**Supplemental Methods**

**Generating simulated reference panels across a range of diversity levels**

To evaluate low-coverage reconstruction for various degrees of genetic diversity, we generated reference panels using haplotypes produced by coalescent models across a range of genetic diversity levels. Haplotypes were generated using the R (R Core Team 2016) package scrm (Staab *et al.* 2015) and subsequently restructured into VCF file format (Danecek *et al.* 2011). We generated ten independent panels for each of all 18 combinations of population size (Ne=104, 105, 106), mutation rate (μ=10-9, 5×10-9, 10-8), and number of haplotypes (32, 128). The value θ for each simulation was defined as 4Neμ. We simulated a chromosome-length locus of 25 Mb with a recombination rate of 1.5 cM/Mb. SNP positions output by scrm (a decimal within the range of 0 to 1) were converted to base pair positions by multiplying the decimal by chromosome length (25 ×106 base pairs for our simulations) and rounding down to the nearest integer. Any sites with more than two alleles were converted to a biallelic site by discarding tertiary or quaternary alleles. Genotype values were re-coded as polarized signed integers: +1 for reference and -1 for alternate alleles. For every position, reference and alternate alleles were defined by randomly selecting one of the twelve non-repeating pairs of nucleotides. Reference genome FASTA files were created with a custom python script that generated a 25 million length string of nucleotide characters with weighted probability to achieve 45% GC-content, followed by replacing variable positions with their respective reference alleles.

**Extended details for simulating GWAS**

*Simulated haplotypes.* Although the forward simulator we developed is efficient, it would not have been computationally feasible to simulate 500 fully independent mapping populations (per parameter combination) in a reasonable amount of time. Instead, we generated ten independent forward-simulated populations, and for each of those, generated fifty randomly permuted subsets. For a single simulated mapping population, we began by sampling (with replacement) a random subset of 5,000 individuals, out of 10,000 total individuals generated by forward-simulation. Then, we performed a permutation of haplotype ancestry with a new, randomly-ordered (equally sized) subset of founders. The permutation of ancestry was one-to-one, e.g. all haplotype blocks that were previously derived from founder X would be translated to founder Y, and blocks previously derived from Y would in turn be mapped to founder Z.

For all simulated GWAS, we began with a set of 129 DGRP haplotypes with the least missing data, e.g. high coverage and low levels of heterozygosity. This allowed us to perform leave-one-out subsampling for the 128-founder populations. In addition to Hybrid Swarm populations, which we ran through the simulated sequencing and mapping pipeline, we generated four additional types of mapping populations for comparing GWAS performance: Highly outbred (F50) populations; Inbred Lines (ILs) to represent the DGRP; and Recombinant Inbred Lines (RILs), similar to the Drosophila Synthetic Population Resource, or DSPR (King *et al.* 2012).

The F50 populations were generated with 128 founders in same manner as the Hybrid Swarm, except that populations were simulated over fifty non-overlapping generations of recombination instead of five generations. The ten resulting forward-simulated populations were resampled and permuted as we did with the Hybrid Swarms.

We simulated ten initial sets of 800 RILs using the same forward-simulator as previously described, each initialized with a random subset of eight DGRP haplotypes. Populations randomly recombined at a population size of 10,000 for fifty non-overlapping generations, after which 800 random male-female pairs of individuals were isogenzied through 25 generations of full-sibling mating. This scenario roughly corresponds to the DSPR. For computational simplicity, after the 25 generations of isogenization we removed any remaining residual heterozygosity by forcing the identity of a second chromosome copy to be identical to the first copy. We then sampled 5,000 draws (with replacement) of the 800 RILs followed by ancestry permutation as described above.

To simulate GWAS on Inbred Lines, no forward-simulation was necessary. For a single simulated population, we first randomly selected 128 DGRP lines, then randomly sample with replacement 5,000 times. As with hybrid swarm and RILs, for any parameter combination we generated a total of 500 mapping populations.

*Liability model.* Using the liability framework, we assign case or control phenotypes to mapping populations derived from DGRP chromosome 2L haplotypes. First, a genotype-dependent risk score is calculated for every individual, which is translated into probabilities for case or control group assignment. We parameterized our model such that individuals begin with equivalent odds of being assigned to case or control groups, and that probability is modified depending on allele status at causal loci.

For a given simulation, N causal loci are randomly selected with equal probability out of all segregating sites in the mapping population. The effect of individual causal loci are simulated as a Gaussian variable drawn with a small extent of noise (sd=0.005) with an expected (mean) value dependent on the allele frequency at that locus. Intermediate frequency loci (i.e. allele frequency = 0.5) are simulated to contribute on average no effect on case/control assignment; alleles near fixation decrease risk; rare alleles increase risk. The linear center of the curve, where most individual risk scores exist under the parameter combinations used in our simulations, approximates additive genetic architecture. The sigmoid tails allow for the continuous risk score to translate to a phenotype necessarily bounded by zero and one. The relationship between allele frequency, risk score, and phenotypic assignment is diagrammed in Supplemental Figure S9. Within the linear portion of the curve (which is the domain for the majority of simulations), a singleton minor allele will modify ‘case’ assignment probability by approximately +5% (small effect) or +10% (large effect), beginning at an initial assignment probability of 50%. The major allele, conversely, would decrease assignment by an opposite amount. An allele at 50% frequency will not modify case/control assignment (aside from a very small degree of random noise equal to about +/- 0.06%). Note that we performed GWAS simulations assuming 100% genotype accuracy.

We implemented a method of efficiently aggregating allele counts compatible with our haplotype map format. Briefly, haplotype map breakpoints across all individuals are sorted in ascending order. When iterating through ascending unique start and stop positions, between any pair of breakpoints, all SNPs will be comprised of the same number of each founding haplotype. Haplotype IDs could then be counted and sorted in the same column position order as the table containing polarized allele status (-1 for alternate, +1 for reference). Multiplying the genotype table by the haplotype count vector results in final allele counts, polarized negative for alternate alleles and positive for reference alleles.