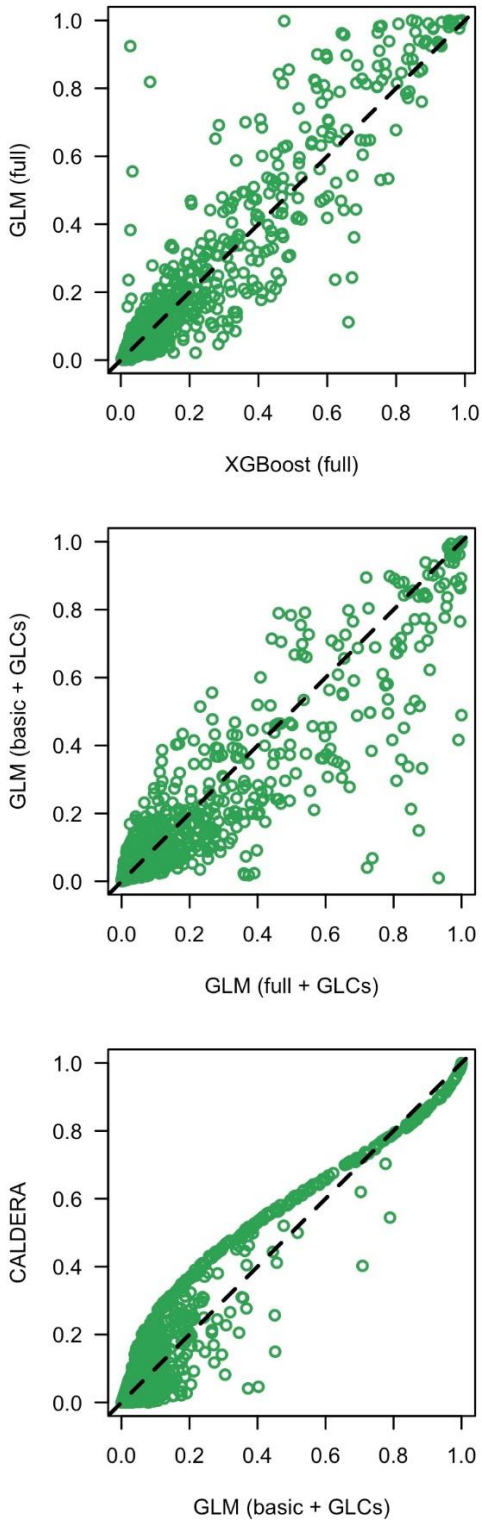

Fig S1. Comparisons of causal probabilities across models. **A.** XGBoost with the full feature set versus logistic regression with the full feature set. **B.** Logistic regression with the full feature set and gene-level covariates (GLCs) versus logistic regression with the basic feature set and GLCs. **C.** Logistic regression with the basic feature set and GLCs versus CALDERA. Each point represents a single trait-gene pair. The solid black line represents an equivalent value for the x- and y-axis variables.

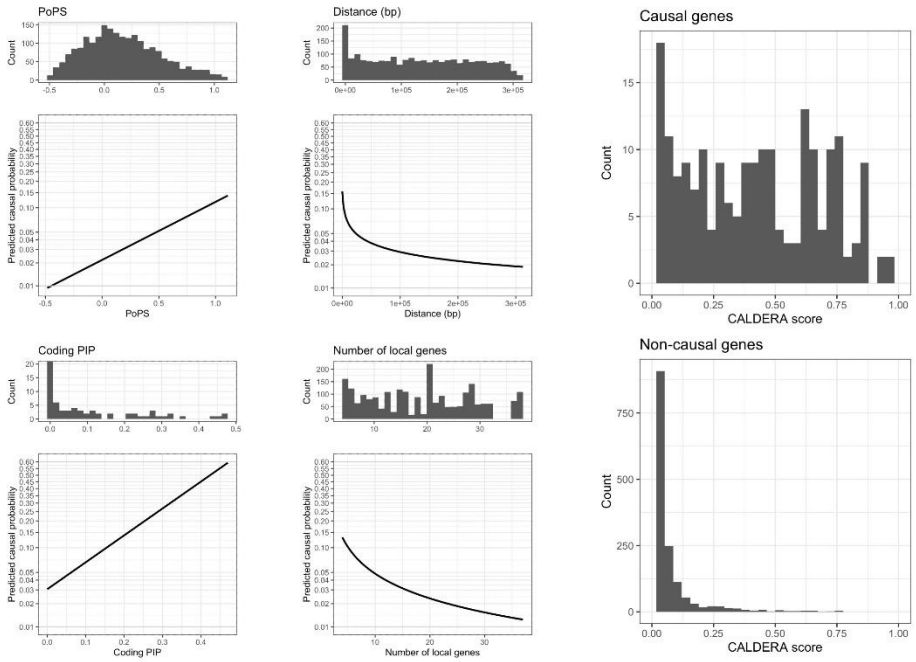

Fig S2. Relationships between predicted causal gene probabilities and PoPS (top left panel), distance between gene and GWAS lead variant (in base pairs, top middle panel), non-synonymous credible set variant posterior inclusion probability (PIP, bottom left panel), and number of local genes (bottom middle panel). The lower y-axis represents the logit-transformed probability that a given gene is causal for a given trait. The x-axis represents global feature values ranging from the 5th to the 95th percentile (except for coding PIP, which ranges from the 0th to the 100th percentile). Histograms showing the global feature distribution are plotted at the top of each panel. For coding PIP, the histogram y-axis was truncated at 20 for clarity (count of first bin = 4,787). All other features were set to their means, leading to low overall probabilities. Although transformed distances were used to train the model, untransformed values are presented to facilitate interpretation. The right panel shows the distribution of CALDERA scores for causal (top) and non-causal (bottom) genes in the training dataset.

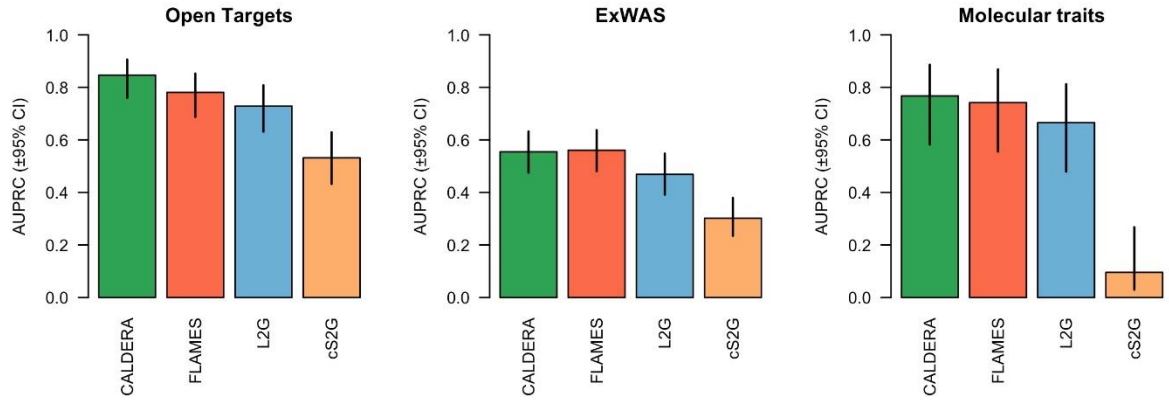

Fig S3. Area under the precision-recall curve (±95% confidence intervals) for CALDERA, FLAMES, L2G, and cS2G model predictions in the Open Targets ground truth dataset (left panel), a ground truth dataset derived from burden tests of rare coding variants in the UK Biobank (middle panel), or a ground truth dataset derived from three well-characterized serum metabolite levels (right panel).

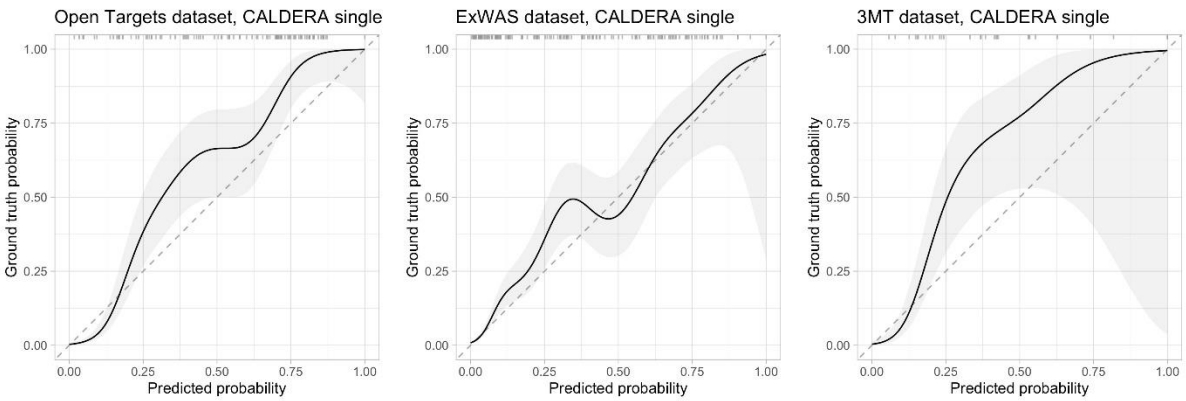

Fig S4. CALDERA calibration plots in the Open Targets (left), ExWAS (middle), and Molecular Traits (right) gold standard datasets. The x-axis represents the model-predicted probability in held-out trait data and the y-axis represents the ground truth causal probability. The solid lines represent the fitted values from generalized additive models with shaded areas representing 95% confidence intervals. The dashed lines represent perfect calibration.

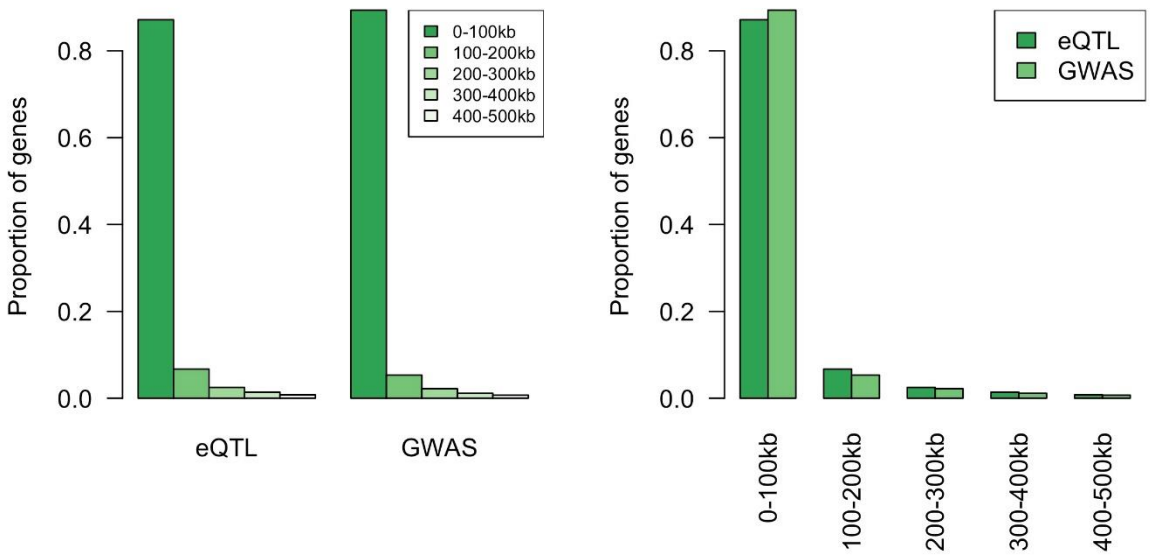

Fig S5. The proportion of genes that lie in various distance bins for eQTLs and their actual effector genes, and for GWAS hits and their nearest genes. The data were reprocessed from Mostafavi *et al*., 2023[18].

