51. Eguizabal C, Aran B, Chuva de Sousa Lopes SM, Geens M, Heindryckx B, Panula S, et al. Two decades of embryonic stem cells: a historical overview. *Hum Reprod Open.* 2019.
  52. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell.* 2008;133: 1106–1117.
  53. Galonska C, Ziller MJ, Karnik R, Meissner A. Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell.* 2015;17: 462–470.
  54. Atlasi Y, Megchelenbrink W, Peng T, Habibi E, Joshi O, Wang S-Y, et al. Epigenetic modulation of a hardwired 3D chromatin landscape in two naïve states of pluripotency. *Nat Cell Biol.* 2019;21: 568–578.
  55. Martello G, Smith A. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol.* 2014;30: 647–675.
  56. Marks H, Kalkan T, Menafrá R, Denissov S, Jones K, Hofemeister H, et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell.* 2012;149: 590–604.
  57. Huang W, Lan M-D, Qi C-B, Zheng S-J, Wei S-Z, Yuan B-F, et al. Formation and determination of the oxidation products of 5-methylcytosine in RNA. *Chem Sci.* 2016;7: 5495–5502.
  58. Lu X, Song C-X, Szulwach K, Wang Z, Weidenbacher P, Jin P, et al. Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J Am Chem Soc.* 2013;135: 9315–9317.
  59. Yin R, Mao S-Q, Zhao B, Chong Z, Yang Y, Zhao C, et al. Ascorbic Acid Enhances Tet-Mediated 5-Methylcytosine Oxidation and Promotes DNA Demethylation in Mammals. *J Am Chem Soc.* 2013;135: 10396–10403.
  60. Minor EA, Court BL, Young JI, Wang G. Ascorbate Induces Ten-Eleven Translocation (Tet) Methylcytosine Dioxygenase-mediated Generation of



- 5-Hydroxymethylcytosine. *Journal of Biological Chemistry*. 2013;288: 13669–13674.
61. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*. 2010;328: 916–919.
  62. Plaza DF, Lin C-W, van der Velden NSJ, Aebi M, Künzler M. Comparative transcriptomics of the model mushroom *Coprinopsis cinerea* reveals tissue-specific armories and a conserved circuitry for sexual development. *BMC Genomics*. 2014;15: 1–17.
  63. Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, et al. Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLoS Genet*. 2016;12: e1006108.
  64. Borgognone A, Castanera R, Morselli M, López-Varas L, Rubbi L, Pisabarro AG, et al. Transposon-associated epigenetic silencing during *Pleurotus ostreatus* life cycle. *DNA Res*. 2018;25: 451–464.

## DISERTACIJOS TEMA PASKELBTI MOKSLINIAI DARBAI

Disertacija paruošta šių publikacijų pagrindu:

1. Liutkevičiūtė, Z., Kriukienė, E., **Ličytė, J.**, Rudytė, M., Urbanavičiūtė, G., Klimašauskas, S. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *Journal of the American Chemical Society*. 2014;136(16):5884–5887. doi: 10.1021/ja5019223. (Cituojamumo rodiklis – 16.383 (2021))
2. **Ličytė, J.\***, Gibas, P.\*, Skardžiūtė, K., Stankevičius, V., Rukšėnaitė, A., ir Kriukienė, E. A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell reports*. 2020;32(11):108155. doi: 10.1016/j.celrep.2020.108155. (Cituojamumo rodiklis – 9,995 (2021))

\* – pirmieji bendraautoriai

Disertacijos rezultatai buvo pristatyti šiose tarptautinėse konferencijose:

1. **Ličytė, J.**, Liutkevičiūtė, Z., Gordevičius, J., Rudytė, M., Urbanavičiūtė, G., Klimašauskas, S., Kriukienė, E. „Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases“. „1st COST-Epigenetic Chemical Biology Meeting“, Budapeštas, Vengrija, 2015 m. rugsėjo 28-29 d., žodinis pranešimas.
2. **Ličytė, J.**, Liutkevičiūtė, Z., Gordevičius, J., Rudytė, M., Klimašauskas, S., Kriukienė, E. „Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases“. „XIV International Conference of Lithuanian Biochemical Society“, Druskininkai, Lietuva, 2016 m. birželio 27-30 d., žodinis pranešimas.
3. **Ličytė, J.**, Liutkevičiūtė, Z., Gordevičius, J., Rudytė, M., Urbanavičiūtė, G., Klimašauskas, S., Kriukienė, E. „Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases“. „Chromatin and the environment“, Spetses, Graikija, 2016 m. rugpjūčio 8-13 d., stendinis pranešimas.
4. **Ličytė, J.**, Liutkevičiūtė, Z., Gordevičius, J., Rudytė, M., Urbanavičiūtė, G., Klimašauskas, S., Kriukienė, E. „Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases“. COST veiklos CM1406 konferencija, Groningen, Nyderlandai, 2016 m. rugsėjo 15-16 d., trumpas žodinis bei stendinis pranešimai.

