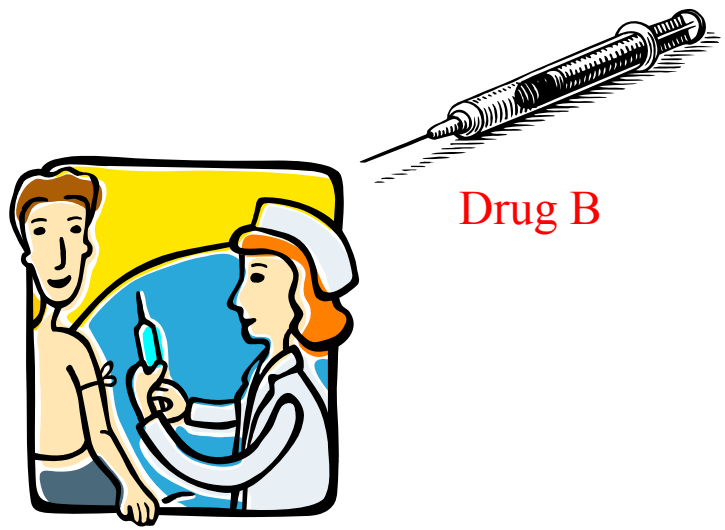
Phasing for Pedigrees when SNPs are Densely Placed

Bonnie Kirkpatrick

University of California Berkeley
International Computer Science Institute, Berkeley, CA

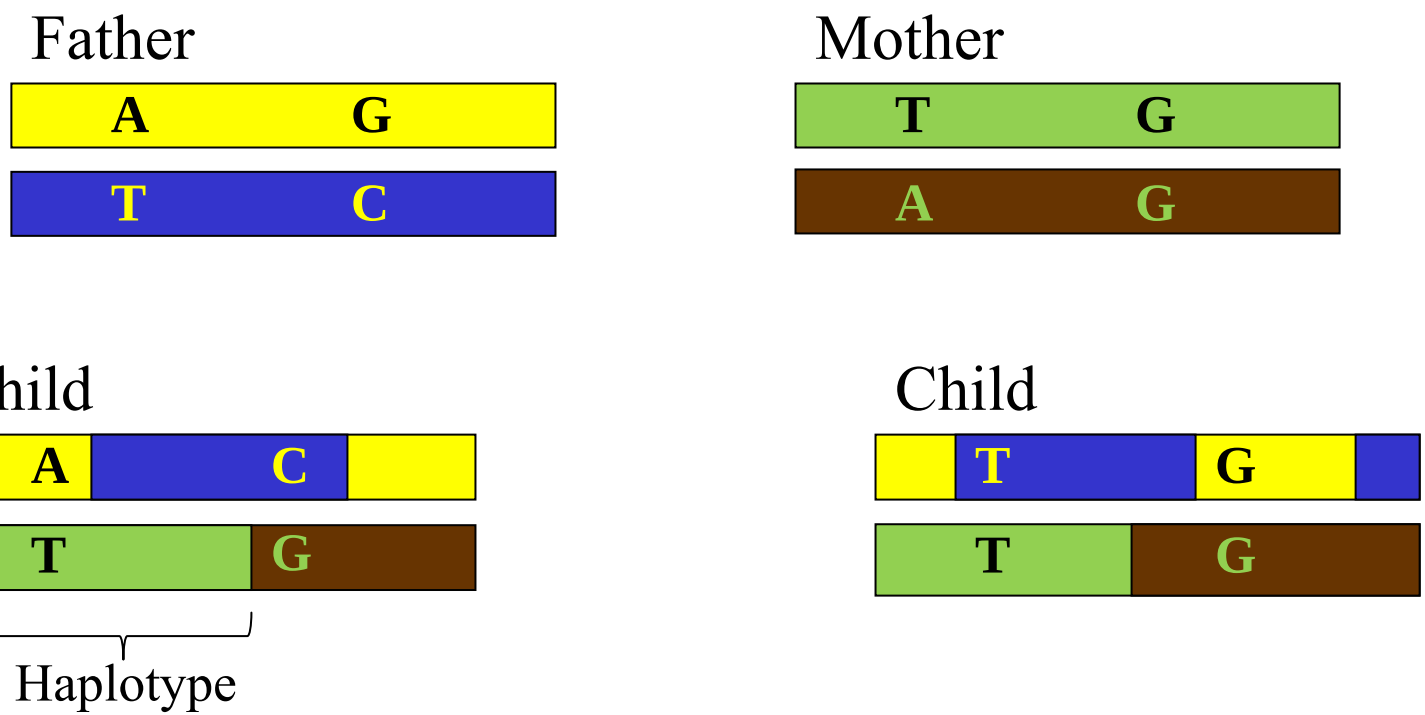
Personalized Medicine

- Understand the mechanisms of genetic diseases
- Diagnose disease according to their causes, rather than symptoms
- Individualize treatment, both to the cause and to the drug response of the individual



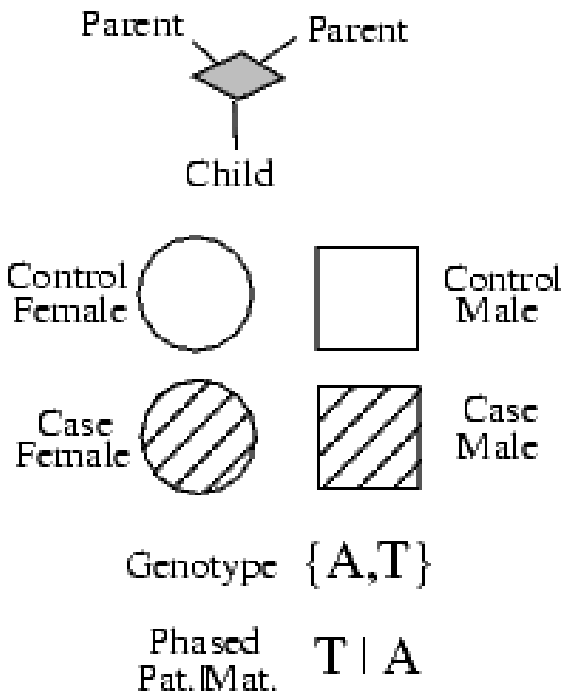
Genetics

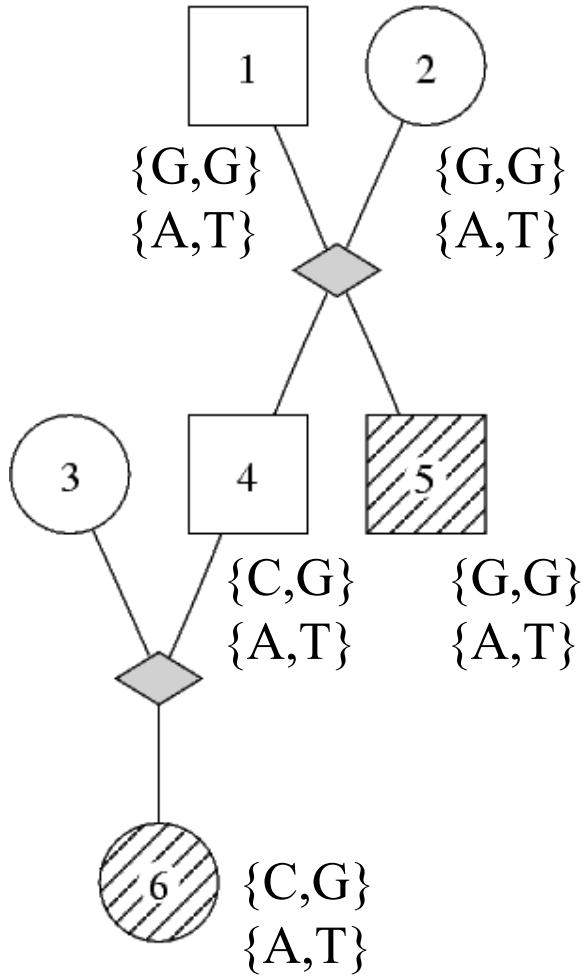
- Two copies of each chromosome (diploid)
- SNPs



Setting

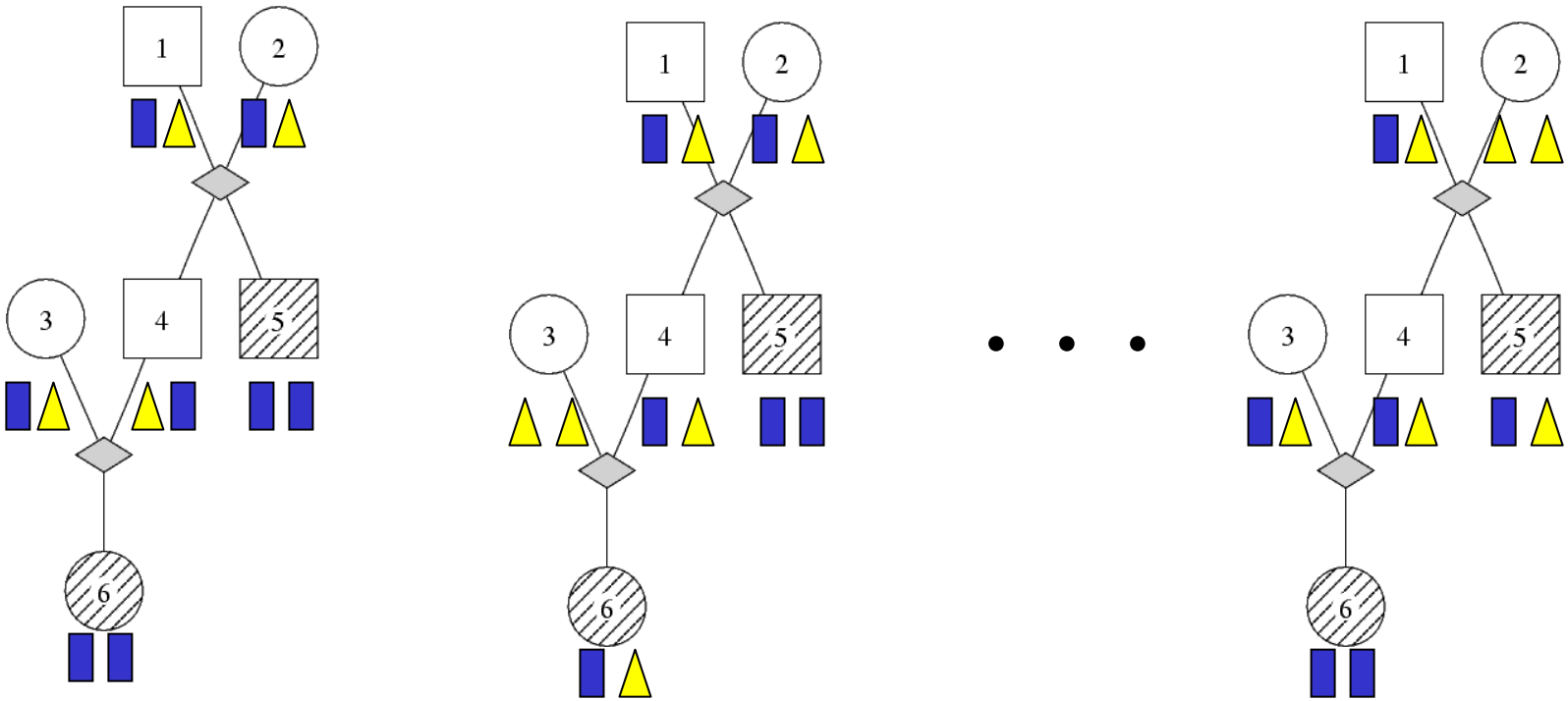
- Related people
- Cases and controls





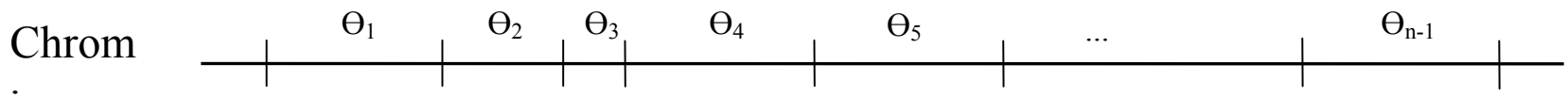
Paradigm

- Locate disease genes
- using association testing or linkage analysis

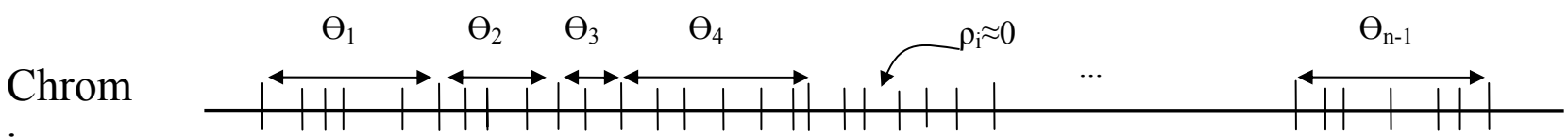


No Recombination

Sparsely located sites:

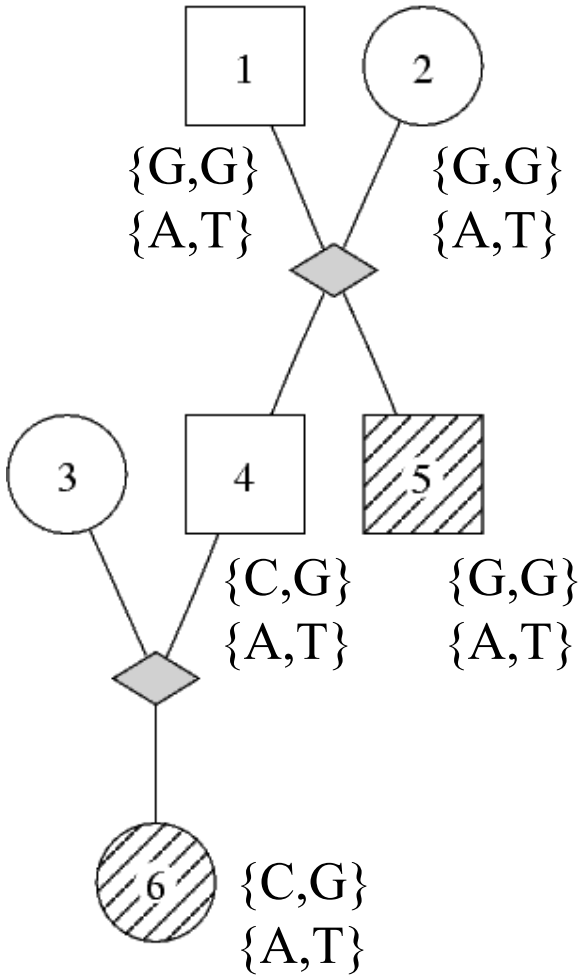


Densely located sites:

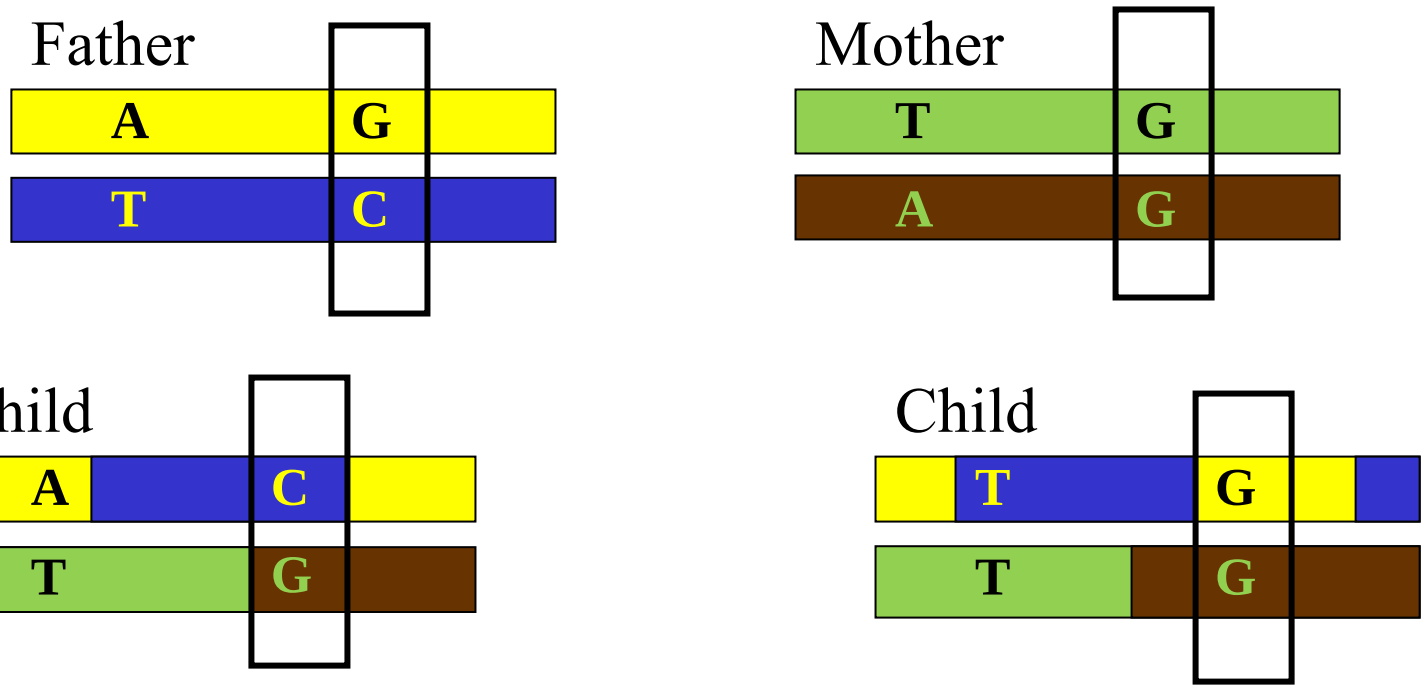


Problem

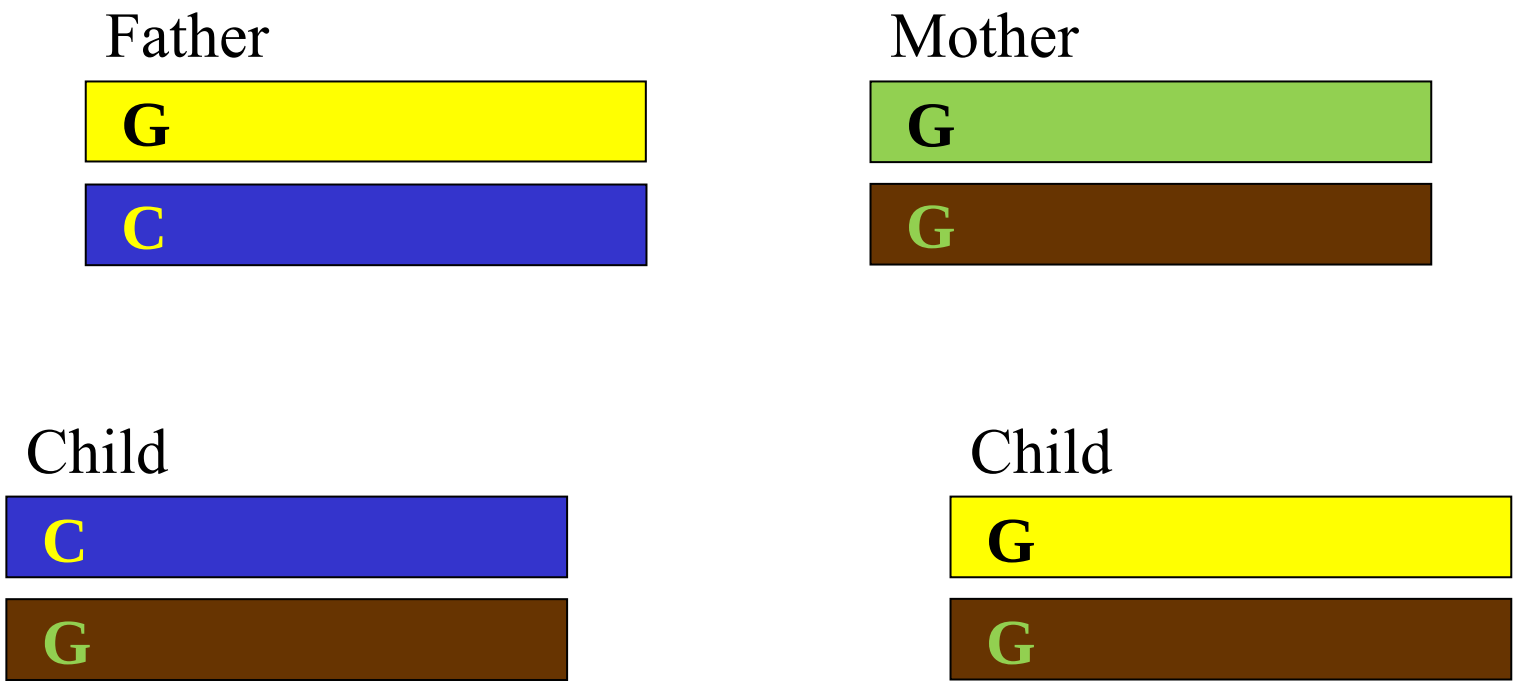
- Find shared genetic variants
- When given:
  - Genotypes for some individuals
  - Mendelian inheritance
  - No recombination



