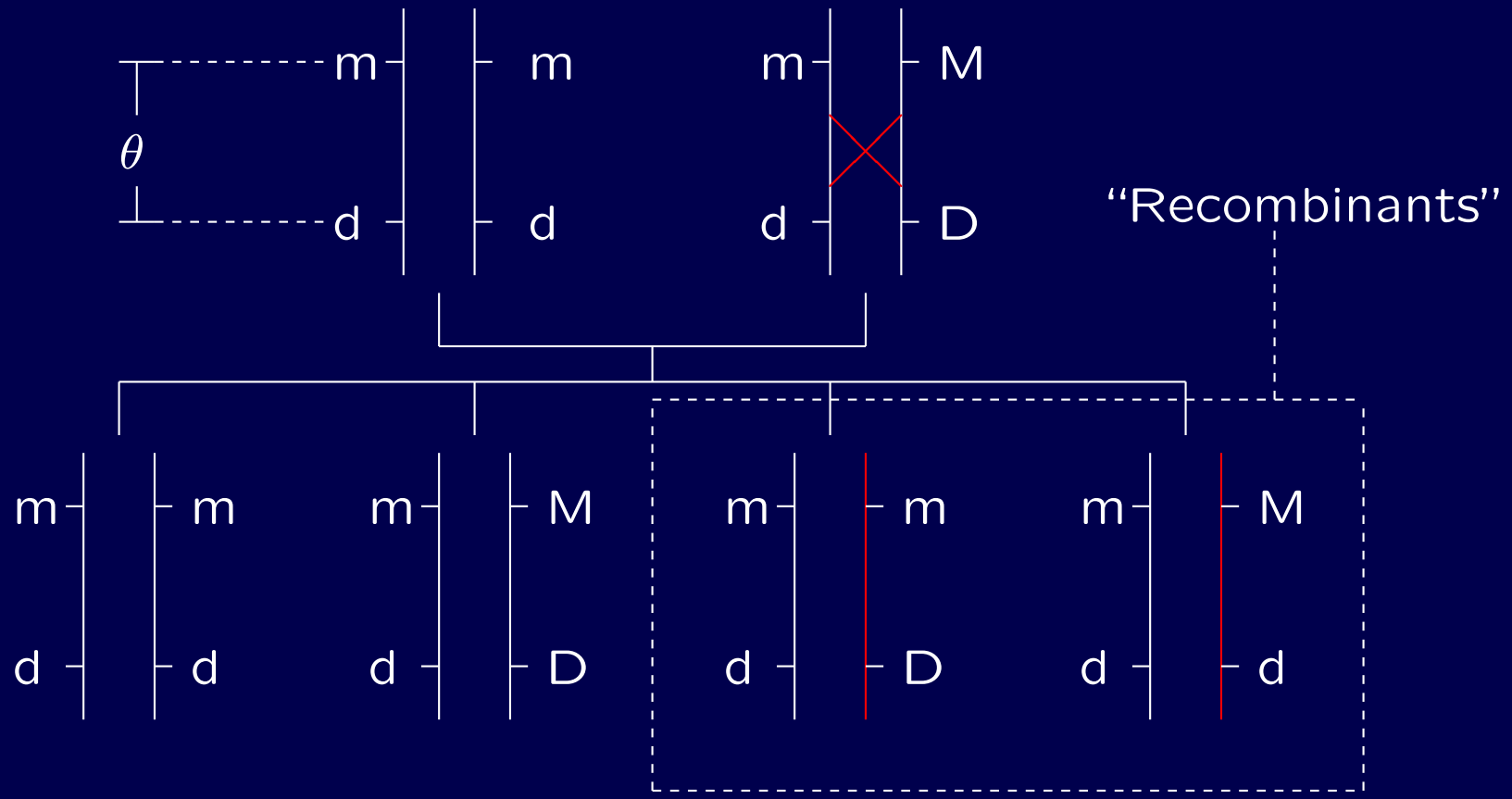
Appyling the Partitioning Principle to Genome Scans

James A. Rogers (Cantata Pharmaceuticals)

joint work with

Jason C. Hsu (Ohio State University)
Shili Lin (Ohio State University)

Genetic Recombination



Model for Simple Data Type

- Suppose one large nuclear family with k children.

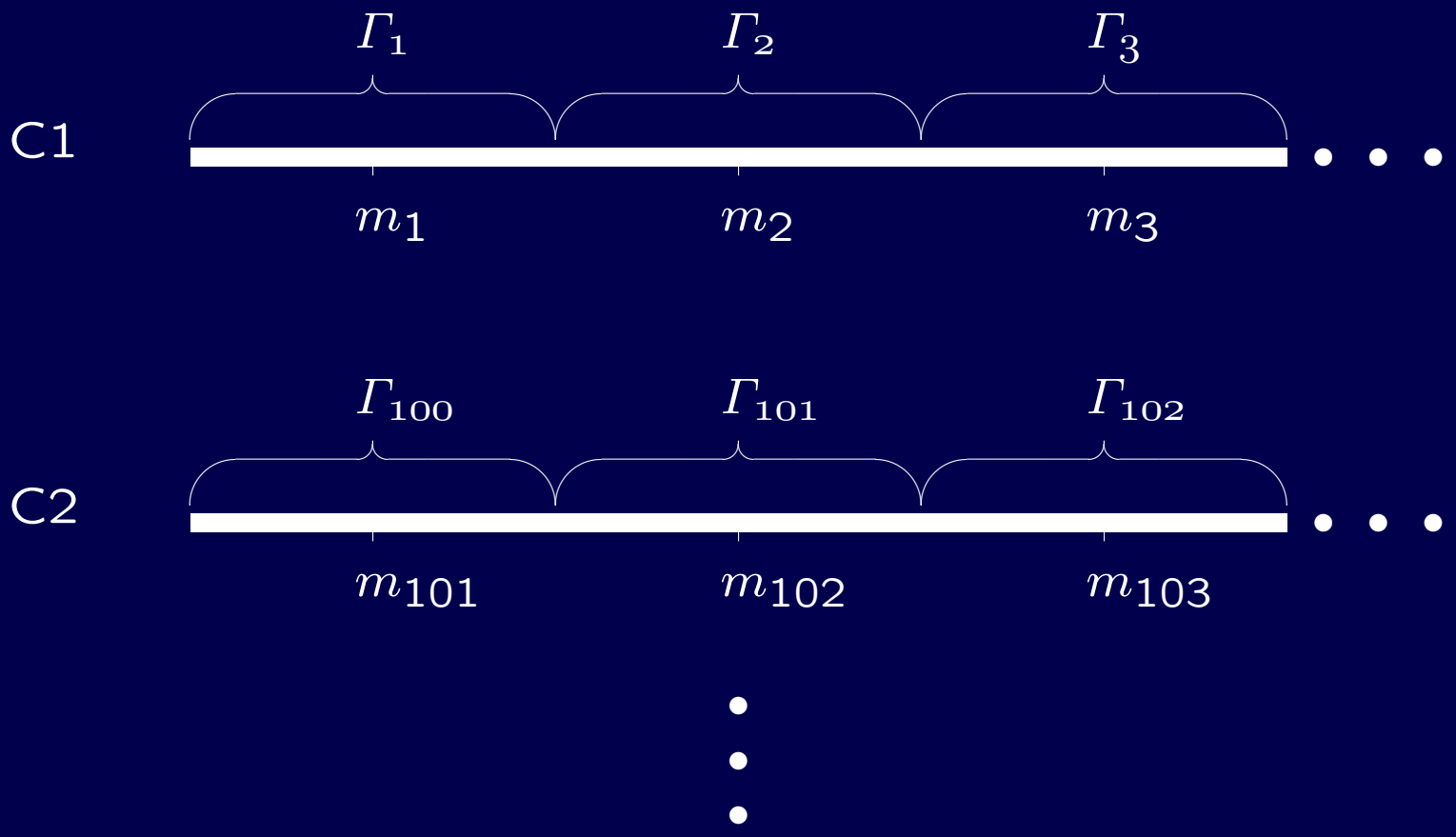
- For each marker m, define

$$R_m = \#\text{ of children in family who are recombinant w.r.t. marker } m.$$

- Then:

$$\begin{aligned} R_1 &\sim \textit{Binomial}(\theta_1, k) \\ &\vdots \quad\quad \vdots \\ R_{1000} &\sim \textit{Binomial}(\theta_{1000}, k) \end{aligned}$$

