TARTU ÜLIKOOL
BIOLOOGIA-GEOGRAAFIATEADUSKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOINFORMAATIKA ÕPPETOOL

Priit Tomson

DNA mikrokiipidel kasutatavate oligote kvaliteeti mõjutavad parameetrid ja meetodid nende disainiks

Bakalaureusetöö

Juhendaja Ph.D. student Reidar Andreson

TARTU 2005

Sisukord

Lühendid ja mõisted

APEX	oligonukleotiidmaatriksil põhinev praimerekstensioon	*Arrayed Primer EXtension*
CDS	kodeeriv järjestus	*coding sequence*
cross-hybridization	risthübridisatsioon - oligo seondumine mittesoovitud kohta	
EST	osaline cDNA järjestus	*expressed sequence tag*
gap	auk – järjestuse joondamisel ühel ahelal puuduolev aluspaar	
HGVbase	Inimese genoomi variatsioonide andmebaas	*Human Genome Variation database*
in silico	arvutis	
microarray	mikrokiip, DNA kiip, geenikiip, biokiip	
oligo	oligonukleotiid, praimer – keemiliselt sünteesitud lühike DNA järjestus	
ORF	avatud lugemisraam – DNA järjestus, mis algab initsiatsioonikoodoniga ja lõppeb stoppkoodoniga	*open reading frame*
probe	proov, märgistatud molekulid, kasutatakse hübridisatsiooniks kiibil olevate sihtmärk molekulidega.	
spot	mikrokiibi element, mis sisaldab substraadile seotud sihtmärkmolekule	
UNG	uratsiil-N-glükosülaas	
VLSIPS	kõrgtihedalt immobiliseeritud polümeeri süntees	*very large scale immobilized polymer synthesis*

Sissejuhatus

Viimasel ajal on mitmetes uurimisvaldkondades laialdaselt kasutama hakatud DNA mikrokiipidel põhinevaid tehnoloogiaid. Mikrokiipide tehnoloogiat rakendatakse efektiivselt teaduslikes uurimustöödes, samuti ka diagnostikas, farmaatsiatööstuses, proteoomikas ning keskkonna puhtuse ja toidu kvaliteedi kontrollimiseks.

Mikrokiibi tehnoloogia on väga tihedas seoses PCR'iga. Enamasti on uuritava materjali kogus väga väike ja seda on vaja amplifitseerida, et hiljem oleks võimalik tulemusi visualiseerida ja kvantifitseerida. Antud juhul on väga oluline, et PCR'is kasutatavad oligod oleksid ülimalt spetsiifilised. Vastasel korral võib amplifikatsiooni käigus tekkida väär(ad) produkt(id), mis vähendavad PCR'i efektiivsust ja võivad põhjustada hilisemates reaktsioonides valesignaalide teket.

Samaväärselt on tähtis ka see, et mikrokiipidel kasutatavad oligod oleksid spetsiifilised. Vastasel juhul võib see viia andmete valele tõlgendamisele. Selliste sündmuste vältimiseks tuleb teha palju erinevaid katsetusi, analüüsimisi ja kontrollimisi, eriti just oligote koha pealt, mis on mikrokiibi rakenduste efektiivsuse alustalaks.

Käesoleva töö eesmärgiks on uurida kirjanduse põhjal mikrokiipide olemust ja nende rakendusi, samuti oligo disaini olulisust mikrokiipidega efektiivseks ja usaldusväärseks manipuleerimiseks ning kirjeldada levinuimaid meetodeid oligote disainimiseks ja nende kvaliteedi kontrollimiseks. Sellele uurimusele toetudes oleks edaspidiseks eesmärgiks arendada programm, mille abil saaks täpsemalt analüüsida APEX oligote parameetreid, mis mõjutavad APEX reaktsiooni kvaliteeti.

Kirjanduse ülevaade

1 DNA mikrokiibid

Mikrokiibid (*microarrays*) on mikroskoopiliste elementide korrastatud reastus tahkel substraadil, tavaliselt klaasil, mis võimaldab nende spetsiifilist seondumist. Mikroskoopilisteks elementideks võivad olla nukleiinhapped, valgud või mõned muud väikesed molekulid. Mikrokiibi elemendid ehk spotid sisaldavad substraadile seotud sihtmärkmolekule (*target*), millel lastakse hübridiseeruda lahuses olevate fluorestsentselt märgistatud komplementaarsete proovimolekulidega (*probe*), tekitades sellega helenduse. Proovideks on rakkudest ekstraheeritud DNA, mRNA või nende derivaadid. Nukleiinhapete mikrokiipidel kasutatakse kiibi elementidena tavaliselt lühikesi oligonukleotiide (15-25 bp), pikki oligonukleotiide (50-120 bp) ja PCR'i abil amplifitseeritud cDNA'd (100-3000 bp) (Stears jt., 2003).

1.1 Mikrokiipide ajalugu

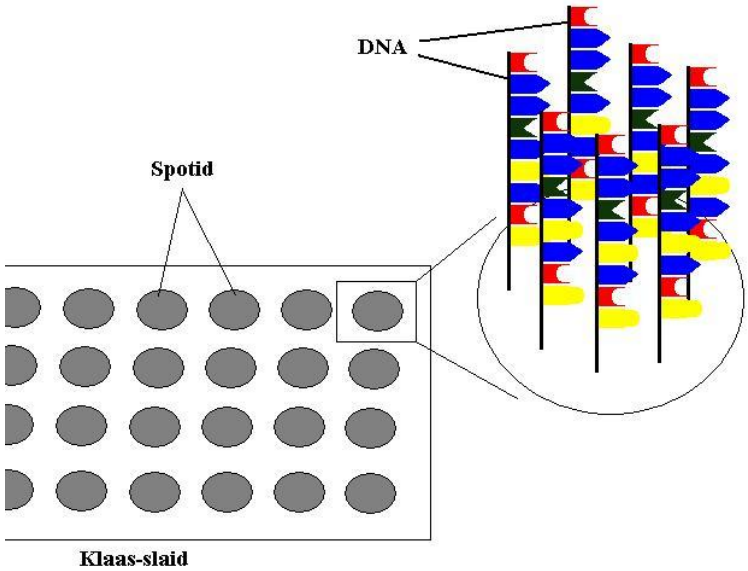
Mikrokiibid arendati välja Stanfordi Ülikoolis Mark Schena ja tema kolleegide poolt 90ndate aastate alguses. Mikrokiibi mõiste esitleti esmakordselt 1994. aasta suvel Hollandis. Esimesed mikrokiibid valmistati koostöös Affymetrix'iga, kasutades kõrgtihedalt immobiliseeritud polümeeri sünteesi (VLSIPS) tehnoloogiat (Schena jt., 1995).

Esimesed komplementaarse DNA (cDNA) mikrokiibid valmistati koostöös Dari Shaloni ja Patrick Browniga. Valmistamisel kasutati nõelprinteril põhinevat robotit, mis pani pisikesed tilgakesed, mis sisaldasid amplifitseeritud *Arabidopsis*'e cDNA'd, eelnevalt töödeldud mikroskoobi alusklaasile. Mikrokiipi hübridiseeriti neljast erineva päritoluga *Arabidopsis*'e mRNA'st valmistatud fluorestsentse cDNA seguga geeni ekspressiooni analüüsimiseks metsiktüüpi ja HAT4 cDNA'd üleekspresseerivas transgeenses liinis (Schena jt., 1995). Prinditud mikrokiipe kasutati esimest korda ka

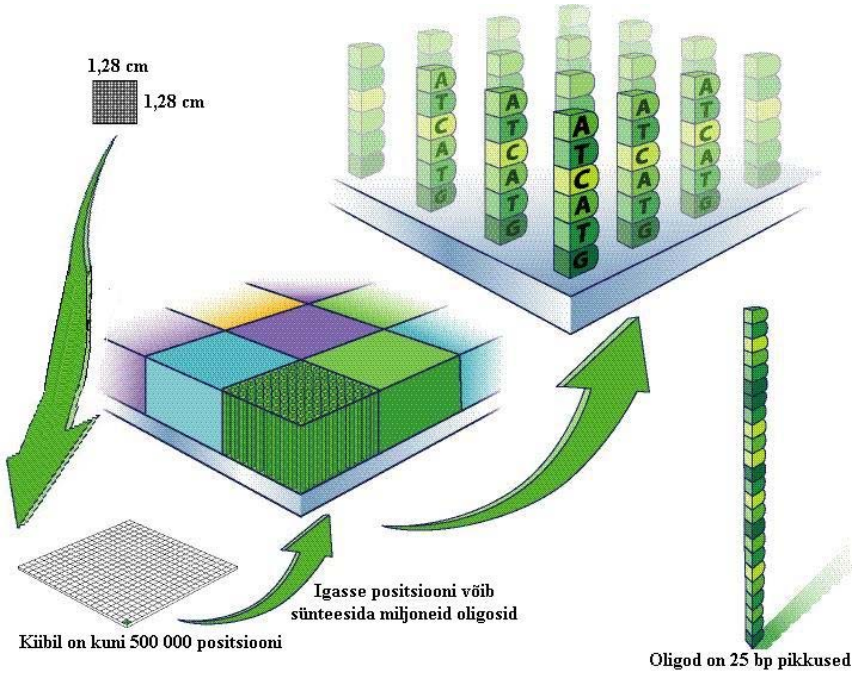
inimese analüüsi katsetes, mille käigus jälgiti paralleelselt 1000 geeni ekspressiooni (Schena jt., 1996).

1.2 DNA mikrokiipide iseloomustus

DNA mikrokiibid koosnevad tahkest kandjast ja sellele korrapärases reastuses kinnitatud sadadest tuhandetest või isegi miljonitest DNA molekulidest (Joonis 1 ja 2). DNA molekulideks võivad olla PCR'i produktid või oligonukleotiidid, mis immobiliseeritakse ruumiliselt eraldatuna enamasti tasapinnalisele pinnale. Immobiliseeritud molekulid peavad jääma aktiivseks ja stabiilseks ning efektiivse hübridisatsiooni toimumiseks olema optimaalse tihedusega. Väga tähtis on ka autofluorestsentsi ja mittespetsiifilise seondumise minimaalsus (Venkatasubbarao, 2004). Et mikrokiipe kiiresti ja lihtsalt toota ja hiljem ka analüüsida, peab mikrokiip olema korrapärane, mikroskoopiline, tasapinnaline ning spetsiifiline (Schena, 2003).



Joonis 1. Spotitud DNA mikrokiip. Klaasist slaidile prinditud spotid on ühesuurused ja korrapärase paigutusega. Iga spott sisaldab kindlaid oligonukleotiide või cDNA fragmente (sihtmärkmolekulid).



Joonis 2. Fotolitograafiliselt valmistaud Affymetrix'i GeneChip®. 1,28 cm^2 pinnale on võimalik korrapäraselt sünteesida kuni 500 000 erinevat oligonukleotiidide positsiooni, millest igaüks sisaldab miljoneid spetsiifilisi 25-meerseid oligonukleotiide. http://www.affymetrix.com/corporate/outreach/lesson_plan/downloads/slides/details_of_a_single_feature.pdf

1.2.1 Korrapärane maatriks mikrokiipidel

Mikrokiibi spotid/positsioonid on kujundatud ridadesse ja veergudesse. Read ja veerud peavad olema rivistatud sirgjooneliselt ning olema üksteisega risti. Selline paigutus võimaldab printeritel, skänneritel ja teistel seadmetel kiipi lugeda ja kirjutada suure kiirusega ning automatiseeritult. Lisaks peab spottidel/positsioonidel olema ühesugune suurus ja jaotus ning unikaalne asukoht. Ühtlane jaotus lihtsustab ja kiirendab mikrokiipide tootmist, nendelt tuvastamist ja andmete analüüsimist. Ühesugune suurus, molekulide samaväärne arv erinevates spottides/positsioonides, on oluline kvantifitseerimiseks ja keskmise signaali intensiivsuse arvutamiseks. Unikaalne asukoht kindlustab kvantifitseerimise õigele järjestustele (Schena, 2003).

1.2.2 Mikroskoopilised elemendid

Spotid/positsioonid on sihtmärkmolekulide kogumikud, mis on võimelised hübridiseeruma spetsiifilise prooviga. Sihtmärkmolekule saab tuletada kas tervest geenist, geeni osadest või keemiliselt sünteesida. Samuti võivad nendeks olla ka genoomne DNA, cDNA, mRNA, valgud, väikeseid molekulid, koed või teised molekulid, mis võimaldavad kvantifitseerida geeni uuringuid. Mikroskoopiliste elementide eeliseks on see, et väikesed spotid/positsioonid võimaldavad väga suurt tihedust, kiiret reaktsiooni kineetikat ning analüüsida kogu genoomi ühel kiibil – võimaldavad minimeerimist ja automatiseerimist (Schena, 2003).

1.2.3 Tasapinnaline substraat

Tasapinnaline substraat on paralleelne ja paindumatu kandja, nagu klaas, plastik või räni, millele mikrokiip kujundatakse. Kõige ulatuslikumalt kasutatakse klaassubstraati. Klaasi eeliseks on odavus, seda on lihtne keemiliselt modifitseerida ning väike autofluorestsents. Tasapinnaline materjal on lame kogu pinna ulatuses. See kindlustab fotomaskide, nõelade (*pin*), tindipritspihustite (*ink-jet nozzle*) ja teiste valmistamisseadmete täpse vahemaa, mis tagab automatiseeritud tootmise ja kindlustab kõrge kvaliteedi. Tasapinnalisus on väga tähtis ka täpseks skaneerimiseks ja pilditehnikas (Schena, 2003). Mikrokiipide substraadid peavad tagama mikrokiibi analüüsimiseks kiibielementide stabiilse seondumise, minimaalse tausta müra, homogeense pinna ning andmete kõrge kvaliteedi (Stears jt., 2003).

1.2.4 Spetsiifiline seondumine

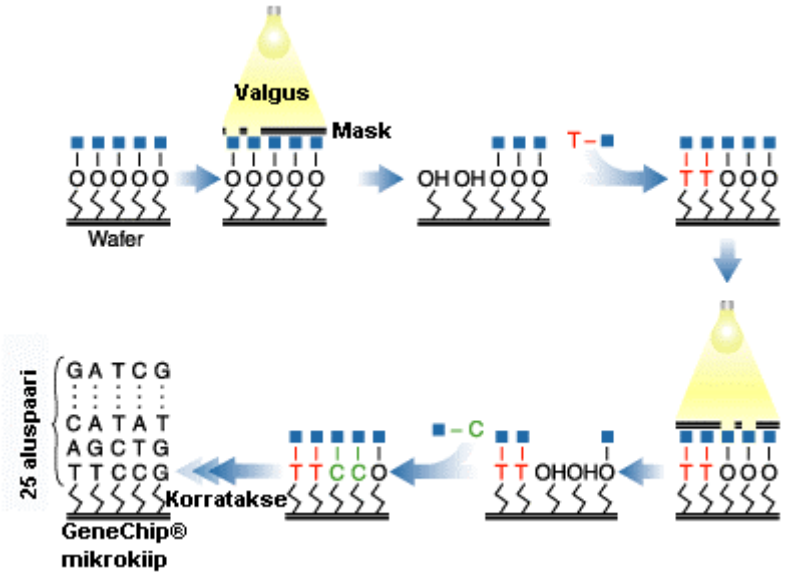
Spetsiifiline seondumine toimub proovimolekuli ja komplementaarse sihtmärkmolekuli vahel biokeemiliste interaktsioonide toimel. Seondumise spetsiifilisus võimaldab geeni või geeniprodukti kvantitatiivset analüüsimist. Kvantitatiivne analüüs on igasugune katse mikrokiipidega, mis võimaldab täpset molekulide arvu, hulga või kontsentratsiooni mõõtmist antud proovis. Niimoodi saab eristada näiteks homo- ja heterosügootseid patsiente (Schena, 2003).

1.3 Tasapinnalised mikrokiibid

Tasapinnalisi mikrokiipe valmistatakse *in situ* sünteesiga, mis on välja töötatud Affymetrix'i (Santa Clara, CA) poolt või spotitakse klaasile (Joonis 7), membraanile või mõnele teisele pinnale, kas mikrospottimise või tindipritsil põhineva printimismeetodiga (Venkatasubbarao, 2004).