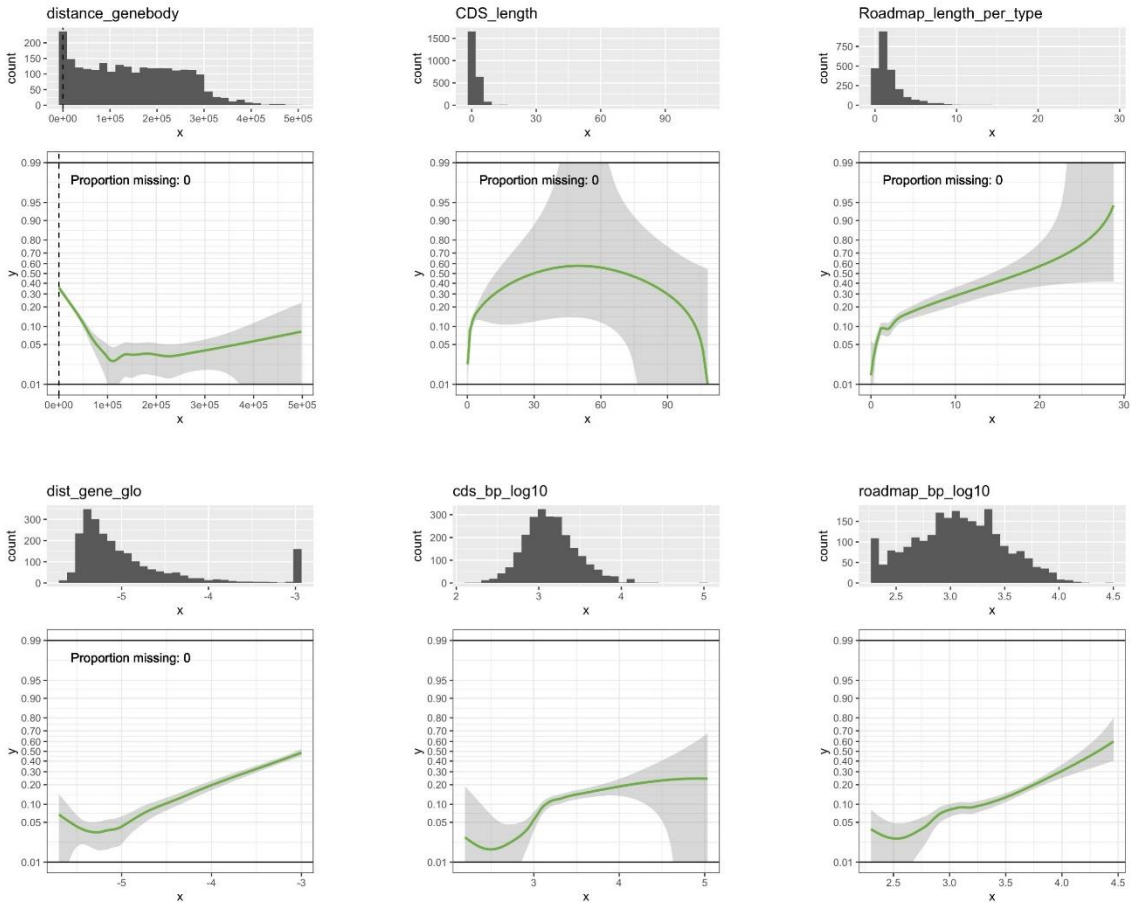

Fig S6. Relationships between predicted causal gene probabilities and distance between gene and GWAS lead variant (left panel), gene coding sequence length (middle panel), and total Roadmap enhancer length (right panel) both before (top panels) and after transformation (bottom panels). The lower y-axis represents the logit-transformed probability that a given gene is causal for a given trait. The x-axis represents feature values across their entire range. Histograms showing the feature distribution are plotted at the top of each panel. Green lines represent the fitted values of a local polynomial regression with grey shading representing 95% confidence intervals. In all cases, transformation improved the linearity of the relationship between the feature and the probability of being a causal gene in the training dataset.