

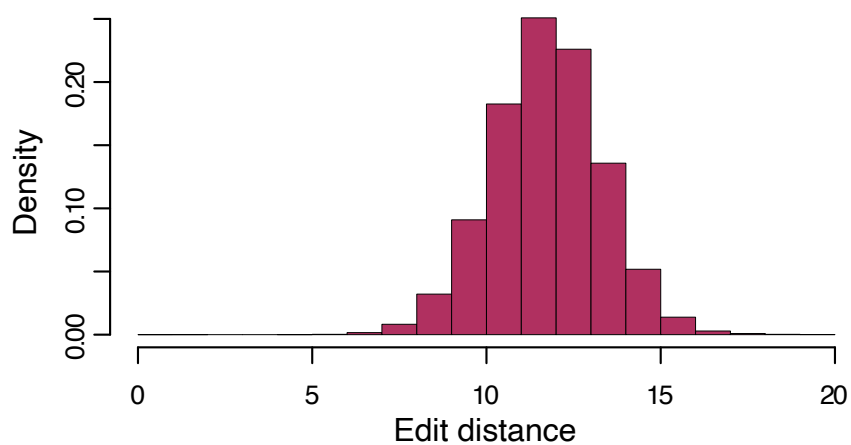
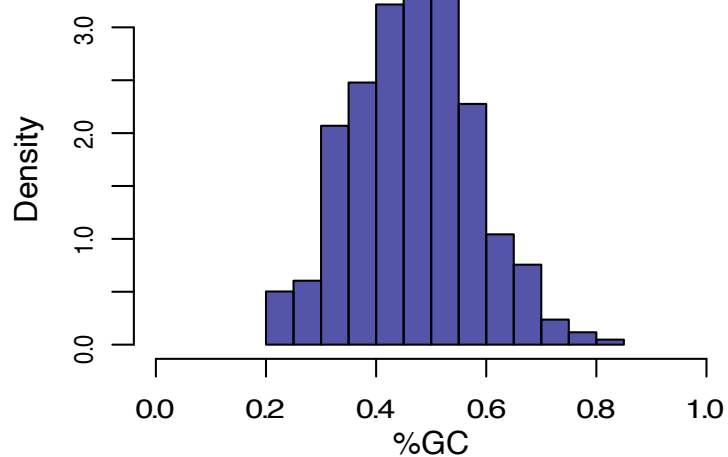
A**B**

Figure S1. Barcode sequence composition and diversity. **A.** Distribution of Levenshtein edit distances among all barcode sequences, indicating high sequence diversity and minimal overlap between barcodes. **B.** Distribution of GC content (%) across all barcode sequences, showing balanced nucleotide composition without GC bias.

