

Detection of SNP-SNP Interactions in Case-Parent Trios

Li Q / Louis TA / Fallin MD / Pulver AE / Ruczinski I

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

October 20, 2009

TDT - Allelic

The transmission disequilibrium test measures the over-transmission of an allele from parents to affected offsprings. For a set of n parents with alleles 1 and 2 at a genetic locus, each parent can be summarized by the transmitted and the non-transmitted allele:

		Non-TA		Σ
		1	2	
TA	1	a	b	a + b
	2	c	d	c + d
Σ		a + c	b + d	2n

Only the heterozygous parents contribute information!

Under the null of no association, $\frac{(b-c)^2}{b+c} \sim \chi_1^2$

→ Even better, use `binom.test()` in R.

TDT - Genotypic

Assume that at a certain locus the father has alleles 11 and the mother has alleles 12. The four *Mendelian children* thus have alleles 11, 12, 11, and 12.

Assume the affected proband has genotype 11.

The three *Pseudo controls* then have the genotypes 11, 12, and 12.

	Y	X
Affected proband	1	11
Pseudo control #1	0	11
Pseudo control #2	0	12
Pseudo control #3	0	12

We can use conditional logistic regression to analyze the data.

Schizophrenia Study

- Case-parent trios of Ashkenazi Jewish descents.
- Diagnosis of SZ and SZA based on DSM IV.
- Dense coverage of 64 candidate genes.
- 375 SNPs on 11 chromosomes genotyped for 312 trios.
- Original analysis through single marker allelic TDT.

→ Fallin MD et al (2005), *Am J Hum Genet* 77(6): 918-36.

Goal:

Explore SNP-SNP interactions for association with SZ and SZA.

	BB	Bb	bb
AA			
Aa			
aa			

$$(\text{SNP}^A^D \wedge \text{SNP}^B^R) \vee (\text{SNP}^B^D \wedge \text{SNP}^A^R)$$

→ Statistical interaction:

Deviation from additivity in a linear statistical model.

→ Epistasis:

Masking of phenotype expressed by one gene by the effects of another gene.

How can we detect models such as

$$\text{logit}(p) = \alpha + \beta \times \text{Ind}\{(\text{SNP}^{24^D} \wedge \text{SNP}^{80^R}) \vee (\text{SNP}^{24^R} \wedge \text{SNP}^{80^D})\} \dots?$$

Logic Regression

- The predictors are the SNPs in dominant and recessive coding.

SNP X	X.R	X.D
AA	0	0
AT	0	1
TT	1	1

- Let X_1, \dots, X_k be binary (0/1) predictors, Y a response variable.
- Fit a model

$$g(E(Y)) = b_0 + \sum_{j=1}^t b_j \cdot L_j,$$

where L_j is a Boolean combination of the covariates, e.g.

$$L_j = (X_1 \vee X_2) \wedge X_4^c.$$

- Determine the logic terms L_j and estimate the b_j simultaneously.

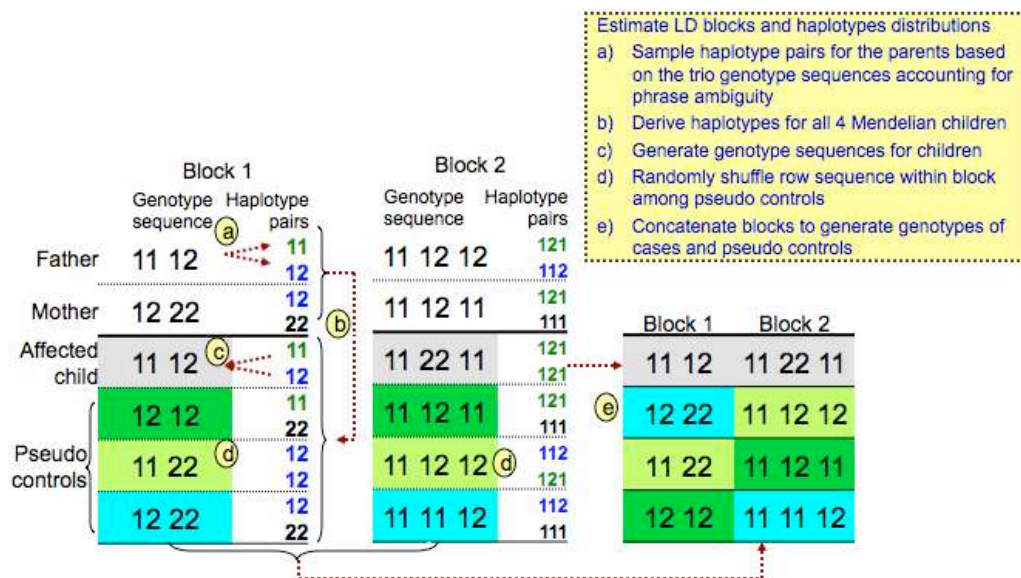
Trio Logic Regression

The rough idea is as follows:

- Pseudo controls are generated from the trio data, taking the LD block structure into account.
- Missing data are handled using haplotype-based imputation.
- The conditional logistic likelihood is used in logic regression to assess differences in cases and pseudo controls (just like in the genotypic TDT).

