

Nálgunaraðferð í tengslagreiningu með skiptingu í hluttré

Óli Þór Atlason og Daníel F. Guðbjartsson

University of Chicago og Íslenskri erfðagreiningu

Vefútgáfa: 19. nóvember 2003

Ágrip NPL er aðferð til að ákvarða hvort sjúkdómur tengist erfðamörkum. Flækjustig hennar vex veldislega með stærð ættartrés þannig að henni er ekki hægt að beita á stór ættartré án nálgunar. Ný nálgunaraðferð er kynnt sem gengur út á að hluta stór tré niður í smærri hluttré sem hægt er að greina nákvæmlega og sameina síðan upplýsingarnar úr þeim. Því er lýst hvernig aðferðin hefur verið útfærð.

Inngangur

Margir sjúkdómar eru algengari í sumum ættum en öðrum og virðast erfast að einhverju leyti. Orðið meingen er notað yfir ákveðinn breytileika í erfðamenginu sem tengist slíkum sjúkdómi. Í tilfelli flókinnna erfðasjúkdóma á borð við heilablóðfall, offitu eða hjartaáfall eru umhverfisþættir ekki síður mikilvægir. Þannig er ekki víst að einstaklingur fái sjúkdóminn þótt hann hafi erfð meingenið og eins er hægt að fá sjúkdóminn þótt meingenið sé ekki að finna á litningi. Einnig getur verið um að ræða fleiri en eitt meingen fyrir sama sjúkdóm [1].

Eitt af fyrstu skrefum við leit að meingeni er tengslagreining. Henni er ætlað að finna svæði á litningi sem inniheldur mjög líklega gen sem tengist sjúkdómnum. Tengslagreining hefur takmarkaða upplausn og því þarf að beita öðrum aðferðum til þess að afmarka litningasvæðið frekar [2].

Tengslagreining er samheiti yfir tölfraðileg próf sem beita má til að ákvarða hvort tiltekin kennileiti á litningnum tengjast sjúkdómi. Kennileitin eru kölluð erfðamörk. Algengustu erfðamörkin í rannsóknum Íslenskrar erfðagreiningar eru langar endurtekningar á einföldum, stuttum basaröðum t.d. CACACACA. Endurtekningarnar koma ávallt fyrir á sama stað á litningnum en fjöldi endurtekninga er breytilegur milli einstaklinga [1]. Þannig er talað um að menn hafi mismunandi útgáfur af erfðaeefni á ákveðnum erfðamörkum.

Erfðamörk eru sögð tengjast sjúkdómi ef þau hafa eftirfarandi eiginleika: Ef við veljum af handahófi tvo einstaklinga sem hafa erfð svæði sem inniheldur erfðamörkin frá sameiginlegum ættföður og ef annar þeirra hefur sjúkdóminn þá eru auknar líkur á að hinn hafi hann.

Til eru fjölmargar mismunandi aðferðir við tengslagreiningu. Sú tegund sem hér er til umfjöllunar er ákveðin tegund stikalausrar tengslagreiningar þar sem einungis eru borin saman erfðamörk sjúklunga [3]. Hún ber heitið NPL, sem stendur fyrir Non-Parametric Linkage.

Upplýsingar sem notaðar eru við NPL tengslagreiningu eru ættartré með fleiri en einum sjúklingi og arfgerðarupplýsingar um meðlimi ættartrésins, jafnt sjúka sem heilbrigða. Með arfgerðarupplýsingum er átt við mælingar á erfðaeefni á mörgum erfðamörkum, t.d. lengd endurtekninga. Þær eru notaðar til þess að ákvarða erfðamynstur fjölskyldunnar, þ.e. hvaðan einstaklingar hafa erfð tiltekin litningasvæði. Í rannsóknum Íslenskrar erfðagreiningar eru notaðar stórar fjölskyldur til þess að auka upplausn og styrk tengslagreiningarinnar. Upplýsingagildi einstakra erfðamarkna er takmarkað þar sem einungis er um að ræða fáar útgáfur erfðaefnis á hverju þeirra. Oft er því ekki hægt að segja til um það með vissu að litningasvæði tveggja einstaklinga hafi erfst frá sama föföður, þó að lengdir erfðamarkanna á svæðinu séu þær sömu

hjá báðum einstaklingum. Þar af leiðandi er yfirleitt ekki hægt að ákvarða einhlýtt erfðamynstur þessara stóru fjölskyldna.

