Human crossover interference

Karl W Broman

Dept of Biostatistics
Johns Hopkins University

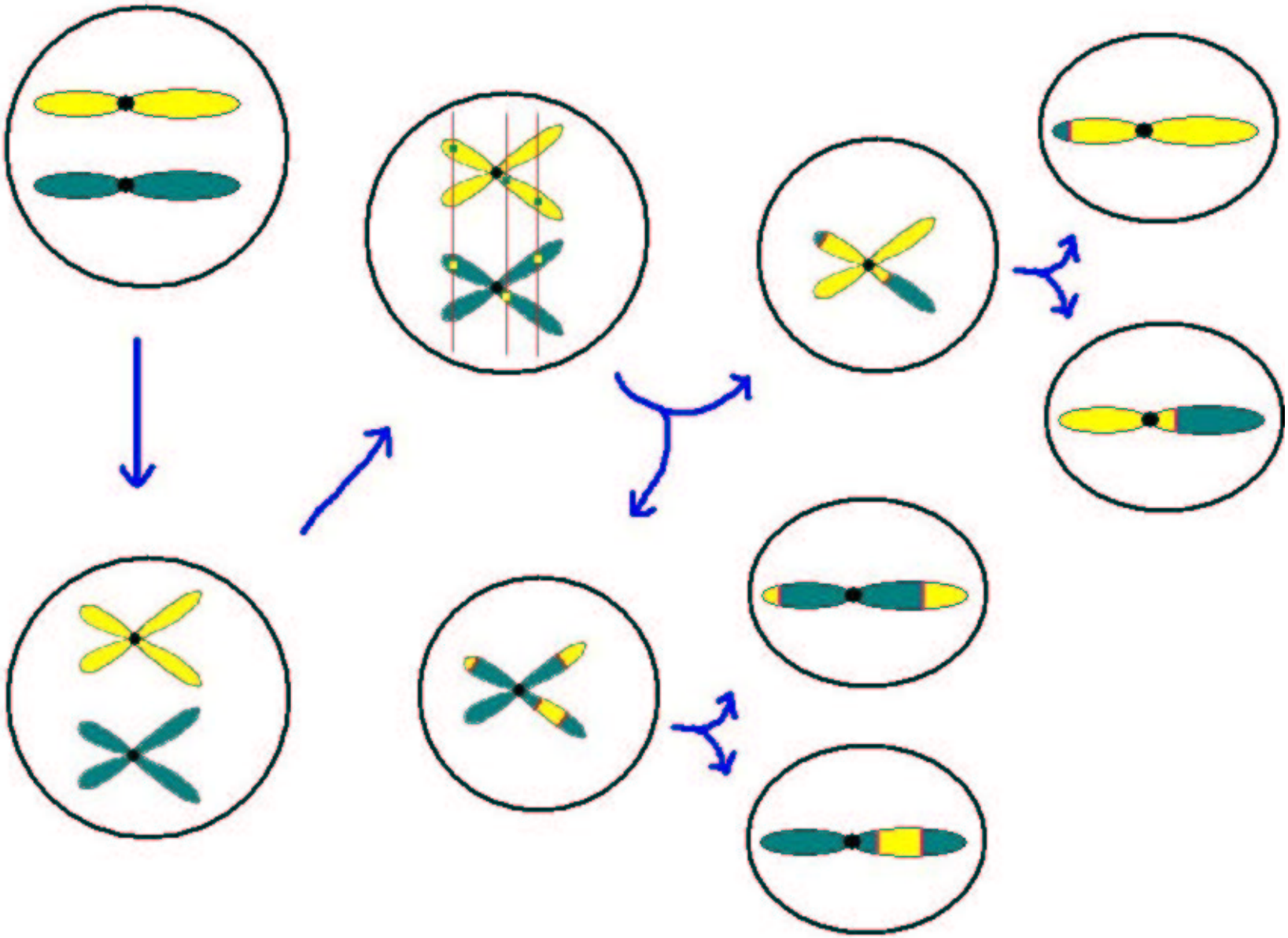
Joint work with James L. Weber,
Marshfield Medical Research Foundation

http://biosun01.biostat.jhsph.edu/~kbroman

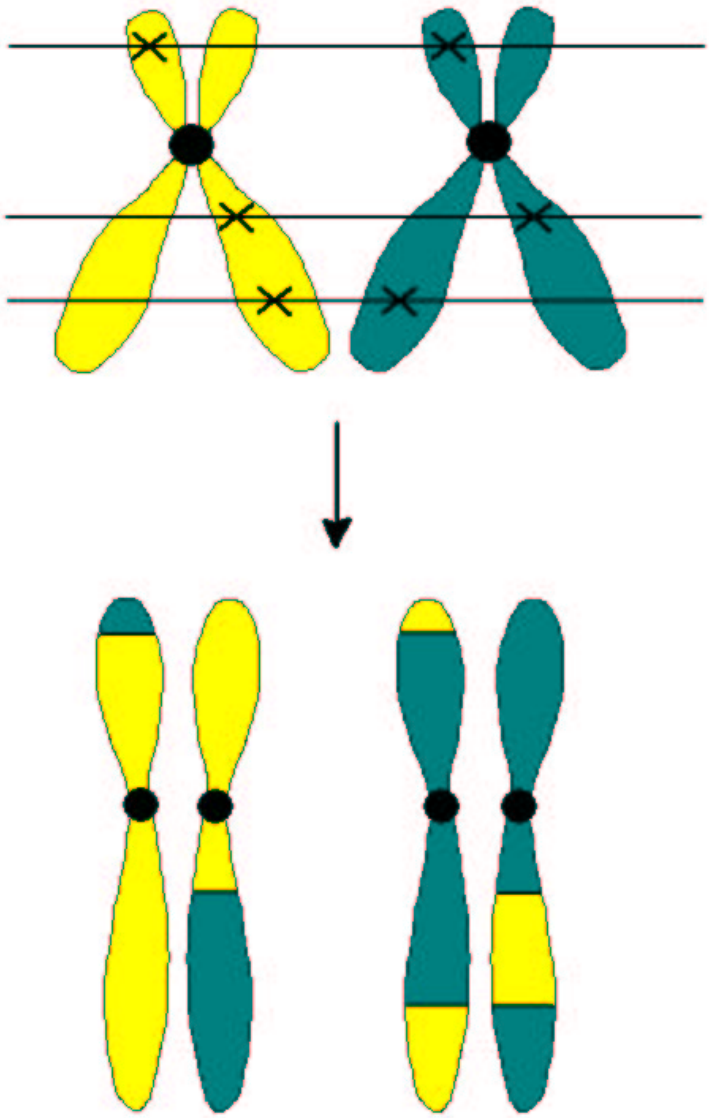
Things I’ve learned

- Don’t be afraid of approximations.
- Study data aberrations.
- Data analysis is often a muddle.
- Consider making plots on the log scale.

Meiosis



Interference



- **Strand choice**

  → Chromatid interference

- **Spacing**

  → Crossover interference

Why study interference?

- Estimate the probability of a double crossover in a small interval.
- Obtain a model of meiosis for simulation and analysis.
- Compare human meiosis to that of other organisms.

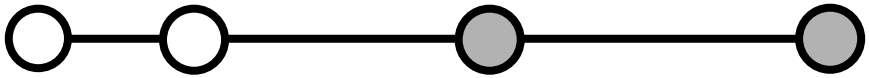
Goals

- Demonstrate the presence of interference in human meiosis.
- Find a good model.
- Estimate the level of interference.

Recombination

Crossovers on a random meiotic product

Typical data: recombination information



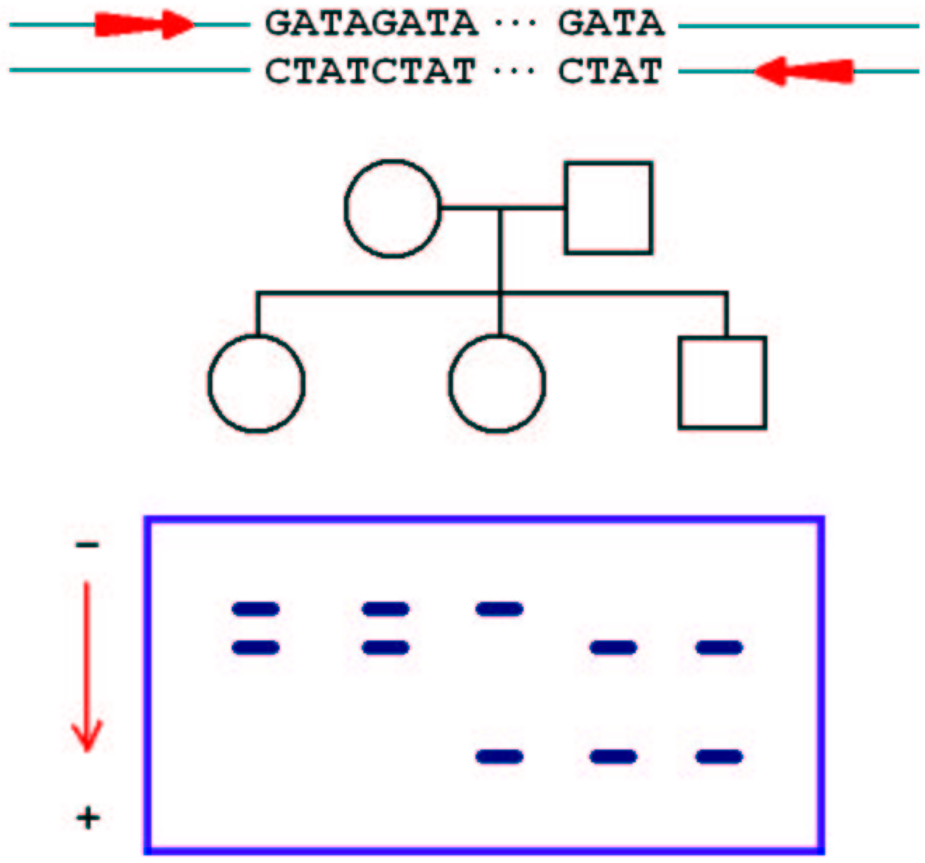
We generally do not observe the locations of crossovers; rather, we observe the grandparental origin of DNA at a set of genetic markers.

