

LATVIJAS UNIVERSITĀTE
MEDICĪNAS FAKULTĀTE

**Datu analīzes algoritma izstrāde šūnas kodolā
internalizējošu aptamēru atlasei**

BAKALaura DARBS

Autors: Rūdolfs Strazdīts

Studenta apliecības Nr.: **rs17069**

Darba vadītājs: Pētnieks (LU MF) **Kārlis Pleiko**

RĪGA 2020

Anotācija

Aptamēri ir moderna zāļu ārstēšanās forma ar lielu potenciālu, kas iekļauj sevī gan jaunas diagnostikas iespējas, tā var tikt lietota kā zāļu forma kurai piemīt agonisma vai antagonisma farmakoloģiskās īpašības pret receptoriem, vai aptamēri var tikt izmantoti kā īpašs zāļu transports. Aptamēru iegūšanas un pētīšanas metodika ir ļoti daudzveidīga un vēl aizvien uzlabojama, jo tai ir daudz posmu, kurus ir iespējams modificēt un pielāgot izpētes vajadzībām, īpaši interesanta ir arī sekvenču datu apstrādes daļa, caur kuru no miljoniem sekvenču beigās, pēc aptamēriem nepieciešamām īpašībām – specifiskuma pret mērķi, saistīšanās spējas jeb afinitātes ar mērķi, oligonukleotīdu stabilitātes un citām īpašībām, tās tiek atlasītas un tikai tad tās var tikt sauktas par aptamēriem. Šajā bakalaura darbā mēs vēlējamies padziļināti izpētīt un apskatīt datu apstrādes iespējas, lai beigās panāktu atkārtojamu rezultātu, ja tiktu izmantoti līdzīgi dati, šajā gadījumā no pētītām sekvencēm, kuras spētu internalizēties šūnas kodolā.

Atslēgas-vārdi: Aptamēri, šūnu kodols, internalizētas sekvences, datu apstrāde, DNS, *high throughput sequencing*, HTS.

Summary

Aptamers are a modern drug form with a very high potential. It can be used in diagnostics, as a drug itself (agonist or an antagonist) or it can be used as a drug delivery tool. Methods that are used to acquire and to also study aptamers are very diverse and still are being improved, because it has many steps between – starting with just a random sequence library and ending with the actual end sequence which in the end can be called an aptamer. These steps can be modified and adjusted to get the sequence that is wanted. Particularly interesting is data processing by which only a couple out of million sequences are filtered out that have the wanted specific characteristics, such as a high specificity against the target, high affinity with the target, structural stability and other specific features that make aptamers a very favorable and promising potential drug class. In this work we looked at data processing and what it can offer, also we completed data processing of existing data and by using our methods, the results can be repeated.

Key-words: Aptamers, internalized sequences, cell nuclei, data processing, DNA, *high throughput sequencing*, HTS.

Apzīmējumi

HTS - *high throughput sequencing*

DNS - dezoksiribonukleīnskābe

ssDNS - *single-stranded* DNS jeb viena-pavediena DNS

dsDNS - *double-stranded* DNS jeb divu-pavediešu DNS

cDNS - *complimentary* DNS jeb komplimentārais DNS pavediens

RNS - Ribonukleīnskābe

siRNS - silencing RNS

SELEX - *systematic evolution of ligands by exponential enrichment*

IgG1 - imūnglobulīns G1

PEGylation - polietilēnglikolēšana

HIV - cilvēka imundeficīta slimība

PCR - *polymerase chain reaction* jeb polimerāzes ķēdes reakciju

QC - *quality control* jeb kvalitātes kontrole

IDE - *integrated development environment*

CPM - *counts per million*

FC - *fold change*

FDR - *false discovery rate*

att. - attēls

Satura rādītājs

Anotācija.....	2
Summary.....	3
Apzīmējumi.....	4
Satura rādītājs.....	5
Ievads.....	6
1. Teorētiskais pamatojums.....	7
1.1. Aptamēri un to tipi.....	7
1.1.1. Aptamēru priekšrocības un nepilnības	8
1.1.2. Aptamēru modifikācijas un to ietekme.....	8
1.2. Aptamēru pirmsākums	10
1.3. PCR.....	10
1.4. SELEX.....	11
1.4.1. <i>Cell-SELEX</i>	13
1.4.2. <i>Capture SELEX</i>	15
1.5. Aptamēri un to pielietojums.....	17
1.6. Intenzalizējoši aptamēri	18
1.7. Aptamēri klīnikā	18
1.8. Pašlaik lietotie analīžu rīki (kas tiek lietoti lai analizētu datus).....	19
1.8.1. Sekvenču sekvencēšana	19
1.8.2. Sekvencēšanas iekārtas.....	20
2. Laboratorija	22
2.1. Izmantotās metodes un materiāli.....	22
2.1.1. Sākotnējā datu apstrāde.....	22
2.1.2. Sekundārā datu apstrāde	24
3. Rezultāti un diskusija	27
3.1. Sekvences, kas sastopamas kodolā vai citoplazmā	27
3.2. Sekvences, kas sastopamas kodolā vai šūnā.....	29
4. Secinājumi.....	32
5. Pateicības	33
6. Izmantotā literatūra un avoti.....	34
DOKUMENTĀRĀ LAPA	38

Ievads

Aptamēri ir veidoti no nukleīnskābju ķēdes (oligonukleotīdi), kas parasti ir sakārtoti RNS ķēdes formā, ssDNS (*single-stranded DNA*) ķēdes formā, kā arī citu struktūru un molekulu aptamēri. Šīs molekulas spēj saistīties ar augstu afinitāti un specifiskumu ar mērķšūnu vai konkrētu molekulu, vienlaicīgi nesaistoties ar citām šūnām vai molekulām, kas tieši padara aptamērus par tik vērtīgiem potenciālajiem terapeitiskajiem līdzekļiem. Ir arī atrasti daudzi citi pielietojumi, kas var tikt izmatoti lai atvieglotu cilvēku dzīves ne tikai medicīnā, piemēram, ir izveidoti precīzi diagnostikas veidi pret mikotoksīniem (Wang, *et al.*, 2011.).

Aptamērus iegūst ar īpašām selekcijas metodēm, kur par pamatu atzīta SELEX (*systematic evolution of ligands by exponential enrichment*) metode, kurai veiktas daudz modifikācijas. Svarīgi ir atlasīt aptamērus pret konkrētiem mērķiem, kas ir iespējams izmantojot mērķi selekcijas metodes laikā.

Aptamērus bieži salīdzina ar antivielu grupas zālēm, jo tās spēj pildīt līdzīgas funkcijas un ir ar līdzīgu definīciju, tas ir, ka svarīga ir specifitāte un afinitāte, tomēr salīdzinot šīs abas klases savstarpēji, var novērot gan daudz priekšrocības, gan arī dažas nepilnības, kas starp šīm divām grupām atšķiras (Chandola & Neerathilingam, 2019.).

Aptamēri vēl ir salīdzinoši jauna tehnoloģija, kas vēl nav pilnvērtīgi izpētīta, un visticamāk, veicot tālākus pētījumus, aptamēriem atradīsies vēl neordinārākas un pārsteidzošākas pielietojamības iespējas daudzās sfērās.

Darba mērķis: izstrādāt datu analīzes algoritmu šūnas kodolā internalizējošu aptamēru atlasei.

Darba uzdevumi:

1. izveidot komandu ķēdi, kas balstās uz vairākām programmām un pakotnēm, un veikt komandu ķēdes izpildi, tādējādi to pārbaudot (svarīgi lai ir iegūstami atkārtojami rezultāti);
2. veicot komandu ķēdi – vajadzētu iegūt sekvences, kuras potenciāli varētu spēt internalizēties mērķa šūnas kodolā;
3. padziļināti iepazīt laboratoriju kas saistīta ar mūsdienīgām metodēm – precīzāk, tieši datu apstrādes posmu.

1. Teorētiskais pamatojums

1.1. Aptamēri un to tipi

Aptamēru galvenā atšķirība starp RNS un ssDNS ir stabilitāte, kur RNS ir mazāk stabils dēļ tā, ka, veidojot nukleotīdu ķēdi, tā struktūrā ir reaktīvas hidroksil ($-OH$) grupas 2' ribozes gredzena pozīcijā, kas ātri spēj reducēties, zaudējot ūdeņraža – H^+ atomu, tādējādi kļūstot nestabils, to var samazināt, atņemot šo, OH^- grupu. Parasti šiem RNS aptamēriem ir veiktas dažādas struktūrālas modifikācijas, tieši lai uzlabotu to stabilitāti. Vēl svarīga atšķirība ir RNS nukleotīdu struktūrā, jo aptamēriem svarīga ir ne tikai divdimensiālā, bet vēl svarīgāka ir to trīsdimensiālā struktūra, šādā ziņā RNS aptamēri veido lielākas dažādības struktūras, kas ir noderīgi lai iegūtu aptamērus ar augstāku specifitāti un afinitāti, taču tieši šī iemesla dēļ ssDNS aptamēri tiek sintizēti ar garākām nukleotīdu sekvencēm, lai arī būtu šī struktūru dažādība. Svarīgs ir arī selekcijas process, parasti veikts ar SELEX (jeb no angļu valodas – *systematic evolution of ligands by exponential enrichment*) un šīs metodes dažādajām variācijām, kur ssDNS aptamēri ir ērtāk atlasāmi par RNS, jo metodes aizņem salīdzinoši mazāk laika (RNS aptamēriem jāveic reversā transkripcija, kā arī *in vitro* transkripcija pēc katras atlasēšanas cikla (jeb angļu valodā *round*), kamēr ssDNS aptamēriem nav vajadzības veikt šos papildus soļus, kas ir arī ekonomiski izdevīgāk) (Zhu, *et al.*, 2015.).

Interesanti, ka RNS un ssDNS aptamēri var būt ar pilnīgi atšķirīgām sekvencēm, jeb no citām nukleotīdu kombinācijām, un arī citām locīšanās struktūrām, bet tāpat šie aptamēri var saistīties ar to pašu mērķi (Song, *et al.*, 2012.).

Ir arī vēl kāds specifisks aptamēru tips – *circular aptamer*, jeb cirkulārie aptamēri, kas ir gan RNS, gan DNS tipa aptamēru paveids, kur parasti aptamēru struktūra būtu salocīta trīsdimensionāla ķēde ar diviem galiem, tā vietā cirkulāra aptamēra struktūra ir arī trīsdimensionāli salocīta nukleotīdu ķēde, bet šīs ķēdes gali ir savienoti – veidojot apļveida struktūru. Šie aptamēri ir arī novēroti dabā, dzīvo organismu šūnās, un līdz galam nav izprasta šo apļveida nukleotīdu ķēdes vajadzība vai nozīme organisma funkcionēšanai. Galvenā cirkulārās struktūras priekšrocība ir stabilitāte, kā arī spēja aizvien būt ar pietiekami augstu specifitāti (Li, *et al.*, 2020.).

Zināms, ka aptamēri ir daudzveidīgi un ar lielu dažādību starp to potenciālajiem mērķiem, kas arī ir aptamēru galvenā priekšrocība, kuru nākotnē varētu izmantot pat ikdienišķu darbu izpildes atvieglošanai arī tādos veidos, kādus šobrīd zinātne vēl nav atklājusi.

1.1.1. Aptamēru priekšrocības un nepilnības

Iepriekš jau tika minēts ka aptamērus bieži salīdzina ar antivielu tipa zālēm, šo abu vielu grupu potenciālā pielietojuma dēļ, kas ir savstarpēji līdzīgs. Aptamēriem ir daudz priekšrocības, salīdzinot ar antivielām, kas padara aptamērus par potenciāli ļoti svarīgu nākotnes zāļu grupu, ar daudz un dažādām izmantošanas iespējām, kā arī šobrīd smagu diagnožu ārstēšanai ar lielāku efektivitāti, salīdzinot ar šī brīža zālēm. Kā galvenās var pieminēt: aptamēru mazais izmērs (ap 3 nm); ļoti maza imūnogenitāte; sintēzes, selekcijas, pavairošanas un modificēšanas vienkāršība ar ne tik daudz prasībām (piemēram, temperatūra, pH, utt.); maza ražošanas partiju atšķirība (atkārtojami rezultāti); daudz plašāka mērķu daudzveidība (Chandola, Neerathilingam, 2019).

Tomēr pastāv arī problēmas, ar kurām saskaras aptamēri, piemēram, tie ir salīdzinoši nestabili, un pret tiem šūnās un ārpus tām – asins serumā, iedarbojas daudz dažādas nukleāzes, tādējādi nemodificēti aptamēri, ievadīti intravenozi, cilvēka organismā nevar izturēt ne vairāk kā par desmit minūtēm, bet to ir iespējams uzlabot, tos modificējot. Aptamēri konjugētā formā ar liposomām piesaista lādēto proteīnu daļiņas serumā, veidojot agregācijas riskus un šādā formā var pat kļūt toksiski; šīs pašas problēmas arī var iespaidot aptamēru pielietojamību, kvalitāti un gala efektivitāti (Raouane, *et al.*, 2012.). Protams, kur ir problēmas, tur arī tiek atrasti risinājumi – daļa šo problēmu jau ir atrisinātas, veicot uzlabojumus, modificējot aptamērus.

1.1.2. Aptamēru modifikācijas un to ietekme

Aptamēru specifitāte, afinitāte un farmakoloģiskās īpašības, piemēram, agonisms vai antagonisms uz proteīnu, stabilitāte u.c., ir atkarīgas no aptamēru ķīmiskās struktūras un modifikācijām, ko aptamēros var ieviest, lai uzlabotu aptamēru potenciālo efektivitāti pret vēlamo mērķi, kā arī lai iegūtu vēlamās aptamēra īpašības. Pamatā modifikācijām un specifisku īpašību iegūšanai no aptamēriem izmanto Ellington un Szostak 1990. gada rakstu, kurā trīsdimensionālā aptamēra formācija nosaka aptamēra specifiskumu, afinitāti un saistīšanos ar ligandu, kas rezultē aptmēra efektu. Aptamēru specifiskumu pret mērķi un to afinitāti var atlasīt ar SELEX metodi un augstāku afinitāti var iegūt veicot atlasē ciklus.

Aptamēru modifikācijas, kā iepriekš minēts, tiek izmantotas, lai iegūtu pēc iespējas labāku rezultātu no aptamēriem. Zināms, ka aptamēriem ir dažas problēmas, kuras tieši varētu izlabot

veicot labojumus oligonukleotīdu sekvencēs. Galvenās aptamēru problēmas: nestabilitāte serumā, nevēlama lādēto proteīnu piesaistīšanās.

Aptamēri saistās ar mērķi, pārsvarā aptamēru elektrostatisko īpašību dēļ, veidojot kristālisku struktūru, kā tas ir novērots aptamēriem saistoties ar nukleāro faktoru (NF)- κ B, bakterofāgu MS2 kapsīdu un asins plazmas enzīmu – trombīnu, bet ne visos gadījumos (Adachi and Nakamura, 2019.). Ir novērots, ka aptamērs, kas tika atlasīts kā specifisks pret IgG1, precīzāk, Fc domēnu, saistījās neitrālā domēna vietā, tātad saistīšanās nenotika elektrostatiski, bet ticams ka tā notiek ar vājāku spēku ietekmē, piemēram, ūdeņražu saites, hidrofobie spēki, bet saistīšanās spēja tika novērota kā ļoti stipra (Nomura, *et al.*, 2010.).

Gadās, ka aptamēri saistās arī ne ar mērķi jeb ir novēroti *false-positive* rezultāti. Labs piemērs būtu Mina, *et al.*, 2009. raksts, kur tiek pētīts aptamērs, kas spēj diagnosticēt arsēna klātbūtni, bet diemžēl pētījumā ir radusies kļūda un rezultāts tagad ir pierādīts kā falšs. Pētnieki nebija kārtīgi veikuši pārbaudes - sekvencēšana un sekvenču dimensionālās ietekmes, utt. Izrādās ka veicot aptamēru meklēšanu tika izmantota aparatūra, kas veido zelta nanodaļiņu artifaktus, kas attiecīgi norāda uz nepatiesu pozitīvo rezultātu. Šādas kļūdas ir bīstamas ne tikai tāpēc, ka nepareizs rādījums var novest pie daudzu cilvēku saindēšanās, bet arī apdraudēt dzīvības. Tāpēc, ja iespējams, kļūdas ir obligāti jāmēģina novērst, jo ir mūsdienās jau ir izgudroti daudz veidi, kā veikt sekvenču analīzi un veikt kvalitatīvas pārbaudes (Zong & Liu, 2019.).

Nestabilitāte serumā tiek novērota divu faktoru dēļ: nukleāžu iespaids uz nukleotīdu sekvencēm, tās degradējot; un renālā filtrācija. Šos faktorus var iespaidot, mainot jeb modificējot aptamērus. Nukleāzes var “apmānīt” veicot izmaiņas aptamēru sekvenču mugurkaulā, precīzāk ribozes 2'-OH grupu aizvietojojot ar 2'-F vai 2'-NH₂ grupām. Tika novērots, ka visstabilākie bija aptamēri, kas modificēti tikai ar 2'-NH₂ grupām (ievietoti serumā, tie izturēja ap 2 nedēļām), tālāk sekoja aptamēri ar 2'-F grupu modifikāciju (tie izturēja 3-4 dienas), bet vismazāko stabilitātes uzlabojumu novēroja miksēti modificētos aptamēros (novērotais uzlabojums bija 1,5 dienas). Renālo filtrācijas ātrumu aptamēriem var samazināt palielinot molekulas masu, veicot uz to *PEGylation* jeb kovalenti pievienojot polietilēnglikola molekulu aptamēram (Adler, *et al.*, 2008.).

1.2. Aptamēru pirmsākums

Aptamēri jau tiek pētīti pēdējos 30 gadus, un ir zināms, ka tiem piemīt augsta specifitāte un afinitāte uz noteiktiem mērķiem, piemēram, šūnām un to proteīniem (gan šūnas iekšpusē, gan ārpusē), kā arī uz šūnas DNS, kas atrodas kodolā. Pats pirmais pētījums bija balstīts uz 1980. gadā veikto HIV un adenovīrusu izpēti, kur uzzināja ka šie vīrusi ir veidoti no nukleīnskābu ķēdītēm, gluži kā RNS, un ir spējīgi nokļūt specifiskā vietā, precīzāk, nokļūt uz specifisku mērķa šūnu, piemēram, HIV vīrusam ir izveidojusies īsa, strukturizēta RNS ķēde, ko sauc par *HIV TAR* elementu, kas spēj saistīties virālajam proteīnam *tat* un šūnas proteīnam *ciklīns T1*, kas atrodas uz imūnšūnām, lai kontrolētu virālo gēnu replikāciju un ekspresiju. Šādi novērojumi noveda zinātniekus uz domu par iespēju izmantot šādas RNS ķēdes lai ārstētu slimības, ieskaitot arī šādus vīrusus. 10 gadus vēlāk, 1990. gadā tika veikts pētījums, kas apstiprināja iespējamību sintezēt īsas RNS sekvences, kuras varētu izmantot terapeitiskiem nolūkiem. Šis – otrs pētījums, bija saistīts ar HIV vīrusa ārstēšanas iespējām (Dollins, *et al.*, 2008.).

Sekojošiem diviem 1990. gadā veiktajiem pētījumiem, kad arī tika atklāta RNS sekvenču priekšrocība un iespējamais pielietojums terapeitisku mērķu sasniegšanai, tika izstrādāta šūnu SELEX metode.

Šī metode ir perfekts piemērs, kā ērti iegūt un izfiltrēt milzīgu daudzumu aptamēru izmantojot to specifitāti un afinitāti uz mērķa šūnu vai citu mērķa elementu. Caur SELEX var iegūt milzīgas datu bibliotēkas ar RNS un DNS, kuri ir potenciāli atbilstoši un ar augstu specifiskumu un afinitāti tiešajam mērķim un nesaistās ar citām šūnām (šīs īpašības atlasa ar īpašām metodēm). SELEX kopš atklāšanas brīža izmanto vēl aizvien, tikai uzlabotā un/vai modificētā veidā, lai izfiltrētu jaunus aptamērus (Dollins, *et al.*, 2008.).

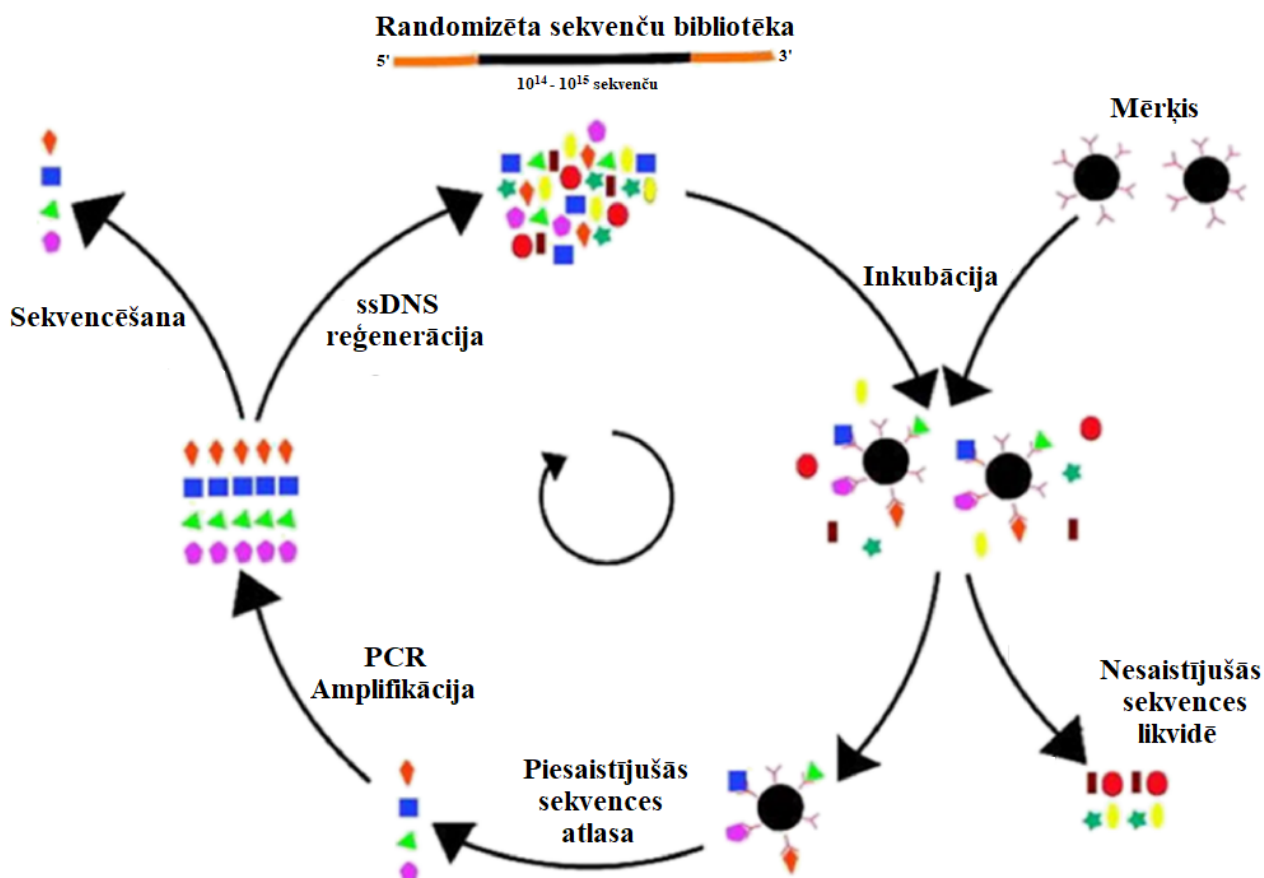
1.3. PCR

PCR - *polymerase chain reaction*, kas latviski tulkojumā tiek saukta par polimerāzes ķēdes reakciju. Šīs metodes mērķis ir pavairot jeb amplificēt DNS sekvenču un specifiskus DNS sekvenču reģionus. Jāņem vērā, par cik PCR metodē strādā ar nukleotīdu ķēdēm – ir daudz svarīgu faktoru, kas var ietekmēt sekvenču struktūru, un citas farmakoloģiskās un fizioloģiskās īpašības (runājot par aptamēru selekcijas metodi). Mūsdienās šos faktorus aprēķina un paredz īpaši datoru algoritmi.

Pamata PCR metode sastāv no četriem soļiem: (1.) pirmapstrāde – dsDNS (divu pavedienu DNS) pārveidošana par ssDNS (viena pavediena DNS), tad paraugu uzgarsē un tad atdzesē – tādējādi pievieno praimerus; (2.) maisīšana un centrifugēšana; (3.) amplifikācija – notiek ar polimerāžu palīdzību, tai ir nepieciešami īpaši apstākļi – praimeru (5' un 3' sekveces galā) un silta temperatūra (ap 70 °C); (4.) amplifikācijas rezultātu novērtēšana – datu apstrāde (Sigma-Aldrich, 2020.).

