FLEXIBLE TWO-STAGE TESTING IN GENOME-WIDE ASSOCIATION STUDIES

A Scherag, H Schäfer, HH Müller

Institute of Medical Biometry & Epidemiology, Philipps-University of Marburg

5th International Conference on Multiple Comparison Procedures, 07/10/2007 Vienna

Overview

- example for a typical genome-wide associations study (GWAS)
- two-stage GWAS - notation
- design modifications and the conditional rejection probabilities
- a new flexible procedure for GWAS
- simulation study
- discussion

Arking et al., 2006

Stage I

115,000 (genetic) markers

in 2 x 100 individuals

Stage II

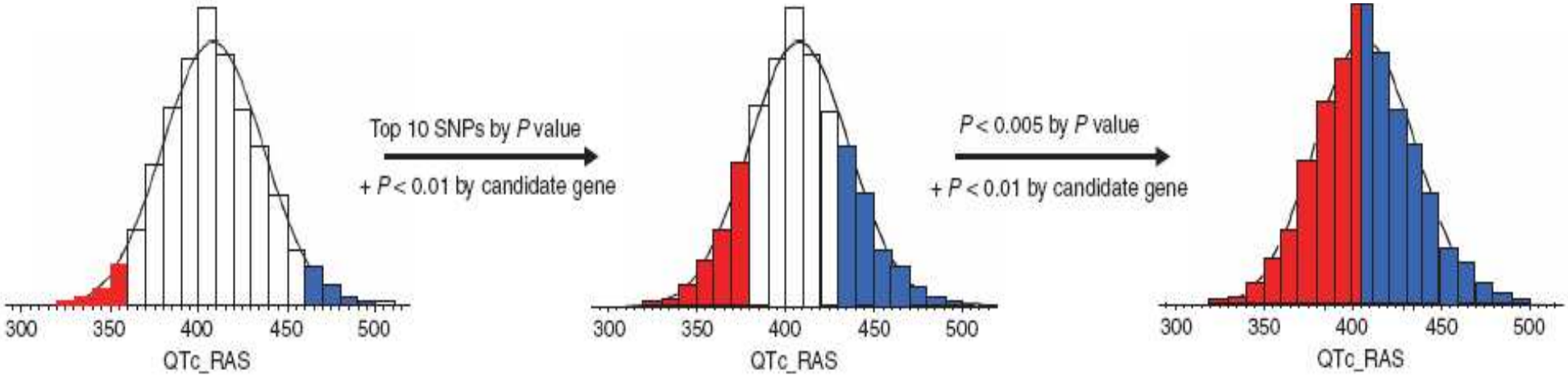
selected markers

in 2 x 200 independent individuals

Stage III

further selected markers

in 3,366 independent individuals



Arking et al., 2006

Stage I

115,000 (genetic) markers

in 2 x 100 individuals

Stage II

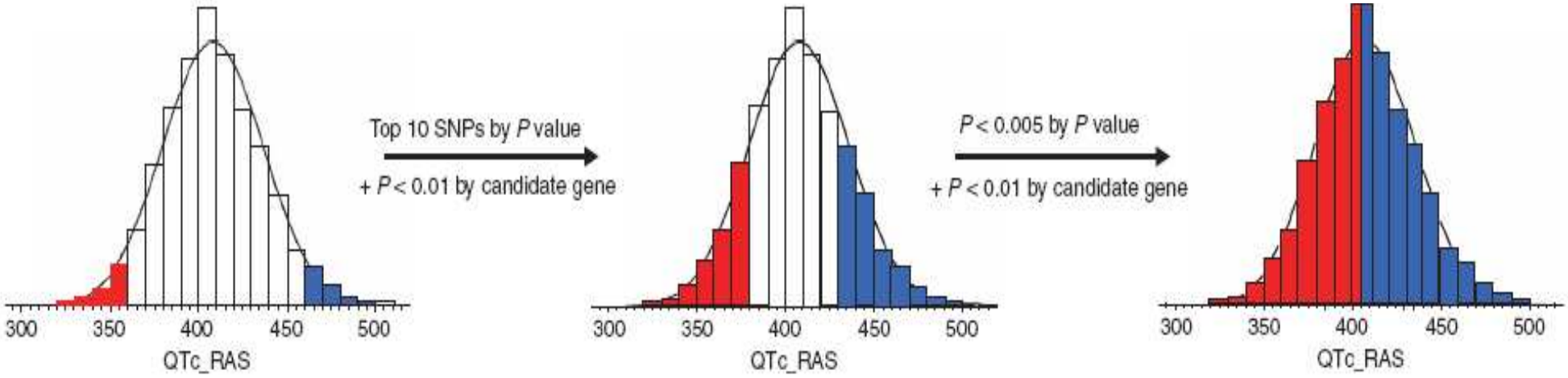
selected markers

in 2 x 200 independent individuals

Stage III

further selected markers

in 3,366 independent individuals



What is the probability for a type I error?

