

TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
Arvutiteaduse Instituut

TAAVI TETLOV

Retseptori internalisatsioonianalüüs

Bakalaureusetöö (6 EAP)

Juhendaja: Sven Laur, PhD
Juhendaja: Andres Tover, PhD
Juhendaja: Raivo Kolde, MSc

Autor: “.....” 2012
Juhendaja: “.....” 2012
Juhendaja: “.....” 2012
Juhendaja: “.....” 2012

Kaitsemisele lubatud

Professor: “.....” 2012

TARTU, 2012

Tänuavaldused

Siinkohal soovib käesoleva töö autor tänada abi eest töö koostamisel juhendajaid Sven Laur, Raivo Kolde ja Andres Tover, ning töös uuritavate katsete ettevalmistajaid Radi Tegova, Kaidi Möll ja Tiiu Männik.

Sisukord

Joonised	vii
1 Sissejuhatus	1
1.1 Töö struktuur	2
2 Cellomics HCS platvorm	3
2.1 High-content screening	3
2.2 Põhikomponendid	4
2.3 Roll	5
3 ET_AR internalisatsioon	7
3.1 GPCR	7
3.2 Uuritav bioanalüüs ja kasutatavad vahendid	7
3.3 Lahendatav probleem	9
4 Pildianalüüs	11
4.1 Objektivastus	12
4.2 Väljundid	15
5 Andmeanalüüs	17
5.1 Kolmogorov-Smirnov test	17
5.2 Z-faktor	18
5.3 SSMD	19
5.4 Pearsoni korrelatsioonikordaja	19
5.5 Manderi kolokalisatsioonikoefitsent	19
5.6 Efekti iseloomustamine ning visualiseerimine	20
5.7 Rakendus	21

SISUKORD

5.7.1	Andmete üldine iseloomustus	21
5.7.2	Analüüs üle plaadi koondandmete	22
5.7.3	Valimite jaotuvuste hinnang	25
6	Kokkuvõte	27
	Resümee (inglise keeles)	29
	Kirjandus	31

Joonised

3.1	Retseptori internalisatsioon ja kolokalisatsioon endosoomidega	8
4.1	Pildinäidis retseptori internalisatsioonist	11
4.2	Pildinäidis erinevatest kanalitest	12
4.3	Isodata lävend	13
4.4	Kolmnurklävend	13
4.5	Fikseeritud ja Isodata lävendi võrdlus	14
4.6	Objektitüübid ning tuvastus	15
5.1	Väljundandmete jaotuvus	21
5.2	Määratud neljaparameetrilised kurvid	23
5.3	Kolmogorov-Smirnov distantside kurvid	25

1

Sissejuhatus

High-content screening (HCS) on kombinatsioon kaasaegsest rakubioloogiast ning automatiseeritud kõrglahutusega mikroskoopiast. HCS meetodite kasutamine võimaldab uurida rakkude füsioloogiat, struktuuri, organelle ning nende reaktsiooni väliskeskkonnale. Fotosensorite ning molekulaarbioloogia areng on andnud võimaluse rakkude tööd visuaalselt jälgida, lisades rakkudesse vastavaid fluorestseeruvaid valke, ning läbi pildianalüüsi selle kohta andmeid koguda. Kuna HCS meetodeid kasutades on võimalik koguda ulatuslikul hulgal andmeid, tuleb bioloogilise hüpoteesi ning statistiliste meetodite alusel sooritada teatav valik, millest oleks võimalik anda inimloetav ülevaade katsete kohta.

Käesolevas bakalaureusetöös on võetud uurimise alla retseptori internalisatsiooni-analüüs (ingl *receptor internalisation assay*), kus uuritakse rakumembraanis paikneva retseptorivalgu raku sisse liikumise sõltuvust retseptori spetsiifilisest ligandist, milleks on mõnekümne aminohappe pikkune polümeerne molekul ehk peptiid. Vaatluse alla võetakse ettevõtte Icosagen poolt arendatavates rakkudes retseptorvalk Endoteliin A (ET_AR). Endoteliin A retseptor mängib inimeses rolli kõrgvererõhutõve tekkes ning on sellepärast potentsiaalne oluline uurimisobjekt ravimiarenduses (1).

Töös tutvustatakse esmalt HCS süsteemi, tema võimalusi ja komponente Cellomics HCS platvormi näitel. Näidete juures tutvustatakse komponentide rolli nii bioloogilisest kui (info)tehnoloogilisest seisukohast. Töö põhieesmärgiks on aga analüüsida Endoteliin A retseptori internalisatsiooni aktiivsuse sõltuvust temaga seonduva peptiidi kontsentratsiooni suhtes. Analüüsi esimeses etapis teostatakse retseptori internalisatsioonianaalüüsi kohta kogutud piltide põhjal pildianalüüs, et tuvastada uuritavad objektid ning

1. SISSEJUHATUS

neilt arvulisi andmeid koguda. Pildianalüüsi käigus loetletakse põhilised etapid ning illustratsioonide abil iseloomustatakse kõige olulisemaid pildianalüüsiks kasutatava protokollide elemente ja parameetreid. Kogutud arvuliste andmete põhjal teostatakse analüüsi teises etapis andmeanalüüs, mille käigus uuritakse retseptorvalgu internalisatsiooni sõltuvust peptiidi kontsentratsiooni suhtes ning proovitakse võimalikke seoseid modelleerida neljaparameetriliste logistiliste mudelite abil. Andmete analüüsimise juures tuuakse ka välja levinumad statistikud ja testid, mida bioloogiliste andmete kirjeldamiseks kasutatakse.

1.1 Töö struktuur

Cellomics HCS platvorm:

Ülevaade High-Content Screening tehnoloogiast Cellomics HCS platvormi näitel, komponentidest ning olulisusest molekulaarses rakubioloogias, biomeditsiinis ning ravimiarenduses.

ET_AR internalisatsioon:

Lühike ülevaade töös uuritavast ET_AR retseptorvalgust, bioloogiline hüpotees, uurimiseks kasutatavate vahendite ülevaade ning oodatav tulemus.

Pildianalüüs:

Uuritava pildimaterjali ülevaade ning neil objektituvastuse teostamine, kasutades Cellomics BioApplication tarkvara.

Andmeanalüüs:

Analüüsi läbiviimiseks kasutatavate meetodite, statistikute ja visualisatsioonide ülevaade ning rakendus pildianalüüsist kogutud andmete põhjal.

2

Cellomics HCS platvorm

See peatükk annab ülevaate töös rakendatavast tehnoloogiavaldkonnast, selle ajaloost, põhikomponentidest ning rollist kaasaegses biotehnoloogias ning ravimiarenduses. Ühtlasi tuuakse välja ülevaade töö hilisemates etappides kasutatavast riistvarast ning tarkvarast.

2.1 High-content screening

High Content Screening (HCS) on defineeritud kui platvorm ja meetodid, sealhulgas ka instrumendid, biotehnoloogiline tarkvara, reagentid, analüüsid ning informaatika-tarkvara, mida kasutatakse rakkude aktiivsuse ja funktsioonide automatiseeritud uurimiseks. Uurimisobjektideks võivad olla nii erinevate organellide tegevus kui ka keerukamad protsessid, kus uuritakse rakus toimuvate erinevate protsesside ahelaid, valkude tootmist või nende lokaliseerimist (2). HCS lahendus võimaldab efektiivselt uurida rohkem kui 10^6 rakku analüüsi kohta, mis on ligi 1000 korda enam kui traditsioonilised rakubioloogia uurimiseks kasutatavad vahendid võimaldavad (3). Pioneeriks ning platvormi loojaks võib pidada biotehnoloogia ettevõtet Biological Detection Inc, kes tutvustas 1996. aastal esimest HCS lahendust nimega ArrayScan (4). ArrayScan süsteemi arendamiseks 1996. aasta oktoobris loodud tütarettevõtte Cellomics, Inc (alates 2010. aastast ThermoFisher Scientificu tütarettevõtte) on tänase päevani juhtivaid HCS süsteemide tootjaid ja arendajaid. Sarnaste omadustega tehnoloogiat pakuvad nüüdseks ka teised suuremad biotehnoloogiliste uuringute läbiviimiseks seadmeid tootvad ettevõtted (5). Lisaks on tekkimas avatud lähtekoodiga tarkvara, mis võimaldab tuntud

2. CELLOMICS HCS PLATVORM

kommertsiaalsete lahendustega samaväärselt hästi HCS käigus omandatavaid pildiandmeid analüüsida (6).