A tech report for the [Methods](#) paper is available, please email ingo@jhu.edu

Trio Logic Regression



Trio Logic Regression

The steps in more detail:

- 1 Estimate the haplotype blocks and the haplotype frequencies using the parents' genotypes.
- 2 For each block and each trio, sample haplotype pairs for the parents and the offspring consistent with the observed genotypes in the trio, allowing for missing data.
- 3 Generate the probands genotype data from the haplotypes that were passed from the parents.
- 4 For each block and each trio, generate genotypes for three pseudo-controls (PC1, PC2, PC3) using the parents' haplotypes that were not passed to the proband. The assignment to PC1, PC2, and PC3 is random.
- 5 Assemble three pseudo-controls for each trio by augmenting the genotypes from the blocks.
- 6 For each locus, translate the genotype data into two binary variables in dominant and recessive coding.

Software available soon!

- 1 The trio logic regression methods are implemented as an augmentation in the logic regression R package `LogicReg`.
- 2 The R package `trio` contains functions to generate logic regression input from pedigree or genotype files, to check for Mendelian errors, to impute missing data, and to simulate case-parent trios.
- 3 A software vignette is also available.

Set up data for trio logic regression or simulate trio data for high order SNP-SNP interaction



Documentation for package 'trio' version 1.0

Help Pages

[trio](#)
[trio.check](#)
[trio.sim](#)

Generate Trio Data Format Suitable for Trio Logic Regression
Check Case-Parent Trio Data for Mendelian Errors
Simulate Case-Parent Trios with Population Disease Risk Dependent on SNP-SNP Interaction

Results

	Logic model	$\exp(\hat{\beta})$
1	$0.67 \times I_{\{302^D\}}$	1.94
2	$0.89 \times I_{\{302^D \vee 166^D\}}$	2.43
3	$1.15 \times I_{\{302^D \vee 166^D \vee 148^D\}}$	3.14
4	$1.30 \times I_{\{302^D \vee 166^D \vee 148^D \vee 368^R\}}$	3.65

SNP 302	Chromosome 12	NOS1	3782219
SNP 166	Chromosome 8	CHRN3	1530848
SNP 148	Chromosome 8	PNOC	3735736
SNP 368	Chromosome 22	COMT	740603

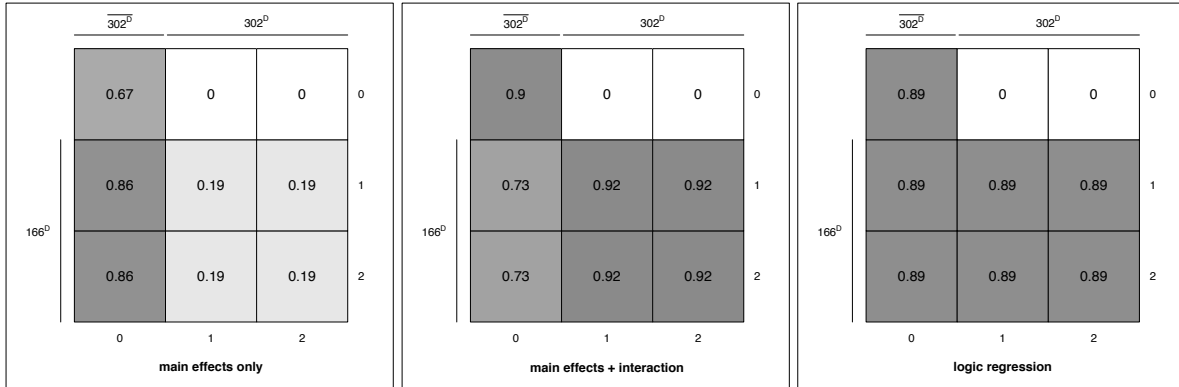
Results

		$\exp(\hat{\beta})$	$\hat{\beta}_{(se)}$	z	p
Marginal	302^D	1.94	0.67 _(0.16)	4.12	4e-05
	166^D	1.17	0.16 _(0.22)	0.71	0.480
Logic	$302^D \vee 166^D$	2.43	0.89 _(0.18)	4.89	1e-06
Additive	302^D	1.95	0.67 _(0.16)	4.15	3e-05
	166^D	1.21	0.19 _(0.22)	0.86	0.390
Additive	302^D	2.46	0.90 _(0.19)	4.86	1e-06
	166^D	2.52	0.92 _(0.33)	2.77	0.006
	$302^D : 166^D$	0.34	-1.09 _(0.34)	-2.87	0.004

Results

$$0.90 \times I_{\{\overline{302^D}\}} + 0.92 \times I_{\{166^D\}} - 1.09 \times I_{\{\overline{302^D}:166^D\}}$$

↓



↑

$$0.67 \times I_{\{\overline{302^D}\}} + 0.19 \times I_{\{166^D\}}$$

↑

$$0.89 \times I_{\{\overline{302^D} \vee 166^D\}}$$

Results

For the three-way interaction model we get

		$\exp(\hat{\beta})$	$\hat{\beta}_{(se)}$	z	p
Marginal	$\overline{302^D}$	1.94	0.67 _(0.16)	4.12	4e-05
	166^D	1.17	0.16 _(0.22)	0.71	0.480
	148^D	1.54	0.43 _(0.25)	1.70	0.088
Logic	$\overline{302^D} \vee 166^D \vee 148^D$	3.14	1.15 _(0.20)	5.67	2e-08

The manuscript for this [Application](#) paper is available, please email ingo@jhu.edu

<http://biostat.jhsph.edu/~iruczins/>
ingo@jhu.edu