Wiener Biometrische Sektion
der Internationalen Biometrischen Gesellschaft
Region Österreich – Schweiz
http://www.meduniwien.ac.at/wbs/

Einladung zum

BIOMETRISCHEN KOLLOQUIUM

Am **Donnerstag, 26. September 2013** um **16:15 Uhr** (s.t.)

in der Informatik-Bibliothek (Ebene 3, Raum 88.03.806) des
Zentrums für Medizinische Statistik, Informatik und Intelligente Systeme (CeMSIIS)
der Medizinischen Universität Wien, Spitalgasse 23, 1090 Wien
(Plan siehe http://www.muw.ac.at/cemsiis/allgemeines/anschrift/)

Vortragender:

GREGORY NUEL

MAP5 (CNRS 8145 / Univ. Paris-Descartes)
http://www.mi.parisdescartes.fr/~nuel/index.html

SIMULATION OF PHENOTYPES UNDER H1 IN GENOME WIDE ASSOCIATION STUDIES AND APPLICATIONS

Wir freuen uns auf zahlreichen Besuch.

Gerhard Svolba
Präsident

Franz König
Sekretär

Simulation of phenotypes under H1 in Genome Wide Association studies and applications

V. Perduca [a,*], C. Sinoquet [b], R. Mourad [b,c], G. Nuel [a,*]

[a] MAP5 - UMR CNRS 8145 Université Paris Descartes, [b] LINA - UMR CNRS 6241 Université de Nantes,

[c] Ecole Polytechnique de l'Université de Nantes

***Context*:** GWAs are widely used to investigate the connection between genotypic and phenotypic variation with respect to a given trait (*e.g.* a given disease). Assessing the statistical power of such studies is crucial. Power is empirically estimated by simulating realistic samples under a disease model H1. For this purpose, the gold standard consists in simulating the genotypes given the observed phenotypes (case or control); thus ensuring that the total number of cases stays unchanged. This method is implemented in the software of reference *Hapgen*. We study an alternative approach for simulating samples under H1 that does not require generating new genotypes for each simulation but only phenotypes. ***Methods*:** In particular, we propose to simulate new phenotypic datasets such that a) the phenotypes are in accordance with the corresponding observed genotypes under the chosen model H1; b) the total number of cases is the same as in the observed dataset. In order to do so, we suggest three algorithms: i) a simple rejection algorithm; ii) a MCMC approach; iii) and an exact and efficient backward sampling algorithm. We validated our three algorithms both on a toy-dataset and by comparing them with Hapgen on a more realistic dataset. As an application, we then conducted a simulation study on a *1000 Genomes Project* dataset consisting of 629 individuals (314 cases) and 8,048 SNPs from Chromosome X. We arbitrarily defined an additive disease model with two susceptibility SNPs and an epistatic effect. ***Results*:** The three algorithms are consistent, but backward sampling is dramatically faster than the others. Our approach also gives consistent results with Hapgen. On our application data we showed that our limited design requires a biological *a priori* to limit the investigated region. We also proved that epistatic effects can play a significant role even when simple marker statistics (*e.g.* trend) are used. We finally showed that the overall performance of a GWAs strongly depends on the prevalence of the disease: the larger the prevalence, the better the power. ***Conclusions*:** Our approach is a valid alternative to Hapgen-type methods; it is not only dramatically faster but also has two main advantages: 1) there is no need for sophisticated genotype models (*e.g.* haplotype frequencies, or recombination rates); 2) the choice of the disease model is completely unconstrained (number of SNPs involved, Gene-Environment interactions, hybrid genetic models, etc.). Our three algorithms will soon be available in an `R` package called `waffect` .

*Corresponding authors: Vittorio Perduca, Gregory Nuel, MAP5 Université Paris Descartes, 45 Rue des Saints Pères, 75006 Paris, France, {vittorio.perduca, gregory.nuel}@parisdescartes.fr