A data mishap

Allele frequencies in sibships

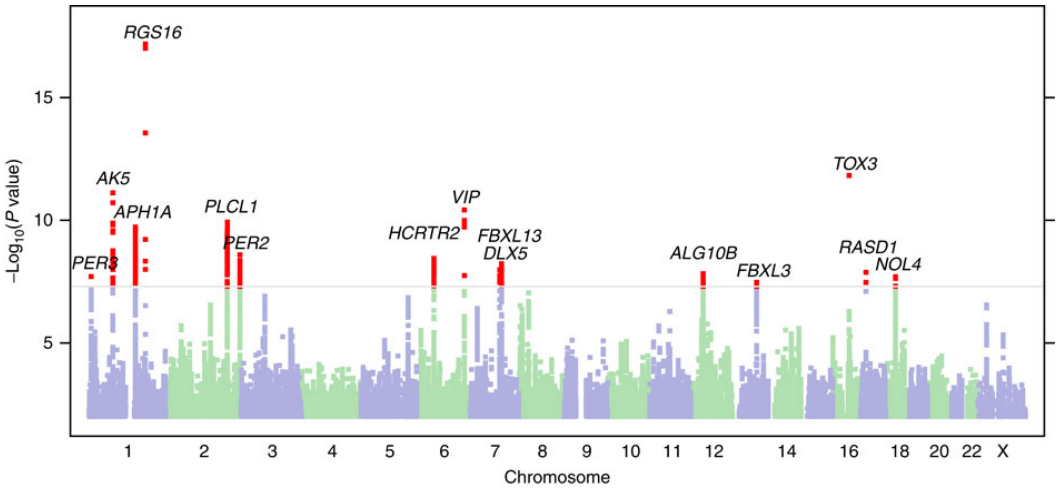
Karl Broman

Biostatistics & Medical Informatics, UW–Madison

kbroman.org
github.com/kbroman
@kwbroman

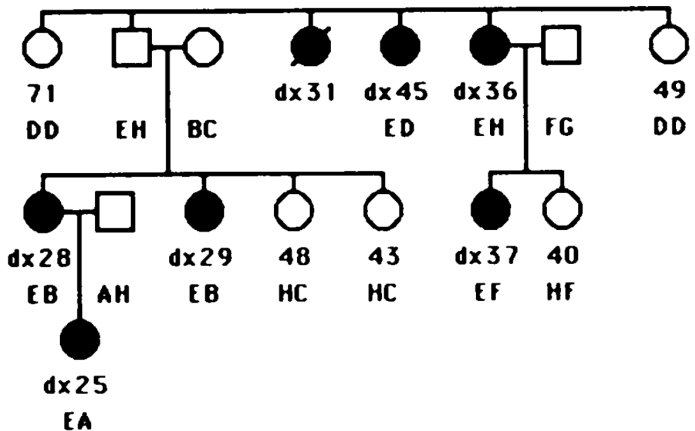
Slides: kbroman.org/Talk_DataMishap

GWAS for “morning person”



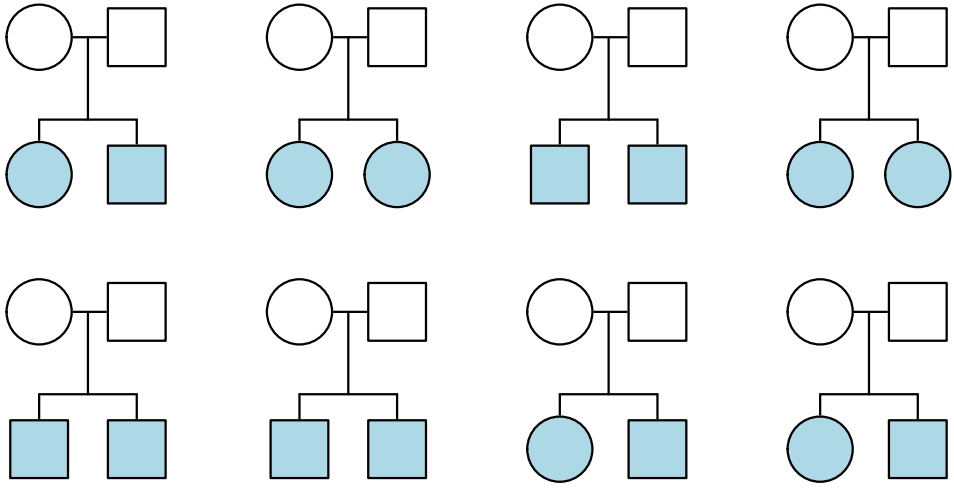
Hu et al (2016) doi:10.1038/ncomms10448

BRCA pedigree

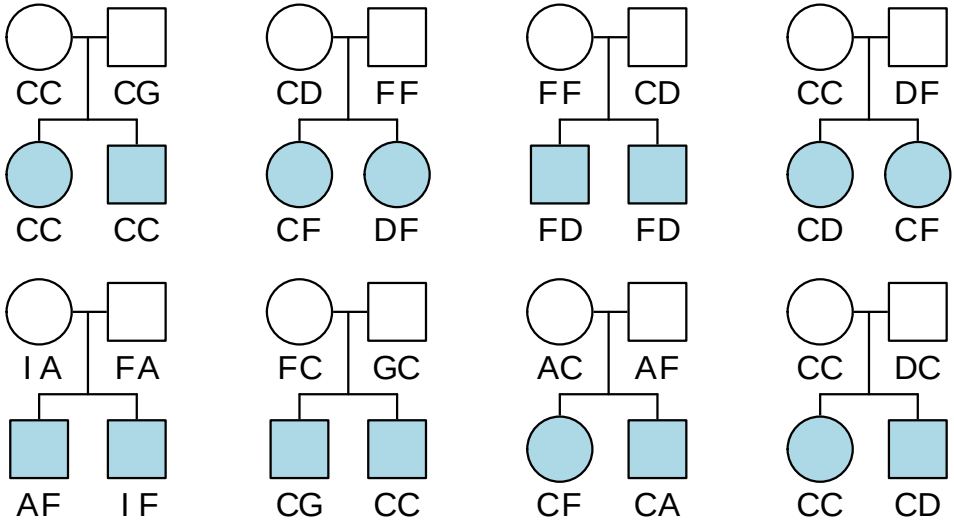


Hall et al (1990) doi:10.1126/science.2270482

Affected sib pairs

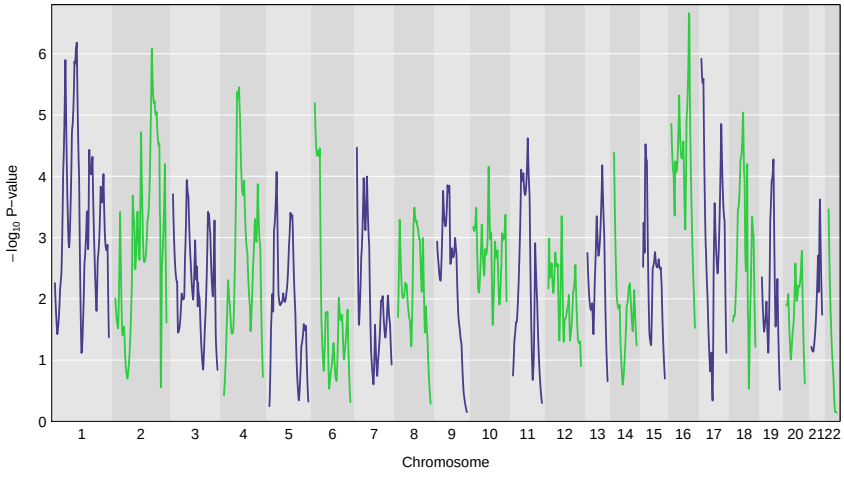


Affected sib pairs



Marshfield, Wisconsin

Prostate cancer genome scan



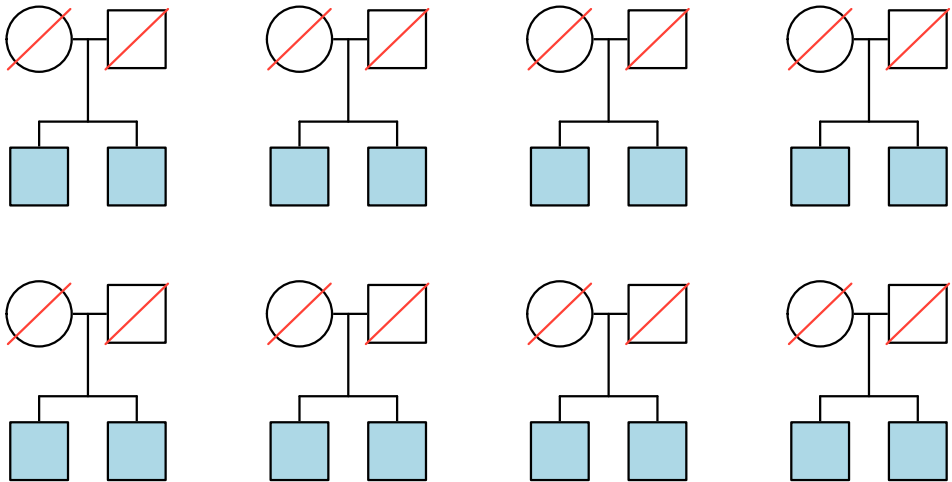

so happy

bit.ly/faxpic

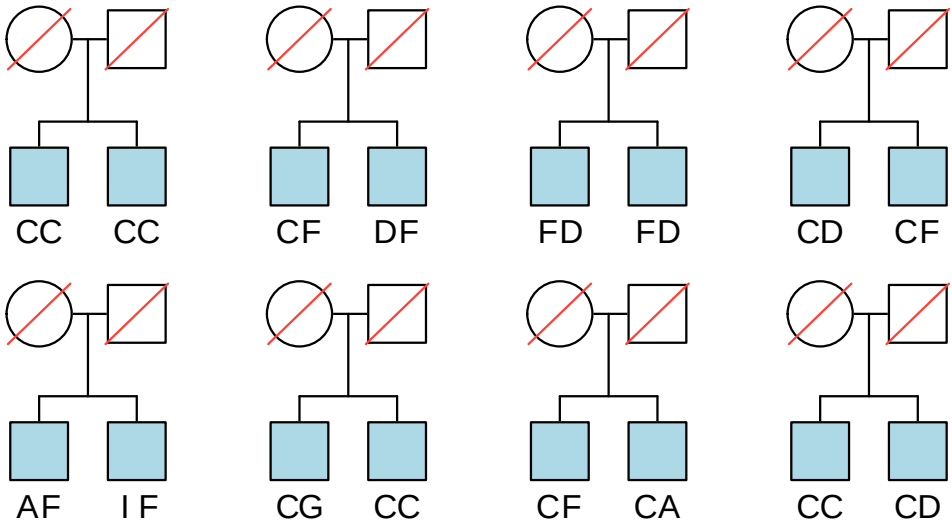
If it seems too good to be true,

it probably is.

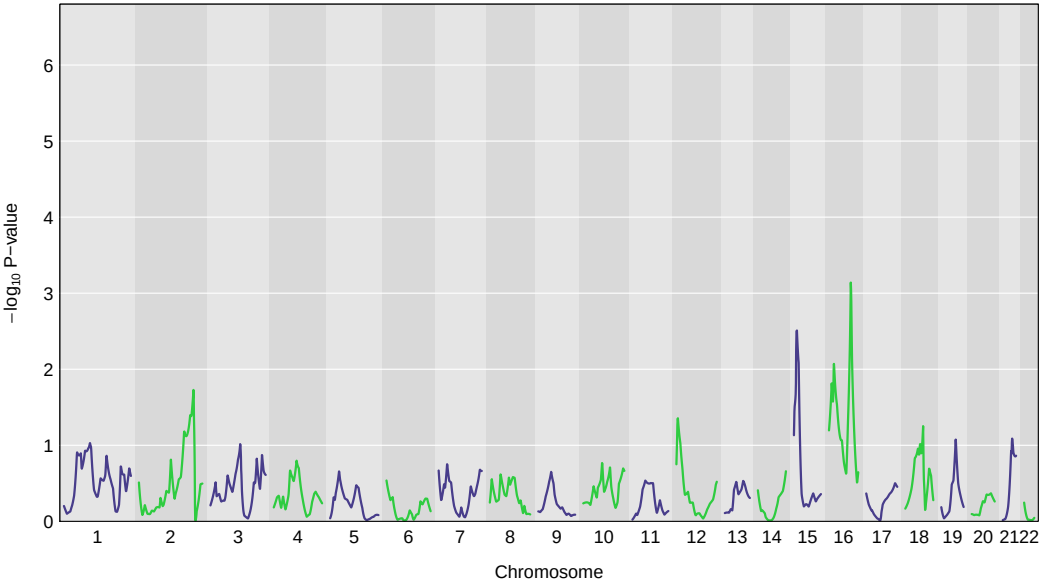
Prostate cancer pairs



Prostate cancer pairs



Prostate cancer genome scan – corrected



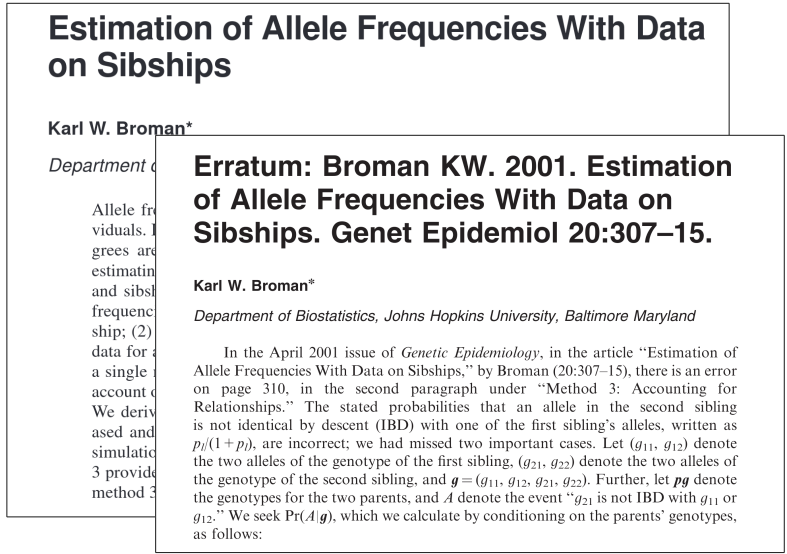
Estimation of Allele Frequencies With Data on Sibships

Karl W. Broman*

Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland

Allele frequencies are generally estimated with data on a set of unrelated individuals. In genetic studies of late-onset diseases, the founding individuals in pedigrees are often not available, and so one is confronted with the problem of estimating allele frequencies with data on related individuals. We focus on sibpairs and sibships, and compare the efficiency of four methods for estimating allele frequencies in this situation: (1) use the data for one individual from each sibship; (2) use the data for all individuals, ignoring their relationships; (3) use the data for all individuals, taking proper account of their relationships, considering a single marker at a time; and (4) use the data for all individuals, taking proper account of their relationships, considering a set of linked markers simultaneously. We derived the variance of estimator 2, and showed that the estimator is unbiased and provides substantial improvement over method 1. We used computer simulation to study the performance of methods 3 and 4, and showed that method 3 provides some improvement over method 2, while method 4 improves little on method 3. Genet. Epidemiol. 20:307–315, 2001.

doi:10.1002/gepi.2



Estimation of Allele Frequencies With Data on Sibships

Karl W. Broman*

Department o

Allele fr
viduals. I
grees are
estimatin
and sibsh
frequenc
ship; (2)
data for a
a single
account o
We deriv
ased and
simulatio
3 provide
method 3

Erratum: Broman KW. 2001. Estimation of Allele Frequencies With Data on Sibships. Genet Epidemiol 20:307–15.

Karl W. Broman*

Department of Biostatistics, Johns Hopkins University, Baltimore Maryland

In the April 2001 issue of *Genetic Epidemiology*, in the article "Estimation of Allele Frequencies With Data on Sibships," by Broman (20:307–15), there is an error on page 310, in the second paragraph under "Method 3: Accounting for Relationships." The stated probabilities that an allele in the second sibling is not identical by descent (IBD) with one of the first sibling's alleles, written as $p_j/(1+p_j)$, are incorrect; we had missed two important cases. Let (g_{11}, g_{12}) denote the two alleles of the genotype of the first sibling, (g_{21}, g_{22}) denote the two alleles of the genotype of the second sibling, and $\boldsymbol{g} = (g_{11}, g_{12}, g_{21}, g_{22})$. Further, let $\boldsymbol{pg}$ denote the genotypes for the two parents, and A denote the event "g_{21} is not IBD with g_{11} or g_{12}." We seek $\Pr(A|\boldsymbol{g})$, which we calculate by conditioning on the parents' genotypes, as follows:

doi:10.1002/gepi.2

Slides: kbroman.org/Talk_DataMishap

kbroman.org

github.com/kbroman

@kwbroman