Haplotype Inferrence

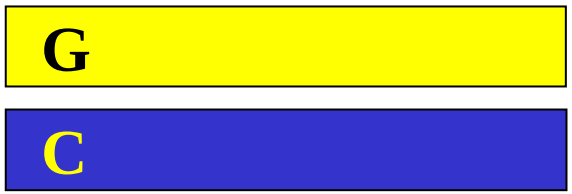


Haplotype Inferrence



Haplotype Inferrence

Father



{T,T}

Mother

{A,T}

Child

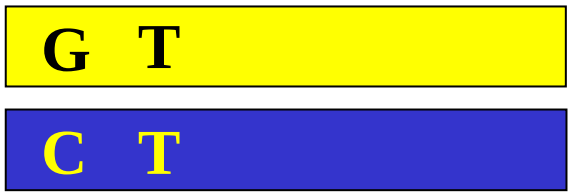
{A,T}

Child

{A,T}

Haplotype Inferrence

Father



{T,T}

Mother

{A,T}

Child

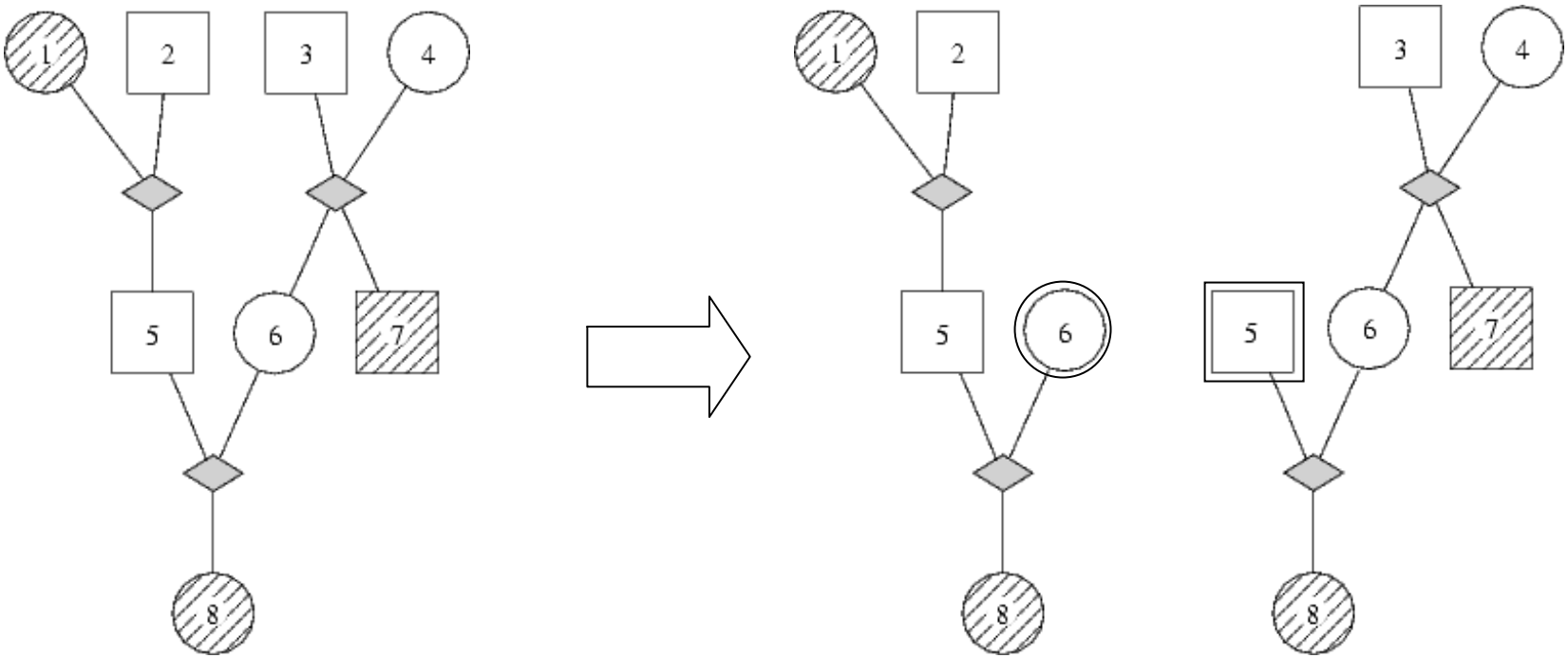
{A,T}

Child

{A,T}

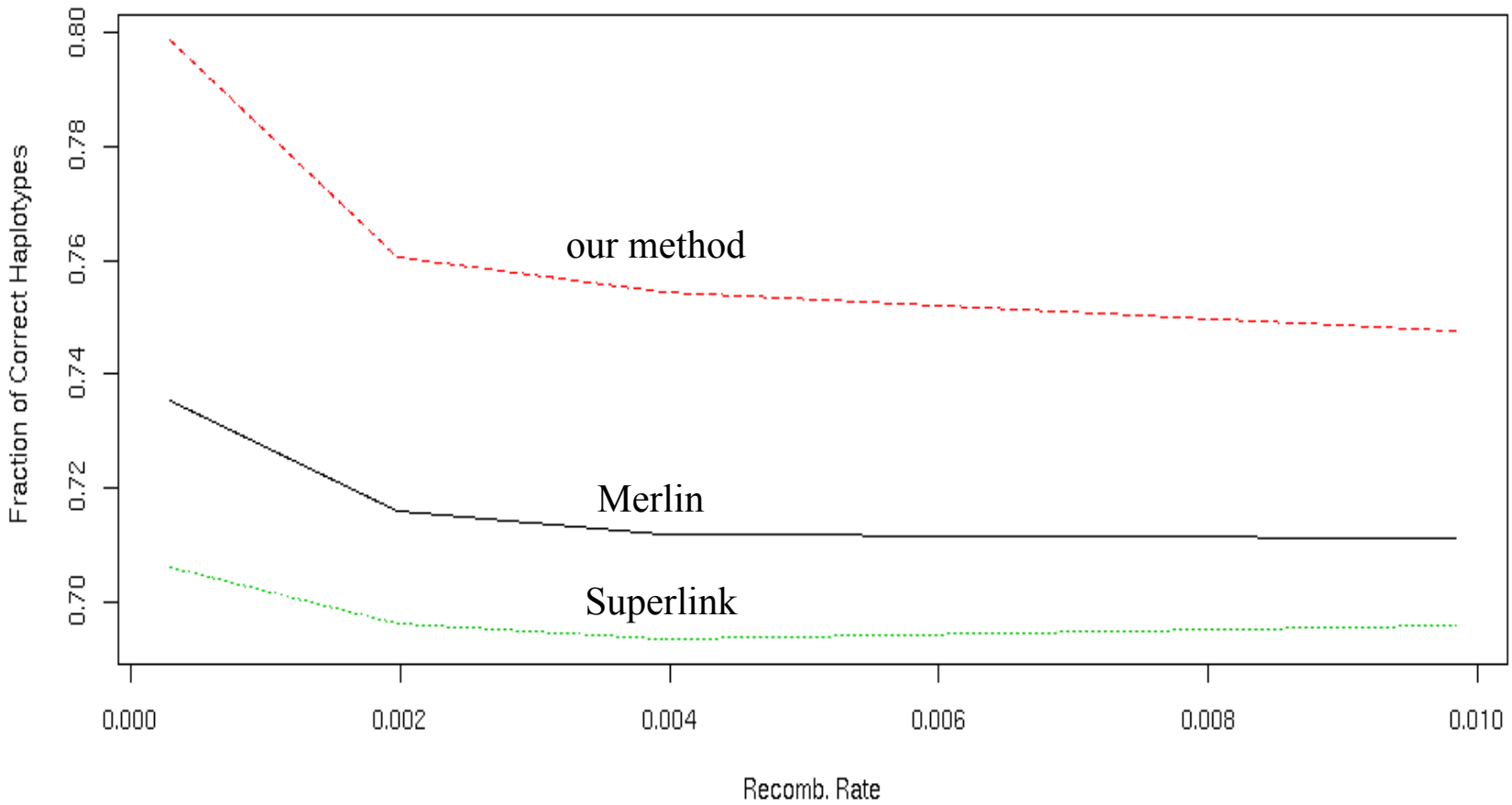
Lineage Decomposition

- Lineage – descendents of a married-pair of founders
- Conditionally independent belief propagation



Input Pedigree **Lineage 1,2** **Lineage 3,4**

Accuracy


Fraction of Correct Haplotypes
0.70
0.72
0.74
0.76
0.78
0.80
our method
Merlin
Superlink
0.000
0.002
0.004
0.006
0.008
0.010
Recomb. Rate

Conclusions

- Novel algorithm for estimating haplotypes for related individuals
- Better accuracy than existing methods

Thanks!

Joint work with:

- Richard Karp
- Eran Halperin
- Javier Rosa

Funding:

- DOE CSGF & NSF Graduate Fellowship (B.K.)
- NSF Grant (R.K. and E.H.)
- NSF REU (J.R.)