A LINKAGE ANALYSIS USING HIGH-DENSITY SNP MARKER DATA

Yuki Sugaya
Keio University, Japan
sugaya@stat.math.keio.ac.jp

Ritei Shibata
Keio University, Japan

Ian Saunders
CSIRO, Australia

Recent works in genomics

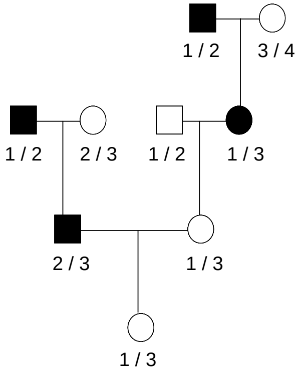
- A large amount of genomic data is available

Post Genome

Relationship between DNA sequence and phenotype?

- Linkage analysis

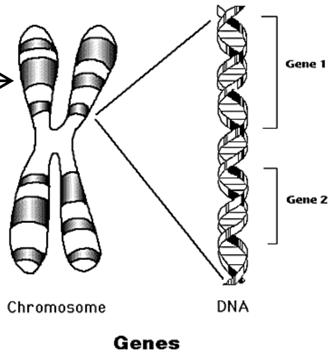
pedigree data



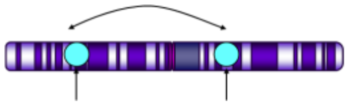
where?

Likelihood

disease locus



linkage



marker locus (known)

disease locus (unknown)

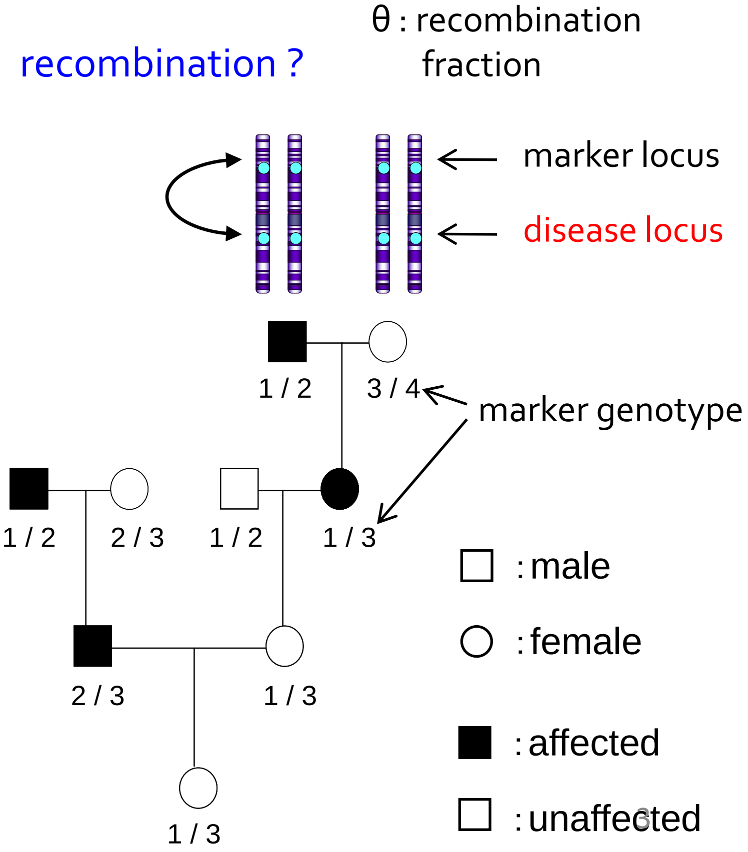
Likelihood of pedigree data

- Pedigree data
  - Relationship between parents and their children
  - Attributes of each individual
    - Gender
    - Affected status ⇨ A_v
    - Marker genotype ⇨ M_v $\quad v \in V$

