Statistics in epigenomics

Karl W Broman
Department of Biostatistics
Johns Hopkins University

`http://www.biostat.jhsph.edu/~kbroman`

What is statistics?

> "We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of the uncertainty may itself be capable of rigorous expression."
>
> — Sir R. A. Fisher

What is statistics?

- Data exploration and analysis
- Inductive inference with probability
- Quantification of uncertainty
- Design of experiments (to do the above well)

Statistics in epigenomics?

The obvious thing: “Classical” imprinting studies

- Parent-of-origin effects in human disease
- Phenotypes within families
- … that plus genotype data

Note: no direct measure of epigenetic marks.

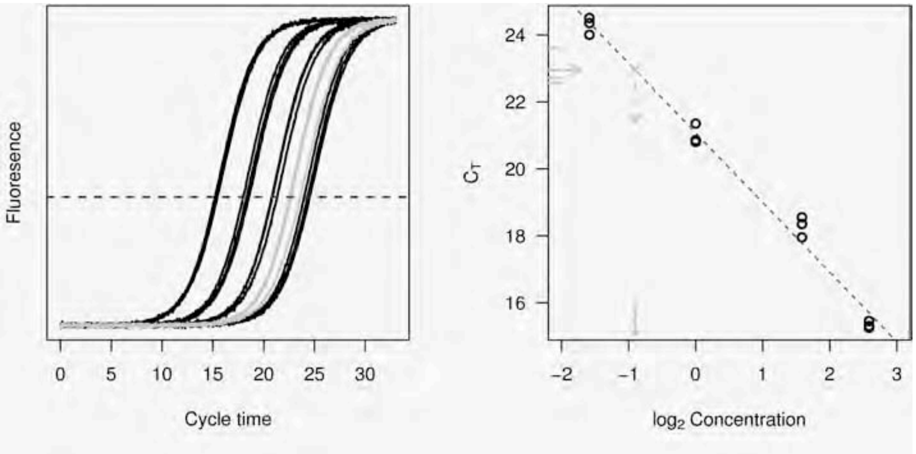
- Allelic expression
- Methylation
- Histone modification

Statistics in epigenomics?

Improve assays to quantify epigenetic marks

- Get the most precise results (cf expr. microarrays)
- Validity and reliability
- Design aspects (eg, # standards, # replicates, how to replicate)

Real-time PCR



Statistics in epigenomics?

Understanding epigenetic marks

- Variation
  - Between individuals
  - Between tissues
  - Across time
- Correlation
  - Between relatives
  - Between loci within an individual
  - Between alleles at a locus
  - Between tissues
  - Across time

Statistics in epigenomics?

Identify polymorphisms contributing to variation in epigenetic marks

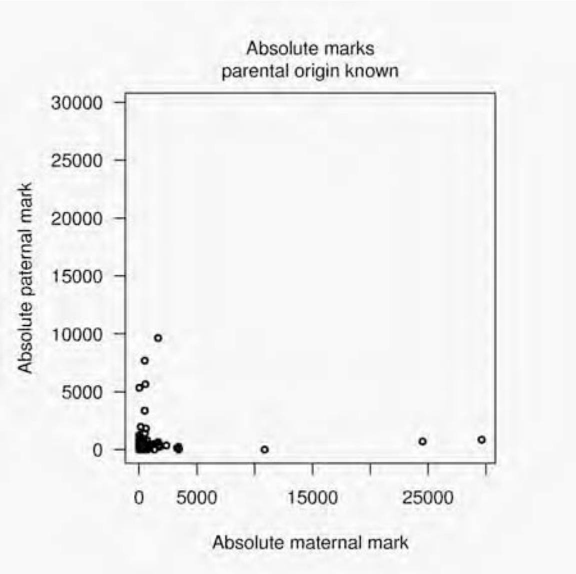
- A quantitative epigenetic mark is a quantitative phenotype.
- Use linkage analysis with pedigrees (or in model organisms) to map the genes responsible for individual variation in an epigenetic mark.

Statistics in epigenomics?

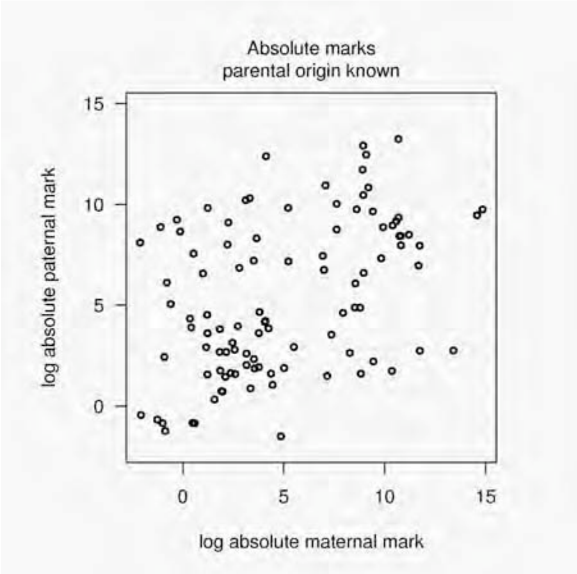
Epigenetic marks and human disease

- A quantitative epigenetic mark may be treated like any other risk factor
- The usual issues:
  - What is the best study design?
  - How to relate the factor to the disease?
  - How to account for other factors?
  - How to deal with associations between relatives?
- Special issues:
  - Special correlation between relatives & within individual
  - Absolute vs. relative measures of epigenetic marks
  - Parental origins of epigenetic marks

