

VILNIAUS UNIVERSITETAS

Juozas
NAINYS

Aukšto našumo pavienių ląstelių sekoskaita ir analizė

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,
Biochemija (N 004)

VILNIUS 2020

Disertacija rengta 2015– 2019 metais Vilniaus universiteto Gyvybės mokslų centro Biotechnologijos institute

Mokslinis vadovas:

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

Gynimo taryba

Pirmininkas:

Prof. dr. Kęstutis Sužiedėlis (Vilniaus universitetas, gamtos mokslai, biochemija - N 004);

Nariai:

dr. Edita Kriukienė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004);

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

dr. Urtė Neniškytė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004);

prof. dr. Sonata Jarmalaitė (Nacionalinis vėžio institutas, gamtos mokslai, biologija – N 010).

Disertacija ginama viešame Gynimo tarybos posėdyje 2020 m. rugsėjo mėn. 25 d. 16 val. Gyvybės mokslų centre (Vilniaus universitetas) R-102 auditorijoje. Adresas: Saulėtekio al. 7, LT- 10257 Vilnius, Lietuva.

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

VILNIUS UNIVERSITY

Juozas
NAINYS

High-throughput single-cell sequencing and analysis

SUMMARY OF DOCTORAL DISSERTATION

Natural Sciences,
Biochemistry (N 004)

VILNIUS 2020

This dissertation was prepared between 2015 and 2019 in Vilnius University Life Sciences Center Institute of Biotechnology.

Academic supervisor:

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

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

Chairman :

Prof. dr. Kęstutis Sužiedėlis (Vilnius University, Natural Sciences, Biochemistry – N 004);

Members:

dr. Edita Kriukienė (Vilnius University, Natural Sciences, Biochemistry – N 004);

prof. dr. Artūras Petronis (Toronto University, Natural Sciences, Biology – N 010);

dr. Urtė Neniškytė (Vilnius University, Natural Sciences, Biochemistry – N 004);

prof. dr. Sonata Jarmalaitė (National Cancer Institute, Natural Sciences, Biology – N 010).

The dissertation shall be defended at a public meeting of the Dissertation Defense Panel at 16:00 (hour)/ on 25th September 2020 in R-102 auditorium of the Life Sciences Center (Vilnius University).
Address: Saulėtekio av. 7, LT- 10257 Vilnius, Lithuania

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

www.vu.lt/lt/naujienos/ivykiu-kalendorius

ĮVADAS

Nedvejojant galima teigti, jog ląstelės yra pamatinis gyvybės vienetas [1]. Itin svarbus žingsnis siekiant pažinti žmogaus ląstelių biologiją buvo žmogaus genomo sekos nustatymas [2]. Tačiau jau nuo ankstyvų žmogaus genomo projekto dienų buvo aišku, jog ši informacija nebus pakankama siekiant suprasti ląstelių biologinę įvairovę žmogaus organizme – visos vieno organizmo ląstelės turi tą pačią DNR seką. Analizuoti ląstelių įvairovę yra kur kas parankiau tyrinėjant jų transkriptomą. Būtent ši informacija gali atskleisti aktyvias genomo dalis bei funkcinius molekulinis junginius, atsakingus už ląstelės fenotipą. Transkriptomo analizės galimybės pirmą kartą pademonstruotos 1995 metais [3]. Kitas svarbus postūmis šioje srityje įvyko 2009 metais, kai mokslininkai pristatė pavienių ląstelių transkriptomo analizės technologiją [4]. Ankstyvieji pavienių ląstelių tyrimai buvo brangūs ir imlūs darbui. Todėl pagrįstai galima teigti, jog technologija buvo nepraktiška ir neprieinama daugeliui tyrėjų. Esminis technologinis proveržis pavienių ląstelių analizės srityje įvyko tik visai neseniai – 2015 metais buvo pristatyta ląstelių mikroskopsčiais paremta pavienių ląstelių transkriptomo analizės platforma [5, 6]. Ši technologija įgalino tyrėjus nebrangiai ir nesudėtingai analizuoti dešimčių tūkstančių pavienių ląstelių transkriptomus vieno eksperimento metu. Tokios metodikos savybės atvėrė iki tol neįsivaizduojamas sudėtingų biologinių organizmų analizės galimybes. Nenuostabu, jog netrukus vienas po kito sekė atradimai: nustatytas naujas dendritinių ląstelių tipas žmogaus organizme, atrastas ląstelių tipas, atsakingas už cistinės fibrozės išsivystymą, bei sudaryti įvairių audinių atlasai, parodę iki tol nežinomus naujus ląstelių tipus. Šie pavienių ląstelių transkriptomo analizės taikymai įrodė tokių tyrimų svarbą ir paskatino Žmogaus Ląstelių Atlaso konsorciumo subūrimą. Šios iniciatyvos tikslas – sukurti visų žmogaus organizmo ląstelių atlasą ir atverti naują žmogaus biologijos pažinimo puslapį [7, 8]. Vis dėlto, siekiant

pavienių ląstelių tyrimus pritaikyti ne tik fundamentiniuose tyrimuose, bet ir medicinos praktikoje, metodikos turi būti tobulinamos. Iš technologinės perspektyvos svarbu padidinti jų efektyvumą, sumažinti tyrimų kainą bei supaprastinti eksperimento eigą. Nepaisant to, jau šiandien yra aišku, jog pavienių ląstelių analizė, leidžianti daug detaliau pažvelgti į sudėtingas biologines sistemas, neabejotinai prisidės prie svarbių atradimų.

Ląstelės būseną yra dinamiška ir gali kisti priklausomai nuo įvairių veiksnių. Vienas iš įdomių ir gerai ištirtų ląstelių būsenos pokyčių yra epitelinių ląstelių virsmas mezenchiminėmis (EMT). Tai yra nuoseklus procesas, kurio metu ląstelės kardinaliai pakeičia savo būseną ir fenotipą ir iš sėslių tampa judriomis. Žmogaus organizme šis virsmas yra stebimas įvairių procesų metu ir pagal tai yra skirstomas į tris tipus [9, 10]. Pirmojo tipo virsmas yra svarbus embriogenezės užtikrinimui. Antrojo tipo EMT stebimas žaizdų gijimo metu. Sutrikus šiam procesui gali išsivystyti audinių fibrozės patologijos. Trečiojo tipo virsmas nustatomas vėžiniame audinyje, kur jis daro įtaką naviko susiformavimui bei metastazių pasklidimui po organizmą. Būtent šis EMT tipas pastaraisiais dešimtmečiais sulaukia vis daugiau susidomėjimo. Dėl svarbaus vaidmens vėžio vystymesi EMT yra patrauklus taikynys naujoms gydymo terapijoms ir vaistams [11]. Tačiau, nepaisant to, jog procesas yra detaliai tiriamas jau virš 30 metų, jo kompleksiskumas ir variabilumas riboja praktinį sukauptų žinių pritaikymą medicinoje.

EMT virsmas yra tik vienas iš daugelio procesų svarbių vėžio biologijoje. Dėl didelės ląstelių tipų įvairovės ir kompleksinės jų tarpusavio sąveikos, vėžinį naviką pagrįstai galima laikyti sudėtinga biologine sistema. Šiame darbe tirtas krūties vėžys yra dažniausias moterims diagnozuojamas vėžio tipas [12]. Skaičiuojama, jog virš 2,1 milijono moterų ši liga buvo diagnozuota 2018 metais ir net 600000 pacienčių pralaimėjo kovą su liga [13]. Be to, yra stebimas 3,1% metinis šio vėžio atvejų augimas pasauliniu mastu [14]. Išsamių tyrimų dėka yra žinoma, jog krūties vėžys – tai heterogeniška liga.

Didelė navikų įvairovė yra nulemiama plataus genetinių mutacijų, susijusių su šia patologija, spektro. Nepaisant to, daugeliu atvejų egzistuoja veiksmingos gydymo strategijos. Nustačius ligą ankstyvoje stadijoje, iki 70-80% pacienčių yra išgydomos. Tačiau krūties vėžį aptikus vėlyvoje stadijoje (susidarius metastazėms) liga yra laikoma neišgydoma, o vidutinė pacientų gyvenimo trukmė yra tik 2-3 metai [15]. Akivaizdu, jog vėlyvos stadijos krūties vėžio gydymas išlieka itin aktualia problema. Viena iš naujų ir daug žadančių vėžio gydymo strategijų yra imunoterapija. Todėl pastaruoju metu krūties vėžio imunologijos tyrimai sulaukia didesnio dėmesio. Tačiau nepaisant to, jog šiandien jau yra daug žinoma apie atskirus imuninių ląstelių tipus krūties vėžio kontekste, bendras paveikslas išlieka neaiškus. Be to, yra stebimas didelis imuninių ląstelių populiacijų variabilumas tarp skirtingų pacientų ir skirtingų krūties vėžio tipų [16]. Sisteminių žinių trūkumas tiesiogiai riboja sėkmingą imunoterapijos taikymą krūties vėžio gydymui.

Darbo tikslas

Optimizuoti ląstelių mikroskysčių technologija paremtą aukšto našumo pavienių ląstelių transkriptomo analizės metodą ir taikyti jį sudėtingų biologinių sistemų tyrimui.

Uždaviniai

- Aprašyti detalų lašelių mikroskysčių technologija paremto aukšto našumo pavienių ląstelių transkriptomio analizės metodo taikymo protokolą.
- Optimizuoti pavienių ląstelių transkriptomio sekoskaitos bibliotekos paruošimo procesą.
- Charakterizuoti EMT virsmą, tiriant HMLE ląstelių modelinę sistemą.
- Nustatyti ir patvirtinti ZEB1 transkripcijos faktoriaus taikinius EMT virsmo metu.
- Sudaryti krūties vėžio pacientų imuninių ląstelių atlasą.
- Charakterizuoti imunines ląsteles, aptinkamas krūties vėžio navikuose.

Mokslinis naujumas

Šiame darbe pirmą kartą išsamiai aprašytas lašelių mikroskysčių technologija paremtas aukšto našumo pavienių ląstelių transkriptomio analizės metodo (scRNA-Seq) protokolą. Šis metodas pavadintas „inDrops“, o aprašyta procedūra buvo pristatyta aukšto lygio moksliniame žurnale. Tai gerokai padidino technologijos sklaidą ir įgalino mokslininkus, neturinčius patirties šioje srityje, atlikti pavienių ląstelių tyrimus. Be to, šiame darbe pristatomas inDrops metodo optimizacijos leidžia iki 10-ies kartų padidinti iš pavienių ląstelių nustatomų transkriptų skaičių, lyginant su pradine metodo versija. Taip pat, aprašytos optimizacijos leidžia sumažinti analizės kaštus, kas yra svarbu vykdant didelės apimties tyrimus. Šiame darbe pristatomi rezultatai yra aktualūs ne tik inDrops metodui, bet ir platesniame pavienių ląstelių transkriptomio technologijų kontekste, kadangi dauguma skirtingų metodų turi bent keletą vienodų techninių žingsnių.

Tolesniame darbo etape scRNA-Seq metodika buvo taikoma analizuojant kompleksinį biologinį procesą, naudojant modelinę sistemą. Tiriant HMLE ląstelių modelinę sistemą, EMT virsmas buvo pirmą kartą charakterizuotas pavienių ląstelių lygmenyje. Atsižvelgiant į scRNA-Seq technologijos naujumą, pirmame darbo etape buvo siekiama įvertinti, ar pavienių ląstelių transkriptomo analizė leidžia aptikti jau žinomas biologines sistemos savybes (pvz.: konkrečių genų tinklų aktyvaciją). Tyrimo rezultatai leidžia teigti, jog scRNA-Seq technologija patikimai charakterizuoja tiriamą sistemą. Be to, darbo metu surinkti duomenys leido patvirtinti iki šiol diskutuojamą hipotezę, jog epitelinės ląstelės pirma įgyja kamieninių ląstelių fenotipą, o tik po to tampa mezenchiminėmis ląstelėmis. Vienas iš technologinių taikyto metodo trūkumų yra duomenų retumas – nustatoma tik dalis (iki 30%) kiekvienos ląstelės transkriptomo informacijos. Dėl šios savybės silpnos genų sąveikos gali būti sunkiai pastebimos pradiniuose duomenyse. Disertacijoje pristatomi rezultatai pirmą kartą parodo, jog ši problema gali būti sprendžiama praturtinant duomenis, taikant difuzinių žemėlapių (*angl.* diffusion maps) algoritmus. Tokiu būdu praturtinti duomenys buvo panaudoti tiksliai kartinio transkripcijos faktoriaus EMT procese – ZEB1 taikinių nustatymui. Rezultatai atskleidė iki šiol neparodytą genų tinklų, formuojančių EMT virsmą, įvairovę ir leido tiksliai nustatyti iki šiol neaprašytus ZEB1 transkripcijos faktoriaus aktyvacijos taikinius. Tokia analizės strategija gali būti plačiai pritaikoma, nes nereikalauja papildomų mėginio manipuliacijų. Ši savybė yra itin patraukli analizuojant klinikinius mėginius, o tokie tyrimai galėtų palengvinti naujų vaistų taikinių atradimą įvairių ligų kontekste.

Parodžius, jog pavienių ląstelių transkriptomo analizė gali būti sėkmingai taikoma sudėtingų biologinių sistemų tyrimams, buvo vykdomi klinikinių mėginių tyrimai. Bendradarbiaujant su MSKCC (JAV) gydytojais pirmą kartą parengtas krūties vėžio pacientų imuninių ląstelių atlasas. Šį atlasą sudaro 62000 pavienių ląstelių, surinktų iš aštuonių pacientų, analizuojant keturis audinius: sveiką bei

vėžinį krūties audinį, kraują bei limfmazgį. Tyrimo rezultatai atskleidė itin didelę įgyto bei įgimto imuniteto ląstelių įvairovę. Disertacijoje pristatomi duomenys patvirtina kitų mokslininkų aprašomą imuninių ląstelių sudėties variabilumą tarp skirtingų pacientų bei audinių. Be to, nustatyta, jog T ląstelių fenotipinė įvairovė navike yra gerokai didesnė lyginant su sveiku krūties audiniu. Atlikti tyrimai leidžia manyti, jog lokalių nišų įvairovė vėžio mikroaplinkoje atlieka svarbų vaidmenį formuojant organizmo imuninį atsaką į naviką. Trys pagrindiniai veiksniai, lemiantys T ląstelių fenotipinės įvairovės padidėjimą, yra T ląstelių aktyvacija, terminalinė diferenciacija bei hipoksija. Be to, nustatyta, jog T ląstelės genų raiškos erdvėje išsidėsto tolydžiai pagal aktyvacijos potencialą. Tokie tyrimo rezultatai prieštarauja nusistovėjusiam modeliui, teigiančiam, jog aktyvacijos metu T ląstelės diskrečiai pereina nuo vieno iki kito aiškiai išreikšto stabilaus fenotipo (pvz.: reguliacinės T ląstelės, atminties T ląstelės, efektorinės T ląstelės ir kt.). Rezultatų aktualumą pabrėžia ir tai, jog panašios hipotezės buvo neseniai pristatytos autoimuninės ligos kontekste [17]. Taip pat darbo metu parodyta, jog makrofagų poliarizacijos būsenos nėra viena kitai prieštaraujančios – ląstelės gali būti tarpinėse bei suminėse poliarizacijos būsenose. Šie duomenys papildo pastaruoju metu keliamas naujas hipotezes apie makrofagų poliarizaciją.

Disertacijoje pristatomi teiginiai:

- Lašelių mikroskopsčių technologija paremta aukšto našumo pavienių ląstelių transkriptomo analizės metodo efektyvumas gali būti padidinamas optimizuojant atskirus protokolo žingsnius.
- Duomenų praturtinimo algoritmai gali efektyviai atkurti genų sąveikų informaciją iš retų pavienių ląstelių transkriptomo duomenų.
- Pavienių ląstelių transkriptomo analizė leidžia tiksliai nuspėti transkripcijos faktorių aktyvacijos taikinius.
- Audinių mikroaplinka formuoja imuninių ląstelių fenotipines būsenas.
- Krūties vėžio navikuose aptinkamų T ląstelių fenotipinių būsenų visuma yra tolydi, o ją formuoja lokaliios naviko mikroaplinkos nišos.

