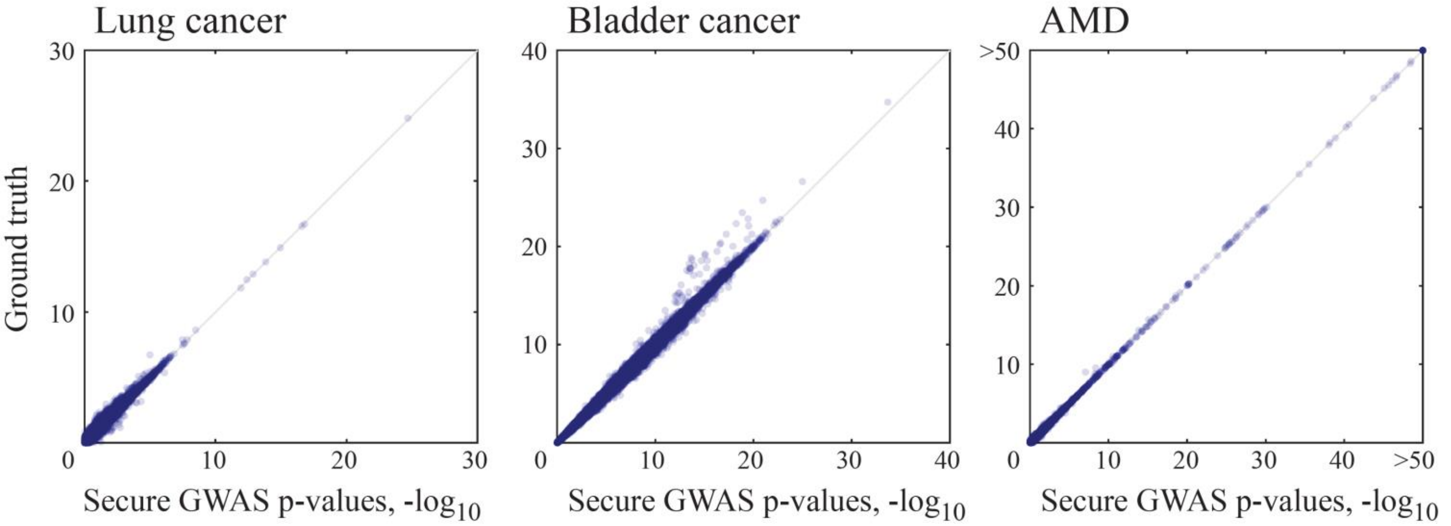
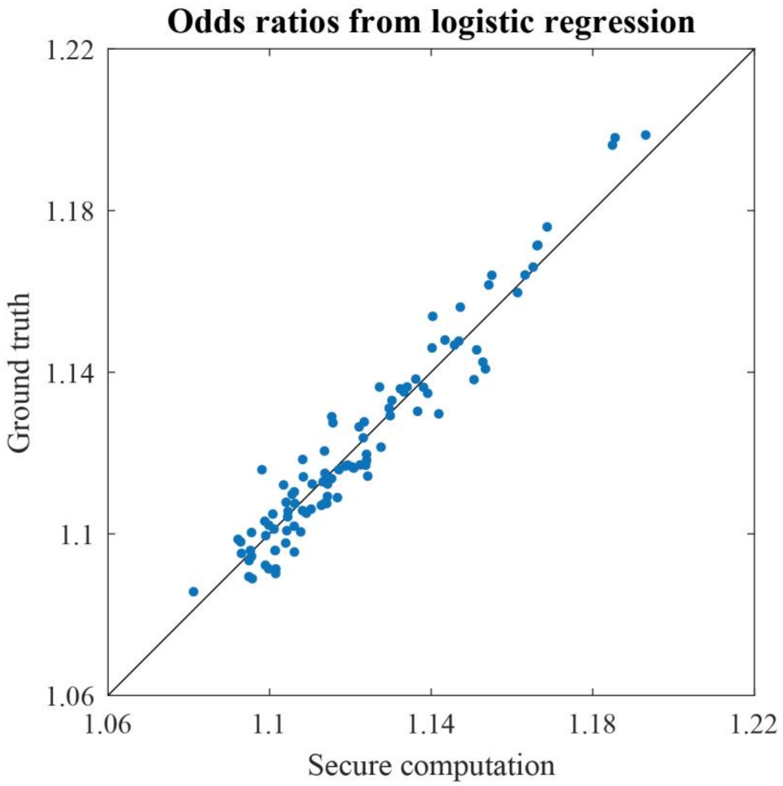


Supplementary Figure 1

Our secure GWAS protocol obtains accurate association statistics.

Using our protocol, we securely performed GWAS on three published case-control data sets for lung cancer (n = 9,098 after quality control), bladder cancer (n = 10,678), and age-related macular degeneration (AMD; n = 20,679). All of the tested SNPs passing quality control are shown in the figure: 378,492 loci for lung cancer, 389,868 loci for bladder cancer, and 221,295 loci for AMD. Our securely computed Cochran-Armitage trend test p-values (one-sided) accurately matched the ground truth we obtained based on plaintext data.



Supplementary Figure 2

Our secure GWAS protocol accurately estimates the effect size of associated SNPs via logistic regression.

We implemented logistic regression in our secure computation framework and applied it to a subset of 100 SNPs (randomly chosen among the top 1000 associations) in the lung cancer data set (*n* = 9,098 after quality control). The odds ratio of a SNP is given by the exponential function evaluated at the estimated weight associated with the SNP's minor allele dosage in a logistic regression model. Analogous to our main GWAS protocol, we included 10 additional phenotypes (e.g., age group) and five principal components securely obtained by our GWAS protocol as covariates in the model. As shown in the scatter plot, the odds ratios securely obtained by our protocol accurately matched those computed based on a plaintext implementation of logistic regression, the latter of which also used a plaintext PCA algorithm to obtain the top principal components. Performing logistic regression on 100 SNPs completed in about a day using our experimental setup. Although performing logistic regression genome-wide is still prohibitively expensive, our method enables a heuristic two-step approach where the odds ratios are computed for only the SNPs passing a certain significance threshold in our main GWAS protocol. Note that our logistic regression pipeline provides the same security guarantees as our main GWAS protocol; namely, no information about the underlying genotypes and phenotypes is revealed during the process other than the final output.