2.2 Põhikomponendid

Cellomics HCS platvorm on modulaarne süsteem erinevatest HCS komponentidest. Süsteemi keskmes on ArrayScan VTI instrument, mis käitub kui automatiseeritud fluorestsentsmikroskoop ning on peamine vahend bioanalüüsides (ingl *bioassay*) uurimiseks. Bioanalüüsis uuritavad rakud paigutatakse standardsuurusega mikrotiiter plaatidele (ingl *microtiter plate*), milles on 96, 384 või 1536 kannu (ingl *well*), kuhu on võimalik erinevaid proove panna. Uuritavad objektid märgistatakse fluorestseeruvate valkudega (antikehad, reagentid), mida on võimalik visuaalselt jälgida ning üksteisest eristada. Instrument on automatiseeritud ning kalibreeritud plaatide täpsete mõõtude järgi ning kasutajal on võimalik määrata vaid uuritavad kannud, ning need pildistatakse automaatselt määratud ulatuses. Plaatide automaatseks seadmesse laadimiseks on võimalik lisada riulist ja robotkäest koosnev CataLyst Express Plate Loading Robot süsteem. See muudab suuremahuliste analüüsides piltide kogumise täielikult automatiseerituks (7).

Bioanalüüsi on kahte tüüpi, esimesel puhul kutsutakse bioloogiliselt välja soovitud efekt ning selles seisus rakud fikseeritakse plaadile, teisel puhul uuritakse elus rakke üle ajaperioodi, mis võib ulatuda minutist kuni päevadeni. Cellomics HCS platvormil on selliseks uurimiseks moodulid KineticScan, mis võimaldab elus rakkudele pipeteerida erinevaid keemilisi ühendeid ja LiveCell Module, mis võimaldab rakkude eluks sobilikke keskkonnatingimusi säilitada, kontrollides CO₂ sisaldust õhus ning pildistamiskambri temperatuuri (7).

Pildianalüüsi läbiviimiseks pakutakse Cellomics BioApplication tüüpi tarkvara, mis võimaldab fluorestsentsmärgistega tähistatud objekte piltidelt ära tunda, ning nende kohta objektide omadusi iseloomustada (8). Omadused on näiteks erinevad morfoloogilised andmed rakkudel (suurus, kuju, rakustruktuuride paigutus ja suurus), erinevate valkude lokaliseerimine ja lokaliseerimiseks kasutatava fluorestsentsruva valguga (antikeha) afiinsus, spetsiifilisus ja fluorestsentsi intensiivsus. Tarkvaraga tuleb kaasa mitu standardprotokolli erinevate spetsiifiliste analüüsides läbiviimiseks nagu näiteks Cytoplasm-to-Nucleus Translocation valkude translokatsiooni uurimiseks tsütoplastist

tuuma (9) ja Cell Cycle rakutsüklite (DNA replitseerumine, mitoos ja jagunemine) uurimiseks (10).

Kogutud andmete analüüsimiseks on võimalik vHCS View tarkvarast saada nii pildi kui ka arvuliste andmete üldine ülevaade või neid analüüsida kasutades mõnd eraldiseisvat tarkvaralahendust. Spetsiaalselt bioloogiliste andmete analüüsimiseks on Cellomicsi süsteemil sisseehitatud tugi SpotFire jaoks, aga võimalik on kasutada ka mõnd üldlevinud andmeanalüüsi tarkvara nagu näiteks GraphPad Prism, MATLAB või R, kuhu on võimalik andmed eksportida komaga eraldatud kujul. Nimetatud üldlevinud tarkvaras on kõigil olemas meetodid spetsiaalselt bioloogiliste andmete analüüsimiseks ning MATLAB ja R puhul on kasutajal võimalik analüüsiks mooduleid ise juurde programmeerida (11) (12) (13).

2.3 Roll

HCS tehnoloogia on tõestanud end kui tõhus viis bioanalüüsides arendamiseks ning nende uurimiseks. HCS vahenditega on uuritud vägagi erinevaid subjekte ning haigusi. Nick Thomase 2009. aastal avaldatud artiklis on loetletud valik publitseeritud rakendusi, mille hulka kuulub hulgaliselt erinevaid diabeedi ning vähi, Alzheimeri ja südamehaigustega seonduvaid uurimusi (14). 2005. aastal tehtud uurimus HCS kasutuse ja huvi selle tehnoloogia vastu näitas, et järgneva aasta jooksul kasvab suuremahuliste HCS uuringute (ingl *screen*) 50% võrra. Suures farmaatsiatööstuses oli tollel aastal ettevõtte kohta tehtavate *screen*-ide arv keskmiselt 6.8 aastas ning eeldati, et see tõuseb 2006. aastal 10.4 peale. Akadeemilistes uuringutes rakendati tehnoloogiat peamiselt onkoloogia, neuroloogia ning toksikoloogia valdkondades ja ravimiarenduses võimalike raviks oluliste bioloogiliste märklaudade (ingl *target*) otsimiseks ning valideerimiseks (14).

2. CELLOMICS HCS PLATVORM

3

ET_AR internalisatsioon

3.1 GPCR

G-valkudega seotud retseptorid (ingl *G protein-coupled receptors* ehk GPCR) on raku pinnal paiknevad valgud, mis aktiveerivad trimeerseid G-valke, ning mida leidub kõikides eukariootsetes rakkudes alates pärmist kuni inimeseni. Inimese genoom kodeerib mitmeid tuhandeid selliseid retseptoreid (15, lk. 535). G-valkudega seotud retseptorid on osalised mitmetes erinevates haigustes ning need retseptorid on hinnanguliselt 40% kaasaegsete retseptiravimite märklauaks (16). Üldiselt on igal retseptoril üks või mitu ligandit, mõnekümne aminohappe pikkust polümeerset molekuli ehk peptiidi, millega ta aktiveeritakse (15, lk. 536). Selles töös on uurimise all Endoteliin A retseptor (ingl *Endothelin A receptor* ehk ET_AR) ning tema aktiivsuse sõltuvus ligandi Endoteliin 1 (ET-1) kontsentratsioonist. Endoteliin A retseptor on seotud hüpertroofia ja hüpertensiooni ehk kõrgvererõhutõve tekkimisega (1).

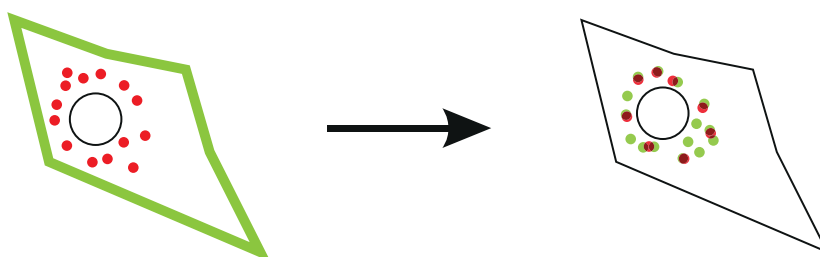
3.2 Uuritav bioanalüüs ja kasutatavad vahendid

Töös uuritavaks bioanalüüsiks on Icosagen Cell Factory poolt arendatavas QMCF süsteemi abil (17) U2OS (ingl *human osteosarcoma cell line*) rakkudes esile kutsutud ET_AR internalisatsioon. Kuna tegemist on esmakordse internalisatsioonianalüüsi uurimisega QMCF süsteemis, siis tuleb esmajärjekorras veenduda tema töös ning õigsuses. Bioloogiline hüpotees on, et piisava koguse ET_AR vastava ligandi ET-1 lisamisel ET_AR internaliseerub ning liigub selle käigus rakutuuma ümbritsevas ehk perinukleaarses alas

3. ET_AR INTERNALISATSIOON

paiknevatesse varajastesse endosoomidesse ja kolokaliseerub seal Rab5 ensüümiga (Joonis 3.1) (18). Kolokalisatsioon Rab5 ensüümiga on hüpoteetiliselt ka üks viis kirjeldamiseks internalisatsiooni aktiivsust ligandi ET-1 kontsentratsiooni suhtes.

Joonis 3.1: Retseptori internalisatsioon ja kolokalisatsioon endosoomidega. Retseptor on tähistatud rohelisega ning endosoomid punasega.



Katsed on ette valmistatud standardsetele mustadele Greiner CellStar 96 kannuga mikrotiiter plaatidele (ingl *microtiter plate*). Igasse kannu on külvatud ligikaudu 15000 raku, indutseeritud peptiidiga kindla aja jooksul ning lisaks on neile kantud vastavad fluorestsentsmärgised kindlate organellide ning piirkondade tuvastamiseks. Uurimise all on kolm bioanalüüsi, mis on koostatud samast kultuurist pärit rakkudega, kuid mis on pisut erineva ülesehitusega märgiste ning ET-1 kontsentratsiooniridade (Tabel 3.1) suhtes. Võimalike vigade ning probleemide vältimiseks on igal katsel mitu kordust iga ligandi kontsentratsiooni kohta. Kasutades ArrayScanVTi instrumenti on esimeses katses pildistatud 20 pilti, teises ja kolmandas katses 9 pilti kannu kohta.

Tabel 3.1: Ligandi kontsentratsioonid erinevates katsetes

Katse	Kontsentratsioonid (nM)
1	0; 0.25; 0.5; 0.75; 1; 1.5; 2; 2.5; 3; 3.5; 4; 5; 7; 10; 12; 15; 25; 50; 75; 100; 125; 150; 200
2	0; 0.25; 0.5; 0.75; 1; 2.5; 5; 7.5; 10; 25; 50
3	0.25; 0.5; 0.75; 1; 2.5; 5; 7.5; 10; 25; 50; 100

Pildianalüüsi läbiviimiseks kasutatakse Cellomics Scan tarkvara koos Cellomics BioApplication tarkvaraga, mille seast tuleb valida pildianalüüsi jaoks sobiv protokoll. Andmete analüüsimiseks on kasutada vHCS View tarkvara, kust on võimalik andmete kohta esmane ülevaade saada, ent andmete põhjalikumaks uurimiseks tuleb valida võimekam tarkvara. Bioanalüüsi peaks saama iseloomustada matemaatilise mudeliga, mis tähendab,

et tarkvara peaks võimaldama teostada *curve-fittingut*, ehk andmete sobitamist matemaatilise funktsiooni kõvera järgi.

3.3 Lahendatav probleem

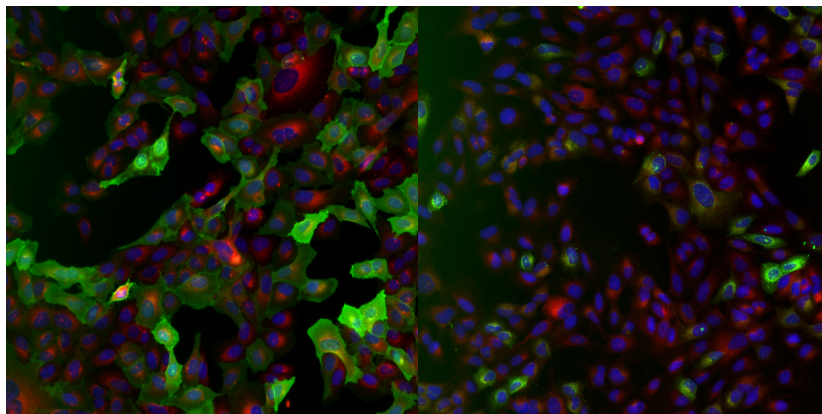
Töö peamiseks eesmärgiks on tõestada, et bioloogiline hüpotees kehtib ning bioanalüüs töötab QMCF tehnoloogiaga loodud süsteemis. Selleks tuleb läbi viia pildianalüüs, kasutades Cellomics BioApplication vahendeid, iseloomustada arvulisi andmeid, tuvastada neist olulised, valida analüüsiks statistikud ja testid ning teostada andmeanalüüs. Hüpoteesi paikapidamisel on eesmärgiks leida retseptori internalisatsiooni suhet ligandiga iseloomustav matemaatiline mudel ning iseloomustada teiste oluliste väljundmuutujate tulemusi selle suhtes. Mudeli abil on eesmärk leida bioanalüüsi suhtes ligandi efektiivsed kontsentratsioonid, millest alates internalisatsioon tööle hakkab, saavutab poole oma efektiivsusest ning enam ei kasva.

3. ET_AR INTERNALISATSIOON

4

Pildianalüüs

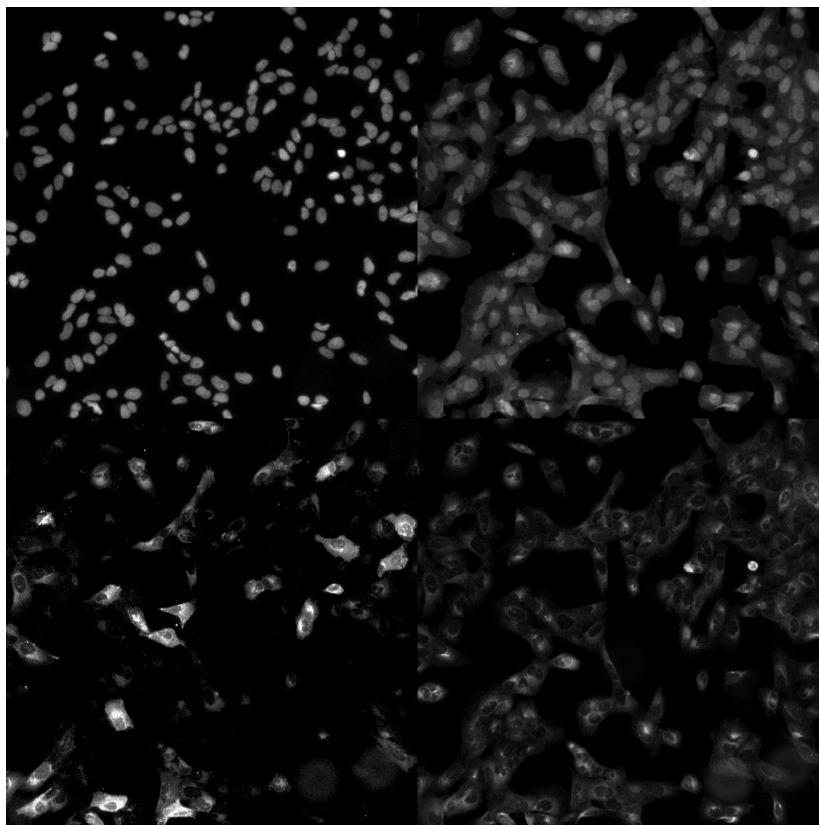
Joonis 4.1: Näidis retseptori internalisatsioonist. Vasakul pildil olevatele rakkudele ei ole lisatud ning parempoolsel pildil olevatele rakkudele on lisatud 10nM kontsentratsiooniga endoteliini. Retseptor on tähistatud rohelisega ning endosoomid punasega.



Pildianalüüsi eesmärgiks on tuvastada ArrayScan VTi instrumendiga pildistatud rakkudelt uuritavad objektid, mille kohta on seejärel võimalik arvulisi andmeid koguda. Retseptori internalisatsioon on nähtav joonisel 4.1. Analüüsi käigus on oluline tuvastada raku tuum, rakumembraani märgis, retseptor ning endosoomid, mis on pildistatud erinevates kanalites (Joonis 4.2). Pildianalüüsi baasiks on töös uuritava juhul kõikide katsete juures samalaadses konfiguratsioonis Cellomics Colocalization V3 BioApplication (19), mis võimaldab defineerida huvipakkuva ala ning uurida ühte või kahte märklauda samaaegselt.

4. PILDIANALÜÜS

Joonis 4.2: Näidis erinevatest kanalitest. Ülemises reas tuumad ning membraanimärgis, alumises retseptor ja endosoomid.

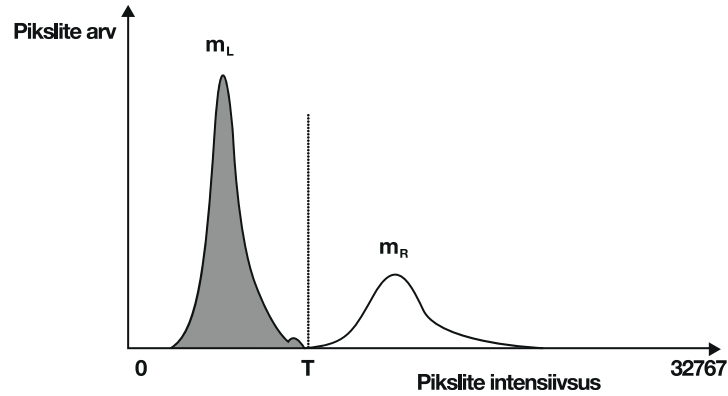


4.1 Objektituvastus

Colocalization V3 võimaldab primaarsete objektidena tuvastada tuuma ringja alana, membraanimärgisest terve raku ala ning retseptor ja endosoomid punktadena (ingl *spots*). Primaarse objektituvastuse aluseks on histogrammipõhine lävend pikslite intensiivsuse suhtes, millest alates pikslit objektiks hakatakse lugema. Cellomics BioApplication tarkvara võimaldab selleks määrata nii fikseeritud intensiivsuse üle terve katse kui ka suhtelise intensiivsuse vastavalt parajasti analüüsitava pildile (2, lk. 363). Kasutamiseks on kolme erinevat tüüpi lävendeid: fikseeritud, Isodata ja Triangle lävend (19, lk. 30).

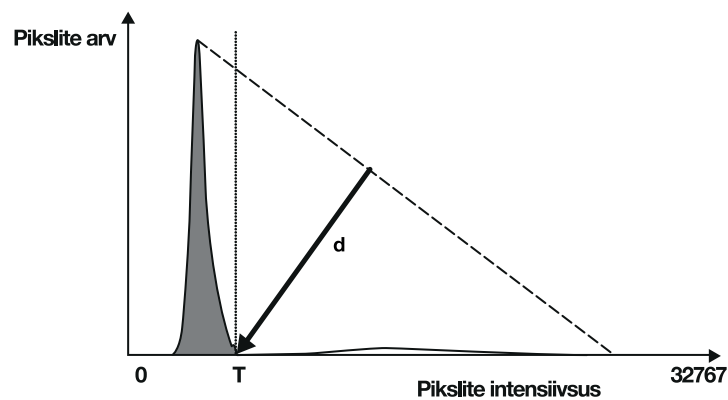
Fikseeritud lävendi (ingl *Fixed threshold*) puhul määratakse absoluutne intensiivsuse piir, millest alates kõik suurema intensiivsusega pikslid objektiks loetakse, sõltumata sealjuures pildist ning sellel leiduvast. Isodata lävend (ingl *Isodata threshold*) tähis-

Joonis 4.3: Isodata lävend. Tausta tippväärtus on tähistatud kui m_L ning objektide tippväärtus kui m_R , eraldusjooneks T on nende keskmine (19).



tab iteratiivset iseorganiseeruvat andmeanalüüsi tehnikat (ingl *Iterative self-organizing data analysis technique*). Isodata klasterdamine arendati 1965. aastal peamiselt ilmastiku modelleerimiseks ning järgimiseks (20). Pildianalüüsis kasutatakse Isodata väärtust eraldamiseks tumedaid objekte eredatest, võttes eralduspiiriks histogrammilt tausta tippväärtuse ning heledate objektide tippväärtuse kauguste keskmise. Selle parameetri muutmise negatiivseks võimaldab kaasata analüüsi tumedamaid objekte, või olemasolevate objektide piire suurendada. Isodata on seega kasulik, kui on tarvis seada piire parajasti analüüsitava pildi suhtes.

Joonis 4.4: Kolmnurklävendi määramine histogrammilt (19).



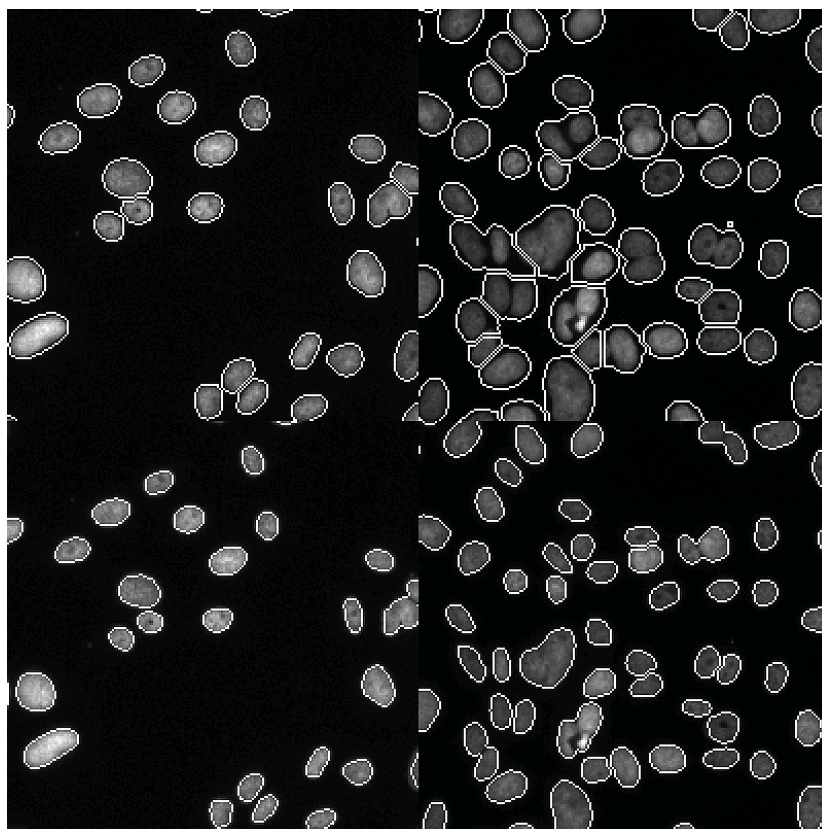
Kolmnurklävend (ingl *Triangle treshold*) (21) on samuti histogrammil põhinev ent on oluliselt robustsem Isodata meetodist. Kolmnurklävend seisneb selles, et lävend määratakse tausta maksimumi ja kõrgeima intensiivsuse vahele tõmmatud joone järgi sellesse

4. PILDIANALÜÜS

punkti, kus histogramm on kõige kaugemal joonest (Joonis 4.4), seega saab nihke alati määrata taustast kõrgemaks, ent piisavalt madalaks, et nõrgema intensiivsusega objekte tuvastada. See on kasulik, kui võib ette tulla olukordi, kus plaadil või kannus leidub tühje alasid.

Rakutuumade tuvastamiseks kasutatakse *Isodata* meetodit, kuna tuumad on küll kõikidel pildidel olemas ning tähistatud, kuid nende signaali intensiivsus võib oluliselt varieeruda. Situatsiooni kirjeldab hästi joonis 4.5, kus on valitud võrdluseks 2 kannu samalt plaadilt ning mõlemal juhul on rakendatud samade parameetritega lävendeid. Fikseeritud lävend annab ühel juhul rahuldava tulemuse ent teisel juhul mõõdetakse objektiks ka märgatavalt palju tausta. *Isodata* seevastu tagab mõlemal pildil korrektse objektituvastuse. Katsetes 2 ja 3 kasutatakse rakkude membraanimärgise tuvastamiseks *Triangle* meetodit. Retseptori tuvastamiseks kasutatakse samuti kolmnurklävendit, kuna seal on võimalik, et mõningatel pildidel pole enam märgist näha.

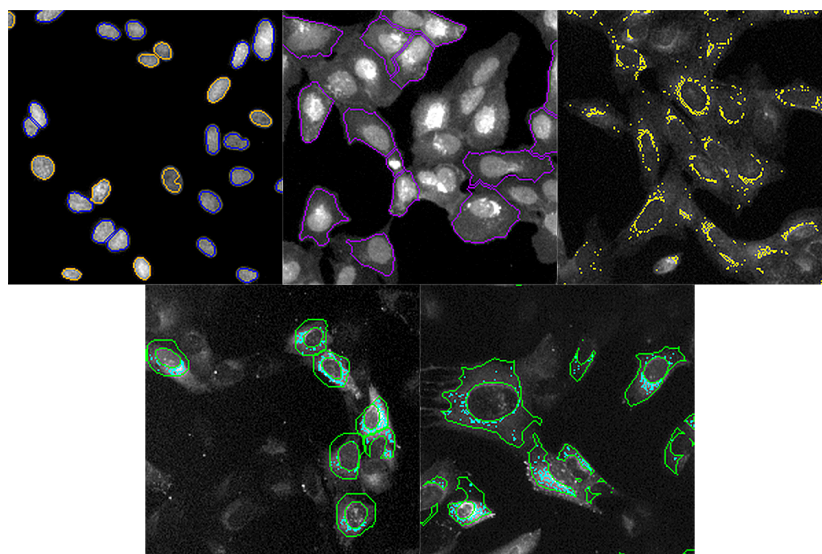
Joonis 4.5: Primaarse objektituvastuse näide kasutades fikseeritud lävendit ülemistel pildidel ning *Isodata* lävendit alumistel.



Peale primaarse objektituvastuse on võimalik määrata segmentatsiooni ja ühtlustamise parameetrid, mille alusel objektid üheks loetakse või üksteisest eraldatakse. Tuumade eraldamiseks kasutatakse geomeetrilist eraldamisviisi, kus tuumad lahutatakse üksteisest maksimaalse raadiuse alusel. See tähendab, et kaks objekti eraldatakse teineteisest, kui üks neist on protokollis määratud segmentatsiooniväärtusest suurema raadiusega. Membraanimärgise abil tuvastatud rakud lahutatakse üksteisest, kasutades intensiivsuste tippväärtust, kus tippväärtus mõõdetakse lähimast madalamast piksli intensiivsuse väärtusest. Tuvastatud spotid ühtlustatakse ühe piksli raadiuses ning võimaluse korral loetakse kokku üheks alaks.

4.2 Väljundid

Joonis 4.6: Tuvastatud objektid. Ülemises reas vasakult paremale: tuumad (sinine,oranž), raku kontuurid (lilla), endosoomid (kollane). Alumises reas mõlemal pildil retseptor (tsüaniid). Rohelistega tähistatud perinukleaarne ala.



Peale primaarsete objektide tuvastamist defineeritakse huvipakkuvad alad, märklauad ning selektsiooniparameetrid. Huvipakkuv ala on kolmes katses mõnevõrra erinevalt defineeritud. Katses 2 ja 3 määratakse bioloogilise hüpoteesi järgi huvipakkuv perinukleaarne ala ROI_A (ingl *region of interest A*) kahandades tuvastatud raku ala ning eemaldades saadud tulemusest tuuma ala (Joonisel 4.6 alumine parempoolne pilt), katses 1 aga kasutatakse lihtsamat meetodit, kus huvipakkuvaks alaks määratakse ühtlaselt

4. PILDIANALÜÜS

suurendatud tuuma ala, millest arvatakse maha esialgne tuvastatud ala (Joonisel 4.6 alumine vasakpoolne pilt). Märklaudadeks määratakse huvipakkuva ala sisse jäävad tuvastatud retseptor (**Target_I**) ja endosoomid (**Target_II**). Katses 1 määratud regioon ei ole täielikult usaldusväärne, kuna selle ala sisse võib mahtuda ka raku või mõne tema naabri membraanist tuvastatud retseptor. Lisaks määratakse ka objektide selektsiooni-parameeter, mis jätab väljundandmetesse ainult need rakud, kust retseptori signaali on võimalik tuvastada.

5

Andmeanalüüs

Esmalt on kirjeldatud analüüsiks valitud kirjanduse ning autori senise kogemuse alusel valitud testid ja statistikud, mida pildianalüüsi käigus kogutud arvilistele andmetele rakendatakse. Andmete analüüsimiseks kasutatakse analüüsitarkvara R (13).

5.1 Kolmogorov-Smirnov test

Kolmogorov-Smirnov test (K-S test) on mitteparameetiline test kahe pideva ühedimensioonilise jaotusfunktsiooni testimiseks. K-S testi loetakse mitteparameetriliseks, kuna test ei eelda spetsiifilist jaotusfunktsiooni (22). See omadus muutub oluliseks, kui uuritav valim ei ole normaaljaotusega, vaid on näiteks tugevalt kallutatud või sisaldab ohtralt nulle (*zero-inflated data*), mis on mõlemad väga levinud jaotused bioloogilistes andmetes. Selliste jaotustega andmete puhul muutuvad kasutuks populaarsed testid nagu näiteks Student's T-test, mis eeldab normaaljaotust, ning Mann-Whitney-Wilcoxon test, mis iseloomustab mediaanväärtuste erinevusi (23).

Kahe valimi Kolmogorov-Smirnov test

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|, \quad (5.1)$$

kus $F_{1,n}$ ja $F_{2,n'}$ on vastavalt esimese ja teise valimi empiirilised jaotusfunktsioonid.

Nullhüpotees kummutatakse α tasemel kui

$$\sqrt{\frac{n+n'}{nn'}} D_{n,n'} > K_\alpha. \quad (5.2)$$

kasutades kriitiliste väärtustena Ian T. Young tabeleid (22).

5. ANDMEANALÜÜS

Kolmogorov-Smirnov kahe valimi test on leidnud HCS analüüsides kasutust nii intensiivsuste võrdlemisel (24) kui ka toksikoloogia uurimustes loendatavate andmete analüüsimisel (25). Andmete analüüsimiseks kasutatavas R tarkvaras on Kolmogorov-Smirnov testid standardmoodulis `stats` nime all `ks.test` (26).

5.2 Z-faktor

Z-faktor (ingl *Z-factor*) on lihtne statistiline parameeter efekti suuruse määramiseks, mis on eelkõige mõeldud HTS analüüsides ja skriinide käigus tabamuste leidmiseks, ent sobilik ka HTS aluseks olevate bioanalüüsides optimeerimiseks ning kvaliteedi hindamiseks (27).

Z-faktor on defineeritud kui

$$\text{Z-faktor} = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}, \quad (5.3)$$

kus σ_p ja σ_n on vastavalt katse positiivse ja negatiivse kontrolli keskmised ning μ_p ja μ_n neile vastavad standardhälbed ning konstant 3 on motiveeritud normaaljaotuse printsiibist, kus 99% väärtustest langevad kuni kolme standardhälbe kaugusele keskmisest.

Praktikas kasutatakse HCS uuringutel tabamuste hindamiseks peamiselt hinnangulist faktorit, kus üks keskmisest ja standardhälbest koosnev paar võetakse kontrollist ning teine huvipakkuvalt valimilt.

Tabel 5.1: Z-faktor väärtuste tõlgendamine vastavalt esmasele publikatsioonile (27)

Z-faktor	Tõlgendus
1	Ideaalne; Z-faktor ei saa kunagi ületada väärtust 1
0.5-1.0	Väga hea bioanalüüs
0-0.5	Piiripealne või vähese kaaluga bioanalüüs
<0	Liiga kõrge kattuvus positiivse ja negatiivse kontrolli vahel

Z-faktor on alates oma esmasest avaldamisest levinud väga kiiresti akadeemilisest ringkonnast välja ning on muutunud standardseks statistikuks bioanalüüsides optimeerimisel ning on märgitud kommertsiaalsete bioanalüüsides juures tihti kui kvaliteedinäitaja (28) (29).

5.3 SSMD

Kuna eelmises lõigus kirjeldatud Z-faktor rõhus liialt normaaljaotuse eeldusele andmetes ja absoluutväärtuse kasutamine raskendab statistiliste järelduste tegemist, siis on alternatiiviks loodud SSMD ehk *strictly standardized mean difference*. SSMD esitati esmalt kui viisi kvaliteedikontrolliks ning HTS skriinides tabamuste leidmiseks (30).

SSMD (tähistatud kui β) kahe sõltumatu valimi võrdlemiseks

$$\beta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (5.4)$$

kus μ_1 ja σ_1 ning μ_2 ja σ_2 on vastavalt esimese ja teise valimi keskmine ja standardhälve.

Põhieelised Z-faktori ees on robustsus - SSMD võtab arvesse ka valimi variatsiooni.

5.4 Pearsoni korrelatsioonikordaja

Colocalization V3 BioApplication üheks võimalikuks väljundväärtuseks on kahe uuritava märklaua `Target_I` ja `Target_II` korrelatsioon, `ROI_A(B)_CorrelationCoef`, kus ROI on pildianalüüsi käigus määratud huvipakkuv regioon (ingl *region of interest*) kasutades Pearsoni korrelatsioonikoeffitsenti.

Pearsoni korrelatsioonikoeffitsent r_p Colocalization V3 tarkvaras on defineeritud kui

$$r_p = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (5.5)$$

kus x ja y on vastavalt defineeritud märklaudade pikslite intensiivsused ning \bar{x} ja \bar{y} on vastavalt instensiivsuste keskmised (19, lk 42-43).

5.5 Manderi kolokalisatsioonikoeffitsent

Lisaks eelpool mainitud korrelatsioonikordajale on üheks Colocalization V3 väljundväärtuseks ka Manderi kolokalisatsioonikoeffitsent (31) `ROI_A(B)_OverlapCoef`, mis kirjeldab uuritavate märklaudade `Target_I` ja `Target_II` kattuvust.

Manderi kolokalisatsioonikoeffitsent r on tarkvaras Colocalization V3 defineeritud kui

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (5.6)$$

kus x ja y on vastavalt defineeritud märklaudade pikslite intensiivsused (19, lk 43).

5.6 Efekti iseloomustamine ning visualiseerimine

Bioanalüüsides eeldatakse, et tulemustes ning käitumises on mingi reeglipära, mida on võimalik väljendada matemaatilise funktsiooniga. Üldiselt võib bioanalüüsi tulemuste kurve iseloomustada nii Gaussiani kumulatiivse jaotusfunktsiooni kui ka logistilise funktsiooniga. Viimase eeliseks on matemaatiline lihtsus, mistõttu on ta ka laialdaselt kasutusel erinevate bio- ja immunoanalüüsides iseloomustamisel (32).

Logistilise funktsiooni üldine kuju on järgmine

$$Y = \frac{a - d}{1 + \left(\frac{X}{c}\right)^b} + d, \quad (5.7)$$

kus Y on reaktsioon; X aritmeetiline doos; a reaktsioon, kui $X = 0$; c on doos, mille juures funktsioon saavutab poole oma maksimaalsest väärtusest ehk ED50 (ingl *effective-dose 50*), mis selle funktsiooni juures tähistab ka sümmeetriapunkti kurvil; b on kalde tegur.

Ülalkirjeldatud logistilist funktsiooni nimetatakse tihti ka neljaparameetriliseks logistikuks (ingl *four-parameter logistic*) (33). Neljaparameetriline logistiline funktsioon on lihtsustatud versioon viieparameetrilisest logistilisest funktsioonist, ehk Richard-i kurvist, kus viies parameeter lubab kurvi teatavat asümmeetriat (34, 308-309). Lisaks on levinud Gompertzi mudel, mis lubab samuti ebasümmeetriat logistilise kurvi ning Brain-Cousens'i mudel, mis võtab arvesse ka madalatel doosidel tekkivaid võimalikke ebäühtlusi (33).

