



Universiteit
Leiden
The Netherlands

Linkage mapping for complex traits : a regression-based approach

Lebrec, J.J.P.

Citation

Lebrec, J. J. P. (2007, February 21). *Linkage mapping for complex traits : a regression-based approach*. Retrieved from <https://hdl.handle.net/1887/9928>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/9928>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

In dit proefschrift worden manieren beschreven om de huidige opzet en analyse van studies naar de koppeling van genen (*linkage*) met complexe eigenschappen te verbeteren. In *linkage* onderzoek wordt gebruik gemaakt van genetische merkers (*markers*). Met deze *markers* kan men genetische overeenkomstigheden tussen verwanten meten. Door deze genetische gelijkenis te vergelijken met fenotypische overeenkomsten, kunnen regio's op het chromosoom geïdentificeerd worden waarin genen liggen die bijdragen tot de vorming van het betreffende fenotype. Hoewel, deze methode erg succesvol bleek bij het in kaart brengen van genen en eigenschappen, die volgens de wet van Mendel overerven, schiet hij bij meer complexe overervingpatronen vaak tekort. Omdat betrokken genen maar een zeer beperkte invloed hebben op complexe eigenschappen is men m.b.t *linkage* studies intrinsiek beperkt. Deze intrinsieke beperking rechtvaardigt een heel eigen statistische benadering, welke de basis vormt van de methodologie beschreven in dit proefschrift. Voor het toetsen van hypothesen kunnen score toetsen worden gebruikt [Cox and Hinkley, 1974]. De lokale optimaliteit eigenschappen van deze toetsen blijken zeer geschikt in de context van complexe eigenschappen. Bovendien hebben zij vaak een herkenbare uitdrukking en kunnen zij geïnterpreteerd worden in termen van regressieanalyse, waardoor zij in principe snel uit te rekenen zijn. Dit laatste is van groot belang in genetisch onderzoek waarbij vaak grote hoeveelheden data geanalyseerd moeten worden.

In **hoofdstuk 1** wordt een inleiding gegeven over de genetische mechanismen die ten grondslag liggen aan *linkage*. Tevens volgt een korte beschouwing over de traditionele methodologie en wordt een samenvatting gegeven van belangrijke nog onopgeloste vraagstukken.

Hoofdstuk 2 behandelt hoofdzakelijk de analyse van kwantitatieve eigenschappen die gemeten zijn in geselecteerde families. Hierbij is de selectie gebaseerd op de waarde van een betreffend kenmerk. Een score toets gebaseerd op de conditionele likelihood gegeven de fenotypische waarden, wordt afgeleid. Deze toets kan gebruikt worden

bij data uit willekeurige stambomen. Hoewel bij de afleiding van de toets wordt aangenomen dat het model normaal verdeelde variatiecomponenten bevat, is de type I fout van de toets robuust tegen afwijkingen van deze normale verdeling. Onder de aanname dat het model de verdeling van het fenotype goed beschrijft, heeft de toets optimale eigenschappen voor lokale alternatieven. Bovendien geeft de waarde van de bijbehorende Fisher informatie van de toets een indicatie in hoeverre elke familie informatief is. Deze Fisher informatie kan gebruikt worden als criterium voor het selecteren van individuen voor genotyperingen. Verderop in het hoofdstuk wordt een aangepaste versie van de toets gegeven voor binaire gegevens. Een model met een onderliggende continue latente variabelen wordt gebruikt waarbij deze variabelen in twee klassen wordt verdeeld door een drempelwaarde te creëren. Dit model wordt liability threshold model genoemd.

Hoofdstuk 3 bepleit het gebruik van geselecteerde families bij het in kaart brengen van genen voor complexe eigenschappen waarbij tweelingen worden gebruikt. Met behulp van de methodologie, welke gebaseerd op het informatiecriterium dat in hoofdstuk 2 is afgeleid, worden potentiële voordelen gekwantificeerd door gebruik te maken van een serie voorbeelden van kwantitatieve en kwalitatieve fenotypen, welke relevant zijn voor het GenomEUtwin project.

Hoofdstuk 4 behandelt het probleem van genotyperingsfouten binnen *linkage* onderzoek. Het effect van genotyperingsfouten op *linkage* studies wordt beschreven door een formule te creëren, die de vertekening die optreedt door genotyperingsfouten weer kan geven. Deze formule geeft inzicht in enkele van de empirische bevindingen, in het bijzonder verklaart het de rol van genotyperingsfouten in onderzoeksopzetten met selecte versus aselecte data. Ten slotte wordt een voorstel gedaan tot een robuuste aanpassing van de gebruikelijke *linkage* toetsen gebaseerd op een genoom wijde controle van het overschot van allelen, die een kopie zijn van een zelfde voorouderlijk allel. Allelen die een kopie zijn van een zelfde voorouderlijk allel worden identical by descent genoemd. Deze aanpassing geeft niet alleen robuustheid tegen genotyperingsfouten, maar ook tegen andere processen die de verwachte waarde van deze fractie verstoren.

Hoofdstuk 5 bespreekt de (on)juistheid van aan aantal standaard methoden welke gebruikt worden als markerinformatie niet volledig is. Het probleem van gevallen waarbij de methode van gegeneraliseerde schattingsvergelijkingen (*generalized esti-*

mating equations) voor het in kaart brengen van genen faalt [Liang et al., 2001] wordt uitgelicht.

Hoofdstuk 6 vertaalt de standaard meta-analyse technieken naar het onderzoeksveld van het in kaart brengen van genen die een rol spelen bij kwantitatieve eigenschappen (*quantitative trait loci (QTLs)*). Dit onderzoeksgebied heeft een aantal specifieke kenmerken waarbij aanpassingen nodig zijn. Het probleem van heterogeniteit in genetische loci wordt nader toegelicht. Wanneer er geen co-variabelen geobserveerd zijn op individu nivo en onder een homogeen model, is de meta-analytische aanpak asymptotisch equivalent aan de analyse van samengevoegde databestanden, maar is logistiek veel eenvoudiger uit te voeren.

Ten slotte wordt in **hoofdstuk 7** een score toets voor *linkage* analyse in de grote klasse van de algemene lineaire modellen beschreven. Deze benadering is gebaseerd op een pseudo-likelihood van de gegevens. Hoewel deze test waarschijnlijk niet optimaal is in alle situaties, heeft deze test het voordeel herkenbaar te zijn en een robuuste type I fout te hebben. Het levert een eenvoudige manier om het bekende effect van co-variabelen te implementeren in *linkage* analyse en is toepasbaar voor willekeurige stambomen.

Het proefschrift wordt afgesloten met conclusie, waarin ik een perspectief schets van de methodologie die een rol speelt in *linkage* bij het in kaart brengen van genen.