- Likelihood of pedigree data

$$L(\theta) = P(\mathscr{A}(V) \cap \mathscr{M}(V))$$

$$\begin{cases} \mathscr{A}(V) = \{A_V = a_V\} \\ \mathscr{M}(V) = \{M_V = m_V\} \end{cases}$$



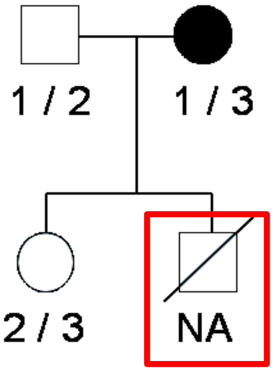
High-density SNP marker data

- Colorectal cancer data
- 113 pedigrees
- 1698 individuals
- 57420 SNP makers

Preliminary analysis

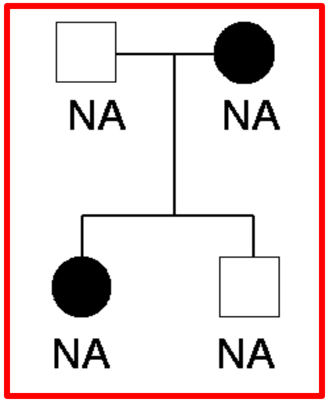
- Pedigree data have a lot of missing values (especially marker genotypes)

Missing individual



Affected status ×
Marker genotype ×

Missing pedigree



Marker genotype ×

no effect on
estimating MLE of θ

→

We can exclude
597 individuals
19 pedigrees
from data

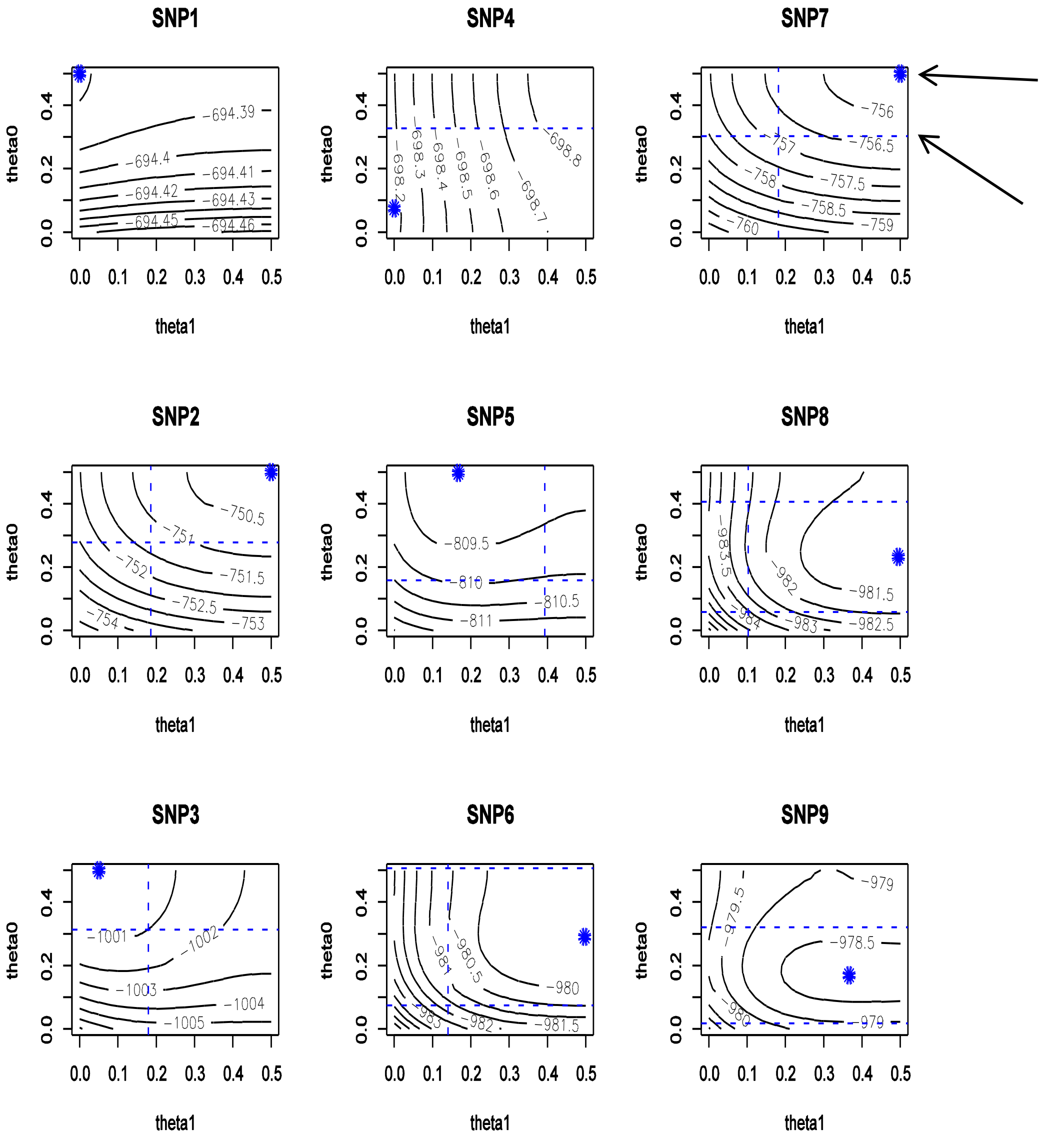
Some results

- Probability Inheritance Algorithm (Sugaya, Shibata)
- Recessive model (penetrance:(0.041, 0.041, 0.85), disease allele frequency:0.27) (Saunders)
- 3 regions examination

sib-pair analysis (Saunders)

- Region 1 (41 markers) ○
- Region 2 (41 markers) ○
- Region 3 (31 markers) ×

Region 1

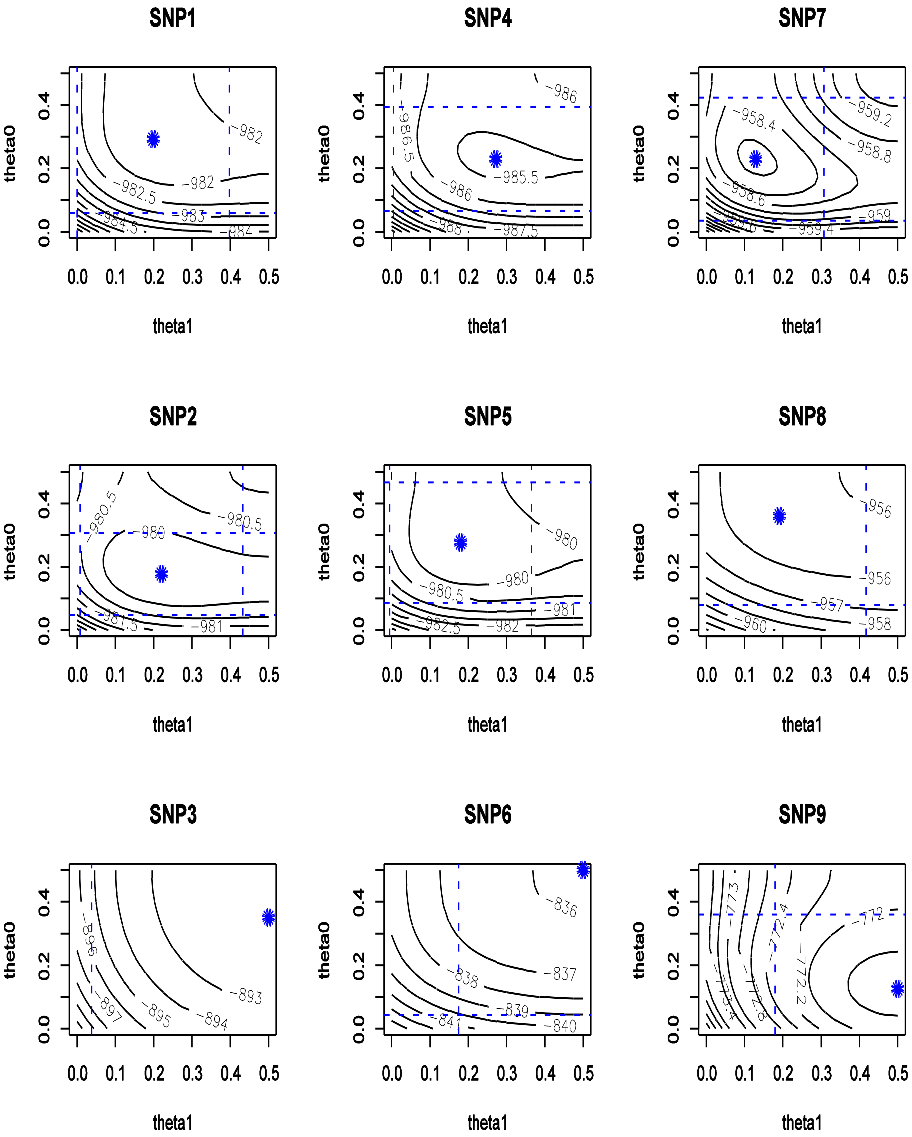


MLE

Standard deviation

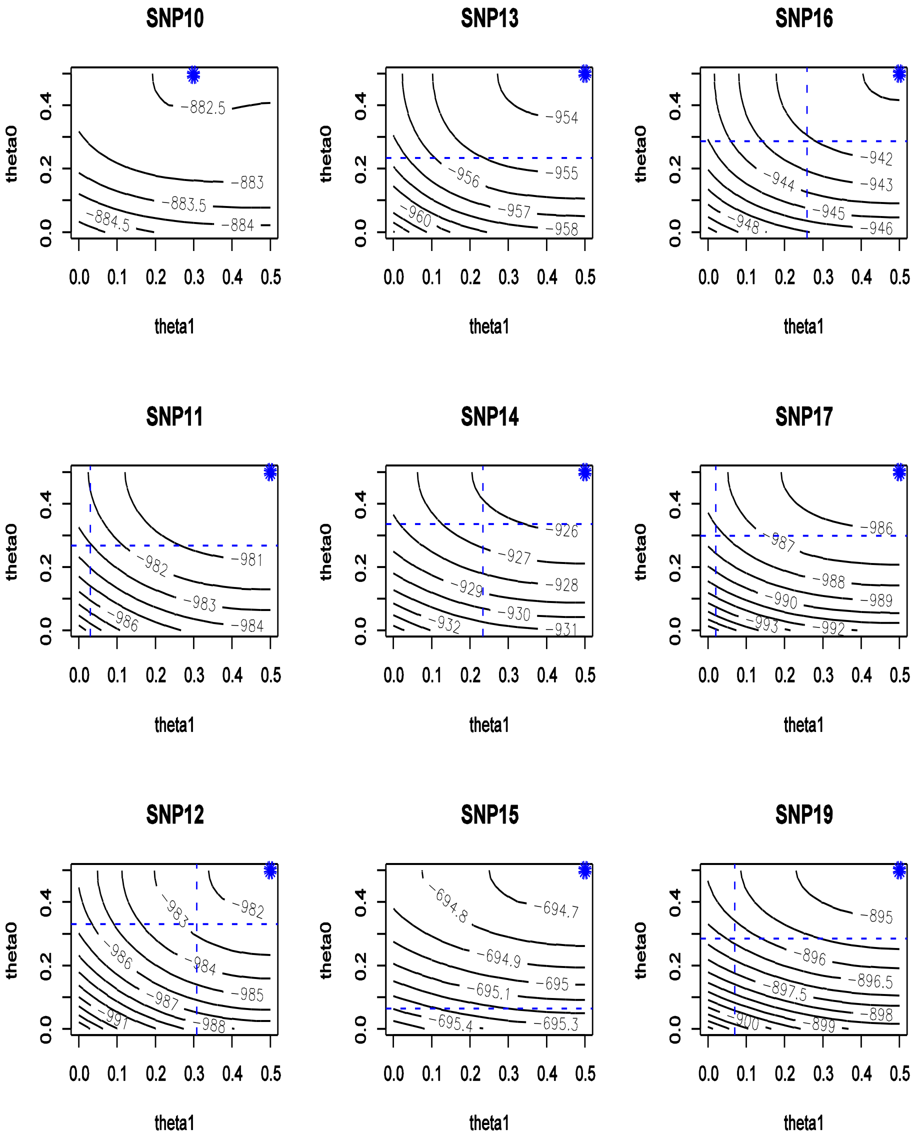
?