two-stage GWAS - notation

- let the proportion of case:controls remain constant
- for each stage:
  - n_1, n_2 the individuals genotyped; $n = n_1 + n_2$
  - M_1, M_2 are marker sets genotyped (with genotyping costs t_1, t_2)
  - α_1, α_2 are the significance levels chosen such that the family-wise error rate in a strong sense (FWER) **is controlled at a level α**

two-stage GWAS - notation

example for a cost-optimal two-stage GWAS (e.g. Wang et al., 2006):

- $|M_1| = 500{,}000$; t_1=\$0.002, t_2=\$0.035
- one true disease marker:
  - allele frequency=0.1; odds ratio=1.35 (mult. model)
- power = 0.9; FWER control at $\alpha = 0.05$ (one-sided) by Bonferroni method $\alpha_i = 0.05/500{,}000$
- $n_1 = 3{,}238$; $n_2 = 7{,}490$; $n = 10{,}728$
- $\alpha_1 = 0.00370$; $\alpha_2 = 1.6 \times 10^{-7}$

design modifications and the conditional rejection probabilities

- Cui et al., 1999; Müller & Schäfer, 2001, 2004

  General condition

  $$\Pr_{H_0}(\text{reject } H_0, \text{ new design} \mid \text{interim data})$$
  $$\leq \Pr_{H_0}(\text{reject } H_0, \text{ initial design} \mid \text{interim data})$$

  These probabilities are called

  conditional rejection probabilities (CRP)

a new flexible procedure for GWAS

- plan a GWAS with a reasonable design including a sample size calculation
- for a marker i define:

$$CRP_i(t^\gamma) := Pr_i(T_{i,n} \geq t^\gamma | \text{ interim data})$$

- for a subset of markers I of M_1 define:

$$CRP_I(t^\gamma) := \sum_{i \in I} CRP_i(t^\gamma)$$

a new flexible procedure for GWAS

- let M_2 be the marker set foreseen for genotyping in stage 2
- determine minimum $\omega := CRP_I(t^{\alpha/|I|})$ for all hypotheses that include the marker set M_2 for **closed testing** (Marcus et al., 1976)
- by sorting the interim test statistics of the set $M_1 \setminus M_2$ (from minimum to maximum)
- ω is calculated for the $|M_1|-|M_2|$ CRP sums with $|I|$ as set size of the currently evaluted set

a new flexible procedure for GWAS

- design modifications must fulfill the inequality

$$CRP_{M_2}(t^{\gamma}) \leq \omega$$

in order to control the FWER for the design as a whole

simulation study

marker	optimal design by Wang et al. (2006)						flexible design	
$\|M_1\|$ ($x10^3$)	α_1	α_2 ($x10^{-5}$)	n_1/n	n	**FWER**	**power**	**FWER**	**power**
1	0.091	7.200	0.257	2,504	**0.051**	**0.887**	**0.048**	**0.887**
10	0.078	0.720	0.224	2,994	**0.045**	**0.893**	**0.042**	**0.892**
30	0.070	0.240	0.209	3,292	**0.050**	**0.886**	**0.048**	**0.886**
100	0.064	0.072	0.200	3,536	**0.053**	**0.889**	**0.051**	**0.888**
500	0.056	0.015	0.187	3,894	**0.050**	**0.890**	**0.045**	**0.889**

- constant per genotype costs, case:control fraction 1
- one true disease marker:
  - allele frequency=0.2; odds ratio=1.5 (mult. model)
- power = 0.9; $\alpha_i = 0.05 / |M1|$ (two-sided)
- 10,000 replicates

discussion

- more flexibility, e.g.:
  - arbitrary selection criteria for the M_2 marker set
  - allows sample size modification in stage 2
- control of FWER
- can be combined with cost-optimal designs

references

- Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, West K, et al. (2006). A common genetic variant in the *NOS1* regulator *NOS1AP* modulates cardiac repolarization. Nat Genet 38(6):644–651.
- Cui L, Hung HM, Wang SJ (1999). Modification of sample size in group sequential clinical trials. Biometrics 55(3):853–857.
- Müller HH, Schäfer H (2001). Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. Biometrics 57(3):886–891.
- Müller HH, Schäfer H (2004). A general statistical principle for changing a design any time during the course of a trial. Stat Med 23(16):2497–2508.

references

- Marcus R, Peritz E, Gabriel KR (1976). On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63(3):655–660.
- Scherag A, Schäfer H, Müller HH (*submitted*). Flexible two-stage testing in genome-wide association studies more flexibility.
- Wang H, Thomas DC, Pe'er I, Stram DO (2006). Optimal two-stage genotyping designs for genome-wide association scans. Genet Epidemiol 30(4):356–368.

Thank you for your attention!