Recombination across an interval indicates an odd number of crossovers.

Genetic markers

Short tandem repeat polymorphisms (STRPs)

- Also known as microsatellites.
- Individuals differ in the number of repeats.
- Use PCR w/ tagged primers to amplify the segment of DNA.
- Use gel electrophoresis to determine the fragment lengths.



Genetic distance

Distance (cM) =

average no. crossovers in 100 meiotic products

Per 100 cM:

2 chiasmata on 4-strand bundle

1 crossover on meiotic product

Map functions

Recombination fraction as a function of genetic distance (Here *d* is in Morgans.)

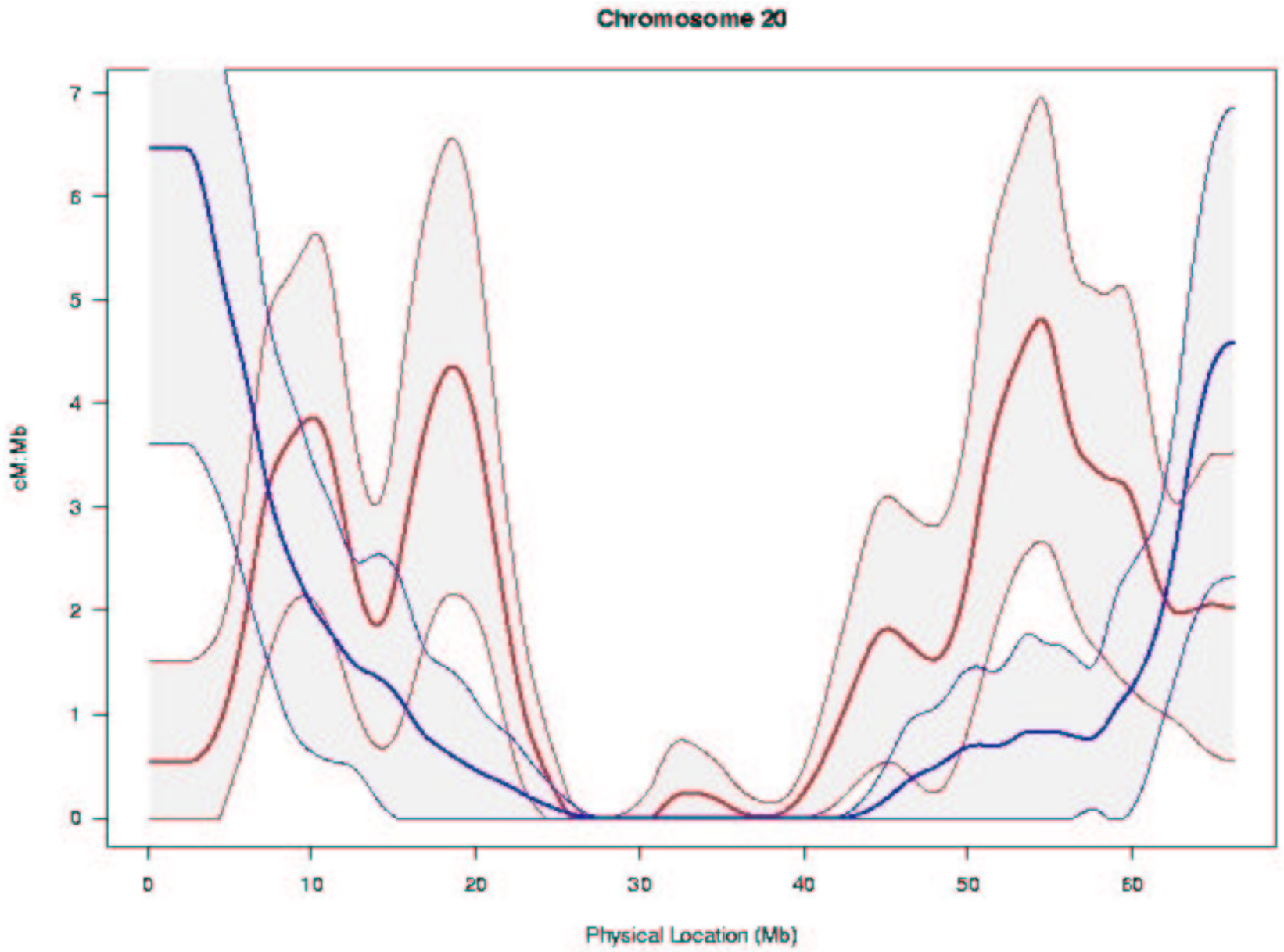
Haldane:

$$r = [\, 1 - \exp(-2d) \,] / 2$$

Kosambi:

$$r = \tanh(2d) / 2$$

Sex-specific recombination rate



Model organisms

***Drosophila* data (Morgan et al 1935)**

Event	Count	Event	Count
0000	10,431	1001	46
1000	771	0101	53
0100	1,579	0011	25
0010	1,221	1110	1
0001	1,994	1101	1
1100	4	1011	1
1010	7	0111	1
0110	4	1111	1

- Many meioses.
- A few linked markers.
- Consider rare multiple recombination events.