Region 2



Disease gene may be close to this region

Region 3

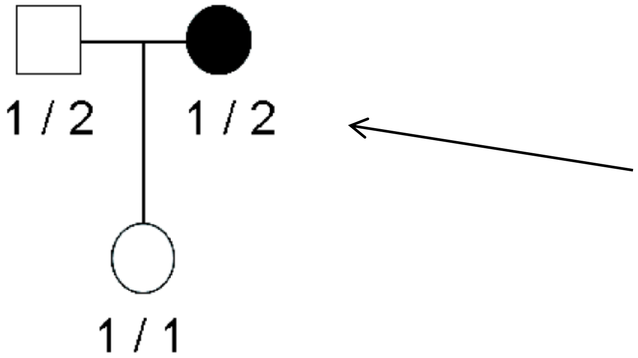


Disease gene may be far from this region

Future works

- Data quality
  - Pedigree data contain genotyping errors (error rate : π)

(Saunders et al; 2007)



1 / 1 : $\pi(1-\pi)$

1 / 2 : $\pi^2 + (1-\pi)^2$

2 / 2 : $\pi(1-\pi)$

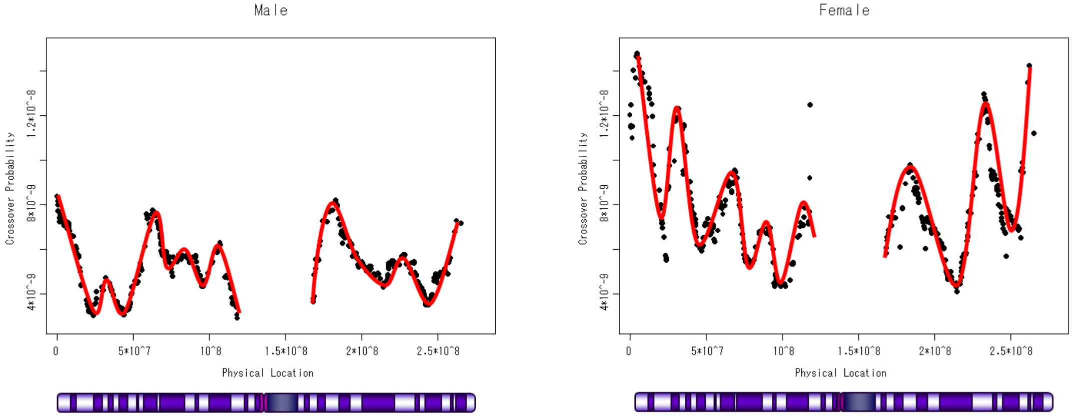
Estimation of physical distance

- A stochastic model of recombination fraction (Karlin ; 1994)

$$\theta = \frac{1}{2}\left\{1-(1-2s)^n\right\}$$

s : crossover probability
θ : physical distance

- Change of crossover probability (Sugaya, Shibata)



$$\hat{n}_i = \frac{\log(1-2\hat{\theta}_i)}{\log(1-2s_i)}$$