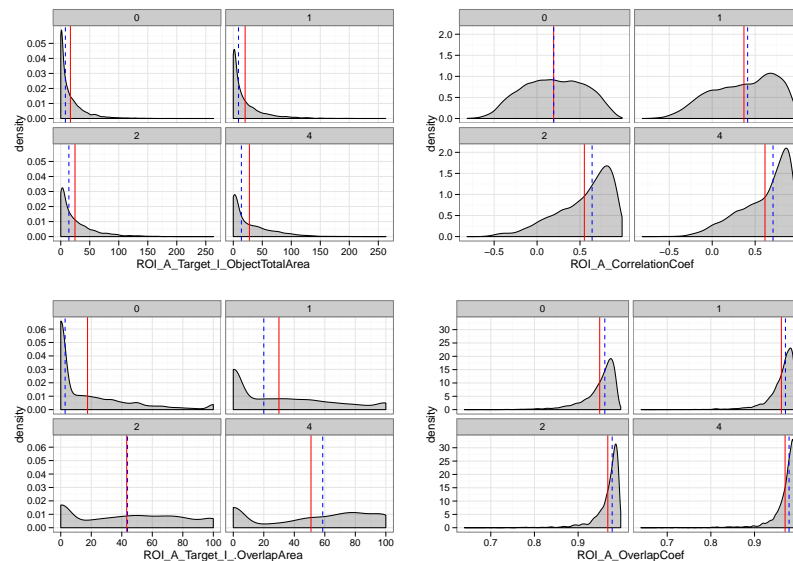
Andmete iseloomustamist matemaatiliste funktsioonide abil nimetatakse kui *curve-fitting*, ehk tulemuste seadmine vastavusse mingi funktsiooni kurviga. Vastavate logistiliste kurvide leidmiseks on andmeanalüüsitarkvaras R standardmoodulis meetod `nls`. Kuna `nls` määrab ainult kurvi, aga tihti on tarvis leida ka sobilikud sisendparameetrid, siis selleks on standardmoodulis neljaparameetrilise logistilise mudeli jaoks meetod `SSfp1` (ingl *self-starting nls four-parameter logistic model*), mis võimaldab leida ka võimalikud funktsiooni parameetrid ette antud valimi põhjal. Spetsiaalselt bioloogiliste andmete analüüsimiseks, mudelite loomiseks, kurvide iseloomustamiseks ning visualiseerimiseks on eraldiseisev moodul `drc` (33).

5.7 Rakendus

ET_AR internalisatsiooni analüüsi läbiviimiseks määratakse esmalt hüpoteesi alusel huvipakkuvad väljundmuutujad. Analüüsi subjektideks on tuvastatud retseptori kogupindala `ROI_A_Target_I_ObjectTotalArea` määratud perinukleaarses alas, märklauaks olevate retseptori ning endosoomide korrelatsioonikoeffitsient `ROI_A_CorrelationCoef`, märklaudade kattuvuskoeffitsient `ROI_A_OverlapCoef` ja retseptori pindala kattuvus endosoomide suhtes `ROI_A_Target_I_OverlapArea`. Uuritavate andmekogude suurusks on katse 1 (plaat 1) 67278 rakku, katse 2 (plaat 2) 31660 rakku ja katse 3 (plaat 3) 35291 rakku.

5.7.1 Andmete üldine iseloomustus

Joonis 5.1: Tihedusdiagrammid uuritavate väljundite kohta katsest 1 jagatuna osadeks dooside (0,1,2 ja 4 nM) kaupa. Punase joonega on tähistatud keskvärtus ning sinise joonega mediaan.



Intuiitiivne lähenemine on esmalt vaadata väljundmuutujaid üldistavaid väärtusi nagu keskvärtus ja mediaan jaotuvuse kontekstis. Koondstatistikute kogumiseks andmetelt kasutatakse tarkvara R moodulit `plyr` (35). Jooniselt 5.1 on näha, et uuritavate väljundmuutujate seas toimuvad muutused peamiselt jaotuses. Retseptori kogupindala ja märklaudade kattuvuse jaotuvust võib iseloomustada kui *zero-inflated data*, kus on

5. ANDMEANALÜÜS

doosiga sõltuvuses 0 väärtuste osakaalu langemine. Korrelatsioonikoefitsentide jaotus moonduv doosiga sõltuvalt 0 koefitsendilt positiivseks ning kattuvuse koefitsent kasvab väga kitsal skaalal. Selliste andmete analüüsimiseks on kaks lähenemist:

Ignoreerida andmeid raku tasemel

Selle lähenemisega vaadeldakse väärtusi plaadi mõistes, kus uuritavaks väärtuseks on mediaanväärtus. Bioanalüüsi kvaliteeti iseloomustavad statistikud arvutatakse samuti üle üldistavate väärtuste. Raku tasemelt plaadi tasemele liikumist võib vaadelda kui statistilise müra vähendamist, ent selle käigus võib kaotada väärtuslikku informatsiooni rakkude kohta.

Alternatiivsete statistikute määramine

Traditsioonilise lähenemise asemel andmete robustne iseloomustamine nende jaotusfunktsioonide alusel. Selline lähenemine võimaldab katseid uurida, säilitades siiski maksimaalselt infot esialgsete andmete kohta. Levinuimaks selliseks robustseks statistikuks on Kolmogorov-Smirnov kahe jaotusfunktsiooni test, mis annab tulemuseks jaotuvuste kauguse üksteisest.

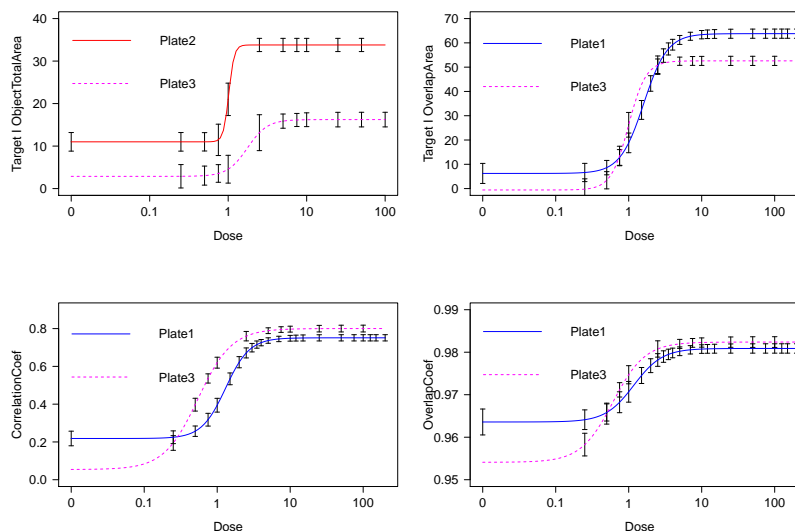
5.7.2 Analüüs üle plaadi koondandmete

Analüüsiks üle koondandmete modelleeritakse potentsiaalsete sõltuvust kirjeldavate muutujate mediaanväärtused doosi kohta neljaparameetrilisteks sigmoidseteks kurvideks, jättes välja retseptori kogupindala väärtuse esimeselt plaadilt, kuna seal ei olnud võimalik kasutada piisavalt usaldusväärset objektituvastust.

Mudelite abil saab anda hinnangu erinevatele efektiivsetele doosidele: bioanalüüs hakkab tööle (aktiivsus 1%), saavutatakse pool aktiivsust ning määratakse mudeli sümmeetriapunkt (aktiivsus 50%) ning doos, millest alates efektiivsus enam oluliselt ei tõuse (aktiivsus 99%). Maksimaalset aktiivsust (100%) pole logistilist mudelit kasutades võimalik määrata, kuna väärtus läheneb lõpmatusse. Visuaalsest hinnangust üksi ei piisa ning dooside hinnangu jaoks peaks vaatama mudelite kvaliteeti. Kvaliteedihinnanguks on drc moodulis võimalik rakendada *lack of fit* testi, mis võrdleb sigmoidset drc mudelit *one-way ANOVA* mudeliga, ning annab hinnangu mudeli sobimatuse kohta.

Tabel 5.2 näitab, et testide tulemused ei ole olulised. See tähendab, et leitud neljaparameetrilised logistilised mudelid kirjeldavad andmeid hästi. Parima mudeli annab

Joonis 5.2: Määratud neljaparameetrilised kurvid mudelipõhiste vigadega.



retseptori ja endosoomide pindalade kattuvus, ning üsnagi sarnase tulemuse Pearsoni korrelatsioonikoeffitsent ja Manderi kolokalisatsioonikoeffitsent.

Tabeli 5.3 alusel võib väita, et lõpliku hinnangu andmiseks on siikohal parim võtta esimese katse kattuvuse hinnang, kuna sealt saadakse sarnased efektiivsed kontsentratsioonid kõikide mudelite suhtes. Bioanalüüsi kohta hinnangu andmiseks saab negatiivseks kontrolliks valida indutseerimata kontrollkannu ning mudeli alusel positiivseks kontrolliks lähima väärtuse 99% efektiivsusele ehk 10.0nM.

Z-faktor ja SSMD üle kattuvuse

$$Z_{overlap} = 1 - \frac{3(8.15892439830629 + 0)}{|55.7692307692307 - 2.85714285714286|} = 0.5374068$$

$$SSMD_{overlap} = \frac{55.7692307692307 - 2.85714285714286}{\sqrt{8.15892439830629^2 + 0^2}} = 6.485179$$

Mõlemad hinnangud näitavad, et tegemist on väga hea bioanalüüsiga. Siinkohal tuleks tähele panna, et arvutatud statistikud on konkreetsel juhul pigem illustratiivse tähendusega, kuna need on arvutatud üle 2 positiivse ja 2 negatiivse mõõdu. Küll aga on korrelatsioonikoeffitsendi alusel võimalik kinnitada bioloogilist hüpoteesi ja kinnitada, et bioanalüüs töötab ning seos ligandi kontsentratsiooni ja internalisatsiooni aktiivsusega on olemas.

5. ANDMEANALÜÜS

Tabel 5.2: Mudelite sobimatuse hinnangud. ModelDF tähistab mudeli vabadusastmeid, RSS *residual sum of squares*, DF hinnangu vabadusastmeid, F-value ANOVA mudeli F-statistik ning p-väärtus olulisusnivoo.

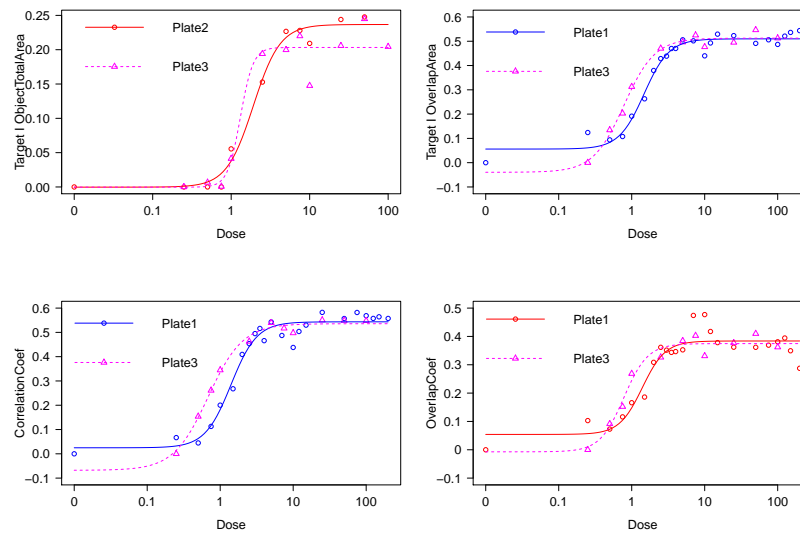
Model	ModelDF	RSS	DF	F-value	p-value
Target I ObjectTotalArea					
ANOVA	66	797.94			
DRC model	80	1172.96	14	2.2157	0.016
Target I OverlapArea					
ANOVA	56	1181.4			
DRC model	82	1617.9	26	0.7957	0.7342
CorrelationCoef					
ANOVA	56	0.08165			
DRC model	82	0.12694	26	1.1947	0.2832
OverlapCoef					
ANOVA	56	0.0004			
DRC model	82	0.0006	26	1.2127	0.2682

Tabel 5.3: Efektiivsed doosid erinevate muutujate ning plaatide kohta

	ED1	stdev	ED50	stdev	ED99	stdev
Overlap						
Plate 1	0.29105	0.0752	1.60575	0.1013	8.85889	1.8053
Plate 3	0.28231	0.1139	0.99453	0.0511	3.50357	1.5669
CorrelationCoef						
Plate 1	0.232094	0.0577	1.303624	0.0802	7.322195	1.5238
Plate 3	0.047141	0.0287	0.544571	0.0859	6.290832	2.0884
OverlapCoef						
Plate 1	0.14969	0.091	1.17611	0.1948	9.24094	4.0161
Plate 3	0.05952	0.0608	0.60771	0.1393	6.20481	3.8972

5.7.3 Valimite jaotuvuste hinnang

Joonis 5.3: Modelleeritud neljaparameetrilised kurvid üle Kolmogorov-Smirnov distant-side kontrollkannude suhtes



Valimi jaotuvuste uurimise eesmärk on anda hinnang raku tasemel andmetele ilma nende kohta eelnevalt üldistusi tegemata. Kolmogorov-Smirnov kahe valimi testi abil saame väljundväärtuseks maksimaalse kauguse kahe jaotuvuse vahel, kus esimene jaotus on alati kontrollkannust kogutud andmete jaotus. Testi jaoks vaadeldakse erinevatest kannudest ent sama doosiga kandmeid tervikuna, kuna tegemist on kahe valimi testiga ja katsetes olevaid korduseid ei saa vaadata kui ühte rida. Seetõttu pole võimalik mudelitele matemaatilist sobivuse hinnangut anda ning joonis 5.3 on pigem illustratiivse tähendusega. On näha, et kurvid on üksteisele oluliselt lähemale toodud ning see annab tulevikus võimaluse Kolmogorov-Smirnov testi abil katseid hilisemaks võrdluseks normaliseerida. Kvaliteedihinnanguks võib K-S testi puhul kasutada kriitilist väärtust usaldusnivool 0.001.

Retseptori kogupindala kriitiline väärtus plaadil 2, kasutades valimite suurusteks 0nM ning 10nM kontsentratsiooni juurest võetud mõõtude koguarvu

$$D_{crit} = \sqrt{\frac{5783 + 1669}{5783 * 1669}} * 1.95 = 0.0541834 \quad (5.8)$$

5. ANDMEANALÜÜS

näitab, et tulemuseks saadud $D \approx 0.23$ on üle eristatavuse piiri. Mõningates artiklites on kriitilise piirina kasutatud ka väärtust 0.2 (25); saadud tulemus ületab ka selle. Lisaks normaliseerviale kõrvalnähule võib öelda, et internalisatsiooni on võimalik uurida, kasutades ainult retseptorit ning rakumärgist, adekvaatse hinnangu bioanalüüsile saab ka ilma kolokalisatsiooni uurimata.

6

Kokkuvõte

Cellomics HCS platvorm on paindlik *high-content screening* lahendus, mis koos oma lisamoodulitega võimaldab uurida mitmesuguseid erinevaid bioanalüüse alates valkude translokatsioonist kuni rakutsükli jälgimiseni elus rakkudes. HCS on tõestanud ennast kui tõhus ja paindlik analüüsiviis ning on kasvava tähtsusega nii akadeemilistes uurimustes kui ka farmaatsias ravimiarenduseks.

Cellomics Colocalization V3 pildianalüüsiprotkolli kasutades piltidelt võetud mõõdud kinnitavad, et Endoteliin A retseptorvalgu ($ET_A R$) internalisatsiooni aktiivsus Icosageni poolt arendatavates rakkudes käitub reeglipäraselt temale lisatud peptiidi Endoteliin-1 (ET-1) suhtes. Matemaatiliselt on võimalik reeglipära kirjeldada neljaparametrilise sigmoidse kurviga. Kuna uuritavate andmete jaotused polnud mitte ühegi uuritava parameetri puhul normaaljaotusega, siis määrati matemaatilised mudelid üle andmete mediaanväärtuste. Mudelid valideeriti võrreldes neid *one-way* ANOVA mudeliga, ning testide tulemusena võib väita, et määratud kurvid iseloomustavad mõõdetud andmeid hästi.

Bioloogiline hüpotees, et bioanalüüs töötab, on võimalik kinnitada retseptori ning tema internalisatsiooni sihtpunktiks olevate endosoomide kattuvuse alusel. Efektiivsed peptiidi kontsentratsioonid on internalisatsiooni tööle hakkamiseks 0.29nM, poole aktiivsuse saavutamiseks 1.6nM ning maksimaalse efektiivsuse saavutamiseks 10nM.

Lisaks tuvastati töö käigus, et internalisatsiooni on võimalik iseloomustada ka endosoomimärget kasutamata, uurides Kolmogorov-Smirnov kahe valimi testist (K-S test) väljundina saadavaid andmete jaotuvuste kaugusi üksteisest. K-S testi tulemused näitasid ka testi potentsiaali erinevate katsete normaliseerimiseks hilisema võrdluse huvides.

Receptor Internalization Assay

Bachelor's Thesis (6 ECTS)

Taavi Tetlov

Summary

High-content screening has proved itself to be a highly efficient method for analyzing bioassays in various different fields from oncology to drug screening. Cellomics HCS platform is a modular system which provides the means to analyze cellular and subcellular processes by using fluorescent dyes and automatic fluorescent microscopy.

The aim of this thesis is to describe receptor internalization bioassay in biological system developed by Icosagen. Main focus of the thesis is to describe both image and data analysis methods and processes behind the final description of the assay.

Studied assay is Endothelin A receptor internalization assay and the goal was to investigate the relation of peptide Endothelin-1 and internalization activity. By using Cellomics Colocalization V3 BioApplication for image analysis and fitting four parameter logistic curve on receptor and early-endosome overlap medians it was found that the Endothelin-1 effective doses are 0.29nM for the assay to show any activity, 1.6nM for half-maximal efficiency and 10nM for maximum internalization.

As an alternative data analysis method Kolmogorov-Smirnov two sample test for difference in variance was also tested. Results of this approach were fitted to four-parameter logistic model and the sigmoidal curves suggested that this method can be used as an alternative to more common study of mean or median values.

Kirjandus

- [1] E. Šchiffirin, Endothelin: potential role in hypertension and vascular hypertrophy, *Hypertension* 25 (6) (1995) 1135–1143. 1, 7
- [2] D. L. Taylor (Ed.), High Content Screening (Methods in Molecular Biology), 1st Edition, Humana Press, 2006. URL <http://amazon.com/o/ASIN/1588297314/> 3, 12
- [3] V. C. Abraham, D. Taylor, J. R. Haskins, High content screening applied to large-scale cell biology, *Trends in Biotechnology* 22 (1) (2004) 15 – 22. doi:10.1016/j.tibtech.2003.10.012. URL <http://www.sciencedirect.com/science/article/pii/S016779903003068> 3
- [4] B. Wire, Arrayscan system introduces high throughput cell-based screening, Tech. rep., Business Wire, http://findarticles.com/p/articles/mi_m0EIN/is_1996_May_28/ai_18320666/ (1996). 3
- [5] F. Žanella, J. Lorens, W. Link, High content screening: seeing is believing, *Trends in biotechnology* 28 (5) (2010) 237–245. 3
- [6] CellProfiler, Cellprofiler (2012). URL <http://www.cellprofiler.org> 4
- [7] C. ThermoFisherScientific, Cellular imaging instrument modules (2012). URL <http://www.cellomics.com/products/product-types/cellular-imaging-instrument-modules.html> 4
- [8] Cellomics, Image analysis tools (2012). URL <http://www.cellomics.com/products/product-types/image-analysis-tools.html> 4
- [9] Cellomics, Cytoplasm to nucleus translocation (2012). URL <http://www.cellomics.com/products/bioapplication/cytoplasm-to-nucleus-translocation-bioapplication.html> 5
- [10] Cellomics, Cell cycle bioapplication (2012). URL <http://www.cellomics.com/products/bioapplications-image-analysis-software/cell-cycle-bioapplication.html> 5
- [11] GraphPad, Graphpad prism (2012). URL <http://www.graphpad.com/welcome.htm> 5
- [12] MathWorks, Matlab (2012). URL <http://www.mathworks.se/products/matlab/> 5
- [13] R-Project, The r project for statistical computing (2012). URL <http://www.r-project.org/> 5, 17
- [14] N. Thomas, High-content screening: A decade of evolution, *Journal of Biomolecular Screening* 15 (1) (2010) 1–9. arXiv:<http://jbx.sagepub.com/content/15/1/1.full.pdf+html>, doi:10.1177/1087057109353790. URL <http://jbx.sagepub.com/content/15/1/1.abstract> 5
- [15] H. Lodish, A. Berk, P. Matsudaira, C. A. Kaiser, M. Krieger, M. P. Scott, L. Žipursky, J. Darnell, *Molecular Cell Biology*, fifth edition Edition, W. H. Freeman, 2003. URL <http://amazon.com/o/ASIN/0716743663/> 7
- [16] D. Filmore, It’s a gpcr world, *Modern drug discovery* 7 (11) (2004) 24–28. 7
- [17] T. Šilla, I. Hääl, J. Geimanen, K. Janikson, A. Abroi, E. Ustav, M. Ustav, Episomal maintenance of plasmids with hybrid origins in mouse cells, *Journal of Virology* 79 (24) (December 15, 2005) 15277–15288. arXiv:<http://jvi.asm.org/content/79/24/15277.full.pdf+html>, doi:10.1128/JVI.79.24.15277-15288.2005. URL <http://jvi.asm.org/content/79/24/15277.abstract> 7
- [18] T. Bremnes, J. Paasche, A. Mehlum, C. Šandberg, B. Bremnes, H. Attramadal, Regulation and intracellular trafficking pathways of the endothelin receptors, *Journal of Biological Chemistry* 275 (23) (2000) 17596–17604. 8
- [19] T. Šcientific, Colocalization v3 bioapplication guide (2010). URL www.cmc-utrecht.nl/facilities/csc/Colocalization_V3_Manual.pdf 11, 12, 13, 19
- [20] G. Ball, D. Hall, Isodata, a novel method of data analysis and pattern classification, Tech. rep., DTIC Document (1965). 13
- [21] G. W. Zack, W. E. Rogers, S. A. Latt, Automatic measurement of sister chromatid exchange frequency., *Journal of Histochemistry & Cytochemistry* 25 (7) (1977) 741–753. arXiv:<http://jhc.sagepub.com/content/25/7/741.full.pdf+html>, doi:10.1177/25.7.70454. URL <http://jhc.sagepub.com/content/25/7/741.abstract> 13
- [22] I. T. Young, Proof without prejudice: use of the kolmogorov-smirnov test for the analysis of histograms from flow systems and other sources., *Journal of Histochemistry & Cytochemistry* 25 (7) (1977) 935–41. arXiv:<http://jhc.sagepub.com/content/25/7/935.full.pdf+html>, doi:10.1177/25.7.894009. URL <http://jhc.sagepub.com/content/25/7/935.abstract> 17
- [23] F. McElduff, M. Cortina-Borja, S.-K. Chan, A. Wade, When t-tests or wilcoxon-mann-whitney tests won’t do, *Advances in Physiology Education* 34 (3) (2010) 128–133. arXiv:<http://advan.physiology.org/content/34/3/128.full.pdf+html>, doi:10.1152/advan.00017.2010. URL <http://advan.physiology.org/content/34/3/128.abstract> 17
- [24] K. A. Giuliano, Y.-T. Chen, D. L. Taylor, High-content screening with sirna optimizes a cell biological approach to drug discovery: Defining the role of p53 activation in the cellular response to anticancer drugs,

- Journal of Biomolecular Screening 9 (7) (2004) 557–568. arXiv:<http://jbx.sagepub.com/content/9/7/557.full.pdf+html>, doi:10.1177/1087057104265387. URL <http://jbx.sagepub.com/content/9/7/557.abstract> 18
- [25] F. Lemaire, C. A. Mandon, J. Reboud, A. Papine, J. Angulo, H. Pointu, C. Diaz-Latoud, C. Lajaunie, F. Chatelain, A.-P. Arrigo, B. Schaack, Toxicity assays in nanodrops combining bioassay and morphometric endpoints, PLoS ONE 2 (1) (2007) e163. doi:10.1371/journal.pone.0000163. URL <http://dx.plos.org/10.1371/journal.pone.0000163> 18, 26
- [26] R-Project, Ks-test `ks.test{stats}` (2012). URL <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/ks.test.html> 18
- [27] J. ZHANG, T. Chung, K. Oldenburg, Validation of high throughput screening assays, Journal of Biomolecular Screening 4 (2). 18
- [28] Invitrogen, Fluo-4 am & fluo-4 nw calcium indicators (2012). URL <http://tinyurl.com/yz7vgyu> 18
- [29] Invitrogen, Z'-lytetm kinase assay kit (2012). URL <http://tinyurl.com/ccqomx6> 18
- [30] X. D. Zhang, A pair of new statistical parameters for quality control in rna interference high-throughput screening assays, Genomics 89 (4) (2007) 552 – 561. doi:10.1016/j.ygeno.2006.12.014. URL <http://www.sciencedirect.com/science/article/pii/S0888754307000079> 19
- [31] E. Manders, F. Verbeek, J. Aten, Measurement of colocalization of objects in dual-colour confocal images, Journal of Microscopy 169 (3) (1993) 375–382. URL <http://fiji.sc/wiki/images/2/24/Manders.pdf> 19
- [32] A. DeLean, P. J. Munson, D. Rodbard, Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves., American Journal of Physiology - Endocrinology And Metabolism 235 (2) (1978) E97–102. arXiv:<http://ajpendo.physiology.org/content/235/2/E97.full.pdf+html>. URL <http://ajpendo.physiology.org/content/235/2/E97.abstract> 20
- [33] C. Ritz, J. C. Streibig, Bioassay analysis using `r`, Journal of Statistical Software 12 (5) (2005) 1–22. URL <http://www.jstatsoft.org/v12/i05> 20
- [34] H. Motulsky, A. Christopoulos, Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting, 1st Edition, Oxford University Press, USA, 2004. URL <http://amazon.com/o/ASIN/0195171802/> 20
- [35] H. Wickham, The split-apply-combine strategy for data analysis, Journal of Statistical Software 40 (1) (2011) 1–29. URL <http://www.jstatsoft.org/v40/i01/> 21