



# A method for genotype phasing, imputation, and meiotic crossover detection from low-coverage single-cell sequencing of gametes

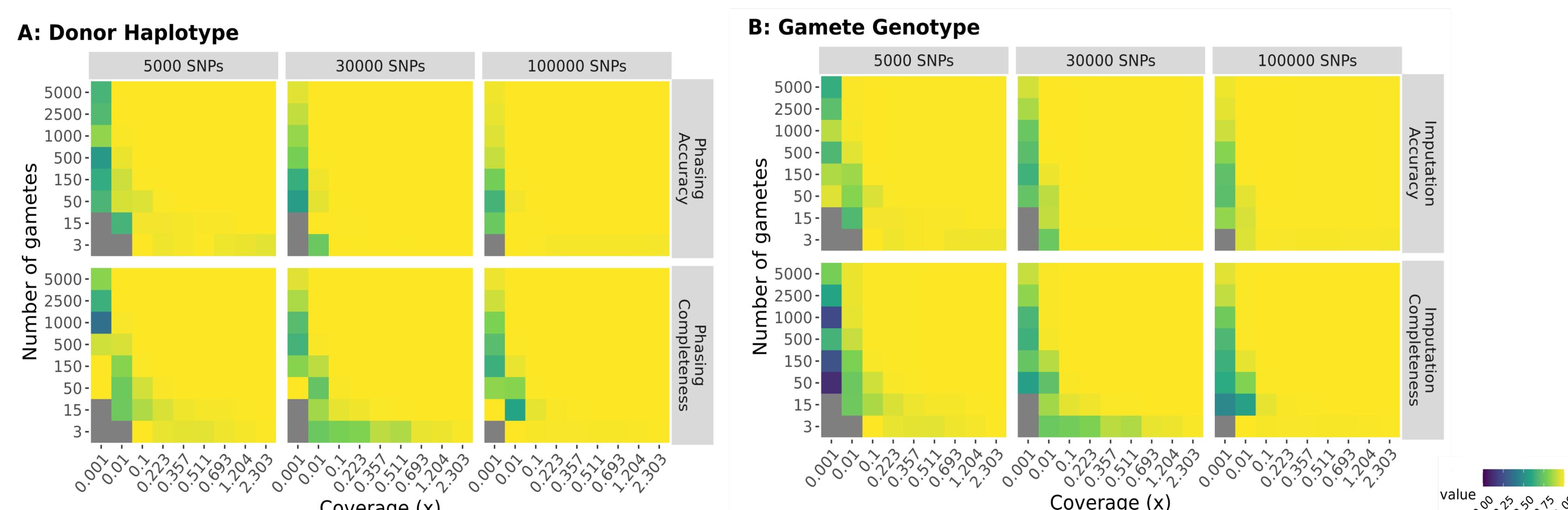
Andrew N. Bortvin, Kathryn J. Weaver, Sara A. Carioscia, Rajiv C. McCoy