Disertacijos turinys:

Disertacija parengta anglų kalba ir susideda iš šių dalių: santrumpų sąrašas, įvadas, mokslinio naujumo aprašymas, literatūros apžvalga, naudotų metodų sąrašas, rezultatai ir jų aptarimas, išvados, publikacijų disertacijos tema sąrašas (3 publikacijos, 5 pranešimai mokslinėse konferencijose), padėka bei literatūros sąrašas (397 šaltinių). Disertacijoje yra pateikiami paveikslai (27) bei lentelės (18). Disertaciją sudaro 138 puslapiai.

METODAI

Pavienių ląstelių transkriptomo barkodavimas

Pavienių ląstelių transkriptomas buvo barkoduojamas naudojant ląstelių mikroskysčių platformą. Šis procesas susideda iš trijų atskirų dalių: i) mikroskysčių lusto ruošimo, ii) barkoduojančių hidrogelinių rutuliukų (BHR) ruošimo ir iii) pavienių ląstelių transkriptomo barkodavimo mikroskysčių ląseliuose. Mikroskysčių lustai buvo ruošiami pagal anksčiau aprašytą metodiką [18]. Detalus protokolas, apimantis barkoduojančių hidrogelinių rutuliukų ruošimą bei pavienių ląstelių transkriptomo barkodavimą mikroskysčių ląseliuose, pateikiamas darbo metu parengtoje publikacijoje [19]. Tyrimo metu naudoti mikroskysčių lustų tėkmės greičiai pateikiami 1-oje lentelėje. Tyrimo metu naudotos pavienių ląstelių transkriptomo barkodavimo reakcijos sąlygos ląseliuose pateikiamos 2-oje lentelėje. Siekiant sumažinti techninį variabilumą mikroskysčių lustai ir BHR buvo ruošiami partijomis ir naudojami keliuose scRNA-Seq eksperimentuose.

Lentelė 1. Darbo metu mikroskysčių ląsteliuose naudoti tėkmės greičiai.

Reagentas	Tėkmės greitis
<i>Hidrogelinių rutuliukų sintezė</i>	
Akrilamido ir pradmenų mišinys	900 µl/val.
Mikroskysčių alyva	1800 µl/val.
<i>Pavienių ląstelių transkriptomo barkodavimas</i>	
Pavienių ląstelių suspensija	250 µl/val.
AT-lizės reagentai	250 µl/val.
Barkoduojančios hidrogeliniai rutuliukai	80±10 µl/val.
Mikroskysčių alyva	700 µl/val. – 900 µl/val.

Lentelė 2. Darbo metu naudotos transkriptomo barkodavimo reakcijos sąlygos.

AT fermentas	Reakcijos sąlygos lašeliuose	Temperatūrinis režimas
SuperScript III (Invitrogen, cat. no. 18080044)	37,6 mM KCl, 45 mM NaCl, 5,8 mM MgCl ₂ , 54 mM Tris-HCl [pH 8,0], 0,3 mM KH ₂ PO ₄ , 0,87 mM Na ₂ HPO ₄ , 0,4% (v/v) Igepal-CA630, 0,017% (v/v) BSA, 3,9% (v/v) Optiprep, 2,17 mM DTT, 0,44 mM dNTPs, 1,16 U/ml RiboLock RNazių slopiklis ir 10,4 U/ml SuperScript-III AT fermentas	<i>Reakcijos iniciacija:</i> 1 min 60°C <i>AT reakcija:</i> 2 val 50°C <i>Inaktyvacija:</i> 15 min 70°C
SuperScript IV (Invitrogen, cat. no. 8090010)	9,8 mM KCl, 45 mM NaCl, 0,4 mM MgCl ₂ , 6,5 mM Tris-HCl [pH 8,0], 0,3 mM KH ₂ PO ₄ , 0,87 mM Na ₂ HPO ₄ , 0,4% (v/v) Igepal-CA630, 0,017% (v/v) BSA, 3,9% (v/v) Optiprep, 0,435 SSIV RT buffer, 2,17 mM DTT, 0,44 mM dNTPs, 1,16 U/ml RiboLock RNazių slopiklis ir 10,4 U/ml Maxima H minus AT fermentas	<i>AT reakcija:</i> 60 min 50°C <i>Inaktyvacija:</i> 10 min 80°C
Maxima H minus (Thermo Scientific, cat. no. EP0751)	43 mM KCl, 45 mM NaCl, 1,7 mM MgCl ₂ , 28 mM Tris-HCl [pH 8,0], 0,3 mM KH ₂ PO ₄ , 0,87 mM Na ₂ HPO ₄ , 0,4% (v/v) Igepal-CA630, 0,017% (v/v) BSA, 3,9% (v/v) Optiprep, 4,4 mM DTT, 0,44 mM dNTPs, 1,16 U/ml RiboLock RNazių slopiklis ir 10,4 U/ml Maxima H minus AT fermentas	<i>AT reakcija:</i> 60 min 50°C <i>Inaktyvacija:</i> 10 min 80°C

Sekoskaitos bibliotekos ruošimas

Darbo metu pavienių ląstelių transkriptomo analizės sekoskaitos bibliotekos buvo ruošiamos remiantis protokolu, pateikiamu darbo metu parengtoje publikacijoje [19]. Sekoskaitos bibliotekos ruošimas susideda iš keleto atskirų žingsnių. Po pavienių ląstelių transkriptomo barkodavimo reakcijos emulsija padalinama į atskirus mėginius, kurių kiekviename yra po ~3000 pavienių ląstelių. Toliau emulsija suardoma ir apjungta kDNR valoma nuo hidrogelinių rutuliukų, reakcijos metu nepanaudotų pradmenų bei susidariusių pradmenų dimerų. Išvalyta kDNR yra naudojama antros grandinės sintezės (AGS) reakcijoje, o gautas produktas dauginamas *in vitro* transkripcijos (IVT) reakcijos pagalba. Padauginta RNR fragmentuojama ir atvirkštinės transkripcijos (AT) reakcijos metu konvertuojama į DNR biblioteką. Toliau ši yra padauginama ir pridedamos sekoskaitai reikalingos specifinės sekos. Viso proceso yra du kartus atliekama bibliotekos kokybės kontrolė vertinant bibliotekos sudėtį pagal molekulių dydžius. Kokybės kontrolės kriterijus atitinkanti biblioteka yra naudojama sekoskaitai atlikti. Darbo metu buvo papildomai optimizuojami atskiri metodikos žingsniai, atliekami iki bibliotekos padauginimo.

Sekoskaita

Darbo metu paruoštų bibliotekų sekoskaita buvo atliekama Illumina kompanijos sekoskaitos prietaisais MiSeq, NextSeq 550 ir HiSeq2500, naudojant „FastQ Only“ režimą bei suderinamus sekoskaitos reagentus. Visais atvejais buvo atliekama abiejų molekulių galų sekoskaita (*angl.* paired end). Dėl skirtingų prietaisų savybių sekoskaitos nuskaitymų ilgis varijavo tarp skirtingų eksperimentų, tačiau visais atvejais pirmojo nuskaitymo ilgis buvo ≥ 51 bp, o antrojo – ≥ 35 bp. Visų sekoskaitos eksperimentų metu pirmojo nuskaitymo sekoje buvo ląstelės barkodo bei molekulių žymens informacija, o antrojo – konkretaus transkripto sekos informacija.

Duomenų analizė

Darbo metu taikytą duomenų analizės procesą galima skirstyti į dvi dalis: i) pradinis duomenų apdorojimas ir ii) specifinių algoritmų taikymas. Pradiniam duomenų apdorojimui buvo naudojama SEQC analizės platforma, kurios veikimo principas detaliam aprašytas vienoje iš publikacijų disertacijos tema [20]. SEQC platformos rezultatas yra ląstelių ir genų matrica, aprašanti visas vieno eksperimento metu tirtas ląsteles. Toliau ši matrica buvo filtruojama pašalinant: i) negyvas ląsteles (mtRNR >20%), ii) žemo kompleksiško ląsteles (remiantis žemu nustatytu unikalių genų kiekiu, lyginant su visomis vienai ląstelei priskiriamomis molekulėmis) ir iii) triukšmingus barkodus (remiantis histograma, apibendrinančia su kiekvienu barkodu susijusių transkriptų kieki). Taip apdoroti duomenys buvo pateikiami kompiuteriniams algoritmams tolesnei tikslinei analizei. Darbo metu duomenų analizei naudotų algoritmų sąrašas pateikiamas 3-oje lentelėje.

Lentelė 3. Darbo metu naudoti duomenų analizės algoritmai ir trumpas jų aprašymas.

Algoritmas	Trumpas apibūdinimas	Šaltinis
EMT duomenų analizė		
MAGIC	Duomenų praturtinimas	[21]
Principal Convex Hull Analysis (PCHA)	Archetipų identifikavimas	[22]
kNN-DREMI	Genų koreliacijų įvertinimas	[21]
DREVI	Genų tinklų identifikavimas	[21]
Krūties imuninių ląstelių analizė		
Biscuit	Duomenų praturtinimas ir grupavimas	[20]
Phenograph	Duomenų grupavimas	[23]
t-SNE	Duomenų vizualizacija	[24]
Diffusion maps	Pagrindinių komponentų nustatymas	[25]

Lentelė 4. Pradmenų, naudotų kPGR tyrime, sekos. RNR kiekio matavimų duomenys yra specifiniai K562 ląstelėms. Duomenys iš Human Protein Atlas duomenų bazės.

Genas	Tiesioginio pradmens seka	Atvirkštinio pradmens seka	Santykinis RNR kiekis
ACTB	CGCCGCCAGTCCACC	TCTCCATGTCGTCACAGTTG	1137
VIM	CGGGAGAAATTGCAGGAGGA	TCTTGGCAGCCACACTTCA	508
B2M	CTCACGTCATCCAGCAGAGAA	TGCTTACATGTCTCGATCCCAC	185
TGFB1	TACCTGAACCCGTGTTGCTC	CCGGTAGTGAACCCGTTGAT	58
STAT3	GGAGAAACAGGATGGCCAA	ACCTGCTCTGAAGAACTGCT	52
AKT1	AAGTCATCGTGGCCAAGGAC	GTTCTCCAGCTTGAGGTCCC	42
SMAD2	GTTCTTCTCTCTCCGCTC	CTTGATCGAACCTCCCGGC	30
EGFR	CGAATGGGCCTAAGATCCCG	CCCTTATACACCGTGCCGAA	0

Kiekybinė PGR

Darbo metu buvo mėginiai iki bibliotekos padauginimo žingsnio buvo lyginami atliekant kiekybinę polimerazės grandininę reakciją (kPGR) pagal gamintojo rekomendacijas (Thermo Scientific, cat. no. K0222). Tyrimo metu buvo matuojami aštuoni skirtingi genai. Jų raiškos lygis ir naudotų pradmenų sekos pateikiamos 4-oje lentelėje.

AT fermentų palyginimas

Darbo metu buvo vertinami trys skirtingi AT fermentai: SuperScript III (Invitrogen, cat. no. 18080044), SuperScript IV (Invitrogen, cat. no. 8090010) bei Maxima H minus (Thermo Scientific, cat. no. EP0751). Fermentų efektyvumo tyrimai buvo atliekami analizuojant pavienių ląstelių transkriptomus scRNA-Seq platformoje pagal aprašytą metodiką.

kDNR valymo metodų palyginimas

Darbo metu buvo vertinamos dvi skirtingos kDNR valymo metodikos. Pirmas žingsnis abiejų metodikų atveju yra vienodas – hidrogeliniai rutuliukai yra pašalinami AT reakcijos mišinį centrifuguojant 1 min 1000 g. greičiu per kolonėlę (Zymo, cat. no.

C1004-50). Kitu žingsniu reakcijos metu nesunaudoti pradmenys bei susidarę pradmenų dimerai yra pašalinami fermentinės DNR hidrolizės pagalba. Reakcija vykdoma 80 μ l tūryje naudojant 1 μ l ExoI (Thermo Scientific, cat. no. EN0581), 2 μ l HinfI (Thermo Scientific, cat. no. FD0804) bei 0,5 μ l FastAP (Thermo Scientific, cat. no. EF0654) fermentų, ir 0,5x FD komercinį buferį (Thermo Scientific, cat. no. B64). Po reakcijos mišinys yra valomas AMPure reagentu (Beckman Coulter, cat. no. A63881), naudojant 1,2x tūrio santykį. Kito metodo atveju fermentinė hidrolizė nebuvo vykdoma, o AT reakcijos mišinys buvo skiedžiamas iki 100 μ l galutinio tūrio ir valomas AMPure reagentu, naudojant skirtingus tūrio santykius. kDNR valymo metodikų efektyvumo tyrimai buvo atliekami analizuojant pavienių ląstelių transkriptomus scRNA-Seq platformoje pagal aprašytą metodiką.

Antros grandinės sintezės reagentų palyginimas

Darbo metu buvo vertinamos dvi skirtingos AGS reakcijos mišinio sudėty. Vienu atveju buvo naudojamas komercinis reagentų rinkinys (NEB, cat. no. E6111S) ir reakcija buvo vykdoma 20 μ l tūryje, pridedant po 1 μ l kiekvieno rinkinyje esančio komponento. Kitu atveju buvo naudojami atskiri komerciškai prieinami reagentai. Analogiškai, reakcija buvo vykdoma 20 μ l tūryje naudojant 1x SSB reakcijos buferį (Thermo Scientific, cat. no. 10812014), 0,133 μ l DNR PolI (Thermo Scientific, cat. no. EP0041), 0,133 μ l T4 DNR ligazės (Thermo Scientific, cat. no. EL0012) ir 0,533 μ l RNazės H (Thermo Scientific, cat. no. EN0202) fermentus bei 200 μ M dNTP. Skirtingų AGS reakcijų sudėčių efektyvumo tyrimai buvo atliekami analizuojant pavienių ląstelių transkriptomus scRNA-Seq platformoje pagal aprašytą metodiką.

IVT reagentų palyginimas

Darbo metu buvo vertinami du skirtingi komerciškai prieinami reagentų rinkiniai IVT reakcijai: HiScribe T7 (NEB, cat. no. E2040S), TranscriptAid T7 (Thermo Scientific, cat. no. K0441). Abiem atvejais IVT reakcija buvo vykdoma 80 μ l tūryje, naudojant 20 μ l AGS reakcijos produkto ir laikantis gamintojo rekomendacijų. Skirtingų IVT rinkinių efektyvumo tyrimai buvo atliekami analizuojant pavienių ląstelių transkriptomus scRNA-Seq platformoje pagal aprašytą metodiką.

K562 ląstelių kultivavimas ir pavienių ląstelių barkodavimas

Darbo metu buvo naudojama komerciškai prieinama K562 ląstelių linija (ATCC, kat. nr. CCL-243). Ląstelės buvo auginamos 25cm² plokščiadugnėse kolbelėse (Corning, cat. no. 430639), naudojant IMDM terpę (Gibco, cat. no. 12440053), papildytą 10% FBS (Gibco, cat. no. 10270-106) bei 1x Pen-Strep (Gibco, cat. no. 15140122). Darbo metu ląstelės buvo kultivuojamos, kol pasiekia 80% tankį (vidutiniškai 2-3 dienas) ir skiedžiamos santykiu 1:6, atlikus ląstelių centrifugavimą 5 min 500g ir perkeltiant į šviežią terpę.

Atliekant scRNA-Seq eksperimentus, K562 ląstelės buvo auginamos, kol pasiekė 70% tankį. Toliau buvo vertinamas ląstelių gyvybingumas naudojant tripano mėlio dažą. Visų eksperimentų metu buvo naudojami ląstelių mėginiai, kurių gyvybingumas viršijo 95%. Prieš atliekant pavienių ląstelių transkriptomų barkodavimą, ląstelės buvo tris kartus plaunamos 5ml 1x PBS (Gibco, cat. no. 20012027) ir skiedžiamos iki 140 000 ląstelių/ml galutinės koncentracijos. Paruošta ląstelių suspensija buvo praturtinama 16% (v/v) Optiprep (Sigma-Aldrich, cat. no. D1556) ir 0,05% (w/v) BSA (Carl Roth, cat. no. 8076.2) ir naudojama sekoskaitos bibliotekos optimizavimo eksperimentuose aprašytuose aukščiau. Viso proceso metu ląstelių mėginiai buvo laikomi +4°C temperatūroje.

HMLE ląstelių kultivavimas ir EMT virsmo indukavimas

Darbo metu buvo naudojamos HMLE ląstelių linijos gautos iš prof. Robert Weinberg laboratorijos (MIT, JAV). Visos HMLE ląstelių linijos buvo auginamos 10cm skersmens lėkštelėse (Corning, cat. no. 430167), naudojant MEGM terpę (Lonza, cat. no. CC-3051). Darbo metu ląstelės buvo kultivuojamos, kol pasiekė 80% tankį (vidutiniškai 2-3 dienas) ir skiedžiamos santykiu 1:7. Visais atvejais ląstelės disocijuojamos nuo paviršiaus naudojant TrypLE Express reagentą (Gibco, cat. no. 12604013), laikantis gamintojo nuorodų.

EMT virsmas buvo indukuojamas ląstelių augimo terpę praturtinant 5ng/ml TGF- β 1 (PeproTech, cat. no. 100-21). Taip pat EMT virsmas buvo indukuojamas, naudojant HMLE ląstelių liniją, pasižyminčią padidinta *ZEB1* geno raiška. *ZEB1* geno raiška indukuojama augimo terpę praturtinant 1 μ g/ml doksiciklinu (Sigma, cat. no. D3447).

EMT mėginių pavienių ląstelių barkodavimas

Atliekant scRNA-Seq eksperimentus HMLE ląstelės buvo auginamos, kol pasiekė 70% tankį. Toliau buvo vertinamas ląstelių gyvybingumas naudojant tripano mėlio dažą. Visų eksperimentų metu buvo barkoduojami ląstelių mėginiai, kurių gyvybingumas viršijo 90%. Prieš atliekant pavienių ląstelių transkriptomų barkodavimą, ląstelės buvo tris kartus plaunamos 5ml 1x PBS (Gibco, cat. no. 20012027) ir skiedžiamos iki 140 000 ląstelių/ml galutinės koncentracijos. Toliau, paruošta ląstelių suspensija praturtinama 16% (v/v) Optiprep (Sigma-Aldrich, cat. no. D1556) ir 0,05% (w/v) BSA (Carl Roth, cat. no. 8076.2) ir pavienių ląstelių transkriptomai barkoduojami aprašyta metodika. Viso proceso metu ląstelių mėginiai buvo laikomi +4°C. Barkodavimo reakcijos detalės taip pat pateikiamos publikacijoje disertacijos tema [21].

Krūties vėžio pacientų mėginių pavienių ląstelių barkodavimas

Krūties vėžio pacientų mėginiai renkami pirminės operacijos metu, laikantis etikos standartų. Mėginiai buvo disocijuojami pagal poreikį ir, naudojant tėkmės citometrijos prietaisus, atrenkamos imuninės ląstelės pagal CD45 žymenį. Prieš atliekant pavienių ląstelių transkriptomų barkodavimą, ląstelės buvo du kartus plaunamos 1ml 1x PBS (Gibco, cat. no. 20012027) ir skiedžiamos iki 140 000 ląstelių/ml galutinės koncentracijos. Toliau, paruošta ląstelių suspensija praturtinama 16% (v/v) Optiprep (Sigma-Aldrich, cat. no. D1556) ir 0,05% (w/v) BSA (Carl Roth, cat. no. 8076.2) ir pavienių ląstelių transkriptomai barkoduojami aprašyta metodika. Viso proceso metu ląstelių mėginiai buvo laikomi +4°C. Barkodavimo reakcijos detalės taip pat pateikiamos publikacijoje disertacijos tema [20].

REZULTATAI IR JŲ APTARIMAS

Disertacijoje pristatomi rezultatai yra suskirstyti į tris tematinės dalis. Pirmoje dalyje yra pristatomi inDrops metodo tyrimai bei protokolo optimizacijos, leidusios pasiekti ženklų efektyvumo padidinimą. Antroje dalyje yra nagrinėjamas EMT virsmas HMLE (krūties vėžio ląstelės) modelinėje sistemoje bei aptariami ZEB1 transkripcijos faktoriaus taikinių tyrimai. Trečioje rezultatų dalyje yra pristatomas krūties vėžio pacientų imuninių ląstelių atlasas ir navikuose nustatomų imuninių ląstelių fenotipų tyrimai. Kiekvienos dalies pabaigoje yra pateikiamas rezultatų aptarimas.

Pavienujų ląstelių transkriptomato analizės metodikos optimizavimas

Pavienujų ląstelių transkriptomato analizės metodas susideda iš daugelio atskirų žingsnių. Juos galima sugrupuoti į keturias atskiras dalis: ląstelių mėginio paruošimas, pavienujų ląstelių transkriptomato barkodavimas, sekoskaitos bibliotekos paruošimas ir duomenų analizė. Kiekviena iš šių dalių turi bent po keletą skirtingų techninių žingsnių, nuo kurių efektyvumo priklauso ir bendras viso proceso efektyvumas. Šiame darbe buvo tiriami pavienujų ląstelių transkriptomato barkodavimo bei sekoskaitos bibliotekos paruošimo etapai. Svarbu paminėti, jog šie etapai sudaro techninį metodikos branduolį ir nekinta analizuojant skirtingus mėginius bei tyrinėjant skirtingus biologinius aspektus. Dėl to šias dvi metodines dalis galima tirti nepriklausomai nuo mėginio tipo ar duomenų analizės pobūdžio.

Šio tyrimo pradžioje scRNA-Seq metodika buvo nauja technologija ir detalūs bei validuoti protokolai nebuvo prieinami mokslinei bendruomenei. Dėl šios priežasties pirma darbo dalis buvo skirta technologijos validavimui ir detalaus protokolo parengimui. Šio darbo rezultatas yra tikslios metodinės gairės ląstelių mikroskysčių technologija paremtai aukšto našumo pavienujų ląstelių transkriptomato

analizei atlikti. Šios gairės buvo pristatytos mokslinei bendruomenei [19] ir jų pagrindu buvo vykdomi tolimesni metodikos optimizavimo tyrimai.

Vykdytų tyrimų tikslas buvo padidinti iš pavienių ląstelių nustatomų transkriptų kiekį. Darbo metu analizuoti pavienių ląstelių transkriptomo barkodavimo bei sekoskaitos bibliotekos paruošimo etapai kartu apima septynias atskiras molekulinės biologijos reakcijas, kurios sudaro nuoseklų procesą. Jo pradžioje pavienės ląstelės mikroskysčių lusto pagalba yra kompartmentalizuojamos mikroskysčių lašeliuose. Juose ląstelės yra lizuojamos ir jų mRNR molekulės (transkriptomas) yra barkoduojamos AT reakcijos pagalba. Šios reakcijos produktas yra barkoduota kDNR, kuri yra apjungiamą (pvz.: iš 3000 pavienių ląstelių), valoma ir dauginama. Tokia padauginta biblioteka yra paruošiama sekoskaitai atliekant tris papildomas molekulinės biologijos reakcijas. Darbo metu buvo postuluojuama, jog neefektyvios molekulinės biologijos reakcijos, atliekamos prieš bibliotekos padauginimą, tiesiogiai daro įtaką iš pavienių ląstelių nustatomų transkriptų kiekiui. Kiekviena molekulė, prarasta iki bibliotekos padauginimo, yra prarandama ir galutiniuose duomenyse. Priešingai, po bibliotekos padauginimo patiriami praradimai turi ženkliai mažesnį poveikį duomenų praradimui, nes kiekvieno unikalaus transkripto yra bent po keletą kopijų. Atsižvelgiant į šią prielaidą darbo metu buvo optimizuojamos tik tos molekulinės biologijos reakcijos, kurios yra atliekamos iki bibliotekos padauginimo įtraukiant ir pastarąjį žingsnį: i) AT reakcija lašeliuose, ii) kDNR valymas, iii) AGS reakcija ir iv) bibliotekos padauginimo – *in vitro* transkripcijos (IVT) reakcija.

Darbo metu buvo naudojamos trys skirtingos metodikos minėtų reakcijų efektyvumui įvertinti. Transkriptų kiekis bibliotekoje iki padauginimo buvo vertinamas kiekybinės PGR (kPGR) metodu. Siekiant kuo tiksliau įvertinti bendrą transkriptų kiekį, buvo analizuojami aštuonių skirtingų genų, pasižyminčių skirtingais ekspresijos profiliais, transkriptai (4 lentelė). Taip pat buvo

vertinamas padaugintos bibliotekos fragmentų kiekis ir jų dydžio pasiskirstymas. Galutinės išvados buvo priimtos, įvertinus konkrečių tiriamų sąlygų įtaką transkriptų kiekio nustatymui naudojant inDrops scRNA-Seq platformą. Atlikus skirtingomis sąlygomis paruoštų bibliotekų sekoskaitą buvo vykdoma pirminė duomenų analizė ir lyginamas vidutinis nustatomas transkriptų skaičius pavienėse ląstelėse. Siekiant išlaikyti kuo pastovesnes eksperimento sąlygas buvo naudojamos komerciškai prieinama ląstelių linija – K562 (žmogaus mielogeninės limfomos ląstelės).

$$\text{Sekoskaitos įsisotinimas} = 100 * \left(1 - \frac{\text{unikalios molekulės}}{\text{kokybiški nuskaitymai}}\right)$$

1 pav. Sekoskaitos įsisotinimo įverčio formulė. Unikaliomis molekulėmis yra laikoma unikali ląstelės barkodo, molekulinio žymens bei transkripto kombinacija. Kokybiški nuskaitymai yra laikomi tie, kurie turi tinkamą unikali ląstelės barkodo, molekulinio žymens bei transkripto seką.

Svarbu atkreipti dėmesį, jog iš pavienių ląstelių nustatomų transkriptų kiekis tiesiogiai priklauso nuo sekoskaitos nuskaitymų kiekio, tekusio konkrečiai ląstelei. Literatūros duomenys rodo, jog sekoskaitos įsisotinimas yra pasiekiamas esant 500 000 – 1 000 000 nuskaitymų kiekvienai individualiai ląstelei [26]. Nepasiekus šio skaičiaus, nustatomų transkriptų kiekis tiesiogiai priklausys nuo sekoskaitos gylio. Tačiau žvelgiant iš ekonominės pusės yra netikslinga atlikti daug nuskaitymų nuo kiekvienos ląstelės, kadangi tokiu atveju labai ženkliai išauga tyrimo kaina. Šiame darbe yra naudojamas „sekoskaitos įsisotinimo“ parametras, leidžiantis kompensuoti sekoskaitos gylio skirtumus duomenyse. Sekoskaitos įsisotinimas yra skaičiuojamas nustatant unikalų molekulių ir visų kokybiškų sekoskaitos nuskaitymų santykis. Šis skaičius po to verčiamas į procentinę išraišką (1 pav.). Dvi tiriamos sąlygos gali būti tiesiogiai lyginamos tarpusavyje, jeigu jos turi tą patį sekoskaitos įsisotinimo įvertį. Tais atvejais, kai lyginamų mėginių įverčiai nesutapo, dalis duomenų iš mėginio su aukštesniu įverčiu buvo

pašalinama atsitiktiniu būdu. Siekiant ekonominio naudingumo scRNA-Seq optimizacijos tyrimams naudoti žemi sekoskaitos įsisotinimai (<20%).

Lentelė 5. kDNR sintezės mikroskysčių lašeliuose reakcijos efektyvumo optimizavimo tyrimų rezultatai. Metodinės sąlygos: MH – Maxima H- AT fermentas, SS – kDNR valymas fragmentų dydžio atrinkimu, NN – NEBNext AGS rinkinys, HS – HiScribe IVT rinkinys.

Tiriamos sąlygos	Vidutinis transkriptų skaičius	Sekoskaitos įsisotinimas	Metodinės sąlygos
SuperScript III	952	11.79 %	SS, NN, HS
SuperScript IV	2220	11.84 %	SS, NN, HS
Maxima H-	3304	10.68 %	SS, NN, HS

Darbo rezultatai yra pristatomi nuosekliai, pradedant pavienių ląstelių transkriptomo barkodavimo žingsniu. Šio etapo metu naudojant mikroskysčių lustą yra apjungiami trys atskiri reakcijos komponentai (pavienės ląstelės, barkoduojantys hidrogeliniai rutuliukai ir AT-lizės reagentai) ir formuojami mikroskysčių lašeliai. Darbo metu buvo optimizuojama kDNR sintezės reakcija, vykstanti susidariusiame lašelyje. Tam tikslui buvo vertinami trys skirtingi AT fermentai. Pradinės sąlygos, kurios yra laikomos atskaitinėmis, yra SuperScript III AT fermentas. AT fermentų efektyvumo tyrimo scRNA-Seq platformoje rezultatai yra pateikiami 5-oje lentelėje ir atskleidžia, jog tirtomis sąlygomis Maxima H- AT fermentas yra 3,4 karto efektyvesnis. Toks ženklus skirtumas tarp skirtingų fermentų gali būti paaiškinamas skirtingomis jų aminorūgščių sekomis. Visi trys fermentai yra dirbtinės evoliucijos būdu patobulintos M-MuLV AT fermento versijos. Įdomu tai, jog pagal gamintojo nurodomą informaciją SuperScript IV fermentas yra pats pažangiausias iš visų trijų pagal termostabilumo bei atsparumo slopikliams savybes. Tačiau šiame darbe pateikiami rezultatai parodo, jog vykdant kDNR sintezę

lašeliuose naudojant pavienes ląsteles gamintojo nurodomi fermento privalumai nebūtinai yra aktualūs. Be to, svarbu paminėti, jog Maxima H- fermentas yra pigiausias iš visų trijų tirtų fermentų.

Toliau pristatomi atskirų sekoskaitos bibliotekos ruošimo reakcijų optimizavimo eksperimentų rezultatai. Šiame darbe buvo tiriami trys atskiri proceso žingsniai: kDNR valymas, antros grandinės sintezės reakcija ir bibliotekos padauginimas *in vitro* transkripcijos reakcijos pagalba. Pradinės sąlygos, kurios yra laikomos atskaitinėmis, yra kDNR valymas DNR nukleazių pagalba, AGS reakcija naudojant NEBNext komercinį rinkinį ir IVT reakcija naudojant HiScribe komercinį rinkinį.

Darbo metu buvo vertinami du kDNR valymo metodai. Pirmuoju atveju AT reakcijos metu nepanaudoti pradmenys ir susidarę pradmenų dimerai yra pašalinami veikiant DNR nukleazėmis. Antruoju atveju kDNR yra valoma atrenkant fragmentus pagal jų dydį – naudojant AMPure reagentą. Darbo metu nustatyta, jog naudojant AMPure reagento ir AT reakcijos mišinio 0,8x tūrio santykį yra efektyviai pašalinami nepanaudoti pradmenys ir susidarę pradmenų dimerai neprarandant analizės jautrumo. Siekiant palyginti dvi skirtingas kDNR valymo metodikas buvo atlikti scRNA-Seq eksperimentai. 6-oje lentelėje pateikti tyrimo rezultatai atskleidžia, jog kDNR valymo metodika, atrenkant fragmentus pagal jų dydį, yra 3 kartus efektyvesnė iš pavienių ląstelių nustatomų transkriptų atžvilgiu. Tokie rezultatai leidžia manyti, jog dėl DNR nukleazių poveikio yra dalinai suardoma ir barkoduota kDNR iš pavienių ląstelių. Svarbu paminėti, jog tyrimo eigoje šis žingsnis optimizuotas pirmiausia. Atitinkamai, visi kiti scRNA-Seq optimizavimo eksperimentai buvo atlikti taikant kDNR valymo metodiką atrenkant fragmentų dydžius.

Kitu etapu buvo vertinami du komerciškai prieinami rinkiniai AGS reakcijai atlikti. Šios reakcijos metu yra naudojamas trijų skirtingų fermentų mišinys. RNR:kDNR duplekso RNR molekulė yra fragmentuojama naudojant RNAzė H. Susidarę viengrandiniai trūkiausiai leidžia DNR polimerazei I inicijuoti antros DNR grandinės sintezę.