1---	5%
-1--	10%
11--	0.04%

Human data

- **http://research.marshfieldclinic.org/genetics**
- **8 CEPH families**

  3 generations; 11 – 15 progeny; 92 meioses

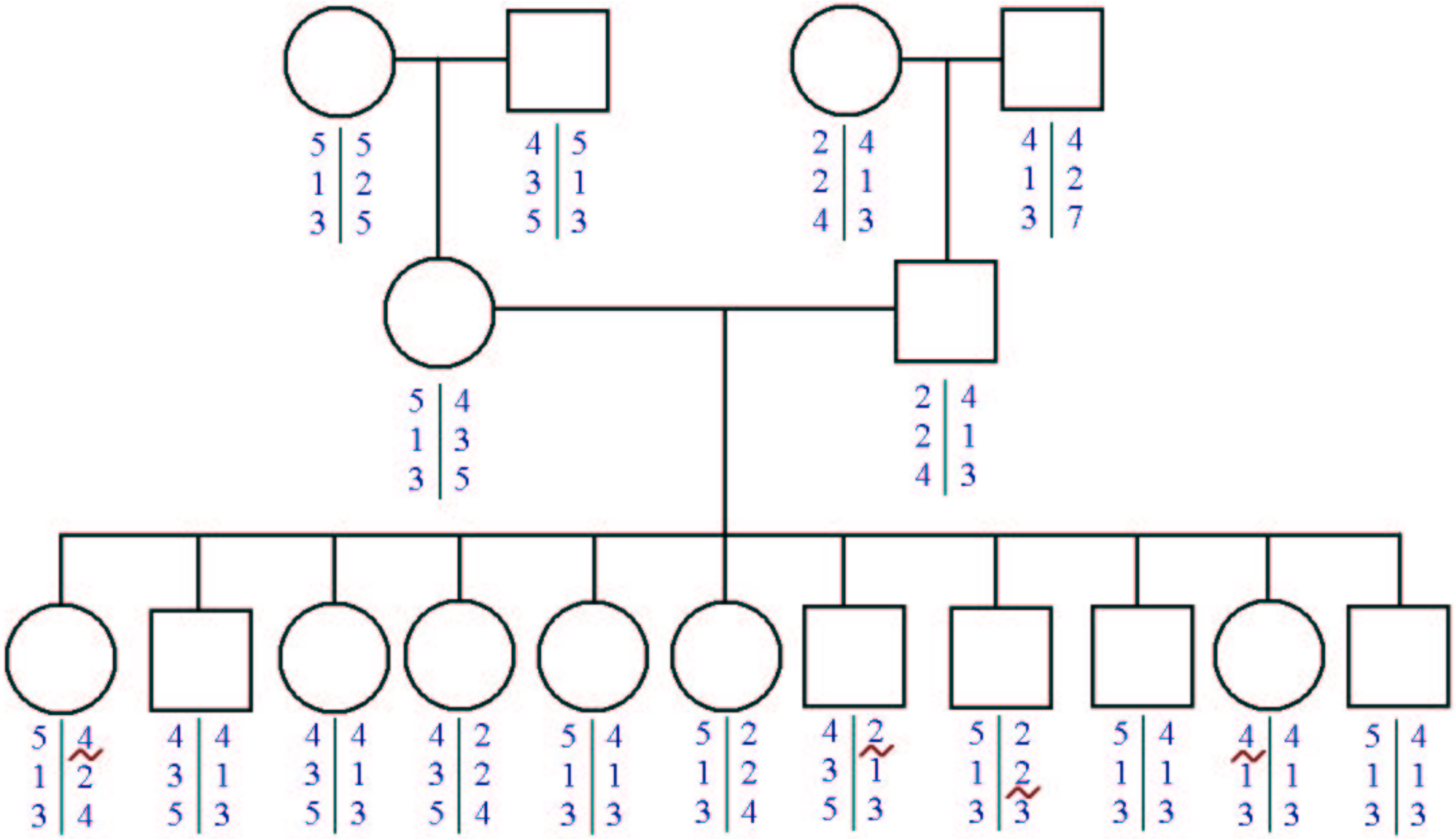
- **~ 8,000 STRPs**

  90% typed; 0.5 cM spacing

- **Data cleaning**

  Removed 764 / 964,425 (~8/10,000) genotypes resulting in tight double recombinants.

CEPH families



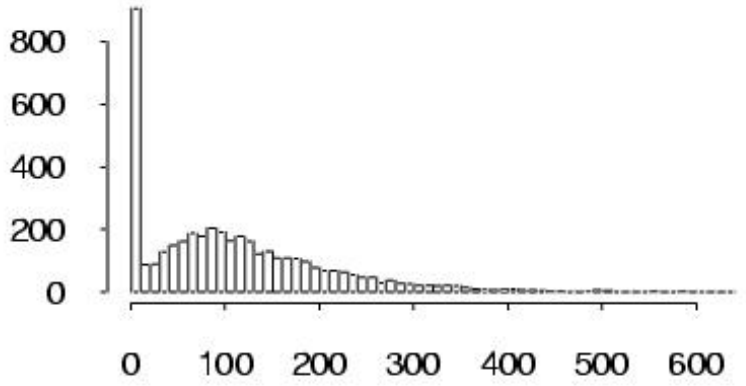
Data cleaning

CRI-MAP *chrompic*

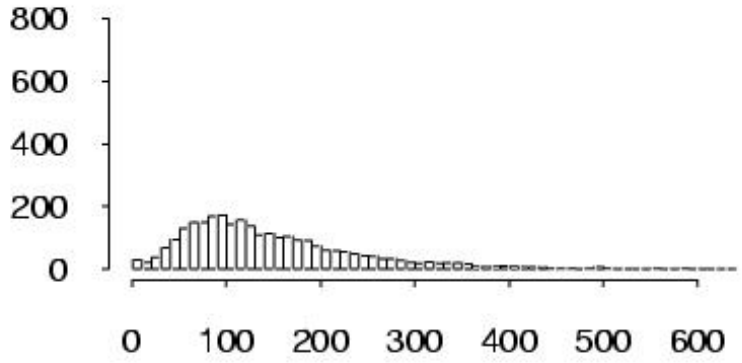
CEPH individual 1331–11; maternal chr 10

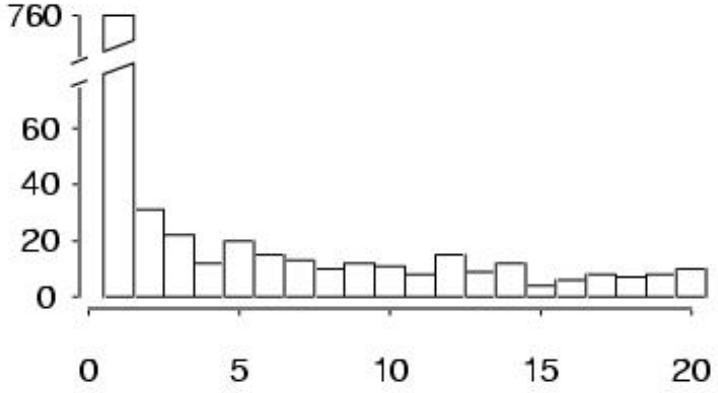
```
1111111--- 11-111-11- --11--111i 1-11---11-
111111--i- 11-1111-11 --11111-11 1111-11111
11--111111 111-111i-i 1111111111 10000-0-00
0--o00-000 0000-00000 0000--0000 0000--0000
0000-00000 o00-0--0-- --0-11-11- -111ii1i-1
---1-i-1-i 1111-i--11 11111-11i1 -11i-11111
-1----i111 1i1111-111 -11i1-111- 11-111111i
111-i111i- 1111111-i- 1111111-1i 1i-111i11-
1i--1-11-1 111-1i-1-1 1-1----1-1 1i-1ii1i11
1i--1--1i- 11i11--111 11--1i111i 1i1i-11111
i-0---0000 00000-000o o0-00o
```