5. **Ličytė, J.**, Liutkevičiūtė, Z., Gordevičius, J., Rudytė, M., Urbanavičiūtė, G., Klimašauskas, S., Kriukienė, E. „Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases“. „2nd Danube conference on epigenetics“, Budapeštas, Vengrija, 2016 m. spalio 5-8 d., standinis pranešimas.
6. **Ličytė, J.**, Skardžiūtė, K., Staševskij, Z., Narmontė, M., Gibas, P., Klimašauskas, S., Kriukienė, E. „Mapping the oxidized forms of 5-methylcytosine“. „EpiPredict End Conference 2018, Conference on Systems Epigenetics: Towards Precision Cancer Medicine“, Amsterdamas, Nyderlandai, 2018 m., lapkričio 27-30 d., standinis ir trumpas žodinis pranešimas.
7. **Ličytė, J.**, Godliauskaitė, E., Rukšėnaitė, A., Masevičius, V., Klimašauskas, S., Kriukienė, E. „Detection of 5-carboxylcytosine in RNA“. „The 24th Annual Meeting of the RNA Society“, Krokva, Lenkija, 2019 m. birželio 11-16 d., standinis pranešimas.
8. **Ličytė, J.**, Gibas, P., Skardžiūtė, K., Stankevičius, V., Rukšėnaitė, A., ir Kriukienė, E. „A bisulfite-free approach for base-resolution analysis of genomic 5-carboxylcytosine“. „5th International Conference on Epigenetics and Bioengineering (EpiBio)“, 2021 m. lapkričio 4-6 d., JAV, nuotolinė konferencija, žodinis pranešimas.

## CURRICULUM VITAE

### Janina Ličytė

DNR modifikacijos tyrimų skyrius

Biotechnologijos institutas, Gyvybės mokslų centras, Vilniaus universitetas

Saulėtekio al. 7, LT-10257 Vilnius

Tel. (8-5) 2234365; el. adresas: janina.licyte@bti.vu.lt

#### Išsilavinimas:

2007-2011 m. molekulinės biologijos bakalauro studijos Vilniaus universitete.

2011-2013 m. molekulinės biologijos magistras *Cum Laude* Vilniaus universitete, baigiamasis darbas „Archėjų C/D ribonukleoproteininio komplekso baltymų Nop5p ir aFib dimero vykdomo RNR metilavimo tyrimas“.

2012.09-2012.12 studijos Suomijos Helsinkio universitete pagal Erasmus mainus.

2013-2017 m. biochemijos krypties doktorantūros studijos Vilniaus universiteto Biotechnologijos institute.

#### Darbinė veikla:

Vilniaus universiteto Biotechnologijos instituto DNR modifikacijos tyrimų skyrius: 2009.07-2011.02 praktikantė; 2011.03-2011.12 laborantė; 2013.03-2017.09 biologė tyrėja; nuo 2017.09 jaunesnioji mokslo darbuotoja.

#### Stažuotės:

2012 spalio-gruodis, praktika Helsinkio universiteto Biotechnologijos institute, Suomijoje. Ląstelių citoskeleto baltymų funkcijų ir lokalizacijos tyrimai prof. Pekka Lappalainen laboratorijoje.

2014 m. gegužės 25-28 d. – COST veiklos TD0905 mokykla „Epigenetics Symposium and Training School“, Kopenhaga, Danija.

#### Publikacijos:

1. Liutkevičiūtė, Z., Kriukienė, E., Ličytė, J., Rudytė, M., Urbanavičiūtė, G., Klimašauskas, S. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *Journal of the American Chemical Society*. 2014;136(16):5884–5887. doi:10.1021/ja5019223 (Cituojamumo rodiklis – 16,383 (2021))
2. Tomkuvienė, M., Ličytė, J., Olendraitė, I., Liutkevičiūtė, Z., Clouet-d'Orval, B., Klimašauskas, S. Archaeal fibrillarlin–Nop5 heterodimer 2'-

*O*-methylates RNA independently of the C/D guide RNP particle. *RNA*. 2017;23(9):1329–1337. doi:10.1261/rna.059832.116 (Cituojamumo rodiklis – 5,636 (2021))