Reakcijos pabaigoje po DNR sintezės likę viengrandiniai trūkliai yra ištaisomi T4 DNR ligazės pagalba. Tokios reakcijos metu susidaro dvigrandinė DNR, kuri gali būti naudojama kaip substratas IVT reakcijos metu. Šiame darbe buvo vertinami du komerciškai prieinami rinkiniai AGS reakcijai atlikti. Jų efektyvumas buvo lyginamas atliekant scRNA-Seq eksperimentus. 6-oje lentelėje pateikti rezultatai atskleidžia, jog SuperScript rinkinys yra 1,2 karto efektyvesnis. Svarbu paminėti, jog šis rinkinys yra tiekiamas atskirais reakcijos komponentais ir vartotojas turi atlikti daugiau techninio darbo reakcijai paruošti. Priešingai, NEBNext rinkinys yra pateikiamas paruoštas darbui. Dėl šios priežasties vartotojui reikia atlikti mažiau techninių žingsnių, kas pagreitina procesą ir sumažina klaidų tikimybę.

Paskutiniame darbo etape buvo vertinami du komerciniai rinkiniai, skirti IVT reakcijai atlikti. Šios reakcijos metu barkoduota pavienių ląstelių kDNR gali būti tiesiškai padauginama T7 RNR polimerazės pagalba dėl barkodavimo pradmenyse esančios T7 promotoriaus sekos. Kaip ir prieš tai aprašytų tyrimų atveju, rinkiniai buvo lyginami atliekant scRNA-Seq eksperimentus. Tyrimo rezultatai, pateikti 6-oje lentelėje, leidžia teigti, jog HiScribe rinkinys yra efektyvesnis už TranscriptAid rinkinį. Dėl žemo sekoskaitos įsisotinimo negalima tiksliai įvertinti šių rinkinių efektyvumo skirtumų. Papildomi eksperimentai nebuvo atlikti, kadangi HiScribe rinkinys buvo naudojamas atskaitinėse reakcijos sąlygose.

Lentelė 6. Sekoskaitos bibliotekos paruošimo žingsnių efektyvumo optimizavimo tyrimų rezultatai. Metodinės sąlygos: SSIII – SuperScript III AT fermentas, SS – kDNR valymas fragmentų dydžio atrinkimu, NN – NEBNext AGS rinkinys, HS – HiScribe IVT rinkinys.

Tiriamos sąlygos	Vidutinis transkriptų skaičių	Sekoskaitos įsisotinimas	Metodinės sąlygos
kDNR valymo metodika			
DNR nukleazių pagalba	312	11.46 %	SSIII, NN, HS
Atrenkant fragmentų dydžius	952	11.79 %	SSIII, NN, HS
Antros grandinės sintezės reakcija			
NEBNext	1392	22.77 %	SSIII, SS, HS
SuperScript	1689	22.91 %	SSIII, SS, HS
Bibliotekos padauginimo IVT reakcija			
HiScribe	434	4.13 %	SSIII, SS, NN
TranscriptAid	388	4.58 %	SSIII, SS, NN

Šioje disertacijos dalyje pristatomi inDrops scRNA-Seq metodo tyrimai. Tikslių metodinių gairių lašelių mikroskysčių technologija paremtai aukšto našumo pavienių ląstelių transkriptomo analizei atlikti pristatymas mokslinei bendruomenei ženkliai prisidėjo prie pavienių ląstelių transkriptomo analizės tyrimų prieinamumo. Po metodinių gairių parengimo atlikti metodikos optimizavimo tyrimai leido iki 10-ies kartų padidinti transkriptų kiekį, nustatomą iš pavienių ląstelių. Pradinėmis sąlygomis yra laikomas „DNR nukleazių pagalba“ mėginys 6 lentelėje, o „Maxima H-“ mėginys 5 lentelėje atitinka optimizuotas sąlygas. Svarbu paminėti, jog nei vienu atveju, atliekant metodikos optimizavimo tyrimus, nebuvo pasiektas

sekoskaitos įsisotinimas. Atitinkamai, šiame tyrime pristatomi rezultatai gali būti lyginami su literatūroje pristatomais duomenimis tik atsižvelgiant į sekoskaitos įsisotinimo įvertį. Dėl darbo metu naudotų santykinai žemų sekoskaitos įsisotinimo lygių, nustatyti transkriptų kiekiai dažnu atveju nėra tiesiogiai prilyginami literatūroje pateikiamiems panašių scRNA-Seq eksperimentų duomenims. Taip pat svarbu atkreipti dėmesį, jog optimizavimo tyrimams atlikti buvo naudojama kultivuojama ląstelių linija (K562). Tai leido užtikrinti žemą biologinio mėginio variabilumą. Iš kitos pusės, šios ląstelės pasižymi aukštu populiacijos vientisumu. Todėl nebuvo galima įvertinti, ar padidėjęs metodo efektyvumas daro įtaką ląstelių populiacijų nustatymui.

Nepriklausomi scRNA-Seq metodų palyginimo tyrimai yra parodę, jog inDrops metodas pasižymi žemu efektyvumu, lyginant transkriptų kiekį, nustatomą iš pavienių ląstelių [27, 28]. Tačiau šiuose palyginimuose buvo naudojama neoptimizuota protokolo versija. Disertacijoje pristatomi inDrops metodo patobulinimai ženkliai padidina metodo efektyvumą, todėl galima pagrįstai manyti, jog tai turėtų ženkliai įtaką analogiškų scRNA-Seq metodų palyginimų rezultatams. Vis dėlto, reikia atkreipti dėmesį į tai, jog tų pačių palyginimų metu buvo parodyta, jog inDrops metodas nenusileidžia kitiems scRNA-Seq metodams lyginant ląstelių tipų nustatymo jautrumą. Nepriklausomai nuo santykinai žemo aptinkamų transkriptų kiekio, inDrops metodu surinkti duomenys leido identifikuoti tuos pačius ląstelių tipus, kaip ir kitomis metodikomis atlikti analogiški eksperimentai. Toks rezultatas verčia pasvarstyti apie nustatomų transkriptų kiekio svarbą biologinių tyrimų kontekste. Apibendrinus literatūros duomenis galima pagrįstai teigti, jog žemo našumo scRNA-Seq metodai pasižymi aukštesniu efektyvumu ir leidžia nustatyti didesnę kiekį transkriptų iš pavienių ląstelių. Tačiau dėl mažo ištiriama ląstelių kiekio, šie metodai neleidžia aptikti retų ląstelių tipų. Pavyzdžiui, jeigu tam tikras tipo ląstelės sudaro 1% visų mėginyje esančių ląstelių – analizuojant šimtą pavienių ląstelių tikimybė aptikti

šià populiacijà yra lygi 0,63. Priešingai, aukšto našumo metodai nepasižymi dideliu nustatomų transkriptų kiekių. Tačiau jis yra pakankamas ląstelių tipų identifikavimui, o dėka didelio analizuojamo ląstelių kiekio galima nesunkiai aptikti net itin retas ląstelių populiacijas. Aukštas nustatomų transkriptų kiekis yra svarbus, kai norima patikimai analizuoti genus, pasižyminčius žemu raiškos lygiu. Tačiau tai yra aktualu ne visų tyrimų strategijose ir dažnu atveju analizuojamų ląstelių kiekis yra svarbesnis analizuojant sudėtingas biologines sistemas nei absoliutus metodo jautrumas. Todėl galima teigti, jog aukšto našumo pavienių ląstelių transkriptomo analizės platformos yra labiau patrauklios sudėtingų biologinių sistemų tyrimams. Disertacijoje pristatomos inDrops metodo optimizacijos ženkliai padidina metodo efektyvumą ir tuo pačiu daro jį patraukliu bendrame scRNA-Seq tyrimų srities kontekste.

Šio darbo metu nustatyta, jog kritiniai proceso etapai, kurių optimizacijos lėmė didžiausią efektyvumo padidėjimą, yra AT reakcija mikroskysčių lašeliuose ir kDNR valymo žingsnis. Tačiau vertinant darbo rezultatus reikia atsižvelgti ir į techninius atlikto tyrimo apribojimus. Darbo nebuvo išanalizuoti visi rinkoje prieinami komerciniai reagentai. Pavyzdžiui, dėl ekonominių apribojimų dalis komerciškai prieinamų AT fermentų nebuvo įtraukti į tyrimą. Taigi, galima manyti, jog darbo metu identifikuoti reagentai ne visais atvejais yra optimaliausi. Taip pat, visos reakcijos buvo atliekamos laikantis gamintojo rekomendacijų. Tokios reakcijos sąlygos nebūtinai sutampa su optimaliomis. Pavyzdžiui, kitos mokslininkų grupės atlikti scRNA-Seq metodikos optimizavimo tyrimai taip pat parodė, jog AGS reakcija gali būti optimizuota [29]. Įdomu yra tai, jog aukštesniam efektyvumui pasiekti reikėjo ženkliai praskiesti gamintojo tiekiamus reagentus. Tokie rezultatai leidžia matyti, jog papildomai optimizuojant reakcijos sąlygas galima būtų pasiekti dar didesnę metodo efektyvumą.

Disertacijoje nagrinėjamas InDrops metodo sekoskaitos bibliotekos padauginimas yra paremtas *in vitro* transkripcijos reakcija.

Tačiau tai yra ne vienintelė bibliotekos padauginimo strategija. Alternatyvūs scRNA-Seq metodai bibliotekos padauginimui yra pagrįsti AT fermento savybe vadinama „substrato keitimu“ (*angl.* template switching). Dėka šios savybės, AT fermentas gali pridėti papildomą seką kiekvienos kDNR molekulės gale, kuri bus bendra visoms reakcijos metu susintetintoms molekulėms. Tokia pavienių ląstelių transkriptomo biblioteka gali būti padauginama PGR pagalba. Ši strategija nereikalauja AGS reakcijos bei ilgai užtrunkančios IVT reakcijos. Todėl vienas akivaizdus privalumas yra sutrumpėjęs bibliotekos ruošimo laikas. Naudojant substrato keitimu paremtą sekoskaitos bibliotekos paruošimo strategiją, biblioteką galima paruošti per vieną dieną. Tuo tarpu naudojant šiame darbe aprašytą metodiką procesas užtrunka dvi dienas. Šis aspektas nėra svarbus, kai reikia analizuoti keletą mėginių, tačiau didėjant tyrimo apimtims laiko sutaupymas gali tapti svarbus. Daryti išvadų apie šių skirtingų metodikų efektyvumo skirtumus negalima, kadangi iki šiol nėra atliktas sisteminis metodikų palyginimas.

Apibendrinant, šioje disertacijos dalyje pristatomi tyrimai atskleidžia, jog, optimizuojant atskirus inDrops metodo sekoskaitos bibliotekos paruošimo žingsnius, galima iki 10-ies kartų padidinti transkriptų kiekį, nustatomą iš pavienių ląstelių. Šis efektyvumo padidinimas yra aktualus atsižvelgiant į nepriklausomų tyrimų metu nustatytus inDrops metodikos trūkumus. Ateityje platforma galėtų būti papildomai tobulinama optimizuojant atskirų molekulinės biologijos reakcijų sąlygas.

EMT virsmo tyrimas

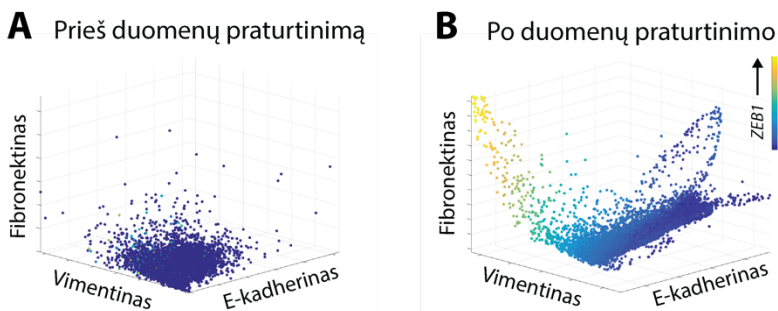
Šioje disertacijos dalyje yra pristatomi EMT virsmo tyrimo rezultatai, kurie yra taip pat aprašyti disertacijos tema parengtoje publikacijoje [21]. Vykdytus tyrimus galima padalinti į dvi dalis. Pirmosios dalies tyrimų tikslas buvo charakterizuoti EMT procesą imortalizuotų krūties vėžio ląstelių (HMLE) modelinėje sistemoje

naudojant scRNA-Seq technologiją. Antroje darbo dalyje buvo siekiama nustatyti ir validuoti kartinio EMT proceso transkripcijos faktorius – ZEB1 aktyvacijos taikinius.

EMT virsmas yra nuoseklus ląstelių būsenos pasikeitimas, kurio metu epitelinės ląstelės palaipsniui tampa mezenchiminėmis. Šio proceso metu keičiasi ir ląstelių žymenys. Literatūros duomenimis epitelinės ląstelės gali būti atpažįstamos pagal E-kadherino žymenį, o mezenchiminės – pagal vimentiną, fibronektiną ir N-kadheriną. Taip pat yra žinoma, jog transkriptomo pokyčius virsmo metu koordinuoja keli kartiniai transkripcijos faktoriai: SNAIL, SLUG, TWIST1, ZEB1 bei ZEB2. Dauguma tyrimų, atliktų EMT srityje, lygino pradinę ir galutinę virsmo būseną, neskirdami dėmesio tarpinėms proceso būsenoms. Be to, beveik visi tyrimai buvo atlikti ląstelių populiacijų lygmeniu (*angl.* bulk). Todėl apie atskiras tarpines proceso būsenas bei proceso dinamiką laiko atžvilgiu šiandien yra žinoma vis dar santykinai nedaug.

Disertacijoje pateikiamų tyrimų metu buvo naudojama HMLE ląstelių modelinė sistema, o EMT virsmas buvo indukuojamas stimuliuojant ląsteles TGF- β signaline molekule. Toks modelis pasirinktas siekiant atkartoti žmogaus organizme vykstančius procesus. EMT srityje atliktų tyrimų dėka yra žinoma, jog TGF- β yra viena iš pagrindinių signalinių molekulių, kontroliuojančių EMT virsmą. Darbo metu EMT virsmas buvo stebimas dvylikos dienų periode. Kas dvi dienas buvo imami ląstelių mėginiai ir pavienės ląstelės analizuojamos inDrops metodu. Šio tyrimo rezultatai parodė, jog EMT procesas yra asinchroniškas – skirtingos ląstelės EMT virsmą patiria skirtingu greičiu. Nustatyta, jog didžiausia ląstelių būsenų įvairovė yra stebima 8 ir 10 dienomis. Šiais laiko momentais atskiros ląstelės užima visą EMT virsmo spektrą (2 pav. B). Atsižvelgiant į tokius rezultatus, tolimesnė analizė vykdyta naudojant duomenų masyvą, susidedantį iš 7523 pavienių ląstelių, išanalizuotų 8 ir 10 dienomis. Taip pat svarbu atkreipti dėmesį į tai, jog norint tirti EMT procesą reikėjo pritaikyti duomenų praturtinimo algoritmą (2

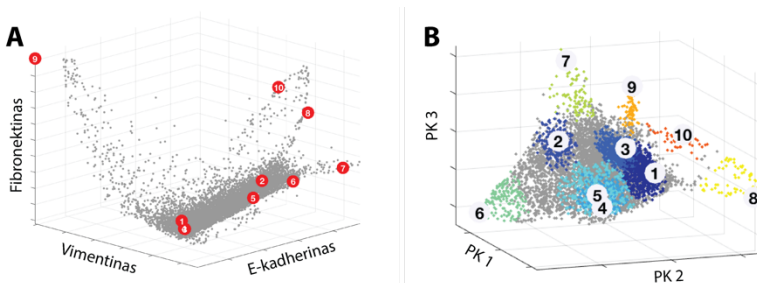
pav.). Šiame darbe duomenų praturtinimui buvo taikomas duomenų difuzija paremtas algoritmas – MAGIC (Markov affinity-based graph imputation of cells). Visos toliau patiekiamos analizės buvo atliktos su praturtintais duomenimis nebent yra nurodoma kitaip.



2 Pav. Ląstelių išsidėstymas pagal fibronektino, vimentino ir E-kadherino genų raišką. Kiekvienas taškas atvaizduoja individualią ląstelę. A – prieš duomenų praturtinimą, B – po duomenų praturtinimo; nuspalvinta pagal *ZEB1* geno raišką.

