Linkage in selected samples

Manuel Ferreira

QIMR

Boulder Advanced Course 2005

1. Which nonparametric linkage statistics to use?
First, type-I error. Then, power.
Influenced by ascertainment, measurement scale, normality.

2. How to estimate the type-I error empirically?
Gene-dropping simulations.

Family ascertainment

Random ascertainment

Selective ascertainment

Proband selection

Concordant affected sib pairs

Discordant sib pairs

Mixture

More complex

Nonparametric linkage statistics

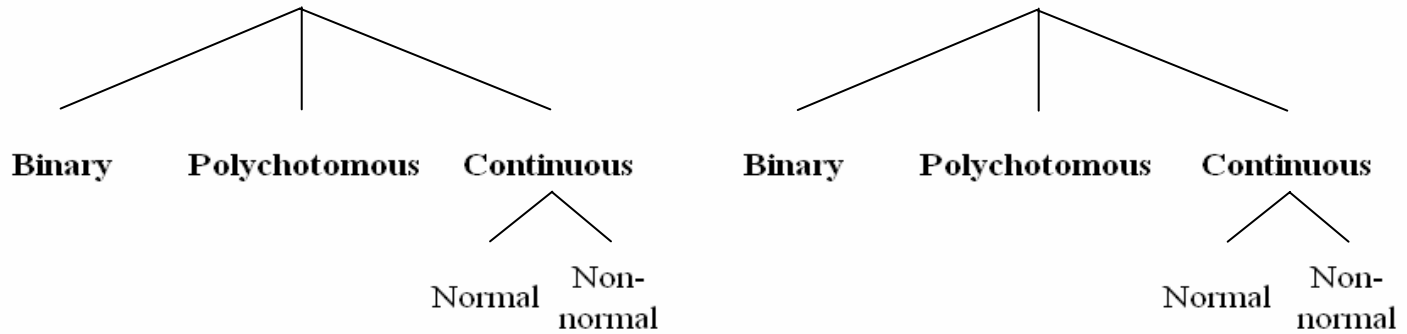References for different Nonparametric linkage statistics

HE	Haseman & Elston 1972, Behav Genet. 1972 Mar;2(1):3-19
VC (binary, polych.)	Duggirala et al. 1997, Genet Epidemiol 14:987-992
VC (continuous)	Almasy & Blangero 1998 Am J Hum Genet 62:1198-1211
Mean IBD	Risch & Zhang 1995, Science. 1995 Jun 16;268(5217):1584-9;
	Zhang & Risch 1996, Am J Hum Genet. 1996 Oct;59(4):951-7
MERLIN-REG	Sham et al. 2002, Am J Hum Genet. 2002 Aug;71(2):238-53
VC-R	Sham et al. 2000, Genet Epidemiol. 2000;19 Suppl 1:S22-8.
NPL	Whittemore & Halpern 1994, Biometrics. 1994 Mar;50(1):118-27;
	Kong & Cox 1997, Am J Hum Genet. 1997 Nov;61(5):1179-88
F&F	Forrest & Feingold 2000, Am J Hum Genet. 2000 May;66(5):1642-60
VC-AC	de Andrade & Amos 2000, Genet Epidemiol. 2000 Dec;19(4):333-44

Software implementing different Nonparametric linkage statistics

HE	GENEHUNTER, QTL EXPRESS, MX
VC (binary)	SOLAR, MX
VC (polychotomous)	SOLAR, MX
VC (continuous)	MERLIN, SOLAR, GENEHUNTER, MX
Mean IBD	GENEHUNTER, MX
MERLIN-REG	MERLIN-REGRESS, MX
VC-R	MX
NPL	GENEHUNTER, MERLIN, MX
F&F	MX
VC-AC	SOLAR, MX

Reviews

Feingold 2002, Am J Hum Gen 71, 217–222
Posthuma et al 2003, Twin Research, 6, 361–376
Ferreira 2004, Twin Research, 7(5), 513–530