Department of Biology, Johns Hopkins University

## Background

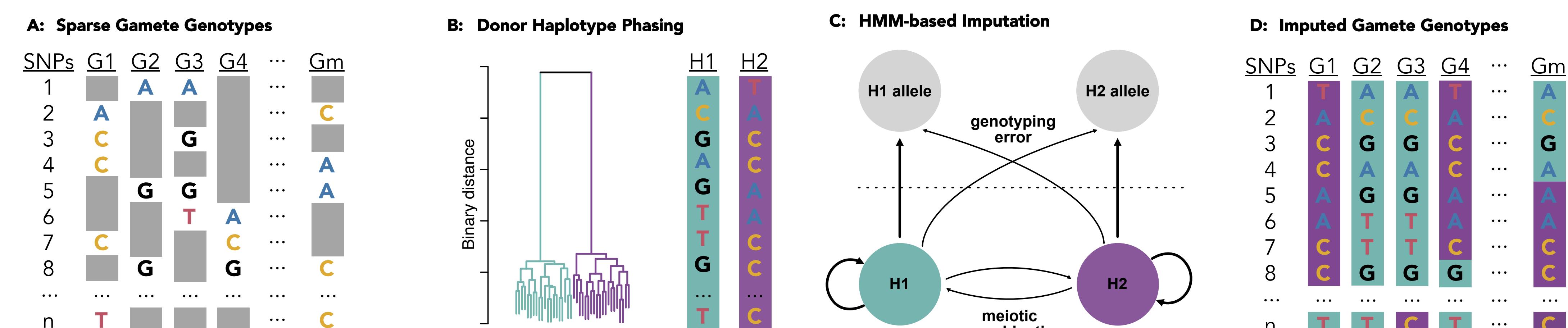
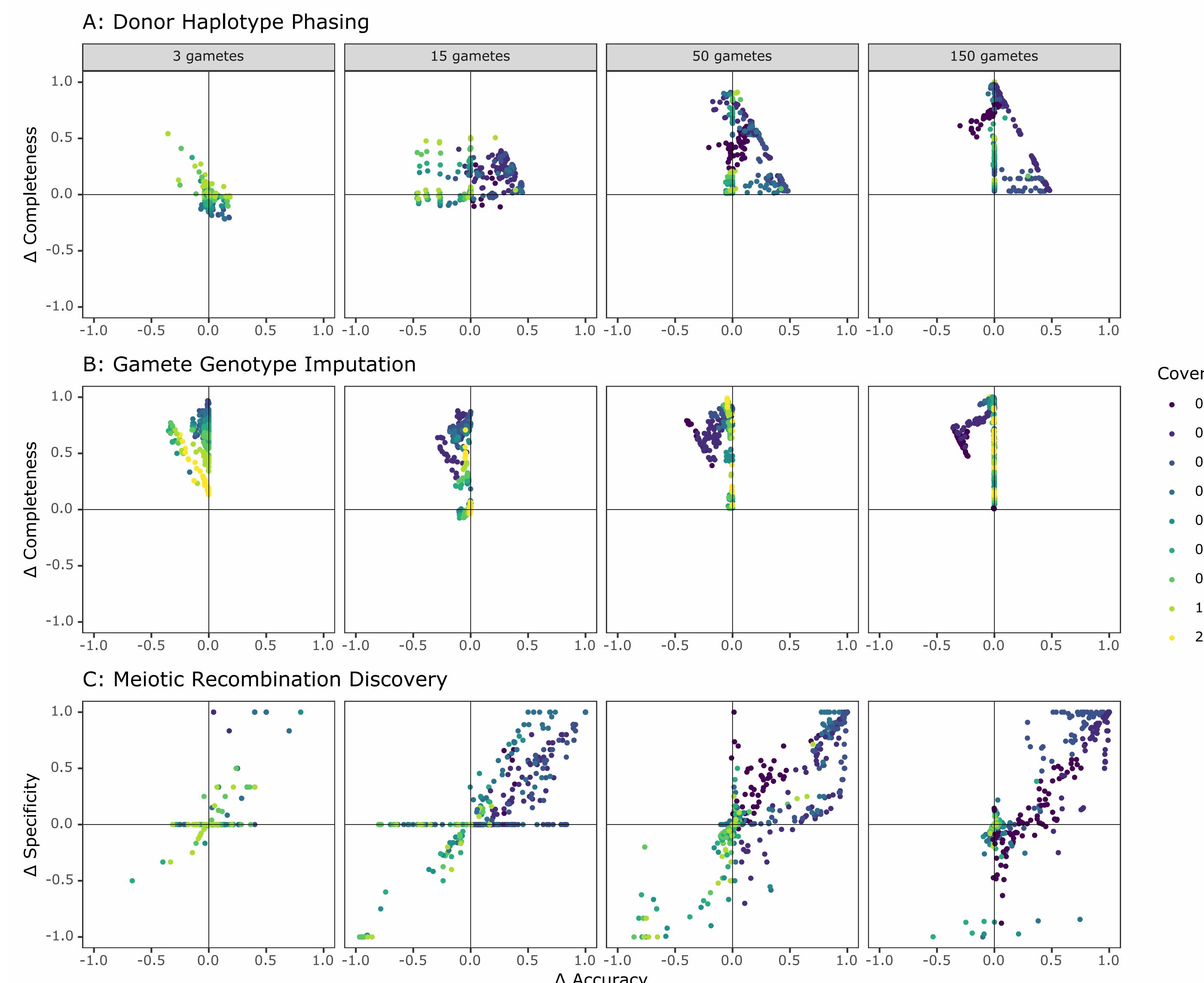
- ◆ Single sperm DNA-sequencing (Sperm-seq) produces low-coverage data for thousands of sperm from a donor.<sup>1</sup>
- ◆ Our package, rhapsodi, phases diploid donor haplotypes, imputes gamete genotypes, and discovers cell-specific meiotic recombination events.
- ◆ We applied rhapsodi to a sample of 41,189 sperm from 25 donors<sup>1</sup> and conducted a genome-wide scan for transmission distortion (TD), or unequal transmission of "selfish" alleles.

