

GroupSeq++/GWAS

Ein R-Programmpaket zur Optimierung und
flexiblen Anpassung von Mehrstufen-Designs
für genomweite Assoziationsstudien

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V.



Impressum

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V.

Charlottenstraße 42/Dorotheenstraße

10117 Berlin

Tel.: +49 (30) 22 00 24 70

Fax: +49 (30) 22 00 24 799

info@tmf-ev.de

www.tmf-ev.de

Stand: Mai 2013

Eine Software für die Planung und das statistische Monitoring genetisch-epidemiologischer Studien

Paket für das frei verfügbare Statistiksystem R mit einer grafischen Benutzeroberfläche

Die Entwicklung von genomweiten dichten Markernetzen und deren populationsgenetische Untersuchung hat der genetisch-epidemiologischen Forschung den Weg von kandidatengenbezogenen Ansätzen zu genomweiten Studien in großen epidemiologischen Kollektiven eröffnet. DNA-Chips mit rund einer Millionen SNP-Marker (SNP = single nucleotide polymorphisms) sind kommerziell verfügbar, und in absehbarer Zeit werden in epidemiologischen Studien die Next-Generation Sequencing-Techniken einsetzbar sein.

Bei solchen genomweiten Screening-Studien ist von vorneherein zu erwarten, dass nur ein sehr kleiner Teil der Marker tatsächlich mit der untersuchten Krankheit bzw. mit dem untersuchten Phänotyp assoziiert ist. Es bietet sich daher schon aus Kostengründen an, das untersuchte Markernetz im Laufe einer Studie aufgrund von Zwischenergebnissen stufenweise auf die vielversprechenden Marker einzuschränken. Hierfür sind in den letzten Jahren geeignete Verfahren entwickelt worden, die nun in einem Softwarepaket des frei verfügbaren Statistiksystems R programmiert wurden und mit einer grafischen Benutzeroberfläche zur Verfügung stehen.

Großer Funktionsumfang

Kein eigener Programmieraufwand notwendig

GroupSeq++ unterstützt die statistischen Berechnungen benutzerfreundlich und ohne Eigenprogrammieraufwand. Es beinhaltet folgende Features:

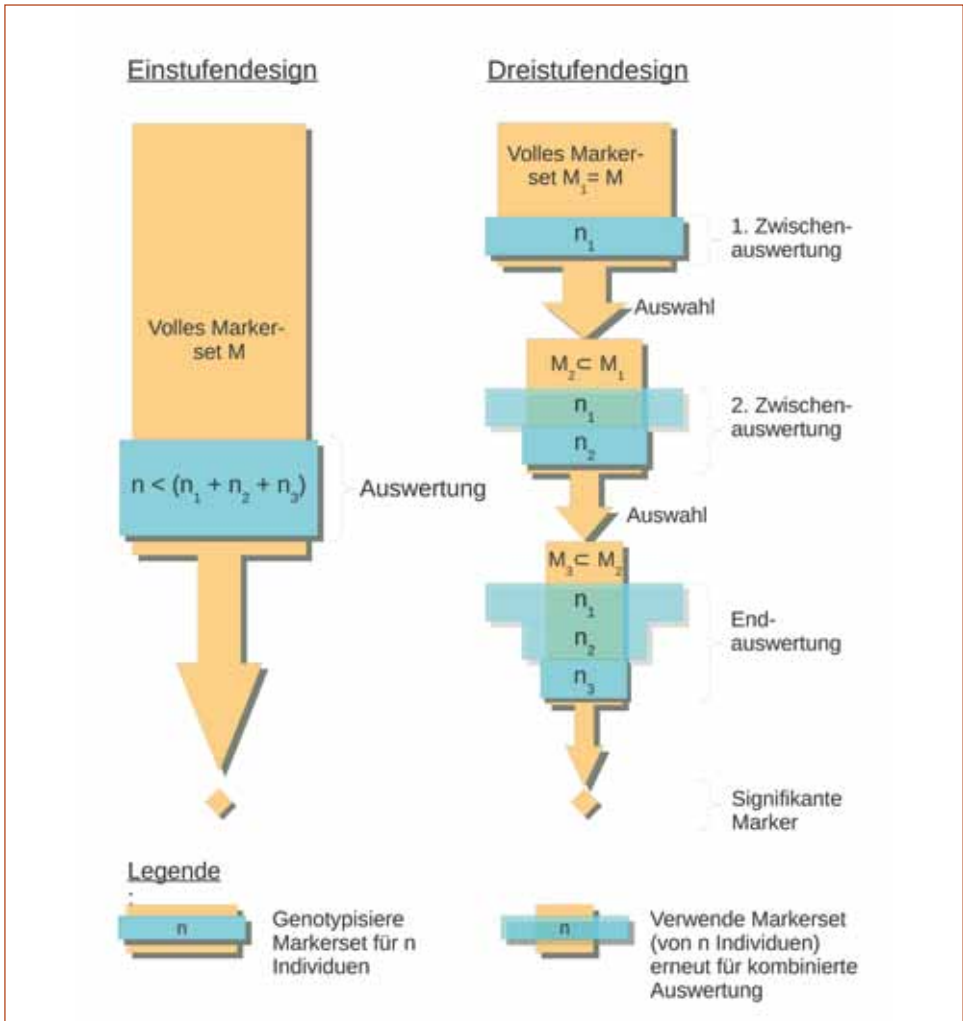
- Konstruktion von optimalen Mehrstufen-Fall-Kontroll-Designs mit bis zu fünf Stufen zur schrittweisen Einschränkung des Markersets auf erfolgversprechende Marker;
- Minimierung der Gesamtkosten der Studie unter Power-Nebenbedingung;
- Maximierung der Power bei gegebenem Budget oder gegebener Fall-Kontroll-Stichprobe;
- Planung der optimalen Fallzahl angepasst an das mehrstufige Auswertungsvorgehen;
- optimierte Aufteilung der Stichprobe auf die einzelnen Stufen und optimale Festlegung der kritischen Grenzen für die Auswahl der Marker anhand von Interims-p-Werten sowie des finalen Signifikanzniveaus bei der Endauswertung.
- Optimierung unter verschiedenen Nebenbedingungen wie
 - fester Stichprobenumfang je Stufe,
 - untere und obere Schranken für den Stichprobenumfang je Stufe,
 - maximal verfügbarer Stichprobenumfang und
 - verfügbares Markersset.

