

TARTU ÜLIKOOL  
LOODUS- JA TEHNOLOOGIATEADUSKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT

Hindrek Teder

## **Geeniekspressiooni uurimine ühe raku tasemel**

Bakalaureusetöö

Juhendaja Kaarel Krjutškov, PhD  
Kaasjuhendaja prof Ants Kurg, PhD

TARTU 2013

# SISUKORD

KASUTATUD LÜHENDID .....	3
SISSEJUHATUS .....	4
1 KIRJANDUSE ÜLEVAADE.....	5
1.1 RNA sekveneerimine .....	5
1.2 Ühe raku RNA-sek.....	6
1.2.1 Tang'i meetod.....	7
1.2.2 SMART-il põhinevad meetodid .....	9
1.2.2.1 STRT meetod.....	10
1.2.2.2 Smart-Seq meetod.....	12
1.2.3 CEL-Seq meetod.....	14
2 UURIMUS .....	16
2.1 Töö eesmärgid.....	16
2.2 Materjal ja meetodika.....	16
2.2.1 Märksõnad.....	16
2.2.2 Andmete analüüsi meetodid .....	16
2.2.3 Uuritavad objektid .....	17
2.3 Tulemused ja arutelu .....	17
2.3.1 Protokollide sarnasused ja erinevused .....	18
2.3.2 Meetodite suutlikkus.....	20
2.3.3 Perspektiiv .....	24
KOKKUVÕTE .....	26
SUMMARY .....	27
TÄNUAVALDUSED .....	28
KASUTATUD KIRJANDUS .....	29
KASUTATUD VEEBIAADRESSID .....	34
LIHTLITSENTS.....	35

## KASUTATUD LÜHENDID

**cDNA** – *complementary DNA* – komplementaarne DNA

**mRNA** – *messenger RNA* – informatsiooni-RNA

**IVT** – *in vitro transcription* – *in vitro* transkriptsioon

**MMLV** – *Moloney murine leukemia virus* – Moloney hiire leukeemia viirus

**SMART** – *switching mechanism at the 5' end of the RNA transcript* – RNA matriitsahela 5' otsa põhine ümberlülitusmeetod

**RNA-sek** – *RNA-sequencing (RNA-seq)* – RNA sekveneerimistehnoloogia

**STRT** – *single-cell tagged reverse transcription* – ühe raku RNA-sek meetod

**CEL-Seq** – *cell expression by linear amplification and sequencing* – ühe raku RNA-sek meetod

## SISSEJUHATUS

Geenide avaldumine ehk geeniekspressioon on eluslooduse fundamentaalne protsess, mille tulemusel sünteesitakse geeni nukleiinhappe järjestusse kodeeritud informatsiooni põhjal kindla funktsiooniga produkt. Geeniekspressiooni käigus sünteesitud produktideks on sageli valgud, kuid proteiini mittekodeerivate geenide korral on produktiks funktsioneeriv RNA, näiteks ribosoomi-, transpordi-, või väikesed tuuma RNA-d. Nimetatud RNA produktide funktsioon on veel paljuski selgusetu, kuid on teada, et osadel neist on oluline roll transkriptsiooni ja translatsiooni regulatsioonis. Ajalooliselt kauem ja tulemuslikumalt on uuritud rakkudes sünteesitud ja seal funktsioneerivaid valke, mis iseloomustavad enda ekspressiooni määra ning paiknemisega raku morfoloogiat ja realiseerivad raku funktsioone.

Oskus kvantitatiivselt määrata, millisel tasemel antud geen avaldub rakus, koes või organismis, avab uued võimalused organismi tundmaõppimiseks ning pakub olulist informatsiooni baasuuringutes, meditsiinilistes uuringutes ja diagnoosimises. Geeniproducti hulka on võimalik määrata, kuid sageli on lihtsam mõõta produkti eelvormi, tavaliselt informatsiooni-RNA-d (mRNA), mille põhjal saab tuletada geeniekspressiooni ulatust ning muutusi. Geeniekspressiooni mõõtmiseks läbi mRNA on kasutusse võetud mitmeid erinevaid tehnoloogiaid, millest nimekamad on reaalkaja-PCR, ekspressiooni geenikiip ning RNA sekveneerimine (RNA-sek). Teise põlvkonna sekveneerimise tehnoloogiate panust RNA-sek arengus on raske ülehinnata, sest RNA-sek on tõestanud end paljudes uuringutes, mistõttu eelistatakse RNA sekveneerimist varasemalt valdkonda valitsenud geenikiibil põhineva meetodi ees.

RNA-sek meetodi edasiarendused võimaldavad uurida geeniekspressiooni väga väikesest lähtematerjalist – ühe raku tasemel, andes uusi teadmisi näiteks raku jagunemise, rakkudevahelise variatsiooni ja geeniregulatsiooni kohta diferentseerunud rakkudes. Uued publikatsioonid pakuvad täiendusi RNA-sek meetoditele ning seetõttu tekib vajadus võrdleva ülevaate järele, mis suudaks välja tuua erinevate meetodite eelised ja puudused, et seeläbi suurendada meie töögrupi teadmiseid tulevaste projektide planeerimisel.

Töö eesmärk on tutvustada siiani avaldatud ühe raku põhiseid RNA-sek protokolle, tuua välja nende sarnasused ja erinevused, võrrelda nende võimekust ning arutleda, miks on ühe raku geeniekspressiooni uurimine üldse oluline.

Töö on koostatud Reproduktiivmeditsiini Tehnoloogia Arenduskeskuses Kaarel Krjutškovi juhendamisel. Töö kaasjuhendajaks on Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituudi molekulaarse biotehnoloogia professor Ants Kurg.

# 1 KIRJANDUSE ÜLEVAADE

## 1.1 RNA sekveneerimine

Transkriptoomi iseloomustamine on olnud laialt kasutatav molekulaarbioloogia uurimise vahend juba enam kui kümme aastat. Sellesse aega jäi geenikiibi tehnoloogia võidukäik, mis põhineb tahkele kandjale kinnitatud proovi molekulide ning lahuses olevate sihtmärk-molekulide vahelisel hübriidsatsioonil. Viimaste aastate jooksul on RNA-sek tehnoloogia välja vahetanud geenikiibid ning eksponentsiaalselt on kasvanud katsetest saadud andmete hulk ja paranenud nende kvaliteet. Moodsad sekveneerimise tehnoloogiad suudavad paralleelselt analüüsida miljoneid järjestusi korraga ja teha seda kõike aina väiksemate kuludega (Church, 2006; Hall, 2007).

RNA-sek, tuntud ka kui kogu transkriptoomi pimejärjestamine (Morin et al., 2008), mRNA-sek ja cDNA-sek, on mRNA järjestamiseks ning geeniekspressiooni analüüsimiseks välja töötatud meetod, mis põhineb moodsatel sekveneerimise tehnoloogiatel. RNA-sek toimub kas otseselt (Ozsolak et al., 2009) või komplementaarse DNA (cDNA) järjestuse alusel pärast pöördtranskriptsiooni (Cloonan et al., 2008; Mortazavi et al., 2008; Wang et al., 2008). Praegused tehnoloogilised piirangud ei võimalda transkriptide järjestust määrata ühe tervikuna, vaid see jaotatakse juhuslikult mõnekümne kuni -saja nukleotiidilisteks järjestusteks. Saadud järjestused joondatakse referentsjärjestusele, mille tulemusel kaetakse kogu geeni kodeeriv järjestus. Järgneb referentsile joondunud järjestuste kokkulugemine, mille põhjal on võimalik hinnata transkriptide hulka ja määrata geeni avaldumise tase. Antud uurimuses keskendutakse just cDNA põhisele RNA sekveneerimise tehnoloogiatele, sest enamik moodsatest sekvenaatoritest võimaldab analüüsida vaid DNA-l põhinevaid raamatukogusid, jättes otsese RNA analüüsi kättesaadavaks seni veel väga vähestele töögruppidele.

Võrreldes ekspressiooni geenikiipidega, on RNA-sek meetodil mitmeid lisaomadusi, mistõttu on kõnealusest tehnoloogiast saanud eelistatud valik transkriptoomi analüüsimisel (Fu et al., 2009). (i) RNA-sek puhul ei ole transkriptide detekteerimine limiteeritud ainult teadaolevate järjestustega, vaid võimaldab järjestuste *de novo* reastamist, mis teeb sellest äärmiselt atraktiivse meetodi ka vähem uuritud organismide analüüsimisel (Vera et al., 2008). (ii) RNA-sek suudab transkriptsiooni äärealasid määrata kuni ühe nukleotiidi täpsusega ning seetõttu on RNA-sek sobilik eksonite ühenduskohtade ja alternatiivse splaissingu uurimiseks. (iii) Geeni kodeeriva ala katmine sekveneerimise järjestustega võimaldab usaldusväärselt tuvastada ühe ja enama nukleotiidi variatsioone. (Cloonan et al., 2008; Morin et al., 2008).

(iv) RNA-sek meetodil on väga madal taustamüra, kuna järjestused joonduvad unikaalsetele genoomi regioonidele üheselt. Nende kvantiseerimine on äärmiselt täpne ja ilma ülempiirita, pakkudes laia dünaamilist vahemikku detekteeritud transkriptide ekspressioonile (Mortazavi et al., 2008; Nagalakshmi et al., 2008). (v) RNA-sek võimaldab leida alternatiivseid promootoralasid ja mõõta nende vahendatud geeniekspressiooni (Islam et al., 2011). Uuringud näitavad, et lisaks sellele on RNA-sek tehniliselt hästi korratav, tulemused on bioloogiliselt sarnased ning usaldusväärsed (Cloonan et al., 2008; Nagalakshmi et al., 2008).

## 1.2 Ühe raku RNA-sek

Rakkude keskmistamisel ehk rakumassi totaalse RNA analüüsiga võrreldes võimaldab individuaalsete rakkude transkriptoomi uurimine hinnata üksiku raku geeniekspressiooni, mõõta rakkudevahelist varieeruvust ning eristada rakutüüpe ühtses koes. Geeniekspressiooni heterogeensus uuritavate üksikute rakkude vahel pakub uut informatsiooni näiteks raku elutsükli, funktsiooni ja ravimitundlikkuse kohta. Rakupopulatsiooni heterogeensust uuritakse mRNA ja proteiini tasemel nii erinevates arengufaasides kui ka erinevates füsioloogilistes tingimustes (Buganim et al., 2012; Bumgarner et al., 2012; Chang et al., 2008; Toyooka et al., 2008).

Koed või teised rakulised kogumid on heterogeensed isegi siis, kui uuritavad rakud on ühest tüübist. Näiteks toimub immuunrakkude korral raku tüübi määramine ekspresseerunud pinnamarkerite põhjal, sest antud rakkude puhul on täheldatud katkematuid ekspressioonivahemikke ja pinnamarkerite üsna piiramatut kombinatoorikat ekspressioonil (Bendall et al., 2011). See tähendab, et piirid rakutüüpide vahel on hägusad ja iga individuaalne rakk on erinev. Sarnane dünaamika esineb ka kasvaja rakkudes ning on teada, et kasvajakud koosnevad heterogeensetest rakutüüpidest (Shackleton et al., 2009) ja rakud võivad üksteisest märkimisväärselt erineda ka genoomi poolest (Navin et al., 2011).

Kasvajaraku heterogeensus on seletatav akumulatsioonidega mutatsioonidega genoomis, sest isegi geneetiliselt identsetel rakkudel, mis asuvad pealtnäha identsetes tingimustes, on geeni- ja proteiini ekspressioonis kõrge varieeruvus. Sellist nähtust defineeritakse kui keskmist varieeruvust või standardhälvet ning nimetatakse „müraks“ (Paulsson, 2005). „Müra“ põhjustele suunatud uuringute tulemusel on leitud, et see tuleneb peamiselt

juhuslikest efektidest, mis on põhjustatud uuritavate molekulide vähesusest, mida omakorda võimendab analüüsimisel kasutatavad amplifikatsiooni etapid (Maheshri & O'Shea, 2007).

Standardsete ekspressiooni analüüsimeetodite rakendamisel saavutatakse uuritavate rakkude keskmine ekspressioonitase. Üksikute rakkude jaotus demonstreerib, et ainult väike hulk neist ekspresseerib mRNA-d sarnaselt keskmisega. Võib juhtuda, et just keskmisest erinevad rakud on funktsionaalselt olulised. Seevastu tavapärasel, keskmistatud rakkude analüüsil, kus kasutatakse eelnevalt puhastatud rakkude totaalset RNA-d, ei ole neid võimalik tuvastada (Munsky et al., 2012).

Ühe raku geeniekspressiooni analüüs ei ole iseenesest midagi uut (Eberwine et al., 1992). Varem on rakulist heterogeensust mõõdetud põhiliselt vaid teatud sihtmärkgeenide suhtes, kasutades ühe raku reaalkva kvantitatiivset PCR meetodit. Geeniekspressiooni ulatuslikuks analüüsiks sobilikud tavapärased RNA-sek protokollid nõuavad enam kui mikrogrammi totaalset RNA-d, millest mRNA-d on vaid mõned protsendid (Mortazavi et al., 2008), mis omakorda vastab sadadele tuhandetele imetajarakkudele. Seetõttu ei ole antud protokollid ühe raku puhul kasutatavad. Lisaks individuaalsete rakkude heterogeensuse uurimisele on ühe raku geeniekspressiooni laialdane analüüsimine traditsioonilistel meetoditel piiratud ka rakkude vähesuse korral. Näiteks embrüonaalsete rakkude uurimisel, kus rakkude arv on limiteeritud ning isegi *in vitro* kasvatatud tüvirakkude puhul on rakkude arv sageli limiteeritud. Seetõttu on antud olukordades vaja tundlikumat RNA analüüsimeetodit, mis oleks võimeline töötama ühe raku tasemel.

Üksiku raku analüüsimiseks on välja töötatud tänapäevased meetodid, mis võimaldavad geeniekspressiooni usaldusväärset uurimist väga madalal totaalse RNA koguse juures (hinnanguliselt 5-20 pg), rakendades moodsaid sekveneerimise tehnoloogiaid (Tang et al., 2010; Tang et al., 2011a). Järgnevalt tutvustatakse detailsemalt ühe raku RNA-sek meetodeid.

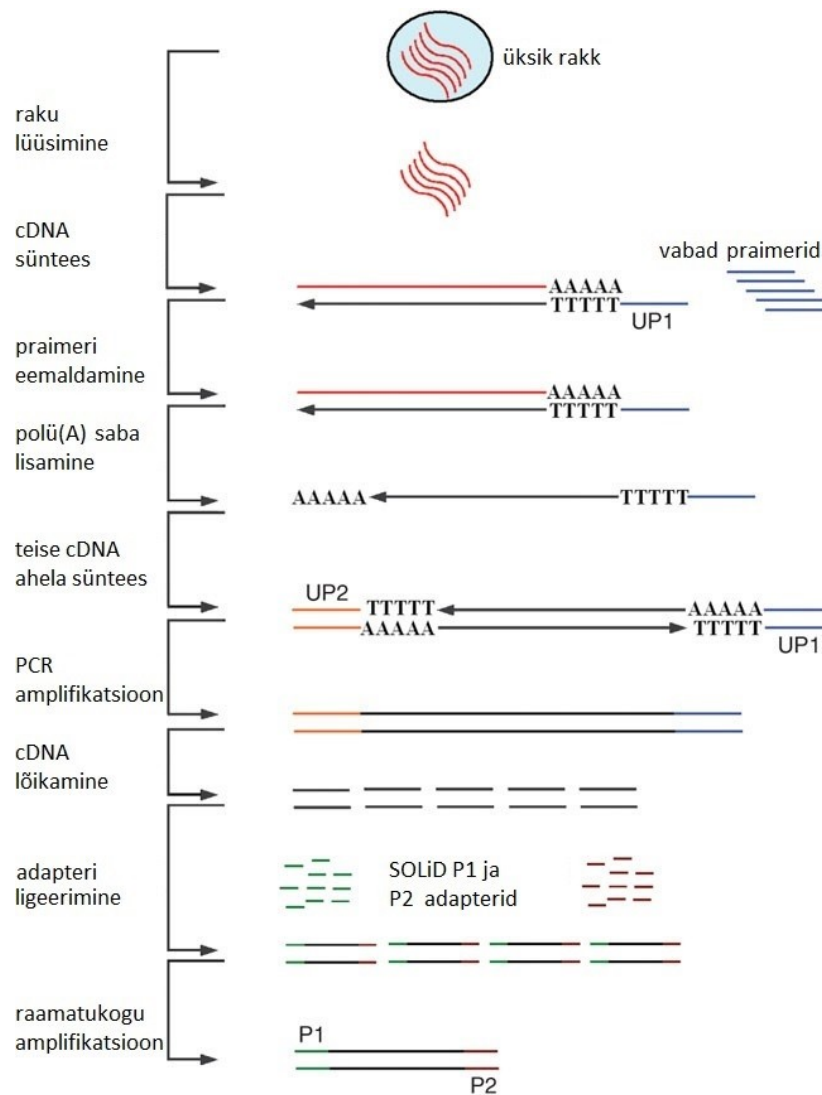
### **1.2.1 Tang'i meetod**

Esimene RNA-sek meetodil põhinev uurimus, kus analüüsiti neljarakulise embrüo staadiumis eraldatud üksikuid blastomeere, publitseeriti juba 2009. aastal (Tang et al., 2009). Tang'i protokoll algab üksiku raku isoleerimisega, millele järgneb raku lüüsimine ja polüadenüleeritud mRNA pöördtranskriptsioon ja cDNA süntees. Selleks kasutatakse polü(dT) praimereid, mis sisaldavad ankurjärjestust. Pöördtranskriptsioonil kasutamata jäänud

praimerid lagundatakse eksonukleas I toimetel. Sünteesitud cDNA 3' otsa lisatakse polü(dA) saba terminaalse desoksünukleotidüül transferaasi vahendusel, et varustada cDNA mõlemad otsad universaalse, kuid omavahel erineva ankurjärjestusega. Järgnevalt viiakse läbi teise cDNA ahela süntees, mille genereerimiseks kasutatakse polü(dT) praimereid, kuid antud juhul erineva ankurjärjestusega. Primerid kinnituvad polü(dA) pikendusele ja initsieerivad teise cDNA ahela sünteesi. Tulemuseks on kaheaahelised cDNA molekulid, mille mõlemas otsas on erinevat tüüpi ankurjärjestused. Lisaks on modifitseeritud PCR-i praimerite 5' otsi, millele on lisatud aminorühm, et kaheaahelise cDNA fragmentide 5' otsadele ei liigeks tulevase sekveneerimise raamatukogu valmistamise käigus mittesoovitavad järjestused, vähendades seeläbi molekulaarsetel manipulatsioonidel saadud kõrvalproduktide hulka. Järgnevalt toimub kaheaahelise cDNA amplifikatsioon, kus primerid seonduvad eelnevalt lisatud ankurjärjestustele. On oluline teada, et teise cDNA ahela süntees algab sealt, kus esimese cDNA ahela süntees lõppes, mistõttu sõltub kogu protsess pöördtranskriptaasi edukusest ning mRNA terviklikkusest (Hebenstreit, 2012).

Amplifikatsioonile järgnevad ühe raku cDNA sekveneerimise raamatukogu ettevalmistamiseks vajalikud protseduurid. Selleks lagundatakse amplifitseeritud cDNA ultraheliõõtluse abil umbes 80-130 bp fragmentideks, millele järgneb kleepuvate otste tõmbistamine ja adenüleerimine. Kaheaahelalisteks taastatud otstele liigeks sekveneerimiseks vajalikud adapterid, millele järgneb orienteeruvalt 150-200 bp pikkuste raamatukogu fragmentide eraldamine kasutades geel-elektroforeesi. Järgneb teistkordne PCR amplifikatsioon eesmärgiga parandada ja korrastada olemasolevaid molekule ning amplifitseerida vajalik kogus materjali järgneva sekveneerimiseks.





**Joonis 1. Tang'i meetodi skemaatiline ülevaade.** Isoleeritud rakk lüüsitakse. mRNA ahelate põhjal pöördtranskribeeritakse cDNA ahelad, kasutades UP1 ankurjärjestusega polü(dT) praimerit. Kasutamata praimerid lagundatakse. Pöördtranskribeeritud cDNA ahelate 3' otsa lisatakse polü(dA) saba, millele kinnitub UP2 ankurjärjestusega polü(dT) praimer ja genereeritakse kaheaahelalised cDNA molekulid. cDNA molekulid amplifitseeritakse PCR-i abil, kasutades UP1 ja UP2 primereid. Amplifitseeritud cDNA molekulid fragmenteeritakse ja otstesse ligeeritakse P1 ja P2 adapterid. Lõpuks viiakse läbi emulsioon-PCR, kus raamatukogudele lisatakse mikrokerad, mille pinnale on kovalentselt seotud P1 praimerid. Joonis modifitseeritud Tang et al., 2009 järgi.

## 1.2.2 SMART-il põhinevad meetodid

Kaks järgnevat ühe raku RNA-sek protokollid põhinevad SMART (ingl *switching mechanism at the 5' end of the RNA transcript*) meetodil (Zhu et al., 2001) ja kasutavad matriitsi ümberlülituse (ingl *template-switching*) mehhanismi (Schmidt & Mueller, 1999). Meetod kasutab Moloney hiire leukeemia viiruse (MMLV) pöördtranskriptaasile iseloomulikku omadust, millega cDNA ahela 3' otsa lisatakse mõned matriitsahelast

sõltumatud nukleotiidid. Kõige tõenäolisemalt lisandub kolm tsütosiini (Zhu et al., 2001). Kohe pärast cDNA ahela sünteesi lõppu kinnitub üleulatuvate tsütosiinide külge komplementaarse (rG)<sub>3</sub> 3' otsaga oligonukleotiid, mille 5' otsas on hilisemaks amplifikatsiooniks vajamineva praimeriseondumise universaalne järjestus. Seejärel jätkub cDNA süntees, kuid edasist matriitsahela funktsiooni täidab eelnevalt kinnitunud kolme riboosi tsütosiiniga varustatud oligonukleotiid. Selline molekulaartehnoloogiline võte võimaldab mRNA 5' otsa lisada soovitud järjestusi ning tekitada seeläbi sekveneerimiseks vajalikke raamatukogusid (Hebenstreit, 2012).

### 1.2.2.1 STRT meetod

STRT on esimene SMART meetodil põhinev ühe raku RNA-sek protokoll (Islam et al., 2011; Islam et al., 2012). STRT meetod näeb ette uuritavate rakkude isoleerimist eraldi proovideks, millele järgneb rakkude lüüsimine väga kiire külmutamise tulemusel pöördtranskriptaasi puhvril. Seega sisaldab iga proov ühest rakust pärinevat totaalset RNA-d ning mRNA-d ei ole tarvis lüüsiisid isoleerida.

Lüüsimisele järgneb pöördtranskriptsioon, mille tulemusena genereeritakse polüadenüleeritud mRNA ahelatele vastavad cDNA järjestused. Selleks kinnitub polüadenüleeritud mRNA sabale oligo(dT) praimer, mille 5' ots on matriitsahela ümberlülitumise vältimiseks biotinüleeritud ja 3' ots sisaldab VN-järjestust, kus „N“ sümboliseerib kõigi nelja aluse segu ja „V“ adeniini, tsütosiini või guaniini segu. Viimane on vajalik, et praimer paarduks vahetult polü(A) saba algusele. Praimer sisaldab lisaks veel universaalset järjestust ja Sall-HF restriktiooni lõikekohta, et elimineerida mRNA 3' sekveneerimise järjestusi. Järgneb cDNA ahela elongatsioon, millega kantakse kogu polüadenüleeritud mRNA järjestus üle cDNA ahelaks.

Polüadenüleeritud mRNA pöördtranskriptsioon toimub SMART meetodi ja matriitsi ümberlülitusmehhanismi alusel (Schmidt & Mueller, 1999; Zhu et al., 2001), kus MMLV pöördtranskriptaasile omaselt lisatakse cDNA ahela 3' otsa eelnevalt mainitud kolm tsütosiini, millega paardub komplementaarse (rG)<sub>3</sub> 3' otsaga oligonukleotiid, millest kujuneb algse mRNA ahela 5' otsa pikendus. Sellega vahetab pöördtranskriptaas RNA-matriitsi DNA-matriitsi vastu. Sama oligonukleotiidi abil viiakse tekkinud konstrukti spetsiifiline indeksjärjestust ning universaalne praimerjärjestust. Indeksjärjestused võimaldavad hiljem, pärast sekveneerimist saadud järjestusi seostada individuaalsete rakkudega, lubades

analüüsida kuni 96 üksikut rakku ühes raamatukogus. Nimetatud oligonukleotiid sisaldab endas ka uratsiili jääke, mis võimaldavad mRNA ning DNA oligonukleotiidi konstrukti ensümaatilist lagunemist enne amplifikatsiooni etappi.

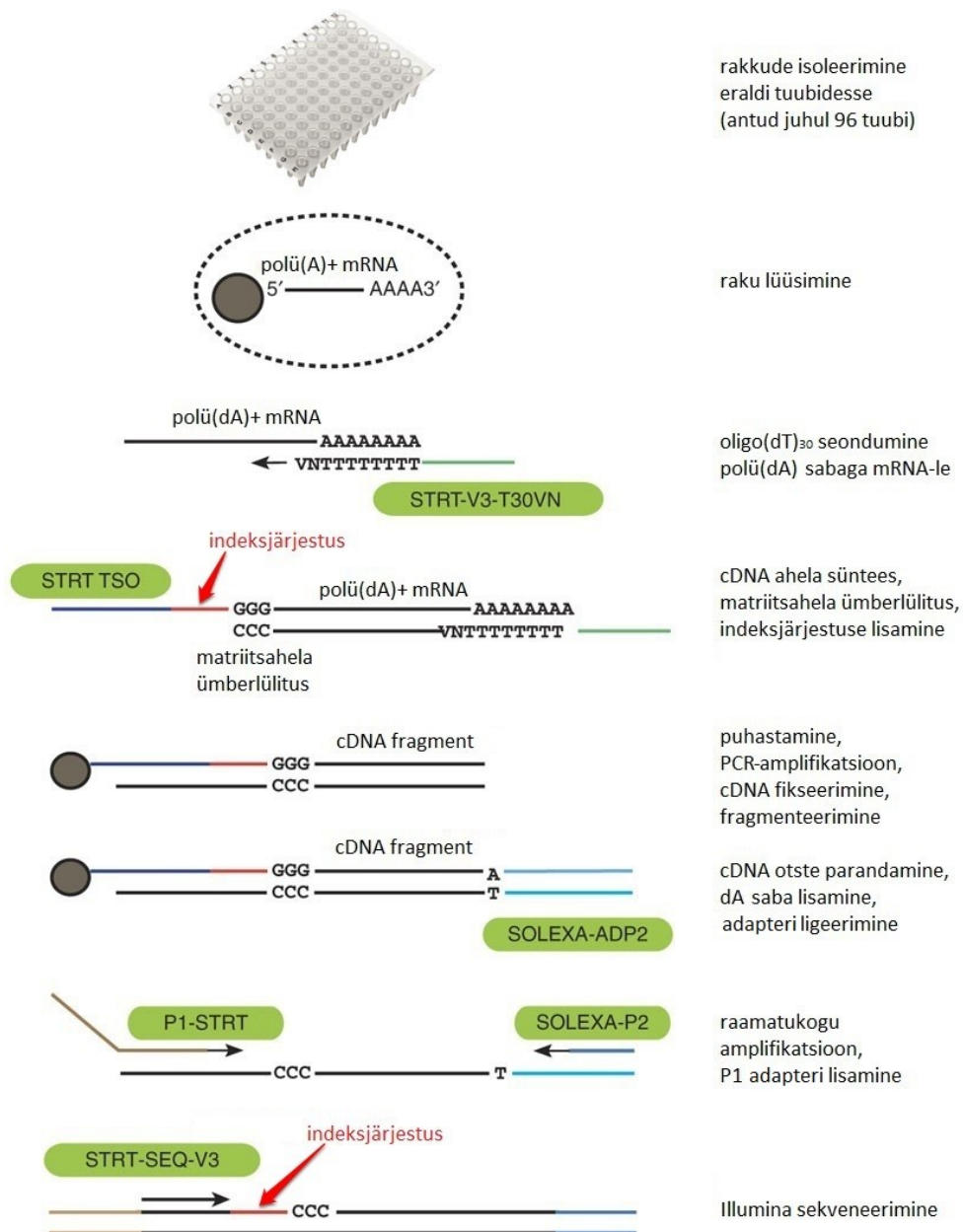
Komplementaarse DNA sünteesi lõppemisel puhastatakse reaktsioon kõrvalisest rakumaterjalist ja eemaldatakse reageerimata jäänud praimerid, kasutades magnetiseeritud ja karboksüleeritud mikrokerakesi. Nii rakkude lüüsimine, pöördtranskriptsiooni reaktsioon kui ka täispika cDNA konstrukti süntees viiakse läbi 96-anumalises analüüsiplaadis. Mikrokeradega seondunud cDNA ahelad koondatakse kokku ühte proovi ja edasised etapid toimuvad uuritavate rakkude puhul ühe raamatukoguna ühes reaktsioonituubis.

Puhastatud cDNA läbib seejärel PCR amplifikatsiooni. Kasutatakse praimerit, mis on komplementaarne varasemalt cDNA 3' otsa lisatud universaalse praimerjärjestusega ja samal ajal identne cDNA 5' otsaga, mis on polü(dT) oligonukleotiidi elongatsioonil saanud konstrukti universaalse järjestuse. Seeläbi amplifitseeritakse cDNA vaid ühte tüüpi praimeriga. Lisaks on praimerit 5' otsa kinnitatud biotiin, mida kasutatakse järgneval etapil produkti puhastamiseks. Selleks immobiliseeritakse amplifitseeritud cDNA kaksikahelad streptavidiini mikrokeradele, kus kõrge afiinsusega biotiin-streptavidiini kompleks võimaldab spetsiifilist cDNA puhastamist nii vahetult pärast PCR-i kui ka pärast teisi ensümaatilisi manipulatsioone.

Amplifitseeritud ja puhastatud cDNA raamatukogu fragmenteeritakse ensümaatiliselt. Tekkinud fragmentide otsad läbivad parandusreaktsiooni, millega tõmbistatakse lõikekohad, ning 3' ots adenüleeritakse, millele ligeeritakse järgnevalt sekveneerimiseks vajalik Illumina<sup>1</sup> adapter. Lisaks, kasutades Sall-HF restriктаasi, eraldatakse cDNA kaksikahelatest polüadenüleeritud mRNA 3' otsa iseloomustavad järjestused. Geeni alguse ehk 5' otsa jaoks vajalik adapter kuulub PCR-i praimerit järjestusse ning raamatukogu amplifikatsiooni etapi kaudu saab adapterist cDNA osa. Produktid sekveneeritakse ainult transkriptide 5' otste poolt. Saadud ahelad algavad indeksjärjestusega, millele järgneb 3-6 tsütosiini ja mRNA-le vastav fragment. See tähendab, et transkriptid säilitavad omavahelise suhte ka pärast sekveneerimist ja kõik järjestused joonduvad geeni promotori regiooni.

---

<sup>1</sup> Illumina, Inc., San Diego, CA, USA



**Joonis 2. STRT meetodi skemaatiline ülevaade.** Rakud isoleeritakse plaadi tuubidesse ja lüüsitakse. mRNA ahelate põhjal pöördtranskribeeritakse cDNA ahelad, kasutades STRT ankurjärjestusega polü(dT) praimerit. cDNA ahelate üleulatuvale 3' otsale paardub STRT ankur- ja indeksjärjestusega matriitsahela ümberlülituse oligonukleotiid. Oligonukleotiidiga seondumata ahelad eemaldatakse ja seondunud ahelad amplifitseeritakse PCR-i abil, kasutades STRT praimerit. Amplifitseeritud cDNA molekulid fragmenteeritakse, kleepuvad otsad tõmbistatakse, adenüleeritakse ja neile kinnitub P2 adapter. Raamatukogu amplifitseeritakse PCR-i abil, kasutades P1 adapteriga seotud STRT praimerit. Lõplik raamatukogu sekveneeritakse algse mRNA 5' otsa poolt. Joonis modifitseeritud Islam et al., 2012 järgi.

### 1.2.2.2 Smart-Seq meetod

Smart-Seq on teine SMART meetodil põhinev ühe raku RNA-Seq protokoll (Ramsköld et al., 2012). Olgugi et Smart-Seq publitseeriti aasta hiljem kui STRT protokoll,

on tegemist siiski STRT lihtsustatud versiooniga, mida demonstreeriti melanoomiga patsiendi verest eraldatud tsirkuleerivate kasvajakude analüüsil.

Smart-Seq meetodi puhul lüüsitakse iga uuritav rakk hüpotoonilises RNAasivabas lahuses, mis RNA stabiliseerimiseks sisaldab lisaks RNAasi inhibiitoreid. Järgneb polüadenüleeritud mRNA pöördtranskriptsioon otse rakulahuses. Pöördtranskriptsioon initsieeritakse mRNA polü(A) otsale seonduva oligo(dT) praimeriga, mille 3' otsas on nukleotiididevahelise fosfodiesterideme loomiseks vajalik hüdroksüülrühm. Järgnevalt kinnitub RNA-praimer dupleksile MMLV pöördtranskriptaas ja algab cDNA süntees. Kui cDNA elongatsioon jõuab matriitsahela 5' lõppu, siis toimub SMART tehnoloogial põhinev matriitsahela ümberlülitus. MMLV pöördtranskriptaasi terminaalse transferaasi aktiivsuse tõttu lisatakse värskelt sünteesitud cDNA 3' otsa mõned matriitsahelast sõltumatud tsütosiinid. Üleulatuvale tsütosiinidest moodustatud otsale paardub komplementaarse (rG)<sub>3</sub> 3' otsaga disainitud oligonukleotiid, millest kujuneb algse polüadenüleeritud mRNA ahela pikendus. Polüadenüleeritud mRNA pöördtranskribeerimine cDNA-ks jätkub kuni lisandunud oligonukleotiidi 5' lõpuni. Tulemusena sünteesitakse täispikk cDNA, mis sisaldab täielikku, kuid komplementaarset polüadenüleeritud mRNA järjestust koos oligonukleotiidide poolt tutvustatud otsmiste järjestustega, mis on universaalse praimeri paardumiskohaks teise ahela sünteesil ja PCR-amplifikatsioonil.

Universaalse PCR-i praimeri seondumisjärjestustega varustatud cDNA amplifitseeritakse PCR-i käigus. Tsükliite arv valitakse sõltuvalt cDNA algsest kogusest nii, et kokku saaks mõni nanogramm amplifitseeritud cDNA-d. Järgneb standardse Illumina sekveneerimise raamatukogu konstrueerimine. Amplifitseeritud cDNA fragmenteeritakse akustiliselt ja üheaahelalised (kleepuvad) otsad parandatakse kaheaahelalisteks (tõmpideks). Fragmentide otsad adenüleeritakse ja nende külge ligeeritakse Illumina adapter. Järgneb PCR-amplifikatsioon, kus tsükliite arv valitakse sõltuvalt cDNA kogusest, ning sekveneerimine.

Hiljuti tehti kättesaadavaks kommertsiaalne Smart-Seq meetodil põhinev toode cDNA genereerimiseks ja amplifikatsiooniks, mida turundab Clontech<sup>2</sup> *SMARTer Ultra Low RNA Kit for Illumina sequencing* nime all. Multipleks on võimalik alles pärast adenüleerimist ja Illumina adapterite ligeerimist, mis teeb meetodi töömahukaks ning kulukaks.

---

<sup>2</sup> Clontech Laboratories, Inc., Mountain View, CA, USA

### 1.2.3 CEL-Seq meetod

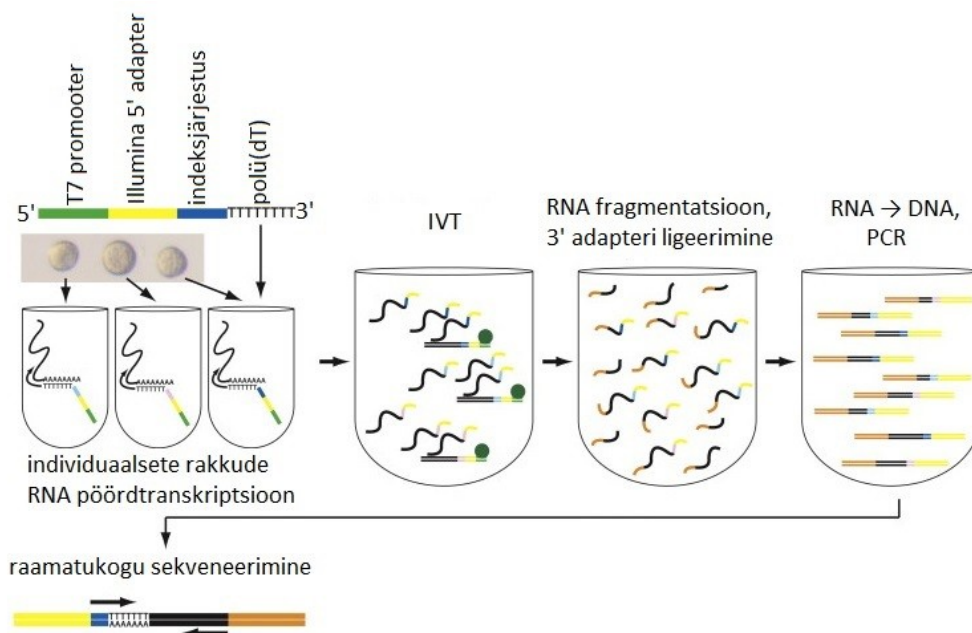
Kõrge läbilaskevõimega RNA-sek tehnoloogiad annavad ekspresseerunud geenidest detailse ülevaate, kuid analüüs nõuab palju rohkem totaalset RNA-d kui üksikust rakust on võimalik saada. Vajaliku RNA koguse saavutamiseks lisatakse analüüsile täiendav etapp – RNA amplifikatsioon. Kasutatakse eksponentsiaalset PCR-põhist või lineaarset *in vitro* transkriptsiooni (IVT) põhise amplifikatsiooni (Eberwine et al., 1992). PCR-i saab kasutada minimaalse koguse RNA amplifitseerimiseks, suurendades tsüklite arvu. Komplitseeritum on lineaarne amplifikatsioon, kuna IVT madala efektiivsuse tõttu on esmaseks amplifikatsiooni tsükliks vaja 400 pg (hinnanguliselt 50 raku materjal) totaalset RNA-d. On näidatud ka ühe raku IVT-d, kuid see on töömahukas, sest vajab kolme sõltumatut amplifikatsiooni ja nõuab ühe raku kohta viis tööpäeva (Eberwine et al., 1992). Nimetatud limitatsioon takistab IVT-põhise meetodi efektiivset rakendamist ühe raku RNA-sek analüüsil (Tang et al., 2011b).

Üksiku raku IVT-põhise amplifikatsioonile on lahenduse pakkunud CEL-Seq meetod (Hashimshony et al., 2012). Tegemist on kõrge läbilaskevõimega sekveneerimise tehnoloogial põhineva ühe raku geeniekspressiooni analüüsimeetodiga, kus pärast indeksjärjestuste lisamist jätkub individuaalsete rakkude uurimine ühises proovis. Antud meetodi efektiivsust demonstreeriti varajase embrüonaalse arengu uuringus, kus kasutati mudelorganismi, *C. elegans*'i blastomeere.

Protokoll algab individuaalsete rakkude eraldamisega üksikuteks proovideks, millele järgneb lüüsimine. Lüsaadile lisatakse pöördtranskriptsiooni praimerid, mis koosnevad neljast osast – T7 promootor, Illumina adapter, indeksjärjestus ja polü(dT). Oligonukleotiidi 5' otsas paikneb T7 promootor, kuhu kinnitub T7 RNA polümeraas, mis on vajalik IVT-ks. Järgneb Illumina 5' adapter sekveneerimiseks ja unikaalne indeksjärjestus, mille abil eristatakse rakkude transkriptid hiljem üksteisest. Indeksjärjestused on disainitud nii, et rakkude jaoks erineksid individuaalsed järjestused vähemalt kahe nukleotiidi võrra, mille tulemusel välditakse üksikutest sekveneerimisvigadest põhjustatud eksimusi. Primaeri viimase osa moodustab polü(dT), millega neljaosaline oligonukleotiid kinnitub mRNA polü(A) sabale. Pärast primaeri kinnitumist toimub pöördtranskriptsioon ja sünteesitakse cDNA ahel.

Pöördtranskriptsioonile järgneb teise ahela süntees, pärast mida segatakse individuaalsete rakkude indekseeritud cDNA üheks prooviks ja puhastatakse reageerimata jäänud praimeritest. Rakumaterjali summeerimisel saavutatakse IVT initsiatsiooniks piisav cDNA kogus. Toimub lineaarne amplifikatsioon, kus matriitsahelaks on cDNA. T7 polümeraasil on võime seonduda korduvalt cDNA 5' otsas asuvale promootorile ning viia läbi piisav amplifikatsioon edasisteks etappideks. Amplifitseeritud ja polüadenüleeritud mRNA

fragmenteeritakse osadeks, mille külge ligeeritakse Illumina 3' adapterid. Fragmenteeritud mRNA pöördtranskribeeritakse uuesti cDNA-ks. Ainult algse mRNA 3' otsa iseloomustavad cDNA järjestused sisaldavad mõlemas otsas Illumina adapterit ja läbivad PCR amplifikatsiooni, mille järel saadud raamatukogu sekveneeritakse.



**Joonis 3. CEL-Seq skemaatiline ülevaade.** Rakud isoleeritakse eraldi tuubidesse koos T7 promootorit, 5' adapterit ja unikaalset indeksjärjestust sisaldavate oligo(dT) praimeritega, mille abil mRNA ahelad pöördtranskribeeritakse cDNA ahelateks. Pärast teise ahela sünteesi proovid ühildatakse, millele järgneb IVT amplifikatsioon. Amplifitseeritud RNA fragmenteeritakse, puhastatakse ja 3' otsa ligeeritakse adapter. RNA pöördtranskribeeritakse ja fragmendid, mis sisaldavad mõlemas otsas adaptoreid, amplifitseeritakse ning sekveneeritakse. Joonis modifitseeritud Hashimshony et al., 2012 järgi.

## **2 UURIMUS**

### **2.1 Töö eesmärgid**

- Anda metoodiline ülevaade siiani avaldatud ühe raku RNA-sek protokollidest. Käesoleva töö valmimise ajaks oli publitseeritud neli meetodit, mis siinkohas ka käsitlemist leiavad.
- Tuua välja ühe raku RNA-sek protokollide omavahelised sarnasused ja erinevused, kuvada laboratoorse töö mahtu, raamatukogu manipulatsioonide läbimõeldust ja keerukust ning multipleksi võimalikkust.
- Võrrelda ühe raku RNA-sek analüüsi tulemusi, anda hinnang andmete iseloomule ning tuua välja aspektid, millistele bioloogilistele küsimustele saab ühte või teist meetodit kasutades vastata.
- Arutleda, miks on ühe raku geeniekspressiooni uurimine oluline, millised puudused hetkel valitsevad ja nende võimalikud lahendused ning mis suunas toimub arendamine.

### **2.2 Materjal ja metoodika**

#### **2.2.1 Märksõnad**

Geeniekspressioon, oligo(dT), polü(A), mRNA, cDNA, indeksjärjestus, MMLV, SMART matriitsahela ümberlülituse meetod, RNA-sek, ühe raku RNA-sek, Tang'i meetod, STRT meetod, Smart-Seq meetod, CEL-Seq meetod, PCR, IVT, Illumina, sekveneerimine.

#### **2.2.2 Andmete analüüsi meetodid**

Käesolev referatiivne uurimustöö on koostatud kirjandusepõhisel võrdleva analüüsi meetodil. Antud töö põhineb teaduslikel artiklidel, mis on leitud otsingumootori Google Scholar (<http://scholar.google.com>) ning RNA-Seq Blog (<http://www.rna-seqblog.com/>) vahendusel.



### 2.2.3 Uuritavad objektid

Eksperimentaalsete ühe raku RNA-sek protokollide esmasel demonstreerimisel kasutatud rakud ja kontrollid:

- Tang – hiire blastomeerid, küpsed ootsüüdid.
- STRT – hiire embrüonaalsed tüvirakud, hiire embrüonaalsed fibroblastid ja inimese aju referentsi RNA.
- Smart-Seq – melanoomipatsiendi verest eraldatud arvatavad tsirkuleerivad kasvajakud, inimese ajurakud, hiire ajurakud, universaalne inimese referentsi RNA, hiire ootsüüdid, eesnäärmevähi rakuliini rakud, põievähi rakuliini rakud, primaarsed melanotsüüdid, melanoomi rakuliini rakud, inimese embrüonaalsed tüvirakud.
- CEL-Seq – *C. elegans*'i embrüo tütararakud, hiire embrüonaalsed tüvirakud ja hiire embrüonaalsed fibroblastid.

## 2.3 Tulemused ja arutelu

RNA-sek võimaldab geeniekspressiooni taseme kvantitatiivset analüüsi, iseloomustab täispika transkripti igat nukleotiidi ning võimaldab määrata kummalt DNA ahelalt on mRNA transkribeeritud. Olenevalt ühe raku RNA-sek avaldatud protokollidest ei ole need eesmärgid olemasolevate meetoditega samaaegselt saavutatavad.

Publitseeritud on neli erinevat ühe raku RNA-sek strateegiat. Esimene uuring, Tang'i meetod, avaldati 2009. aastal ja seal rakendati modifitseeritud versiooni eelnevalt laialdaselt kasutatud ühe raku transkriptoomi amplifikatsiooni meetodist (Kurimoto et al., 2006; Kurimoto et al., 2007). Järgnevad kaks ühe raku RNA-sek strateegiat, STRT ja Smart-Seq, on olemuselt sarnased, sest mõlemad lähtuvad SMART-il põhineva matriitsahela ümberlülituse mehhanismist. Neljas ühe raku RNA-sek protokoll, CEL-Seq, on publitseeritud 2012. aastal ja erinevalt varasematest meetoditest, rakendatakse sellel IVT-põhist amplifikatsiooni. Uurimuse järgnevas osas tuuakse välja nelja nimetatud ühe raku RNA-sek meetodi põhimõttelised erinevused ja võrreldakse nende mõõdetavaid karakteristikuid.

### 2.3.1 Protokollide sarnasused ja erinevused

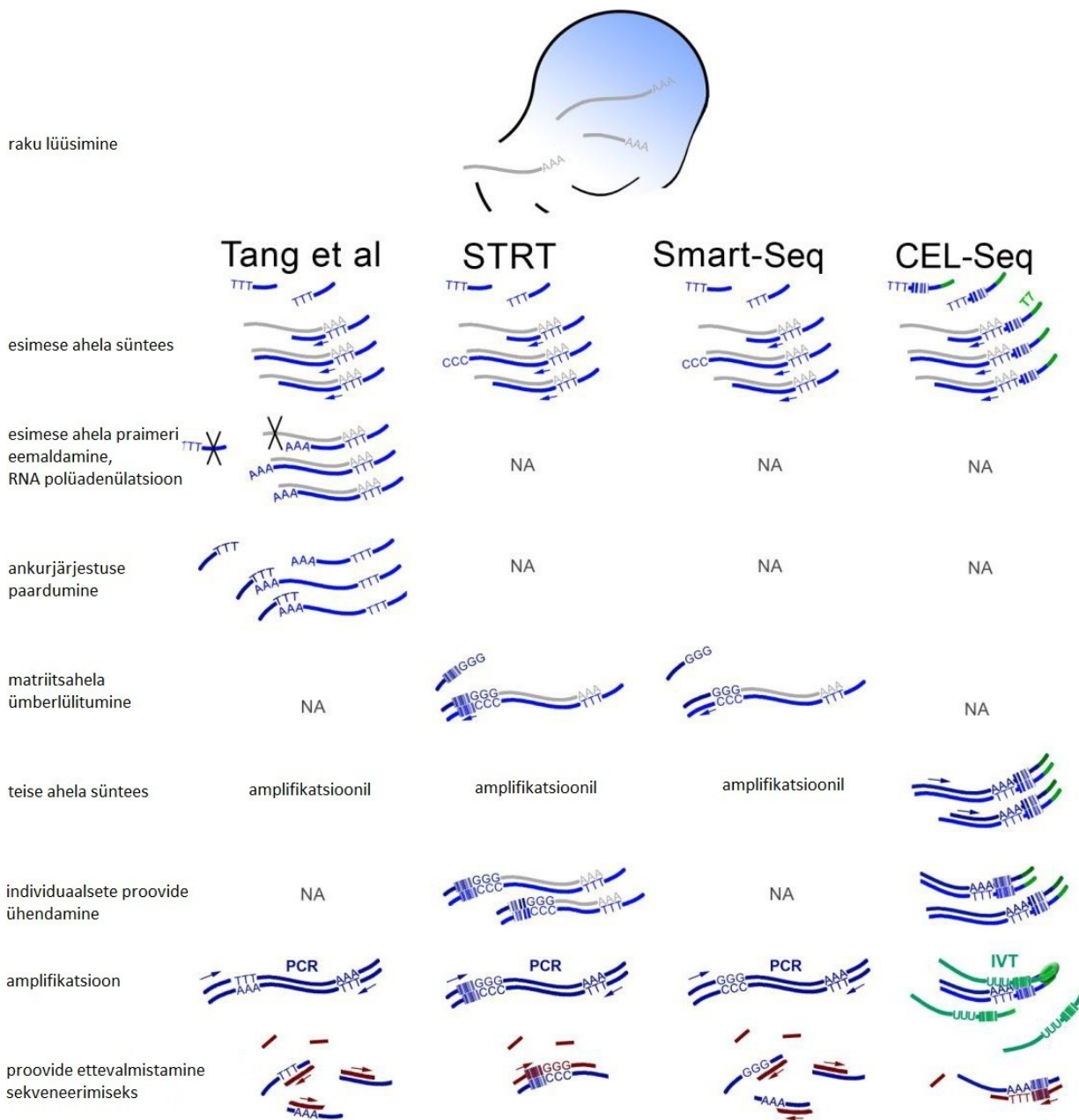
Originaalartiklitele põhinedes on keeruline võrrelda olemasolevate ühe raku RNA-sek meetodite tõelist suutlikkust ning peab piirduma silmnähtavate erinevuste nagu näiteks ahela spetsiifilisus, transkriпти katvus, multipleksi võimalikkus ja laboratoorse töö maht. Autorid kasutasid suutlikkuse hindamiseks küllaltki erinevaid viise – võrreldi geenikiibi, sisemiste kontroll-RNA-de, algse RNA massi tiitrimise ja teiste süsteemidega. Samuti kasutati erinevaid parameetreid nagu PCR-i tsüklite arv, millel võib olla tugev efekt lõppresultaadile. Hea oleks võrrelda kõiki meetodeid üksteisega identsetes tingimustes, mida on varasemalt kasutatud tavapäraste RNA-sek protokollide puhul (Levin et al., 2010).