## Benchmarking with Simulated Data



## Comparison to Existing Software

- ◆ We benchmarked rhapsodi against Hapi<sup>2</sup>, an existing program designed for the same task and demonstrated to outperform all prior tools.

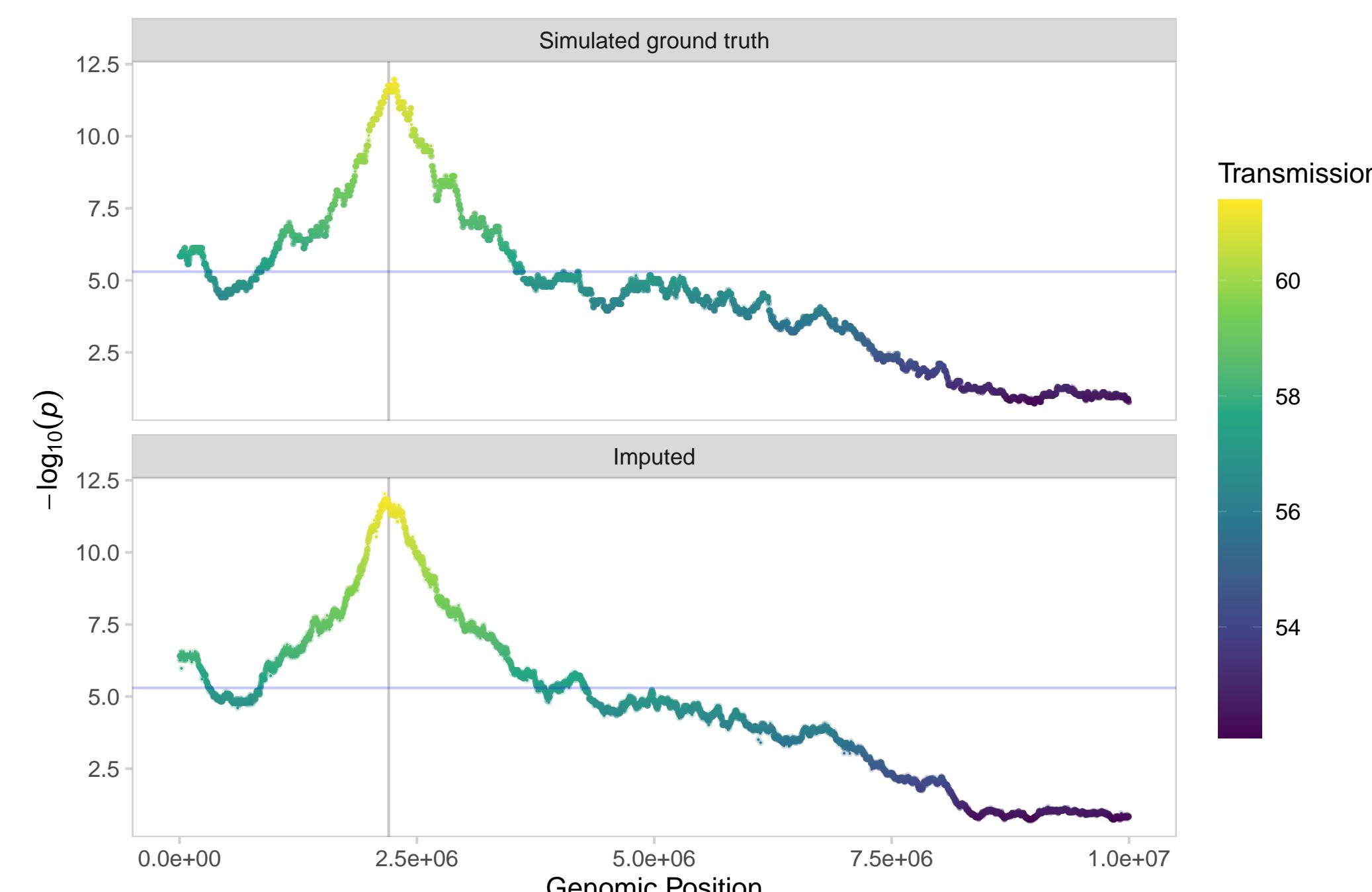


## Method

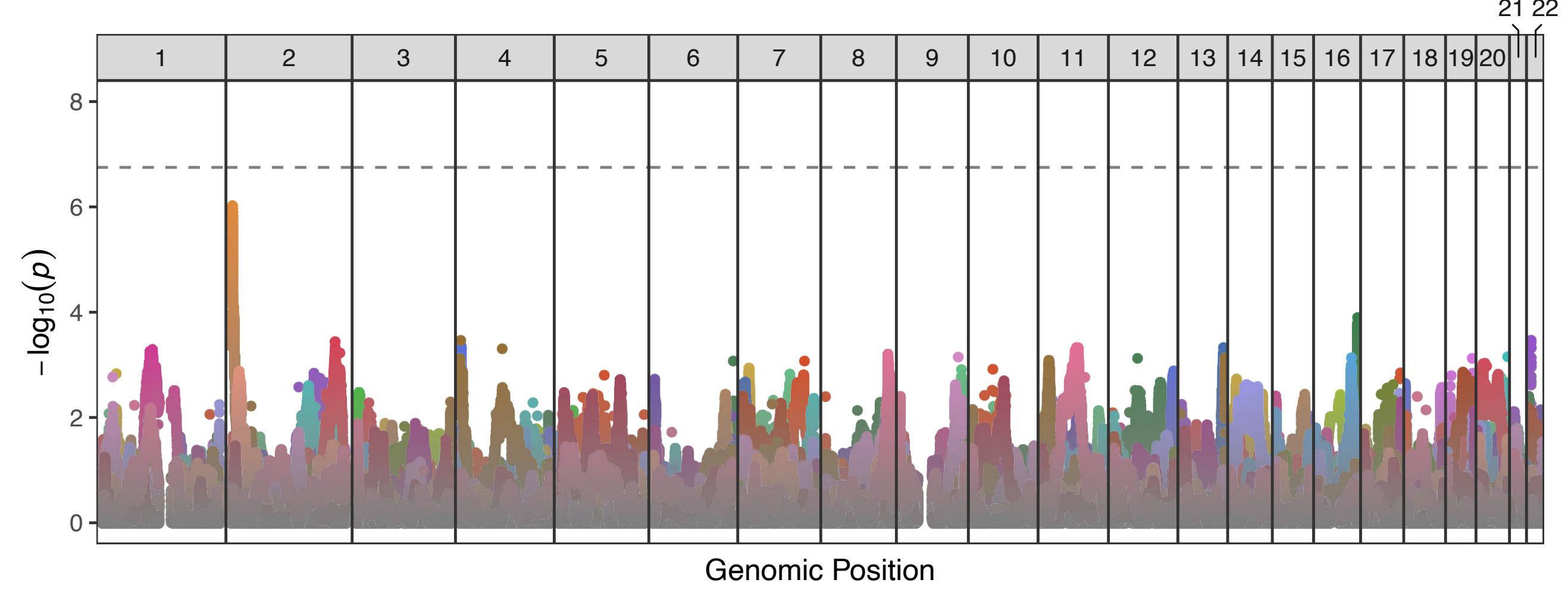
- ◆ Binary clustering of extremely low coverage data from individual gametes (**A**) phases the two donor haplotypes (**B**), whose sequence of alleles can be reconstructed by majority vote.
- ◆ A hidden Markov model (**C**) is used to trace the most likely path along the phased haplotypes for each gamete.
- ◆ Transitions between the haplotype states indicate the occurrence of meiotic crossovers (**C**).
- ◆ Missing data are imputed based on the inferred sequence of haplotype states (**D**), recovering the dense genotype matrix.

## Genome-wide scan for TD

- ◆ We simulated Sperm-seq data with known levels of TD and evaluated rhapsodi's ability to correctly phase and impute these chromosomes, and our ability to subsequently detect signatures of TD.



- ◆ We applied rhapsodi to a sample of 41,189 sperm from 25 donors. We scanned imputed gamete genomes for TD. After multiple testing correction, we find that all gametes in this cohort adhere to Mendelian expectations.



## Conclusions

- ◆ Using sparse genotypes, like those from Sperm-seq, we reconstruct individual gamete haplotype structure by exploiting the shared content within sperm cohorts.
- ◆ Accuracy and completeness of phasing and imputation increase with coverage, as does resolution for detecting recombination breakpoints. Yet high accuracies are achieved even at low coverage when many gametes are included.

- ◆ Once missing genotypes are imputed, this data can be used for analyses such as identifying recombination hotspots, exploring signatures of selection, or finding *de novo* mutations.

- ◆ We apply this method to human Sperm-seq data, demonstrating adherence to Mendel's First Law.

Download the R package, rhapsodi:

