



Figure 2-figure supplement 4 Automatic phasing window size calculation. A beta regression model relating optimal phasing window size (represented as window size / of SNPs) to number of gametes, coverage, genotyping error rate, and recombination rate was fit on the training data and applied to the held-out test set. The number of gametes, coverage, and recombination rate were significantly associated with optimal window size. The model performed well with both the training and test sets, with no obvious loss when generalizing to new data. This model is implemented as an optional feature for automatic phasing window size calculation, given the input data parameters, within the rhapsodi software package.