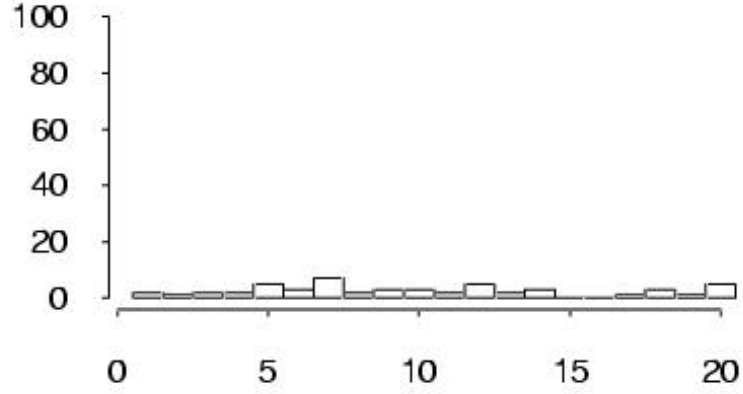
Raw Data



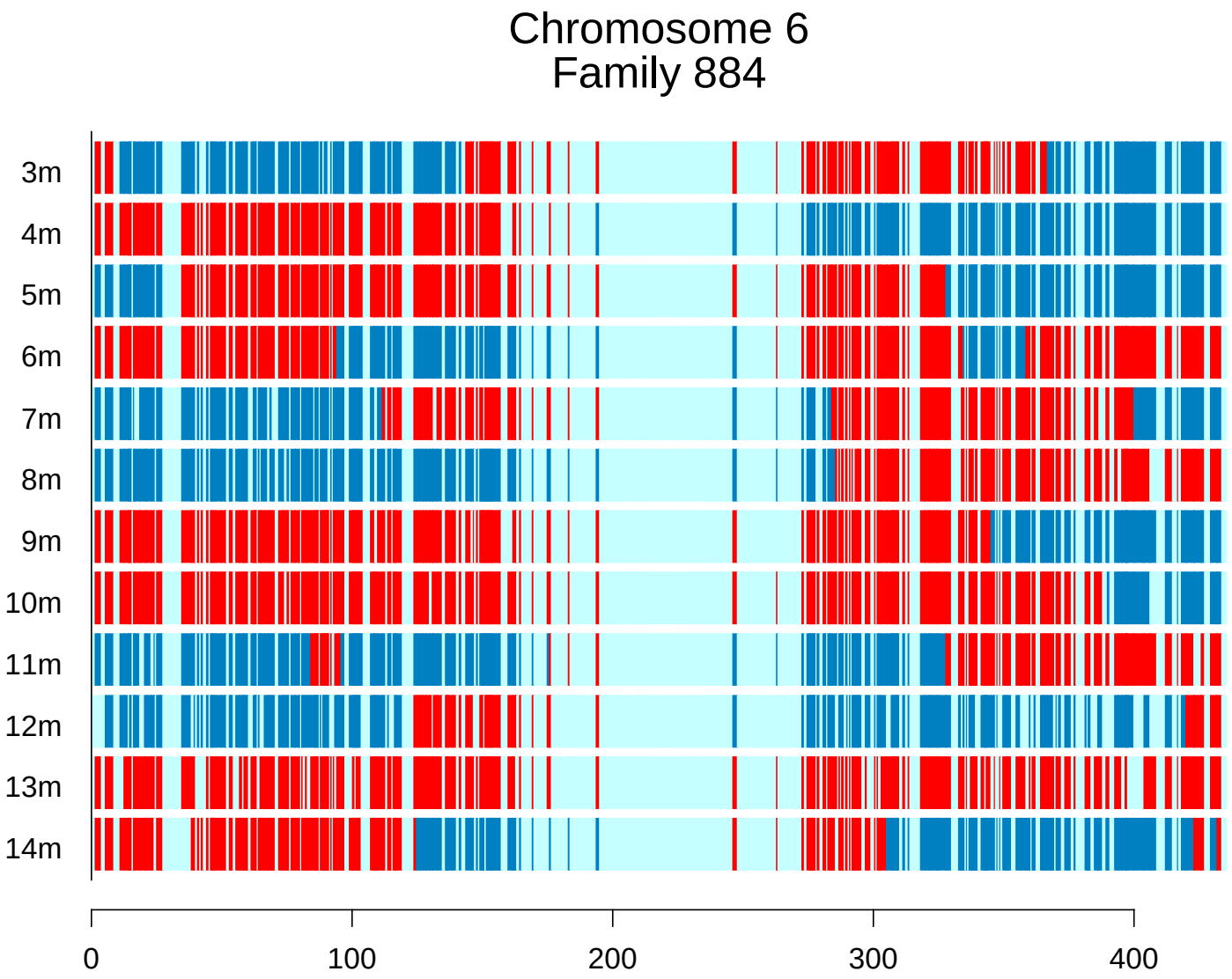
Clean Data



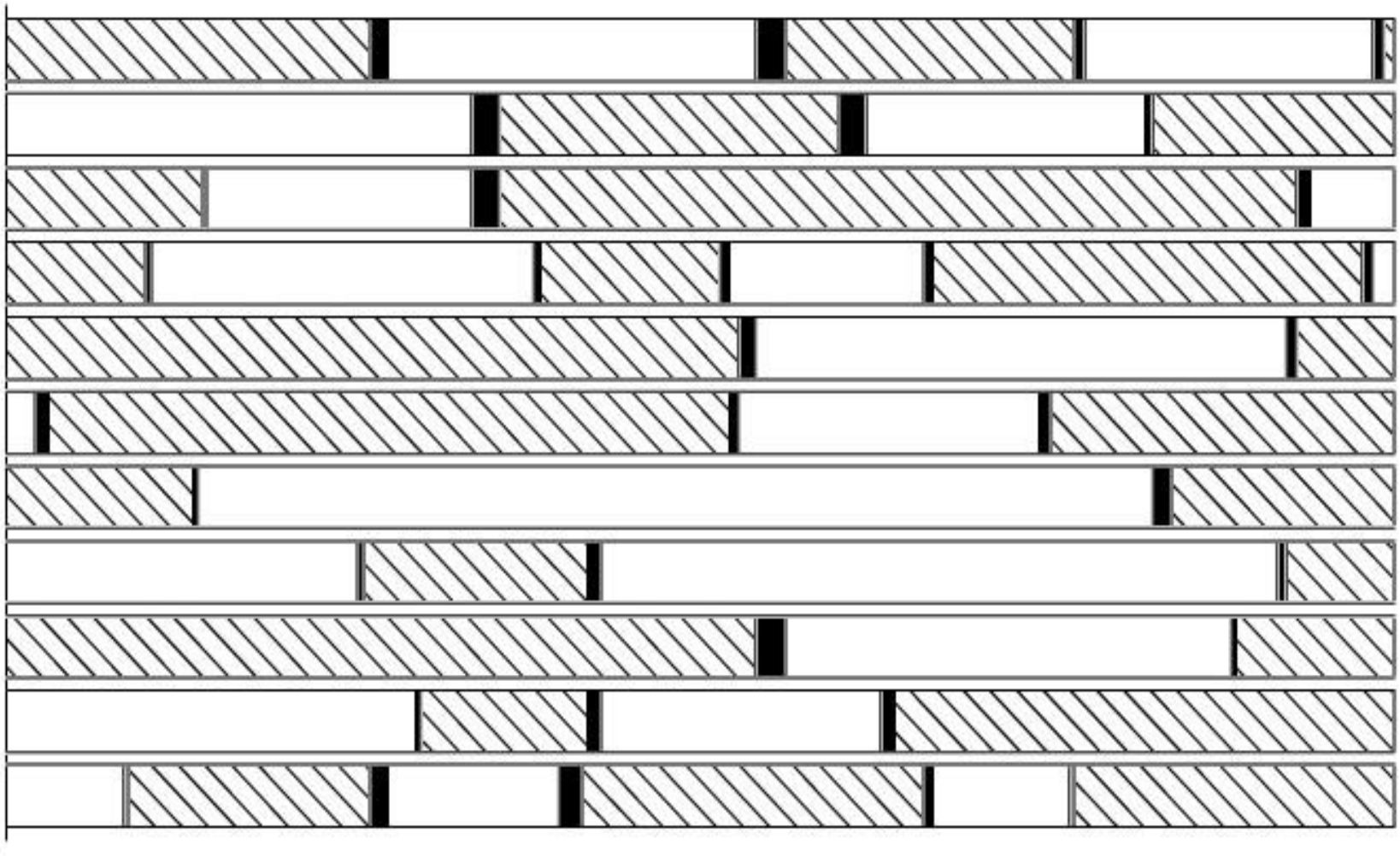




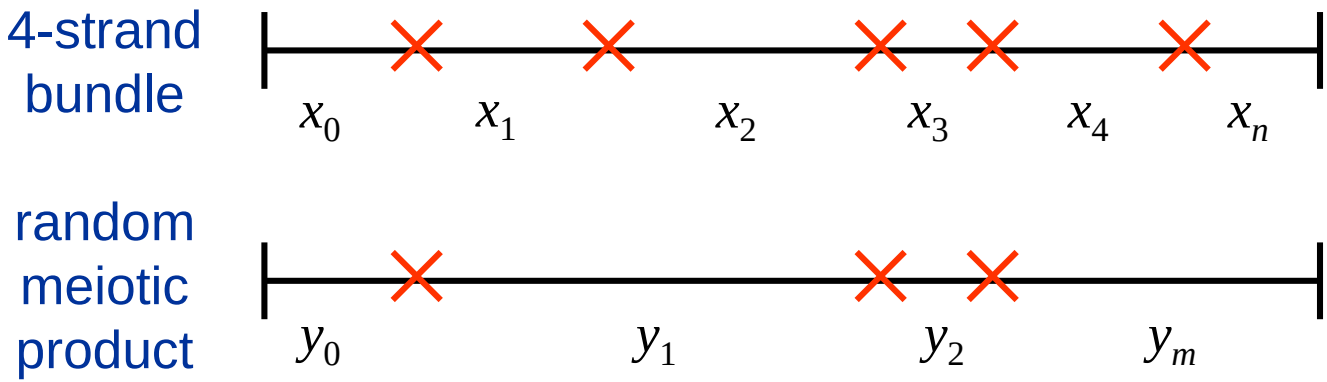
Autozygosity



The data



Models



Count-location model

$n \sim (p_0, p_1, p_2, \ldots)$

locations | n ~ iid uniform

Gamma model

x_i's ~ stationary gamma renewal process (shape = u, rate = $2u$)

y_i's ~ mixtures of gammas

Model fit: C-L model

m_i = # crossovers

n_i = underlying # chiasmata

$n_i \sim (p_0, p_1, p_2, \ldots)$

$m_i \mid n_i \sim \text{binomial}(n_i, 1/2)$

MLEs by a version of the EM algorithm

Model fit: Gamma model

$x_1, x_2, \ldots \sim f(u, 2u)$

$x_0 \sim g = 2[1 - F(u, 2u)]$

x_i's independent

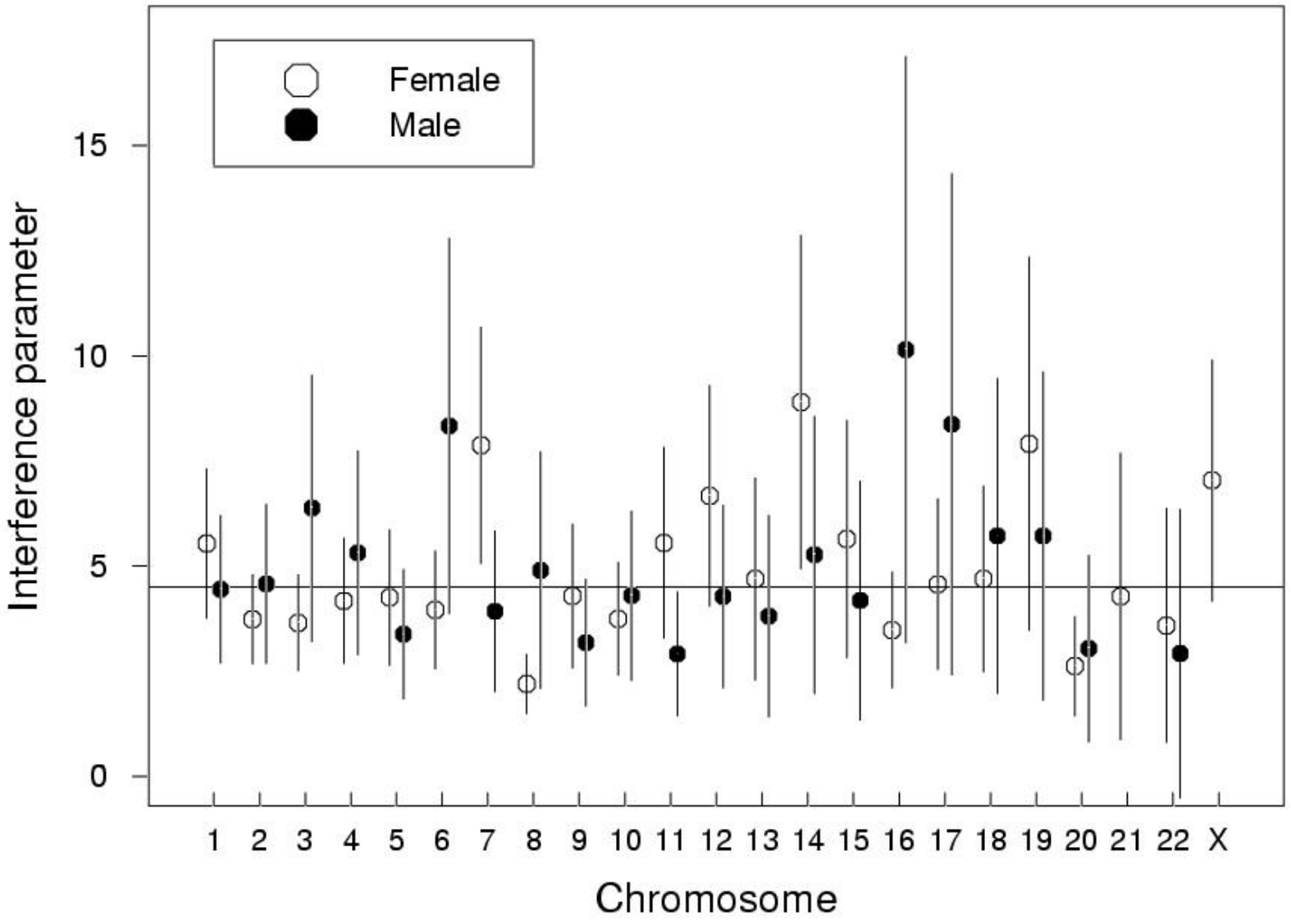
$y_1, y_2, \ldots \sim \Sigma\, (½)^k f(ku, 2u)$