Kõigi nelja ühe raku RNA-sek meetodi puhul kasutatakse proovi rikastamiseks oligo(dT) praimereid, mis seonduvad mRNA polü(A) sabale. Positiivseks küljeks on, et bioloogiliselt informatiivset mRNA-d ei pea lisaetapina üleliigsest rakumaterjalist puhastama, vaid reaktsioonidega jätkatakse otse lüsaadis. Negatiivne on, et polüadenüleerimata mRNA-d, näiteks histoone kodeerivad geenid (Marzluff et al., 2008), jäävad uuringust välja. Alternatiivina kasutatakse ka totaalset RNA-st suurima osakaaluga ribosomaalse RNA eemaldamist, kuid see pole nii spetsiifiline kui polü(dT) praimeri kasutamine. Rikastamine on vajalik, sest huvipakkuva mRNA osakaal moodustab raku totaalset RNA-st vaid mõne protsendi. Vastasel juhul jääks mRNA ülejäänud RNA varjus vähemärgatavaks.

Kõigi meetodite puhul kasutatakse amplifikatsiooni kahes etapis, sealhulgas sekveneerimise tehnoloogiale spetsiifilise proovi ettevalmistamisel. Teistest meetoditest erineb CEL-Seq, kus esimene amplifikatsioon põhineb *in vitro* transkriptsioonil. IVT eelis seisneb amplifikatsiooni lineaarses toimemehhanismis, mis erinevalt PCR-ist ei kahanda eksponentsiaalselt järjestusi, mida on raske töödelda (Hashimshony et al., 2012). T7 polümeraasi võime korduvalt seonduda promootorile lubab saavutada märkimisväärset amplifikatsiooni samas reaktsioonis, ilma et oleks vaja denatureerida ja praimerit uuesti siduda. Samas nõuab IVT-põhine amplifikatsioon CEL-Seq meetodilt kahte pöördtranskriptsiooni – algne mRNA teisendatakse cDNA-ks, siis tagasi komplementaarseks RNA-ks ja lõpuks jälle cDNA-ks. Tõenäoliselt toob mitmekordne ahela teisendamine kaasa lisahälbeid.

SMART-il põhinevad STRT ja Smart-Seq meetodid kasutavad mõlemad matriitsahela ümberlülituse mehhanismi, mille vahendusel lisatakse mRNA 5' otsa soovitud universaalne praimerjärjestus. Ümberlülituse mehhanismiga välditakse ka puudulikult sünteesitud ahelate analüüsimist, mis on omane Tang'i meetodile. Samas ei tähenda see, et transkriпти katvus oleks parem tänu sellele, et kogu pikkuses cDNA-d sünteesitakse sagedamini. Vastupidiselt

on ebatäielikult sünteesitud cDNA-l väiksem tõenäosus jätkata protokollis edasistes etappides. Seejuures on negatiivne, et transkriptid, mis võiks vähemalt osaliselt panustada eksperimendil kogutud andmetesse, jäetakse arvestamata (Hebenstreit, 2012).



**Joonis 4. Nelja ühe raku RNA-sek meetodi erinevuste skemaatiline võrdlus.** Tegemist on lihtsustatud skeemiga, mis keskendub protokollide omavahelistele erinevustele ja toob välja nende iseloomulikud omadused. Tang'i protokollis kinnituvad adeniinnukleotiidid esimese ahela 3' lõppu, kus neid kasutatakse ankur-oligonukleotiidide seondumissaitidena ja RNA ning esimese cDNA ahela praimerid eemaldatakse ensümaatilisel (mustad ristid). STRT ja CEL-Seq sisaldavad proovide ühendamisetappi, mis arvestades algseid transkripte, baseerub vastavalt esimese cDNA ahela 5' ja 3' regioonide indeksjärjestustel. CEL-Seq kasutab esimesel amplifikatsiooni etapil IVT-d, mis nõuab otseselt RNA adapteri ligeerimist, millele järgneb proovi sekveneerimiseks ettevalmistamisel teine pöördtranskriptsioon. Smart-Seq on STRT lihtsustatud versioon ilma indeksjärjestusteta. Kõik PCR-i praimerid kinnituvad oligonukleotiidide abil tutvustatud kindlatele järjestustele. Nooled viitavad praimerit poolt initsieeritud DNA polümerisatsioonile. AAA ja TTT järjestused on illustratiivsed ja ei viita nukleotiidide tegelikule arvule. Joonis modifitseeritud Hebenstreit, 2012 järgi.

STRT meetodi puhul lisatakse ümberlülituse mehhanismi abil ka indeksjärjestused, tänu millele saab erinevaid proove pärast cDNA sünteesi kokku segada ilma raku individuaalsuses kaotamata. Indeksjärjestust kasutab samal eesmärgil ka CEL-Seq, kuid sel juhul lisatakse indeks oligo(dT) praimeriga mRNA 3' lõppu iseloomustavale otsale. Võrreldes Tang'i ja Smart-Seq meetoditega, vähendab mitme raku analüüsimine ühtse proovina töö, aja ning raha kulu ühe raku kohta märgatavalt. Lisaks välistatakse multipleksiga individuaalsete reaktsioonide erinevustest tingitud lahknevused.

### 2.3.2 Meetodite suutlikkus

Esimestel teedrajavatel ühe raku RNA-seq meetoditel puudusid eelnevad võrdlusaspektid ja seetõttu on erinevatel alustel saadud tulemusi raske võrrelda. Hilisemates uuringutes kasutatakse eelnevate meetoditega analoogseid katseid, kus võrreldakse samaliigilisi rakke ja rutiinselt on kasutusele võetud soovituslikud sisemised kontrollid, mis võimaldavad määrata algse mRNA kogust ning analüüsida proove omavahel.

Sekveneerimisel sõltub transkriptide katvus paljuski protokollide eripärast. Tang meetodi korral langeb katvus eksponentsiaalselt 5' suunas. Transkriptide piirkonnad, mis on 3' otsast kaugemal kui 3 kb, on kaetud väga limiteeritult (Tang et al., 2009).

STRT meetodil sekveneeritakse mRNA 5' otsa usaldusväärselt vähemalt 2 kb pikkustel transkriptidel (vastavalt pikimale kontrollile), kuid märkimisväärset kadu ei olnud ka pikemate transkriptide hulgas. Tänu 5' otsa järjestamisele saab määrata ka täpse transkriptsiooni alguskoha ja võimaliku alternatiivse promootori. Kontroll-RNA korral jäi 85% järjestustest esimesse 5% RNA pikkusest, millest peaaegu kõik joondusid transkripti 5' otsa mõne nukleotiidilisse vahemikku. Endogeense mRNA korral oli näha bimodaalset jaotust, kus lisaks 5' maksimumile kerkis esile ka transkriptide 3' transleerimata piirkond. Antud nähtus vajaks edasist uurimist. (Islam et al., 2011).

Erinevalt STRT strateegiast ei kasuta Smart-Seq meetod indeksjärjestusi ja sarnaselt Tang'i meetodile kaotatakse fragmenteerimisel *sense*- ja *antisense* DNA ahela informatsioon. Ahela informatsioon puudumine raskendab järjestuste joondamist transkriptsioonilistele üksustele, kuna vastasahelatel paiknevad geenid võivad kattuda ja 5' ning 3' transleerimata piirkonnad on sageli eelnevalt ebapiisavalt kirjeldatud, kuid autorite sõnul paraneb transkriptide katvus märgatavalt üle kogu pikkuse – ligi 40%-line katvus transkriptide 5' otsas. Samuti on joondunud järjestused keskmiselt pikemad kui Tang'i meetodi korral

(Ramsköld et al., 2012). Võrreldes teiste meetoditega, on Smart-Seq strateegial parem katvus üle kogu transkripti, mis tõstab alternatiivsete transkripti isovormide detailse analüüsi kvaliteeti ja võimaldab SNP-de identifitseerimist.