Atlikto pavienių ląstelių tyrimo rezultatai sutampa su literatūroje aprašytais tyrimais. EMT virsmas yra charakterizuojamas E-kadherino (epitelinų ląstelių žymens) raiškos sumažėjimo ir atitinkamo vimentino bei fibronektino (mezenchiminių ląstelių žymenys) raiškos padidėjimo. Mezenchiminių ląstelių būseną pagal žymenis sutampa su stebima *ZEB1* geno, kuris yra kritinis EMT virsme, raiška (2 pav. B). Tolesniu etapu buvo siekiama detaliau charakterizuoti EMT virsmą. Analizė parodė, jog net 79% ląstelių yra tarpinėje būsenoje ir nėra nei epitelinės, nei mezenchiminės. Literatūroje tokia ląstelių būseną vadinama daline EMT būseną. Yra žinoma, jog šios ląstelės yra svarbios įvairiuose biologiniuose procesuose, tame tarpe ir vėžinio naviko vystymuisi. Šios tarpinės būsenos ląstelės yra tarpusavyje heterogeniškos ir gali būti sugrupuojamos pagal savo fenotipą. Siekiant nustatyti konkrečias ląstelių būsenas buvo atliekama archetipinė analizė, kuri leido

identifikuoti 10 ląstelių archetipų (AR) (3 pav.). Kiekvienas iš jų sutampa su konkrečia ląstelių būsenos virsmo metu. Svarbu atkreipti dėmesį į tai, jog ne visas ląstelės galima patikimai priskirti tam tikram archetipui (3 pav. B, pilka spalva). Tikėtina, jog šios ląstelės yra aktyviame būsenos virsme.



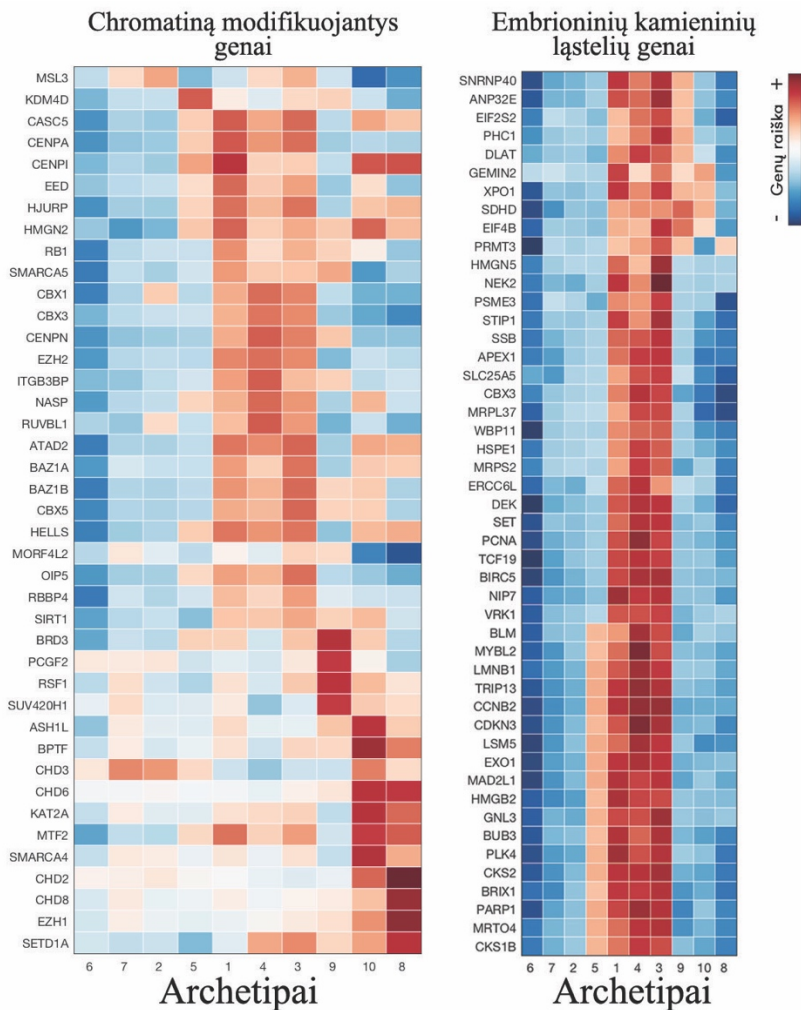
3 Pav. EMT proceso vizualizacija pagal genų fibronektino, vimentino ir E-kadherino raišką (A) arba pagal principines komponentes (B). Kiekvienas taškas atvaizduoja individualią ląstelę. A – raudoni taškai žymi archetipus, B – juodi numeriai žymi archetipus. Ląstelės nuspalvintos pagal archetipus, o pilkos ląstelės nepriskirtos jokiame archetipui.

Pagal ląstelių tipą identifikuoti archetipai gali būti skirstomi į: epitelines ląsteles (AR6 ir AR7), dalinį EMT (AR1-AR5), mezenchimines ląsteles (AR9) ir apoptotines ląstelės (AR8 ir AR10). Kiekvienas archetipas pasižymi konkrečiam ląstelių tipui būdinga genų raiška. Įdomu tai, jog epiteliniuose archetipuose buvo nustatyta transkripcijos faktoriaus SOX4 geno raiška. Šis genas yra svarbus TGF- β signalinio kelio veiksnys [30], todėl galima teigti, jog tiriamoje sistemoje pilnai epitelinių ląstelių nebuvo. Toks rezultatas nėra netikėtas, atsižvelgiant į tai, jog buvo analizuojamos ląstelės surinktos 8-tą ir 10-tą TGF- β stimuliacijos dienomis.

Darbo metu buvo nustatyti net penki archetipai siejami su daline EMT būsenos (AR1-AR5), o dalis ląstelių nebuvo priskirtos jokiame archetipui. Tokie rezultatai sutampa su neseniai pasiūlytu EMT proceso modeliu, teigiančiu, jog virsmo metu ląstelės pereina per

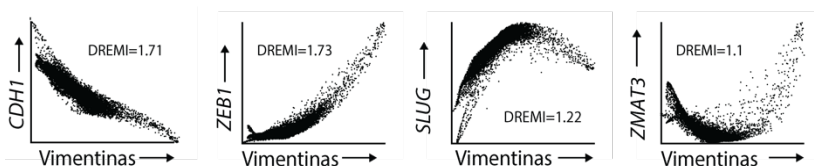
metastabilias būsenas [31]. Atlikus išsamią genų raiškos analizę pagal archetipus, nustatyta, jog tarpiniai ląstelių tipai pasižymi įvairių chromatiną modifikuojančių genų aktyvacija (4 pav. kairėje). Šie rezultatai leidžia manyti, jog EMT virsmo metu įvyksta ženklūs chromatinio persitvarkymai. Atitinkamai mezenchiminės ląstelės (AR9) taip pat yra charakterizuojamos savitu chromatiną modifikuojančių genų profiliu. Tokie rezultatai nėra netikėti, atsižvelgiant į didelio masto genų raiškos pokyčius, vykstančius EMT virsmo metu. Genų raiškos pokyčio analizės tyrimai taip pat atskleidė ženklų su kamieninėmis embrioninėmis ląstelėmis siejamų genų raiškos padidėjimą tarpinės būsenos ląstelėse (4 pav. dešinėje). Tai leidžia patvirtinti iki šiol diskutuojamą hipotezę, teigiančią, jog EMT virsmo metu epitelinės ląstelės dediferencijuojasi prieš įgydamos mezenchiminį fenotipą [32].

Aprašyta pradinė duomenų analizė parodo, jog scRNA-Seq technologija yra tinkama sudėtingų biologinių procesų analizei. Atlikus santykinai nedidelės apimties eksperimentus buvo galima atkurti kitų mokslininkų išvalgas, padarytas atliekant daugelį skirtingų tyrimų. Be to, žinomų biologinių signalų nustatymas duomenyse patvirtina analizės jautrumą ir patikimumą. Svarbus šio darbo dalies aspektas – nustatyta gausi tarpinių būsenų, besiskiriančių tarpusavyje, įvairovė. Tokie rezultatai leidžia manyti, jog EMT procesą kontroliuoja įvairūs tarpusavyje persipynę signaliniai keliai, nulemiantys ženklius genų raiškos pokyčius ir atskirų būsenų susiformavimą.



4 Pav. Genų raiškos pokyčių analizė pagal nustatytus archetipus. Pateikiama chromatiną modifikuojančių ir su embrioninėmis kamieninėmis ląstelėmis siejamų genų raiškos pokyčių analizė. Raudona spalva žymi padidėjusią raišką, mėlyna – sumažėjusią.

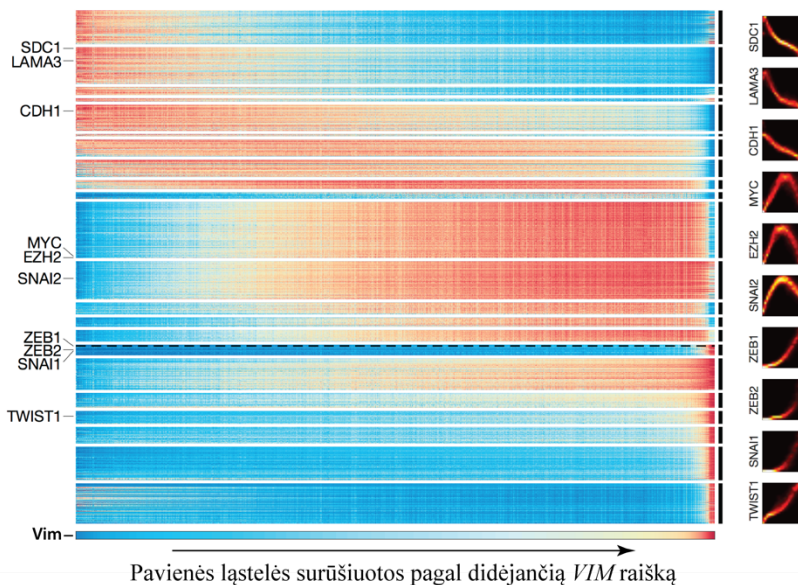
Tolesniame tyrimo etape buvo analizuojama genų raiškos pokyčių dinamika EMT virsmo metu. Tuo tikslu visos ląstelės buvo surūšiuotos pagal jose nustatytą vimentino geno raišką, kuri koreliuoja su virsmo stadija. Toks ląstelių surūšiuavimas gali būti prilyginamas surūšiuavimui EMT virsmo pseudolaikais. Ši analizė leidžia stebėti atskirų genų raiškos pokyčio dinamiką EMT proceso metu. Tiriamoje sistemoje buvo stebima keletas skirtingų raiškos pokyčio dinamikų (5 pav.). Pavyzdžiui, virsmo eigoje genų, susijusių su epiteliniu fenotipu, raiška nuosekliai mažėja (5 pav. *CDHI*), o genų, susijusių su mezenchiminiu fenotipu – nuosekliai didėja (5 pav. *ZEB1*). Taip pat, nemažos dalies genų raiška yra aktyvuojama arba slopinama virsmo viduryje (5 pav. *SLUG* ir *ZMAT3*). Tikėtina, jog šie genai yra susiję su tarpinėmis metastabiliomis dalinio EMT fenotipo būsenomis.



4 Pav. Genų raiškos pokyčio analizė pagal nustatytus archetipus. Pateikiama chromatiną modifikuojančių ir su embrioninėmis kamieninėmis ląstelėmis siejamų genų raiškos pokyčio analizė.

Siekiant sistemaiškai ištirti genų raiškos pokyčio dinamiką, buvo buvo skaičiuojamas tiriamų genų ir vimentino geno raiškos koreliacijos įvertis. Šiai analizei buvo pritaikytas DREMI algoritmas, leidžiantis įvertinti funkcinių dviejų genų sąryšį pavienių ląstelių duomenyse [33]. Atliekant genų raiškos dinamikos tyrimą, pirmiausia iš duomenų buvo pašalintos apoptuojančios ląstelės (pagal *MT-ND1* raišką). Toks žingsnis atliktas norint sukcentruoti analizę tik į EMT proceso dinamiką. Tuomet buvo skaičiuojamas individualus DREMI įvertis kiekvieno tyrime nustatyto geno ir vimentino geno raiškos koreliacijai aprašyti. Pagal nustatytą įvertį atrenkami genai, kurių raiškos pokytis buvo susijęs su EMT eiga ($DREMI > 0,5$). Šios

analizės metu, nustatyta, jog netikėta dauguma genų (2/3 visų genų) yra susiję su EMT virsmu. Pagal nurodytą įvertį tolimesnei analizei buvo atrinkti 13 487 genai, kurie buvo sugrupuoti į 22 atskiras grupes pagal koreliacijos su vimentino genu dinamiką. Genų grupės tuomet buvo išrūšiuotos pagal koreliacijos su vimentino geno raiška dinamiką, tokiu būdu sudarant globalų transkriptomą pokyčius EMT virsmo metu aprašanti žemėlapi (6 pav.). Analogiška analizė buvo pakartota su trimis papildomais genais, siejamais su mezenchiminiu fenotipu – *CHD2*, *ITGB4* ir *CD44*. Visi gauti žemėlapiai pasižymėjo aukštu tarpusavio atitikimu. Tokie rezultatai patvirtina, jog atlikta analizė yra patikima, o nustatytos genų raiškos pokyčio dinamikos atitinka tiriamoje sistemoje vykstančius virsmus.



6 Pav. Genų raiškos pokyčio dinamikos EMT virsmo metu žemėlapis sudarytas iš 13 487 atskirų genų. Genų grupės (Y ašyje) yra surūšiuotos pagal raiškos piką lyginant su vimentino genu raiška (X ašyje). Brūkšniuota linija žymi *ZEB1* geną. Dešinėje pateikiami reprezentatyvūs atskirų genų iš vimentino genu raiškos koreliacijos profiliai.

Sudarytas genų raiškos dinamikos žemėlapis atskleidžia, jog EMT proceso metu vyksta kardinalūs genų raiškos pokyčiai. Nustatytas genų grupes galima suskirstyti į tris tipus. Pirmojo tipo genų raiška nuosekliai mažėja EMT virsmo eigoje. Šie genai yra būdingi epiteliniam fenotipui – pavyzdžiui *SDC1* ir *LAMA3* genai, susiję su ląstelių adhezija. Antrojo tipo genų raiška padidėja prasidėjus virsmui ir sumažėja ląstelėms tampant mezenchiminėmis. Šie genai yra svarbūs dalinio EMT ląstelių fenotipo palaikymui. Tokių genų pavyzdžiai – *MYC* ir *EZH2*. Paskutinis, trečiasis genų tipas yra susijęs su mezenchiminiu fenotipu. Šių genų raiška palaipsniui didėja EMT virsmo metu. Tipiniai tokių genų pavyzdžiai yra kertiniai EMT virsmo transkripcijos faktoriai *ZEB1*, *TWIST* ir *SNAIL*. Atlikta analizė leidžia geriau suprasti EMT virsmo metu vykstančius genų raiškos pokyčius. Globalus genų surūšiavimas proceso pseudolaike leidžia daryti įžvalgas apie transkripcijos faktorių poveikį. Pavyzdžiui, tam tikro transkripcijos faktoriaus taikiniai turėtų būti aktyvuojami arba slopinami tik po konkretaus faktoriaus raiškos padidėjimo. Šią genų sąveiką analogiškai galima įvertinti DREMI algoritmo pagalba. Tokia analizė leistų nuspėti konkrečius transkripcijos faktoriaus taikinius neatliekant papildomų tiriamos biologinės sistemos modifikacijų. Paskutiniame EMT virsmo tyrimo etape buvo siekiama patvirtinti šią hipotezę.

$$\text{poveikio įvertis}(G) = \frac{1}{N} \sum_{j=1}^N \left(\sum_{i=1}^S r_z(g_i^j) - \sum_{i=1}^S r_t(g_i^j) \right)$$

7 Pav. Poveikio įverčio skaičiavimo formulė. N – atsitiktinių imčių skaičius; r – geno rangas. S – atsitiktinė genų imtis iš G genų rinkinio.