3. Ličytė, J.\*, Gibas, P.\*, Skardžiūtė, K., Stankevičius, V., Rukšėnaitė, A., ir Kriukienė, E. A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell reports*. 2020;32(11):108155. doi:10.1016/j.celrep.2020.108155 (Cituojamumo rodiklis – 9,995 (2021))
4. Ličytė, J.\*, Kvederavičiūtė, K.\*, Rukšėnaitė, A., Godliauskaitė, E., Gibas, P., Tomkutė, V., Petraitytė, G., Masevičius, V., Klimašauskas, S., ir Kriukienė, E. Distribution and regulatory roles of oxidized 5-methylcytosines in DNA and RNA of the basidiomycete fungi *Laccaria bicolor* and *Coprinopsis cinerea*. *Open biology*. 2022;12(3):210302. doi:10.1098/rsob.210302 (Cituojamumo rodiklis – 7,124 (2021))

\* – pirmieji bendraautoriai

Tarptautinės konferencijos:

2014 m. – „XIII International Conference of Lithuanian Biochemical Society“, Birštonas, Lietuva, žodinis pranešimas; „FEBS-EMBO 2014“, Paryžius, Prancūzija, standinis pranešimas.

2015 m. – „The Seventh NEB Meeting on DNA Restriction and Modification“, Gdanskas, Lenkija, standinis pranešimas; „1st COST-Epigenetic Chemical Biology Meeting“, Budapeštas, Vengrija, žodinis pranešimas.

2016 m. – „The 24th Annual Meeting of the RNA Society“, Krokva, Lenkija, standinis pranešimas; „Vita Scientia 2016“, Vilnius, Lietuva, standinis pranešimas; „XIV International Conference of Lithuanian Biochemical Society Druskininkai, Lietuva, žodinis pranešimas; „Chromatin and the environment“, Spetses, Graikija, standinis pranešimas; COST veiklos CM1406 konferencija, Groningen, Nyderlandai, trumpas žodinis bei standinis pranešimai; „2nd Danube conference on epigenetics“, Budapeštas, Vengrija, standinis pranešimas.

2018 m. – „EpiPredict End Conference 2018, Conference on Systems Epigenetics: Towards Precision Cancer Medicine“, Amsterdamas, Nyderlandai, standinis ir trumpas žodinis pranešimas.

2021 m. – „5th International Conference on Epigenetics and Bioengineering (EpiBio)“, lapkričio 4-6 d., JAV, nuotolinė konferencija, žodinis pranešimas.

Dalyvavimas moksliniuose projektuose:

2013.04-2015.01, „Molekulinės įrankinės epigenomikai ir Rnomikai“, Nr. VP1-3.1-ŠMM-07-K-01-105. Europos socialinio fondo visuotinė dotacija, S.Klimašauskas.

2013.09-2014.06, Lietuvos mokslų tarybos projektas Nr. MIP-45/2013, E. Kriukienė.

2014.07-2014.12, ES struktūrinių fondų projektas, „Biotechnologija ir biofarmacija: fundamentiniai ir taikomieji tyrimai“, Nr. VP1-3.1-ŠMM-08-K-01-005/KS-560000-1757, K.Sasnauskas.

2018.01-2022.01, „Vienos molekulės TOP-SEQ – inovatyvi technologinė platforma ankstyvai neinvazinei vėžio ir kitų epigenetinių susirgimų diagnostikai“, Europos socialinio fondo dotacija Nr. 09.3.3-LMT-K-712-01-0041, E. Kriukienė.

2018.01-2020.09, „Vienos ląstelės technologija genominio DNR modifikavimo tyrimui ir neuroblastomos epigenetinio heterogeniškumo analizė“, LMT mokslinių grupių projektas Nr. MIP-58-17, E. Kriukienė.

Stipendijos:

2010, 2012 m. – LMT Studentų mokslinė praktika. 2011 m. – LMT Studentų moksliniai tyrimai. 2015 m. – Lietuvos mokslų tarybos doktoranto stipendija už akademinis pasiekimus. 2014 m. – FEBS stipendija dalyvauti „FEBS-EMBO 2014“ konferencijoje. 2015 m. – „New England Biolabs“ stipendija „The Seventh NEB Meeting on DNA Restriction and Modification“ konferencijai. 2016 m. – organizatorių stipendija dalyvavimui konferencijoje „2nd Danube conference on epigenetics“. 2016 m. – FEBS stipendija dalyvavimui „Chromatin and the environment“ konferencijoje.

Moksliniai interesai:

Epigenetika, DNR ir RNR modifikacijos bei jų funkcijos, metodai jų nustatymui, DNR/RNR metiltransferazių savybės.

Pedagoginė ir mokslo populiarinimo veikla:

2007, 2008 m. – dalyvavimas Lietuvos gamtos mokslų asociacijos savanoriškoje veikloje.

2016 m. – ekskursijų vedimas mokiniam ir mokytojams Vilniaus universiteto Gyvybės mokslų centro Biotechnologijos instituto DNR modifikacijos tyrimų skyriuje.