$y_0 \sim ½\, g + \Sigma\, (½)^{(k+1)}\, g * f(ku, 2u)$

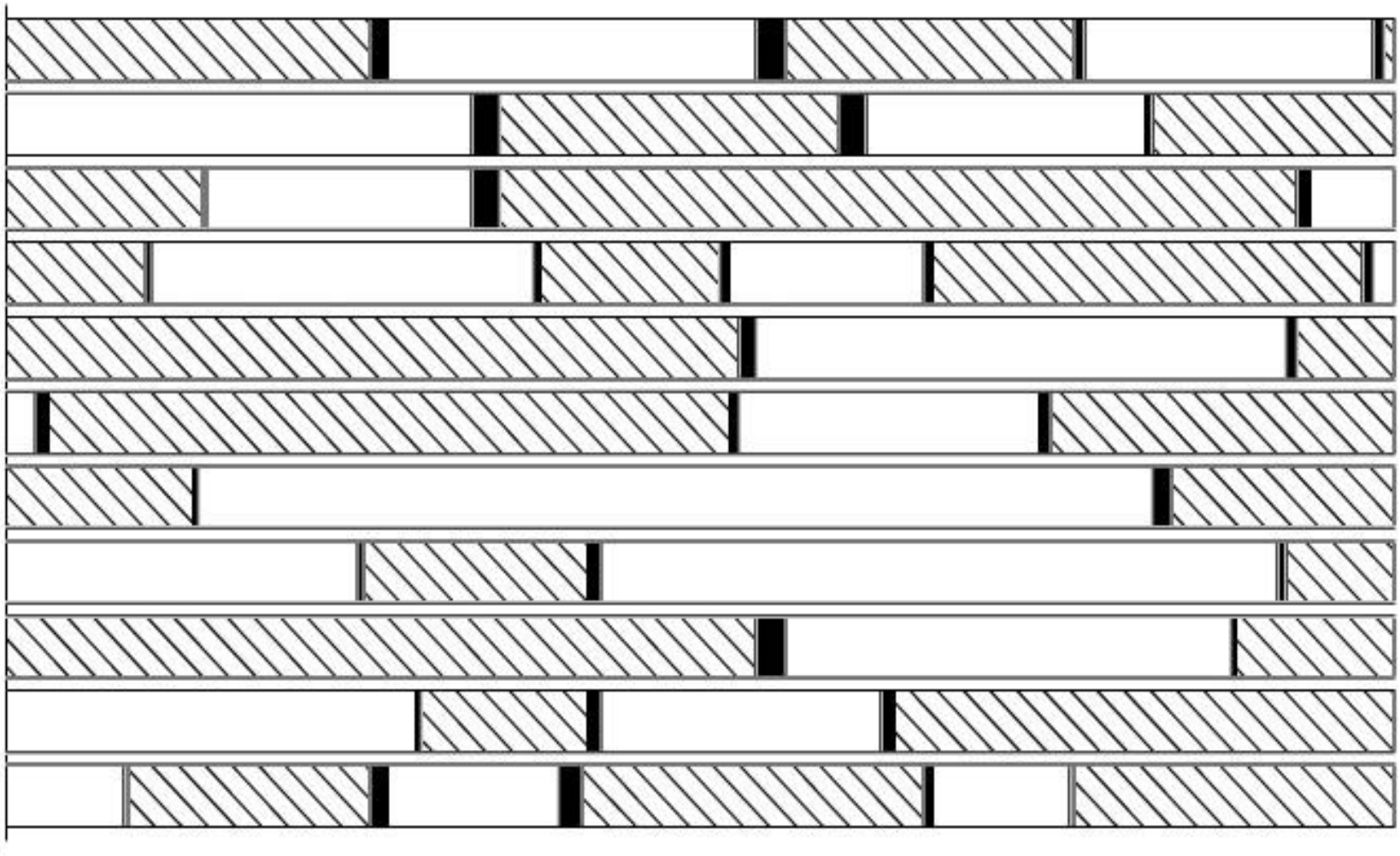
y_i's independent

- MLE of u using y_i's.
- g calculated numerically.
- Convolutions calculated numerically.
- Maximization performed using a quasi-Newton algorithm.