	RANDOM ASCERTAINMENT				SELECTED SAMPLE			
	Binary	Polychotomous	Continuous Normal	Continuous Non-normal	Binary	Polychotomous	Continuous Normal	Continuous Non-normal
HE	-	-	✓	✓	-	-	✓	✓
VC	✓	✓	✓	✗	✗	✗	✗	✗
Mean IBD	✓	-	-	-	✓	-	-	-
MERLIN-REG	-	-	✓	✓	-	-	✓	✗
VC-R	?	?	✓	✓	?	?	✓	✓
NPL	✓	-	-	-	✓	-	-	-
F&F	-	-	✓	✓	-	-	✓	✓
VC-AC	✓	✓	✓	✗	✓	✓	✓	✗

Correct type-I error (~0.05)   Inflated type-I error (>0.05)

HE: Haseman-Elston (SD, SS, CP, COM); **VC**: Variance Components; **Mean IBD**: Mean IBD sharing statistic; **MERLIN-REG**: Pedwide-regression analysis ("reverse regression"); **VC-R**: Reverse Variance components; **NPL**: Non Parametric Linkage Analysis that is based on IBD scoring functions; **F&F**: Forrest & Feingold composite statistic; **VC-AC**: Variance Components with ascertainment correction.

Complex traits = nonparametric linkage analysis   Sibships only: HE, Mean IBD, NPL, Composite
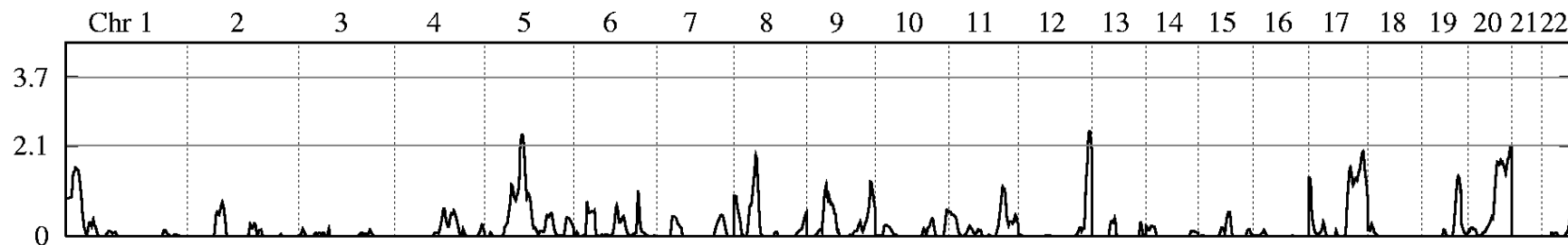
Assessing type-I error empirically

Selected sample (proband selection), Continuous trait, Slightly non-normal (k=1.5)

```
asthma.ped
22    1    3    4    2    0 151/155  133/133  237/237  226/230  175/179  251/257  248/256  107/111    1.540
22    2    3    4    2    0 151/151  125/133  237/237    0/  0  175/185  243/251  248/248  109/111    1.890
22    3    0    0    1    0 151/155  125/133  237/237    0/  0  179/185  243/257    0/  0  107/109       x
22    4    0    0    2    0 151/153  133/133  237/243  226/226  173/175  243/251  248/256  109/111       x
65    1    3    4    2    0 149/151  125/129  243/245  226/232  175/181  243/255  248/256  107/113    2.140
65    2    3    4    2    0 149/151  125/129  245/247  226/232  175/181  243/255  248/256  107/109    2.070
65 |  3    0    0    1    0 151/151  129/133  245/247  226/232  175/181  243/243  248/260  107/107       x
65    4    0    0    2    0 149/155  125/133  243/245  226/226  175/181  243/255  256/260  109/113       x
65   51    3    4    1    0 149/151  125/133  243/245  226/226  175/181  243/243  256/260  107/113    1.600
```

▷ **1. Analyse your dataset**

merlin –d asthma.dat –p asthma.ped –m asthma.map --start:0 --grid:2 --vc --usecovariates

Assessing type-I error empirically

▷ **2. Generate replicate dataset with same phenotypes but unlinked genotypes**

Gene-dropping simulation

merlin –d asthma.dat –p asthma.ped –m asthma.map --simulate --save -r $seed

```
22    1    3    4    2    0                                              1.540
22    2    3    4    2    0                                              1.890
22    3    0    0    1    0                                              x
22    4    0    0    2    0                                              x
65    1    3    4    2    0                                              2.140
65    2    3    4    2    0                                              2.070
65 |  3    0    0    1    0                                              x
65    4    0    0    2    0                                              x
65   51    3    4    1    0                                              1.600
```
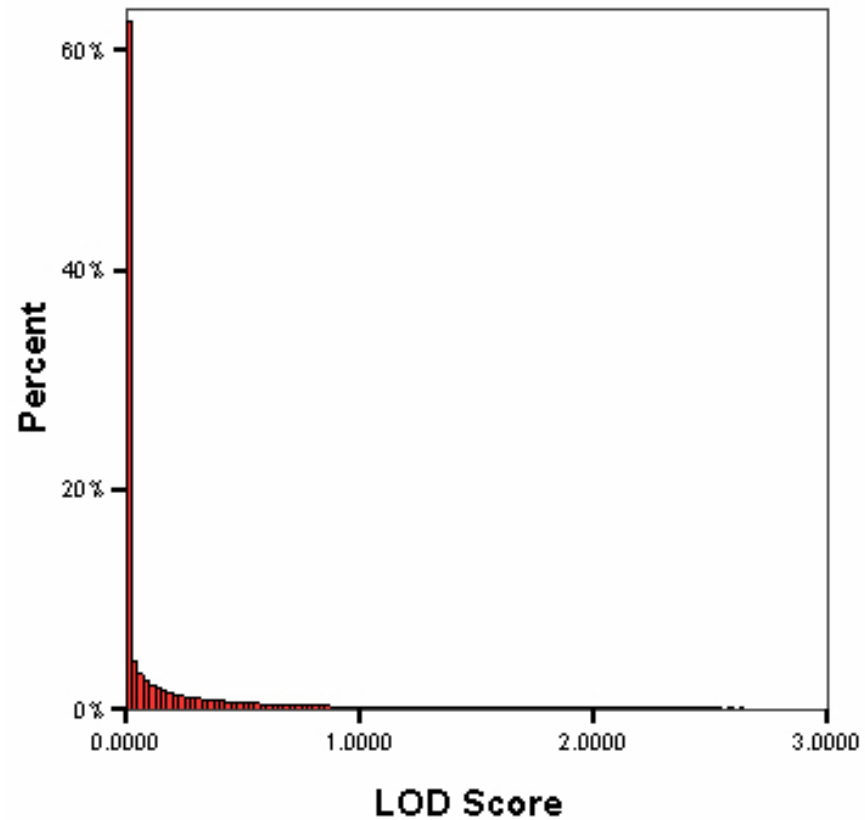
New Simulated Genotypic data

▷ **3. Analyse replicate dataset EXACTLY like the real dataset**

merlin –d replicate.dat –p replicate.ped –m replicate.map --start:0 --grid:2 --vc --usecovariates
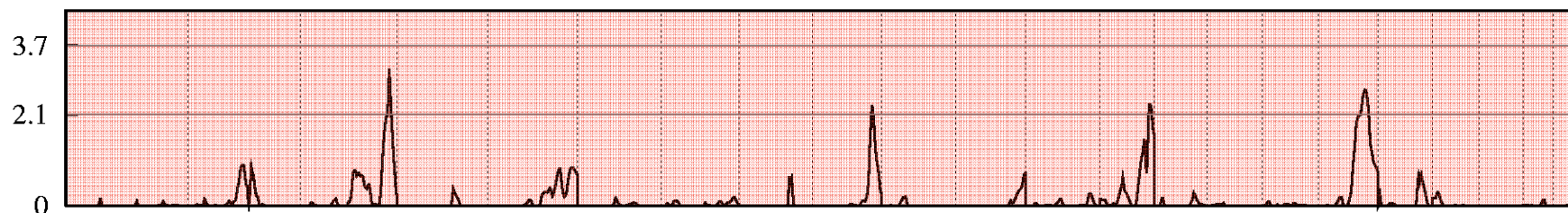
Assessing type-I error empirically