Partitioning the Parameter Space



Testing Every Possible Disease Locus

- Let γ denote the (unknown) locus of the disease gene.
- For $\gamma' \in \Gamma_i$,
  - $\gamma = \gamma' \iff \theta_i = \theta_i'$ for some θ_i'.
  - we test $H_{0\gamma'} : \gamma = \gamma'$ by testing

$$H_{0\theta_i'} : \theta_i \leq \theta_i' \text{ versus } H_{A\theta_i'} : \theta_i > \theta_i'$$

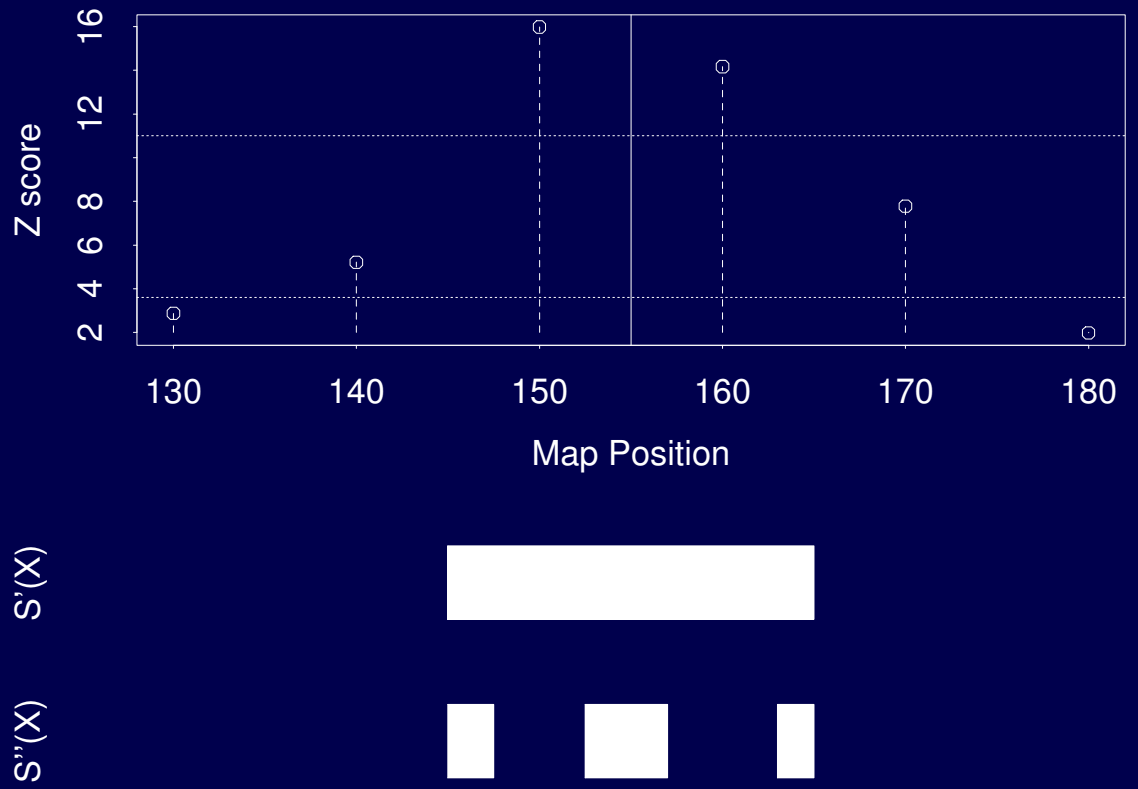
  - use R_i to test $H_{0\gamma'}$ at level α.
- Do this for every γ' on the genome.

Assembling the Confidence Set

- $S(R_1, \ldots, R_{1000}) = \{\gamma' : H_{0\gamma'}$ is accepted$\}$ is a $100(1-\alpha)\%$ confidence set for γ.

- No Multiplicity Adjustment ! (relevant multiplicity is number of disease genes, not number of markers or number of hypotheses).

Application to Simulated Data



To Learn More ...

- Lin, Rogers and Hsu (2001). A Confidence Set Approach for Finding Tightly Linked Genomic Regions. *American Journal of Human Genetics* **68**, 1219–1228.

- email me at: rogers@cantatapharm.com