1.3.1 *In situ* sünteesiga valmistatud mikrokiibid

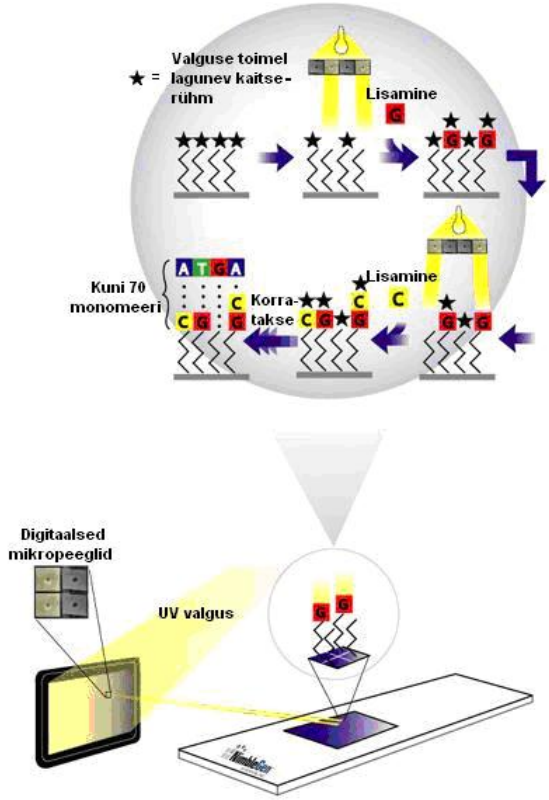
Affymetrix tehnoloogias kasutatakse mikrokiipide valmistamisel fotolitograafia ja kombinatoorse keemia kombinatsiooni (Joonis 3). Kasutatakse kroomist fotomaske ning täpseid maski joondajaid, mis teeb tootmise väga kulukaks. Mikrokiibile disainitakse 25-aluspaarilised proovid paaridena, millest üks on täpselt komplementaarne sihtmärgiga ning teine sisaldab täpselt keskkohas ühte mutatsiooni ja on ettenähtud sisemiseks kontrolliks (Venkatasubbarao, 2004).



Joonis 3. Affymetrix GeneChip® tootmine. Oligod sünteesitakse otse tahkele kandjale. Sinised ruudud tähistavad kaitserühmi. Oligote sünteesimisel kasutatakse fotomaske, mille abil juhitakse UV valgust soovitud positsioonidele. Valguse toimel eemaldatakse kaitserühmad ja seejärel lisatakse kaitserühmadega nukleotiide, mis reageerivad molekulidega, millel kaitserühmad eemaldati. Järgnevalt kasutatakse uut

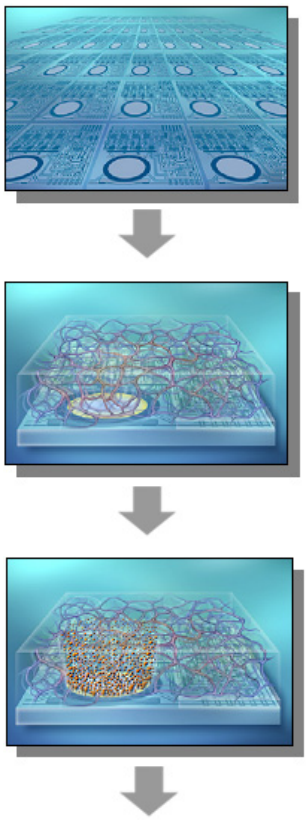
maski, mille abil suunatakse valgus järgmisele soovitud positsioonile. Lisatakse uued soovitud kaitserühmadega nukleotiidid ja protsessi korratakse kuni kiibile saadakse 25 bp pikkused oligod. http://www.affymetrix.com/corporate/outreach/lesson_plan/downloads/slides/photolithography_basics.pdf

NimbleGen tehnoloogias (Madison, WI) kasutatakse maskita mikrokiibi süntesaatorit (MAS), kasutades nn virtuaalset maski, mis koosneb sadadest tuhandetest arvuti abil suunatavatest mikropeeglitest, millega suunatakse UV kiirgust soovitud kohtadesse mikrokiibi pinnal (Joonis 4) (Venkatasubbarao, 2004).



Joonis 4. NimbleGen mikrokiipide valmistamine. Oligod sünteesitakse etappide kaupa otse tahkele kandjale. Meetod on põhimõtteliselt samasugune nagu Affymetrix'il, kuid UV valguse suunamiseks soovitud kohale kiibil kasutatakse fotomaskide asemel tuhandetest arvuti abil suunatavatest peeglitest koosnevat nn virtuaalse maski. http://www.nimblegen.com/technology/manufacture.html

CombiMatrix (Mukilteo, WA) kasutab mikrokiipide valmistamisel elektrokeemiat (Joonis 5). Iga elektrood asub bioühilduvas kihis, mis soodustab sünteesitud molekulide kinnitumist. Oligonukleotiid sünteesitakse virtuaalses kambris, mis hoiab keemilised reagendid mikroelektroodide kohal. Sünteesitud molekulid püütakse bioühilduvasse maatriksisse, mis võimaldab seda kasutada mikrokiibina (Venkatasubbarao, 2004).



Lab-on-a-Chip

Eraldi paiknevate mikroelektroodidega maatriksid võimaldavad sünteesida sadu või tuhandeid erinevaid molekule paralleelselt.

Porous Reaction Layer (PRL)

Poorne reaktsiooni kiht. Bioühilduv kiht, mis soodustab sünteesitud molekulide kinnitumist.

Virtuaalne kamber

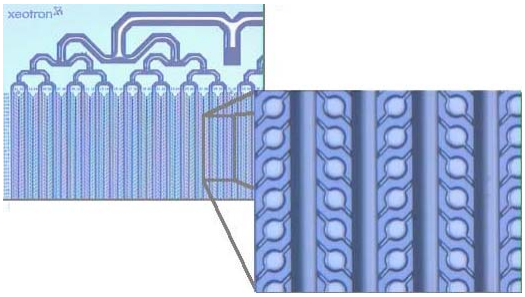
Keemilised reagendid hoitakse virtuaalses kambris mikroelektroodi kohal.

DNA mikrokiip

Kiibil sünteesitud molekulid on eraldatud bioühilduvasse maatriksisse.

Joonis 5. CombiMatrix mikrokiipide valmistamine.
http://www.combimatrix.com/tech_microarrays.htm

Xeotron (Houston, TX) kasutab valguse suunamiseks sarnaselt NimbleGen'ile digitaalset mikropeegel projektorit. Oligonukleotiidid sünteesitakse 3D mikrofluidilistest (*microfluidic*) nanokambritest (Joonis 6) koosnevasse mikrokiipi. Seda kasutatakse pikkade, kuni 150-nukleotiidiliste oligonukleotiide sünteesimiseks, aga ka RNA järjestuste, peptiidide ning orgaaniliste molekulide mikrokiipide tootmiseks (Venkatasubbarao, 2004).

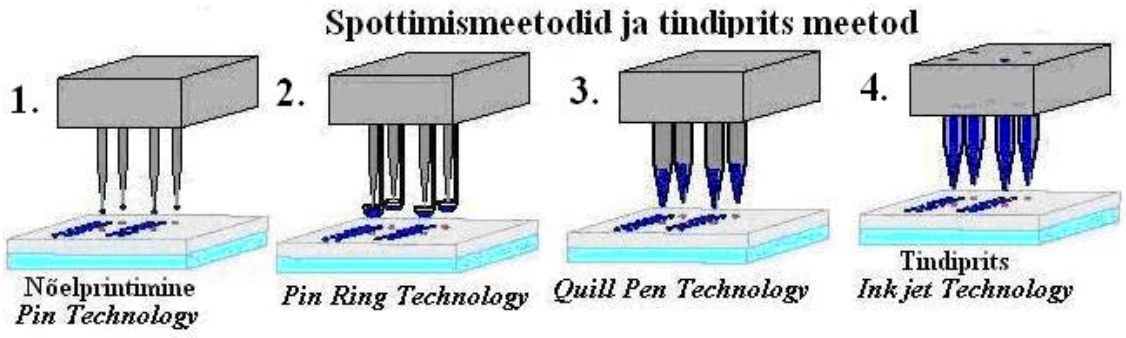


Joonis 6. Xeotron 3D mikrofluidiline kiip. Mikrokiip koosneb väikestest nanokambritest, kuhu sünteesitakse mikropeeglite poolt suunatud valguse toimel oligonukleotiidid. www.invitrogen.com/content.cfm?pageid=10620

1.3.2 Spottimise ja *ink-jet* tehnoloogiaga valmistatud mikrokiibid

Spottimisega valmistatakse nii oligonukleotiidide kui ka cDNA mikrokiipe. cDNA mikrokiipide puhul spotitakse tavaliselt PCR'i abil amplifitseeritud cDNA'd, mille pikkus on enamasti vahemikus 0,5 kuni 2,0 kb. Sihtmärgid võib valida ka andmebaasidest, nagu näiteks GenBank, dbESt, ja UniGene, või kasutada nendeks osaliselt sekveneeritud cDNA'de kollektsioone (EST) ja sünteesida need keemiliselt. Amplifitseerimiseks kasutatakse spetsiifilisi või universaalseid oligosid. Enne printimist fragmendid puhastatakse. Prinditakse tavaliselt polü-L-lüsiiniga kaetud mikroskoobi slaidile. DNA fragmendid seotakse kovalentselt UV valguse toimel. Pärast fikseerimist lisatakse suktsiin anhüdriidi, millega reageerivad amiini jäägid, et vähendada pinna positiivset laengut (Xiang ja Chen, 2000).

Spottimiseks kasutatakse erinevaid spotterite tüüpe (Joonis 7) ja substraadi pindasid. Derivatiseeritud substraadi pind määrab biomolekulide immobiliseerimise, ning kinnitumise tüübi (Van der Waals'i jõud, ioonne, kovalentne seondumine). Tavaliselt võimaldavad oligonukleotiidide, cDNA, valkude ja teiste molekulide stabiilset seondumist reaktiivne amiin, aldehüüd või epoksiid molekulid klaaspinnal (Stears jt., 2003).



Joonis 7. Spottimise ja tindiprits meetodid. 1. Tavaline nõelprintimine. Selle meetodi puhul saab korraga võtta vähem materjali, kuid nõelade pesemine on lihtsam, kui teiste meetodite puhul. 2. Nõelprintimise tõhusam meetod, mis võimaldab kiiremini proovi juurde laadida.3. *Quill Pen* meetodi puhul saab korraga rohkem materjali ja seega ka korraga rohkem spottida. 4. Tindiprits võimaldab korraga kasutada väga palju materjali.

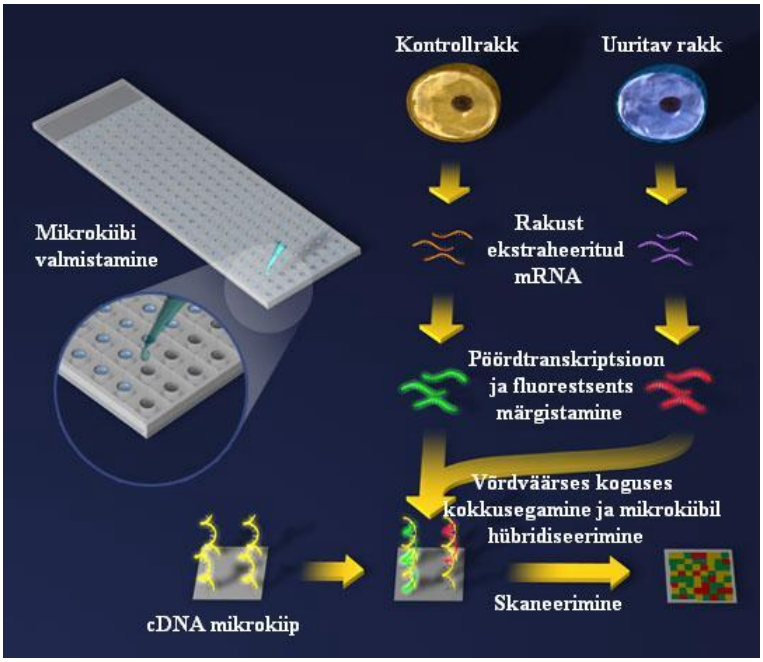
DNA immobiliseerimiseks kasutatakse tavaliselt fosfaatrühmasid, mis seonduvad ioonsete interaktsioonidega positiivselt laetud substraadi aminorühmadega. Aminiseeritud pindadel on neutraalse pH juures positiivne laeng, mis moodustab negatiivselt laetud DNA molekulidega ioonseid sidemeid. DNA kovalentseks sidumiseks kasutatakse UV või termilist töötlust. Sellised slaidid sobivad pikkade oligonukleotiidide ja cDNA immobiliseerimiseks. Sidumise parandamiseks kasutatakse sageli ka pikaahelalisi speissermolekule (Venkatasubbarao, 2004).

1.4 Mittetasapinnalised mikrokiibid

Metrigenix tehnoloogias (Gaithersburg, MD) kasutatakse mikrokiibina läbivoolu-kiibi platvormi. Mikrokiip koosneb poorsetest mikrokanalite võrgustikust. Selline disain võimaldab suuremat pindala ja seostumise tihedust. Tehakse ka kuulikestega mikrokiipe (*Bead arrays*). Need kõik on madala tihedusega mikrokiibid (Venkatasubbarao, 2004).

2 Mikrokiipide kasutusalad

DNA mikrokiibid koosnevad tuhandetest või sadadest tuhandetest klaasile kinnitatud hübridiseerumisvõimelistest cDNA fragmentidest või sünteetilistest oligonukleotiididest. Fluorestsentselt märgistatud RNA või DNA proovi hübridiseerumisel kiibil olevate fragmentidega võimaldab vaadelda geeni ekspressiooni või polümorfismide esinemist genoomses DNA's (Gerhold jt., 1999). Mikrokiibid võimaldavad paralleelselt analüüsida komplekssseid biokeemilisi proove. DNA mikrokiibi analüüsimine seisneb mikrokiibi tootmises, proovi prepareerimises, hübridisatsioonis ja andmete analüüsimises (Joonis 8). Praegu on laialt kasutusel kaks DNA kiibi vormi: cDNA kiibid (Schena jt., 1995) ja suure tihedusega sünteetiliste oligonukleotiididega kiibid (Lockhart jt., 1996; Pease jt., 1994; Gerhold jt., 1999). Oligonukleotiidide ekspressiooni kiipide hulka kuuluvad lühikestest oligotest (20-25 bp) (näiteks Affymetrix) koosnevad kiibid ja pikkadest oligotest (50-120 bp) koosnevad kiibid (Kane jt., 2000).



Joonis 8. cDNA mikrokiibi analüüsi skeem. Kõigepealt valmistatakse mikrokiip ja seejärel proov. Proovi valmistamiseks eraldatakse kahest erinevast rakust mRNA, mis konverteeritakse fluorestsentsete nukleotiidide kasutamisega cDNA'ks. Kaks proovi

segatakse võrdses mahus kokku ja lastakse hübridiseeruda mikrokiibil olevate sihtmärkmolekulidega. Pärast hübridisatsiooni pestakse seondumata molekulid maha ja kiipi skaneeritakse. Saadud pildi põhjal sooritatakse andmete analüüsimine. http://www.bioteach.ubc.ca/MolecularBiology/microarray/

Mikrokiibid võimaldavad uurijatel läbi viia suuremastaabilisi kvantitatiivseid eksperimente, võimaldades paralleelselt määrata väga suurel hulgal ühenukleotiidilisi polümorfisme (SNP) (Fan jt., 2003). Kõige enam rakendatakse mikrokiipe mRNA ekspressiooni profileerimiseks, mis 2003. aastal hõlmas 81,5% kõikidest mikrokiipide rakendustest (Schena, 2003). Mikrokiipe kasutatakse ka uute geenide, transkriptsioonifaktorite seondumissaitide, DNA koopiaarvu muutuste, üksikute järjestusvariatsioonide või komplekssete mutatsioonide identifitseerimiseks haigusi põhjustavates geenides (Stoughton, 2004). Samuti võimaldavad mikrokiibid analüüsida valgu taset rakkudes, kudesid, rakke ning teisi bioloogilisi ja keemilisi molekule massiivselt ja paralleelselt (Stears jt., 2003).

2.1 Ekspressiooni profileerimine – ekspressiooni analüüs

Ekspressiooni profileerimine võimaldab uurijatel saada kvantitatiivset geenieksprессiooni informatsiooni paljude geenide kohta paljudes proovides paralleelselt (Stears jt., 2003). DNA mikrokiibid on saanud standardseteks vahenditeks molekulaarbioloogia uurijatel ja kliinilises diagnostikas. Kõige enam kasutatakse mikrokiipe mRNA määramiseks. mRNA annab rikkalikku informatsiooni geenide funktsiooni kohta nii rakkudes kui kudedes. RNA ekstraheeritakse paljudest rakkudest, ideaalis samast rakutüübist, ja konverteeritakse cDNA'ks või cRNA'ks. Kontsentratsiooni suurendamiseks amplifitseeritakse seda PCR'i abil. Fluorestsentsmärgised lisatakse kas ensümaatiliselt sünteesitavatele ahelatele või liidetakse pärast keemiliselt. cDNA või cRNA ahelad, mis on komplementaarsed spotil olevate immobiliseeritud fragmentidega, hübridiseeruvad aluspaaride paardumisega. Spotid annavad fluorestsentssignaali, kui neid skaneerida mikrokiibi skanneriga (Schena, 2003).

Lühikesed oligod ei kindlusta alati ühe geeni spetsiifilisust ja seetõttu kasutatakse tihti mitut oligonukleotiidi ühe geeni kohta. SNP'de detekteerimine vajab ühe valepaardumise (*mismatch*) eristamist, mille puhul kasutatakse enamasti lühikesi oligonukleotiide, et maksimaliseerida valepaardumisest põhjustatud keemilise destabilisatsiooni efekti. PCR'i produktid on pikemad ning annavad seetõttu suuremat signaali ja suuremat täpsust. Pikad oligonukleotiidid võimaldavad samuti tugevat hübridisatsiooni signaali, head täpsust ja ühetähenduslikku proovi tuvastamist. Oligonukleotiidide puhul on eelnevalt vaja teada järjestuse informatsiooni (Stears jt., 2003).

2.2 Mikrokiipide rakendused mikroobsetes süsteemides

Tervet mikroobi genoomi saab lihtsalt esindada ühel mikrokiibil, mis teeb võimalikus analüüside sooritamise kogu genoomi ulatuses (DeRisi jt., 1997). Bakteriaalsetel analüüsidel märgistatakse tavaliselt kogu rakuline RNA, kuna bakteritel puudub polü(A) saba mRNA 3' otsas ja seetõttu on seda raske eraldada kogu RNA'st. Üldiselt on ka bakteriaalse mRNA pooleluiga palju lühem, kui eukarüootidel, mis teeb intaktse mRNA populatsiooni isoleerimise väga raskeks (Ye jt., 2001).

Mikrokiipide üldisteks rakendusteks prokarüootsetes süsteemides on transkriptsiooni aktiivsuse uurimine kogu genoomi ulatuses, regulonide (*regulon*) kindlaks tegemine, operonide struktuuri kirjeldamine, tundmatute DNA regioonide uurimine, DNA-valk interaktsioonide uurimine ning võrdlev genoomika ja genotüpiseerimine. Spetsiifilisemad rakendused on patogeenide virulentsusfaktorite määramine, peremehe vastused patogeenidele või normaalsele mikrofloorale, ravimite, inhibiitorite ja toksiliste ühendite geeniekspressiooni profiilid, mikroobse evolutsiooni ja epidemoloogia analüüsid. Kasutatakse ka diagnostilise vahendina (Ye jt., 2001).

2.3 cDNA mikrokiibid ja nende rakendamine

cDNA mikrokiibid on võimsad vahendid geeniekspressiooni uurimiseks. Neid on rakendatud edukalt samaaegselt mitmete tuhandete geenide ekspressiooni ja suuremastaabilistes genoomi uuringutes, samuti genoomse DNA polümorfismide

sõeltestimisel (*screening*) ja kaardistamises. Võimaldavad kvantitatiivselt analüüsida nii teadaolevatelt kui ka tundmatutelt geenidelt transkribeeritud RNA'd. Kuna samaaegselt on võimalik jälgida ekspressiooni taset paljudel geenidel, saab uurida ka komplekssed raku kontrollsüsteeme (Xiang ja Chen, 2000).

cDNA mikrokiibi tehnoloogiat on kasutatud komplekssete haiguste ja uute haigustega seotud geenide identifitseerimiseks. Samuti on neid kasutatud geeni ekspressioonimustrite analüüsimiseks vähi rakkudes (DeRisi jt., 1996). Heller ja kaasautorid uurisid geenieksрessiooni iseloomu põletikulises reumatoidartriidis ja põletikulises südame haiguses (Heller jt., 1997). Welford ja kaasautorid uurisid geeni ekspressiooni erinevusi algses vähikoes (Welford jt., 1998). Iyer ja kaasautorid rakendasid mikrokiipe inimese kasvu ja rakutsükli kulgemise uurimiseks, uurides 8600 erinevat geeni (Iyer jt., 1999).

Geenieksрessiooni rajad pakuvad kaudset informatsiooni geeni funktsiooni kohta. Geeni ekspressiooni tundmine, samuti järjestuste homoloogia tuntud geeni perekondades võib pakkuda sobivat otseteed sihtmärgi rolli kohta antud rajas või haiguses. cDNA mikrokiibi tehnoloogia abil on suutnud paljud farmaatsia firmad identifitseerida sobivaid sihtmärke terapeutiliseks sekkumiseks (Xiang ja Chen, 2000). Mikrokiipe on kasutatud ka geenieksрessiooni muutumiste jälgimiseks vastusena ravimenetlusele. Marton ja kaasautorid sooritasid ravimi ratifitseerimise uurimust ja identifitseerisid ravimi sekundaarseid efekte (Marton jt., 1998).

2.4 Teised rakendused

Mikrokiipe on kasutatud ka uute geenide, transkriptsioonifaktorite sidumissaitide, DNA koopiaarvu muutuste ja aluspaaride variatsioonide identifitseerimiseks. Neid saab kasutada haiguse seisundi määramiseks, võrreldes haige ja normaalse inimese olukorda. Haiguse olukorras võivad olla osad geenid kas üle- või alareguleeritud. Samuti saab uurida splaissingut ja leida eksoneid (Stoughton, 2004).

Kui geeni koopiaarv on muutunud, siis sellele vastab ka mRNA taseme muutus. Sel teel on mikrokiipidega avastatud aneuploidsus pärmis (Hughes jt., 2000). Koopiaarvu muutust on võimalik näha otse genoomse DNA fragmentide

kontsentratsioonist spetsiifilises genoomi regioonis ja seda on rakendatud vähiga seotud muutuste skaneerimiseks (Lucito jt., 2003).

Kõrgelt paralleelne patogeenide genoomide „ülekuulamine" mikrokiipidega võimaldab põhjalikult määrata infektsioonhaiguste diagnoose, jälgida tekkivaid infektsioone, jälgida toidu, vee ning õhu saastust (Stoughton, 2004).

Mikrokiipe on hakatud palju kasutama proteoomikas. Väike reaktsiooni maht vähendab kulusid reagentide jaoks ja võimaldab väga kiiret seondumise kineetikat ning paralleelsust. Prinditud valkude suur kontsentratsioon stabiliseerib ka valkude struktuuri ja printimise puhvrid kaitsevad valku oksüdatsiooni eest. Mikrokiipidel on võimalik uurida valk-valk interaktsioone, valk-ravim interaktsioone, posttranslatsioonilisi modifikatsioone jne (Stears jt., 2003).

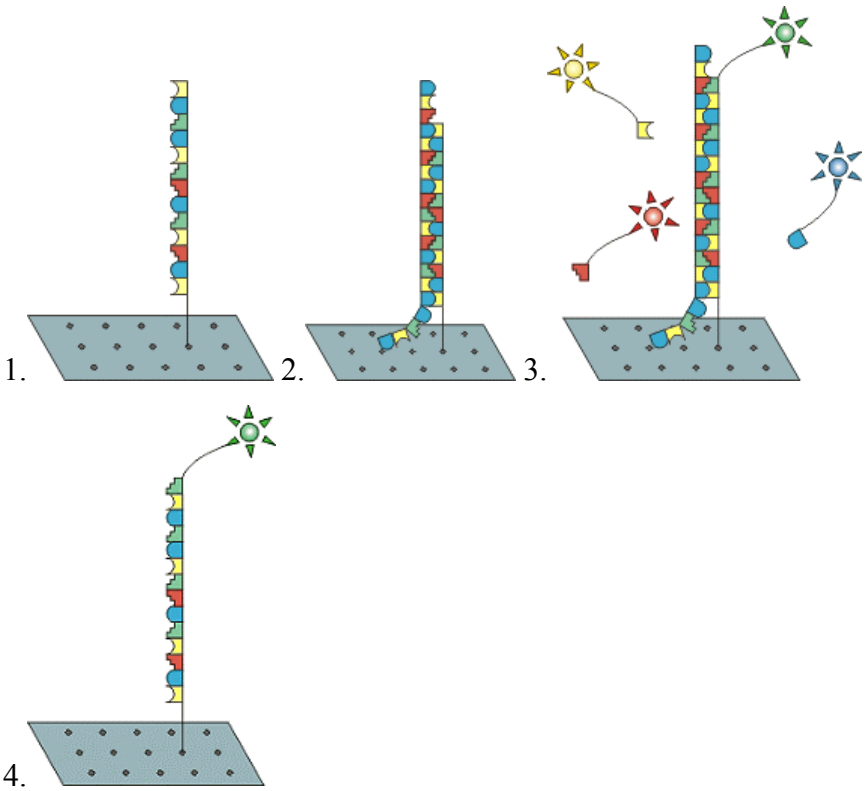
Väga laialt kasutatakse mikrokiipe sõeltestimisel. Uuritakse ühenukleotiidilisi polümorfisme, deletsioone ja teisi mutatsioone geenides, mis on aluseks geneetilistes haigustes. Sõeltestimiseks ja diagnoosimise rakendamiseks on arendatud kaks peamist mikrokiibi vormi: „*Single-patient*" ja „*Multipatient*" mikrokiibid. „*Single-patient*" oligonukleotiidide kiibid on mõeldud ühe patsiendi mutatsioonide määramiseks. Selle kiibi eeliseks on suure hulga järjestuste määramine ühe testiga. Puuduseks on kõrge hind patsiendi kohta. „*Multipatient*" mikrokiibid võimaldavad määrata paljudel patsientidel väiksemal hulgal lookuseid. Eeliseks on tuhandete või isegi kümnete tuhandete patsientide analüüsimine ühel kiibil. Puuduseks on, et patsiendi proovi peab printima enne genotüpiseerimist mikrokiibile, milleks kulub lisaaega (Stears jt., 2003).

3 APEX meetod

Arrayed Primer EXtension ehk praimerekstensioon oligonukleotiidmaatriksil on DNA analüüsi meetod, mis põhineb komplementaarsete fragmentide hübridisatsioonil ja DNA polümeraasi vahendatud praimerekstensiooni reaktsioonil (Joonis 9). APEX'it saab kasutada SNP'de, deletsioonide ja insertsioonide analüüsimiseks. Selle meetodi puhul pole võimalik analüüsida pikki kordusjärjestusi, nagu mikrosatelliidid (Kurg jt., 2000).

APEX on kompleksne süsteem, mis hõlmab kiipide ja proovide valmistamist, APEX reaktsiooni, fluorestsentssignaali detekteerimist ning andmete analüüsimist.

Kiipide valmistamisel immobiliseeritakse klaasslaidile 5' otsa kaudu uuritavate lookuste suhtes komplementaarsed 25-meersed oligonukleotiidid. APEX reaktsiooni jaoks vajaliku koopiaarvu saamiseks amplifitseeritakse uuritavat materjali PCR'i abil. PCR'i produktid on suured (enamasti 100-800 bp) ja pole eriti efektiivsed hübridisatsiooniks. Seetõttu kasutatakse PCR'i reaktsiooni segus dTTP nukleotiidide koguhulgast ≈20% ulatuses dUTP nukleotiide. See võimaldab fragmenteerida DNA'd uratsiil-N-glükosülaasi (UNG) abil, mis eemaldab DNA ahelast uratsiili. Sellise DNA denatureerimisel tekivad hübridisatsiooniks sobiliku pikkusega fragmendid. APEX reaktsiooni käigus lastakse proovil hübridiseeruda immobiliseeritud oligotega, millele järgneb matriits ahelast sõltuv ühenukleotiidiline ekstensioon DNA polümeraasi poolt, kasutades nelja unikaalset fluorestsentselt märgistatud terminaator- ehk didesoksünukleotiidi. Seejärel detekteeritakse fluorestsentssignaal, mille järgi analüüsitakse arvuti abil mutatsioone (Kurg jt., 2000).



Joonis 9. APEX meetodi põhimõtteline skeem. 1. Klaasist slaidile on immobiliseeritud 5’ otsa kaudu kuni 6000 teadaoleva järjestusega 25-aluspaarilist oligonukleotiidi (DNA kiip). 2. PCR’iga amplifitseeritud DNA seondub komplementaarse oligonukleotiidiga. 3. DNA polümeraasi abil lisatakse oligole sõltuvalt sihtmärk DNA järjestusest üks neljast märgistatud terminaatornukleotiidist. 4. Seondumata DNA fragmendid ja terminaatornukleotiidid pestakse maha. Signaali detekteerimisel saab määrata paralleelselt tuhandeid SNP’sid. http://www.asperbio.com/APEX.htm

4 Oligo disaini vajalikkus

Biomeditsiini uurimislaborid, biotehnoloogia- ja farmaatsiatööstused on välja arendanud väga palju suure jõudlusega meetodeid genoomide uurimiseks, millest paljud sõltuvad nukleiinhapete amplifitseerimisest PCR'i abil. PCR'i tegemiseks on vaja DNA'le kinnituvaid spetsiifilisi oligosid või oligopaare, mis moodustaksid stabiilse dupleksi ainult spetsiifilise saidiga huvipakkuval sihtmärk DNA'l. Spetsiifiline oligo on selline oligo, mis seondub antud hübridisatsiooni tingimustel ainult soovitud sihtmärgiga (Benita jt., 2003).

PCR'i produktidel põhinevate meetodite puhul on väga tähtis saada kvaliteetset produkti. Selle saavutamiseks on vaja kasutada kvaliteetseid PCR'i oligosid, mida tuleb hoolikalt *in silico* kontrollida, vältimaks oligote seondumist mittesoovitud järjestustele ning mittespetsiifiliste produktide teket. Probleemiks on just korduvad järjestused, mis võivad põhjustada risthübridisatsiooni. Seetõttu on vaja hoolikalt kontrollida kõiki võimalikke alternatiivseid seondumiskohti. Väga tähtis on ka see, et oligod ei seonduks iseendaga ega teiste oligotega, mis vähendab nende hulka PCR'i reaktsioonisegus. Samuti peavad PCR'is kasutatavad oligopaarid olema võimalikult lähedase sulamistemperatuuriga (Matveeva jt., 2003).

SNP'd on kõige sagedasemad DNA järjestuse variatsioonid inimese genoomis, keskmise sagedusega 1–2 kb tagant (Sachidanandam jt., 2001) ja seetõttu on nad leidnud laialdast kasutamist markeritena paljudes geneetilistes uuringutes. Selleks, et koostada suure tihedusega SNP kaarte või edukalt läbi viia eelnimetatud uuringuid, on vaja disainida kõrgekvaliteedilisi PCR'i ja sekveneerimis oligosid, eelistatavalt kiiresti ja automatiseeritult (Weckx jt., 2005).

Oligonukleotiidide kiipide puhul sõltub andmete kvaliteet valitud oligotest kiibil. Kiibile kinnitatud oligod on sihtmärkideks hübridisatsiooniproovis olevatele märgistatud komplementaasetele DNA või mRNA proovimolekulidele. Enamasti on oligotel ainult üks sihtmärk, nad on unikaalsed vastavas proovide segus, mis klaasile kantakse. Mitteunikaalsetel oligotel võib olla alternatiivseid sihtmärke, isegi kui nende vahel on üksikuid valepaardumisi või valepaardumised paiknevad proovi otste lähedal. Selle vältimiseks on väga oluline disainida optimaalsed oligod. Optimaalne oligo peaks

olema minimaalse hübridisatsiooni vabaenergiaga spetsiaalse proovi jaoks (määratud hübridisatsiooni tingimustel) ja maksimaalse energiaga kõikide teiste proovide jaoks (Li ja Stormo, 2001).

Oligote disainimisel mikrokiipide jaoks on väga tähtis, et oligod ei annaks soovimatuid risthübridisatsioone ja et kõikidel oligotel oleks minimaalne sulamistemperatuuri varieeruvus. Madalama hübridisatsiooni temperatuuri kasutamine soodustab kõrgema sulamistemperatuuriga oligotel alternatiivseid seondumisi, mis annab andmete analüüsimisel vale signaali. Samuti on mikrokiibi oligote puhul tähtis, et nad ei moodustaks stabiilseid sekundaarstruktuure, mis pärsib soovitud hübridisatsiooni ja annab jällegi vale signaali (Reymond jt., 2004).

5 Oligo kvaliteeti mõjutavad parameetrid

Üheks tähtsamaks parameetriks oligo disaini kvaliteedi juures on oligo stabiilse dupleksi moodustamine eranditult spetsiifilise saidiga sihtmärk DNA'l. Sellise olukorra saavutamiseks on vajalik disainida oligo, millel oleks unikaalne nukleotiidne järjestus. Lisaks on oligo disainimisel olulised oligo sulamistemperatuur (Tm) ja termodünaamika, oligo G/C sisalduse %, iseendaga komplementeerumine, oligo pikkus ning oligo 3' terminaalne järjestus (Haas jt., 1998).

5.1 Oligo sulamistemperatuur

Oligo sulamistemperatuur on selline temperatuur, mille juures pooled DNA ahelatest on üheahelalised ja pooled kaheahelalised. Oligote ideaalne sulamistemperatuur peaks olema vahemikus 50-70 °C. PCR'i jaoks kasutatavad oligote paarid tuleb disainida võimalikult lähedase sulamistemperatuuriga. Oligonukleotiidide kiipide puhul tuleb disainida aga mitmed sajad või tuhanded oligod võimalikult lähedase sulamistemperatuuriga. Sulamistemperatuur sõltub oligo pikkusest ja G/C sisaldusest (Haas jt., 1998).

Sulamistemperatuuri arvutamiseks on välja töötatud mitmeid erinevaid valemid. Kõige lihtsamaks võimaluseks lühikeste järjestuste (16-28 nukleotiidi) sulamistemperatuuri arvutamiseks võib kasutada valemit, mida nimetatakse ka Wallace'i reegliks:

$$Tm = 4*(C+G)+2*(A+T)$$

ehk G ja C nukleotiidile omistatakse 4°C, A ja T nukleotiidile 2°C (Suggs jt., 1981).

Sulamistemperatuuri täpsemaks arvutamiseks kasutatakse dermodünaamilist *nearest-neighbor* (NN) mudelit (SantaLucia, 1998). Selle meetodi puhul arvestatakse, et energia, mida on vaja paardunud ahelas ühe aluspaari vahelise vesiniksideme lõhkumiseks, sõltub kõrvalolevast aluspaarist. Selle jaoks on välja arvutatud dimeerdupleksite termodünaamilised suurused (Tabel 1).

Tabel 1. Oligonukleotiidide ΔH° ja ΔS° *nearest-neighbor* parameetrid.

Interaktsioon	ΔH°(kcal/mol)	ΔS°(cal/K mol)
AA/TT	7,9	22,2
AT/TA	7,2	20,4
TA/AT	7,2	21,3
CA/GT	8,5	22,7
GT/CA	8,4	22,4
CT/GA	7,8	21,0
GA/CT	8,2	22,2
CG/GC	10,6	27,2
GC/CG	9,8	24,4
GG/CC	8,0	19,9

ΔH° on entalpia heeliksi formeerumisel ja ΔS° entroopia. Kehtivad 1 M NaCl kontsentratsiooni, 25 °C ja pH 7 juures (SantaLucia, 1998).

Valem sulamistemperatuuri arvutamiseks näeb välja järgmine:

$$Tm = \frac{\Delta H^\circ}{\Delta S^\circ + R * \ln(C/4)},$$

kus ΔH° on entalpia, ΔS° on entroopia muutus (Tabel 1), R on universaalne gaasi konstant (1,987 cal/kmol) ja C on oligonukleotiidide molaarne kontsentratsioon lahuses. Valem kehtib 1 M Na^+ ioonide kontsentratsiooni juures (SantaLucia, 1998).

Mõned autorid ei soovita seda valemit kasutada üle 50-nukleotiidiste proovide sulamistemperatuuri arvutamiseks, kuna pikemate proovide puhul on sulamistemperatuurid tegelikult natukene madalamad (Rouillard jt., 2003; Haas jt, 2003).

Kasutaja poolt soovitud soola kontsentratsiooni korral saab kasutada täpsemat valemit:

$$Tm = \frac{T^\circ * \Delta H^\circ}{(\Delta H^\circ - \Delta G^\circ + R * T^\circ \ln[C/4])} + 16{,}6 * \log_{10}\{\frac{[Na^+]}{(1 + 0{,}7[Na^+])}\} - 269{,}3,$$

kus ΔG° on standartne vabaenergia, T° on temperatuur (Lee jt., 2004).

ΔG° (dupleksi stabiilsus) arvutamiseks kasutatakse valemit ΔG° = ΔH° – TΔS°, kus T = 293 K. ΔG° väärtused vastavad energiale, mis kulub vastavate nukleotiidipaaride vesiniksidemete lõhkumiseks (Breslauer jt., 1986). Näiteks järjestuse GGAT korral on *nearest-neighbor* termodünaamika: ΔH°(GGAT) = ΔH°(GG) + ΔH°(GA) + ΔH°(AT) = 8,0 + 8,2 + 7,2 = 23.4 (kcal/mol).

5.2 Oligo G/C sisalduse %

Suurema G/C sisaldusega oligotel on kõrgem sulamistemperatuur, mis on tingitud rohkematest vesiniksidemete arvust. Madalama sulamistemperatuuriga oligote disainimiseks kasutatakse väiksemat G/C sisaldust ja vastupidi. Optimaalne G/C sisaldus peaks oligotel olema ≈50, kuna sellisel juhul on väiksem tõenäosus sekundaarsete seondumiskohtade leidumiseks (Haas jt., 1998).

PCR'i oligote puhul peaks oligote G/C sisaldus olema ligikaudu sama või natukene suurem kui amplifitseeritaval DNA'l (Benita jt., 2003).

Hübridisatsioonil seonduvad G ja C nukleotiidid omavahel tugevamini kui A ja T nukleotiidid. G ja C nukleotiidid moodustavad omavahel paardumisel kolm, A ja T nukleotiidid kaks vesiniksidet. On näidatud, et paljude lühikeste G/C rikaste regioonide sisaldumine proovis mõjutab negatiivselt proovi ja sihtmärgi hübridiseerumist, soodustades mittespetsiifilist risthübridisatsiooni (Rouillard jt., 2003).

5.3 Oligo sekundaarsed seondumiskohad

Oligote disainimisel tuleb kontrollida kõiki võimalikke seostumiskohti sihtmärgil, vältimaks lisaproduktide tekkimist ja sellega kaasnevaid ebatäpseid tulemusi. Enamasti on see probleemiks just kordusjärjestuste puhul. Kandidaatoligosid on võimalik mitmetes andmebaasides teadaolevate kordusjärjestuste suhtes kontrollida, kuigi andmebaasid pole veel täielikud. Suhe originaalse ja kõige stabiilsema sekundaarse seondumissaidi dupleksi vahel on oligo suhtelise unikaalsusele määrajaks. Oligotel, mis sisaldavad vähemalt kahte samasugust lühikest kordusjärjestust, on suurem tõenäosus termodünaamiliselt stabiilsete sekundaarsete sidumissaitide olemasoluks nendes piirkondades (Haas jt., 1998).

5.4 Iseendaga komplementeerumine

Hübridisatsiooni proovide disainimisel on oluline kontrollida, et proovid ei annaks stabiilseid sekundaarstruktuure. Iseendaga hübridiseerumisel (*self annealing*) võivad moodustuda nn *stem-loop* või *hairpin* struktuurid või oligo 3' otsa endaga seostumine (*self-end annealing*), mille tulemusel oligod ei saa enam sihtmärgile seonduda. Oligod võivad hübridiseeruda ka vastasoligoga täispikkuses või poole järjestusega. Iseenda ja teiste oligotega hübridiseerumise tulemusel väheneb sihtmärk DNA'ga seostuvate oligote kontsentratsioon (Haas jt., 1998).

5.5 Oligo pikkus

Oligo pikkus on erinevate meetodite puhul kõige varieeruvam parameeter. Mikrokiipidel läbiviidavatel katsetel peaksid ilma speisseriteta kiibile seotud oligod olema vähemalt 60 nukleotiidi pikad, kuna lühemad järjestused ei taga piisavat hübridisatsiooni efektiivsust (Hughes jt., 2001). Pikemad proovid seonduvad tavaliselt efektiivsemalt, kui lühikesed (Relógio jt., 2002). Sünteetilised hübridisatsiooni proovid on tavaliselt 25-70 bp pikad (Kane jt., 2000; Chen jt., 2002).

5.6 Oligo 3' terminaalne järjestus

Oligo 3' otsas võiks stabiliseerimiseks olla G või C nukleotiid (GC-*clamp*), kuna oligo spetsiifilise seostumise DNA'ga määrab 3' ots. Samas on näidatud, et kõrge G/C nukleotiidide sisaldus oligo 3' otsas võib põhjustada vale seostumist ja vähendada oligo spetsiifilisust (Li jt., 1997).

6 Oligo disaini levinuimad meetodid

Oligo disainimiseks on avalikult saadaval väga palju erinevaid programme. Enamus nendest kasutab oligo disainimiseks sarnast algoritmi või kriteeriumeid. Ükski olemasolevatest programmidest ei arvesta kõiki olulisi kriteeriumeid oligo valimisel (Rimour jt., 2005). Järgnevalt annan lühikese ülevaate mõnest üldlevinud praimeri disaini programmist (Tabel 2) ja nende algoritmidest.

Tabel 2. Mõned oligo disaini programmid ja nende saadavus.

Programm	**Saadavus**
OLIGOARRAY 2.0	jmrouill@umich.edu
OLIGODB	http://oligodb.charite.de
PERLPRIMER	http://perlprimer.sourceforge.net
PRIDE	http://www.dkfz-heidelberg.de/tbi_old/services/Pride/search_primer
PRIMEARRAY	christoph.dehio@unibas.ch
ROSO	http://pbil.univ-lyon1.fr/roso/
SNPBOX	jurgen.delfavero@ua.ac.be
UNIFRAG ja GENOMEPRIMER	http://molgen.biol.rug.nl/molgen/research/molgensoftware.php
GOARRAYS	http://www.isima.fr/bioinfo/goarrays

6.1 OLIGOARRAY 2.0

OLIGOARRAY 2.0 (Rouillard jt., 2003) on oligonukleotiidide disainimise programm kogu genoomi ulatuses geeni ekspressiooni profileerimiseks mikrokiipidel. Programmil saab seadistada maksimaalset oligonukleotiidide arvu, pikkuse vahemikku, G/C sisaldust ja sulamistemperatuuri, piirväärtust stabiilseid sekundaarstruktuure moodustavate oligonukleotiidide kõrvaldamiseks, piirväärtust risthübridisatsioonide arvestamiseks, paralleelselt protsessitavate järjestuste arvu mitme protsessori olemasolul ja minimaalset vahemaad kahe külgneva oligonukleotiidi 5' otsa vahel.

Sisend fail, mis võib sisaldada mRNA, CDS'i (*coding sequences*) või eksonijärjestusi, peab olema FastA formaadis. Igal sisendfaili kirjel määratakse pikkus. Kui see on pikem kui maksimaalne lubatud vahemik oligonukleotiidi 3' otsa ja

sisendjärjestuse 5’ otsa vahel, kahandatakse seda 5’ otsast kuni lubatud maksimumini ja kahandatud nukleotiidid asendatakse „N“ tähtedega ning järjestus maskeeritakse keelatud järjestuste olemasolu tõttu. Kõik keelatud järjestustele vastavad aluspaarid asendatakse „N“ tähega.

Oligonukleotiidide spetsiifilisuse tagamiseks määratakse kõigepealt järjestuse sarnasust sihtmärgi ja teiste järjestustega. Maskeeritud järjestusi võrreldakse teiste tuntud järjestustega BLAST (vt. peatükk 8.1) programmi abil. DUST (Tatusov ja Lipman, unpublished) filter on inaktiveeritud, et võrrelda kõiki järjestusi, ka neid mis eelnevalt maskeeritud. Otsingut on piiratud ainult kodeeriva ahela jaoks, kuna ainult sellelt saab pöördtranskriptsioonil märgistatud proove. Maksimaalse sarnasuse määramiseks kasutatakse vähimat lubatud sõna pikkust. E-väärtus (vt. peatükk 8.1) valitakse vastavalt päringu pikkusele, et kindlustada ka lühikeste joonduste raporteerimine. BLAST’i väljundis säilitatakse sisendi ja teiste järjestuste vahelised sarnasused.

Sisendjärjestust loetakse tagurpidi 3’ otsast kasutades minimaalse oligonukleotiidi pikkust liikuvat akent. Kõigepealt kontrollitakse keelatud järjestuste puudumist ja kasutaja poolt määratud G/C sisaldust. Kui üks nendest tingimustest ei sobi, liigub aken 5 nukleotiidi võrra 5’ otsa suunas ja kontrollitakse jälle kuni mõlemad tingimused on täidetud. Seejärel arvutatakse sulamistemperatuur, kasutades *nearest-neighbor* mudelit. Kui sulamistemperatuur on väljaspool valitud vahemikku, püüab programm kohandada pikkust ja sulamistemperatuuri sobivasse vahemikku. Kui sobivat kombinatsiooni ei leita, liigub aken 1 nukleotiidi võrra edasi ja testi korratakse. Kui see tingimus on täidetud, kontrollitakse sekundaarstruktuuride olematust. Minimaalse vabaenergia arvutamiseks kõikidel võimalikel sekundaarstruktuuridel kasutatakse MFOLD programmi (Zuker jt., 1999) ja SantaLucia termodünaamilisi parameetreid 1 M Na^+ kontsentratsioonil ja kasutaja poolt määratud temperatuuri vahemikus. Negatiivse vabaenergiaga oligonukleotiidid heidetakse kõrvale.

Järgnevalt otsitakse sarnasusmaatriksist (*similarity matrix*) sarnasusi oligonukleotiidide endi ja teiste järjestuste vahel, et arvutada kõikide võimalike hübridisatsioonide termodünaamilisi väärtusi (Tm, vabaenergia, entalpia, entroopia), kasutades jällegi MFOLD programmi. Arvutusi saab sooritada nii perfektse kattuvusega kui ka valepaardumistega kahe järjestuse vahel. Kui võimalikke risthübridisatsioone määratud temperatuuri vahemikus ei esine salvestatakse oligonukleotiid väljundfaili.

Võimalike risthübridisatsioonide korral salvestatakse andmed edasiseks kasutamiseks mällu. Kui oligonukleotiide leitakse vähem kui soovitud, arvestatakse ka mittespetsiifilisi oligonukleotiide, mis esitatakse järjekorras väiksemast arvust võimalike risthübridisatsioonidega.

OLIGOARRAY 2.0 genereerib kolm väljundfaili. Log fail sisaldab programmi olekut ja disainimise protsessi ja seda saab kasutada disaini ebaõnnestumise selgitamiseks. Teine fail, rejected.fas, sisaldab järjestusi, kuhu pole võimalik oligonukleotiide disainida, salvestatakse FastA formaati ja saab kasutada vähem rangete parameetritega protsessimiseks. Oligonukleotiidide andmed on salvestatud teksti faili, milles andmed on tabuleeritud formaadis. Iga oligonukleotiidi jaoks on antud geeni identifikaator, pikkus, hübridisatsiooni vabaenergia sihtmärgiga 37°C juures (kcal/mol), entalpia (kcal/mol), entroopia (cal/kmol) ja Tm (°C) dupleksil ning sihtmärgi (sihtmärkide) loetelu. Iga võimaliku risthübridisatsiooni jaoks on antud vabaenergia, entalpia, entroopia ja Tm.

6.2 OLIGODB

OLIGODB (Mrowka jt., 2002) on veebipõhine süsteem oligonukleotiidide disainimiseks transkriptsiooni profileerimiseks kasutaja poolt määratud parameetritega. Programm kasutab inimese geeni transkripte ENSEMBL projektist. Veebiliides võimaldab leida oligosid inimese transkrtiptide jaoks paindlikul viisil.

Oligo valimise kriteeriumiteks on minimaalne sarnasus teiste geenidega, stabiilsete sekundaarsete seondumiskohtade olematus, oligo pikkus, vähese keerukusega regioonide puudumine, mis võib põhjustada mittespetsiifilist seondumist. Oligosid on võimalik valida vastavalt soovitud positsioonile 5'–3' transkripti suunas. Sekundaarstruktuure moodustavad oligod välistatakse.