Sérhver afkomandi í ættartré erfir eina samsætu frá móður sinni og aðra frá föður sínum. Hvort þeirra hefur tvær samsætur þannig að möguleikarnir eru fjórir. Þannig er fjöldi mögulegra erfðamynstra fjölskyldu 4^{n_a} , þar sem n_a er fjöldi afkomenda. NPL tengslaggreiningaraðferðin byggist á að meta líkur sérhvers erfðamynsturs. Það er hinsvegar ekki hægt að beita henni á mjög stórar fjölskyldur vegna þess að minnis- og reikningaþörfin helst í hendur við fjölda mynstra og stækkar því ört með fjölda afkomenda í fjölskyldunni. Það er hvítleitt því að við leit að meingenum eru tengslapróf sem byggjast á stórum fjölskyldum oft sterkari en próf sem byggjast á litlum fjölskyldum. Nálgunaraðferðin sem lýst er í þessari grein gerir vísindamönnum í erfðafræði kleyft að beita NPL-aðferðinni á stórar fjölskyldur.

Nálgunaraðferðin byggist á því að hluta stór ættartré niður í smærri búta. Þá er hægt að beita NPL aðferðinni á sérhvert hluttré. Síðan þarf að sameina niðurstöðurnar fyrir hluttrén þannig að hægt sé að túlka þær á svipaðan máta og ef aðferðinni hefði verið beitt á upphaflega tréð.

NPL skorið

Fyrir sérhvert erfðamark einstaklings er um að ræða tvö eintök af erfðaefni, eitt frá hvoru foreldri. Eintökin eru kölluð samsætur af því að þau hafa sömu staðsetningu. Tvær samsætur eru sagðar samstofna ef þær eru komnar af sömu samsætu sameiginlegs forföður. Samstofna samsætur geta ýmist verið í sama einstaklingi eða sín í hvorum einstaklingnum. Ef erfðamörk eru nálægt meingeni á litningnum má búast við að samsætur sjúklinga séu oftast samstofna þar en á öðrum erfðamörkum. NPL aðferðin byggist á því að meta að hversu miklu leyti samsætur sjúklinga á tilteknum erfðamörkum eru samstofna. Til þess eru skilgreind skorföll sem eru mælikvarði á því hve samstofna samsætur sjúklinga í ættartré eru.

Til eru nokkur mismunandi skorföll, t.d. S_{pairs} , S_{all} , S_{robdom} o.fl. [4]. Skorföll gefa mismikinn styrk eftir því með hvaða hætti sjúkdómurinn erfist. Þannig henta t.d. mismunandi skorföll fyrir víkjandi sjúkdóma annars vegar og ríkjandi hins vegar. Í leit að flóknum sjúkdómum er ekki mikið vitað fyrirfram um hvernig sjúkdómurinn erfist og því er best að nota

skorföll sem reynast vel fyrir margar mismunandi tegundir sjúkdóma. Sýnt hefur verið að skorfallið S_{pairs} uppfyllir þetta skilyrði [4]. Nálgunaraðferðina sem hér er lýst er aðeins hægt að nota ef valið er skorfallið S_{pairs} .

Athugum ættartré með fleiri en tveimur sjúkum meðlimum og festum ákveðin erfðamörk. Að tilteknu erfðamynstri gefnu er skorfallið S_{pairs} skilgreint sem

$$S_{\text{pairs}} = \sum_{\substack{a, b \text{ par af} \\ \text{sjúklingum, } a \neq b}} S_{ab},$$

þar sem S_{ab} táknar fjölda óraðaðra para samsætna á erfðamörkunum sem a og b eiga samstofna. Hann getur verið 0, 1, 2 eða 4. Að gefnu safni n ættartrjáa táknum við S_{pairs} skorfalið i -ta trésins með S_i . Núlltilgátan í tengslaggreiningu, H_0 , er að engin tengsl séu á milli sjúkdóms og erfðamarkanna. Ef H_0 er rétt þá eru erfðir Mendelskar sem þýðir að öll erfðamynstur eru jafnlíkleg. Í því tilfalli er unnt að reikna væntigildi og fervik skorfallsins fyrir gefið ættartré til dæmis með því að nota skyldleikastuðla. Fjallað er um þá síðar í greininni. Staðlaða skorfallið er skilgreint með:

$$Z_i = \frac{S_i - E[S_i|H_0]}{\sqrt{V[S_i|H_0]}}$$

Gefum okkur nú að við höfum safnað upplýsingum um arfgerðir einhverra einstaklinga í ættartrénu. Þá má reikna skilyrta væntigildið af Z_i :

$$\bar{Z}_i = E[Z_i|H_0, g]$$

þar sem g stendur fyrir arfgerðarupplýsingarnar. Það er þessi útreikningur sem krefst mikils reikniafls. NPL skorið \bar{Z} er

$$\bar{Z} = \sum_{i=1}^n \bar{Z}_i.$$

Til eru forrit sem reikna \bar{Z} fyrir ættartré af takmarkaðri stærð, t.d. Allegro [5], GENEHUNTER [3] og Merlin [6]. Upphaflega byggðust tölfræðileg próf í þessari tegund tengslaggreiningar á því að kanna dreifingu \bar{Z} þegar núlltilgátan er rétt og nota þannig NPL-skorið beint sem prófstærð. Þegar $n \rightarrow \infty$ stefnir dreifing NPL skorsins á normaldreifingu með meðalgildi 0. Ef við gerum ráð fyrir að fervikið sé 1 þá má bera \bar{Z} saman við stöðluðu normaldreifinguna og var það gert til að byrja með. Slík próf eru mjög veik í flestum tilvikum þar sem fervikið er minna en 1 ef

upplýsingarnar eru ekki fullkomnar [3]. Önnur aðferð er sú að kanna dreifingu NPL skorsins fyrir gefið ættartré með hermun. Það er gríðarlega tímafrek aðferð og í raun ónothæf fyrir stór ættartré.

Í stað þessara aðferða má notast við tölfraðileg líkön yfir möguleg erfðamynstur sem voru innleidd af Kong [7]. Ef gengið er út frá slíku líkani má framkvæma líknahlutfallspróf. Þannig má reikna bæði p -gildi og LOD skor, sem eru svipuð p -gildum en hafa skírskotun til gamalreindra aðferða í erfðafræði. Um er að ræða tvö líkön, annarsvegar línulega líkanið og hinsvegar veldisvísis líkanið. Þessi líkön eru skilgreind með jöfnunum

$$P_{vel}(w|\delta) = c_i r_i(\delta) \exp \frac{\delta \gamma_i(S_i(w) - \mu_i)}{\sigma_i}$$

$$P_{lin}(w|\delta) = c_i \left(1 + \frac{\delta \gamma_i(S_i(w) - \mu_i)}{\sigma_i} \right).$$

Hér er w tiltekið erfðamynstur, δ er stíki fjölskyldunnar og er rauntala í ákveðinni grennd um 0, og c_i og r_i gegna því hlutverki að staðla líkindadreifinguna. Tilfellið þegar $\delta = 0$ samsvarar því að erfðir séu Mendelskar. Jákvæð gildi á δ gefa erfðamynstrum þar sem S_i skorið er hátt hærri líkur.

Línulega líkanið hefur \bar{Z} sem tæmandi hendingu [7]. M.ö.o. er nóg að reikna NPL skorið til þess að geta framkvæmt tölfraðileg tengslapróf á grundvelli línulega líkansins. Veldisvísislíkanið hefur því miður ekki einfalda tæmandi hendingu. Af því leiðir að sú nálgunaraðferð sem hér er lýst er ekki nothæf fyrir það.

Samleitni dreifingar líknahlutfallsins að markdreifingunni er talsvert hraðari þegar veldisvísislíkanið er notað en þegar línulega líkanið er notað. Þar af leiðandi er veldisvísislíkanið yfirleitt valið fram yfir línulega líkanið. Það er því nokkur ágalli á nálgunaraðferðinni að hún nýtist aðeins fyrir línulega líkanið. Afleiðingin er sú að nota verður fleiri fjölskyldur til að fá sambærilegar niðurstöður með línulega líkaninu og fást með veldisvísislíkaninu.