1.4. SELEX

SELEX ir ļoti svarīga un parocīga metode, ar kuras palīdzību viegli un salīdzinoši ātri var atklāt un izfiltrēt milzīgus daudzumus aptamēru. (SELEX metode pārskatāma *1.4.1 att.*). Parasti tiek apstrādāts ap 10^{14} – 10^{15} aptamēru, kas ir sākotnēji nejaušas 30-100 nukleotīdu garas sekvenču, kuru sintēze ir standartizēta un atkārtojama. Šīs sekvenču ievieto kopējā inkubētā traukā ar mērķšūnām vai mērķa molekulām. Tad ļauj notikt aptamēru saistīšanās reakcijām ar šūnām vai molekulām, un tās atdala no pārējām, kas nav saistījušās. Šo procesu sauc par pozitīvo atlasīšanu, jo tiek atlasīti aptamēri, kuri ir spējīgi saistīties ar pietiekamu afinitāti - mērķim. Pastāv arī negatīvā atlase, kura tiek izmantota, lai, pretēji pozitīvajai atlasei, atlasītu aptamērus kas nesaistās ar mērķi, un kad šos oligonukleotīdus noskalo jeb no tiem atbrīvojas, paliek piesaistījušies aptamēri. Šīs izfiltrētās, saistīties spējīgās nukleotīdu ķēdes kalpo par pamatu jaunām bibliotēkām, un tās tiek pavairinātas jeb amplificētas, ko veic ar PCR. Galvenā SELEX priekšrocība ir iespēja metodes laikā kā mērķi izmantot dzīvas šūnas, jo tā var precīzāk imitēt un veikt pētījumus pēc iespējas līdzīgākai videi, kāda ir organismā. Noderīgi arī tas, ka ar šo metodi par mērķi ir iespējams izmantot molekulas, kas ir pāris atomu lielumā, līdz pat multidomēnu proteīniem un veselām šūnām (Ohuchi, 2012.).



1.4.1. att. **SELEX metode**, (redīgēts). Šī ir pamata metode, kuras soļi pamatā būs līdzīgi visām šīs metodes “atvasēm” jeb modificētajām SELEX metodēm. Randomizētas ssDNS (*single-stranded* DNS) sekvenses ar praimeriem 5' un 3' galos tiek inkubētas kopā ar mērķi (mērķis var būt gandrīz jebkas – sākot no mazām molekulām, proteīniem, līdz pat veselām šūnām), kur pret mērķi afinitīvi saistīties spējīgās sekvenses saistās ar mērķi, bet tās kuras nespēj, tiek likvidētas jeb aizvāktas. Piesaistījušās sekvenses tālāk tiek atlasītas un amplificētas. Gala rezultātā notiek sekvenču sekvenču un ssDNS bibliotēkas reģenerācija (šoreiz jau ar mērķi saistīties spējīgu sekvenču būs vairāk, kā iepriekš. Šādi atlasē cikli tiek veikti vairākas reizes, tā nodrošinot sekvenču afinitāti un savā ziņā arī specifiskumu (kas saistās ar noteikto mērķi) (Mehta, *et al.*, 2012.).

SELEX var tikt veikts gan *in vivo*, gan *in vitro*, ar to domāts, ka šī aptamēru selekcija var notikt gan uz mērķiem, kas ir dzīvos organismos, kā piemēram aptamēri, kas tiek injecēti dzīvā žurkā (*in vivo*), gan uz mērķiem, kas atrodas, piemēram petri traukā, kuri var būt dzīvas šūnas, proteīni vai pat molekulas. *In vivo* ir salīdzinoši sarežģītāka un komplicētāka selekcijas metode,

galvenokārt jo aptamēru iegūšana ir salīdzinoši bīstamāka – jāiegūst mērķa šūnas, uz kurām ir piesaistījušies aptamēri, tā lai netiktu bojāti iegūtie paraugi, kā arī šādas procedūras atkārtošana ir laikietilpīgāka (Doria-Rose & Vogt, 1998.).

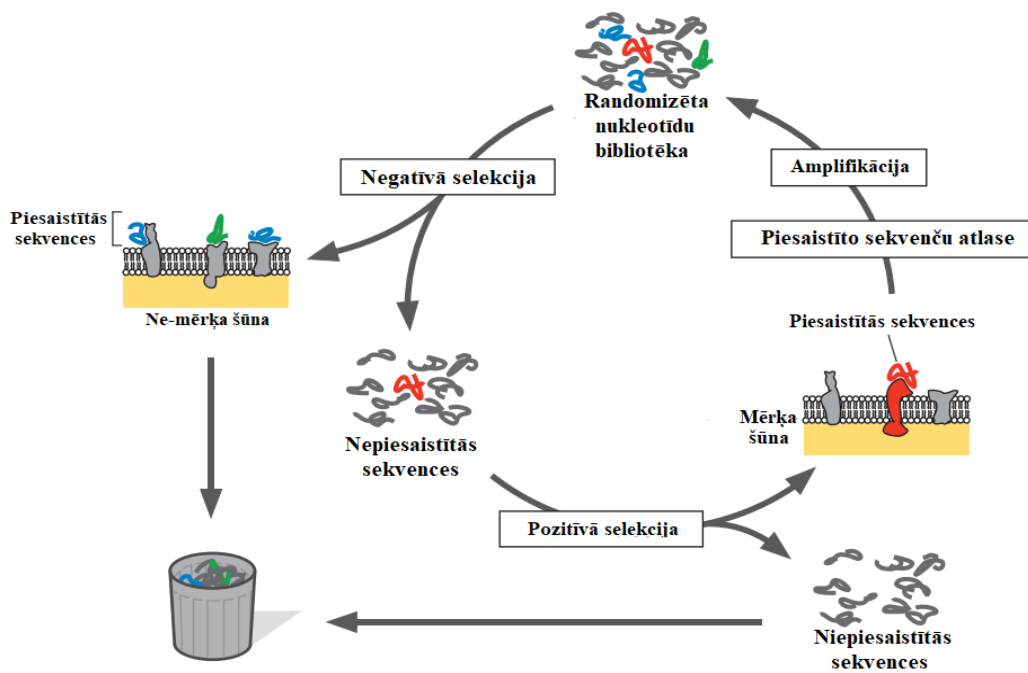
Pēc SELEX metodes izmantošanas, kas parasti iekļauj vairākus atlasēšanas ciklus (dažviet minēts, ka tiek veikti līdz pat 20 cikli), tiek iegūti jeb izfiltrēti vairāki aptamēri ar vajadzīgo afinitāti un specifiskumu uz mērķi. Ļoti labi, ka šī metode ir modificējama, lai pielāgotos aptamēra terapeitiskajam, diagnostikas vai cita veida pielietojumam.

Svarīgi pieminēt, ka SELEX metodei ir daudz dažādu variāciju, kas ir modificētas no pamata SELEX metodes, lai varētu ērtāk veikt aptamēru selekciju, attiecīgi aptamēru prasībām. Kā nozīmīgākos piemērus var pieminēt, (1.) *Cell-SELEX*, (2.) *Capillary Electrophoresis-SELEX* (jeb CE-SELEX) (kapilārās elektroforēzes SELEX), (3.) *rapid-SELEX*, (4.) *microfluidic SELEX*, (5.) *Magnetic Bead-SELEX*, (6.) SPR-SELEX (angliski - *surface plasmon resonance SELEX* jeb latviski - virsmas plazmonu rezonanses SELEX).

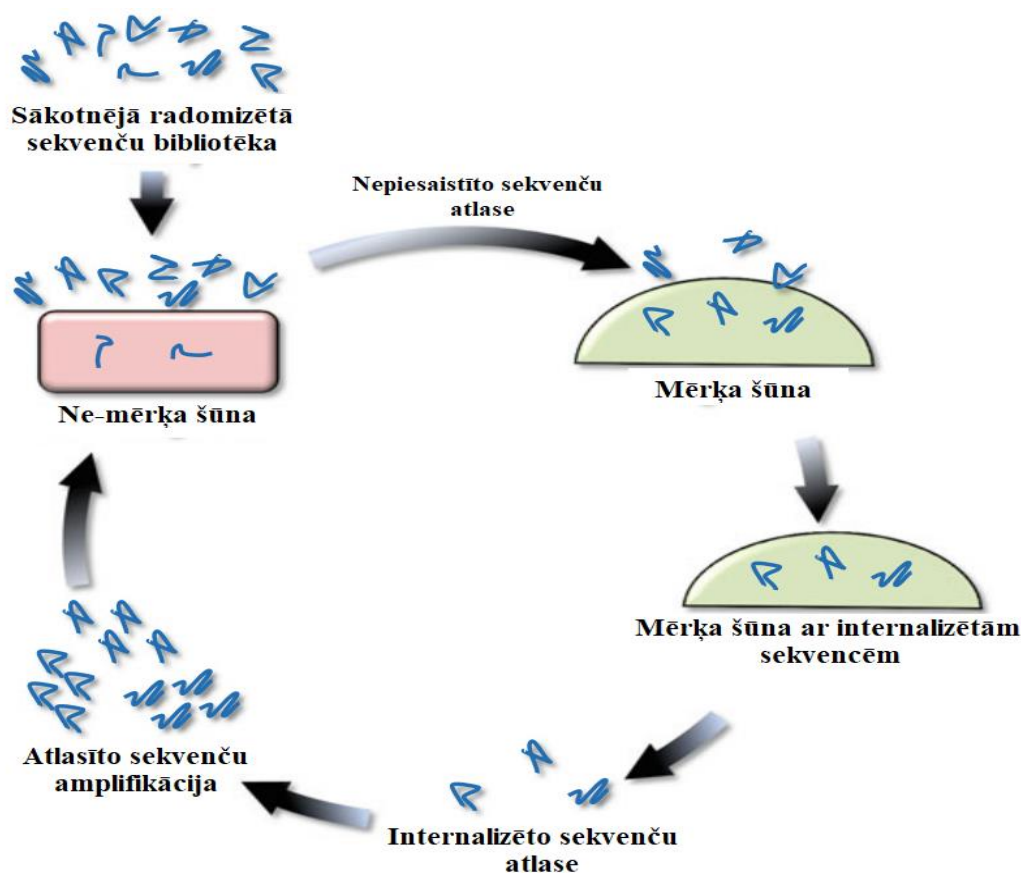
1.4.1. *Cell-SELEX*

Cell-SELEX, kas latviski tiek saukts par šūnu-SELEX. Metodes gaita attēlota 1.4.1.1. att. Galvenās izteiktās pazīmes jeb īpašības ir, ka mērķis ir šūnas proteīni, kas atrodas uz vai iekšā dzīvā šūnā; otra īpašība ir tāda, ka tiek veikta gan pozitīvā selekcija, kurā ir šūnas ar ekspresētu mērķ-proteīnu (citi proteīni ir “aizvākti”, tātad šūnas ir modificētas), gan negatīvā selekcija, kurā izmanto tāda paša veida šūnas, tikai šīm šūnām nav mērķ-proteīna, tādējādi nespecifiskās nukleotīdu ķēdes tiek atlasītas un netiek tālāk uz PCR pavairošanu jeb amplifikāciju. Pozitīvā atlasēšana paredzēta, lai atlasītu aptamērus ar pēc iespējas augstāku afinitāti, un negatīvā atlasēšana ļauj atlasīt aptamērus ar pēc iespējas augstāku specifiskumu pret mērķi (Ohuchi, 2012.).

Cell-SELEX metodei līdzīga ir *Cell-internalization SELEX*. Šīs metodes pamatā ir *Cell-SELEX*, bet tās gaita ir izmainīta jeb metode ir modificēta, lai to var izmantot internalizētu aptamēru iegūšanai. Metodes gaita internalizētu sekvenču iegūšanai attēlota 1.4.1.2. att. Galvenā atšķirība no *Cell-SELEX* ir, ka veicot internalizēto sekvenču atlasēšanu, netiek atlasītas sekvences - kas nav saistījušās ar šūnas membrānas proteīniem, un arī sekvences - kas ir saistījušās ar šūnas membrānas proteīniem, bet nav veiksmīgi internalizējušās šūnā. Pārējās darbības norisinās tāpat kā *Cell-SELEX* (Thiel *et al.*, 2015).



1.4.1.1. att., **Cell-SELEX metode** (rediģēts). *Cell-SELEX* metodes pamatā ir iepriekš jau pieminētā SELEX metode, tikai šī metode ir modificēta jeb mainīta, lai atlasītu aptamērus, kuri specifiski saistās ar specifiskiem šūnu membrānu proteīniem. Sākumā randomizēta bibliotēka (tādai pašai kā lielākajai daļai citu SELEX metožu) iziet negatīvajai selekcijai, kur bibliotēka tiek inkubēta ar ne-mērķa šūnu – šoreiz piesaistījušās sekvenču tiek aizvāktas, bet nepiesaistījušās sekvenču tiek atlasītas un atkal tiek inkubētas, bet šoreiz ar šūnu uz kuras atrodas mērķa proteīns. Piesaistījušās sekvenču tiek atlasītas un uz tām tiek veikta amplifikācija, kamēr sekvenču kas atkal nepiesaistās – tiek likvidētas (šo sauc par pozitīvo selekciju (Ohuchi, 2012.).