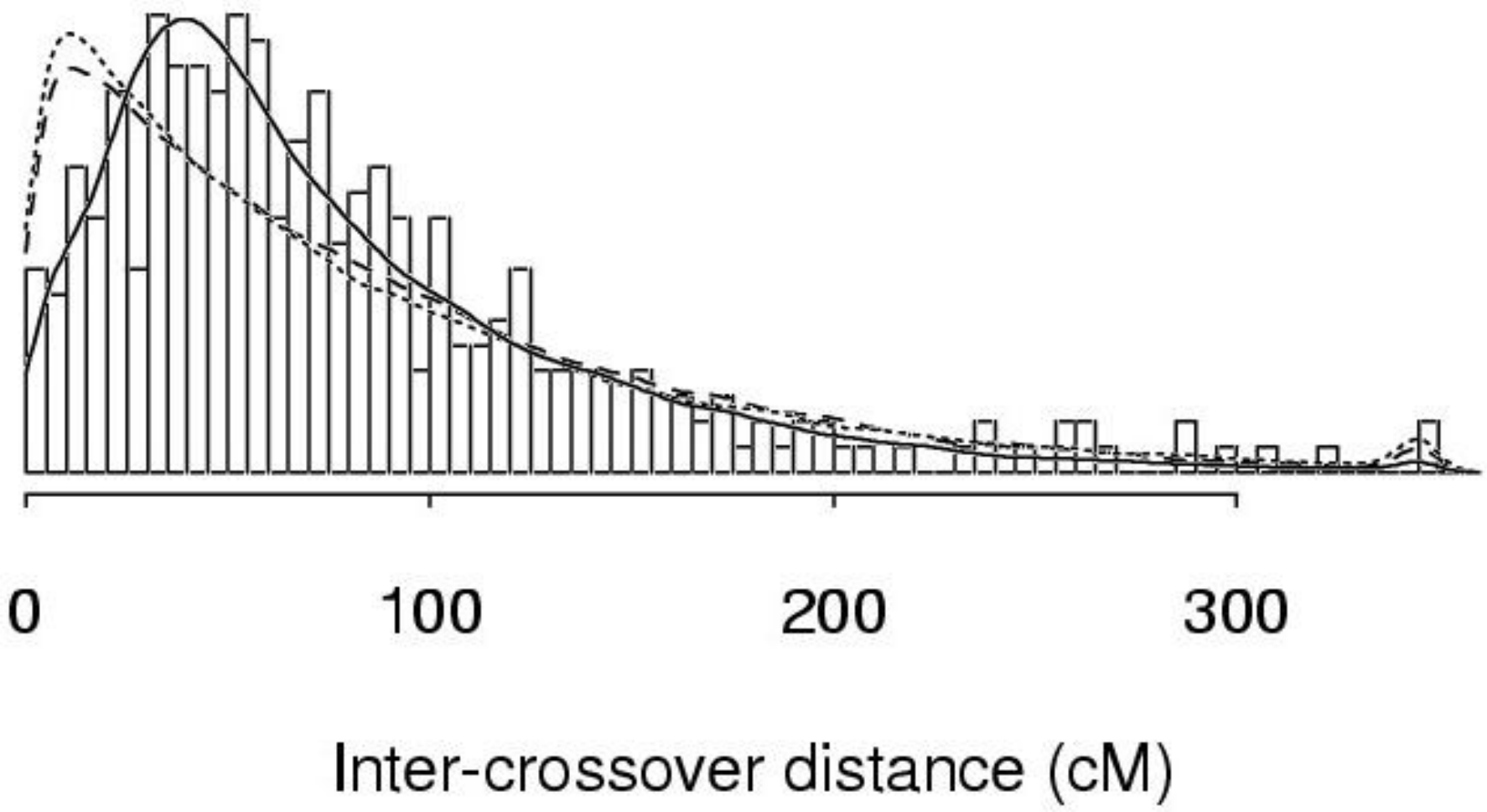
Estimated level of interference



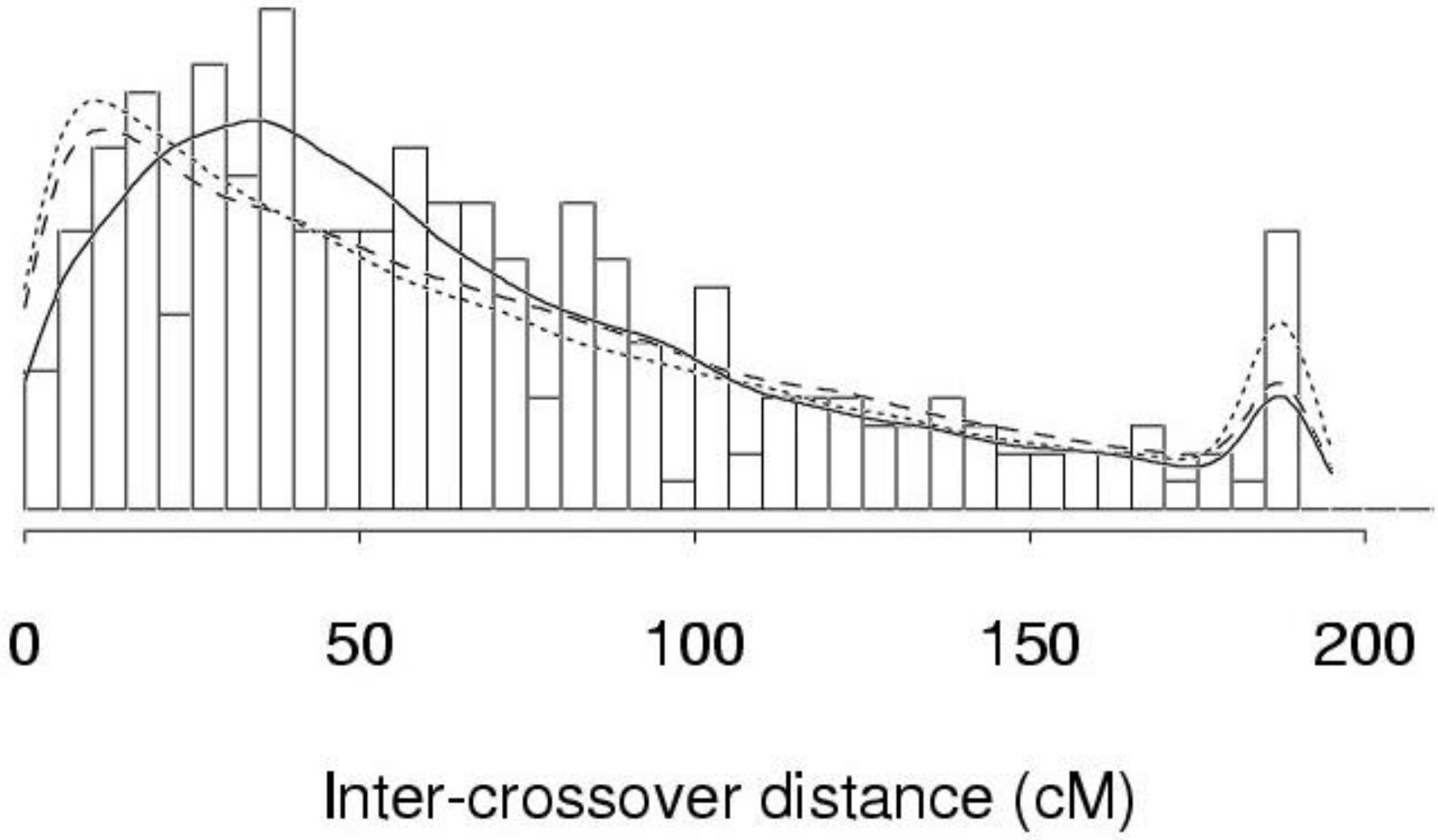
The data



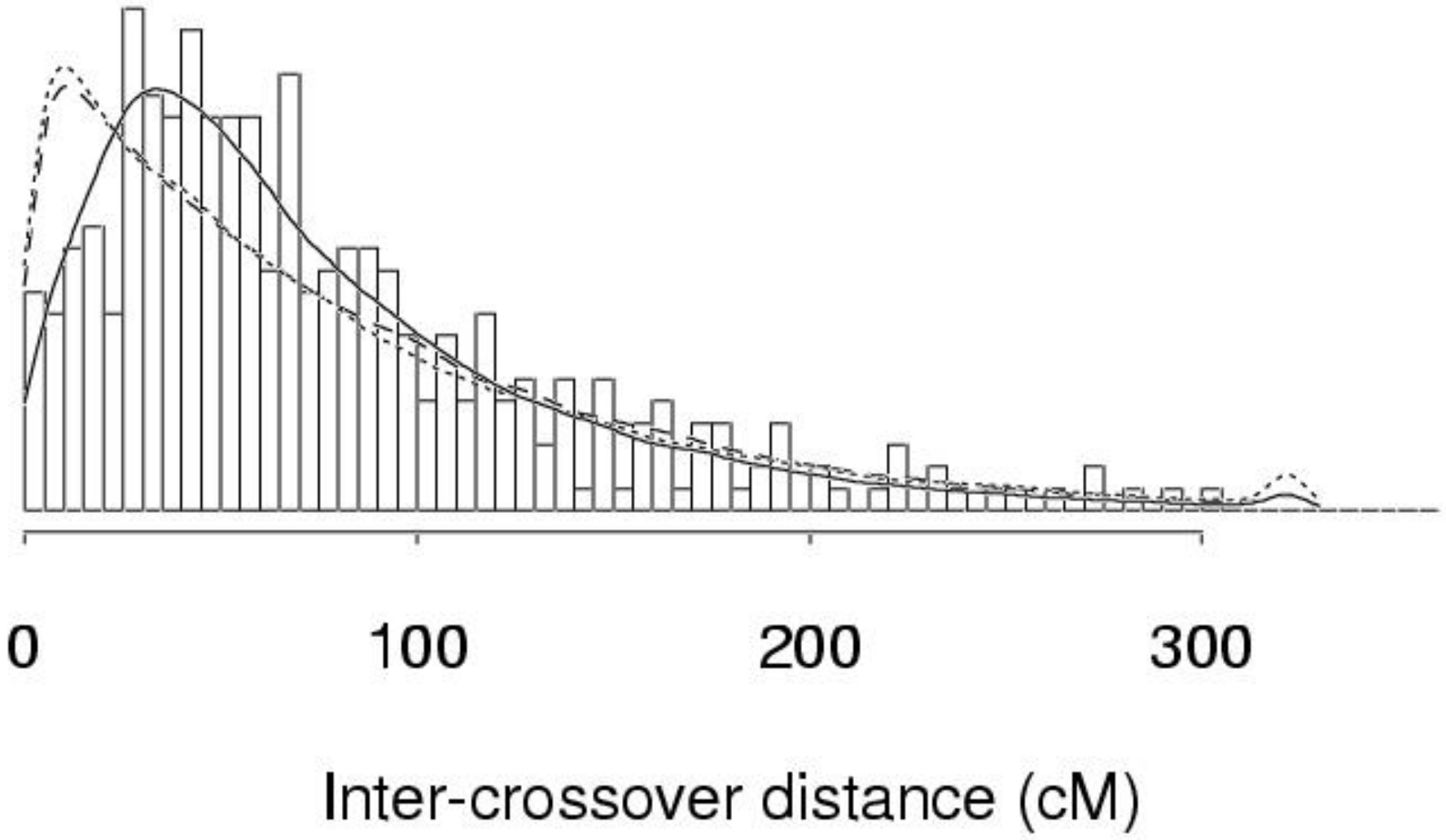
Maternal chr 1



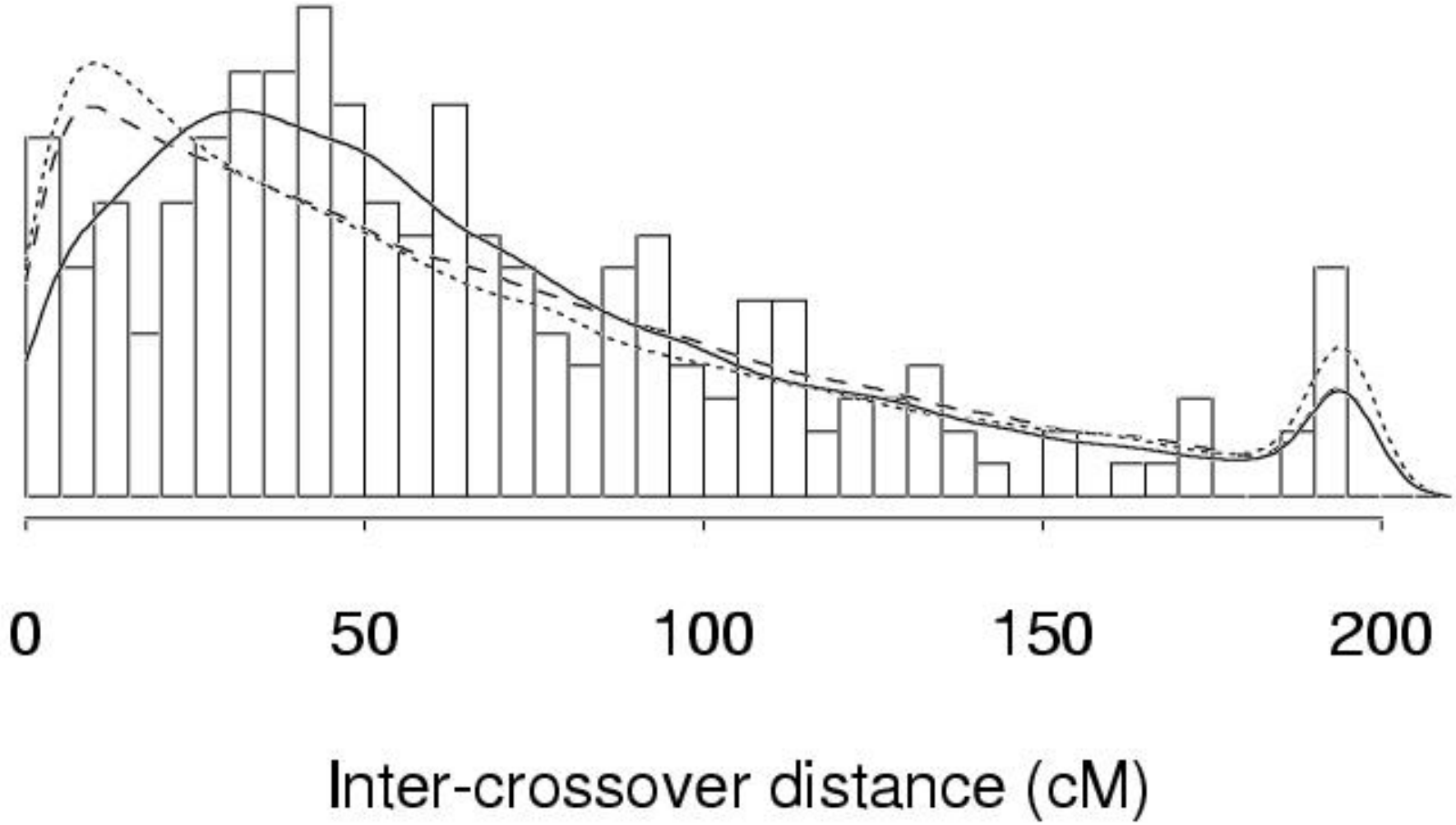
Paternal chr 1



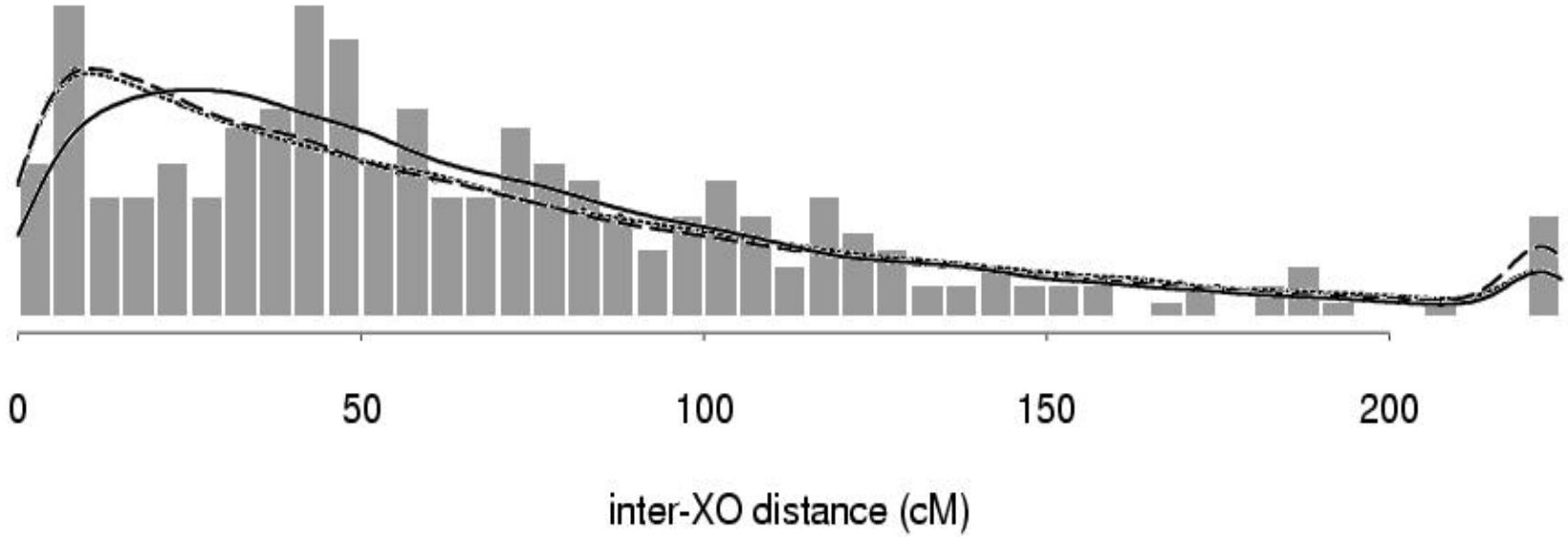
Maternal chr 2



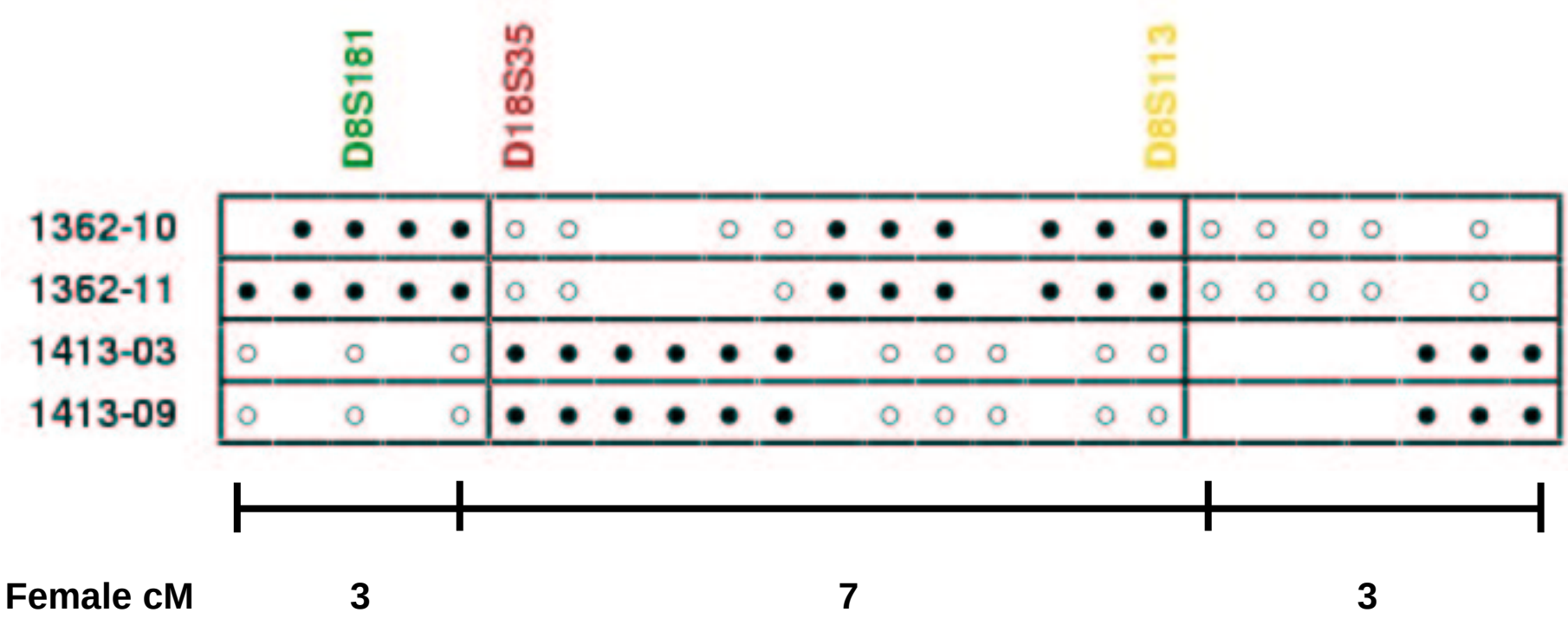
Paternal chr 2



Maternal chr 8



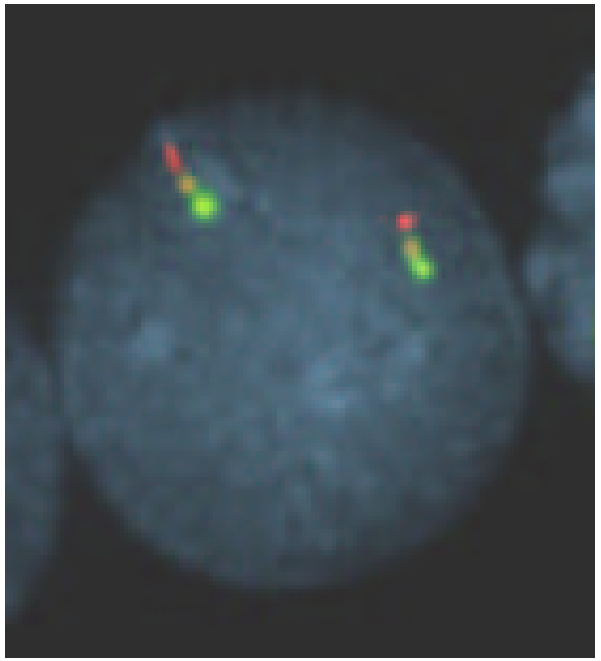
Chr 8p: maternal chromosomes



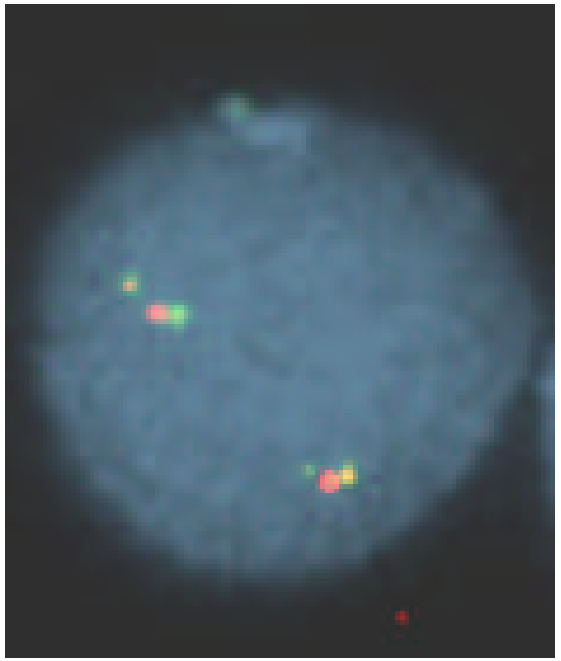