CEL-Seq meetodiga säilitatakse järjestuste ahela informatsioon ja sarnaselt Tang'i meetodile langeb katvus 5' suunas. 90% sekveneerimisel saadud järjestustest joondub 3' otsa, millega kaetakse ligikaudu esimesed 500 bp (Hashimshony et al., 2012).

Esimesel eksperimentaalsel ühe raku RNA-sek meetodil saadud andmeid võrreldi Affymetrix'i<sup>3</sup> geenikiibi tehnoloogiaga (Tam et al., 2008). Ühe hiire blastomeeri analüüsil Tang'i meetodiga detekteeriti 75% (5270) rohkem ekspresseerunud geene kui geenikiibiga sadade blastomeeride puhul. Samuti avastati 1027 transkripti, mille proovid geenikiibilt puudusid.

Hiire embrüonaalsete tüvirakku analüüsil detekteeriti CEL-Seq kasutades keskmiselt ligikaudu poole rohkem ekspresseerunud geene kui STRT meetodiga. Fibroblasti puhul jäi viimane alla vaid ligikaudu kümnendiku jagu. CEL-Seq meetod produtseeris kõrgemaid korrelatsioone embrüonaalsete tüvirakkude korral ning tegi võrreldes rakutüüpidel selgemini vahet siis, kui uuriti kõrgelt ekspresseerunud geene kummaski rakutüübis.

CEL-Seq edestas näitajate poolest STRT meetodit nii robustsuse, tundlikkuse kui ka reprodutseeritavuse poolest ning kannatas märgatavalt vähem tehnilise müra all. Kuigi eelnevalt nimetatud meetodit kirjeldavad erinevaid transkripti otsi ja analüüsitud rakutüüpide kasvutingimuste erinevust ei saa täielikult välistada, on siiski ebatõenäoline, et need aspektid põhjendaksid meetodite demonstreerimisel täheldatud erinevusi, kuna meetodeid võrreldi ka sisemiste kontroll-RNA-dega, mida bioloogilised faktorid ei mõjuta.

Üks laialdaselt kasutatud RNA-sek eelis geenikiibi ees on selle võime tabada splaissingu variante. Seda on demonstreeritud kõigi kolme PCR-põhise ühe raku RNA-sek protokollil puhul. Näiteks identifitseeriti Tang'i meetodiga ainult ühe blastomeeri kohta 1753 eelnevalt tundmatut splaissingu ehk kokkupõime liitekohta. See näitab RNA-sek võimet leida uusi splaissingu isovorme *de novo*. Samuti leiti 2070 uut splaissingu liidet küpse ootsüüdi analüüsimisel. Ühe raku tasemel pole seda varasemalt geenikiibiga teha suudetud. Lisaks leiti, et 8-19% geenidest ekspresseerisid kahte või enam transkripti isovormi samas raku, mis demonstreerib transkriptide varieeruvuse kompleksust individuaalsetes rakkudes kogu genoomi tasandil.

Ühe raku RNA-sek meetoditele seni lahendamata probleemiks on madalalt ekspresseerunud transkriptide kvantifitseerimine. STRT meetod detekteerib usaldusväärselt ainult kõrgelt ja keskmiselt ekspresseerunud geenid – 10 või rohkem koopiat raku kohta

---

<sup>3</sup> Affymetrix, Inc., Santa Clara, CA, USA

(Islam et al., 2012). Smart-Seq meetodi puudused avalduvad väga vähese lähte-RNA korral. Mõned madalama taseme juures ekspresseeruvad transkriptid näitavad juhuslikku kadu (Ramsköld et al., 2012). CEL-Seq meetodi puhul on molekulide arv otseselt proportsionaalne sisend-RNA-ga. Tundlikkus on üle kogu vahemiku lineaarne ja ei lange vähese algmaterjali korral. Väga madalalt ekspresseeritud geenid, kus transkriptid on esindatud nelja kuni viie koopiaga, detekteeritakse 50% tõenäosusega ja 50 koopiaga geenid praktiliselt alati sõltumata analüüsitud RNA hulgas (Hashimshony et al., 2012). Olukorra muudab paremaks veel mitme proovi ühendamise etapp, kus haruldased transkriptid summeeruvad ja ei jää täiesti märkamata. Lisaks on võimalik, et mõningate madalalt ekspresseeritud geenide puhul on avaldumine juhuslik – kord avaldub, kord mitte. Transkriptid, mis ekspresseeruvad madalatel tasemetel võivad kodeerida väga olulisi proteiine, mis on doositundlikud ja mida seetõttu hoitakse madala taseme juures. Meetodite edasiarendused võiks tõsta tundlikkust veelgi, et analüüsida ka madala avaldumisega geene.

**Tabel 1. Ühe raku RNA-Sek meetodite võrdlus.** Andmed pärinevad (Hashimshony et al., 2012; Islam et al., 2011; Islam et al., 2012; Ramsköld et al., 2012; Tang et al., 2009).

	<b>Tang</b>	<b>STRT</b>	<b>Smart-Seq</b>	<b>CEL-Seq</b>
Sihtmärk RNA	Polüadenüleeritud mRNA			
Proovi rikastamine	PCR			IVT
Raamatukogu amplifikatsioon	PCR			
Sekvenerimise platvorm	Applied Biosystems SOLiD	Illumina		
Multipleksi võimalikkus	Puudub	Olemas	Puudub	Olemas
Indeksjärjestus	Puudub	5' otsas	Puudub	3' otsas
Ahela informatsioon	Ei säili	Säilib	Ei säili	Säilib
Matriitsahela ümberlülitamine	Ei kasutata	SMART tehnoloogia		Ei kasutata
Raamatukogu ettevalmistamine	Mõned päevad	2-3 päeva	3 päeva	2-3 päeva
Raamatukogu hind	Võrreldav Smart-Seq'ga	~160 € / 96 rakku	~80 € / rakk	Võrreldav STRT-ga
Transkripti katvus	3' ots	5' ots	Kogu ulatus	3' ots
Transkripti efektiivne pikkus	Kuni 3 kb	Vähemalt 2 kb	Suurem kui Tang'il	NA
Kvantifitseerimise tundlikkus	75% rohkem geene võrreldes geenikiibiga	50% 10 koopia korral	Sarnane Tang'le	50% 4-5 koopia korral
Varieeruvus	Tehnilise varieeruvuse hindamiseks puudusid kontrollid	Bioloogiline varieeruvus ületas tehnilist igal tasemel	Bioloogiline varieeruvus väiksem kui Tang'il	Märgatavalt väiksem tehniline varieeruvus kui STRT-l
Sobilikud rakendused	3' eksonite splaiisingu isovormide määramine	Transkriptsiooni alguskoha ja alternatiivse promotori määramine	Alternatiivse splaiisingu isovormide, SNP-de ja mutatsioonide määramine	Suuremahuline kvantitatiivne analüüs
Puudused	3' otsa tendents, multipleks	Tugev 5' otsa tendents	Multipleks	Tugev 3' otsa tendents

### 2.3.3 Perspektiiv

Esimesed ühe raku RNA-sek uurimused on tõestanud seda tüüpi analüüsi teostatavust. Meetodi tegelik tugevus ilmneb tõenäoliselt alles siis, kui see võetakse kasutusele uurimaks mRNA ekspressioonitasemete jaotust individuaalsete rakkude seas ning selle karakteristikuid. Samuti on objektiivse hinnangu andmiseks tarvis enam publikatsioone teiste töögruppide poolt, kes on suutnud mõnda neist meetodist juurutada ja kasutada.

Ühe raku analüüsi täieliku potentsiaali kasutamiseks peab esmalt olema täidetud mitu tingimust. Täpse parameetri hinnang (ingl *precise parameter estimation*) transkriptsiooni regulatsiooni matemaatilistele mudelitele põhineb tavapäraselt ühe molekuli fluorestsents *in situ* hübriidisatsioonil, mida kohaldatakse vähemalt mõnesajale rakule (So et al., 2011). See teeb hädavajalikuks ühe raku RNA-sek edendamise, mida tõenäoliselt on kõige lihtsamalt teha mitme proovi ühildamise teel nagu STRT ja CEL-Seq meetodite korral. Jätkuvalt on lahendamata ka ühe raku RNA-sek kvantitatiivne täpsus. Sisemised kontrollid võimaldavad tuletada, kuidas sekveneerimisel saadud järjestuste arv transleeritakse transkriptide algseks hulgaks (Hebenstreit, 2012).

Üks viis, kuidas jälgida amplifikatsioonil põhjustatud hälvet, on kasutada teist tüüpi indeksjärjestusi, mis identifitseeriks individuaalseid molekule (Fu et al., 2011). Sama indeksjärjestuse arvukas ilmumine viitab sellele, et vastavad sekveneeritud fragmendid tulenesid kõik üksnes ühest algmolekulist. See tõstab esile PCR-iga kaasneva hälbe, mille arvesse võtmine suurendab RNA-sek täpsust (Fu et al., 2011; Kivioja et al., 2011), kuid ei kaitse molekulide täieliku kao eest. Seega oleks veel kasulik omada täiendavaid matemaatilisi mudeleid, mis kirjeldaksid proovide ettevalmistamise protsesse.

Teine strateegia meetodite efektiivsuse tõstmiseks oleks protokollide etappide vähendamine. Katse selles suunas tehti Helicos<sup>4</sup> süsteemil põhineva otsese sekveneerimise tehnoloogiaga, mis lubab cDNA sekveneerimist ilma proovi ettevalmistamiseta (Lipson et al., 2009) ning võimaldab otse RNA sekveneerimist (Ozsolak et al., 2009; Ozsolak & Milos, 2011). Antud tehnoloogia piiranguks on sekveneerimisel saadavate järjestuste madalam arv (Ozsolak & Milos, 2010). Kavandatavad tehnoloogiad nagu Pacific Biosciences<sup>5</sup> poolt arendatav meetod lubab ühe molekuli sekveneerimist ning on potentsiaalselt kohandatav ühe raku uuringuteks (Eid et al., 2009; Uemura et al., 2010).