- Weitere Features:
 - Berücksichtigung von »kostenfreien Samples« beispielsweise aus vorangegangenen Studien (zur Konstruktion von follow-up Designs),
 - praxisnahe Spezifikation der Chipkosten durch Angabe verfügbarer Chips inklusive möglicher Rabatte,
 - variable Rekrutierungskosten pro Individuum.

- Tool für zwischenzeitliche flexible Designanpassungen bei GWAS nach einer erfolgten ersten Stufe: flexible datenabhängige Anpassung der Stichprobengröße und freie Auswahl des Markersets für die zweite Stufe (CRP-Prinzip).

Eine ausführliche Dokumentation führt User an die Bedienung des Programms heran

Um dem Funktionsumfang des Programmpakets gerecht zu werden, wurde eine ausführliche, über die üblichen R-Minimalanforderungen hinausgehende Dokumentation erarbeitet. Diese enthält neben der allgemeinen Beschreibung auch zahlreiche Beispiele, die den User schrittweise an die Bedienung des Programms heranführen sollen.



Einstufendesign vs. Dreistufendesign (schematisch)

Download und Kontakt

Die Software steht auf der TMF-Website zum Download zur Verfügung:
<http://www.tmf-ev.de/Produkte/P100301>

Das Programm wird bei den Verantwortlichen der R-Software eingereicht und später dann auch über das Comprehensive R Archive Network (CRAN) als Teil der R-Software bereitgestellt.

Projektleitung

Prof. Dr. Helmut Schäfer

Institut für Medizinische Biometrie und Epidemiologie, Universität Marburg

E-Mail: hsimbe@staff.uni-marburg.de

Kontakt:

Roman Pahl

Institut für Medizinische Biometrie und Epidemiologie, Universität Marburg

E-Mail: pahl@staff.uni-marburg.de

Ansprechpartner bei der TMF

Ulrich Gauger

Telefon: +49 (0) 30 22 00 24 744

E-Mail: ulrich.gauger@tmf-ev.de

www.tmf-ev.de

Verfahren zur flexiblen Planung genetisch-epidemiologischer Studien

Optimale Mehrstufendesigns für genomweite Assoziationsstudien (GWAS), die im Laufe der Studie oder der Genotypisierung eine schrittweise Reduktion des genotypisierten Markersets auf die Erfolg versprechenden Marker ermöglichen. Ziel ist eine Minimierung der Genotypisierungs- bzw. Gesamtstudienkosten, eine Maximierung der Power der Studie bei limitiertem Budget oder eine maximale Ausschöpfung der Power bei Genotypisierung einer vorhandenen Stichprobe. Außerdem Verfahren nach CRP-Prinzip, mit dem aufgrund von Zwischenergebnissen die Marker für die nächste Stufe flexibel ausgewählt werden können, ohne dabei an vordefinierte formale Entscheidungskriterien (kritische Grenze für p-Werte) gebunden zu sein; ermöglicht nötigenfalls auch die Neufestlegung der weiteren Fallzahl. Das Programmpaket (Programmsystem R) unterstützt die statistischen Berechnungen nach diesen Methoden benutzerfreundlich und ohne Eigenprogrammieraufwand.



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung