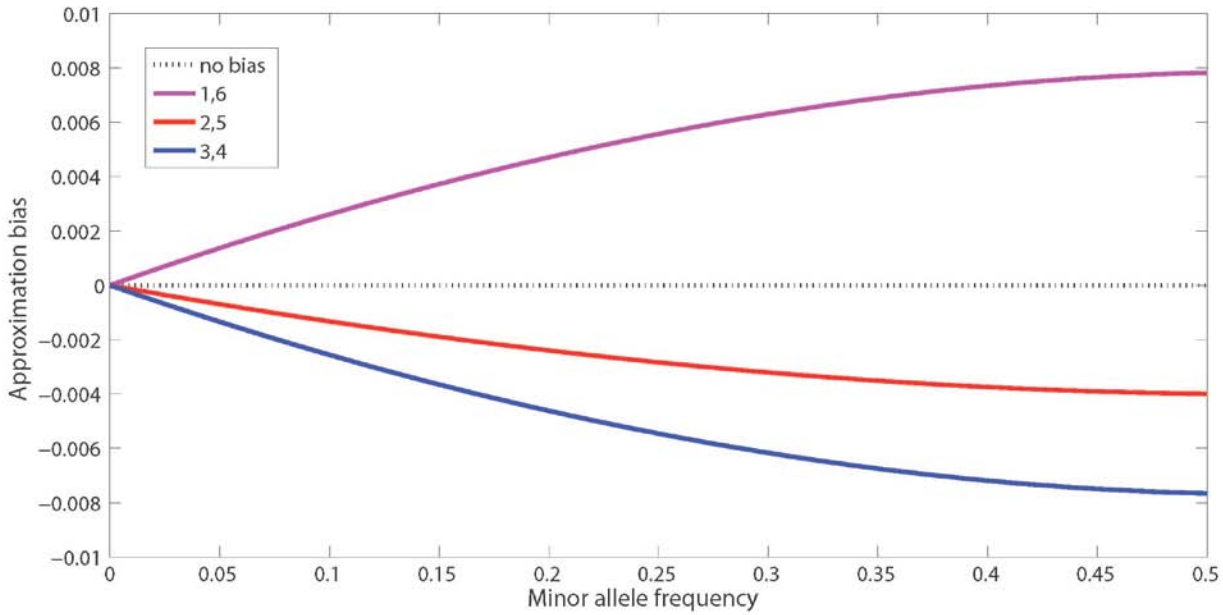
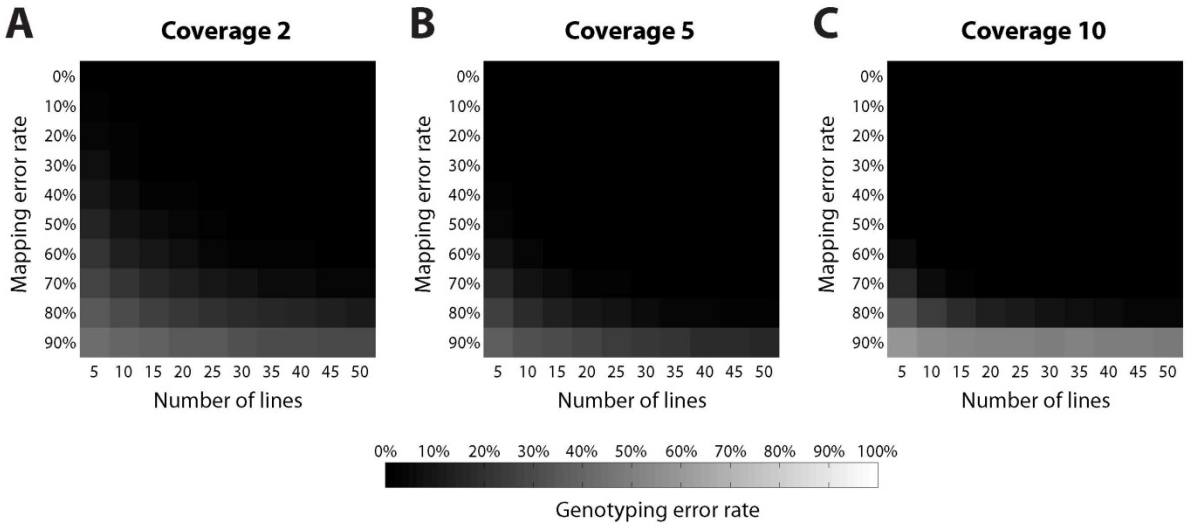
SUPPLEMENTAL FIGURE 1 | Approximation bias in $\Pr(\mathbf{G}\,|\,\mathbf{p})$ as a function of allele frequency

Under the assumption that two alleles, A and a, are segregating in the population with frequencies p and $q = 1 - p$ respectively, the figure plots approximation bias against q for the G20 probabilities of (1) AA x AA, (2) AA x Aa, (3) AA x aa, (4) Aa x Aa, (5) Aa x aa, and (6) aa x aa. The exact values are given by the vector $\mathbf{x} = \left(p^4, 4p^3q, 2p^2q^2, 4p^2q^2, 4pq^3, q^4\right)\mathbf{Q}^{20}$ where $\mathbf{Q}$ is the the Markov transition matrix introduced in Methods. The approximations for states 1 and 6 are $p^2(1-F)+pF$ and $q^2(1-F)+qF$ respectively, and their deviation from the exact values is plotted in magenta. The approximations for states 2 and 5 are both $(3/5)\,pq(1-F)$ and their deviation from the exact values is plotted in red. States 3 and 4 describe the scenario in which two of each allele are represented in G20. These are treated collectively and approximated by $(4/5)\,pq(1-F)$. Shown in blue is the deviation of this approximation from the sum of the third and fourth entries of $\mathbf{x}$. The figure demonstrates that the approximation causes a weak bias towards homozygosity. This has a very modest effect on the estimates of $\boldsymbol{\theta}$ and the calculation of posterior probabilities.