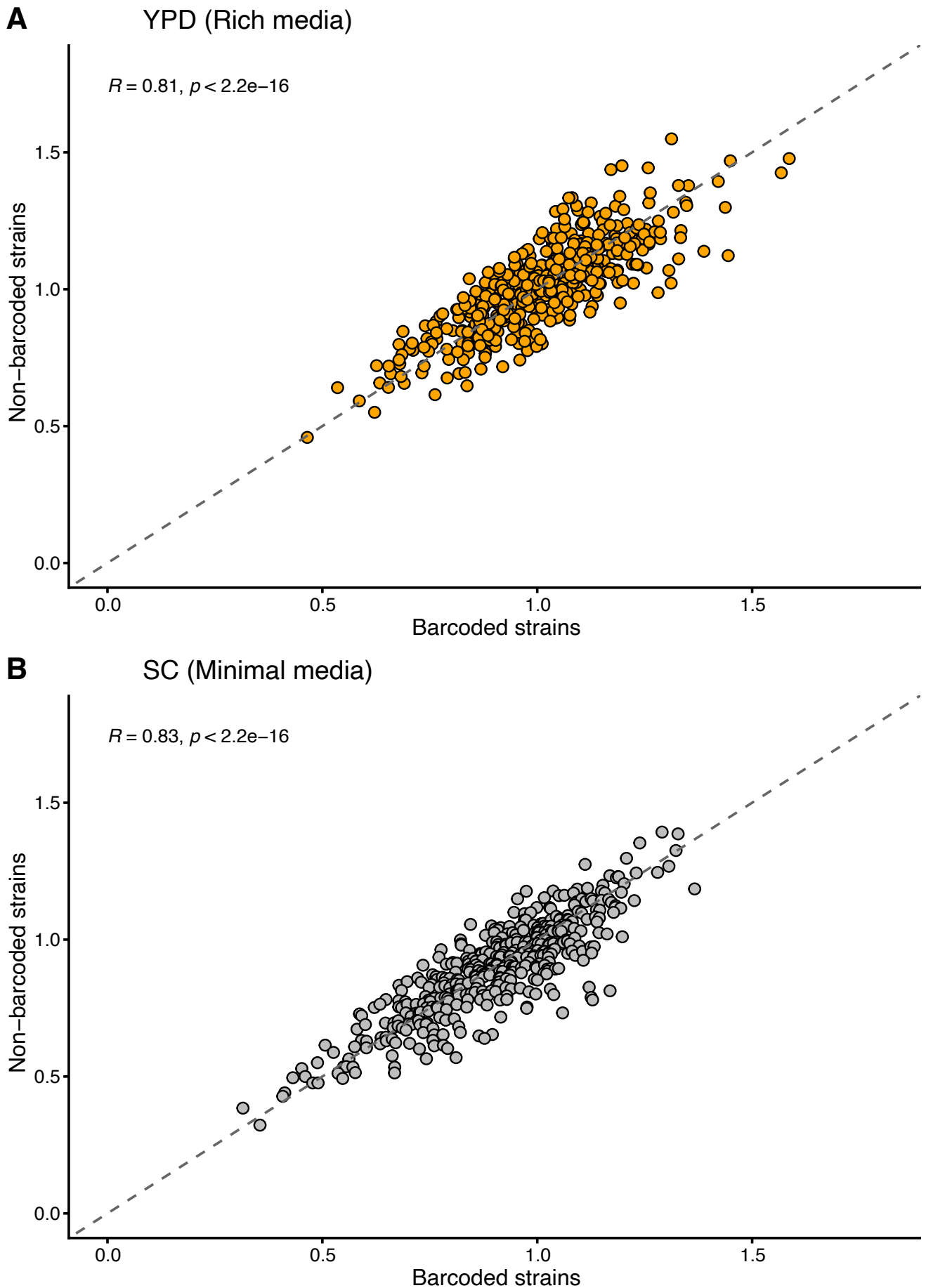


Figure S2. Barcoding does not alter strain fitness under standard growth conditions. **A.** Scatterplot of fitness (colony size relative to the BY reference) between barcoded and non-barcoded isolates grown in YPD (2% glucose). **B.** Corresponding comparison in minimal medium (SC). High correlations (Spearman's $R = 0.81$ – $0.83, p < 2.2e-16$) indicate that barcode integration does not measurably affect strain fitness in either condition.