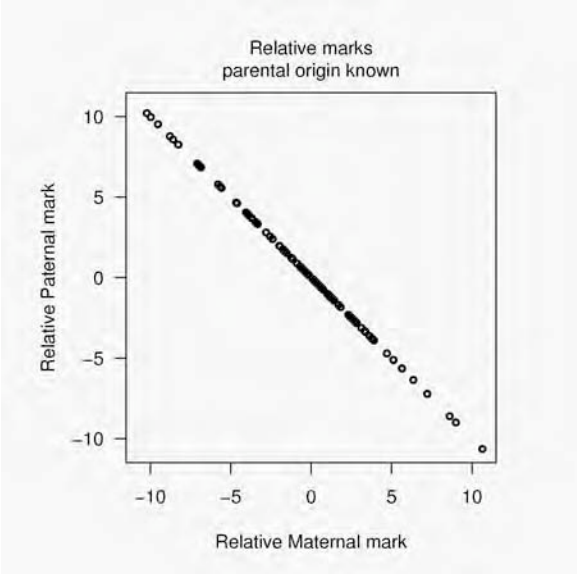
Generic epigenetic marks



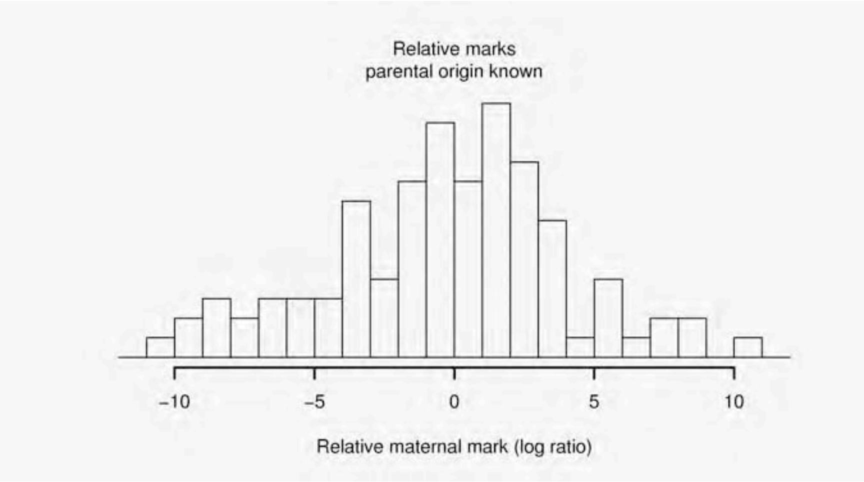
Generic epigenetic marks



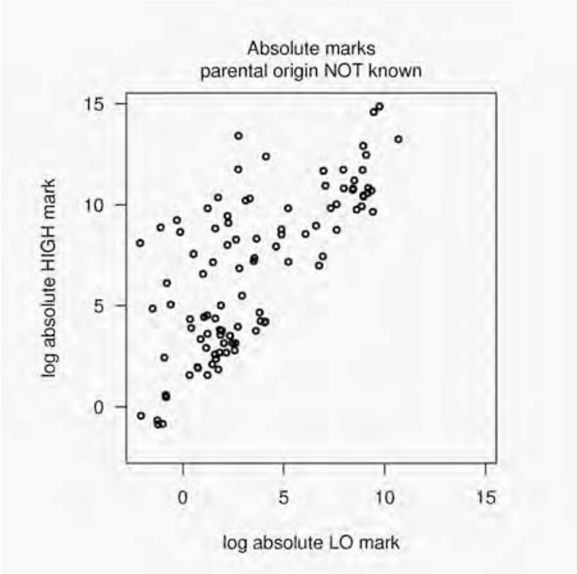
Generic epigenetic marks



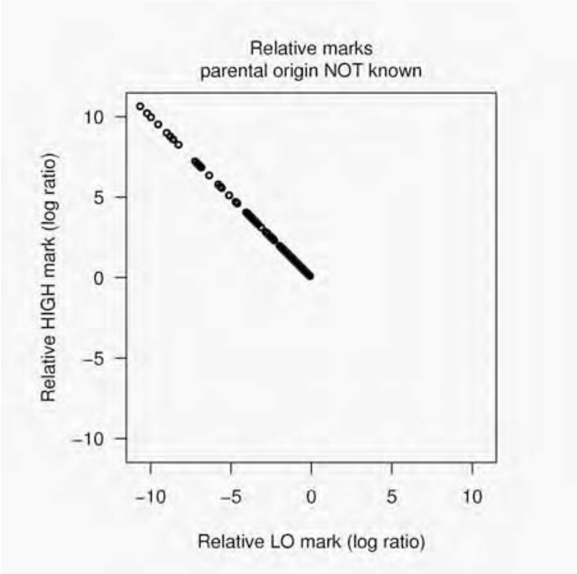
Generic epigenetic marks



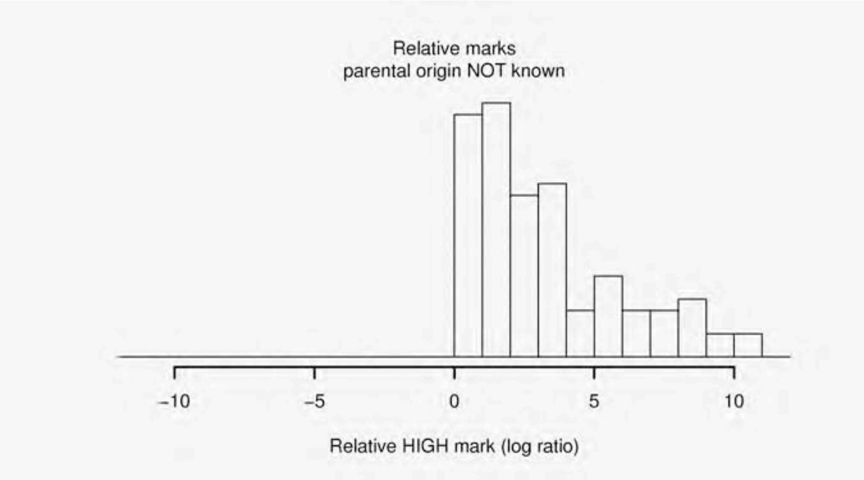
Generic epigenetic marks



Generic epigenetic marks



Generic epigenetic marks



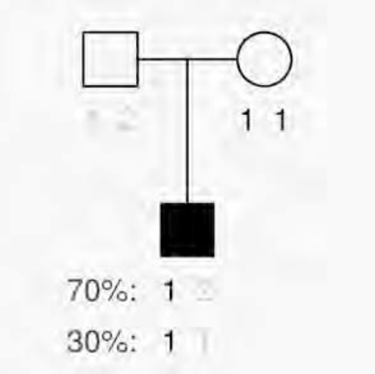
Statistics in epigenomics?

Epigenetic mark in parent ↔ disease in child?

- Can an aberrant epigenetic mark influence a child’s disease (when the corresponding allele is transmitted to the child)?

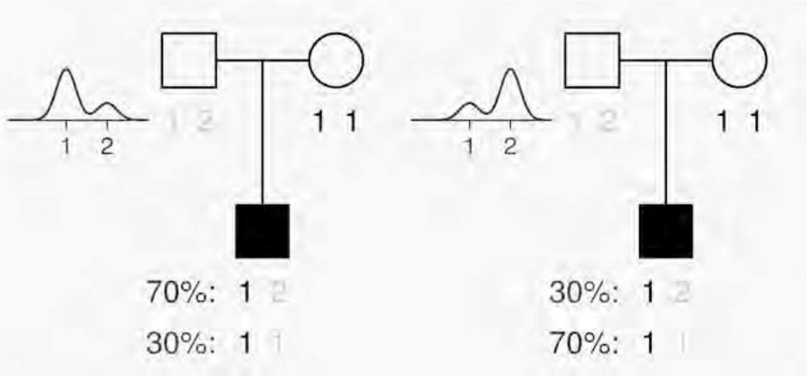
Transmission/disequilibrium test

- Establish association between an allele and risk of disease.



Epigenetic transmission test

- Establish association between an epigenetic mark and risk of disease.



Summary

- Establishing the connection between epigenetic marks and human disease is inherently statistical.
- Statisticians can assist with several aspects of epigenomic studies.
  - “Classical” imprinting
  - Improvement of epigenetic assays
  - Understanding epigenetic marks
  - Mapping genes which influence epigenetic marks
  - Connecting epigenetic marks to human disease
- We need to integrate epigenetics, genetics, environments, and phenotypes.

Acknowledgments

Johns Hopkins University

Andrew Feinberg
Hans Bjornsson
James Potash
Rafael Irizarry
Ingo Ruczinski
Dani Fallin

University of Pennsylvania

Richard Spielman