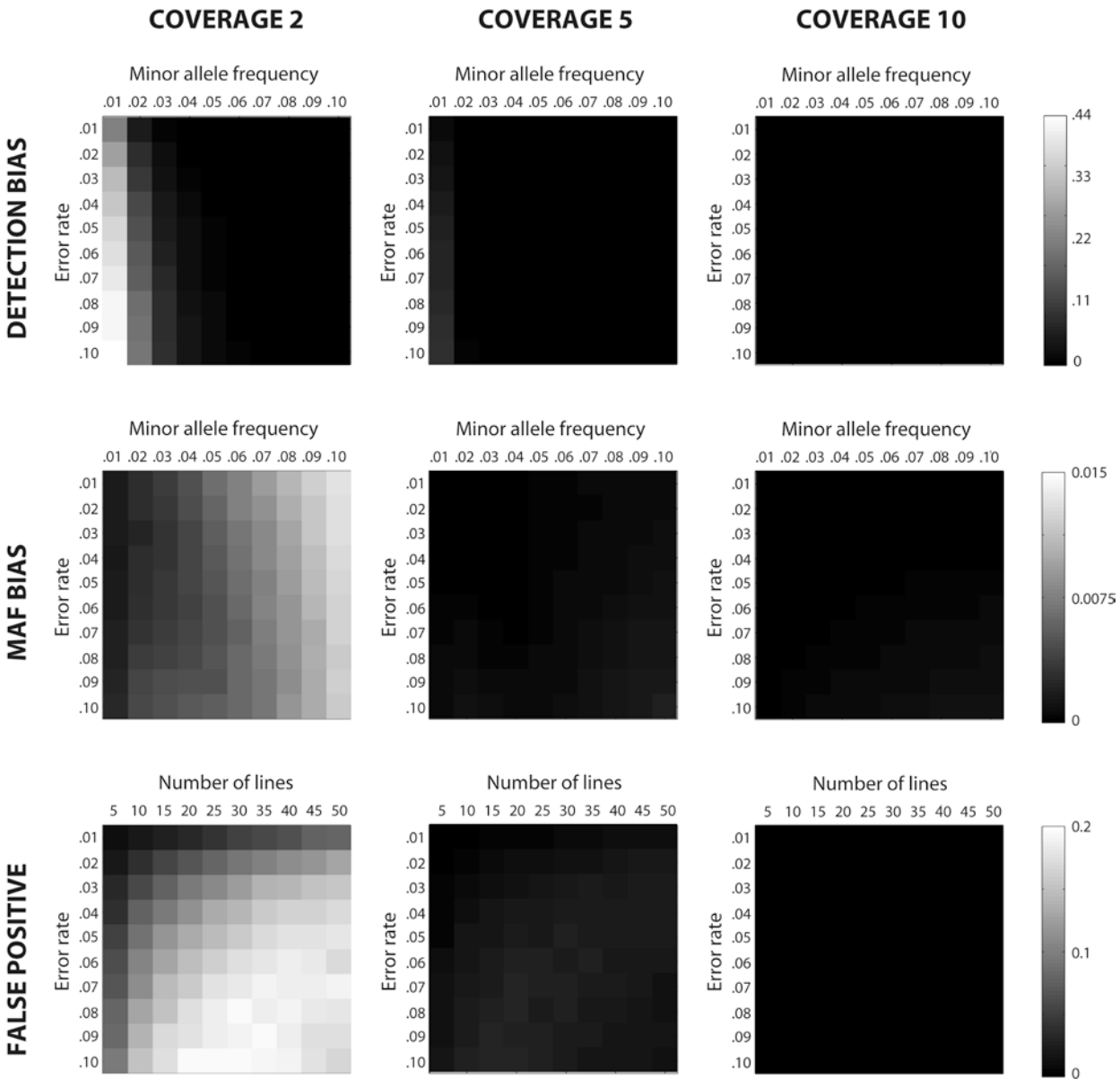
Supplemental Figure 2 | Genotyping error rate as a function of mapping error and number of lines

At a single site, sequencing read data was simulated for a varying number of lines assumed to be homozygous for the nucleotide "A". The expected coverage for "A" was set at either (*A*) 2 reads, (*B*) 5 reads, or (*C*) 10 reads. Mapping error was simulated by setting the expected coverage for a second erroneous nucleotide as a percentage of the expected coverage of the correct base. For erroneous coverage levels ranging from 0% to 90%, the JGIL genotyping error rate was calculated across 10,000 replicate simulations.



Supplemental Figure 3 | Bias in SNP detection and frequency estimation with varying coverage and error rate

The top row of the figure shows ascertainment bias as a function of error rate (vertical axis), minor allele frequency (horizontal axis), and coverage (by figure column). The simulation was based on 100 lines and the value reported is the proportion of 10,000 replicates in which the SNP went undetected by JGIL. Under the same simulation conditions, the middle row of the figure reports the bias of the minor allele frequency estimate. The value reported is the specified minor allele frequency minus the mean minor allele frequency across the 10,000 replicates. In the bottom row of the figure, the minor allele frequency is set to zero to indicate a monomorphic site, and the horizontal axis indicates a varying number of lines. The value reported is the proportion of 10,000 replicate simulations in which JGIL mistakenly called a SNP.


COVERAGE 2
COVERAGE 5
COVERAGE 10
DETECTION BIAS
MAF BIAS
FALSE POSITIVE
Minor allele frequency
Number of lines
Error rate
.01 .02 .03 .04 .05 .06 .07 .08 .09 .10
5 10 15 20 25 30 35 40 45 50
.44
.33
.22
.11
0
0.015
0.0075
0.2
0.1