1.4.1.2. att., **internalizētu sekvenču atlase izmantojot Cell-SELEX** (redigēts). Šajā attēlā tiek attēlota Cell-SELEX metode, kur specifiski tiek atlasītas sekvenses, kas ir spējīgas saistīties un internalizēties jeb iekļūt mērķa šūnā. Galvenā atšķirība no parastajām sekvencēm – likvidētas tiek visas sekvenses, kuras ir internalizējušās ar mērķa šūnu (pozitīvās selekcijas laikā) (Thiel, *et al.*, 2015.).

1.4.2. Capture SELEX

Iepriekš jau tika minēts, ka SELEX kopumā ir ļoti svarīga un parocīga metode, bet tai ir arī savas problēmas un ierobežojumi, kas ietekmē aptamēru atlases pilnvērtīgu izmantošanu. Kā galveno no problēmām – sekvenču imobilizāciju, kas tiek veikta ekstrēmu faktoru apstākļos, un var pat denaturēt oligonukleotīdu ķēdes, piemēram, pH, šķīdinātāju ietekme un citi faktori, kas var ietekmēt sekvenses, pat pirms ir sāka selekcija, tādējādi būtiski ietekmējot bibliotēkas, kuras iegūst veicot šo metodi, padarot sekvenses nepilnīgas un pat ar samazinātu specifiskumu un afinitāti pret mērķi. Sekvenses, šādu apstākļu klātbūtnē, var arī zaudēt spēju saistīties ar *linker* molekulu (piemēram biotīna molekulu), kas piesaista sekvenses, piemēram, matricas, lai uz tām varētu veikt PCR un citas SELEX metodes procedūras (Boussebayle, *et al.*, 2019.). Pašas metodes pamatā, protams, aizvien ir SELEX metodes princips, bet tā ir izmainīta jeb pielāgota, lai

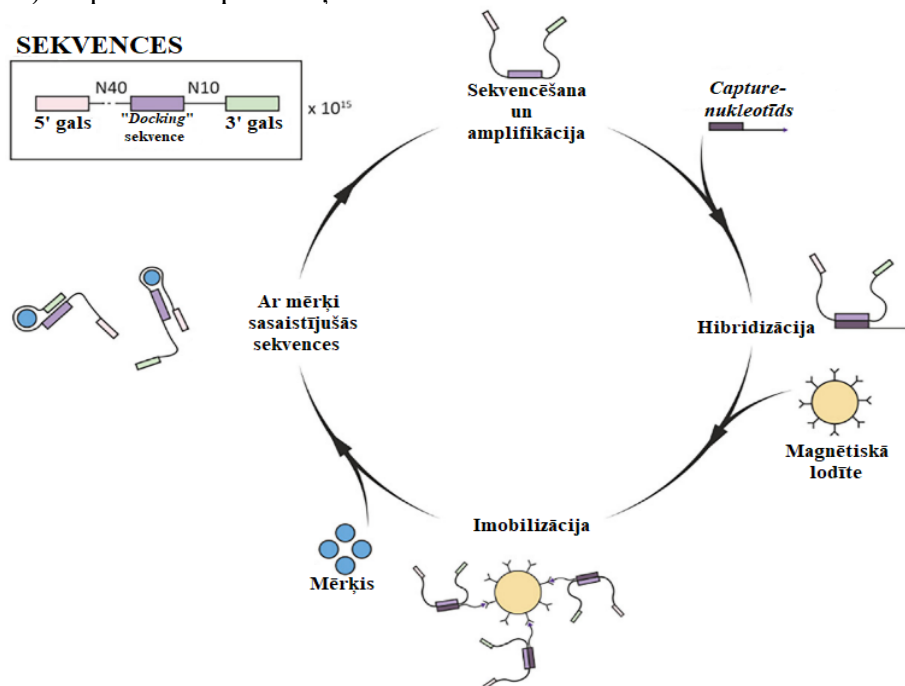
samazinātu tieši šo imobilizācijas soli un samazinātu negatīvos faktorus, kas ietekmē oligonukleotīdu sekvences.

(Metodes gaita attēlota 1.4.2.1. att.) Ir trīs galvenās atšķirības no citām SELEX metodēm:

Šīs metodes pirmā atšķirība ir sekvenču bibliotēkā, kas nav vienkāršas sekvences, ar konstantām praimeru piesaistīšanās daļām 5' un 3' galos (paredzētas PCR pavairošanai), bet vēl pa vidu, randomizētajai sekvenču daļai, ir pievienota “*docking*” daļa, kas nodrošina metodes izpildīšanos, tātad ir 5' gals, kur ir praimeru piesaistīšanās sekvenču daļa, tai seko četrdesmit nukleotīdus gara randomizēta sekvenču daļa, tad šī “*docking*” daļa, otra randomizēta nukleotīdu daļa, kura ir desmit nukleotīdiem gara, un galā 3' gala “*primer*” sekvenču daļa.

Otra atšķirība no pārējām SELEX metodēm ir imobilizācijas procesā – sekvenču tālāk tiek savienotas “*docking*” daļā ar *capture* nukleotīdu, kas atbilst par tālāku saistīšanos ar “*magnetic beads*” jeb magnētiskajām lodītēm, kas kalpo kā sekvenču imobilizācijā izmantotās matricas aizvietotājs. Nesaistījušās sekvenču daļas aizskalo no parauga, paliek tikai imobilizētās sekvenču daļas.

Trešā atšķirība ir, ka paraugam, kas satur imobilizētās sekvenču daļas, tiek pievienots mērķis (parasti SELEX metodē, sekvenču bibliotēku pievieno mērķi saturošajam paraugam). Sekvenču daļas, kas ir spējīgas saistīties ar mērķi, atsaistās no *capture* nukleotīda daļas (to atstāj piesaistītu ar magnētisko lodīti) un piesaistās pie mērķa.



1.4.2.1. att. **Capture SELEX metodes gaita** (redīgēts). Attēlota *capture* SELEX metode, kuras īpašas vairāku iemeslu dēļ, bet kā galveno var nosaukt – īpašo sekvenču uzbūvi, kuras sastāda

randomizēto bibliotēku (5' galā praimeris, kam seko 40 nukleotīdu gara randomizēta sekvenca, tad pa vidu *Docking* sekvenca, kas nodrošina imobilizāciju, kas neietekmē sekvenci tik ļoti, kā citu SELEX metožu imobilizācijas metodes, tai seko otra randomizētā sekvences daļa, kas ir desmit nukleotīdus gara, beigās tai pievienots 3' gala praimeris). Sekvences savieno iepriekš jau pieminētajā *docking* daļā ar *capture*-nukleotīdu – notiek hibridizācija, un jaunizveidojusies hibrīdmolekula pievienojas pie vienas no magnētiskajām lodītēm, pie kuras pievienojas arī vēl citas šādas hibrīdmolekulu sekvences. Nākamais solis ir mērķa pievienošana imobilizētajām sekvencēm, kas ir arī pretēji citām SELEX sekvencēm, kuras parasti pievieno sekvenču bibliotēku mērķim. Saistīties spējīgas (pret mērķi specifiskās sekvences saistās ar mērķi vienlaicīgi atdaloties no *capture* nukleotīda, kas paliek pie magnētiskās lodītes. Tad šīs sekvences, kuras saistījušās ar mērķi un vienlaicīgi atdalījušās no *capture* nukleotīda tiek atlasītas, sekvencētas un amplificētas (Boussebayle, *et al.*, 2019.).

1.5. Aptamēri un to pielietojums

Pateicoties SELEX ir iespējams iegūt lielu variāciju aptamēriem, un tādējādi palielinot specifitāti un aptamēru pielietojumu. Šobrīd aptamērus var pielietot daudzveidīgi, iegūstot dažādus rezultātus. Piemēru ir daudz, bet svarīgi ir uzskaitīt galvenos aptamēru pielietošanas ceļus, kā arī parādīt to pielietojumu daudzveidību un, iespējams, pat spekulēt par nākotnes iespējām.

Šobrīd pasaulē ir izsludināta slimības Covid-19 pandēmija un ir problemātiski veikt testēšanu pietiekami ātri un efektīvi (ekonomiski un pēc iespējas vienkāršākā veidā), lai diagnosticētu slimības klātbūtni pacientos. Iespēju izmantot aptamērus, kas saistīti kopā ar fluoriscējošu modifikāciju, kas zem ultravioletās gaismas fluoriscē, un atšķirībā no krāsas var noteikt vai rezultāts ir negatīvs – aptamērs nav saistījies ar mērķi, vai rezultāts ir pozitīvs – aptamērs ir saistījies ar mērķi. Šis ir mūsdienu augstākais sasniegums, ņemot vērā cik ātri šī diagnosticēšanas metode tika izstrādāta (kopš slimība tika atklāta). Šo metodi sauc par Mango, izteikto krāsu dēļ, kas novērojamas veicot metodi (Cawte, Unrau and Rueda, 2020).

Izrādās, ka aptamēri jau ir atrodami dzīvos organismos, tie atrodami veidojumos, ko sauc par *riboswitch* – RNS sekvenca, kas ir kā sastāvdaļa mRNS (matricas jeb informācijas RNS) kuru nolasa ribosoma, lai veiktu jaunu proteīnu sintēzi, šī sekvenca spēj “ieslēgt” vai “izslēgt” gēnus, kas ir uz noteiktās mRNS, ja tai piesaistās noteikta molekula, šādi veicot auto-regulāciju. Šos *riboswitches* var arī modificēt, lai tos izmantotu diagnostikā, vai arī laboratorijas pētījumos, varbūt ar laiku šis atklājums būs praktiski izmantojams citās bioķīmijas apakšnozarēs (Bédard, *et al.*,

2020.). You, *et al.*, 2015. gadā publicētajā rakstā minēts īpatnējs pielietojums, kur šo *riboswitch* modificē tā, lai tas saistoties ar ligandu mainītu struktūru un sāktu radīt signālu, kas ir nolasāms ar īpašām metodēm.

Atromdami arī alternatīvi veidi, kas uzlabo cilvēku dzīves kvalitāti un palielina drošību, piemēram, Zhu, *et al.*, 2018. gada rakstā minēts, ka sēnīšu producētā aflatoksīna, kas ir ļoti bīstams cilvēkiem, un it īpaši ietekmē bērnus, klātbūtni graudos var viegli un īsā laikā identificēt ar aptamēru palīdzību, tādā veidā izglābjot daudz cilvēku dzīvību.

Aptamēri šobrīd tiek izmatoti ārstnieciskiem terapijas mērķiem – gan kā transports, kas transportē zāļu vielu, kas piesaistīta tam, piemēram, ķīmijterapijā (Dou *et al.*, 2018), gan arī kā tiešas zāles, kas saistās ar receptoriem kā agonists vai antagonists, vai ietekmē organismu un šūnas citos veidos (Shahid, *et al.*, 2017.). Šobrīd šāda veida pielietojums un arī diagnostikas mērķi ir galvenie interesējošie aptamēru nākotnes pielietojuma veidi, kuros sagaidāms vislielākais potenciāls.

1.6. Intenalizējoši aptamēri

Internalizācija šūnā notiek, ja aptamērs saistās ar mērķa proteīnu, kas atrodas uz šūnas membrānas. Pēc saistīšanās ar proteīnu tas iekļūst šūnā un var veikt attiecīgo funkciju, dēļ kuras tika radīts, piemēram, atbrīvoties no sev piesaistītā siRNS (*silencing* RNA) vai zāļu vielas, kas uzsāk savu darbību, tādējādi ietekmējot tikai tās šūnas, kurām uz membrānas ir noteiktais proteīns, tādējādi nodrošinot specifiskumu un drošāku zāļu lietošanu (Thiel, *et al.*, 2015.).

Kā piemēru var nosaukt prostatas vēža šūnu specifisku aptamēru, precīzāk, tas ir specifisks uz antigēnu, kas atrodas tikai uz prostatas vēža šūnu virsmas (specifiskuma nodrošinātājs kā mērķis), tad šim aptamēram tika pievienota doksorubicīna molekula, kas šādā kombinācijā dosies tieši uz prostatas vēža šūnām, kurās internalizējas endocitozes ceļā. Kad doksorubicīna molekula, kas saistīta ar aptamēra sekvenci iekļūst šūnā, tā tiek atdalīta no aptamēra sekvences un aktivizējas, tādējādi iznīcinot vēža šūnu caur apoptozes ceļu, un neietekmējot veselās prostatas šūnas (Min, *et al.*, 2011.).

1.7. Aptamēri klīnikā

Aptamēri šobrīd jau tiek lietoti lai ārstētu dažādas pataloģijas, kā arī šobrīd ir vairākas zāles, kas jau tiek pētītas klīniskajos pētījumos. Pats pirmais apstiprinātais (gan Eiropas Savienībā, gan

Amerikas Savienotajās Valstīs) un klīnikā izmantotais medikaments, kas ir aptamērs, ir pegaptanib (tirdzniecības nosaukums – Macugen), ko izstrādāja Eyetech Pharmaceuticals un Pfizer. To izmanto neovaskulārās (mitrās), ar vecumu saistītās makulas deģenerācijās. Šī aptamēra mērķis ir vaskulārā endotēlija augšanas faktora inhibīcija (M.D. Gragoudas *et al.*, 2004). Pegatanib ir 28 nukleotīdo bāzu garš RNS oligonukleotīds, kam kovalenti piesaistītas divas polietilēnglikola grupas jeb, var teikt ka, veikta *PEGylation* modifikācija, kas nodrošina labāku stabilitāti serumā (Kourlas, Schiller, 2006.).

1.8. Pašlaik lietotie analīžu rīki (kas tiek lietoti lai analizētu datus)

1.8.1. Sekvenču sekvencēšana

1.8.1.1. Sanger sekvencēšanas metode

Pirmā veiksmīgā un daudzviet pasaules laboratorijās izmantotā sekvencēšanas metode bija Sanger sekvencēšana, kas nosaukta metodes radītāja Frederika Sangers (*Frederick Sanger*) vārdā. Pateicoties Sangers metodei tika radītas un attīstītas jaunas - efektīvākas, lētākas un ātrākas metodes, kas ir padarījušas sekvencēšanu viegli pielietojamu un par daļu no daudzām citām kompleksām metodēm, it īpaši aptamēru selekcijā.

Šī metode balstās uz četriem soļiem: (1.) DNS ekstrakcija un atšķaidījuma (jeb DNS sadalīšana par ssDNS); (2.) PCR – šo ssDNS pavairošana, lai varētu veikt tālākus soļus; (3.) pavairoto sekvenču purifikācija no PCR produktiem, kas var traucēt un samazināt DNS kvalitāti tālākai apstrādei; (4.) DNS sekvencēšana – DNS fragmentus sekvencē ar īpašu metodi, kur sekvencēm tiek pievienoti īpaši bāzu pāri – katram šādam īpašajam nukleotīdam ir savas krāsas molekula, kas pretējai bāzei piesaistoties ķīmiski izreagē un izdala specifisku gaismas signālu, ko nolasa īpašais sensors jeb jūtīga kamera, kas spēj nolasīt vienlaicīgi milzīgu daudzumu šādu signālu vienlaicīgi, tad sekvenču sašķiro pēc to izmēriem (izmanto elektroforēzi), tad rezultātus apstrādā saliekot DNS fragmentus kopā, līdz tiek iegūta pilnā sekvenču (Carpente, *et al.*, 2019.).

1.8.1.2. Augstas veikspējas sekvencēšanas metodes izmantošana aptamēru atlasei

Augstas veikspējas sekvencēšana jeb, no angļu valodas, *high throughput sequencing*, ir metode, kuru izmanto nukleotīdu sekvencēšanai noteiktās sekvenču, piemēram cilvēka DNS sekvencēšanai, aptamēru sekvenču sekvencēšanai. Aptamēru selekcijas veikšanā šo metodi

izmanto starp katru atlasē ciklu SELEX metodē. To dara lai novērotu sekvencēšanas gaitu, tādējādi laicīgi, jau pirmajos atlasē ciklos, spētu identificēt uz mērķi visspecifiskākos aptamērus, ar augstāko saistīšanās spēju, vai izsekotu līdzī vispārējai SELEX metodes gaitai; lai dziļāk izprastu - SELEX metodes procesus un uzlabotu selekcijas efektivitāti, aptamēru-mērķu saistīšanos un citas savstarpējas īpašības (Guo, et al., 2014.).

1.8.2. Sekvencēšanas iekārtas

Gan *short-read*, gan *long-read* cDNS (*complimentary*DNS jeb komplimentārais DNS pavediens) sekvencēšanas metodes sastāv no līdzīgiem izpildes punktiem, bet visās metodēs ir adaptoru pievienošana sekvencēm, un visu metožu rezultāti ir atkarīgi no paraugu kvalitātes un datu, paraugu apstrādes gan pirms sekvencēšanas veikšanas, gan pēc tās, apstrādājot sekvenču bibliotēkas (Stark, et al., 2019.).

***Illumina workflow* sekvencētājs** – šī ierīce veic nākamās paaudzes (*next-generation*) sekvencēšanu uz iepriekš apstrādātu bibliotēku, kurā jau ir izolētas un attīrītas sekvenses, kuras tālāk veidos bibliotēku, kā arī RNS sekvenses ir jāpārveido par cDNS sekvencēm (ar reverso transkriptāzi), un vēl strukturāli ietekmētas sekvenses tiek, tām pievienojot praimerus 3' un 5' galos, lai varētu veikt PCR pavairošanu. Svarīgi, lai tiktu veikta kvalitātes kontrole, pirms sāk veikt sekvencēšanu, piemēram ultravioletās spektrometrijas izmantošana tīrības noteikšanai.

Tālāk seko klasteru veidošana, kurā sekvenses (ssDNS) tiek piesaistītas ar likeru molekulām un pavairotas ar sekvenču polimerāzi, tādējādi nodrošinot tālākās sekvenses pavairošanu sekvencēšanai, šī daļa notiek automātiski, bez cilvēka palīdzības (Stark, et al, 2019.).

Tad nākamais solis ir sekvencēšana. *Illumina workflow* sekvencētājs izmanto *flow cell* jeb plātnītes, uz kurām ir izveidoti un izvietoti iepriekš minētie sekvenču klasteri, kas tiek nolasīti izmantojot īpašu nukleotīdu pievienošanu. Šie nukleotīdi ir īpaši ar to, ka tiem ir piesaistīta fluorescējoša molekula, kas nukleotīdam saistoties ar pretējo nukleotīdu bāzi sekvencē, kura tiek sekvencēta, ķīmiskas reakcijas rezultātā iespīdas – radot nolasāmu, īpašu signālu, kuru uztver iekārta, attiecīgajam signālam iespīdoties, un atzīmē sekvencē nākamo nukleotīdu bāzi (ar C vai T, vai G, vai A burtu). Šādi tiek iegūti vairāki sekvenču fragmenti, kas tālāk tiek apstrādāti datu analizē, lai no šo sekvenču fragmentiem izveidotu kopējo sekvenci. Jāpiemin, ka katru reizi kad fragments ir nolasīts no 5' gala uz 3' galu, šo fragmentu aizskalo, un nākamā sekvencēšanas reize tiek veikta otrādāk (no 3' uz 5' sekvenses galu), šādu metodi sauc par *paired-end* sekvencēšanu.

Un, protams, pašās beigās ir datu apstrāde, kur šie sekvenču fragmenti tiek sapāroti un salikti kopā, reversie lasījumi kopā ar ne-reversajiem, galā veidojot kopēju, vienu sekvenci, gluži kā saliekot puzzles gabaliņus lai veidotu vienu attēlu, tikai parasti rezultātā nav viena sekvence bet ap miljoniem sekvenču, kuras jāapstrādā vēl tālāk, lai uzzinātu sekvenču potenciālo efektu, lai sauktu kādu no tām par aptamēru (Illumina Inc., 2020.). Sekvences var būt no 50 līdz 500 bāzu pāriem garas (Stark, *et al.*, 2019.).

2. Laboratorija

2.1. Izmantotās metodes un materiāli

Pirms darba izpildes jā sagatavo programmatūra: Ubuntu 20.04 LTS release on Windows (tālāk Ubuntu); Anaconda for Python 3.7 (tālāk Anaconda); Cutadapt v2.10 (tālāk Cutadapt); fastQC 0.11.9 (tālāk fastQC); FASTAptamer v1.0.14 (tālāk FASTAptamer); R version 4.0.0 (tālāk R); RStudio Desktop 1.2.5042 (tālāk RStudio).

2.1.1. Sākotnējā datu apstrāde

Sākotnējai datu ievadei izmantots .fastq standarta fails, kas sekvencēšanas rezultātā iegūts, izmantojot *Illumina MiSeq* augstas veiktspējas sekvencēšanas iekārtu. Dati tiek apstrādāti Ubuntu terminālī, tā lai tos var analizēt tālāk, RStudio programmā. Lai tos varētu apstrādāt būs nepieciešamas dažas papildus programmas un pakotnes (angliski *packages*), ar kurām, pēc pareizas izmantošanas secības un pareizajām komandu virknēm var iegūt nolasāmus datus, kurus varēs tālāk analizēt.

Šīs programmas ir: **fastQC**; **cutadapt**; **FASTAptamer**.

FastQC – ir programma, kuru izmanto uzreiz pēc sekvencēšanas paveikšanas, iegūstot pirmos datus. Tās pamata funkcija ir kvalitātes kontroles veikšana (var noprast no nosaukuma - “QC” daļas, kas simbolizē no angļu valodas *quality control* jeb kvalitātes kontrole). Šī programma apstrādājot iegūtos sekvencēšanas datus, norāda uz nepilnībām, kuras ieviesušās bibliotēkas veidošanas, apstrādes un sekvencēšanas procesā (Andrews, 2010.).

Pēc sekvencēšanas iegūtie faili ir “.gz” formātā, kurus nepieciešams apstrādāt – šī programma ir pirmā, kas veic datu apstrādi – veicot kvalitātes kontroli uz tiem. Rezultātā tika iegūti faili kas ir mainījuši formātu uz “.fastq” un ir pārbaudīti, un izfiltrēti. Kopā no divpadsmit failiem, aizvien ir divpadsmit, un šis skaits nesamazināsies līdz pat sākotnējās apstrādes beigām.

Cutadapt – ir nākamā programma, kas atlasa sekvenču, kurām pēc sekvencēšanas ir nolasīti, iepriekš, bibliotēkas sagatavošanas laikā, pievienotie praimeru jeb adapteru 3’ un 5’ galos. Ar šo programmu iespējams, kā jau iepriekš minēts, atlasīt šīs sekvenču un no tām noņemt adapterus – konstantās sekvenču daļas, lai dati būtu vienkāršāk apstrādājami un sekvenču, kuras ir ar meklētajām bioloģiskajām īpašībām, varētu nolasīt bez, visās oligonukleotīdu ķēdēs konstantajām, praimeru sekvenču (Marcel, 2011.).

Palaižot programmu caur Ubuntu termināli, tā apstrādā failus visus divpadsmit failus, atdalot adapterus no 5' un 3' gala un izfiltrējot zemas kvalitātes bāzes no šiem galiem. Programma veicot atlasī un adapteru atdalīšanu izveido ziņojuma failu “*report*”, kur var pārskatīt dažādu informāciju par veikto aptamēru atlasī jeb izfiltrēšanu caur šo programmu, kā galvenā nolasāmā informācija, ko var nolasīt šajā ziņojuma failā ir sekojoša:

=== Summary ===

1. Total reads processed: 1,365,266
2. Reads with adapters: 1,076,323 (78.8%)
3. Reads written (passing filters): 1,365,266 (100.0%)

4. Total basepairs processed: 206,155,166 bp
5. Quality-trimmed: 28,849,574 bp (14.0%)
6. Total written (filtered): 60,689,376 bp (29.4%)

(Ziņojuma faila redzamā informācija ņemta no veiktās analīzes faila “*f_rna_1_1.fastq*” (fails kas iegūts pēc fastQC programmas datu apstrādes), šis ir viens no visiem divpadsmit šādiem failiem, uz kuriem arī tika veikta šāda apstrāde.)

FASTAptamer – programma ir pēdējā, kura izmantota primārajā datu apstrādes daļā. Tās pamata funkcijas un izpildes mērķi ir iepriekš minēto programmu atlasīto datu apstrāde un sakārtošana, tālākai lietošanai. Tā veic vienkāršu datu filtrēšanu un kārtošānu – saskaitot, normalizējot, sarindojot (*ranking*) un izšķirojot katru iegūto sekvenci, arī tās salīdzinot starp populācijām, sagrupējot sekvenču kopā - izmantojot *Levenshtein* attāluma īstenošanu, aprēķinot *fold-enrichment* un meklējot deģenerācijas oligonukleotīdu sekvencēs (Alam, *et al.*, 2015.).

Veicot šo (jau ar fastQC un cutadapt apstrādāto) divpadsmit failu apstrādi ar FASTAptamer iegūst failus, kas, kā jau iepriekš minēts, veic datu sakārtošanu, kā arī beigās ar īpašu komandu “*saspiež*” visus šos divpadsmit failus par tikai sešiem failiem “.tsv” formātā, kurus var tālāk izmantot sekundārajā apstrādē – Rstudio programmā. Šie seši faili viegli tika sadalīti, lai vieglāk varētu veikt secinājumus un atpazīt attiecīgās sekvenču grupas: 1_1 un 1_2 faili saturēja informāciju par veselās šūnās (gan membrānā, gan citoplazmā, gan kodolā) atrodamām sekvenču, 2_1 un 2_3 saturēja datus par citoplazmā atrastām sekvenču, 3_1 un 3_2 faili sastāvēja no kodolā internalizētām sekvenču.

2.1.2. Sekundārā datu apstrāde

Kad primārā jeb sākotnēja datu apstrāde ir pabeigta, un ir iegūti dati, kas ir sagatavoti, lai tos varētu apstrādāt RStudio programmā, varam uzsākt sekundāro apstrādi (.tsv fails, kurš satur rindas ar individuālām sekvencēm un kolonnas, kuras raksturo sekvencēšanas nolasījumu skaitu katrai individuālajai sekvencei atsevišķos paraugos). RStudio programmatūra izveidota, lai lietotājam draudzīgā veidā strādātu ar R programmēšanas valodu. R ir īpaša programmēšanas valoda, kas izmantota analīžu, statistikas un citu zinātnisku mērķu sasniegšanai (R Core Team, 2016.).

RStudio – IDE jeb *integrated development environment*, kas nozīmē, ka tā ir programmēšanas aplikācija ko izmanto dažādām programmēšanas programmu pielietošanai kā arī izstrādei. Šīs programmas pamatā ir iepriekš minētā R valoda un R programma, varētu teikt, ka Rstudio ir kā ērts rīks R programmas izmantošanai, padarot lietošanu ērtāku, pārskatāmāku un paplašinot iespējamās darbības izmantošanas iespējas, piemēram, programmu ērtāka pārvalde, ievadīto programmu vēsture, grafiku izveide, īpašu komandu pakotņu izmantošana - vieglākai datu pārvaldei un citas parocīgas funkcijas (RStudio Team, 2015.).

Rstudio tika izmantota īpaša komandu rinda, kura balstījās gan uz pamata R valodas komandām, gan uz īpašām pakotņu komandām: BiocManager 1.30.10. (tālāk BiocManager); edgeR 3.30.0. (tālāk edgeR); readr 1.3.1 (tālāk readr). Katra pakotne ir izveidota, lai vienkāršāk apstrādātu datus, izmantojot vienkāršākas – īpašas komandas, kuras balstās uz pamata R valodas komandām, un zem sevis var pat slēpt komandu ķēdes. Pakotnes arī bieži ir citu pakotņu salikums, tās apvienojot zem viena nosaukuma.

Sekundārā datu apstrāde balstās uz primārās apstrādes jau apstrādātajiem datiem. Primārās datu apstrādes beigās mēs ieguvām sešus “.tsv” formāta failus (1_1 un 1_2 faili – veselas šūnas (membrāna, citoplazma, kodola) sekvences, 2_1 un 2_3 – citoplazmas sekvences, 3_1 un 3_2 faili – kodolā internalizētās sekvences).

```

sample_1_pl_replicates <- Read_delim("C:/SAMPLE_13/INPUT_FAILI/sample_1_pl_replicates.tsv",
                                   "\t", escape_double = FALSE, trim_ws = TRUE)

sample_1_pl_needed <- sample_1_pl_replicates[,c(1,4,7)]

colnames(sample_1_pl_needed) <- c("Sequence", "I_1", "I_2")

```

Šādu komandu virkni ievada programmā, tādējādi izvēloties failus tālākai apstrādei, var jau redzēt, ka informācija no faila ar pēdējām divām komandām jau tiek apstrādāta:

1. (*sample_1_pl_needed <- sample_1_pl_replicates[,c(1,4,7)]*) komanda tiek izmantota, lai atlasītu tikai vajadzīgās kolonnas no faila un šo atlasi nosauc par “*sample_1_pl_needed*”);
2. Komanda “*colnames(sample_1_pl_needed) <- c("Sequence", "I_1", "I_2")*” tiek ievadīta, lai atlasītās kolonnas nosauktu attiecīgi par: “*Sequence*”, “*I_1*”, “*I_2*”. Tas tiek darīts, lai vēlāk būtu vieglāk nolasīt iegūtos datus, par cik to dara cilvēks.

Šādi turpina līdz ir iegūtas visas grupas 1_1-4, 2_1-4 un 3_1-4 (jeb vesela šūna, citoplazma un kodola sekvenču grupas). Tālāk sekvences apstrādā, sakārto un saskaita kopā līdzīgās sekvences un atfiltrē pavisam maz sastopamās sekvences (retāk par piecām sekvencēm uz visiem paraugiem) un sekvences, kuras sastopamas mazāk kā divos no šiem sešiem failiem (arī ja būtu kāda sekvence kas ir milzīgā daudzumā vienā no datu grupām, bet nav nevienā citā – tā uzskatāma par nelietderīgu). Tiek ievadītas šādas komandas, lai iegūtu šādu rezultātu:

```

keep_f_rna <- rowSums(cpm(f_rna_dge_nuclei_vs_non_nuclei$counts) >5) >=2

f_rna_dge_nuclei_vs_non_nuclei_filt <-
f_rna_dge_nuclei_vs_non_nuclei[keep_f_rna, , keep.lib.sizes = FALSE]

```

1. Pirmā komanda “*keep_f_rna <- rowSums(cpm(f_rna_dge_nuclei_vs_non_nuclei\$counts) >5) >=2*” tiek izmantota, lai atlasītu datus (augstāk jau tika pieminēts). Tiek atlasītas sekvences, kuras ir mazāk par 5 > CPM, un tiek atlasītas sekvences, kas sastopamas mazāk kā 2 failos.
2. “*f_rna_dge_nuclei_vs_non_nuclei_filt <- f_rna_dge_nuclei_vs_non_nuclei[keep_f_rna, , keep.lib.sizes = FALSE]*” komanda arī tiek izmantota, lai izfiltrētu liekos datus.

Veicot šo apstrādi no 2516960 sekvencēm tika iegūtas 17824 sekvences – grupā kurā ir kodolā internalizētās un citoplazmā atrodamas sekvences, un 19135 sekvences – grupā kurā ir

kodolā internalizētas sekvences un sekvences kas atrodamas veselās šūnās, kuras atbilst pamata kritērijiem.

Iegūtājām sekvencēm tālāk tiek aprēķināts, datu apstrādei ļoti svarīgais, normalizācijas skaitlis – tas stipri samazina tālāku nepatiesi pozitīvo datu parādīšanos turpmāk sekojošajā datu apstrādes daļā. Šis skaitlis nodrošina, ka rezultātā iegūtie dati ir ticamāki, un rezultātā iegūtās sekvences patiešām arī izpildīs vēlamo funkciju.

Tālāk informācija tiek apstrādāta ar sekojošajām komandām:

```
f_rna_de <- estimateGLMCommonDisp(f_rna_dge_nuclei_vs_non_nuclei_filt_norm,  
f_rna_design)  
f_rna_de <- estimateGLMTrendedDisp(f_rna_de, f_rna_design)  
f_rna_de <- estimateGLMTagwiseDisp(f_rna_de, f_rna_design)
```

Ar šīm komandām tiek veikta datu dispersija, tādējādi īpaši tiek atlasīti dati, kas ļoti izceļas no kopējā informācijas fona.

Tad tiek izmantotas īpašas komandas, kuras izmanto lai apstrādāto grupu datus varētu savstarpēji salīdzināt un veikt pēdējo datu apstrādi – ko veic pētnieks *Excell* failā

```
f_rna_fit <- glmQLFit(f_rna_de, f_rna_design)  
nuclei_vs_non_nuclei <- glmQLFTest(f_rna_fit)
```

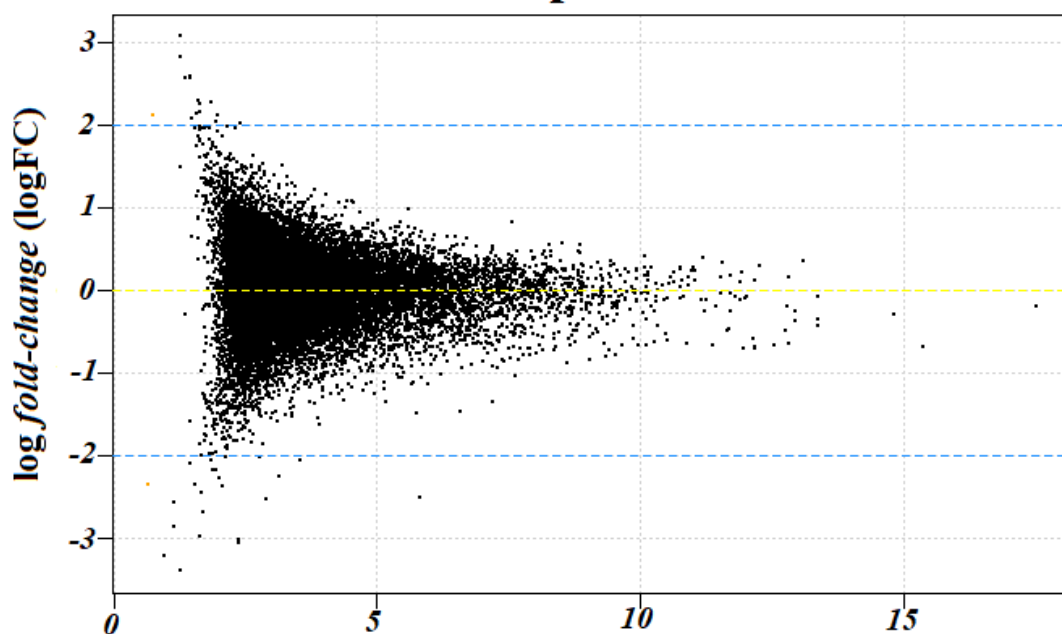
Tiek iegūts *Excell* fails, kuru var izmantot, lai, izmantojot īpašus filtrus, iegūtu mūsu meklētās sekvences un tiktu pie rezultātiem.

3. Rezultāti un diskusija

3.1. Sekvences, kas sastopamas kodolā vai citoplazmā

Izmantojot Rstudio programmu un ievadot komandu rindu tiek iegūti rezultāti – programmas *Excell* fails, kurā iegūstama visa apstrādātā informācija par izfiltrētajām sekvencēm, un grafiki, kurā attēlots logaritms no sekvenču skaita (sastopamība skaita veidā), angliiski *counts per million* (CPM), un sekvenču atrašanās jeb sastopamība noteiktā grupā (kodols vai citoplazma un kodols vai veselas šūnas) jeb sastopamība noteiktajā grupā, ko angliiski apzīmē ar *fold change* (FC)(precīzāk logaritms no šī lieluma – $\log FC$).

Sekvences sastopamība - kodolā vai citoplazmā



Vidējais log counts-per-million (logCPM)

3.1.1. att. **Sekvences sastopamība – kodolā vai citoplazmā.** Redzamajā attēlā *log fold-change* ($\log FC$) ass attēlojami gan pozitīvi, gan negatīvi rezultāti – punkti virs pozitīvās divi (2) vērtības apzīmē sekvences, kas sastopamas pārsvarā kodolā, jo augstāks skaitlis, jo stiprāk sekvence ir internalizējusies kodolā, punkti zem negatīvās divi (-2) vērtības atrodami pārsvarā citoplazmā – un ne kodolā. Zonā, starp pozitīvo divi (2) un negatīvo divi (-2), atrodami punkti (apzīmē sekvences), kas atrodamas gan kodolā, gan citoplazmā (jo tālāk no nulles (0) vērtības, gan pozitīvā, gan negatīvā virzienā, jo lielāka nosliece uz kādu no grupām). Vidējais log counts-per-million

(logCPM) norāda uz sekvenču logaritmisku sastopamību no miljons sekvencēm – šis ir reprezentatīvs lielums, lai vieglāk varētu pārskatīt, cik sastopama ir kāda sekvence.