▷ **4. Repeat 2 and 3 e.g. 1000 times** (Shell script)

50% Point mass at zero and 50% χ^2_1

Assessing type-I error empirically

▷ **5. Estimate empirical genome-wide _P_ value for peak LOD score**

Replicate	Highest LOD
1	2.1
2	1.3
3	0.8
4	4.2
…	…
1000	1.3

_P_ = proportion of replicates with a highest LOD score ≥ your peak LOD score

e.g. LOD = 3.8, P = 20/1000 = 0.02 (Significant linkage, since < 0.05)

Assessing type-I error empirically

▷ **6. Estimate empirical genome-wide thresholds for suggestive and significant**

	Highest LOD						Count LODs					
Replicate	Chr1	Chr2	Chr3	Chr4	…	Chr22	≥ 0.5	≥ 1	≥ 1.5	≥ 2	…	≥ 3.5
1	2.1	0.5	1.2	3.1	…	1.4	12	9	3	1	…	1
2	0.2	1.3	2.5	1.8	…	1.3	11	6	1	0	…	0
3	0.8	2.1	2.6	0.1	…	0.8	12	7	3	2	…	0
4	3.1	0.6	2.0	1.3	…	4.2	10	5	2	1	…	0
…	…	…	…	…	…	…	9	2	0	0	…	0
1000	1.3	0.9	0.2	2.5	…	1.3	12	8	4	3	…	1
				Average (N=1000)			11.0	6.2	2.2	1.2	…	0.3

Assessing type-I error empirically
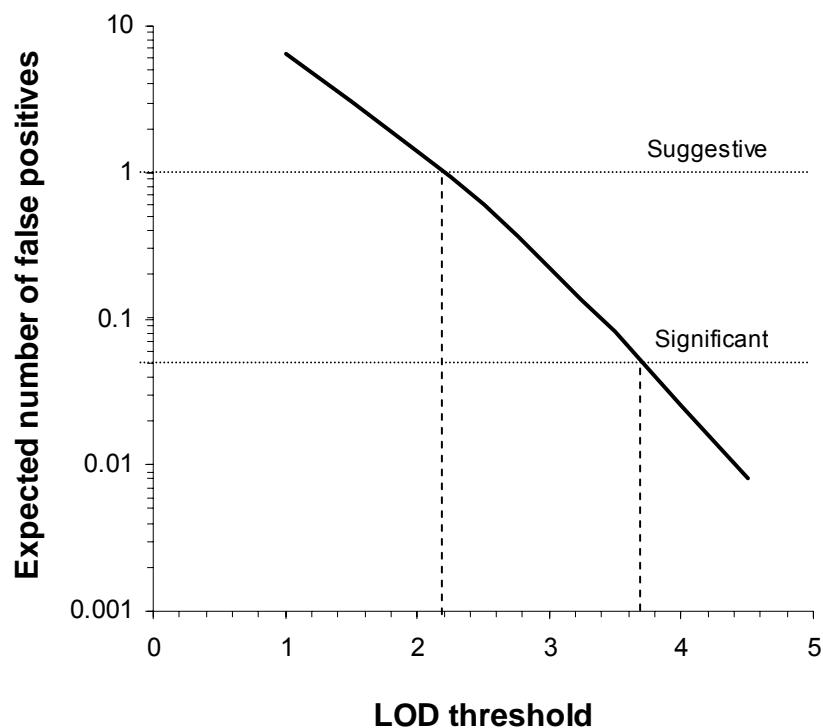
▷ <u>Genome-wide threshold for significant linkage</u> (e.g. 3.7)

LOD score that occurs by chance alone on average once per 20 scans

▷ <u>Genome-wide threshold for suggestive linkage</u> (e.g. 2.1)

LOD score that occurs by chance alone on average once per scan

References for Gene-dropping simulations

Kruglyak & Daly Am J Hum Genet. 1998 Apr;62(4):994-7

Abecasis et al. Am J Hum Genet 2004; 74: 403-417.