Tyrimui pasirinktas *ZEB1* transkripcijos faktorius, apie kurio taikinius yra žinoma santykinai nedaug, nepaisant jo svarbos EMT procese [34]. Remiantis aukščiau aprašytu žemėlapiu buvo identifikuoti 4509 genai, kurių raiška pasiekia maksimalią vertę kartu

su arba po *ZEB1* geno raiškos piko. Toliau iš šio genų rinkinio buvo atrinkti 1085 genai, kurių DREMI įvertis su *ZEB1* genu buvo ≥ 1 . Tikėtina, jog būtent šie genai yra tiesiogiai arba netiesiogiai aktyvuojami *ZEB1* transkripcijos faktoriaus. Siekiant pavirtinti šią hipotezę buvo atlikti papildomi eksperimentai. Jų metu, naudojant modifikuotą HMLE ląstelių liniją, sistemoje buvo tiesiogiai indukuojama *ZEB1* transkripcijos faktoriaus raiška. Tokioje sistemoje stebimas EMT skirsis nuo TGF- β indukuojamo EMT virsmo. Nepaisant to, dėl *ZEB1* geno raiškos padidėjimo bus stebimas ir *ZEB1* transkripcijos faktoriaus taikinių raiškos padidėjimas. Tad galima teigti, jog ši sistema yra tinkama išskeltos hipotezės patvirtinimui. Atliekant tyrimą, *ZEB1* geno raiška buvo indukuojama dvi dienas ir po to pavienės ląstelės buvo analizuojamos inDrops metodu. Siekiant patvirtinti *ZEB1* aktyvacijos taikinius buvo skaičiuojamas „poveikio įvertis“ (7 pav.). Šis dydis leidžia palyginti genų raiškos pokyčius tarp skirtingų eksperimentų (*ZEB1* raiškos indukcija ir TGF- β indukuojamas EMT). Visi genai kiekvieno eksperimento atveju yra išrūšiuojami pagal savo vidutinę raišką tarp visų pavienių ląstelių, taip suteikiant kiekvienam genui konkrečią poziciją raiškos eilėje (kitai tariant – rangą). Poveikio įvertis tuomet yra skaičiuojamas lyginant rinkinio G genų (nuspėti *ZEB1* aktyvacijos taikiniai) rangų sumų skirtumą tarp lyginamų eksperimentinių sąlygų. Siekiant atsižvelgti į didelį genų kiekį rinkinyje, užuot tiesiogiai susumavus genų iš rinkinio G rangus, kiekvienai sąlygai paskaičiuojamas vidutinis rangų sumų skirtumas fiksuotam genų skaičiui $S=200$, kuris $N=1000$ kartų atsitiktinai parenkamas iš rinkinio G. Aukštas poveikio įvertis reiškia, jog tiriamo genų rinkinio raiška yra padidėjusi *ZEB1* indukcijos sąlygomis lyginant su TGF- β indukcija. Siekiant įvertinti analizės patikimumą poveikio įvertis buvo skaičiuojamas analogiškai imant visų genų, dalyvaujančių EMT (DREMI su $VIM>0,5$), rinkinį. P-vertė yra dalis atsitiktinių genų imčių (S), kurių poveikio įvertis yra didesnis arba lygus nuspėto genų rinkinio (G) poveikio įverčiui.

Naudojant aprašytą metodiką nustatyta, jog tiesiogiai indukuojant *ZEB1* geną, nuspėtų *ZEB1* aktyvacijos taikinių genų rinkinio raiškos padidėjimas stebimas $P = 3,1e^{-73}$ patikimumu. Tokie rezultatai leidžia teigti, jog naudojant aprašytą metodiką galima tiksliai nuspėti tiesioginius ir netiesioginius transkripcijos faktorių aktyvacijos taikinius. Analogiška metodika įvertinus genų rinkinio, kurio raiškos padidėjimas pseudolaikė stebimas kartu arba po *ZEB1* geno raiškos (4509 genai), patikimumą nustatytas statistiškai patikimas, tačiau gerokai mažesnis patikimumas $P=0,004$. Toks rezultatas leidžia teigti, jog šio genų rinkinio aktyvacija priklauso ne tik nuo *ZEB1* transkripcijos faktoriaus, bet ir nuo papildomų signalinių kelių, aktyvių TGF- β indukuojamo EMT virsmo metu.

Tarp nuspėtų *ZEB1* aktyvacijos taikinių galima atrasti daugumą literatūroje jau aprašytų šio geno tiesioginių taikinių (*SNAIL*, *ZEB2*, *BMP*, *MMP3*). Aktyvuojami genai yra svarbūs citoskeleto persitvarkyme, užląstelinio matrikso pertvarkyme, ląstelės migracijos procese bei ląstelės ciklo pokyčiams. Tokie rezultatai sutampa su literatūroje aprašytais pokyčiais, vykstančiais EMT virsmo metu. Taip pat buvo nustatyta ir aktyvacijos taikinių, kurie iki šiol nebuvo aprašyti EMT virsmo kontekste. Pavyzdžiui *RHOA* genas, kuris dalyvauja aktino citoskeleto persitvarkyme. Vienas iš netikėtų aktyvacijos taikinių buvo genas *NTN4*, kuris yra aprašytas neuronų migracijos procese. Išvalgomis iki galo suprasti ir patvirtinti reikalingi papildomi tyrimai. Vis dėlto, apibendrinus atliktą analizę galima teigti, jog pavienių ląstelių transkriptomų tyrimai leidžia patikimai nuspėti transkripcijos faktorių aktyvacijos taikinius.

Šioje disertacijos dalyje pristatomi tyrimai parodo, jog scRNA-Seq technologija yra tinkama sudėtingų biologinių procesų tyrimams. Vienas iš tyrimo tikslų buvo nuspėti transkripcijos faktorių aktyvacijos taikinius. Šį uždavinį galima spręsti keletu skirtingų būdų. Disertacijoje pristatomi tyrimai leidžia nuspėti aktyvacijos taikinius, remiantis pseudolaikio analize be papildomų sistemos manipuliacijų. Genų taikinių tyrimus taip pat galima atlikti apjungiant CRISPR ir

scRNA-Seq technologijas [35, 36]. Taikant šią metodiką galima tiksliau ištirti ląstelėje veikiančius genų tinklus. Iš kitos pusės, tokie tyrimai remiasi genų inaktyvacija, kas gali turėti nenumatytų šalutinių pasekmių tiriamoje sistemoje ir gauti rezultatai nebūtinai atitiks *in vivo* procesus. Be to, ženklus disertacijoje pristatomo tyrimo privalumas yra ir tai, jog analizuojamoje sistemoje nereikia atlikti papildomų manipuliacijų (pvz.: genų inaktyvavimas). Dėl šios priežasties aprašyta metodika yra suderinama su klinikiniais pacientų mėginiais ir galėtų būti naudinga nustatant genų tinklų pokyčius, nulemiančius įvairias patologijas.

Šioje disertacijos dalyje pristatoma duomenų analizė rėmėsi duomenų praturtinimu. Darbo metu buvo taikomas MAGIC algoritmas, kurio vystymas nebuvo šio darbo dalis. Vis dėlto, yra svarbu aptarti naudoto duomenų praturtinimo algoritmo poveikį ir apribojimus. scRNA-Seq analizės duomenys pasižymi retumu. Tokia jų savybė yra susijusi su techniniais metodikos apribojimais. Dėl neefektyvaus proceso yra nuskaitoma tik dalis kiekvienos individualios ląstelės transkriptomo. Duomenų retumas nedaro didelės įtakos ląstelių tipų nustatymui ar populiacijų vizualizavimui, tačiau norint analizuoti kelių konkrečių genų tarpusavio sąveikas, ši duomenų savybė apriboja analizės galimybes (2 pav.). Atitinkamai, šiame darbe pristatoma analizė būtų neįmanoma be duomenų praturtinimo algoritmų. Iš kitos pusės, duomenų praturtinimas turi ir trūkumų. Pavyzdžiui, MAGIC algoritmo poveikyje gali būti prarandami subtilūs genų raiškos pokyčiai. Galima teigti, jog duomenų praturtinimo algoritmai pasižymi tendencija išryškinti dominuojančius genų tinklus silpnų sąsajų sąskaita. Todėl yra svarbu tinkamai parinkti algoritmo parametrus, siekiant užtikrinti optimalų duomenų praturtinimo ir jautrumo praradimo balansą. Atliktos analizės rezultatai sutampa su literatūroje aprašytais įžvalgomis apie EMT virsmą, o *ZEB1* aktyvacijos taikinių tyrimo metu keliamos hipotezės buvo validuotos papildomais eksperimentais. Be to, po duomenų praturtinimo nustatyti genų raiškos pokyčiai leido analizuoti

EMT virsmo dinamiką. Tokie rezultatai leidžia teigti, jog parinkus tinkamus parametrus duomenų praturtinimo algoritmas (MAGIC) tinkamai atkuria genų-genų sąveikos informaciją, kuri prarandama dėl pradinė duomenų retumo.

Šio darbo metu buvo analizuojama imortalizuotų krūties vėžio ląstelių (HMLE) modelinė sistema. Toks pasirinkimas leido vykdyti nuoseklią EMT virsmo stebėseną dvylikos dienų bėgyje bei nesunkiai atlikti papildomus validacijos eksperimentus (atliekant *ZEB1* geno raiškos indukciją). Tačiau reikia pripažinti, jog ne visos modelinėje sistemoje nustatytos išvalgos tiesiogiai atitinka EMT virsmą, stebimą *in vivo*. Pavyzdžiui, šiame darbe EMT procesas buvo indukuojamas TFG- β poveikyje, tačiau yra žinoma, jog krūties vėžio mikroaplinkoje veikia įvairūs papildomi signaliniai keliai, kurie daro įtaką ir EMT virsmui. Viena iš svarbiausių išvalgų EMT proceso tyrime yra dalinių EMT būsenų įvairovės patvirtinimas. Šių būsenų svarba pabrėžia ir tai, jog net tiesiogiai indukuojant EMT virsmą modelinėje sistemoje tik nedidelė dalis ląstelių turėjo mezenchiminį fenotipą. Dauguma literatūroje aprašomų EMT virsmo modelių nurodo, jog *in vivo* sąlygomis procesas tik išimtiniais atvejais pasiekia galutinį fenotipą. Be to, svarbus proceso aspektas yra jo plastiškumas, leidžiantis ląstelėms sugrįžti į pradinę epitelinę būseną. Šis atgalinis virsmas yra vadinamas mezenchiminių ląstelių virsmu epitelinėmis (MET) ir yra svarbus embriogenezės bei vėžio metastazių susidarymo metu. Tačiau šiame darbe EMT virsmo plastiškumas negalėjo būti analizuojamas, kadangi ląstelės buvo pastoviai stimuliuojamos EMT kryptimi. Atliktų tyrimų tikslas buvo ištirti scRNA-Seq ir kompiuterinių algoritmų pritaikomumą sudėtingų biologinių procesų tyrimui. Šioje disertacijos dalyje pristatomi rezultatai įrodo, jog taikant pažangius technologinius ir bioinformatinius metodus galima ne tik patikimai charakterizuoti sudėtingą biologinį virsmą, bet ir pateikti naujų išvalgų apie detalai ištyrinėtą procesą.

Krūties vėžio pacientų imuninių ląstelių tyrimas

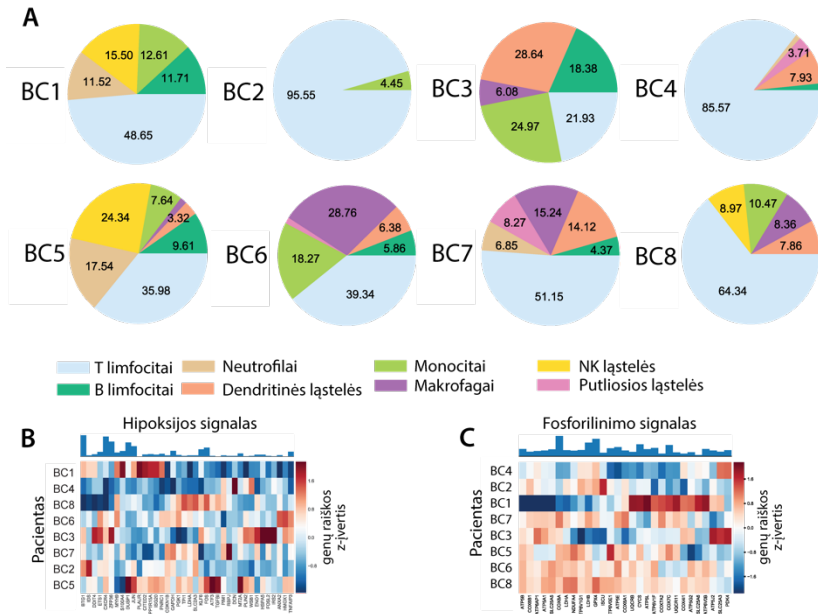
Šioje disertacijos dalyje yra pristatomi krūties vėžio pacientų imuninių ląstelių tyrimo rezultatai, kurie taip pat yra aprašyti disertacijos tema parengtoje publikacijoje [20]. Vykdytus tyrimus galima padalinti į dvi dalis. Pirmosios dalies tyrimų tikslas buvo sudaryti krūties vėžio pacientų imuninių ląstelių žemėlapi. Kitame tyrimo etape buvo siekiama charakterizuoti navike aptinkamas imunines ląsteles.

Lentelė 7. Darbo metu ištirtų krūties vėžio pacientų mėginių apibendrinimas. ER ir PER žymenų atveju nurodoma vėžinių ląstelių dalis, kurioje buvo nustatytas žymuo klinikinėje praktikoje naudojamais metodais.

Pacientas	Audinys				Žymuo		
	Navikas	Sveikas audinys	Kraujas	Limfmazgis	ER	PR	Her2
BC1	+	+	+	-	0.95	0.95	-
BC2	+	+	-	+	0.9	0.1	-
BC3	+	+	-	-	0	0	-
BC4	+	-	+	-	0.95	0.95	-
BC5	+	-	-	-	0.05	0.01	-
BC6	+	-	-	-	0.99	0.01	-
BC7	+	-	-	-	0	0	+
BC8	+	-	-	-	0.2	0.05	-

Atlikto tyrimo metu buvo ištirti aštuonių pacientų mėginiai (lentelė 7). Pacientai prieš operaciją nebuvo papildomai gydyti. Tais atvejais, kai operacijos aplinkybės sudarė galimybę, buvo analizuojami ir sveiko audinio, kraujo bei limfmazgių pavienės ląstelės iš to paties paciento. Darbo metu tirti pacientai pasižymėjo skirtingais krūties vėžio tipais, kurie buvo nustatyti remiantis standartiniais žymenimis (lentelė 7). Švieži klinikiniai mėginiai buvo disocijuojami, imuninės ląstelės atrenkamos FACS citometrijos pagalba (pagal CD45 žymenį) ir analizuojamos inDrops scRNA-Seq metodu. Siekiant užtikrinti aukštą tyrimo jautrumą visų sekoskaitos

eksperimentų metų buvo pasiekas >91% sekoskaitos įsisotinimas. Iš viso tyrimo metu buvo išanalizuotos 62 024 pavienės ląstelės.



8 Pav. A paveikslas: Skritulinės diagramos, apibendrinančios kiekvieno paciento naviko imuninių ląstelių sudėtį. Skirtingos spalvos žymi skirtingus ląstelių tipus. B paveikslas: Kairėje nurodomi hipoksijos metabolinio signalo vidurkis visoms kiekvieno paciento pavienėms ląstelėms. Dešinėje – konkrečių genų, susijusių su metaboliniu signalu, raiška kiekvienam pacientui. C paveikslas: kairėje nurodomi oksidacinio fosforilimo metabolinio signalo vidurkis visoms kiekvieno paciento pavienėms ląstelėms; dešinėje – konkrečių genų, susijusių su metaboliniu signalu, raiška kiekvienam pacientui.