Potentsiaali omavad ka nanopooridel põhinevad sekveneerimise tehnoloogiad, mis esindavad täiesti uut lähenemist. DNA või RNA ahel juhitakse nanomeetrise läbimõõduga

---

<sup>4</sup> Helicos BioSciences, Corp., Cambridge, MA, USA

<sup>5</sup> Pacific Biosciences, Corp., Menlo Park, CA, USA



pooride kaudu läbi membraanide. Alused, mis läbivad poore, detekteeritakse elektrooniliselt või optiliselt (Niedringhaus et al., 2011).

Hoolimata kirjeldatud edasiminekutest, ei ole ühe raku sekveneerimise rakendamine kliinilistes uuringutes majanduslikult taskukohane ning diagnostiliselt usaldusväärne. Siiski annab näiteks teadmine kasvaja heterogeensusest suurt lootust selle praktiliseks kasutuselevõtuks, mis lubaks täpsemaid diagnoose, sihtmärkidele suunatud ravi ja personaliseeritud meditsiini, parandamaks vähipatsientide kliinilist ravi. Ühe raku sekveneerimine on vaid üks molekulaar-tehnilistest meetoditest, mis omab potentsiaali nimetatud valdkondades, kuid sedagi vaid koos teiste meetodite, analüüsi protokollide ning usaldusväärsete publikatsioonide koosmõjul.

## KOKKUVÕTE

Antud referatiivne uurimustöö annab ülevaate hetke seisuga publitseeritud ühe raku RNA-sek protokollidest alates raku isoleerimisest kuni sekveneerimiseni. Välja on toodud meetodite omavahelised sarnasused ja erinevused, laskudes protokoll-spetsiifiliste etappideni nagu matriitsi ümberlülitamine, indeksjärjestuste lisamine ja *in vitro* transkriptsioon, millest tingitud erinevustest lähtudes võrreldakse meetodite suutlikkust ja sellest tulenevat andmestikku. Lisaks keskendutakse praeguste meetodite kitsaskohtadele ja tutvustatakse perspektiivikaid arendusi.

Nelja erineva ühe raku RNA-sek võrdlemisel selgus, et lisaks ühe raku kvantitatiivsele analüüsile, on kõigil meetoditel teatud piirangud olenevalt protokolleriipärasest ja kindla bioloogilise probleemi lahendamisel pole alati ühtset valikut. Kas leppida näiteks multipleksi puudumise või transkriptide lühema katvuse vahel, valida 5' või 3' otsa iseloomustava meetodi vahel, tuleb otsustada lähtudes uuringu olemusest, kuna hetkel pole kõik eemärgid samaaegselt saavutatavad.

# Gene expression analysis at the single cell level

Hindrek Teder

## SUMMARY

Gene expression analysis has been a popular tool in molecular biological research for more than a decade. Mostly implemented by microarray technology, it has led to numerous insights and discoveries. Following its recent introduction, RNA-sequencing (RNA-seq) is rapidly replacing microarrays as the method of choice for transcriptomics. Besides the superior accuracy in the quantification of expression, RNA-seq offers other advantages, such as the possibility to detect novel transcripts, splice variants or allele-specific expression.

To understand the basis and importance of heterogeneity and the stochastic aspect of gene expression, it is essential to examine transcriptomes of individual cells. Several recent studies demonstrate the feasibility of RNA-seq at the single cell level. This allows new biological insights into cell differentiation, cell-to-cell variation and gene regulation, and how these aspects depend on each other.

In this study, I give an overview of the current single cell RNA-seq experimental protocols, reveal the similarities and differences between them, bring out the special features, like template-switching, barcoding and *in vitro* transcription, and discuss the efforts, challenges and potentials.

Despite of the fact that methods were demonstrated on different cells in different conditions, I compared the results of the four single cell RNA-seq protocols and found that all the strategies have some technical limitations which define their suitability for certain applications.

Depending on the experimental protocols, the goals of “traditional” RNA-seq are not simultaneously achievable with current single cell RNA-seq methods. Artifacts and biases still exist and that needs to be identified and controlled for. The next few years will hopefully see a consensus emerge on the best analysis pipeline.

# TÄNUAVALDUSED

Avaldan tänu kõigile, kes panustasid selle töö valmimisse oma nõu ja jõuga. Eriti sooviksin tunnustada oma juhendajat Kaarel Krjutškovi, kelle käe all see uurimus teoks sai, Maria Tederit, Vidrik Tederit ja Epp Kalevistet, kes on aidanud kaasa töö viimistlemisel ning Karin ja Tõnu Tederit toetuse eest.

## KASUTATUD KIRJANDUS

- Bendall, S. C., Simonds, E. F., Qiu, P., Amir, E. D., Krutzik, P. O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., & Ornatsky, O. I. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*, 332, 687-696.
- Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., van Oudenaarden, A., & Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150, 1209-1222.
- Bumgarner, S. L., Neuert, G., Voight, B. F., Symbor-Nagrabska, A., Grisafi, P., van Oudenaarden, A., & Fink, G. R. (2012). Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Molecular Cell*, 45, 470-482.
- Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., & Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453, 544-547.
- Church, G. M. (2006). Genomes for ALL. (cover story). *Scientific American*, 294, 47.
- Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., & Bethel, G. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, 5, 613-619.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., & Coleman, P. (1992). Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences*, 89, 3010-3014.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., & Bettman, B. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133-138.

- Fu, G. K., Hu, J., Wang, P., & Fodor, S. (2011). Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proceedings of the National Academy of Sciences*, *108*, 9026-9031.
- Fu, X., Fu, N., Guo, S., Yan, Z., Xu, Y., Hu, H., Menzel, C., Chen, W., Li, Y., & Zeng, R. (2009). Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics*, *10*, 161.
- Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology*, *210*, 1518-1525.
- Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*.
- Hebenstreit, D. (2012). Methods, Challenges and Potentials of Single Cell RNA-seq. *Biology*, *1*, 658-667.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J., Lönnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, *21*, 1160-1167.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J., Lönnerberg, P., & Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5 [prime] end sequencing. *Nature Protocols*, *7*, 813-828.
- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., & Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K. D., Yamada, R. G., Ueda, H. R., & Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Research*, *34*, e42-e42.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., & Saitou, M. (2007). Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nature Protocols*, *2*, 739-752.

- Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A., & Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, 7, 709-715.
- Lipson, D., Raz, T., Kieu, A., Jones, D. R., Giladi, E., Thayer, E., Thompson, J. F., Letovsky, S., Milos, P., & Causey, M. (2009). Quantification of the yeast transcriptome by single-molecule sequencing. *Nature Biotechnology*, 27, 652-658.
- Maheshri, N., & O'Shea, E. K. (2007). Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu.Rev.Biophys.Biomol.Struct.*, 36, 413-434.
- Marzluff, W. F., Wagner, E. J., & Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly (A) tail. *Nature Reviews Genetics*, 9, 843-854.
- Morin, R. D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T. J., McDonald, H., Varhol, R., Jones, S. J., & Marra, M. A. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques*, 45, 81.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5, 621-628.
- Munsky, B., Neuert, G., & van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. *Science*, 336, 183-187.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320, 1344-1349.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., & Esposito, D. (2011). Tumour evolution inferred by single-cell sequencing. *Nature*, 472, 90-94.
- Niedringhaus, T. P., Milanova, D., Kerby, M. B., Snyder, M. P., & Barron, A. E. (2011). Landscape of next-generation sequencing technologies. *Analytical Chemistry*, 83, 4327.

- Ozsolak, F., & Milos, P. M. (2010). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, *12*, 87-98.
- Ozsolak, F., & Milos, P. M. (2011). Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdisciplinary Reviews: RNA*, *2*, 565-570.
- Ozsolak, F., Platt, A. R., Jones, D. R., Reifengerger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M., & Milos, P. M. (2009). Direct RNA sequencing. *Nature*, *461*, 814-818.
- Paulsson, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews*, *2*, 157-175.
- Ramsköld, D., Luo, S., Wang, Y., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebtkova, I., Loring, J. F., & Laurent, L. C. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, *30*, 777-782.
- Schmidt, W. M., & Mueller, M. W. (1999). CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Research*, *27*, e31-i-e31-iv.
- Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell*, *138*, 822-829.
- So, L., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., & Golding, I. (2011). General properties of transcriptional time series in *Escherichia coli*. *Nature Genetics*, *43*, 554-560.
- Zhu, Y., Machleder, E., Chenchik, A., Li, R., & Siebert, P. (2001). Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *BioTechniques*, *30*, 892-897.
- Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., & Schultz, R. M. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, *453*, 534-538.



- Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., & Surani, M. A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell*, *6*, 468-478.
- Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B. B., Heard, E., Lao, K., & Surani, M. A. (2011a). Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PloS One*, *6*, e21208.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., & Siddiqui, A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*, 377-382.
- Tang, F., Lao, K., & Surani, M. A. (2011b). Development and applications of single-cell transcriptome analysis. *Nature Methods*.
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., & Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*, *135*, 909-918.
- Uemura, S., Aitken, C. E., Korlach, J., Flusberg, B. A., Turner, S. W., & Puglisi, J. D. (2010). Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature*, *464*, 1012-1017.
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*, 470-476.
- Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I., & Marden, J. H. (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*, *17*, 1636-1647.

## KASUTATUD VEEBIAADDRESSID

Wikipedia <http://en.wikipedia.org>

Google Scholar <http://scholar.google.com>

RNA-Seq Blog. RNA-Seq from Single Cells <http://www.rna-seqblog.com/publications/rna-seq-from-single-cells/>

# LIHTLITSENTS

## **Lihlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Hindrek Teder (sünnikuupäev: 19.01.1989)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Geeniekspressiooni uurimine ühe raku tasemel“, mille juhendajad on Kaarel Krjutškov ja Ants Kurg.

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 28.05.2013