Nálgunaraðferðin

Látum A_i vera mengi allra tveggja punkta mengja $\{a, b\}$ þar sem a og b eru sjúkír meðlimir i -ta ættartrésins með $a \neq b$ og táknum ættartréð sem til atvagnar er með P . Þá má skrifa \bar{Z}_i fyrir skorfallið

S_{pairs}

$$\bar{Z}_i = \frac{\sum_{A_i} E[S_{ab}|H_0, P, g] - E[S_i]}{\sqrt{V[S_i]}}. \quad (1)$$

Í þessari jöfnu og þeim sem á eftir koma tákna $E[\cdot]$ og $V[\cdot]$ væntigildi og fervik þegar gert er ráð fyrir að erfðir eru Mendelskar. Þannig er $E[S_i] = E[S_i|H_0]$ og svo framvegis.

Nálgunin felst í að líta fram hjá þeim hlutum ættartrésins sem gefa líklega ekki miklar upplýsingar um S_{ab} . Fyrir hvert par $\{a, b\}$ veljum við eitt ákveðið hluttré $P_{ab} \subset P$ sem inniheldur bæði a og b og hefur einn samhengisþátt. Táknnum með $g_{ab} \subset g$ arfgerðarupplýsingarnar um einstaklingana sem eru í P_{ab} . Þá er nálgunin þessi:

$$E[S_{ab}|H_0, P, g] \approx E[S_{ab}|H_0, P_{ab}, g_{ab}]. \quad (2)$$

Forritið Allegro er notað til að reikna stærðina hægra megin í jöfnu 2 að gefnum P_{ab} og g_{ab} . Til þess að nálgá \bar{Z}_i er nálgunin 2 sett inn í 1, þ.e. notuð er jafnan

$$\bar{Z}_i \approx \frac{\sum_{A_i} E[S_{ab}|H_0, P_{ab}, g_{ab}] - E[S_i]}{\sqrt{V[S_i]}}.$$

Til þess að reikna út þessa stærð þarf einnig að reikna $E[S_i] = \sum E[S_{ab}]$ og $V[S_i]$. Fervikið má rita með eftirfarandi hætti:

$$V[S_i] = V \left[\sum_{A_i} S_{ab} \right]$$

$$= \sum_{A_i} V[S_{a,b}] + 2 \sum_{B_i} \text{Cov}[S_{ab}, S_{cd}],$$

með $B_i = \{(\{a, b\}, \{c, d\}) \in A_i^2 \mid \{a, b\} \neq \{c, d\}\}$. Það er því nóg að reikna $E[S_{ab}]$ og $E[S_{ab}S_{cd}]$ fyrir allar mögulegar ferndir meðlima P , (a, b, c, d) til að fá $E[S_i]$ og $V[S_i]$.

Skyldleikastuðlar

Ein leið til að reikna þessi væntigildi og fervik byggir á notkun skyldleikastuðla. Þeir eru skilgreindir með eftirfarandi hætti sem föll af hópum meðlima a, b, c, d í ættartré: Festum ákveðin erfðamörk. Í hverjum einstaklingi er þá um að ræða tvær samsætur. Ef valin er ein samsæta af handahófi úr hverjum einstaklingi þá tákna Φ_{ab} líkurnar á því að samsæturnar sem valdar voru úr a og b séu samstofna. $\Phi_{ab,cd}$ tákna líkurnar á því að samsæturnar sem valdar voru úr a og b séu

samstofna og að samsæturnar sem valdar voru úr c og d séu líka samstofna. Til er almennari skilgreining á stuðlum af þessu tagi [8].

Skyldleikastuðlarnir eru nátengdir stærðunum sem við höfum áhuga á. Að gefnum meðlimi ættartrés, a , og erfðamörkum köllum við samsæturnar G_a^i , $i = 1, 2$. Í útreikningunum notum við jafnaðarmerkið til að tákna að samsætur séu samstofna. Ef við veljum fjórar samsætur $G_a^i, G_b^j, G_c^k, G_d^l$ af handahófi frá fjór-um meðlimum ættartrés, a, b, c, d þá er (i, j, k, l) sér í lagi hending með jafna dreifingu yfir $\{1, 2\}^4$ og við fáum:

$$\begin{aligned}\Phi_{ab} &= P(G_a^i = G_b^j) \\ &= \sum_{s=0}^4 P(G_a^i = G_b^j | S_{ab} = s) P(S_{ab} = s) \\ &= \sum_{s=0}^4 \frac{s}{4} P(S_{ab} = s) = \frac{E[S_{ab}]}{4}.\end{aligned}$$

Með svipuðum hætti fæst:

$$\begin{aligned}\Phi_{ab,cd} &= \sum_{r=0}^4 \sum_{s=0}^4 P(G_a^i = G_b^j, G_c^k = G_d^l \\ &\quad | S_{ab} = r, S_{cd} = s) P(S_{ab} = r, S_{cd} = s) \\ &= \sum_{r=0}^4 \sum_{s=0}^4 \frac{rs}{16} P(S_{ab} = r, S_{cd} = s) \\ &= \frac{E[S_{ab}S_{cd}]}{16}.\end{aligned}$$

Þannig má reikna væntigildi, fervik og samdreifni stuðlanna S_{ab} ef við þekkjum skyldleikastuðlana. Hægt er að reikna skyldleikastuðlana Φ_{ab} og $\Phi_{ab,cd}$ einn af öðrum með því að nota endurkvæmar jöfnur [9, 10].

Eiginleikar nálgunarinnar

Gæði nálgunarinnar eru háð því hvernig hluttrén eru valin. Ef við veljum t.d. $P_{ab} = P$, þ.e. höldum öllu trénu fyrir sérhvert par, þá er ekki um nálgun að ræða, heldur nákvæman útreikning. Við val á hluttré verður að gæta þess að halda öllum sameiginlegum forfeðrum a og b og öllum leiðum frá slíkum forfeðrum niður í a eða b í hluttrénu. Þá er dreifing S_{ab} þegar núlltilgátan er rétt varðveitt og einu áhrif þess að velja fleiri einstaklinga í hluttréð eru að bæta matið á $E[S_{ab}|g]$.

Áhrif þess að bæta einstaklingum í hluttré eru að fjölgja arfgerðum sem notaðar eru til þess að greina

á milli erfðamynstra. Þannig er minni óvissa um það hvaða erfðamynstur er rétt. Með því að búa til mismunandi hluttré, herma arfgerðarupplýsingar og meta upplýsingagildi ættartrésins eftir hverja hermun má búa til reynslulögmál um hvaða einstaklinga eigi að velja í hluttrén [11]. Slíkar rannsóknir hafa leitt í ljós að nákvæmnin eykst yfirleitt mest þegar bætt er við náskyldum einstaklingum. Foreldrarnir eru yfirleitt sjálfkrafa með, því án þeirra er ekki hægt að tengja a og b um sameiginlegan forföður. Matið batnar talsvert með því að bæta við systkinum og börnum a og b . Einnig eru verðmætar upplýsingar um þá sem eru á leiðinni á milli a og b , t.d. börn sameiginlegs forföður.

Tölvuforritið Allegro framkvæmir nákvæma tengslaggreiningu án nálgunaraðferðarinnar. Það hefur flækjustigið $O(2^{2n-3f})$, þar sem n er fjöldi meðlima ættartrésins og f fjöldi ættfedra, meðlima sem ekki hafa foreldra í ættartrénu sem til skoðunar er. Nálgunaraðferðin krefst þess að við framkvæmum nákvæma tengslaggreiningu á hverju hluttré. Fjöldi þeirra er í mesta lagi $\binom{k}{2} = O(k^2)$, þar sem $k \leq n$ er fjöldi sjúklinga í trénu og líta má svo á að það taki fastan tíma að framkvæma hverja tengslaggreiningu á hluttré, vegna þess að reynt er að velja hluttré af takmarkaðri stærð. Flækjustig þess að reikna vægi S_{ab} er $O(m^4)$, þar sem $m \leq n$ er fjöldi meðlima upphaflega ættartrésins sem ómissandi eru til að tengja saman sjúku meðlimina. Þannig er flækjustig nálgunaraðferðarinnar í mesta lagi $O(n^4)$ ef hægt er að velja hluttrén af takmarkaðri stærð.

Takmarkanir

Helsti gallinn við nálgunaraðferðina liggur í línulega líkaninu. Táknnum líknahlutfallið sem fæst með líkaninu með λ . Dreifing $-2 \ln \lambda$ stefnir á χ^2 dreifingu með einni frígráðu þegar fjölskyldunum fjölgar. Frávikin frá χ^2 dreifingunni eru hinsvegar tiltölulega stór þegar fjölskyldurnar eru fáar og sérstaklega þegar sterk tengsl eru fyrir hendi. Þessi frávik eru í íhaldssama átt, þ.e. líklegra er að missa af tengslum en að fá of sterka vísbendingu um tengsl.

Önnur takmörkun er sú að nauðsynlegt er fyrir sérhverja tvo sjúklinga sem eru skyldir að geta valið hluttré sem inniheldur öll ættartengsl milli þeirra og er nægilega lítið til þess að hægt sé að reikna líkindadreifingu erfðamynstra þess nákvæmlega. Til eru stórar fjölskyldur þar sem mikið er um barneignir náskyldra. Í slíkum fjölskyldum er oft ekki hægt

að velja nægilega lítil hluttré vegna þess hve flóknar tengingar eru í fjölskyldunni. Þá er ekki hægt að beita nálgunaraðferðinni. Sem dæmi um slíka fjölskyldu má nefna Hutterites ættbálkinn í Bandaríkjunum sem hefur verið mikið rannsakaður af vísindamönnum við Háskólann í Chicago. Í íslenskum fjölskyldum er hinsvegar ekki óhóflega mikið um barneignir náskyldra og því er yfirleitt hægt að beita aðferðinni á þær.

Þriðja takmörkun aðferðarinnar liggur í að velja verður skorfallið S_{pairs} . Hún er þó ekki ýkja hvítleið þar sem mælt er með notkun þess umfram önnur skorföll, eins og kemur fram í kaflanum um skorföll.

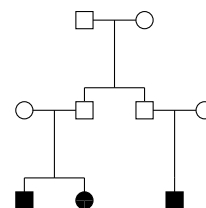
Loks verður að hafa í huga að hér er um að ræða nálgunaraðferð. Þannig verða niðurstöðurnar ekki nákvæmlega þær sömu og ef ekki væri nálgad. Aðferðin gefur hinsvegar góða raun þegar fjöldi fjölskyldna sem athugaður er samtímis er nægilega mikill.

Útfærsla

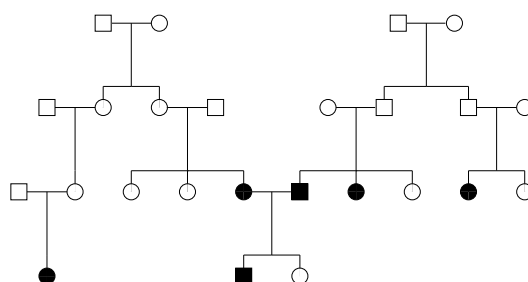
Skrifað hefur verið forrit sem skiptir stórum ættartrjám sjálfkrafa í hæfilega stór hluttré. Reikniritið virkar þannig að fyrir hvert par sjúklinga er valið minnsta hugsanlega hluttré sem tengir sjúklingana og varðveitir dreifingu S_{ab} þegar núlltilgátan er rétt. Síðan er bætt við aukameðlimum með því að styðjast við reynslulögmál sem eru svipuð þeim sem getið var hér að ofan. Ennfremur hefur skyldleikastuðlaaðferðin verið útfærð og því hægt að beita nálgunaraðferðinni.

Nálgunaraðferðin var prófuð á tiltölulega einföldu ættartré (mynd 1). Arfgerðir voru hermdar með því að gera ráð fyrir að núlltilgátan sé rétt og líknahlutfallspróf voru framkvæmd með því að nota nálgunaraðferðina. Niðurstaðan er að þegar fjölskyldunum fjölgar virðist markdreifing líknahlutfallsins vera sú sama og þegar ekki er nálgad. Sér í lagi má treysta þgildum af stærðinni 10^{-3} þegar fjölskyldurnar eru 25 eða fleiri. Þegar útkomur nálgunaraðferðarinnar eru bornar saman við útkomur með nákvæmum reikningum kemur í ljós talsverð fylgni, t.d. er fylgni NPL skoranna 0.980 þegar fjölskyldurnar eru 25.

Á mynd 2 má sjá ættartré sem er við efri mörk þess sem fýsilegt er að reikna án nálgunaraðferðar. Fyrstu útfærslurnar á nálgunaraðferðinni ráða við ættartré með a.m.k. 6 sinnum fleiri meðlimi. Beita má



Mynd 1. Einföld prófunarfjölskylda. Svarti liturinn táknar sjúkan meðlim.