Kā redzams, diagrammā (3.1.1. att.) attēlotie dati norāda par lielu daudzumu sekvenču, kas atrodas gan kodolā, gan citoplazmā, arī ieskaitot trīs sekvences, kuras potenciāli varētu būt vērtīgas un tiks izpētītas tālāk – dēļ to lielā sastopamības CPM skaita un tās sastopamas abās grupās, kas nozīmē, ka iespējams sekvences vēl nebija paspējušas internalizēties kodolā, tas varētu būt dēļ tā, ka paraugi tika inkubēti ar mērķi tikai uz stundu, un sekvences burtiski nespēja paspēt internalizēties un vēl atradās citoplazmā, nevis jau kodolā.

3.1.1. tabula.

Potenciāli izmantojamās sekvences (log CPM)

Nr.	Sekvence	log FC	log CPM	F	P vērtība	FDR
219299.	GGTAATACGACTCACTATAGG GAGATACCAGCTTATTCAATT	-0,67466	15,40622	15,83758	0,000239	0,034343
161174.	GTGGGTTGGTAAATTTTTTA TGA TATT CAGGGTGCCGGC	-0,42339	13,4034	258,7689	1,1*10 ⁻²⁰	1,00*10 ⁻¹⁶
103978.	GTTTTAGCTAATAGAGTTAA ATTTTTTATGTGTCGTGTGG	-0,35367	13,40325	103,1701	2,07*10 ⁻¹³	7,38*10 ⁻¹⁰

Tabulā redzamas trīs iepriekš minētās sekvences, kurām ir liels potenciāls. Svarīgi izcelt logFC, logCPM un FDR, jo pēc šiem lielumiem varēja izfiltrēt jeb atrast attiecīgās sekvences, starp 17824 sekvencēm kas atrodamas šajā grupā. Lai gan logFC ir vairāk vērsts uz negatīvu skaitli, kā jau iepriekš minēts, šīs sekvences logCPM ir ļoti liels, salīdzinot ar pārējām sekvencēm, un logFC ir gandrīz nulle un arī veicot ilgāku inkubāciju ar mērķi varētu novirzīties uz pozitīvo logFC ass rezultātu. FDR jeb *false discovery rate* izmanto lai norādītu, cik ticams ir rezultāts – ticams rezultāts, ja FDR < 0,05 – līdzīgi kā visiem zināmā p vērtība (tabulā “P vērtība”), kas arī ir pievienota tabulā, lai varētu redzēt, ka p vērtības ir pat neticami labas, bet strādājot ar datiem, kuru apstrādei tiek izmantots liels salīdzinājumu skaits, p vērtības nozīme pazūd.

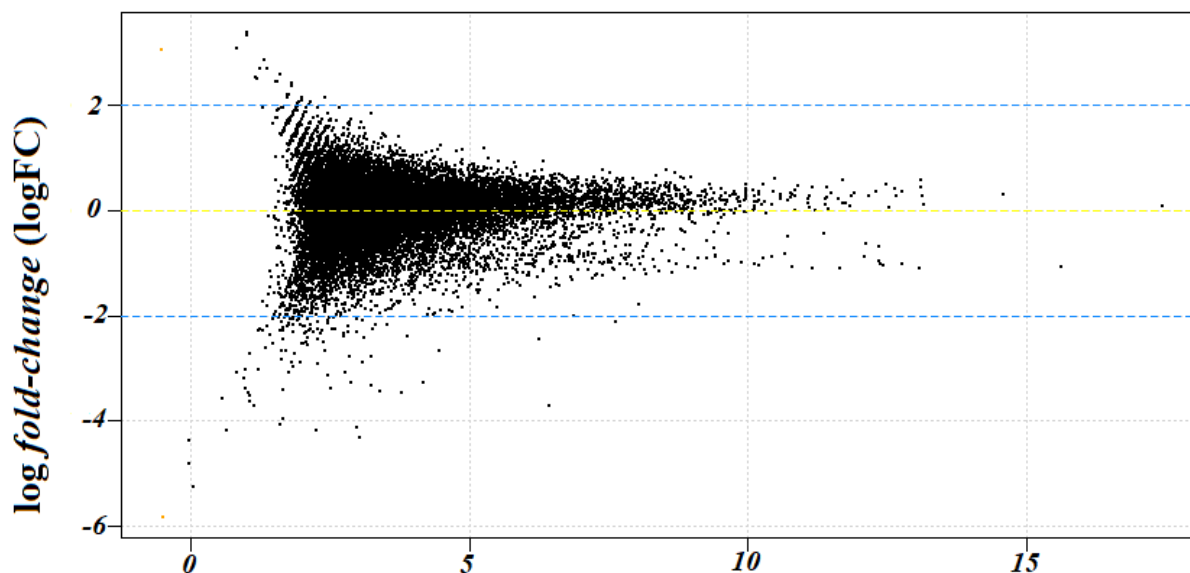
Potenciāli izmantojamās sekvenses (log FC)

Nr.	Sekvence	log FC	log CPM	F	P vērtība	FDR
44958.	TTTTGACCGATAACATACTCT CTGTCTTGTCACGGGCCGCG	2,789455	2,64123	14,29789	0,000442	0,047788

Šī tabula sastāv tikai no vienas sekvenses, kuras logFC ir ļoti augsts un pozitīvs rādītājs, tātad sekvence atrodama kodolā. Lai gan grafikā redzama arī sekvence, kas ir ar pat augstāku rādītāju (virs 3), bet acīmredzot neatbilda kritērijam ($FDR < 0,05$). Jāizceļ ka, lai gan logCPM ir salīdzinoši, ar augstāk pieminētajām (3.1. tabulas) sekvencēm, zems, tāpat ir vērts veikt tālākus pētījumus.

3.2. Sekvences, kas sastopamas kodolā vai šūnā

Sekvences sastopamība - kodolā vai šūnā



Vidējais log counts-per-million (logCPM)

3.2.1. att. Sekvences sastopamība – kodolā vai šūnā. Redzamajā attēlā log fold-change (logFC) ass attēlojami gan pozitīvi, gan negatīvi rezultāti – punkti virs pozitīvās divi (2) vērtības apzīmē sekvenses, kas sastopamas pārsvarā kodolā, jo augstāks skaitlis, jo stiprāk sekvence ir internalizējusies kodolā, punkti zem negatīvās divi (-2) vērtības atrodami pārsvarā veselās šūnās (sastāv no šūnas membrānas, citoplazmas un kodola) – un ne kodolā. Zonā, starp pozitīvo divi (2) un negatīvo divi (-2), atrodami punkti (apzīmē sekvenses), kas atrodamas gan kodolā, gan citoplazmā (jo tālāk no nulles (0) vērtības, gan pozitīvā, gan negatīvā virzienā, jo lielāka nosliece uz kādu no grupām). Vidējais log counts-per-million (logCPM) norāda uz sekvenču logaritmisku

sastopamību no miljons sekvencēm – šis ir reprezentatīvs lielums, lai vieglāk varētu pārskatīt, cik sastopama ir kāda sekvenca.

Kā redzams, diagrammā (3.2.1. att.) attēlotie dati norāda par lielu daudzumu sekvenču, kas atrodas gan kodolā, gan veselās šūnās, līdzīgi kā pirmajā diagrammā (3.1.1. att.). Ar šiem datiem veiktās atlasēs rezultātā tika iegūtas šādas sekvences, kuras izcēlās ar log CPM (tāpat kā tika izceltas sekvences – pēc tādiem pašiem kritērijiem):

3.2.1. tabula.

Potenciāli izmantojamās sekvences (log CPM)

Nr.	Sekvenca	log FC	log CPM	F	P vērtība	FDR
219299.	GGTAATACGACTCACTATAGG GAGATACCAGCTTATTCAATT	-1,08262	15,70826	41,60563	1,11*10 ⁻⁸	2,41*10 ⁻⁶
232752.	GTTTTAGCTAATAGAGTTAA AATTTTTATGTGTCGTGTGG	0,290207	14,58616	16,73107	0,000111	0,005134
3860.	GGTAATACGACTCACTATAGG GAGATACCAGCTTATTCAATTT	-1,0374	13,17316	190,0407	6,42*10 ⁻²²	4,09*10 ⁻¹⁸
103978.	GTTTTAGCTAATAGAGTTAA AATTTTTATGTGTCGTGTGG	0,241994	13,12606	68,43842	4,63*10 ⁻¹²	3,45*10 ⁻⁹
182617.	CGACTGTTGTTGAAATTGAT TGTGGTAAATTTTTTTGGGC	0,516809	13,12336	225,0518	6,79*10 ⁻²⁴	1,30*10 ⁻¹⁹
2412.	GTTGTCGAGAAAACGGGGA CAAATTTTTTCAGTTGCTGC	0,374012	13,12336	129,0728	9,65*10 ⁻¹⁸	2,05*10 ⁻¹⁴

Tabulā redzamas sešas sekvences, kuras potenciāli varētu veikt tālākus padziļinātus pētījumus – sekvences ar lielu potenciālu. Līdzīgi kā iepriekš šajā tabulā sakopotas sekvences kurām ir augsts logCPM un pietiekami zems FDR (zem 0,05), bet diemžēl arī šīs sekvences ir ar zemu logFC (iemesls visticamāk ir tāds pats, kā minēts pie iepriekšējās sekvenču tabulas (3.1.1. tabulas)).

Salīdzinoši ar iepriekšējās grupas log CPM atkarīgajām sekvencēm (3.1.1. tabula) šajā grupā, kurā tika sakopotas kodolā internalizējošas un veselā šūnā atrodamas sekvences, ir vairāk sekvenču, un šīm sekvencēm ir diezgan labs log FC (vairāk kā pusei pozitīvs), bet diemžēl neviena sekvenca nesasniedz log FC ≥ 2 , kas ir sākuma sliekšnis, kas norādītu uz ticamu specifiskumu uz šūnas kodolu, kas ir šo sekvenču bibliotēkas datu apstrādes mērķis.

Potenciāli izmantojamās sekvences (log FC)

Nr.	Sekvence	log FC	log CPM	F	P vērtība	FDR
1075749.	TTTTATTAGACTGATTTTCGT AACCTCCCTCGCTGTTCCCC	5,266443	2,424876	13,92374	0,000386	0,014434
687607.	GCGGTTTGGAGTTGTCGACT GCAATTATTATTTTACTGCG	3,663945	2,683809	11,17466	0,001316	0,035619
227380.	TGGTTTCAAAAATTTTTCTAT AGCCTGCTTACTACTGTGGGG	2,923907	2,738349	11,7444	0,00101	0,0297
199252.	GCGTGTGCTTGCCTAATACT AATATGGTATTGTTTACGGG	2,834595	2,658224	10,30351	0,001982	0,047161

Tabula, atšķirībā no pirmās grupas sekvencēm (kodols vai citoplazma), kur log FC sastāvēja tikai no vienas sekvenču, sastāv no četrām sekvencēm. Šo sekvenču logFC ir augstāks par pirmās grupas (3.1.2. tabulā *apkopotajām*) sekvencēm, kas nozīmē, ka šī grupa sastāv no daudz vairāk sekvencēm kas spēj internalizēties kodolā, un kurām FDR ir pietiekami zems (zem 0,05). Šo sekvenču log CPM diemžēl ir (salīdzinoši ar 3.3. tabulas sekvencēm) zems.

Šajā tabulā (3.2.2. tabula) ir atzīmētas sekvenču, kuras nav atrodamas grafikā (3.1.2. att.), kur log FC beidzas pie pozitīva divi (2), bet šeit log FC ir līdz pat 5,266443. Tas izskaidrojams ar grafiku veidošanai lietotās programmas automātisko filtrāciju, kas pati papildus izfiltrē datus. Vizuāli attēlot datus par laimi nav tik svarīgi, kā tos spēt izfiltrēt un atrast – lai tos identificētu (ko nav iespējams izdarīt skatoties uz grafiku).

4. Secinājumi

Darba beigās var secināt, ka balstoties uz iepriekš rakstītajiem rezultātiem, bakalaura darba uzdevumi ir veiksmīgi izpildīti:

1. veiksmīgi tika izveidota un izpildīta komandu ķēde, kas balstās uz vairākām dažādām programmām un to pakotēm, kuru var izmantot, lai iegūtu atkārtojamus rezultātus;
2. darba galā ir iegūtas potenciāli vērtīgas sekvences, kuras būs nepieciešams izpētīt dziļāk, citā pētījumā;
3. padziļināti iepazīt laboratoriju kas saistīta ar mūsdienīgām metodēm – precīzāk, tieši datu apstrādes posmu, kas mūsdienās paliek aizvien efektīvāks – prasot mazāk cilvēka darba un kļūst vispārēji mazāk laikietilpīgs, bet ir svarīga daļa pētniecības darbu veikšanā un rezultātu iegūšanā.

Aplūkojot visus šos trīs izpildītos uzdevumus, var secināt, ka ir izstrādāts datu analīzes algoritms šūnas kodolā internalizējošu aptamēru atlasei, kas arī bija šī darba mērķis.

5. Pateicības

Īpaši vēlos pateikties Kārlim Pleiko, par mana bakalaura darba vadīšanu, par manu apmācīšanu datu apstrādē, par ko man pat iepriekš nebija nekāda veida pieredzes vai īpašu priekšzināšanu, un arī par vispārēju atsaucību.

Ļoti arī gribu pateikties Valentīnai Strazdītei, Aivaram Strazdītam un Andrim Strazdītam par atbalstu manā mācību procesā, kā arī par to pašu arī maniem tuviniekiem un draugiem.

6. Izmantotā literatūra un avoti

1. **Adler, A., Forster, N., Homann, M., Goring, U. H.**, Post-SELEX Chemical Optimization of a Trypanosome-Specific RNA Aptamer, (2008.), *Combinatorial Chemistry & High Throughput Screening*, 11(1), pp. 16–23. doi: 10.2174/138620708783398331.
2. **Alam, K. K., Chang, J. L. and Burke, D. H.**, FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections, (2015.), *Molecular Therapy - Nucleic Acids*. IOP Publishing, 4(3), p. e230. doi: 10.1038/mtna.2015.4.
3. **Andrews, S.**, FastQC: A Quality Control Tool for High Throughput Sequence Data, (2010.), skatīts: 07.05.2020., Pieejams: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
4. **Bédard, A. S. V., Hien, E. D. M., Lafontaine, D. A.**, Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins, (2020.), *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms.*, 1863(3), p. 194501. doi: 10.1016/j.bbagr.2020.194501.
5. **Boussebayle, A., Groher, F., Suess, B.**, RNA-based Capture-SELEX for the selection of small molecule-binding aptamers, (2019.), *Methods.*, 161(January), pp. 10–15. doi: 10.1016/j.ymeth.2019.04.004.
6. **Carpente, L., Cerdeira-Pena, A., Lorenzo-Freire, S., Places, Á. S.**, Optimization in Sanger sequencing, (2019.), *Computers and Operations Research*, 109, pp. 250–262. doi: 10.1016/j.cor.2019.05.011.
7. **Cawte, A. D., Unrau, P. J., Rueda, D. S.**, Live cell imaging of single RNA molecules with fluorogenic Mango II arrays, (2020.), *Nature Communications.*, 11(1). doi: 10.1038/s41467-020-14932-7.
8. **Chandola, C., Neerathilingam, M.**, Aptamers for Targeted Delivery: Current Challenges and Future Opportunities, (2019.), *Role of Novel Drug Delivery Vehicles in Nanobiomedicine*. doi: 10.5772/intechopen.84217.
9. **Dollins, Claudia M., Nair, Smita, and Sullenger, B. A.**, Aptamers in immunotherapy, (2008.), *Hum Gene Ther.*, 19(5), pp. 443–450. doi: 10.1038/jid.2014.371.
10. **Doria-Rose, N. A., Vogt, V. M.**, In Vivo Selection of Rous Sarcoma Virus Mutants with Randomized Sequences in the Packaging Signal, (1998.), *Journal of Virology*, 72(10), pp. 8073–8082. doi: 10.1128/jvi.72.10.8073-8082.1998.
11. **Dou, X. Q., Wang, H., Zhang, J., Wang, F., Xu, G. L., Xu, C. C., Xu, H. H., Xiang, S. S., Fu, J. and Song, H. F.**, Aptamer–drug conjugate: Targeted delivery of doxorubicin in a HER3 aptamer-functionalized liposomal delivery system reduces cardiotoxicity, (2018.),

- International Journal of Nanomedicine*, 13, pp. 763–776. doi: 10.2147/IJN.S149887.
12. **Ellington, A. D., Szostak, J. W.**, In vitro selection of RNA molecules that bind specific ligands, (1990.), *Nature*, 346(6287), pp. 818–822. doi: 10.1038/346818a0.
 13. **Guo, W. M., Kong, K. W., Brown, C. J., Quah, S. T., Yeo, H. L., Hoon, S., Seow, Y.**, Identification and characterization of an eIF4e DNA aptamer that inhibits proliferation with high throughput sequencing, (2014.), *Molecular Therapy - Nucleic Acids*. American Society of Gene & Cell Therapy, 3(September), p. e217. doi: 10.1038/mtna.2014.70.
 14. **Illumina Inc.**, *Understanding the NGS workflow.*, (2020.), Pieejams: <https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html>.
 15. **Kourlas, H., Schiller, D. S.**, Pegaptanib Sodium for the Treatment of Neovascular Age-Related Macular Degeneration: A Review, (2006.), *Clinical Therapeutics*, 28(1), pp. 36-44. doi: 10.1016/j.clinthera.2006.01.009.
 16. **Li, X., Yang, Y., Zhao, H., Zhu, T., Yang, Z., Xu, H., Fu, Y., Lin, F., Pan, X., Li, L., Cui, C., Hong, M., Yang, L., Wang, K. K. and Tan, W.**, Enhanced in Vivo Blood-Brain Barrier Penetration by Circular Tau-Transferrin Receptor Bifunctional Aptamer for Tauopathy Therapy, (2020.), *Journal of the American Chemical Society*, 142(8), pp. 3862–3872. doi: 10.1021/jacs.9b11490.
 17. **Gragoudas, S. E., Adamis, A. P., Cunningham, E. T. J., Feinsod, M., Guyer, D. R.**, Pegaptanib for Neovascular Age-Related Macular Degeneration, (2004.), *The new england journal of medicine original*, 351(27), pp. 2805-2816.
 18. **Marcel, M.**, Cutadapt removes adapter sequences from high-throughput sequencing reads., (2011.), *EMBnet.journal*. doi: <https://doi.org/10.14806/ej.17.1.200>.
 19. **Mehta, J., Rouah-Martin, E., Dorst, B. V., Maes, B., Herrebout, W., Scippo, M. L., Dardenne, F., Blust, R., Robbens, J.**, Selection and characterization of PCB-binding DNA aptamers., (2012.), *Anal Chem*, (84), pp. 1669–1676.
 20. **Min, K., Jo, H., Song, K., Cho, M., Chun, Y.-S., Jon, S., Kim, W. J., Ban, C.**, Dual-aptamer-based delivery vehicle of doxorubicin to both PSMA (+) and PSMA (-) prostate cancers., (2011.), *Biomaterials*, 32(8), pp. 2124–2132. doi: 10.1016/j.biomaterials.2010.11.035.
 21. **Mina, K., Hyun-Ju, U., Sunbaek, B., Sang-Hee, L., Suk-Jung, O., Ji-Hye, H.**, Arsenic Removal from Vietnamese Groundwater Using the Arsenic-Binding DNA Aptamer, (2009.),

- Environmental Science and Technology*, 43(24), pp. 9335–9340.
22. **Nomura, Y., Sugiyama, S., Sakamoto, T., Miyakawa, S., Adachi, H., Takano, K., Murakami, S., Inoue, T., Mori, Y., Nakamura, Y., Matsumura, H.**, Conformational plasticity of RNA for target recognition as revealed by the 2.15 Å crystal structure of a human IgG-aptamer complex, (2010.), *Nucleic Acids Research*, 38(21), pp. 7822–7829. doi: 10.1093/nar/gkq615.
 23. **Ohuchi, S.**, Cell-Selex technology, (2012.), *BioResearch Open Access*, 1(6), pp. 265–272. doi: 10.1089/biores.2012.0253.
 24. **R Core Team**, R: A language and environment for statistical computing. R Foundation for Statistical Computing, (2016.), Vienna, Austria: R Core Team. Pieejams: <https://www.r-project.org/>.
 25. **Raouane, M., Desmaële, D., Urbinati, G., Massaad-Massade, L., Couvreur, P.**, Lipid conjugated oligonucleotides: A useful strategy for delivery, (2012.), *Bioconjugate Chemistry*, 23(6), pp. 1091–1104. doi: 10.1021/bc200422w.
 26. **RStudio Team**, RStudio: Integrated Development for R. RStudio, Inc., (2015.), Boston. Available at: <https://rstudio.com/>.
 27. **Nimjee, S. M., White, R. R., Becker, R. C., Sullenger, B. A.**, Aptamers as Therapeutics, (2017.), *Physiology & behavior*, 176, pp. 61–79. doi: 10.1146/annurev-pharmtox-010716-104558.
 28. **Sigma-Aldrich**, *Polymerase Chain Reaction (PCR) Basics.*, (2020.) Pieejams: <https://www.sigmaaldrich.com/life-science/molecular-biology/pcr/routine-amplification.html>.
 29. **Song, K. M., Lee, S., Ban, C.**, Aptamers and their biological applications, (2012.), *Sensors*, 12(1), pp. 612–631. doi: 10.3390/s120100612.
 30. **Stark, R., Grzelak, M., Hadfield, J.**, RNA sequencing: the teenage years, (2019.), *Nature Reviews Genetics.*, 20(11), pp. 631–656. doi: 10.1038/s41576-019-0150-2.
 31. **Stuart, C. H., Riley, K. R., Boyacioglu, O., Herpai, D. M., Debinski, W., Qasem, S., Marini, F. C., Colyer, C. L., Gmeiner, W. H.**, Selection of a Novel Aptamer Against Vitronectin Using Capillary Electrophoresis and Next Generation Sequencing, (2016), *Molecular Therapy - Nucleic Acids.*, 5(July), p. e386. doi: 10.1038/mtna.2016.91.
 32. **Thiel, W. H., Thiel, K. W., Flenker, K. S., Bair, T., Dupuy, A. J., McNamara II, O. J., Miller, F. J., Giangrande, P. H.**, Cell-Internalization SELEX: Method for Identifying Cell-

- Internalizing RNA Aptamers for Delivering siRNAs to Target Cells, (2015.), *Methods in molecular biology* (Clifton, N. J.), 1218, pp. 187-199. doi: 10.1007/978-1-4939-1538-5_11.
33. **Wang, L., Ma, Wenwei, Chen, W., Liu, L., Ma, Wei, Zhu, Y., Xu, L., Kuang, H., Xu, C.,** An aptamer-based chromatographic strip assay for sensitive toxin semi-quantitative detection, (2011.), *Biosensors and Bioelectronics*. Elsevier B.V., 26(6), pp. 3059–3062. doi: 10.1016/j.bios.2010.11.040.
 34. **You, M., Litke, J. L., Jaffrey, S. R.,** Imaging metabolite dynamics in living cells using a Spinach-based riboswitch, (2015.), *Proceedings of the National Academy of Sciences of the United States of America*, 112(21), pp. E2756–E2765. doi: 10.1073/pnas.1504354112.
 35. **Zhu, C., Zhang, G., Huang, Y., Yang, S., Ren, S., Gao, Z., Chen, A.,** Dual-competitive lateral flow aptasensor for detection of aflatoxin B1 in food and feedstuffs, (2018.), *Journal of Hazardous Materials.*, 344, pp. 249–257. doi: 10.1016/j.jhazmat.2017.10.026.
 36. **Zhu, Q., Liu, G., Kai, M.,** DNA aptamers in the diagnosis and treatment of human diseases, (2015.), *Molecules*, 20(12), pp. 20979–90997. doi: 10.3390/molecules201219739.
 37. **Zong, C., Liu, J.,** The arsenic-binding aptamer cannot bind arsenic: critical evaluation of aptamer selection and binding, (2019.), *Analytical Chemistry.*, 91(16), pp. 10887–10893. doi: 10.1021/acs.analchem.9b02789.

DOKUMENTĀRĀ LAPA

Bakalaura darbs Datu analīzes algoritma izstrāde šūnas kodolā internalizējošu aptamēru atlasei .

izstrādāts LU Medicīnas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Rūdolfs Strazdīts _____

(vārds, uzvārds)

(paraksts)

(datums)

Rekomendēju darbu aizstāvēšanai

Vadītāja: Pētnieks (LU MF), Kārlis Pleiko _____

(amats, vārds, uzvārds, grāds)

(paraksts)

(datums)

Recenzents: _____

(amats, vārds, uzvārds, grāds)

(paraksts)

(datums)

Darbs iesniegts LU Medicīnas fakultātē _____

(datums)

Vecākā lietvede Juta Bārtule _____

(paraksts)

Bakalaura darbs aizstāvēts bakalaura studiju programmas „Farmācija” Bakalaura gala pārbaudījuma komisijas sēdē _____ 2020., prot. Nr. _____.

Komisijas sekretāre: docente Kristīne Saleniece, dr.pharm. _____

(paraksts)