Vadovavimas ruošiant 2 baigiamuosius bakalauro darbus; konsultavimas ruošiant 2 baigiamuosius magistro darbus.

## PADĖKA

Dėkoju vadovei dr. Editai Kriukienei už didelį indėlį į šį darbą, pasiūlytą temą, perduotas žinias, patarimus, įžvalgas ir idėjas, koordinavimą, suteiktas galimybes atlikti eksperimentus ir dalyvauti konferencijose, viltį tęsiant sudėtingus darbus ir pagalbą rašant.

Prof. habil. dr. Sauliui Klimašauskui už patarimus, suteiktas galimybes atlikti darbus ir dalyvauti konferencijose.

Visiems kolegoms, be kurių indėlio nebūtų pasiekti šie rezultatai. Dr. Povilui Gibui, Kotrynai Kvederavičiūtei bei Vitai Tomkutei už sekoskaitos duomenų analizę ir vizualizavimą. Audronei Rukšėnaitei už sekoskaitas, dalį atliktų HPLC-MS/MS, optimizuotas sąlygas, pamokas, gerumą bei palaikymą sugaudant modifikacijas. Kotrynai Skardžiūtei už kai kuriuos eksperimentus ir pakartojimus, puikų darbą kartu ir draugiškumą. Prof. Viktorui Masevičiui už Ado-6-N<sub>3</sub> ir chemijos konsultacijas. Prof. V. Masevičiui ir dr. Gražinai Petraitytei už (4-aminometil)benzilazido sintezę. Prof. Francis M. Martin už dovanotą *L. bicolor*. Prof. Guo-Liang Xu už pelės ESC, dr. Vaidotui Stankevičiui už jų auginimą. „Thermo Fisher Scientific Baltics“ už *L. bicolor* RNR bibliotekas ir jų sekoskaitą. Zdislav Staševskij už TOP-seq sąlygas ir pirkimus. Eglei Godliauskaitei už pirmuosius darbus su RNR, derivatizacijos optimizavimą. Ievai Stirblytei ir Mildai Narmontei už kai kuriuos pirmuosius bandymus, taškinio imunobloto sąlygas. Dr. Zitai Liutkevičiūtei už masių spektrometrijos pradžiamokslį ir atrastas reakcijas. Giedrei Urbanavičiūtei už eM.SssI ir kitų reagentų skolinimą.

DMTS kolektyvo žmonėms už skolintus reagentus, patarimus, perduotas žinias, pamokas, pagalbą. Visiems nuo kurių pradinių darbų ir patirties galėjau atsispirti tolesniuose tyrimuose. Visiems studentams, prisidėjusiems prie eksperimentų, varginančių gryninimų ir mokiusiems mane kantrybės. Visiems padėjusiems, palaikiusiems, praskaidrinusiems laiką ir sunkumus, ypač dr. Mildai Mickutei.

Įkvėpusiems ir skatinusiems mokslo bendražygiams, dėstytojams, biologijos mokytojai Reginai Mongirdienei. Gerus pagrindus suteikusiai pirmajai vadovei dr. Miglei Tomkuvienei.

Šeimai ir draugams už palaikymą, supratimą, optimizmą, paramą, paskatinimus ir dar daug daugiau. Visiems sutiktiems nuostabiems žmonėms ypač palaikiusiems mane sunkiais momentais, stiprinusiems ir auginusiems mane šiame kelyje.

Finansinė parama:

Europos socialinio fondo visuotinė dotacija Nr. VP1-3.1-ŠMM-07-K-01-105, S. Klimašauskas.

Europos socialinio fondo lėšos pagal priemonės Nr. 09.3.3-LMT-K-712 veiklą „Mokslininkų kvalifikacijos tobulinimas vykdant aukšto lygio MTEP projektus“, Nr. 09.3.3-LMT-K-712-01-0041, E. Kriukienė.

LMT mokslinių grupių projektas Nr. MIP-58-17, E. Kriukienė

Europos mokslo tarybos dotacija Nr. ERC-AdG-2016/742654, S. Klimašauskas.

2015 m. Lietuvos mokslų tarybos doktoranto stipendija.

COST veiklą CM1406, TN1302 parama konferencijoms. FEBS ir „2nd Danube conference on epigenetics“ organizatorių stipendija konferencijoms.



UŽRAŠAMS

UŽRAŠAMS

UŽRAŠAMS

Vilniaus universiteto leidykla  
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius  
El. p. [info@leidykla.vu.lt](mailto:info@leidykla.vu.lt), [www.leidykla.vu.lt](http://www.leidykla.vu.lt)  
[bookshop.vu.lt](http://bookshop.vu.lt), [journals.vu.lt](http://journals.vu.lt)  
Tiražas 15 egz.