Sisendparameetriteks on transkript, oligo pikkus (L) ja maksimaalne kogu tabamuste arv (N). Kõigepealt võrreldakse BLAST'i abil kogu transkripti cDNA järjestust kõigi teiste ENSEMBL'is (Hubbard jt., 2002) olevate inimese cDNA järjestustega, et määrata teiste geenidega sarnased olevad transkriptide regioonid. Sarnased järjestused märgistatakse.

Järgnevalt määratakse kõikide võimalike L pikkuste oligote maksimaalne aluspaaride kattuvus (MO) teiste transkriptidega. Väiksema MO väärtusega oligod sorteeritakse LI listi, mida kasutatakse edasises töötluses. LI listist kõrvaldatakse vähem keeruliste piirkondadega (määratakse DUST programmiga) ja sekundaarstruktuure moodustavad (kasutatakse MFOLD programmi) kandidaatoligod. Negatiivse voltimisenergiaga (*folding energy*), 65°C ja 1M Na^+ kontsentratsiooni juures, oligod eemaldatakse LI listist. Tulemustena esitatakse LI listist N parimat oligot.

Väljundfail sisaldab oligo järjestust, minimaalset voltimisenergiat, sulamistemperatuuri ja oligo asukohta transkriptil. Sulamistemperatuur arvutatakse *nearest-neighbor* meetodiga programmi MELTING (Le Novère, 2001) abil.

Korraga on võimalik protsessida ka suuremal arvul transkripte üheaegselt. Tulemusi võib jälgida brauseris ja saab ka allalaadida tabuleeritud teksti failina.

6.3 PERLPRIMER

PERLPRIMER (Marshall, 2004) on mitmeplatvormiline programm praimerite disainimiseks standardseks, bisulfit ja reaalaja PCR'i (Quantitative PCR, QPCR) ning sekveneerimise jaoks.

Programm sisaldab suure täpsusega sulamistemperatuuri ja praimeri dimeeride ennustamise algoritme ning võimaldab järjestuse otsinguid ENSEMBL'ist ning BLAST'i otsingut.

Standard PCR'i praimerite disainimiseks kasutab programm avatud lugemisraami (ORF) detekteerimise algoritmi, mis leiab sisendfailist suurima ORF'i ja määrab amplifikatsiooni piirid leitud geenile. Programm võimaldab kasutajal lisada igale praimerile ekstra 5' järjestustusi. Kui lisatud järjestus sisaldab restriktsiooniensüümi kloneerimissiati ja see leidub ka sihtmärkvektoris, siis lisatakse sinna otsa automaatselt adeniin.

Praimeri disainimisel bisulfit PCR'i jaoks valitakse vaikimisi disainitud praimeritest sellised, mille 3' otsas on tsütosiin – võimaldab spetsiifilist bisulfitis konverteeritud sihtmärkide amplifikatsioni. Lubatud ei ole CpG järjestused. Programmis sisaldub CpG saari otsiv algoritm, mis seadistab leitud CpG saarte

piiritlemiseks automaatselt amplifikatsiooni piirid. Väikeste produktide (300-600 bp) puhul ei limiteerita amplifikatsiooni järjestusi automaatselt, soovi korral saab seda manuaalselt teha.

Nii bisulfit kui standard PCR'i puhul on määratud tingimustega sobivate praimerite leidmiseks võimalik automaatselt suurendada või vähendada amplifikatsiooni piire.

QPCR'i praimerite disainimine on täielikult automaatne. Kasutatakse SPIDEY rakendust (Wheelan jt., 2001), et leida intronite/eksonite piirid ja kõik võimalikud praimeripaarid neile. Praimeripaari valimiseks peab üks praimer ulatuma introni/eksoni piirile ja teine paiknema üle piiri. Vaikimisi on amplikoni suurus limiteeritud 100 - 300 bp. Kui geeni järjestus on võetud ENSEMBL'i andmebaasist (genoomne või cDNA) käib QPCR'i praimerite disainimine lihtsalt ja väga kiiresti.

Leitud praimeripaarid kantakse tabelisse, mis sisaldab järjestust, pikkust, asukohta, sulamistemperatuuri, amplikoni suurust ja $\Delta G°_{37}$ juures kõige stabiilsema praimeri dimeeri vabaenergiat. Praimerid saab sorteerida erinevate kategooriate järgi. Praimeripaare saab kasutada ka BLAST'i otsingus. Projekti failid saab salvestada tabuleeritud teksti faili, mis sisaldab praimeri termodünaamilisi detaile, praimeri dimeere ja täielikku järjestuste joondusi.

Sulamistemperatuur arvutatakse *nearest-neighbor* termodünaamilise meetodiga. Entroopia arvutamiseks saab kasutaja määrata Mg^{2+}, dNTP'de ja praimerite kontsentratsiooni. Praimeri dimeeride kalkuleerimiseks tehakse sidumise maatriks (*binding matrix*) iseenda, teiste ning vastaspraimerite jaoks, kust loetakse komplementaarsus. Ligikaudne $\Delta G°_{37}$ arvutatakse *nearest-neighbor* meetodiga, arvestades ka valepaardumise andmeid.

$\Delta G°$ sõltub entroopiast, mis omakorda sõltub soolakontsentratsioonist PCR'i reaktsioonis. Kui kasutaja muudab soola kontsentratsiooni, arvutatakse need väärtused ümber. Praimeri dimeeride stabiilsus arvutatakse nii pikendatavatele praimeri dimeeridele, mis annavad amplifitseeritavaid produkte kui ka mittepikendatavatele dimeeridele, mis vähendab vabade praimerite hulka reaktsioonisegus.

6.4 PRIDE

PRIDE (Haas jt., 1998) on praimeri disainimise programm üksikutele kontiigidele või terve sekveneerimisprojekti jaoks. Praimeri kvaliteedi kalkuleerimiseks kasutatakse häguse loogika (*fuzzy logic*) süsteemi. Arvutuslik jõudlus on täiustatud sufiksipuude (*suffix tree*) (Weiner, 1973) kasutamisega suurte koguste andmete salvestamisel. Kasutaja poolt on määratav ainult soovitud optimaalne sulamistemperatuur.

Optimaalsete praimerite otsimine on jagatud kahte ossa. Esiteks skaneeritakse kogu sihtmärkjärjestus, et leida üheahelalised regioonid. Praimeri disainimiseks määratakse alaregioonid üksikahelaliste regioonide servast kuni 500 aluspaarini naabruses olevast kaheahelalisest järjestusest. Alaregioonid määratakse ka kontiigide lõpus.

Teiseks kasutatakse iga alaregiooni konsensusjärjestusi oluliste parameetrite arvutamiseks kõigile võimalikele praimeri kandidaatidele. Luuakse sufiksipuu, mis sisaldab kõiki vajalikke järjestusi sekundaarsete seondumiskohtade määramiseks ja võimaldab kiiret juurdepääsu järjestuse osadele. Vaikimisi sisaldab see sihtmärgi järjestust ja vektorit, kuid saab lisada ka korduvaid elemente või teisi järjestusi andmebaasidest. Määratakse kõik võimalikud sekundaarsed seondumiskohad sekveneerimisvektoreites, teadaolevates sihtmärkjärjestustes ning kontiigidega seotud järjestustes ja soovi korral kordustes (Alu, LINE, MER jne). Sekundaarsete seondumiskohtade stabiilsus määratakse *nearest-neighbor* meetodiga. Praimeri dimeeride stabiilsus määratakse suurima arvu naabruses olevate komplementaarsete aluspaaride põhjal. Sekveneerimisel määramata jäänud aluspaaride kohtadele disainitud praimerid hinnatakse madala kvaliteediga ja nende aluspaaride asemele arvestatakse kõige halvemini sobivad aluspaarid. Ei eelistata järjestusi mis sisaldavad järjest kolme või enamat identset nukleotiidi. Praimerid püütakse valida ≈50 nukleotiidi ulatuses teadaoleva järjestuse 3’ otsast või kaheahelalistes regioonides ≈50 nukleotiidi ulatuses enne üksikahelalisi regioone. Sulamistemperatuur arvutatakse Suggs’i reegliga. Praimerite kalkuleerimisel ei arvestata lühikesi korduseid, silmuste moodustumist, G/C sisaldust, praimeri pikkust ega 3’ otsa aluspaari.

Järgnevalt määratakse häguse loogika abil kõigile praimeri kandidaatidele lõplik kvaliteet. Häguse loogika puhul võetakse sisendiks kasutaja poolt kategooriatesse liigendatud olulised parameetrid. Näiteks, termodünaamiline stabiilsus on jagatud „väga madal“, „madal“, „optimaalne“, „kõrge“ ja „väga kõrge“. Disainitakse matemaatiline mudel, mis kujundab sisendi parameetrid vastavalt kasutaja kvalitatiivsete kirjeldustega väljundparameetritesse. Praimeri kvaliteediks saadakse 0-100%. Lõpuks kirjutatakse nimekiri parimatest praimeritest iga sihtmärkregiooni jaoks.

PRIDE võimaldab praimereid disainida ka korduvatesse regioonidesse. Korduvad elemendid võivad mõnikord üksteisest erineda mõne aluspaari võrra ja nende vahel võib olla ka lühikesi mittekorduvaid järjestusi ja seetõttu on nendesse regioonidesse võimalik praimeried leida, kui pole stabiilseid sekundaarseid seondumiskohti teistes sama tüüpi elementidel. Korduvaid regioone saab kontrollida andmebaasidest.

6.5 PRIMEARRAY

PRIMEARRAY (Raddatz jt., 2001) on oligonukleotiidi paaride disainimise programm ulatuslikuks geenide amplifitseerimiseks PCR’i abil DNA mikrokiipide tootmise eesmärgil. Programm võimaldab kõiki parameetreid kasutaja poolt määrata. Kodeerivad järjestused saab ekstraheerida erinevatest sisendfaili formaatidest (näiteks GenBank-formaat).

Praimeripaaride kalkuleerimist sooritatakse praimeri kandidaate korduva nihutamisega mööda järjestust, kuni on saavutatud soovitud sulamistemperatuur, G/C sisaldus ja soovimatute iseendaga ja praimer-praimer interaktsioonide olematus. Sulamistemperatuuri arvutamiseks kasutatakse Suggs’i reeglit. Samas võimaldab programm rakendada ka *nearest-neighbor* meetodit. Praimeripaaride vahelised interaktsioonid ja isekomplementeerumine määratakse suurima arvu ümberkaudsete komplementaarsete aluspaaride järgi. Kontrollitakse ka G/C sisaldust ja sekundaarsete seondumiskohtade leidumist. Lisaks on võimalus kasutaja poolt määratud praimeripaaride analüüsimiseks, kasutaja poolt kontrollitav individuaalsete praimerite optimeerimine erinevate parameetritega ja maksimaalse huvipakkuva regiooni piiritlemine. Praimeripaaride optimeerimisel nihutatakse päripidist praimerit (*forward primer*) mööda kodeerivat järjestust kuni kõik soovitud tingimused ühe sobiva praimeri

jaoks on täidetud. Järgnevalt nihutatakse äraspidist praimerit (*reverse primer*) kodeeriva järjestuse lõpust kuni kõik tingimused ühe sobiva praimeri jaoks on täidetud ning on saavutatud minimaalsed interaktsioonid praimeripaaride vahel. Kui ühtegi rahuldavat äraspidist praimerit ei leita, siis nihutatakse päripidist praimerit edasi kuni on täidetud ühe praimeri jaoks kõik tingimused jne. Seda strateegiat korratakse korduvalt. Kui amplifitseeritava DNA fragment on limiteeritud, siis saab määrata ka vastavad piirväärtused (*cut-off*).

Disainitud praimeri paarid ja prognoositava PCR'i produkti suurused saab kirjutad Excel'i arvutustabelisse või ASCII faili.

6.6 ROSO

ROSO (Reymond jt., 2004) on optimaalsete oligonukleotiidi proovide disainimise programm mikrokiipide jaoks. Rakendamiseks on vaja kahte erinevat FastA formaadis sisendfaili: huvipakkuv fail, mis sisaldab kiibile spotitavate geenide järjestusi ja fakultatiivne väline fail, mis sisaldab kõiki geenide järjestusi, mida ei spotita. Sellega võimaldatakse vältida risthübridisatsioone teadaolevate geenidega. Kui disainimine sooritatakse mittekompleksse täiendava välise failiga, on kasulik eemaldada huvipakkuvast failist korduvad järjestused.

Proovi valiku alustamiseks on vaja määrata sihtmärkide ja soola kontsentratsioon, proovi suurus ja orientatsioon, proovi asukoht geeni järjestusel, proovide arv geeni kohta, piirväärtused sekundaarstruktuuride stabiilsuse jaoks ning hübridisatsiooni temperatuur. Saab kasutada ka vaikimisi parameetreid.

Algoritm on jagatud viide järjestikulisse etappi tähtsuse kahanemise järjekorras.

1. Huvipakkuvat faili kärbitakse 3 viisil. Esiteks eemaldatakse >100 aluspaari pikkused järjestused, mille sarnasus on üle 98% või EST järjestuste puhul >95%. Igast sarnaste geenide grupist jäetakse huvipakkuvasse faili ainult pikemad ja sarnaste geenide nimed raporteeritakse väljundfailis. Teiseks saab kasutaja valida proovi asukoha n nukleotiidi piirides geeni 3' või 5' otsast. Kolmandaks, edasisteks toiminguteks ei valita kindlaksmääramata nukleotiididega proove. Kasutaja võib nõustuda või keelduda nelja järjestikku identset nukleotiidi sisaldavate järjestuste kaasamisest.

2. BLAST programmi abil kontrollitakse eeldatavaid risthübridisatsioone iga geeni jaoks huvipakkuvas ja välises failis, määrates minimaalselt 70% identsust 20 aluspaari vahel. Iga eeldatava proovi jaoks arvutatakse risthübridisatsiooni homoloogsuse skoor (*score*) (BLAST'i maksimaalsel tundlikkusel).

3. Parimatel proovidel otsitakse stabiilseid sekundaarstruktuure. Analüüsitakse kõiki sekundaarstruktuuride konformatsioone ja vabaenergia arvutatakse kasutaja poolt määratud hübridisatsiooni temperatuuri jaoks. Stabiilseid sekundaarstruktuure ilmutavad proovid eemaldatakse nimekirjast.

4. Arvutatakse järelejäänud proovide sulamistemperatuur kasutades *nearest-neighbor* termodünaamilist mudelit.

5. Optimaalsete proovide valimine hõlmab G/C sisalduse kontrolli (eelistatult 40% -65%), esimese ja viimase aluspaari vaatlemist (eelistatult G või C), lühikeste kordusmotiivide vältimist (GGG või CCC rida) ja oligo otstes vabaenergia määramist (GC *clamp* mõlemas otsas).

Kasutaja võib määrata sulamistemperatuuri ka valepaardumistega kontrollproovidele risthübridisatsiooni hindamiseks, kuid see funktsioon pole veebis saadaval. Kasutaja võib välja jätta ka aeganõudva BLAST'i otsingu, võimaldades sooritada korduvaid otsinguid proovide termodünaamiliseks ühtlustamiseks. Järkjärguliseks kvaliteedi parandamiseks saab kasutada erinevaid suvandeid ja piirväärtuseid

ROSO't saab sobivate parameetrite kasutamisel kasutada ka PCR'i proovide valimiseks (>300 aluspaari).

6.7 SNPBOX

SNPBOX (Weckx, jt., 2005) on modulaarne tarkvara pakett, mis võimaldab automatiseeritud PCR praimerite disaini ning on kasutusel kõrgtihedate SNP kaartide koostamisel. Praimereid saab disainida kas avalikustatud SNP'de kontrollimiseks või uute SNP'de avastamiseks resekveneerimisega. Lisaks saab praimerieid disainida ka mutatsioonide analüüsimiseks, STR markerite genotüpiseerimiseks ning oligote disainimiseks mikrokiipide jaoks. SNPBOX kombineerub mitme avalikult saadaoleva programmidega, nagu näiteks BLAST, SPIDEY ja REPEATMASKER.