Cytogenetics (FISH)

David H. Ledbetter	U Chicago
Naomichi Matsumoto	U Chicago, Nagasaki U
Sabrina Giglio	U Chicago, U di Pavia
Christa M. Lese	U Chicago
Jessica A. Roseberry	U Chicago

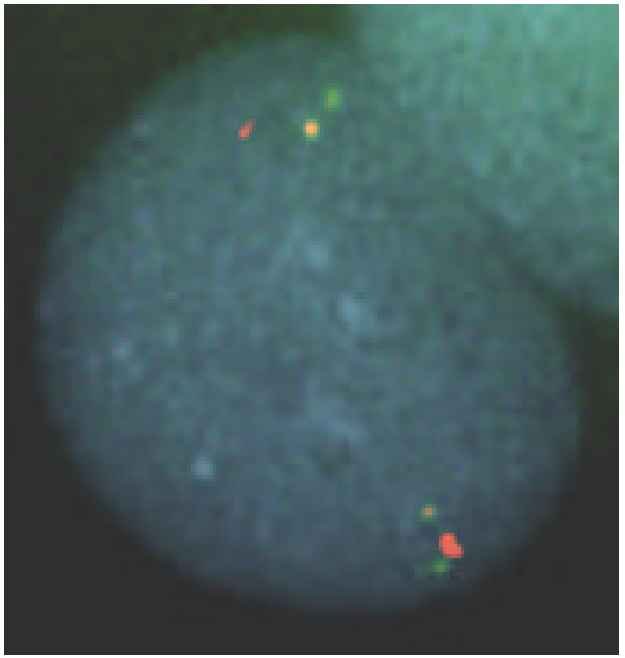
Chr 8p: Interphase FISH



Homozygous inversion

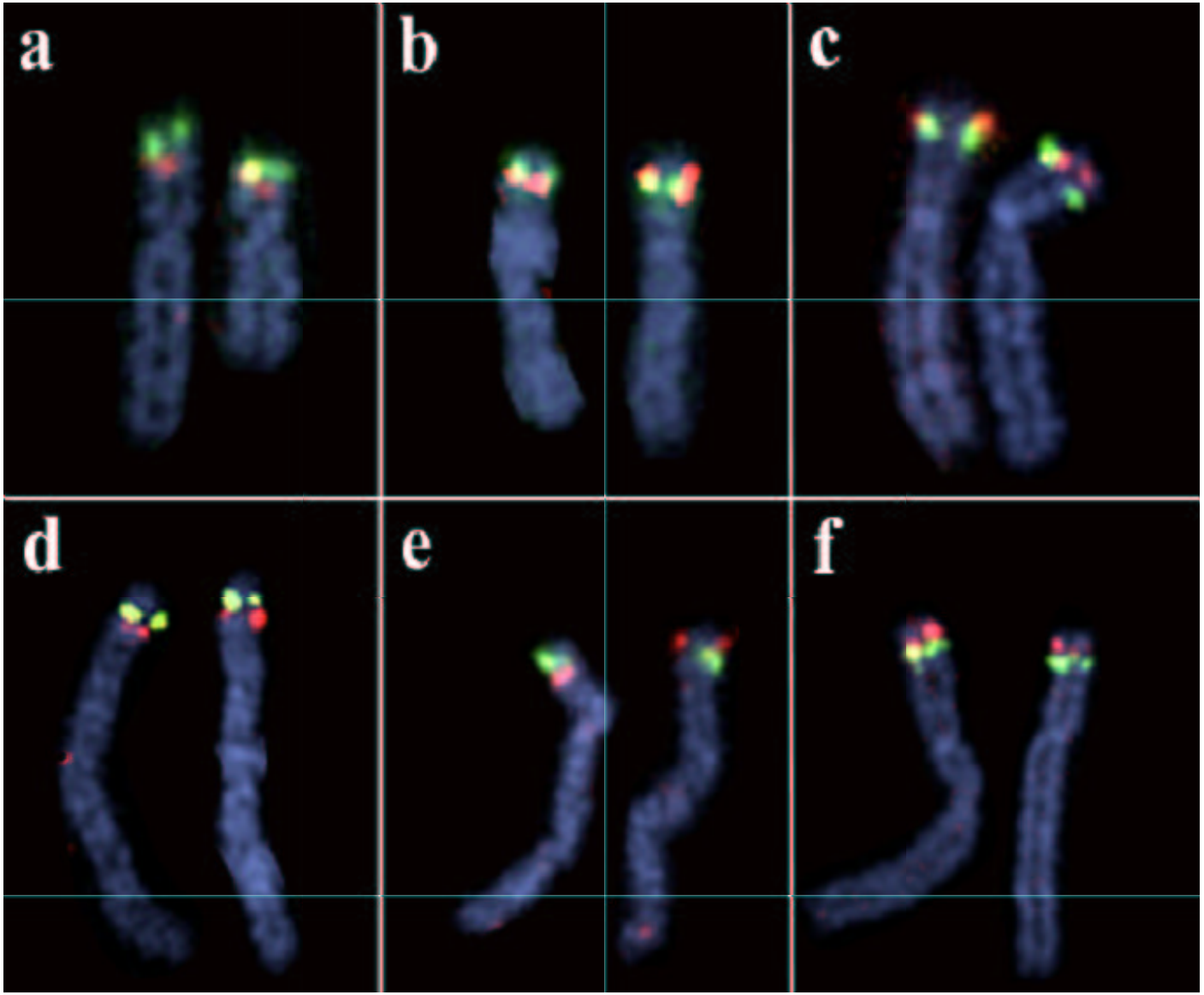


Homozygous normal



Heterozygous

Chr 8p: Metaphase FISH



Approximations

- Correct marker order.
- Correct inter-marker distances.
- All crossovers observed.
- Interval censoring unimportant.
- No individual variation in recombination.
- Interference constant along chromosomes.

Summary

- Gamma model fits well.
- Count-location model fits poorly.
- Gamma parameter, $\nu \approx 4.3$
  (stronger than Kosambi, $\nu \approx 2.6$)
- No significant variation in interference between chromosomes.
- Possible individual variation in interference among mothers.
- Other findings: autozygosity; inversion on 8p

Pr(double XO | no recombination)