ibdreg:
An R package for Genetic Linkage with Covariates

Jason Sinnwell
Daniel Schaid

August 10, 2007

ibdreg: outline

- Intro to Linkage (brief)
- Why do linkage with covariates?
- Methodology in ibdreg
- Demo of ibdreg on prostate cancer data

Linkage Intro:
Genetic Analysis Types

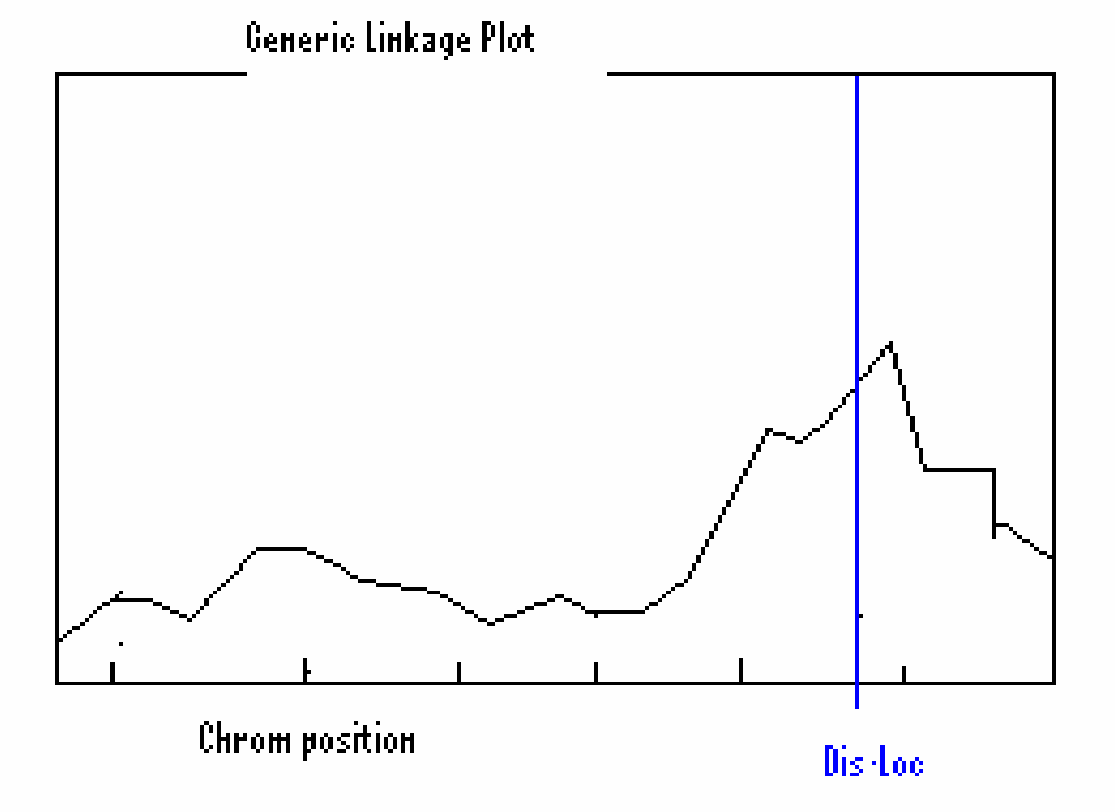
- Segregation Analysis - Use family data to determine the extent to which a disease is genetic
- Linkage Analysis – Use family data to locate a chromosome segment that is inherited jointly with a disease locus
- Association Analysis – Use unrelated data to find a genetic effect of a particular DNA marker(s) on a disease

Intro to Linkage: Recombination

- Mendel proposed genes (on peas) are inherited independently of one another
- Not quite right -- genes / markers which are "close" together show associations

- During meiosis, corresponding DNA segments can recombine in any location
- Recombination is somewhat constant over the whole genome
- In general, "close" markers have fewer recombinations

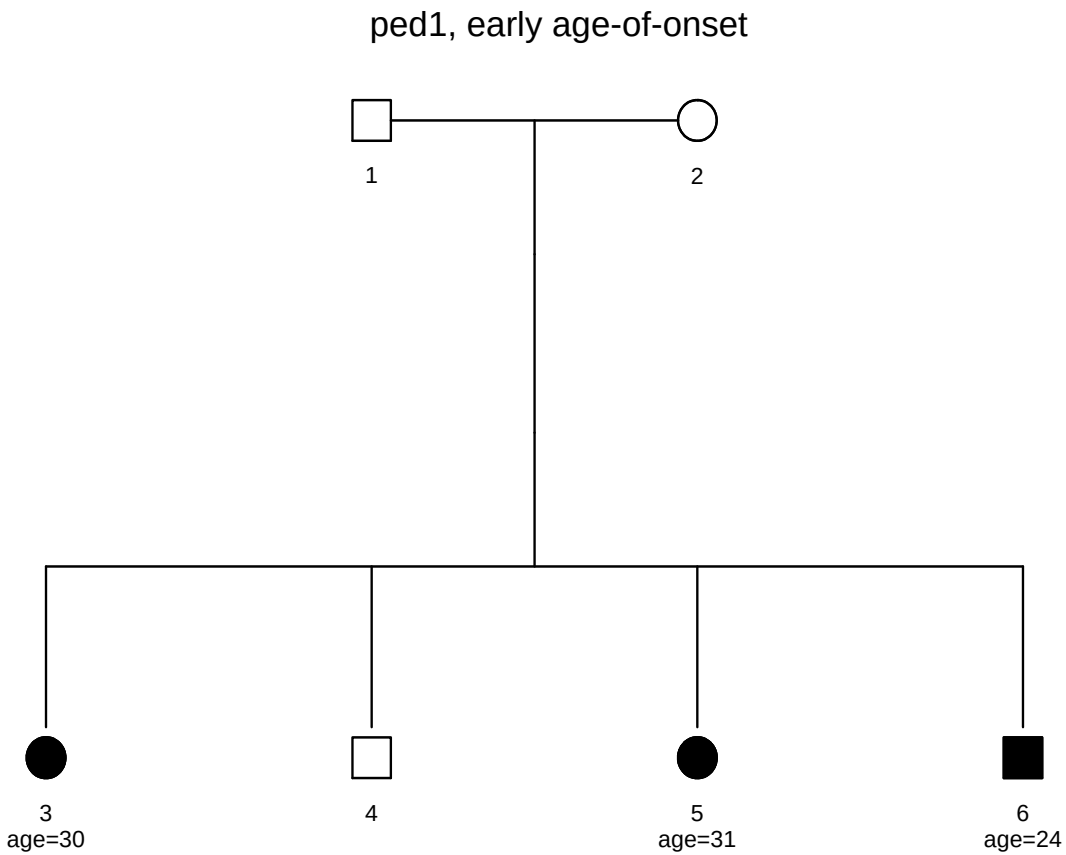
Intro to Linkage, basic plot



Linkage with Covariates: illustrated

Nuclear family with 4 children, 3 affected with disease

If genetic linkage:
- persons 3,5 share more alleles near disease locus
- persons 3,4 share fewer alleles at disease locus



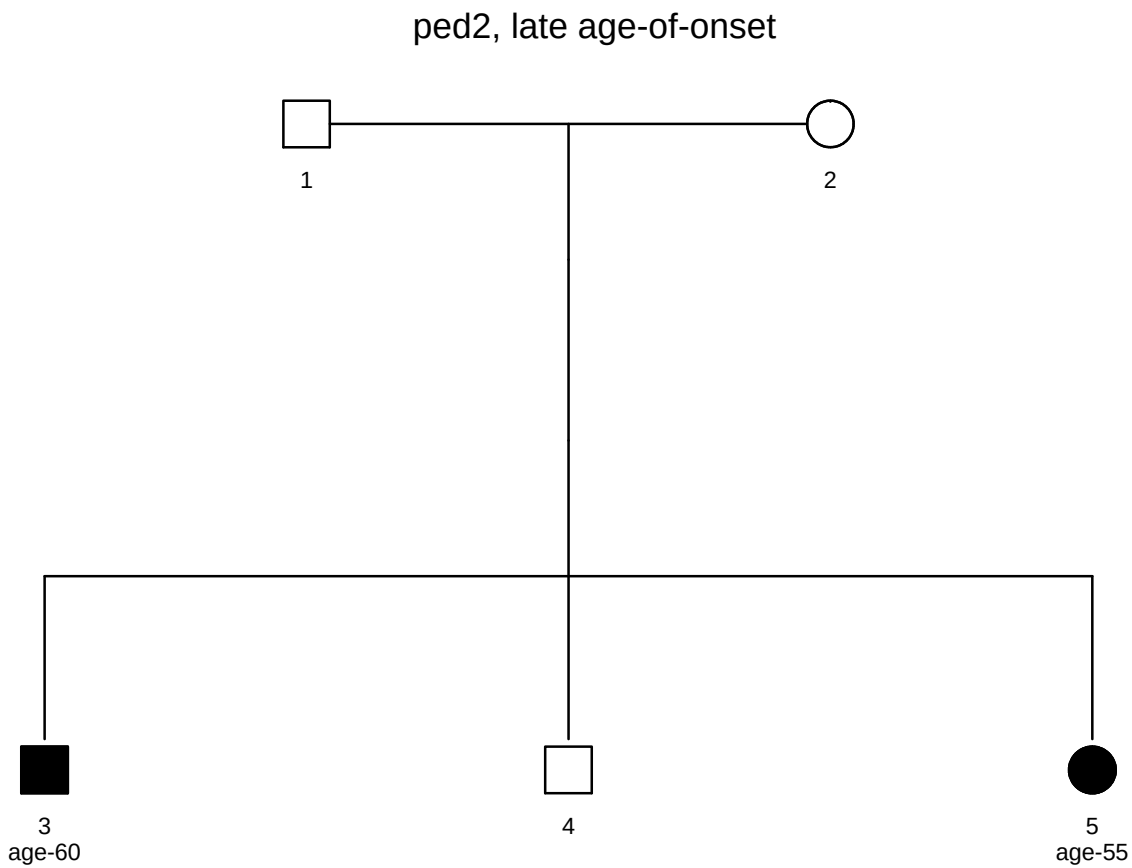
Linkage with Covariates: IBD Sharing

- Identity-By-Descent (IBD): alleles that have been inherited from the same ancestor
- Estimate probability of sharing 0, 1, and 2 alleles IBD: f_0, f_1, f_2
- Use Merlin, Genehunter, etc.
- Estimated IBD sharing for a relative pair: $$s_r = 2 f_{r,2} + 1 f_{r,1}$$

Linkage with Covariates: illustrated

Nuclear family, two affected siblings

Late-onset form of the disease



Linkage with Covariates

- Ped1 and Ped2 could be two forms of disease: early and late onset
- Mixture of gene and environment effects, i.e. disease heterogeneity
- Assume the early onset is genetically linked at locus **L**
- IBD sharing at **L** is:
  - ped1: greater than null
  - ped2: somewhere near null

Linkage with Covariates

- The more alike the covariates, the more allele sharing is expected
- Consider a function of the age of onset for affected siblings:
  1. age1 + age2
  2. age1 * age2
  3. $(|age1-60| * |age2-60|)^2$
- Define any function that applies to the disease and covariate impact

Linkage with Covariates

- Intuitively, affected relative pairs with similar covariates share more alleles than the null
- Linear relationship for IBD sharing and pair-specific covariates (X):

$$m_r = m_r^0 + c_r X\beta = X_r^* \beta$$

m_r, m_r^o: Expected s_r under linkage, no linkage

c_r = scaling factor for the relationship types

$$X_r^* = c_r[1 \mid X]$$

ibdreg method

Quasi-likelihood score function:

$$U = \sum_{i=1}^{n} D_i^T V_i^{-1} (S_i - M_i)$$

where

i = 1..n pedigrees

S_i, M_i: vectors of estimated and expected (H_o) allele sharing

D_i: derivative of M_i with respect to β

V_i: covariance matrix of the S_i vector

ibdreg method

Develop a chi-square test:

$$T = U^T V_u^{-1} U$$

where $V_u = \sum_{i=1}^{n} X_i^{*T} V_{0,i}^{-1} X_i^*$

With q degrees of freedom:

- q=1 for intercept-only (linkage)
- q=2 for intercept with 1 covariate
- etc.

ibdreg method

Available tests (degrees of freedom)

1. Linkage without covariates (q=1)
2. Linkage with p covariates (q=1+p)
3. Effect of p covariates on IBD sharing, adjusting for linkage (p)

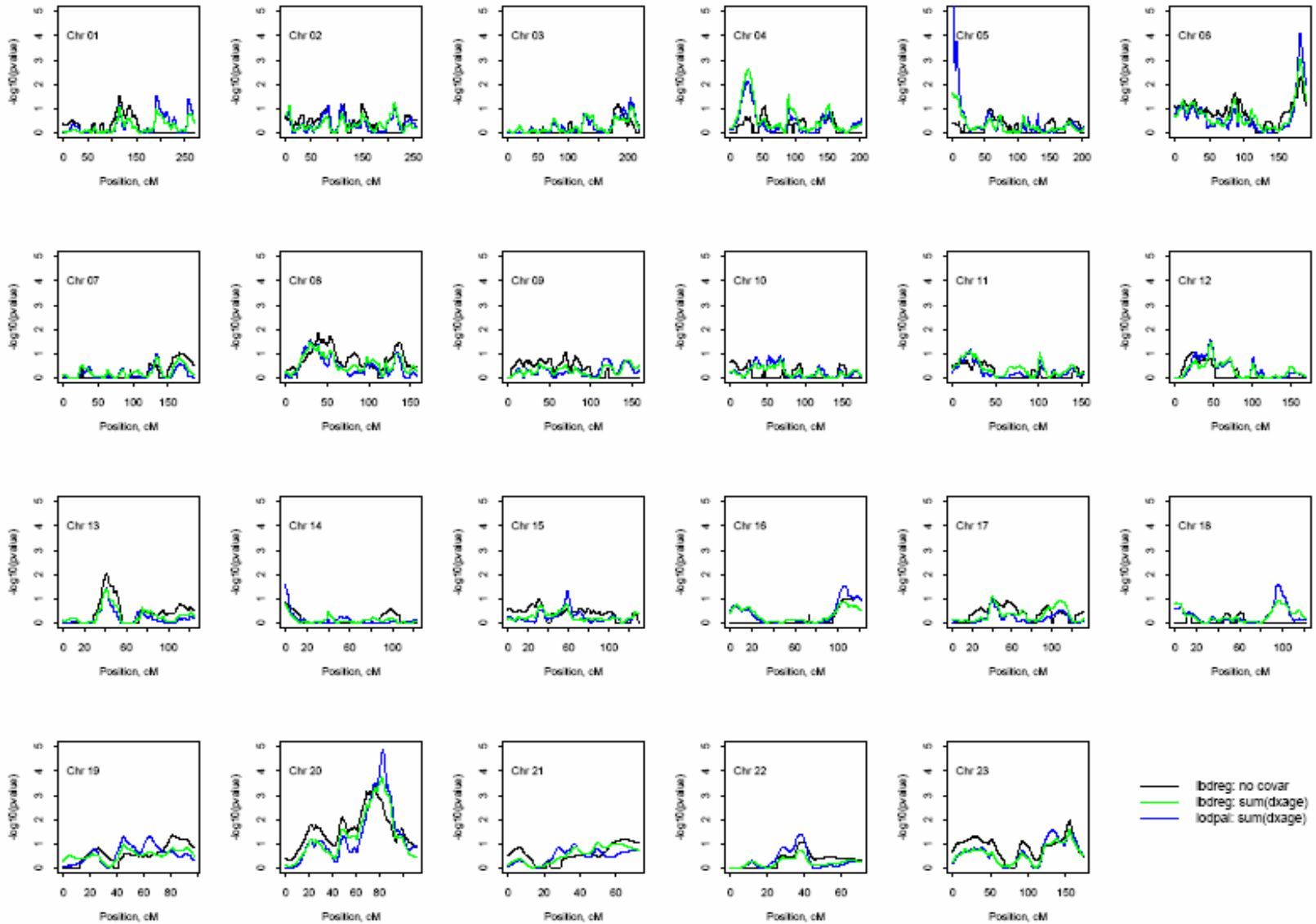
ibdreg applied

- Data: sample of pedigrees ascertained at Mayo Clinic (SE MN) for having multiple cases of prostate cancer
- 159 pedigrees (max 21 people)
- 495 affected relative pairs (ARPs)
- Covariate: age-of-onset (dxAge)

ibdreg applied

- LODPAL, of S.A.G.E., approximates a pseudo-likelihood of IBD sharing probabilities by a trinomial logistic regression model (**blue line**)
- Compare to ibdreg (**green line**), both using sum(dxAge) as a covariate
- Also include ibdreg, linkage-only (**black line**)
- Same c scale (genetic effect between dominant and recessive)

Figure 1


Chr 01
Chr 02
Chr 03
Chr 04
Chr 05
Chr 06
Chr 07
Chr 08
Chr 09
Chr 10
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15
Chr 16
Chr 17
Chr 18
Chr 19
Chr 20
Chr 21
Chr 22
Chr 23
-log10(pvalue)
Position, cM
lbdreg: no covar
lbdreg: sum(dxage)
lodpal: sum(dxage)

ibdreg applied

For figure above, note a few things:

- LODPAL and ibdreg generally close
- Differences:
  1. Narrow peaks (ch6, ch20)
  2. Ends of chromosomes (ch5)
- Both may be attributable to instability of maximizing pseudo-likelihood of LODPAL

ibdreg applied: R code

```
# load the library
R> library(ibdreg)

# load covariate data, containing ped id and person id to match file
R> cov.data <- read.table("cov.data.csv", sep=",")

# create data.frame with id (ped and person) and IBD data with Merlin results
R> ibd.dat.obj <-
create.ibd.dat(postfile="chr20.post.ibd",
                        priorfile="chr1.prior.ibd")

# simulate ibd.var object using "gene-dropping"
# requires a .pre file for pedigree specs
R> ibd.var.obj <- sim.ibd.var("ch.20.pre",
                           n.sim=10000)
```

ibdreg applied: R code

```
# define a function for the sum of covariates
R> pairSum <- function(cov1,cov2) {cov1+cov2}

# run ibdreg for AA relatives,
# with covariate sum(dxage), minimax c scaling
R> sum.dxage.AA <-
ibdreg(formula=~pairSum(dxage),
   status.method="AA", c.scale="minimax",
   status=pcstat, ped.id=ped.id,
   person.id=person.id, data=cov.data,
   ibd.dat=ibd.dat.obj, ibd.var=ibd.var.obj)
```

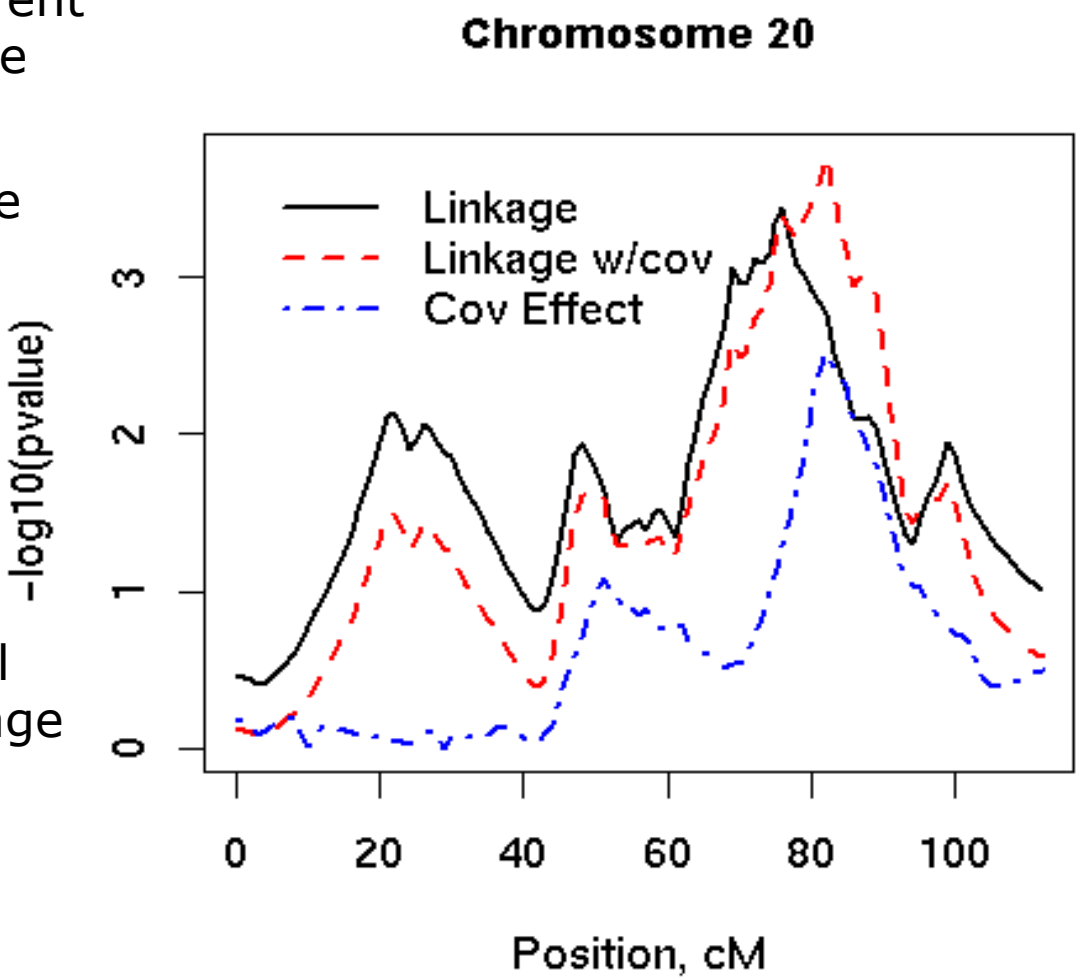
ibdreg applied: chromosome 20

Linkage signal different location from linkage w/cov

In non-peak, linkage w/cov is smaller

--1 more deg. of freedom

Cov Effect has small contribution to linkage signal



ibdreg applied: all ARPs

- Disregarding covariates, we can test linkage on all 495 relative pairs in the example, which contain sub-groups
  - AA: Affected-Affected (429)
  - UU: Unaffected-Unaffected (7)
  - AU: Affected-Unaffected (59)

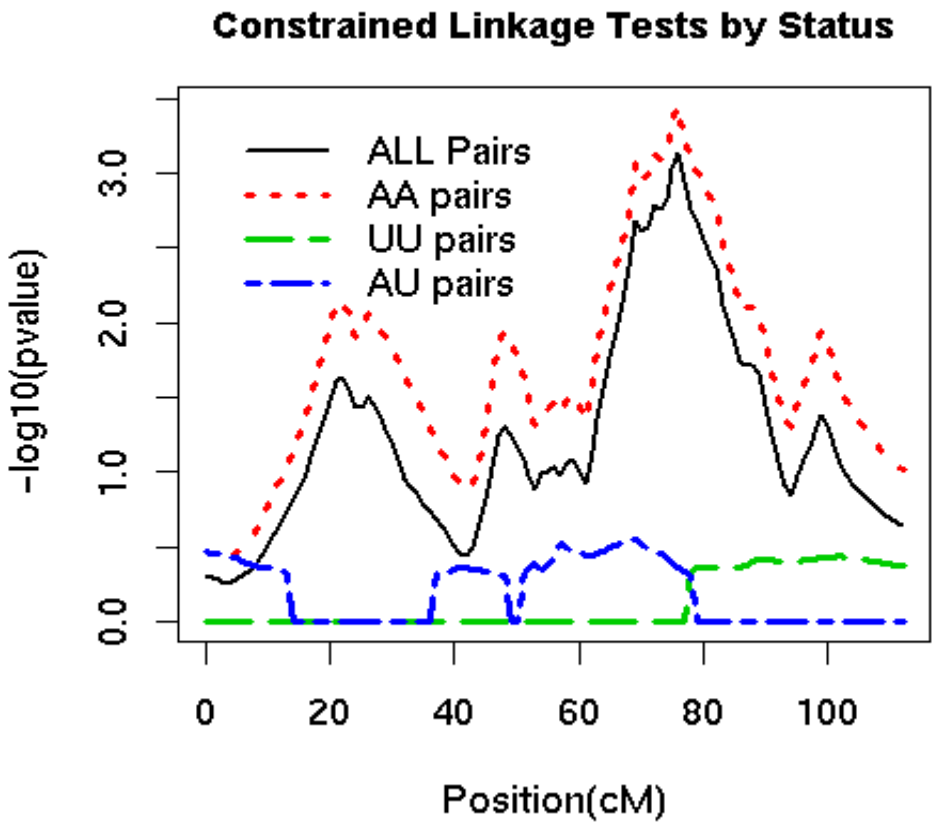
- Sharing under linkage is expected as
  - AA: More than null
  - UU: More than null, but less than AA
  - UU: Less than null

Ibdreg applied, all ARPs

Linkage tests for all relative pairs

Test for linkage in direction expected under linkage:
- AA-more
- UU-more
- AU-less

Overall linkage is driven by AA pairs



ibdreg applied: unexpected sharing

- IBD allele sharing can be invalid by:
  1. Mis-specify marker allele freq when parents missing
  2. Incorrect pedigree relationships
  3. Unaccounted inbreeding
- The χ^2 test for linkage given as both unconstrained and constrained to favor linkage (1-sided, smaller p-value)
- If $p_{unconstrained} < p_{constrained}$, indicates unexpected IBD sharing (chroms 1, 7, 12)

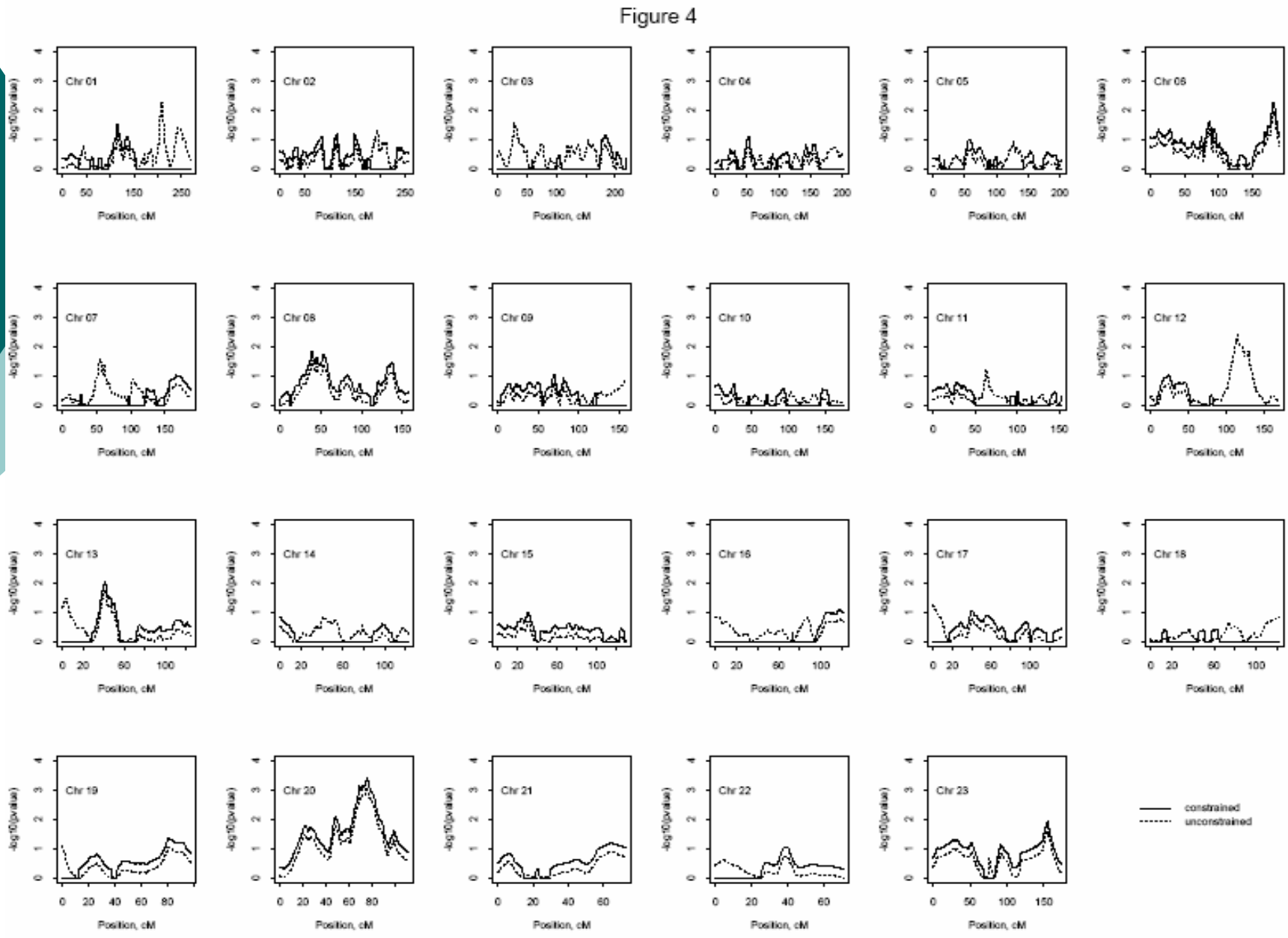


Figure 4
Chr 01
Chr 02
Chr 03
Chr 04
Chr 05
Chr 06
Chr 07
Chr 08
Chr 09
Chr 10
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15
Chr 16
Chr 17
Chr 18
Chr 19
Chr 20
Chr 21
Chr 22
Chr 23
-log10(pvalue)
Position, cM
constrained
unconstrained

ibdreg, concluded

- Allows multiple tests for linkage
- Provides options for specifying covariates and scaling (c_r)
- Scaling avoids fitting new regression line for each relationship type
- Linear regression is easy to fit, and easy to apply weights to account for dependence of variance on covariates
- Returned S3 object with print and plot methods
- Perl scripts supplied to work with Merlin.

Thank You

For more details and references see:

Schaid DJ, Sinnwell JP, Thibodeau SN. Testing genetic linkage with relative pairs and covariates by quasi-likelihood score statistics.
Hum Hered. 2007;64(4):220-33. Epub 2007 Jun 12.