SNPBOX automatiseerib praimeri disainimise tervenisti määratud genoomsetele järjestustele (edaspidi objektid). Objektid on sihtmärgi määramise alguseks, mille otstesse 70 bp ulatuses disainitakse PRIMER3 (Rozen ja Skaletsky, 2000) abil praimerid. Sihtmärgi pikkus on kasutaja poolt määratav. Kui objekt on lühem kui optimaalne sihtmärgi pikkus, siis pikendatakse seda sümmeetriliselt mõlemalt poolt optimaalse pikkuseni. Kui objekt on pikem kui sihtmärk, siis üleulatused piiritletakse. Väikesed ja ümberkaudsed objektid ühendatakse üheks või rohkemaks sihtmärgiks. Kordusjärjestused võib sõltuvalt nende pikkusest, olemusest ja paigutusest sihtmärkjärjestusse kaasa arvata. Kaasa arvatakse <300 bp pikkused vahelduvate korduste (*interspersed repeats*) hulka kuuluvad kordused. Polümorfsed kordused välistatakse ning praimerieid ei disainita kordusjärjestustele.

Sisendiks on FastA formaadis failid või GenBank gi numbrid, viimase puhul laetakse järjestus otse NCBI'st kasutades EFETCH (http://www.ncbi.nih.gov/entrez/query/static/eutils_help.html) programmi. Kõigepealt maskeeritakse kordusjärjestused, kasutades REPEATMASKER'it (Smit ja Green, http://repeatmasker.org) ning mikrosatelliitide ja ühealuspaariliste korduste leidmiseks kohandatud versiooni SPUTNIK'ust (http://espressosoftware.com).

SNPBOX sisaldab 3 moodulit: SNP moodul, eksoni moodul ja küllastus (*saturation*) moodul. SNP moodul võimaldab praimereid disainida avalike SNP'de kontrollimiseks. SNP'de kaardistamiseks kasutatakse genoomse DNA joondamiseks HGV andmebaasis olevate SNP'dega BLAST programmi. SNP'd valitakse ainult juhul kui maksimaalse E-väärtuse 1E-10 puhul on järjestuse sarnasus ≥95% minimaalselt 40 bp kohta. Genoomsel DNA'l määratakse iga SNP asukoht ning SNP mõlemale poolele 30 bp objektid. 300 bp ulatuses olevad objektid ühendatakse üheks objektiks. Eksonimoodulis määratakse genoomses DNA's olevad kodeerivad järjestused, joondades cDNA ja/või EST järjestused SPIDEY programmi abil. Objekti piiritlemiseks pikendatakse eksoneid mõlemast otsast 50 bp võrra, et ära katta hargmik (*branch point*) ja splaissimissaidid. Kordusjärjestuste leidumisel eksoni lähedal võib pikendada vähem või üldsegi mitte. Objektid 250 bp ulatuses ühendatakse üheks objektiks. Küllastumismoodulis, objektid võivad koosneda regulaatorpiirkondadest, intronitest, tervest geenist või kromosoomi regioonist. Sihtmärgid on vaikimisi määratud 35 bp kattuvustega. Koos praimerite disainimiseks määratud aladega on maksimaalne

amplikoni kattuvus 175 bp (70+70+35). Selles moodulis üritatakse praimerid disainida võimalikult lähedale optimaalsele pikkusele.

SNPBOX'i väljundiks on HTML lehekülg, mis sisaldab tabuleeritud praimeri järjestust, genoomset positsiooni ja PCR'i amplifikatsiooni tingimusi ning G/C sisaldust.

6.8 UNIFRAG ja GENOMEPRIMER

UNIFRAG ja GENOMEPRIMER (van Hijum jt., 2003) on teineteist täiendavad programmid, mis koos võimaldavad leida unikaalseid regioone DNA järjestuses ja automaatselt disainida praimereid minimaalse kasutajapoolse vahelesekkumise ning maksimaalse paindlikkusega. Eelnimetatud meetod on mõeldud bakteri genoomi mikrokiipide tootmiseks ja pole arendatud praimerite disainimiseks introneid sisaldavate geenide jaoks.

UNIFRAG'iga töötamiseks on vaja eelnevalt installitud BLAST programmi. Sisendiks on UNIFRAG'il 3 faili. 1. FastA fail DNA järjestustega, kust valitakse unikaalsed regioonid. 2. referentsidega FastA fail, DNA järjestuste formaatimiseks FORMATDB programmi abil BLAST'i jaoks. Referentsides peab olema vähemalt UNIFRAG'i sisendina kasutatavad DNA järjestused, risthübridisatsioonide vältimiseks võib lisada ka teisi genoomseid järjestusi. 3. kasutaja poolt valitud suvanditega konfiguratsiooni fail.

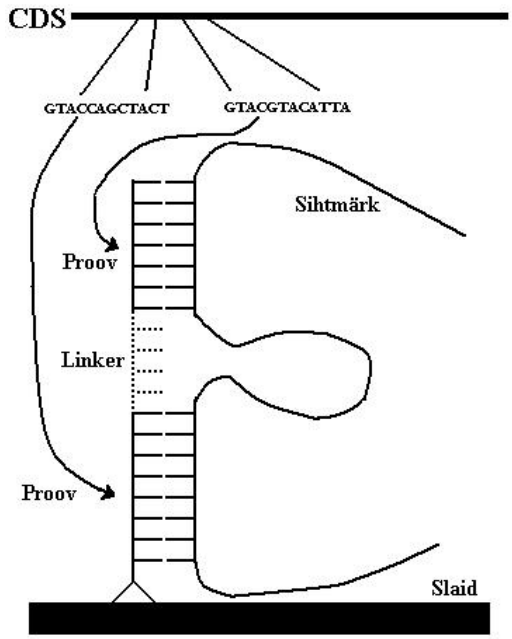
UNIFRAG genereerib kattuvad fragmendid kasutaja poolt määratud pikkusega „akna" (*window*) nihutamisega mööda igat sisendjärjestust. Samuti on kasutaja poolt määratavad kattuvused akende vahel. Minimaalsest fragmendi suurusest lühemad fragmendid salvestatakse „jäänuste" (*leftovers*) listi. Järelejäänud kattuvaid DNA fragmente võrreldakse vastavalt referentsile BLAST programmiga. Filtreeritud BLAST'i väljundist valitakse iga sisend DNA järjestuse jaoks kõrgeima E-väärtusega fragmendid. Kui E-väärtus ületab unikaalsuse piirväärtuse (*cutoff unique*), siis kasutatakse sisend DNA järjestust praimeri disainimiseks, vastupidisel juhul salvestatakse see „jäänuste" listi. Fragmentide ja kattuvuste suurust võib vähendada ja protsessi võib korrata seni, kuni maksimaalne arv unikaalseid fragmente on leitud.

Sobilikud andmed võib vahepeal salvestada FastA formaadis fragmentide tulemuste listi, mida kasutatakse GENOMEPRIMER'iga töötlemisel.

GENOMEPRIMER võimaldab kasutajal määrata praimeri eelistatud asukohta geenil, pikkust ja selle vähendamist, säilitades samal ajal seondumise omadusi, 3' otsa G või C nukleotiidi ja G/C sisaldust ning seejärel valitakse tuhanded praimerid sekunditega. Praimeritel kontrollitakse palindroomsete järjestuste esinemist ja praimeripaaride vahelist homoloogiat. Sulamistemperatuuri arvutamiseks saab valida Suggs'i või *nearest-neighbor* meetodi. Mooduli REPORTGENERATOR poolt raporteeritakse üldiste tulemuste statistika, nagu õnnestumise protsent, keskmine praimeri pikkus, sünteesiks vajaminevate nukleotiidide ning amplikonide arv. Kui praimeri disainimine ebaõnnestub võib sätteid kohandada ja sooritada otsimisi seni, kuni kõik praimerid/amplikonid sobivad soovitud parameetritega.

6.9 GOARRAYS

GOARRAYS (Rimour jt., 2005) on oligo disainimise programm mikrokiipide jaoks. Meetod on suure täpsuse ja kõrge tundlikkusega, võimaldades oligonukleotiidide disainimist peaaegu kõikidele kodeerivatele järjestustele (CDS), välja arvatud täpsetele geeni duplikaatidele. Selle meetodi puhul määratakse CDS'i kohta kaks lühikest (näiteks 25 bp) spetsiifilist järjestust, mille määramine on palju lihtsam kui ühe pikema järjestuse otsimine. Kahe spetsiifilise järjestuse vahele lisatakse juhuslikest aluspaaridest koosnev lühike (3–6 bp) linker järjestus. Sellise komposiitproovi stabiilsel hübridisatsioonil cDNA'ga moodustub silmus (Joonis 10).



Joonis 10. **GOARRAYS komposiitproov**. Kiibile kinnitatud oligonukleotiidproov koosneb kahest mittekülgnevast komplementaarsest alast CDS'iga, mille vahel on juhuslikest nukleotiididest koosnev lühike linker (3-6 bp), mis soodustab silmuse moodustumist stabiilsel hübridisatsioonil (Rimour jt., 2005).

GOARRAYS vajab autonoomset BLAST programmi ja MFOLD tarkvara. Programmil on kaks sisendfaili. Esimene fail peab sisaldama kõiki järjestusi, kuhu oligonukleotiide soovitakse disainida ja teine fail järjestusi, mis võivad esineda hübridisatsiooni ajal sihtmärk mRNA mikstuuris. Kasutaja poolt on määratavad oligonukleotiidi kahe alamjärjestuse pikkused *l*, minimaalne lo_{min} ja maksimaalne lo_{max} silmuse suurus, kahe alamjärjestuse vahel oleva linkeri suurus ja piirväärtus järjestikku komplementaarsete nukleotiidide jaoks mittesihtmärk järjestusega (vaikimisi 15). Lisaks saab kontrollida sulamistemperatuuri paikapidavust, stabiilsete sekundaarstruktuuride olematust jne. Sellisel juhul tuleb määrata soovitud minimaalne ja maksimaalne sulamistemperatuur ning maksimaalne sulamistemperatuur sekundaarstruktuuride jaoks.

Sisendjärjestustest spetsiifiliste järjestuste otsimiseks liigutakse mööda järjestusi 3' otsast *l* pikkuse aknaga. Alamjärjestuste spetsiifilisust kontrollitakse BLAST'i otsinguga. Kui alamjärjestus sisaldab 15 järjestikku komplementaarset nukleotiidi mittesihtmärk järjestusega või kui nende joondus on ≥75% sarnane, siis on see

alamjärjestus mittespetsiifiline. Spetsiifilisele alamjärjestusele ostitakse teine spetsiifiline alamjärjestus, arvestades lo_{min} ja lo_{max}. Kui kaks spetsiifilist alamjärjestust on leitud lisatakse nende vahele juhuslikult valitud nukleotiididest lühike linker. Järgnevalt kontrollitakse, kas lisatud linkerjärjestus ei või põhjustada täiendavat risthübridisatsiooni.

Vajadusel kontrollitakse ka oligonukleotiidide sulamistemperatuure, kasutades *nearest-neighbor* meetodit. Sekundaarstruktuuride esinemist oligonukleotiididel saab vajadusel kontrollida MFOLD programmi abil, mille olemasolul oligonukleotiidid kõrvaldatakse. Lõpuks on võimalus kontrollida ka kasutaja poolt määratud keelatud järjestuste olemasolu (näiteks korduvad aluspaarid). Kui järjestus on kõrvaldatud, siis liigub alamjärjestuse aken mööda järjestust edasi, seni kuni leitakse korrektne oligonukleotiid.

7 Probleemid oligote disainimisel

Peamiseks probleemiks oligote disainimisel on genoomis leiduvad kordusjärjestused ja homoloogsed geenid, mis soodustavad oligote risthübridisatsiooni. Samuti on suurte genoomide puhul suurem tõenäosus sekundaadersete seondumiskohtade leidumiseks. Risthübridisatsioon võib põhjustada PCR'i puhul soovimatute produktide teket, nõrgendada õige produkti amplifitseerimist või pärssida täielikult selle amplifitseerimist. Mikrokiipide puhul annab risthübridisatsioon valesignaale, mis võib viia bioloogiliste tulemuste valesti mõistmisele (Rimour jt., 2005).

Lühikeste oligonukleotiidide mikrokiipide puhul on risthübridisatsioon väga suureks probleemiks. Vale seondumist põhjustavad peamiselt prooviga komplementaarsed olevad 10-16 nukleotiidilised alad. Valepaardumisi hinnatakse energeetiliselt palju tähtsamaks kui geenispetsiifilist risthübridisatsiooni, mis tähendab et risthübridisatsiooni mõju on erinevatel kattuvatel ja valepaardumistega proovipaaridel erinev. Valepaardumiste vabaenergia sõltub samuti ka naabruses olevatest aluspaaridest (Wu jt., 2005).

Vaata veel peatükk 4, 5.2, 5.3, 5.4 ja 5.6

8 Alternatiivsed meetodid nende probleemide kõrvaldamiseks

On arendatud mitmeid programme, mis võimaldavad kontrollida praimerite või praimeripaaride seondumisi kogu genoomi vastu. Nende algoritmidega on võimalik ennustada praimerite sekundaarseid seondumiskohti ja samuti ka võimalikke PCR'i produkte. Olemasolevate meetoditega on võimalik sooritada nii perfektse kattuvusega kui ka valepaardumiste ja aukudega joondamisi. Järgnevalt vaatleme nelja sellist programmi, mis on muidugi väike osa olemasolevatest meetoditest.

8.1 GAPPED BLAST

BLAST (*Basic Local Alignment Search Tool*) programmid on laialt kasutusel valkude ja DNA andmebaasides järjestuste sarnasuse otsimiseks ja neid on liidetud paljudesse programmidesse. Originaalne BLAST (Altschul jt., 1990) kasutab ühetabamusega (*one-hit*) meetodit, otsides lühikesi sõnapaare skooriga vähemalt T ning pikendades igat sellist tabamust. Pikendamise samm on kõige aeganõudvam etapp BLAST'i otsingus, tavaliselt >90% BLAST'i teostamise ajast. Protsessi kiirendamiseks tuleb vähendada pikendatavate tabamuste arvu.

GAPPED BLAST (Altschul jt., 1997) kasutab kahetabamusega (*two-hit*) meetodit ning võimaldab ka aukudega joondamist (*gapped alignment*) ning töötab originaalsest BLAST'ist umbes kolm korda kiiremini. Selle meetodi puhul skaneeritakse andmebaasi ja otsitakse sõnu pikkusega W, mille joondamisel päringus olevate sõnadega saadakse skoor vähemalt T. Iga sõnapaar, mis rahuldab seda tingimust nimetatakse tabamuseks. Järgnevalt kontrollitakse, kas tabamused on piisava skooriga, et neid raporteerida. Selleks tuleb pikendada igat tabamust mõlemas suunas, kuni joonduse skoor ei lange rohkem kui X alla seni saadud parimast skoorist. Aukudega joondamisel kasutatakse mõlema suunaliseks pikendamiseks dünaamilist programmeerimist (Smith ja Waterman, 1981). Valitakse aken pikkusega A ja pikendamine käivitatakse ainult juhul, kui samal diagonaalil leidub kaks mittekattuvat tabamust distantsis A. Teisisõnu, pikendamine käivitatakse ainult juhul, kui ühel diagonaalil leidub kaks mittekattuvat sõnapaari distantsis A. Tundlikkuse säilitamiseks kasutatakse madalamat T piirväärtust,

mis toob kaasa rohkem tabamusi, kuid ainult väike osa nendest on seotud teise tabamusega samal diagonaalil ning kuuluvad pikendamisele. Lisaks heidetakse kõrvale tabamused, mille viimase tabamuse koordinaat jääb distantsi A hetketabamusega. Selleks kasutatakse maatriksit, kuhu salvestatakse iga viimase tabamuse esimene koordinaat, mis asendatakse igal järgneval tabamusel uue koordinaadiga. Aukudega joondus raporteeritakse ainult juhul, kui E-väärtus on piisavalt madal. E-väärtus (*expected value, e-value*) näitab kui palju antud skooriga järjestusi võib antud suurusega andmebaasis juhuslikult leiduda. Suuremate andmebaaside puhul on E-väärtus suurem.

Skoori arvutamiseks kasutatakse valemit:

$$S' = \frac{\lambda S - \ln K}{\ln 2},$$

kus λ ja K on statistilised parameetreid ning S′ esialgne skoori. Ühik on bittides.

E-väärtuse arvutamiseks on valem:

$$E = \frac{N}{2^{S'}},$$

kus N=mn, m on andmebaasi järjestuse pikkus ja n on päringu pikkust.

8.2 PRIMEX

PRIMEX (*PRimer Mach EXtractor*) (Lexa ja Valle, 2003) on programm, mis on võimeline leidma oligonukleotiidide järjestusi kogu genoomi ulatuses, võimaldades ka valepaardumisi. Seda programmi võib kasutada oligonukleotiidide proovi valimisel hübridisatsiooni eksperimentideks või PCR praimerite disainiks genoomsetelt DNA järjestustelt. Sõna otsingu tabeli (*lookup table*) kasutamine ning klient-server

funktsionaalsus võimaldavad kliendi päringule tagastada kattuvuse tulemused väga kiiresti.

Kasutatakse määratud suurusega akent, mis paigutatakse otsitava genoomi alguspunkti. Akent nihutatakse mööda järjestust ja iga aknasuurune sõna salvestatakse otsingu tabelisse. Otsingu tabel on maatriksiks, mille viited esindavad akna suuruseid sõnu. Otsingu tabel salvestatakse edasiseks kasutamiseks. Pärast päringu vastuvõtmist ekstraheeritakse sõnad päringu järjestusest ja otsitakse kattuvusi, kasutades otsingu tabelit ning arvestades lubatud valepaardumiste arvu. Saadud kattuvused filtreeritakse vastavalt lubatud valepaardumistele, kõrvaldatakse duplikaadid ja tulemused kirjutatakse välja.

Päringu tulemused sisaldavad praimeri numbrit, päringu järjestust, kattuvaid järjestusi, klooni/genoomi nime, positsiooni kloonis/genoomis, orientatsiooni ning kattuvate aluspaaride arvu.

8.3 SSAHA

SSAHA (*Sequence Search and Alignment by Hashing Algorithm*) (Ning jt., 2001) on kiire DNA järjestuste otsimise algoritm suurtest DNA andmebaasidest, mis võib olla kolm kuni neli korda kiirem kui BLAST. Programmi alternatiivseteks kasutusaladeks on ühenukleotiidiliste polümorfismide detekteerimine ja suurte *shotgun* järjestuste osade kokkupanemine.

DNA järjestuste otsimiseks eelprotsessitakse kõigepealt andmebaasi järjestused, jagades need k külgnevast aluspaarist koosnevatesse *k-tuple*'tesse ning salvestatakse paisktabelisse (*hash table*). Paisktabel salvestatakse mällu, mis võimaldab sooritada suurtest andmebaasidest (inimese genoom või veelgi suuremad) väga kiireid otsinguid. Päringjärjestuse otsimiseks paisktabelist kasutatakse päringjärjestuse *k-tuple*'de kattuvusi andmebaasi *k-tuple*'dega ning tabamused sorteeritakse tulemusse.

Järjestuse otsimisel saab otsida täpset või osalist päringjärjestuse Q esinemist andmebaasi $D = \{S_1, S_2, \ldots, S_d\}$ alamjärjestustes. Iga alamjärjestus S on tähistatud täisarvulise indeksiga i. *k-tuple* tähistab k aluspaari pikkust DNA järjestust. S on n aluspaari pikkune DNA järjestus, mis sisaldab $(n - k + 1)$ kattuvat *k-tuple*'t. Lähe

(*offset*) tähistab *k-tuple* esimese nukleotiidi positsiooni *S* järjestusel. $w_j(S)$ tähistab *k-tuple*'t, mille lähe on *j*. Seega *k-tuple* positsioon andmebaasis on tähistatud (*i*; *j*).

Andmebaasi *D* järjestus konverteeritakse paisktabelisse, mis sisaldab 4^k *pointer*'i jada ja nende positsioonide loetelu, ning salvestatakse mällu. Paisktabel koostatakse kahe töötiiruga läbi andmete, arvestades ainult mittekattuvaid *k-tuple*'sid. Esimesel töötiirul loendatakse kõigi võimalike 4^k *k-tuple* esinemised, ning määratakse nende *pointer*'ite positsioonid. Järgnevalt võib ignoreerida sõnasid, mille esinemissagedus ületab piirväärtuse *N*, vähendamaks paisktabeli suurust ja sellega omakorda mälu vajadust. See on kasulik, kui paisktabel luuakse ühekordseks otsinguks.

Näide järjestuse otsimise kohta paisktabelist. Olgu otsitavaks järjestuseks *Q* = TGCAACAT, *k* olgu 2 ning andmebaasis *D* olgu 3 järjestust:

S_1 = GTGACGTCACTCTGAGGATCCCCTGGGTGTGG,

S_2 = GTCAAC**TGCAACAT**GAGGAACATCGACAGGCCCAAGGTCTTCCT,

S_3 = GGATCCCCTGTCCTCTCTGTCACATA.

Esmalt tehakse andmebaasi *D* järjestustest paisktabel (Tabel 4).

Tabel 4. Andmebaasi *D* 2-*tuple* paisktabel.

w	**Pointer**	**Positsioon (*i*, *j*)**						
AA	0	(2, 19)						
AC	1	(1, 9)	(2, 5)	(2, 11)				
AG	2	(1, 15)	(2, 35)					
AT	3	(2, 13)	(3, 3)					
CA	4	(2, 3)	(2, 9)	(2, 21)	(2, 27)	(2, 33)	(3, 21)	(3, 23)
CC	5	(1, 21)	(2, 31)	(3, 5)	(3, 7)			
CG	6	(1, 5)						
CT	7	(1, 23)	(2, 39)	(2, 43)	(3, 13)	(3, 15)	(3, 17)	
GA	8	(1, 3)	(1, 17)	(2, 15)	(2, 25)			
GC	9							
GG	10	(1, 25)	(1, 31)	(2, 17)	(2, 29)	(3, 1)		
GT	11	(1, 1)	(1, 27)	(1, 29)	(2, 1)	(2, 37)	(3, 19)	
TA	12	(3, 25)						
TC	12	(1, 7)	(1, 11)	(1, 19)	(2, 23)	(2, 41)	(3, 11)	
TG	14	(1, 13)	(2, 7)	(3, 9)				
TT	15							

Tabelis on 4^2 pointerit, iga võimaliku 2-*tuple* jaoks. Positsioonid tähistavad kõigi võimaliku 2-*tuple* (4^k = 16) positsioone andmebaasis *D*. Sulgudes esimene liige (*i*) tähistab alamjärjestuse indeksit ning teine liige (*j*) tähistab 2-*tuple* esimest nukleotiidi alamjärjestusel.

Päringu otsimiseks liigutakse esmalt mööda otsitavat järjestust, *k* pikkuse sõnaga, ühe aluspaari kaupa 0 aluspaarist kuni aluspaarini *n* - *k*, kus *n* tähistab *Q* pikkust. Aluspaari *t* juures saadakse list *r*, mis sisaldab päringu *k-tuple* leidumise positsioone *D*'s [(i_1, j_1), (i_2, j_2), …, (i_r, j_r)], millest arvutatakse tabamuste list [$H_1 = (i_1, j_1 - t, j_1)$, $H_2 = (i_2, j_2 – t, j_2)$, …, $H_r = (i_r, j_r - t, j_r)$]. Tabamuste list lisatakse *master* listi *M*, mis kujutab endast sorteeritud *H* listi. *Master* list on tähistatud (indeks, nihe (*shift*), lähe). *Master* list sorteeritakse järjekorras indeks, nihe ja siis lähe (Tabel 5). Nihke muutmisel mõne üksiku aluspaari võrra on võimalik lubada ka deletsioonide ja insertsioonide arvestamist. Lähte järgi sorteerides saab teada täpsete kattuvuste vahelised regioonid ning piisavalt lähedal asuvate regioonide ühendamisel saab tekitada aukudega kattuvused. Selliselt saab leida kattuvused edaspidises suunas. Tagurpidi suunas kattuvuste leidmiseks tuleb võtta pööratud *Q* järjestus ja seejärel protseduuri korrata.

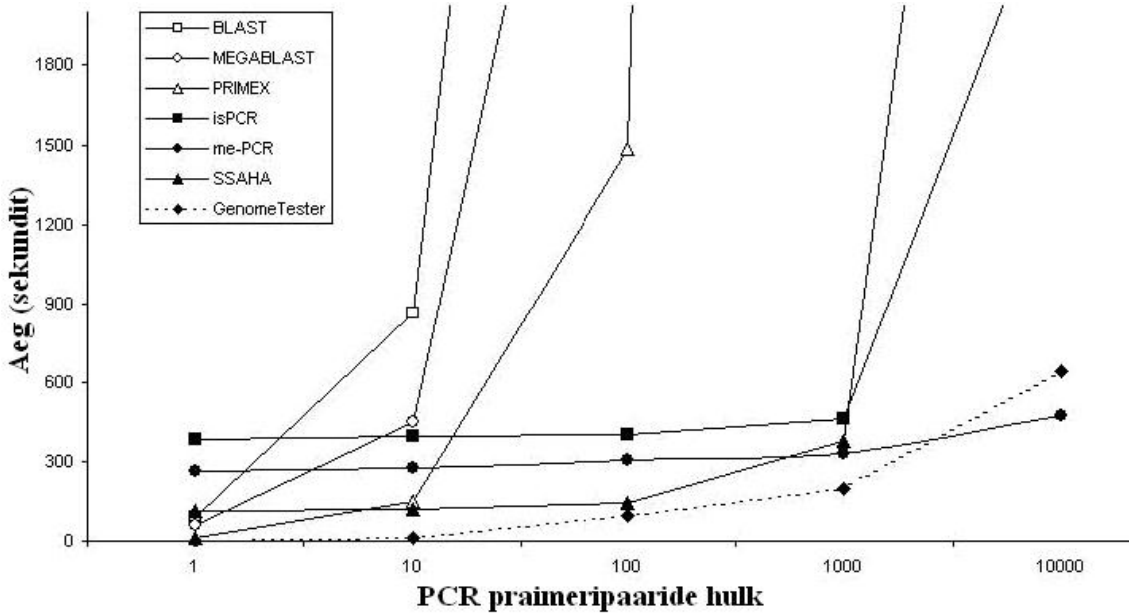
Tabel 5. Kattuvuste tabel.

t	***w*$_t$*(Q)***	**Positsioon**	***H***	***M***
0	TG	(1, 13)	(1, 13, 13)	(1, 5, 9)
		(2, 7)	(2, 7, 7)	(1, 13, 13)
		(3, 9)	(3, 9, 9)	(2, -2, 3)
1	GC			(2, 1, 3)
2	CA	(2, 3)	(2, 1, 3)	(2, 1, 5)
		(2, 9)	(2, 7, 9)	(2, 4, 9)
		(2, 21)	(2, 19, 21)	**(2, 7, 7)**
		(2, 27)	(2, 25, 27)	**(2, 7, 9)**
		(2, 33)	(2, 31, 33)	**(2, 7, 11)**
		(3, 21)	(3, 19, 21)	**(2, 7, 13)**
		(3, 23)	(3, 21, 23)	(2, 16, 19)
3	AA	(2, 19)	(2, 16, 19)	(2, 16, 21)
4	AC	(1, 9)	(1, 5, 9)	(2, 19, 21)
		(2, 5)	(2, 1, 5)	(2, 22, 27)
		(2, 11)	(2, 7, 11)	(2, 25, 27)
5	CA	(2, 3)	(2, -2, 3)	(2, 28, 33)
		(2, 9)	(2, 4, 9)	(2, 31, 33)
		(2, 21)	(2, 16, 21)	(3, -3, 3)
		(2, 27)	(2, 22, 27)	(3, 9, 9)
		(2, 33)	(2, 28, 33)	(3, 16, 21)
		(3, 21)	(3, 16, 21)	(3, 18, 23)
		(3, 23)	(3, 18, 23)	(3, 19, 21)
6	AT	(2, 13)	(2, 7, 13)	(3, 21, 23)
		(3, 3)	(3, -3, 3)	

t tähistab päringjärjestuse 2-*tuple* esimese nukleotiidi positsiooni. $w_t(Q)$ on päringu 2-*tuple* positsioonis *t*. Positsioon näitab päringu 2-*tuple* esinemise positsioone andmebaasis *D*. *H* tulbas on päringu tabamused. *M* on sorteeritud tabamuste list, kust on võimalik lugeda järjestikuseid tabamusi (tumedalt). *M* listis sulgudes esimene liige tähistab alamjärjestuse indeksit andmebaasis, teine liige nihet ning kolmas päringu 2-*tuple* esimest positsiooni andmebaasi alamjärjestusel. Kattuvus on päringu *Q* ja alamjärjestuse S_2 vahel, mis algab S_2 järjestuse seitsmendast positsioonist ja kulgeb edasi 8 aluspaari võrra.

8.4 GENOMETESTER

GENOMETESTER (Andreson jt., submitted) on programm oligo seondumiskohtade ja PCR produktide ennustamiseks. Programmis kasutatakse kolme alamprogrammi. (i) *gindexer*'i sisend failiks on FastA formaadis genoomi või kromosoomi järjestus, millest tehakse 4 erinevat indeksfaili kõigi *n*-meersete järjestuste jaoks – üks järjestustele, mis algavad A'ga, teine C'ga, kolmas G'ga ja neljas T nukleotiidiga algavate järjestuste jaoks. See on selleks, et vähendada *gtester*'i mälu nõudlust. Need binaarsed indeksfailid sisaldavad kõiki võimalikke seondumiskohti antud järjestuse failis. (ii) *gtester*'i abil luuakse praimeripaaridest järjend, mis koosneb etteantud pikkusega *n*-meersetest praimeri 3' otsadest, nende originaal- ning pöördkomplementidest. Järgnevalt sooritatakse binaarne otsing kõikide maatriksis olevate praimeritega indeksfailide vastu. Siis luuakse loetelu kõikide seondumiskohtade informatsioonidega ja liikudes üle praimeripaari saitide määratakse võimalikud PCR'i produktid. Kõige lõpuks leitakse kõik võimalikud produktid, mida üks praimer võib anda hübridisatsiooniga nii *sense* kui *antisense* DNA ahelale. (iii) *gt2multiplex* ekstraheerib *gtester*'i väljundi informatsiooni põhjal sihtmärkjärjestusest kõikide PCR'i produktide järjestused, mida saab kasutada edasistel analüüsidel. Praimeri seondumiskohti saab modelleerida erinevate sõna pikkustega (8 - 16 bp). GENOMETESTER on tunduvalt kiirem kui teised hetkel saadaolevad programmid (Joonis 11).



Joonis 11. GENOMETESTER'i jõudluse võrdlus erinevate meetoditega. Katses kasutati 5 erinevat andmehulka, mis sisaldas 1, 10, 100, 1000 ja 10000 juhuslikult valitud praimeripaari.

Arutelu

Antud töö eesmärgiks oli uurida kirjanduse põhjal DNA mikrokiipide olemust ning nende rakendusi, APEX meetodit, oligo kvaliteeti määravaid parameetreid ning levinumaid oligo disainimise ja kvaliteedi kontrollimise algoritme ja meetodeid, et selle põhjal edaspidi arendada programm, mis on mõeldud APEX oligote kvaliteeti mõjutavate parameetrite uurimiseks.

Oligote disainimiseks ja seostumiskohtade leidmiseks on olemas väga palju erinevaid parameetreid kasutavaid algoritme ja meetodeid. Enamus oligo disaini programme võimaldavad efektiivselt vältida oligote omavahelisi interaktsioone, kuid ei paku seda võimalust genoomse DNA interaktsioonide jaoks. On ka mitmeid algoritme oligote alternatiivsete seondumiskohtade leidmiseks. Varem arvati, et põhiline tegur risthübridisatsiooni põhjustamisel on komplementaarsus, kuid nüüd on hakatud üha enam mõistma ka termodünaamika olulisust. Kuid ikkagi pole veel ühtegi eriti tõhusat meetodit, mis võimaldaks väga kiiresti leida paljusid olulisi parameetreeid suurte genoomsete järjestuste vastu. Praeguseks on sekveneeritud väga palju terveid genoome, mis pakub võimalusi ja tekitab vajadust selliste meetodite väljatöötamiseks.

Praeguseks pole veel täpselt teada, millised parameetrid mõjutavad APEX reaktsiooni kvaliteeti. Et seda mõista, oleks vaja valmistada programm, mille abil saaks läbi viia täpsemat uuringut, et analüüsida APEX oligote olulisi parameetreid. Nendeks parameetriteks peaksid olema:

1. APEX oligote seostumiskohtade leidmine genoomis
   a. 100% identsete järjestuste leidumine genoomis oligo 3’ otsast 8, 10, 12, 14, 16, 18 nukleotiidi ja kogu oligo ulatuses
   b. 100% identsete järjestuste leidmine oligo 3’ otsast, mille ΔG_{37} <-10, <-15, <-20, <-25 ja <-30 kcal/mol
   c. 1 valepaardumise lubamisel punkti a. ja b. parameetrites
   d. 2 valepaardumise lubamisel punkti a. ja b. parameetrites } (va 3’ otsa nukleotiid)
   e. 1 augu (*gap*)lubamisel punkti a. ja b. parameetrites
2. APEX oligo omadused
   a. G/C sisaldus

b. 3’ otsa stabiilsus ΔG_{37} põhjal (eraldi 3-9 viimase nukleotiidi jaoks)
c. Lihtsate korduste olemasolu (näiteks korduvad nukleotiidid)
d. 3’ otsa viimase nukleotiidi mõju (4 varianti)
e. 3’ otsa 2 viimase nukleotiidi mõju (16 varianti)
f. Sekundaarstruktuuride mõju

Vajalikeks sisendfailideks peaks olema kaks faili. Üks eraldi tulpadena tabuleeritud teksti fail APEX oligote andmetega, mis sisaldab APEX oligo ID numbrit ja vasakpoolse ning parempoolse oligo järjestust. Teine fail peab olema genoomse järjestusega FastA formaadis fail. Väljundfail sisaldaks iga APEX oligo kohta kõiki eelnevalt nimetatud parameetreid.

Identsete, valepaardumiste ja aukudega seondumiskohtade leidmiseks võiks kasutada BLAST programmi otsingut. ΔG_{37} jaoks võiks hakata kasutama hetkel veel väljatöötamisel olevat programmi GENOMETESTER2, mis määrab seondumissaidid termodünaamiliselt dimeeride vabaenergia (ΔG) kaudu. Programm kasutab vabaenergia arvutamiseks *nearest-neigbor* meetodit. Seondumissaidid salvestatakse binaarsel kujul paisktabelisse, kust GENOMETESTER2 on võimeline kiiresti leidma kõik potentsiaalsed seondumiskohad genoomis. Sekundaarstruktuuride määramiseks oleks sobiv kasutada MFOLD ning lihtsate korduste otsimiseks DUST programmi.

Parameetrite kontrolliks, lisaprogrammide käivitamiseks ja nende tulemuste interpreteerimiseks on plaanis kirjutada programmide pakett PERL’is. Valmistatava programmide paketi abil oleks võimalik reaalsetele eksperimentaalsetele andmetele põhinedes leida APEX reaktsiooni kvaliteeti mõjutavaid parameetreid.

Kokkuvõte

Käesolevas töös anti ülevaade DNA mikrokiipidest, nende kasutusaladest, APEX meetodist, oligo kvaliteeti mõjutavatest parameetritest ja levinuimatest oligo disaini ja kvaliteedi kontrollimise programmidest.

Mikrokiibid võimaldavad analüüsida komplekssseid biokeemilisi proove paralleelselt ja uurijatel läbi viia suuremastaabilisi kvantitatiivseid eksperimente. Rakendusi, kus mikrokiipe kasutada, on mitmeid ja paljud nendest rakendustest sõltuvad nukleiinhapete amplifitseerimisest PCR'i abil. PCR'i tegemiseks on vaja DNA'le kinnituvaid spetsiifilisi oligosid või oligo paare. Sama probleem on ka oligote disainimisel mikrokiipide jaoks. Väga tähtis on, et oligod ei annaks soovimatuid risthübridisatsioone s.t. moodustaksid stabiilse dupleksi ainult spetsiifilise saidiga huvipakkuval sihtmärk DNA'l ja et kõikidel oligodel oleks minimaalne sulamistemperatuuri varieeruvus.

Oligo disainimiseks on avalikult saadaval väga palju erinevaid programme. Enamus nendest kasutab oligo disainimiseks sarnast algoritmi või kriteeriumeid. Ükski olemasolevates programmidest ei arvesta kõiki olulisi kriteeriumeid oligo valimisel. On arendatud ka programme, mis võimaldavad kontrollida oligo või oligo paaride seondumisi kogu genoomi vastu.

Täpsema uuringu läbiviimiseks, mis leiaks olulisi seoseid oligo disaini parameetrite ja APEX reaktsiooni edukuse vahel, on järgmiseks tööetapiks planeeritud vastava programmi kirjutamine.

Resümee

DNA microarrays represent a glass surface bearing thousands of DNA fragments (cDNA microarray) at discrete sites or oligonucleotides (oligonucleotide microarray) which are available for hybridization. Hybridization of fluorescently labeled RNA and DNA-derived samples to DNA chips have enabled biology researchers to conduct large scale quantitative experiments. Microarrays have been used to identify novel genes, binding sites of transcription factors, changes in DNA copy number, and variations from a baseline sequence, such as in emerging strains of pathogens or complex mutations in disease-causing human genes. Most of these methods require PCR amplification to achieve sufficiently strong signals. Therefore, there is a growing need for automatic oligo design and PCR primer design methods.

cDNA microarrays containing cDNAs amplified by the polymerase chain reaction (PCR) ranging 0,5 to 2,0 kb. The efficient selection of genespecific primers is of the utmost importance for the successful production of such chips. In the short oligo (20-25-mer) microarray technology is one of the critical problems howto deal with cross-hybridization that produces spurious data. Little is known about the details of cross-hybridization effect at molecular level.

Although there are many publicly available programs for microarray oligonucleotide design, most of them use the same algorithm or criteria to design oligos, with only little variation. They primarily differ by the criteria chosen to select the oligonucleotides: some will take into account the possibility of an oligonucleotide having a stable secondary structure under certain conditions, others will make it possible to exclude certain sequence patterns determined by the user, such as the repetition of a single base. None of them gathers the whole set of criteria.

APEX (Arrayed Primer EXtension) is an integrated system with DNA chip and template preparation, multiplex primer extension on the array, fluorescence imaging, and data analysis. It is not yet fully known which parameters affect the success of APEX reaction. In order to gain better understanding a computer program should be created to analyse those parameters more thoroughly. The parameters to investigate should be:

1. The search for APEX oligo binding sites in genomes

a. The presence of sequences that are 100% identical to oligo 3' terminus to the extent of 8, 10, 12, 14, 16, 18 nucleotides or the whole oligo

b. The presence of sequences that are 100% identical to the oligo 3' terminus and the ΔG_{37} of which is <-10, <-15, <-20, <-25, <-30 kcal/mol

c. Allowing a single mismatch to the parameters of a. and b. (excluding the 3' terminal nucleotide)

d. Allowing two mismatches to the parameters of a. and b. (excluding the 3' terminal nucleotide)

e. Allowing a gap in the parameters of a. and b. (excluding the 3' terminal nucleotide)

2. Properties of the APEX oligos

a. G/C content

b. stability of the 3' terminus, based on ΔG_{37} (separately for the last 3-9 nucleotides)

c. Presence of simple repeats (e.g. repeating nucleotides) (DUST)

d. The effect of 3' terminal nucleotide (4 possibilities)

e. The effect of the last two nucleotides of the 3' terminus (16 possibilities)

f. The effect of secondary structures (MFOLD)

Kasutatud kirjandus

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment search tool. J Mol Biol. 215 (3): 403-410.

2. Altschul, S. F., Madden, T. L., Schaffer, A, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17): 3389-3402.

3. Andreson, R., Reppo, E., Kaplinski, L., Möls, M., Remm, M. Submitted. GenomeMasker and GenomeTester: tools for design of high quality genomic PCR primers.

4. Benita, Y., Oosting, R. S., Lok, M. C., Wise, M. J., Humphery-Smith, I. 2003. Regionalized GC content of template DNA as a predictor of PCR success. Nucleic Acids Res. 31 (16): e99.

5. Breslauer, K. J., Frank, R., Blocker, H., Marky, L. A. 1986. Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A. (11): 3746-3750.

6. Chen, Y., Kamat, V., Dougherty, E. R., Bittner, M. L., Meltzer, P. S., Trent, J. M. 2002. Ratio statistics of gene expression levels and applications to microarray data analysis. Bioinformatics. 18 (9): 1207-1215.

7. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., Trent, J. M. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer.at Genet. 14 (4): 457-460.

8. DeRisi,. J. L., Iyer, V. R., Brown, P. O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 278 (5338): 680-686.

9. Fan, J. B., Oliphant, A., Shen, R., Kermani, B. G., Garcia, F., Gunderson, K. L., Hansen, M., Steemers, F., Butler, S. L, Deloukas, P. jt. 2003. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol. 68: 69-78.

10. Gerhold, D., Rushmore, T., Caskey, C. T. 1999. DNA chips: promising toys have become powerful tools. Trends Biochem Sci. 24 (5):168-173.

11. Haas, S. A., Hild, M., Wright, A. P., Hain, T., Talibi, D., Vingron, M. 2003. Genome-scale design of PCR primers and long oligomers for DNA microarrays. Nucleic Acids Res. 31 (19): 5576-5581.

12. Haas, S., Vingron, M., Poustka, A., Wiemann, S. 1998. Primer design for large scale sequencing. Nucleic Acids Res. 26 (12): 3006-3012.

13. Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D. E., Davis, R. W. 1997. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci U S A. 94 (6): 2150-2155.

14. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T. jt. 2002. The Ensembl genome database project. Nucleic Acids Res. 30 (1): 38-41.

15. Hughes, T. R., Mao, M., Jones, A. R., Burchard, J., Marton, M. J., Shannon, K. W., Lefkowitz, S. M., Ziman, M., Schelter, J. M., Meyer, M. R. jt. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nature Biotechnology. 19 (4): 342-347.

16. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J. jt. 2000**.** Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet. 25 (3): 333-337.

17. Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., Trent, J. M., Staudt, L. M., Hudson, J. Jr., Boguski, M. S. jt. 1999. The transcriptional program in the response of human fibroblasts to serum. Science. 283 (5398): 83-87.

18. Kane, M. D., Jatkoe, T. A., Stumpf, C. R., Lu, J., Thomas, J. D., Madore, S. J. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res. 28 (22): 4552-4557.

19. Kurg, A., Tõnisson, N., Georgiou, I., Shumaker, J., Tollett, J. and Metspalu, A. 2000. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. Genet Test. 4 (1): 1−7

20. Le Novere, N. 2001. MELTING, computing the melting temperature of nucleic acid duplex. Bioinformatics. 17 (12): 1226-1227.

21. Lee, I., Dombkowski, A. A., Athey, B. D. 2004. Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray. Nucleic Acids Res. 32 (2): 681-690.

22. Lexa, M., Valle, G. 2003. PRIMEX: rapid identification of oligonucleotide matches in whole genomes. Bioinformatics. 19 (18): 2486-2488.

23. Li, F., Stormo, G. D. 2001. Selection of optimal DNA oligos for gene expression arrays. Bioinformatics. 17 (11): 1067-1076.

24. Li, P., Kupfer, K. C., Davies, C. J., Burbee, D., Evans, G. A., Garner, H. R. 1997. PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing. Genomics. 40 (3): 476-485.

25. Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton. H., Brown, E. L.

1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol. 14 (13): 1675-1680.

26. Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J. jt. 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res. 13 (10): 2291-2305.

27. Marshall, O. J. 2004. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics. 20 (15): 2471-2472.

28. Marton, M. J., DeRisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H. jt. 1998. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med. 4 (11): 1293-1301.

29. Matveeva, O. V., Shabalina, S. A., Nemtsov, V. A., Tsodikov, A. D., Gesteland, R. F., Atkins, J. F. 2003. Thermodynamic calculations and statistical correlations for oligo-probes design. Nucleic Acids Res. 31 (14): 4211-4217.

30. Mrowka, R., Schuchhardt, J., Gille, C. 2002. Oligodb--interactive design of oligo DNA for transcription profiling of human genes. Bioinformatics. 18 (12): 1686-1687.

31. Ning, Z., Cox, A. J., Mullikin, J. C. 2001. SSAHA: a fast search method for large DNA databases. Genome Res.11 (10): 1725-1729.

32. Pease, A. C., Solas, D., Sullivan, E. J, Cronin, M. T., Holmes, C. P., Fodor, S. P. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A. 91 (11): 5022-5026.

33. Raddatz, G., Dehio, M., Meyer, T. F., Dehio, C. 2001. PrimeArray: genome-scale primer design for DNA-microarray construction. Bioinformatics. 17 (1): 98-99.

34. Relogio, A., Schwager, C., Richter, A., Ansorge, W., Valcarcel, J. 2002. Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Res. 30 (11): e51.

35. Reymond, N., Charles, H., Duret, L., Calevro, F., Beslon, G., Fayard, J. M. 2004. ROSO: optimizing oligonucleotide probes for microarrays. Bioinformatics. 20 (2): 271-273.

36. Rimour, S., Hill, D., Militon, C., Peyret, P. 2005. GoArrays: highly dynamic and efficient microarray probe design. Bioinformatics. 21 (7): 1094-1103.

37. Rozen, S., Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132: 365-386.

38. Rouillard, J. M., Zuker, M., Gulari, E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res. 31 (12): 3057-3062.

39. Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L. jt. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature.409 (6822): 928-933.

40. SantaLucia, J. Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A. 95 (4): 1460-1465.

41. Schena, M, Shalon, D., Davis, R. W., Brown, P. O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 270 (5235): 467–470.

42. Schena, M. 2003. Preface, p. ix-xiv, 4-9. Microarray Analysis. Wiley-Liss, United States of America. John Wiley & Sons, Inc.

43. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., Davis, R. W. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci U S A. 93 (20): 10614-10619.

44. Smith, T. F., Waterman, M. S. 1981. Identification of common molecular subsequences. J Mol Biol. 147 (1): 195-197.

45. Stears, R. L., Martinsky, T., Schena, M. 2003. Trends in microarray analysis. Nat Med. 9 (1): 140-145.

46. Stoughton, R. B. 2004. Applications of DNA Microarrays in Biology. Annu Rev Biochem.

47. Suggs, S. jt. 1981. ICN-UCLA Symposia on Developmental Biology Using Purified Genes. Academic Press Inc., New York, NY, Vol 23, pp683-693.

48. Zuker, M., Mathews, D. H. and Turner, D. H. 1999. Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical. Guide, NATO ASI Series. Kluwer Academic Publishers, Dordrecht, NL.

49. Tatusov, R. L, and Lipman, D. J. unpublished, DUST.

50. van Hijum, S. A., de Jong, A., Buist, G., Kok, J., Kuipers, O. P. 2003. UniFrag and GenomePrimer: selection of primers for genome-wide production of unique amplicons. Bioinformatics. 19 (12): 1580-1582.

51. Weckx, S., De Rijk, P., Van Broeckhoven, C., Del-Favero, J. 2005. SNPbox: a modular software package for large-scale primer design. Bioinformatics. 21 (3): 385-387.

52. Weiner, P. 1973 In Proceedings of the IEEE 14th Annual Symposium on Switching and Automata Theory, Linear Pattern Matching Algorithms. IEEE, New York, pp. 1–11.

53. Welford, S. M., Gregg, J., Chen, E., Garrison, D., Sorensen, P. H., Denny, C. T., Nelson, S. F. 1998. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. Nucleic Acids Res. 26 (12): 3059-3065.

54. Venkatasubbarao, S. 2004. Microarrays – status and prospects. Trends Biotechnol. 22 (12): 630-637.

55. Wheelan, S. J., Church, D. M., Ostell, J. M. 2001. Spidey: a tool for mRNA-to-genomic alignments. Genome Res. 11 (11): 1952-1957.

56. Wu, C., Carta, R., Zhang, L 2005. Sequence dependence of cross-hybridization on short oligo microarrays. Nucleic Acids Res. 33 (9): e84.

57. Xiang, C. C., Chen, Y. 2000. cDNA microarray technology and its applications. Biotechnol Adv. 18 (1): 35-46.

58. Ye, R. W., Wang, T., Bedzyk, L., Croker, K. M. 2001. Applications of DNA microarrays in microbial systems. J Microbiol Methods. 47 (3): 257-272.

Kasutatud veebiaadressid

Affymetrix:

http://www.affymetrix.com/index.affx

http://www.affymetrix.com/technology/manufacturing/index.affx

http://www.affymetrix.com/technology/index.affx

Asper Biotech:

http://www.asperbio.com/APEX.htm

http://www.asperbio.com/index_testing.htm

CombiMatrix:

http://www.combimatrix.com/tech_microarrays.htm

Metrigenix:

http://www.metrigenix.com/site802/microflu.shtml

Molcular Biology:

http://www.bioteach.ubc.ca/MolecularBiology/microarray/

NimbleGen:

http://www.nimblegen.com/technology/manufacture.html

Tartu Ülikooli Molekulaar – ja Rakubioloogia Instituut:

http://www.biotech.ebc.ee/praktikum/juhend.html

Wikipedia, the free encyclopedia:

http://en.wikipedia.org/wiki/DNA_microarray

Xeotron:

http://www.invitrogen.com/content.cfm?pageid=10620