

Selection of the Most Informative Individuals From Families With Multiple Siblings for Association Studies

Chunyu Liu, Qiong Yang, L. Adrienne Cupples, James B. Meigs
and Josée Dupuis

Presenter: Jingyuan Yang
Statistical Genetics Journal Club
Department of Statistics
The Ohio State University

May 1, 2009

Introduction

Methods

Simulations

Application to FHS Data

Goal

Utilize **linkage information** and **quantitative trait** values to select the most informative individuals for association studies for which follow up a linkage analysis and are for fine mapping.

Score Statistic

$$S(j) = \sum_{i \neq j} \frac{(x_i - \mu)(x_j - \mu)(\hat{\pi}_{i,j} - \frac{1}{2})}{N_{sib} - 1},$$

where

- ▶ the summation is over sibling pairs that include sibling j ;
- ▶ x_i and x_j are phenotypic values of the two siblings in the pair;
- ▶ $\hat{\pi}_{i,j}$ is the estimated multipoint IBD sharing for siblings i and j in a sibship at the locus with MLS;
- ▶ N_{sib} is the number of siblings in the sibship.

Score Statistic

$$S(j) = \sum_{i \neq j} \frac{(x_i - \mu)(x_j - \mu)(\hat{\pi}_{i,j} - \frac{1}{2})}{N_{sib} - 1},$$

where

- ▶ the summation is over sibling pairs that include sibling j ;
- ▶ x_i and x_j are phenotypic values of the two siblings in the pair;
- ▶ $\hat{\pi}_{i,j}$ is the estimated multipoint IBD sharing for siblings i and j in a sibship at the locus with MLS;
- ▶ N_{sib} is the number of siblings in the sibship.

Tend to select a sibling

- (1) with large absolute trait value and high IBD sharing with other siblings with similar trait values
- (2) with large absolute trait value and low IBD sharing with other siblings with dissimilar trait values

Six Selection Strategies

- (1) Randomly select one sibling from each sibship (AR)
- (2) Select the sibling with the highest absolute phenotype value $|x_j|$ (EV)
- (3) Select the sibling with the highest $S(j)$ score, using population mean (IBD-EV)
- (4) Select the sibling with the highest $S(j)$ score, using sibship-specific mean (IBD-EV2)
- (5) Select the sibling with the highest absolute phenotype value, restricted to linked families (ML-EV)
- (6) Select the sibling with the highest $S(j)$ score, restricted to linked families (ML-IBD-EV)

In addition, there is a reference group that consists of all siblings

Tests for Association

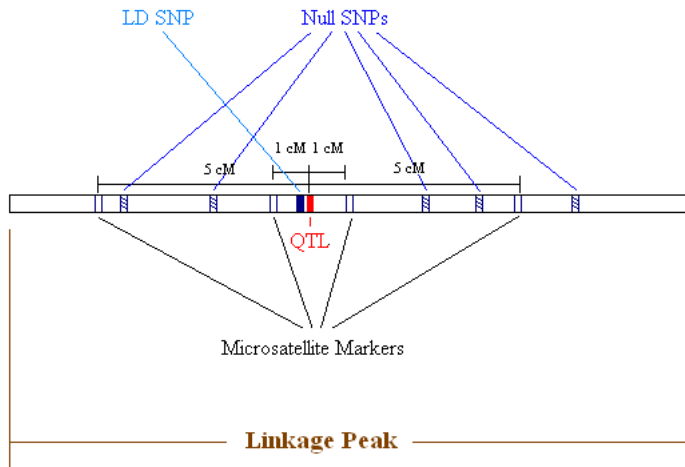
For selected siblings:

- ▶ ANOVA - F test of the difference among the phenotype means of different genotype groups
- ▶ Permutation test - randomly shuffle phenotypic values to obtain an estimate of null distribution

For all siblings:

- ▶ Generalized estimating equations (GEE) - account for correlation among siblings within the same sibship

Simulation Settings



Simulation Settings

Three types of genetic model: additive ($d=0$), dominant ($d=a$), and recessive ($d=-a$).

Two types of sample:

▶ Homogeneous:

- ▶ a QTL with 20% or 30% total heritability; $p=0.1$, 0.2 and 0.4 for additive and dominant models and $p=0.3$, 0.4 and 0.7 for recessive models

▶ Heterogeneous:

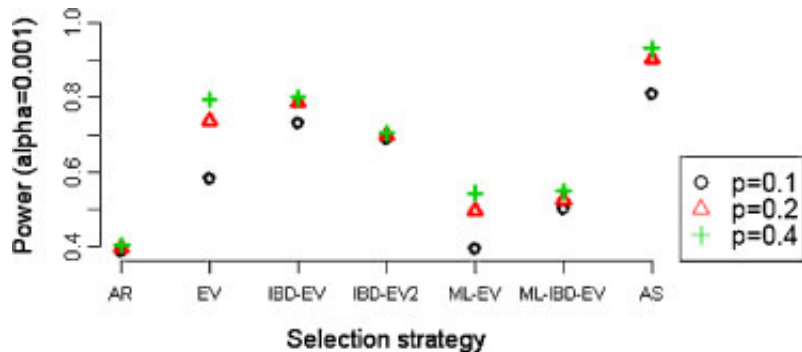
- ▶ Het 1-3: a QTL with 40% total heritability is present in subpopulation 1; no QTL effect in subpopulation 2;
- ▶ Het 4: four equally sampled subpopulations with a QTL effect explaining 0,10,20 and 40% of the total variance in subpopulations 1-4

Type I Error Rate

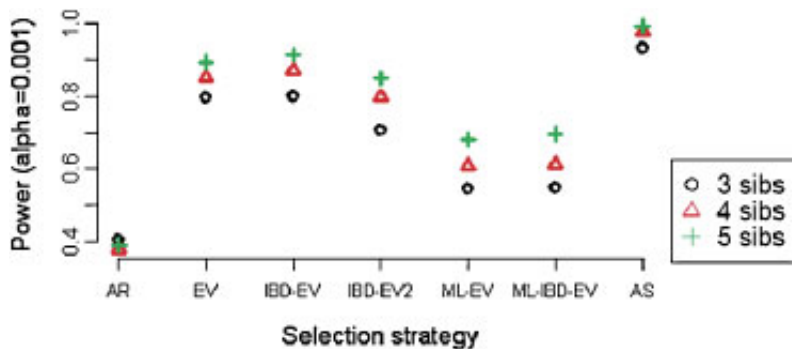
TABLE IV. Type I error (%) of the five selection strategies and the approach that uses all siblings for five SNP markers not associated with the quantitative trait

	α	AR	EV	IBD- EV	ML- EV	ML-IBD- EV	AS (GEE)
SNP1	0.10%	0.12	0.10	0.09	0.06	0.06	0.23
SNP2		0.13	0.07	0.11	0.10	0.10	0.12
SNP3		0.11	0.09	0.10	0.11	0.11	0.15
SNP4		0.11	0.09	0.10	0.11	0.11	0.13
SNP5		0.11	0.10	0.10	0.10	0.10	0.11
SNP1	1.00%	0.96	0.97	0.98	0.91	0.91	1.45
SNP2		1.03	1.00	0.94	0.99	0.99	1.21
SNP3		0.96	0.90	0.93	0.96	0.96	1.20
SNP4		0.97	0.96	1.02	1.05	1.06	1.16
SNP5		0.99	0.99	1.02	1.01	1.01	1.10
SNP1	5.00%	5.02	4.91	4.95	4.89	4.89	6.03
SNP2		5.13	5.00	4.94	4.95	4.95	5.39
SNP3		4.87	4.88	4.83	4.88	4.87	5.51
SNP4		4.89	5.09	5.04	5.03	5.03	5.35
SNP5		5.05	5.00	4.95	5.02	5.02	5.26

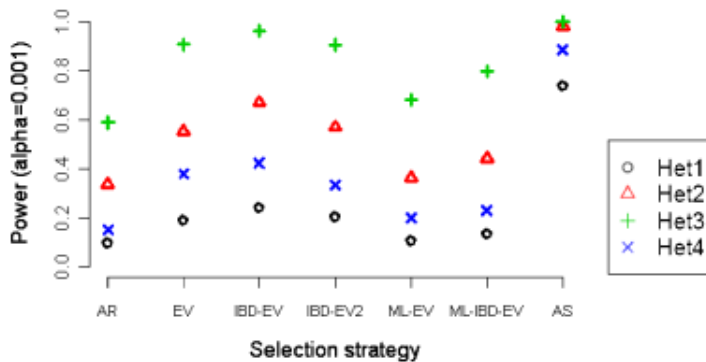
Power - Homogeneous Sample



Power - Homogeneous Sample



Power - Heterogeneous Sample



Two Best Selection Strategies

Select one sibling per sibship with multiple siblings:

- (1) One sibling with the most extreme trait value (EV)
- (2) One sibling using a combination score statistic based on trait values and IBD-sharing information (IBD-EV)

Analysis

- ▶ Quantitative trait: Fasting plasma glucose (FPG)
- ▶ Region: 72-77cM on chromosom 10
- ▶ Sibships: 441 sibships with two or more siblings with phenotypes and genotypes available.
- ▶ SNPs: 467 SNPs with an average distance 289.5 kb.
- ▶ Selection strategy: (1) AR (2) EV (3) IBD-EV

Result

TABLE V. The 10 smallest *P*-values from EV, IBD-EV strategies and full sample (GEE) within 1.0 - LOD support interval for linkage peak on chromosome 10 with fasting plasma glucose at exam 5

SNP ID	IBD-EV	IBD-EV Permutation	EV	EV Permutation	AR	GEE	Known Gene
(a) Ordered by IBD-EV <i>P</i> value							
rs4074715	0.0005	0.0011	0.0002	0.0002	0.2215	0.0228	CTNNA3 (intron)
rs10509276	0.0022	0.0026	0.0008	0.0013	0.3289	0.0859	CTNNA3 (intron)
rs5030938	0.003	0.0023	0.0237	0.0236	0.3181	0.1451	HKDC1/HK1 (flanking)
rs2043089	0.0036	0.0036	0.0047	0.0042	0.1622	0.0288	
rs4082516	0.0053	0.0061	0.0118	0.0119	0.3594	0.0955	COL13A1 (intron)
rs876705	0.0067	0.0065	0.0059	0.0057	0.1405	0.0113	
rs10509378	0.0085	0.0094	0.0681	0.0664	0.3365	0.2838	KCNMA1 (intron)
rs10509379	0.0112	0.0106	0.056	0.0526	0.3987	0.553	KCNMA1 (intron)
rs3851252	0.0113	0.0106	0.0057	0.0061	0.3053	0.1041	
rs780654	0.0126	0.0137	0.0037	0.0043	0.281	0.0065	SLC29A3 (intron)
(b) Ordered by EV <i>P</i> value							
rs4074715	0.0005	0.0011	0.0002	0.0002	0.2215	0.0228	CTNNA3 (intron)
rs10509276	0.0022	0.0026	0.0008	0.0013	0.3289	0.0859	CTNNA3 (intron)
rs4074716	0.0191	0.0173	0.0015	0.0016	0.3236	0.1131	CTNNA3 (intron)
rs4622198	0.014	0.0137	0.0016	0.0017	0.3138	0.1234	KCNMA1 (intron)
rs1516510	0.019	0.0199	0.0026	0.0026	0.3378	0.158	KCNMA1 (intron)
rs780654	0.0126	0.0137	0.0037	0.0043	0.281	0.0065	SLC29A3 (intron)
rs2043089	0.0036	0.0036	0.0047	0.0042	0.1622	0.0288	
rs3851252	0.0113	0.0106	0.0057	0.0061	0.3053	0.1041	
rs876705	0.0067	0.0065	0.0059	0.0057	0.1405	0.0113	
rs2140391	0.0589	0.0565	0.0099	0.008	0.2739	0.0093	
rs1880065	0.0589	0.0565	0.01	0.008	0.2744	0.0097	
(c) Ordered by GEE <i>P</i> value							
rs780654	0.0126	0.0137	0.0037	0.0043	0.281	0.0065	SLC29A3 (intron)
rs2140391	0.0589	0.0565	0.0099	0.008	0.2739	0.0093	
rs1354038	0.068	0.0656	0.1175	0.1183	0.2639	0.0096	
rs1880065	0.0589	0.0565	0.01	0.008	0.2744	0.0097	
rs2140390	0.0589	0.0565	0.01	0.0082	0.2746	0.0097	
rs7895188	0.0499	0.0477	0.0368	0.0349	0.2132	0.0105	DLG5 (intron)
rs876705	0.0067	0.0065	0.0059	0.0057	0.1405	0.0113	
rs1500737	0.0867	0.0803	0.1438	0.1426	0.2753	0.0119	
rs4399260	0.0274	0.0256	0.0367	0.0396	0.1755	0.012	
rs2120989	0.1735	0.1706	0.3323	0.3287	0.4128	0.012	CCDC6 (intron)