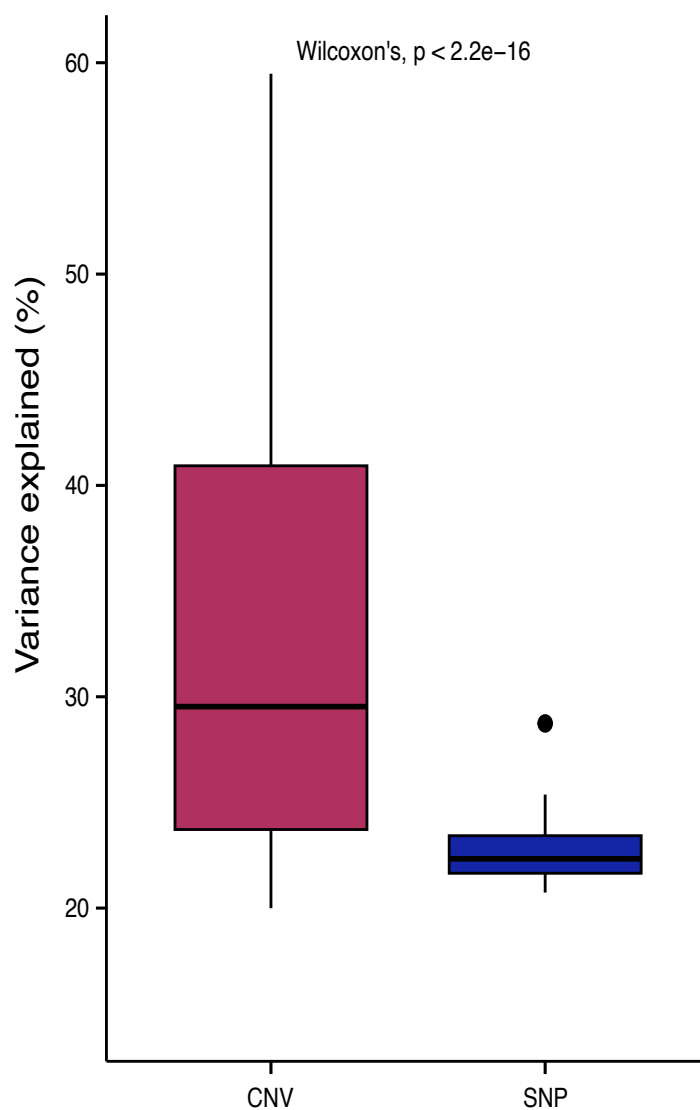


Figure S3. Copy number variants explain a greater proportion of fitness variance than SNPs. Boxplot showing the distribution of variance explained by significant CNV- and SNP-based associations across 13 compounds in our benchmarking experiment. CNV [maxima-59%, minima-20%, Q1-24%, Q3-41%, median-29.5%]; SNP [maxima-25%, minima-21%, Q1-22%, Q3-23.5%, median-23%, outlier-29%]