Mynd 2. Efri mörk á stærð fjölskyldu fyrir nákvæma tengslaggreiningu. Ættartréð er byggt á raunverulegri fjölskyldu en hefur verið breytt lítillega.

aðferðinni á enn stærri ættartré með því að reikna skyldleikastuðlana ekki nákvæmlega heldur að meta þá með hermunaraðferðum. Þó er óvíst hvort máli skipti að geta framkvæmt tengslaggreiningu á margfalt stærri ættartrjám en þessi útfærsla ræður við, m.a. vegna þess að erfitt er að sannfæra sig um að ættfræðiupplýsingar langt aftur í tímann eru réttar.

Summary NPL is a method used to test for linkage of a disease to a genetic marker. The method can not be directly applied to large pedigrees since the computational complexity grows exponentially in the size of the pedigree. We propose a novel approximation method that breaks large pedigrees into subpedigrees. The subpedigrees are analyzed directly and the information from the subpedigrees combined to obtain a statistic approximating NPL.

Heimildir

- [1] Grunnhugtök í erfðafræði, kynningarefni gefið út af Íslenskri erfðagreiningu.
- [2] Hreinn Stefánsson og Þorgeir H. Þorgeirsson, Meingen og flóknir erfðasjúkdómar, *Lýfjatíðindi*, **5:2**, 64–70 (1998).

- [3] L. Kruglyak, M.J. Daly, M.P. Reeve-Daly og E.S. Lander, Parametric and nonparametric linkage analysis: A unified multipoint approach, *American Journal of Human Genetics*, **58**, 1347–1363 (1996).
- [4] M.S. McPeck, Optimal Allele-Sharing Statistics for Genetic Mapping Using Affected Relatives, *Genetic Epidemiology*, **16**, 225–249 (1999).
- [5] Daníel F. Guðbjartsson, Kristján Jónasson, M. Frigge og A. Kong, Allegro, a new computer program for multipoint linkage analysis, *Nature Genetics*, **25**, 12–13 (2000).
- [6] G.R. Abecasis, S.S. Cherny, W.O. Cookson og L.R. Cardon, Merlin - rapid analysis of dense genetic maps using gene flow trees, *Nature Genetics*, **30**, 97–101 (2002).
- [7] A. Kong og N.J. Cox, Allele-sharing models: LOD scores and accurate linkage tests, *American Journal of Human Genetics*, **61**, 1179–1188 (1997).
- [8] K. Lange og J.S. Sinsheimer, Calculation of genetic identity coefficients, *Annals of Human Genetics*, **56**, 339–346 (1992).
- [9] G. Karigl, A recursive algorithm for the calculation of identity coefficients, *Annals of Human Genetics*, **45**, 299–305 (1981).
- [10] M. Abney, M.S. McPeck og C. Ober, Estimation of variance components of quantitative traits in inbred populations, *American Journal of Human Genetics*, **66**, 629–650 (2000).
- [11] D.L. Nicolae, Allele sharing models in gene mapping: a likelihood approach, *doktorsritgerð*, Department of Statistics, University of Chicago (1999).

Um höfundana: Óli Þór Atlason er fæddur 1978 í Reykjavík. Hann lauk stúdentsprófi frá Verzlunarskólanum 1998 og BS prófi í stærðfræði frá Háskóla Íslands 2001. Hann hóf doktorsnám í tölfraði við Háskólann í Chicago árið 2002. Óli hefur starfað hjá Íslenskri Erfðagreiningu, Fjárfestingabanka atvinnulífsins og hjá Skímu/Miðheimum.

Daníel Fannar Guðbjartsson er fæddur 1973 í Hafnarfirði. Hann lauk stúdentsprófi frá Fjölbrautaskóla Suðurnesja 1993, BS prófi í stærðfræði frá Háskóla Íslands 1996 og doktorsprófi í tölfraði frá Duke-háskóla 2001. Daníel hefur starfað hjá Íslenskri erfðagreiningu frá 1999.

University of Chicago
 Department of Statistics
 5734 S. University Avenue
 Chicago, IL 60637
 oli@uchicago.edu
 Íslensk erfðagreining
 Sturlugötu 8
 IS-101 Reykjavík
 dfg@decode.is

Móttekin: 5. mars 2002