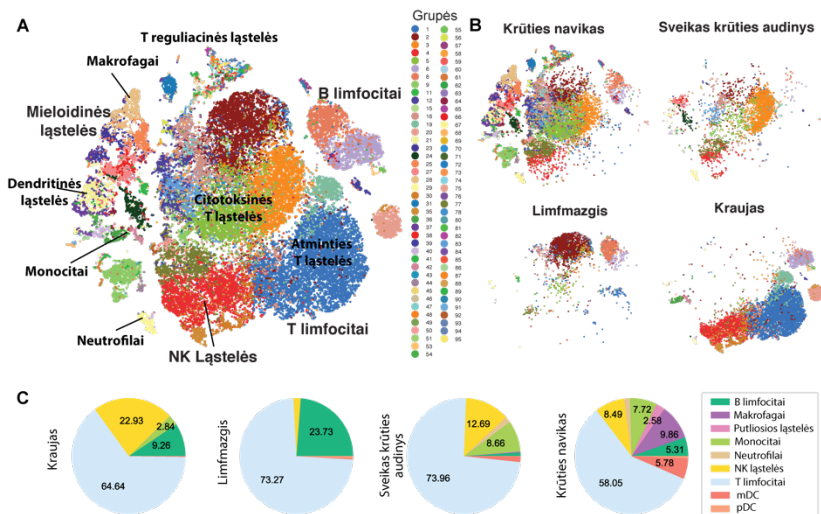
Pirmuoju darbo etapu kiekvieno paciento mėginiai buvo analizuojami atskirai. Pavienės ląstelės buvo grupuojamos ir kiekvienos grupės buvo nustatomos atliekant vidutinės genų raiškos profilio palyginimą su literatūroje aprašomais imuninių ląstelių genų raiškos profiliais. Atliekant šią analizę buvo siekiama patvirtinti, jog visų pacientų atveju buvo nustatomi tikėtini imuninių ląstelių tipai.

Atlikta analizė atskleidė, jog tarp pacientų yra stebimas ženklus imuninių ląstelių sudėties variabilumas (8 pav. A). Pavyzdžiui, mieloidinių ląstelių frakcija varijuoja 4-55% ribose, o T limfocitų frakcija – 21-96% ribose. Siekiant įvertinti šio variabilumo priežastis buvo atliekama vėžio metabolinių signalų analizė. Remiantis GSEA duomenų bazėje pateikiamais duomenimis buvo analizuojama hipoksijos, riebalų metabolizmo, glikolizės ir oksidacinio fosforilavimo signalai. Rezultatai parodė, jog pacientų mėginiuose buvo aktyvūs skirtingi signalo genai (8 pav. B ir C). Tokie rezultatai leidžia manyti, jog daliniai imuninių ląstelių sudėties skirtumai tarp pacientų yra nulemiami skirtingų naviko mikroaplinkos signalų.

Kitu tyrimo etapu buvo konstruojamas krūties vėžio pacientų imuninių ląstelių atlasas. Dėl ženklų techninių tyrimo artefaktų (operacijos eigos, audinių disociacijos bei mėginių transportavimo sąlygų skirtumų poveikio) pradiniai duomenys pasižymėjo ženkliai specifiskumu pacientui. Siekiant pašalinti stebimus artefaktus buvo pritaikytas duomenų normalizavimo ir praturtinimo algoritmas Biscuit [37]. Po normalizavimo, praturtinimo ir duomenų apjungimo buvo nustatytos 95 ląstelių grupės. Dalis iš jų buvo žemos kokybės ląstelės ir buvo pašalinti iš analizės. Po šio pradinio duomenų apdorojimo sudarytas pavienių ląstelių atlasas apėmė 57 143 pavienes ląsteles. Toliau kiekvienai grupei buvo priskiriamas ląstelės tipas lyginant vidutinę grupės genų raišką su literatūroje aprašomais imuninių ląstelių genų raiškos profiliais. Šios analizės metu nustatyta, jog dalis grupių atitiko vėžines naviko ląsteles. Šios grupės buvo pašalintos iš tolimesnės analizės. Darbo rezultatas yra 47 016 pavienes ląsteles iš aštuonių pacientų apimantis imuninių ląstelių atlasas (9 pav. A). Iš viso nustatytos 83 ląstelių grupės atitinkančios šiuos ląstelių tipus: 38 T limfocitų grupės, 27 mieloidinių ląstelių grupės, 9 B limfocitų grupės ir 9 NK ląstelių grupės. Svarbu paminėti, jog išskiriant šias ląstelių grupes buvo naudojami viso transkriptomo duomenys ir tik po to buvo nustatomas konkretus ląstelių tipas pagal specifinius literatūroje nurodomus žymenis. Tolesnio tyrimo metu pagrindinis

dėmesys buvo skirtas T limfocitų ir mieloidinių ląstelių charakterizavimui dėl šių ląstelių svarbos vėžio imunoterapijos kontekste.

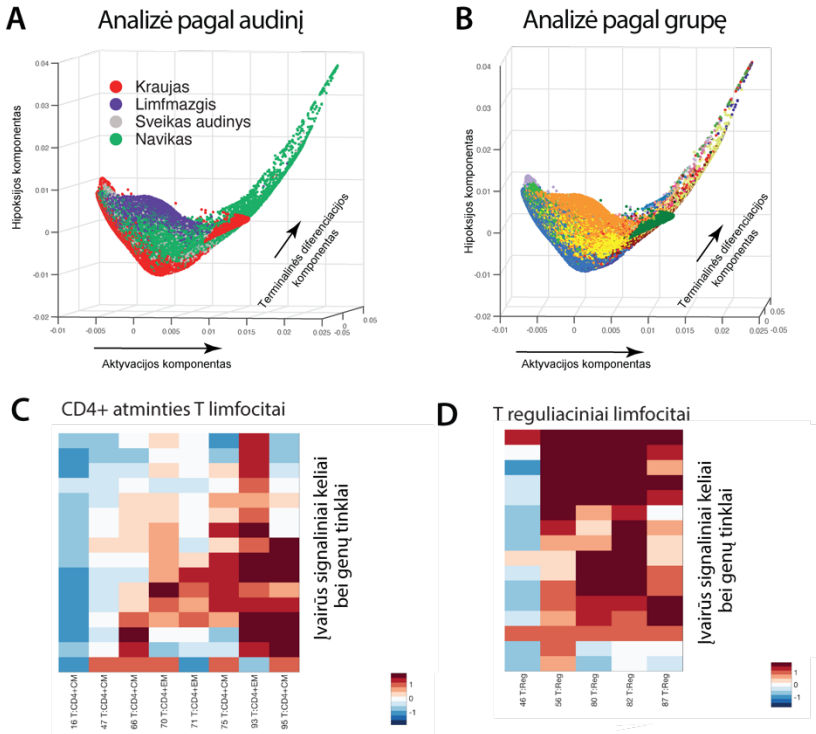
Analizuojant imuninių ląstelių sudėties skirtumus tarp skirtingų audinių buvo nustatyti akivaizdūs skirtumai (9 pav. B ir C). Statistiškai patikimai galima daryti keletą išvadų. Pavyzdžiui, naivūs T limfocitai daugiausiai aptinkami kraujyje (χ^2 , $p = 3 \times 10^{-80}$), o B limfocitai yra dažniau stebimi limfmazgyje (χ^2 , $p = 0,0$). Šios įžvalgos nėra naujos, tačiau aukštas statistinis patikimumas įrodo atliktos analizės tikslumą. Lyginant sveiką ir vėžinį krūties audinius galima pastebėti akivaizdų imuninių ląstelių sudėties įvairovės padidėjimą navike (9 pav. B). Įdomu tai, jog nustatytas tiek citotoksinių T limfocitų (χ^2 , $p = 3 \times 10^{-25}$), tiek reguliacinių T limfocitų (χ^2 , $p = 5 \times 10^{-91}$) kiekio padidėjimas. Tokie rezultatai atskleidžia sudėtingą pusiausvyrą tarp imuninės sistemos aktyvacijos ir slopinimo naviko mikroaplinkoje. Analogiškai, stebimas makrofagų įvairovės padidėjimas navike (χ^2 , $p = 0,0$) lyginant su sveiku audiniu. Svarbi įžvalga yra ir tai, jog didelė dalis sveikame krūties audinyje aptinkamų imuninių ląstelių tipų nebuvo nustatomi kraujyje bei limfmazgyje. Iš kitos pusės, visos sveikame krūties audinyje aptinkamos imuninės ląstelės yra naviko imuninės sudėties dalis. Tokie rezultatai rodo, jog navike vyksta imuninių ląstelių fenotipų įvairovės padidėjimas. Apibendrinant šias įžvalgas, galima teigti, jog audinio aplinka yra svarbi formuojant imuninių ląstelių sudėtis.



9 Pav. A paveikslas: krūties vėžio pacientų imuninių ląstelių atlaso vizualizacija tSNE algoritmu. Skirtingos spalvos žymi skirtingas ląstelių grupes. B paveikslas: skirtingų audinių imuninių sudėties vizualizacija tSNE algoritmu. Skirtingos spalvos žymi skirtingas ląstelių grupes. C paveikslas: skritulinės diagramos apibendrinančios kiekviename tirtame audinyje nustatomus ląstelių tipus ir jų santykinį kiekį. Pateikiami apibendrinti visų tirtų mėginių duomenys. Skirtingos spalvos žymi skirtingus ląstelių tipus.

Tyrimo metu nustatyta padidėjusi imuninių ląstelių fenotipo įvairovė naviko audinyje buvo analizuojama detaliau. Siekiant įvertinti svarbiausius veiksnius, nulemiančius fenotipo įvairovės padidėjimą, buvo atliekama daugiamatės genų ekspresijos erdvės dimensijų mažinimo analizė. Šiuo tikslu buvo taikoma difuzinių žemėlapių analizė. Žemesnio dimensiškumo erdvėje pagrindinės sistemą aprašančios komponentės atspindi svarbiausius biologinius veiksnius, nulemiančius ląstelių fenotipų įvairovę. Ši analizė buvo atliekama atskirai T limfocitams ir mieloidinėms ląstelėms. Tyrimo metu nustatyta, jog svarbiausi veiksniai, lemiantys T limfocitų įvairovę yra ląstelių aktyvacija, terminalinė diferenciacija ir hipoksija (10 pav.). Ląstelių išsidėstymas pagal šiuos komponentus buvo nagrinėjamas dviem aspektais. Analizuojant ląstelių išsidėstymą

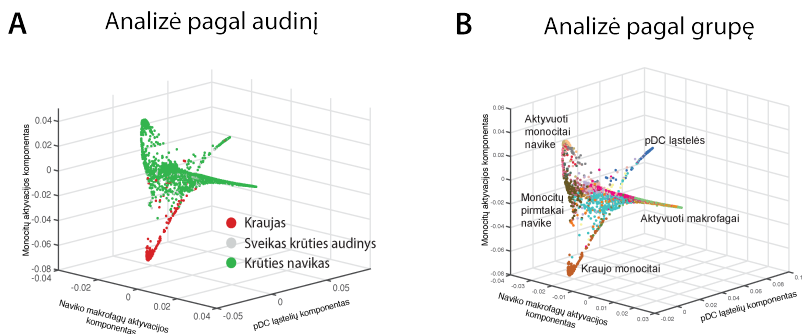
difuzijos komponentų erdvėje pagal audinio lokalizaciją (10 pav. A), galima teigti, jog sveikuose audiniuose yra stebima santykinai žema ląstelių terminalinė diferenciacija. Tokie rezultatai nėra netikėti ir sutampa su konkrečiuose audiniuose aptinkamais ląstelių tipais. Pavyzdžiui, didelė dalis kraujo T limfocitų buvo naivios ląstelės, kurios pasižymi itin žemu terminalinės diferenciacijos signalu. Analogiškai, hipoksijos įtaka taip pat stebima dažniausiai naviko audinyje. Be to, tarp aktyvacijos ir terminalinės diferenciacijos yra stebima santykinai aukšta koreliacija. Panašios išvados buvo pristatytos ir kitame scRNA-Seq tyrime [38]. Įdomu tai, jog analizuojant naviko T limfocitų išsidėstymą pagal jų priskyrimą grupėms (10 pav. B) nesimato aiškaus dėšningumo – ląstelės iš tos pačios grupės tolydžiai išsidėsto pagal tiriamus difuzijos komponentus. Tokie rezultatai leidžia teigti, jog priešingai įprastam T limfocitų aktyvacijos modeliui, ląstelės esančios navike neturėtų būti skirstomos į kelis griežtai apibrėžtus aktyvacijos fenotipus. Be to, tokia analizė parodė, jog pagrindiniai difuzijos komponentai iki galo nepaaiškina ląstelių atsiskyrimo į grupes. Siekiant suprasti faktorius, nulemiančius T limfocitų grupių įvairovę, buvo atlikta papildoma analizė, kurios metu buvo vertinama signalinių kelių bei atskirų genų tinklų aktyvacija konkrečiuose ląstelių grupėse (10 pav. C ir D). Rezultatai atskleidžia, jog skirtingos to paties ląstelių tipo grupės tarpusavyje stipriai skiriasi savo transkripcinėmis programomis. Tokius rezultatus galėtų paaiškinti naviko mikroaplinkoje susidaranti lokalios nišos, tarpusavyje besiskiriančios savo signaline aplinka, ir nulemia stebimą ląstelių fenotipų įvairovę. Be to, papildomos analizės rezultatai leidžia teigti, jog ne tik naviko mikroaplinkos signalai, bet ir imuninių ląstelių tarpusavio sąveikos signalai daro įtaką fenotipinių būsenų įvairovės susiformavimui.



10 Pav. A paveikslas: T limfocitų išsidėstymas difuzijos komponentų erdvėje. Spalvinimas pagal ląstelių kilmės audinį. B paveikslas: T limfocitų išsidėstymas difuzijos komponentų erdvėje. Spalvinimas pagal ląstelių grupes atitinka 9 pav. spalvas. C paveikslas: CD4+ atminties T limfocitų grupių analizė pagal įvairių signalinių kelių bei genų tinklų aktyvaciją. D paveikslas: reguliacinių T limfocitų grupių analizė pagal įvairių signalinių kelių bei genų tinklų aktyvaciją.

Analogiška difuzijos komponentų analizė buvo atlikta ir mieloidinėms ląstelėms (11 pav.). Pagrindiniai šių ląstelių įvairovę paaikšinantys veiksniai: naviko makrofagų aktyvacija, monocitų aktyvacija ir plazmocitoidinių dendritinių ląstelių (pDC) komponentas. Nustatyta, jog audinio aplinka taip pat atlieka svarbų vaidmenį nulemiant ląstelių fenotipą (11 pav. A). Priešingai nei T limfocitų analizės atveju, mieloidinių ląstelių grupės pagal difuzijos komponentus yra išsidėčiusios labiau lokalizuotai (11 pav. B). Įdomu

ir tai, jog tyrimo duomenyse nustatyta M1 ir M2 makrofagų genų raiškos koreliacija tose pačiose ląstelėse. Visos trys naviko makrofagų grupės pasižymėjo aukšta abiejų būsenų genų aktyvacija. Panašios išvalgos neseniai buvo pristatytos ir melanomos kontekste [39]. Tokie rezultatai sustiprina pastaruoju metu keliamas hipotezes, jog esamas alternatyvios makrofagų poliarizacijos modelis nepilnai paaikšina makrofagų aktyvaciją naviko mikroaplinkoje.



11 Pav. A paveikslas: mieloidinių ląstelių išsidėstymas difuzijos komponentų erdvėje. Spalvinimas pagal ląstelių kilmės audinį. B paveikslas: mieloidinių ląstelių išsidėstymas difuzijos komponentų erdvėje. Spalvinimas pagal ląstelių grupes atitinka 9 pav. spalvas.

Šioje disertacijos dalyje pristatomas krūties vėžio pacientų imuninių ląstelių atlasas atskleidžia didelę imuninių ląstelių fenotipinę įvairovę bei aukštą jų sudėties variabilumą tarp skirtingų pacientų. Vienas iš tyrimo apribojimų yra tai, jog ne visais atvejais naviko mėginiai buvo analizuojami poroje su sveiku audiniu iš to paties paciento (lentelė 7). Dėl to galima daryti tik ribotas išvadas apie imuninių ląstelių skirtumus tarp sveiko ir suvėžėjusio krūties audinio. Nepaisant to, turimų duomenų pakanka teigti, jog tarp šių audinių imuninių ląstelių sudėčių yra stebimi esminiai skirtumai. Be to, verta atkreipti dėmesį į tai, jog visi sveikame audinyje stebimi imuninių ląstelių tipai buvo aptinkami ir to paties paciento naviko mėginyje. Toks rezultatas leidžia manyti, jog naviko mikroaplinkos signalų

poveikyje ženkliai praplečiama fenotipinė ląstelių įvairovė. Tokia hipotezė buvo patvirtinta ir tolesnių tyrimų rezultatų. Analogiškai, nors analizuotų kraujo bei limfmazgių mėginių kiekis yra dar mažesnis, gauti rezultatai leidžia manyti, jog ląstelių tipai tarp šių audinių ir vėžinio audinio turi mažai panašumų. Tokios išvalgos verčia pasvarstyti apie kraujyje aptinkamų imuninių ląstelių biožymenų aktualumą vėžio diagnostikoje. Deja, dėl riboto mėginių skaičiaus patikimų išvadų daryti negalima, o siekiant atsakyti šį klausimą reikalingi papildomi tyrimai.

Svarbus pavienių ląstelių transkriptomo tyrimų aspektas yra ne visuomet aiški riba tarp ląstelės būsenos ir tipo. Remiantis disertacijoje pristatomais rezultatais galima teigti, jog ląstelių skirstymas pagal kanoninius jų tipus neatspindėtų visos sudėties įvairovės. Iš kitos pusės, šio tyrimo duomenys nėra pakankami, siekiant atsakyti į klausimą, ar visos aptiktos ląstelių fenotipinės būsenos atlieka aiškų funkcinį vaidmenį. Galimai dalis ląstelių būsenų yra tarpinėse virsmo stadijose ir nepasižymi efektoriniu fenotipu. Taip pat galima manyti, jog dalis nustatytų būsenų pasižymi vienodu efektoriniu fenotipu, o stebimi transkriptomo skirtumai yra skirtingų signalinių aplinkų lokaliuose naviko mikroaplinkos nišose rezultatas.

Šioje disertacijos dalyje pristatomų tyrimu metu papildomas dėmesys buvo skirtas navike aptinkamų T limfocitų ir mieloidinių ląstelių analizei. Gauti rezultatai leidžia manyti, jog modelis, teigiantis, kad aktyvacijos metu T limfocitai užima diskrečias stabilias būsenas (pavyzdžiui reguliaciniai, atminties, ar efektoriniai T limfocitai) nėra iki galo tikslus. Daugeliu atveju vienos grupės ląstelės (atitinkančios vieną fenotipinę būseną) nuosekliai išsidėstė palei aktyvacijos komponentą. Tokie rezultatai leidžia manyti, jog T limfocitų aktyvacija naviko mikroaplinkoje yra sudėtingas ir netiesinis procesas. Mieloidinių ląstelių atveju fenotipinės būsenos aiškiau atsiskyrė pagal pagrindinius difuzijos (variabilumo) komponentus. Šis skirtumas gali būti paaiškinamas tuo, jog mieloidinių ląstelių įvairovė didele dalimi yra formuojama jų vystymosi eigoje [40]. Tyrimo metu

taip pat parodyta, jog esamas makrofagų poliarizacijos modelis nepilnai paaiškina stebimus genų raiškos profilius. Nustatyta, jog su M1 ir M2 poliarizacijos fenotipais siejami genai yra nustatomi toje pačioje ląstelėje. Toks rezultatas papildo analogiškas kitų mokslininkų išvalgas inkstų bei plaučių vėžio kontekste [41, 42]. Atlikti tyrimai atskleidžia scRNA-Seq technologijos potencialą sudėtingų biologinių sistemų tyrimams. Vis dėlto, keliamos hipotezės turėtų būti papildomai validuojamos atliekant tikslinius eksperimentus.

Disertacijos tyrimų metu parengtas krūties vėžio pacientų imuninių ląstelių atlasas taip pat yra svarbus įrankis krūties vėžio tyrimų bendruomenei. Šiame atlase pirmą kartą yra atskleidžiama itin didelė imuninių ląstelių fenotipinė įvairovė, o pristatomi duomenys leidžia geriau pažinti krūties vėžio imuninę aplinką. Tačiau reikia pripažinti, jog atlikti tyrimai nėra pakankami visiems keliamiems klausimams atsakyti. Svarbus aspektas, nenagrinėtas šiame tyrime, yra vėžinių ir imuninių ląstelių tarpusavio sąveika. Taip pat, tolesnių tyrimų metu būtų tikslinga analizuoti pacientus gydymo eigoje. Tokie tyrimai galėtų padėti suprasti įvairių priešvėžinių terapijų poveikį bei apribojimus ir padėti efektyviau gydyti pacientus.

IŠVADOS

- Optimizuojant atskirus inDrops metodo sekoskaitos bibliotekos paruošimo žingsnius, buvo iki 10-ies kartų padidintas transkriptų kiekį, nustatomą iš pavienių ląstelių.
- Difuzijos žemėlapiams paremtas duomenų praturtinimo algoritmas (MAGIC) atkuria genų-genų sąveikos informaciją, kuri prarandama dėl pradinių duomenų retumo.
- EMT procesą indukuojant TGF- β transkripcijos faktorius ZEB1 tiesiogiai ir netiesiogiai aktyvuoja 1085-ies genų raišką.
- Tarp skirtingų pacientų krūties navikų yra stebima skirtinga naviką infiltruojančių imuninių ląstelių įvairovė. Šiuos skirtumus lemia skirtingi vėžio mikroaplinkos signalai. Mieloidinių ląstelių frakcija varijuoja 4-55% ribose, o T limfocitų frakcija – 21-96% ribose.
- Krūties vėžio audinyje yra aptinkama ženkliai didesnė T limfocitų fenotipinių būsenų įvairovė lyginant su sveiku krūties audiniu. Trys svarbiausi veiksniai lemiantys įvairovės padidėjimą yra T ląstelių aktyvacija, terminalinė diferenciacija bei hipoksija.

TRUMPOS ŽINIOS APIE DISERTANTĄ

Vardas Juozas Nainys
Adresas Gyvybės mokslų centras, Saulėtekio al. 7, 10257,
Vilnius, Lietuva
El. pašto adresas juozas.nainys@bti.vu.lt

Išsilavinimas

2015-2019 Doktorantūros studijos, Biochemija, Vilniaus universitetas
2013-2015 Magistras, Biochemija, Vilniaus universitetas
2009-2013 Bakalauras, Biochemija, Vilniaus universitetas

Profesinė patirtis

nuo 2017/09 Lektorius, Vilniaus universitetas, Gyvybės mokslų centras
Nuo 2016/09 Jaun. mokslo darbuotojas, Vilniaus universitetas, Biotechnologijos institutas
2015/09-2016/09 Vizituojantis mokslininkas, Columbia university (JAV)
2015/04-2015/12 Biologas tyrėjas, Vilniaus universitetas, Biotechnologijos institutas
2013/01-2015/04 Specialistas, Vilniaus Universitetas, Biotechnologijos institutas
2011/07-2013/01 Laborantas, Vilniaus Universitetas, Biotechnologijos institutas

Publikacijų sąrašas

1. Leonavicius K, **Nainys J**, Kuciauskas D, Mazutis L. Multi-omics at single-cell resolution: comparison of experimental and data fusion approaches. (2019) *Current opinion in biotechnology* 55, 159-66, Feb 1
2. **Nainys J**, Milkus V, Mažutis L. Single-cell screening using microfluidic systems. (2019) *In Microfluidics for Pharmaceutical Applications*. William Andrew Publishing. 353-367, Jan 1
3. Azizi E, Carr A, Plitas G, Cornish EA, Konopacki C, Prabhakaran S, **Nainys J.**, Wu K., Kiseliovas V., Setty M., Choi K., Dao P, Mazutis L, Rudensky YA & Peer D, Single-cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment (2018), *Cell* 174, 1–16, August 23
4. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr A, Burdziak C, Moon RK, Chaffer C, Pattabiraman D, Bieri B, Mazutis L, Wolf G, Krishnaswamy S, Peer D. Recovering gene interactions from single-cell data using data diffusion, (2018) *Cell*, 174, 716-729, July 29
5. Zilionis R, **Nainys J**, Veres A, Savova V, Zemmour D, Klein MA, and Mazutis. (2017) Single-cell barcoding and sequencing using droplet microfluidics, *Nature Protocols* 12, 44–73
6. **Nainys J**, Timinskas A, Schneider J, Ulrich RG, Gedvilaite A. (2015). Identification of Two Novel Members of the Tentative Genus Wukipolyomavirus in Wild Rodents. *PLoS one* 10.10, e0140916.
7. Zaveckas M, Snipaitis S, Pesliakas H., **Nainys J**, Gedvilaite A. (2015). Purification of recombinant virus-like particles of porcine circovirus type 2 capsid protein using ion-exchange monolith chromatography. *Journal of Chromatography B*.991, 21-28.
8. **Nainys J**, Lasickiene ., Petraityte-Burneikiene R, Dabrisius J, Lelesius R, Sereika V, Gedvilaite A. (2014). Generation in yeast of recombinant virus-like particles of porcine circovirus type 2 capsid protein and their use for a serologic assay and development of monoclonal antibodies. *BMC biotechnology*.14(1), 100.
9. Šimoliūnas E., Kaliniene L, Stasilo M, Truncaitė L, Zajančauskaitė A, Staniulis J, **Nainys J**, Meškys R. (2014). Isolation and Characterization of

vB_ArS-ArV2–First *Arthrobacter* sp. Infecting Bacteriophage with Completely Sequenced Genome. *PLoS One* .9(10), e111230.

Pranešimai konferencijose:

Žodiniai:

- 2017-06. VIB Novel tools for transcriptional profiling, Belgija. Pranešimo pavadinimas: “Single-cell barcoding and sequencing using droplet microfluidics”
- 2017-12, Bioateitis, Lietuva. Pranešimo pavadinimas: “Studying cancer metastasis models with single-cell transcriptomics”
- 2018-02, The COINS, Lietuva. Pranešimo pavadinimas: “Full-stack single-cell transcriptomics”

•

Stendiniai:

- 2016-02, Genomics and Personalized Medicine, Kanada. Pranešimo pavadinimas: “Single-cell barcoding and sequencing using droplet microfluidics”
- 2017-06, Single Cell Omics, Švedija. Pranešimo pavadinimas: “Single-cell barcoding and sequencing using droplet microfluidics”

LITERATŪROS SĄRAŠAS

1. Arendt, D., et al., *The origin and evolution of cell types*. Nature Reviews Genetics, 2016. **17**(12): p. 744.
2. Collins, F.S., M. Morgan, and A. Patrinos, *The Human Genome Project: lessons from large-scale biology*. Science, 2003. **300**(5617): p. 286-290.
3. Velculescu, V.E., et al., *Serial analysis of gene expression*. Science, 1995. **270**(5235): p. 484-487.
4. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell*. Nature methods, 2009. **6**(5): p. 377.
5. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells*. Cell, 2015. **161**(5): p. 1187-1201.
6. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets*. Cell, 2015. **161**(5): p. 1202-1214.
7. Rozenblatt-Rosen, O., et al., *The Human Cell Atlas: from vision to reality*. Nature News, 2017. **550**(7677): p. 451.
8. Regev, A., et al., *Science forum: the human cell atlas*. Elife, 2017. **6**: p. e27041.
9. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition*. The Journal of clinical investigation, 2009. **119**(6): p. 1420-1428.
10. Zeisberg, M. and E.G. Neilson, *Biomarkers for epithelial-mesenchymal transitions*. The Journal of clinical investigation, 2009. **119**(6): p. 1429-1437.
11. Dongre, A. and R.A. Weinberg, *New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer*. Nature reviews Molecular cell biology, 2019. **20**(2): p. 69-84.
12. Harbeck, N., et al., *Breast cancer*. Nature Reviews Disease Primers, 2019. **5**(1): p. 66.
13. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers*

- in 185 countries*. CA: a cancer journal for clinicians, 2018. **68**(6): p. 394-424.
14. Bray, F., et al., *Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration*. International journal of cancer, 2015. **137**(9): p. 2060-2071.
 15. Cardoso, F., et al., *4th ESO–ESMO international consensus guidelines for advanced breast cancer (ABC 4)*. Annals of Oncology, 2018. **29**(8): p. 1634-1657.
 16. Segovia-Mendoza, M. and J. Morales-Montor, *Immune tumor microenvironment in breast cancer and the participation of estrogens and its receptors into cancer physiopathology*. Frontiers in Immunology, 2019. **10**.
 17. Gaublomme, J.T., et al., *Single-cell genomics unveils critical regulators of Th17 cell pathogenicity*. Cell, 2015. **163**(6): p. 1400-1412.
 18. Mazutis, L., et al., *Single-cell analysis and sorting using droplet-based microfluidics*. Nature protocols, 2013. **8**(5): p. 870.
 19. Zilionis, R., et al., *Single-cell barcoding and sequencing using droplet microfluidics*. Nature protocols, 2017. **12**(1): p. 44.
 20. Azizi, E., et al., *Single-cell map of diverse immune phenotypes in the breast tumor microenvironment*. Cell, 2018. **174**(5): p. 1293-1308. e36.
 21. Van Dijk, D., et al., *Recovering gene interactions from single-cell data using data diffusion*. Cell, 2018. **174**(3): p. 716-729. e27.
 22. Mørup, M. and L.K. Hansen, *Archetypal analysis for machine learning and data mining*. Neurocomputing, 2012. **80**: p. 54-63.
 23. Levine, J.H., et al., *Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis*. Cell, 2015. **162**(1): p. 184-197.
 24. Maaten, L.v.d. and G. Hinton, *Visualizing data using t-SNE*. Journal of machine learning research, 2008. **9**(Nov): p. 2579-2605.
 25. Coifman, R.R., et al., *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion*

- maps*. Proceedings of the national academy of sciences, 2005. **102**(21): p. 7426-7431.
26. Svensson, V., et al., *Power analysis of single-cell RNA-sequencing experiments*. Nature methods, 2017. **14**(4): p. 381.
 27. Ding, J., et al., *Systematic comparison of single-cell and single-nucleus RNA-sequencing methods*. Nature Biotechnology, 2020: p. 1-10.
 28. Zhang, X., et al., *Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems*. Molecular cell, 2019. **73**(1): p. 130-142. e5.
 29. Keren-Shaul, H., et al., *MARS-seq2. 0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing*. Nature protocols, 2019. **14**(6): p. 1841.
 30. David, C.J., et al., *TGF- β tumor suppression through a lethal EMT*. Cell, 2016. **164**(5): p. 1015-1030.
 31. Nieto, M.A., et al., *Emt: 2016*. Cell, 2016. **166**(1): p. 21-45.
 32. Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors*. Nature genetics, 2008. **40**(5): p. 499.
 33. Krishnaswamy, S., et al., *Conditional density-based analysis of T cell signaling in single-cell data*. Science, 2014. **346**(6213): p. 1250689.
 34. Stemmler, M.P., et al., *Non-redundant functions of EMT transcription factors*. Nature cell biology, 2019. **21**(1): p. 102-112.
 35. Dixit, A., et al., *Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens*. Cell, 2016. **167**(7): p. 1853-1866. e17.
 36. Datlinger, P., et al., *Pooled CRISPR screening with single-cell transcriptome readout*. Nature methods, 2017. **14**(3): p. 297.
 37. Prabhakaran, S., et al. *Dirichlet process mixture model for correcting technical variation in single-cell gene expression data*. in *International Conference on Machine Learning*. 2016.
 38. Tirosh, I., et al., *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq*. Science, 2016. **352**(6282): p. 189-196.

39. Müller, S., et al., *Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment*. *Genome biology*, 2017. **18**(1): p. 234.
40. Perdiguero, E.G. and F. Geissmann, *The development and maintenance of resident macrophages*. *Nature immunology*, 2016. **17**(1): p. 2.
41. Chevrier, S., et al., *An immune atlas of clear cell renal cell carcinoma*. *Cell*, 2017. **169**(4): p. 736-749. e18.
42. Lavin, Y., et al., *Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses*. *Cell*, 2017. **169**(4): p. 750-765. e17.

UŽRAŠAMS

UŽRAŠAMS

UŽRAŠAMS

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