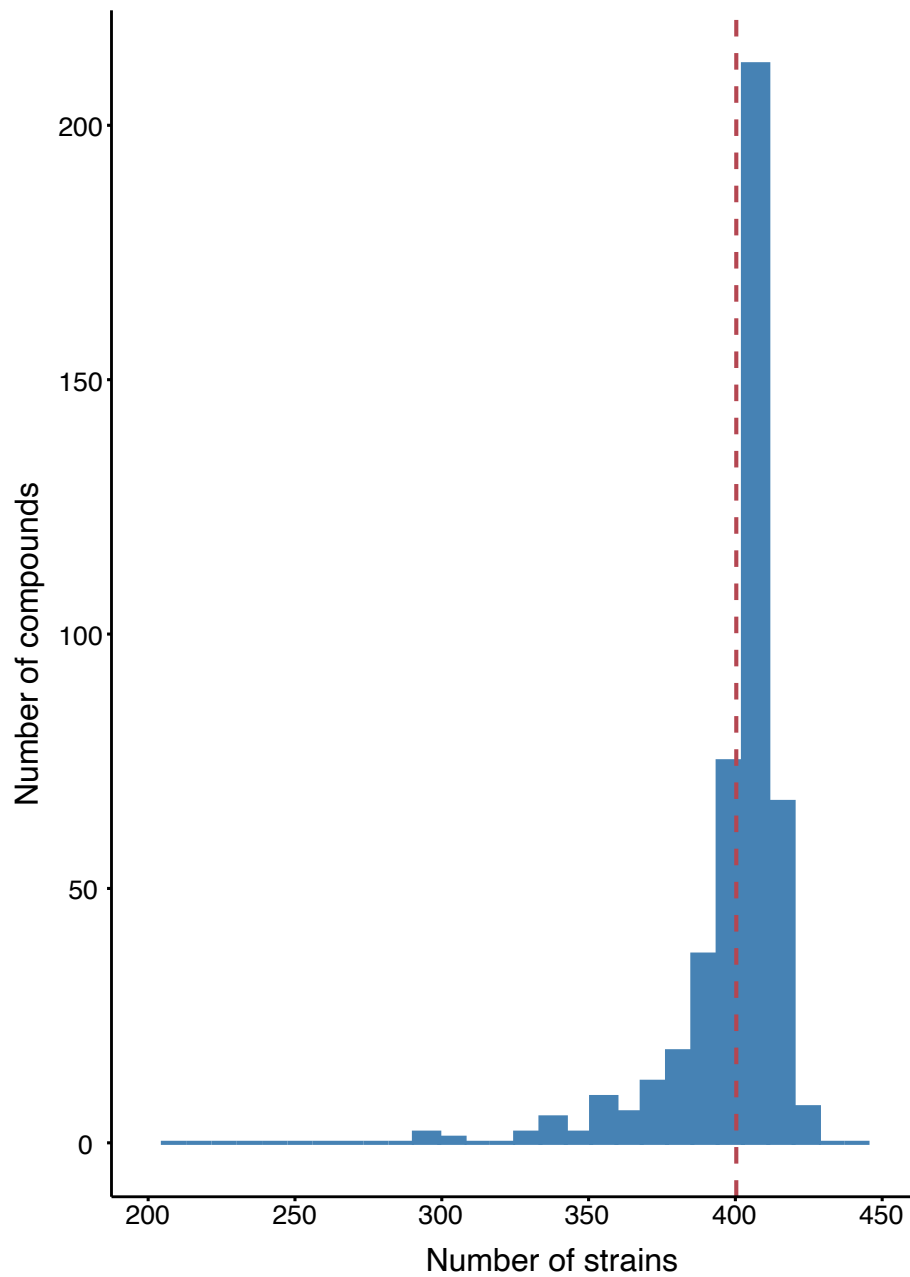


Figure S4. Distribution of strain detection across the compound library. Histogram showing the number of strains detected at the 24-h timepoint across 455 compounds that passed our QC threshold. On average, ~400 strains were detected per condition.

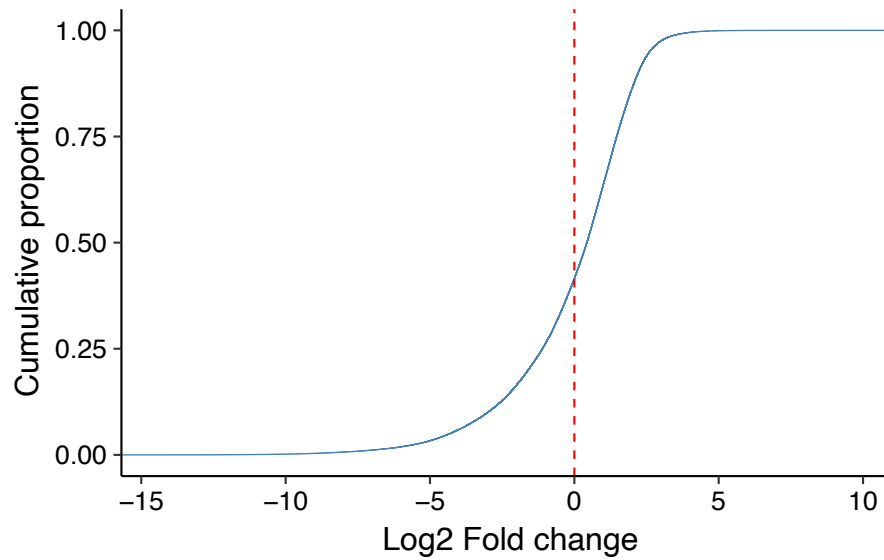


Figure S5. Global distribution of strain fitness across all compound treatments. Empirical cumulative distribution function (ECDF) of strain fitness values (log₂ fold change) across 455 compounds. The fitness distribution spans both positive and negative values but is asymmetric relative to neutrality (0), indicating unequal representation of fitness gains and losses and suggesting that subsets of strains respond non-randomly to chemical perturbations.

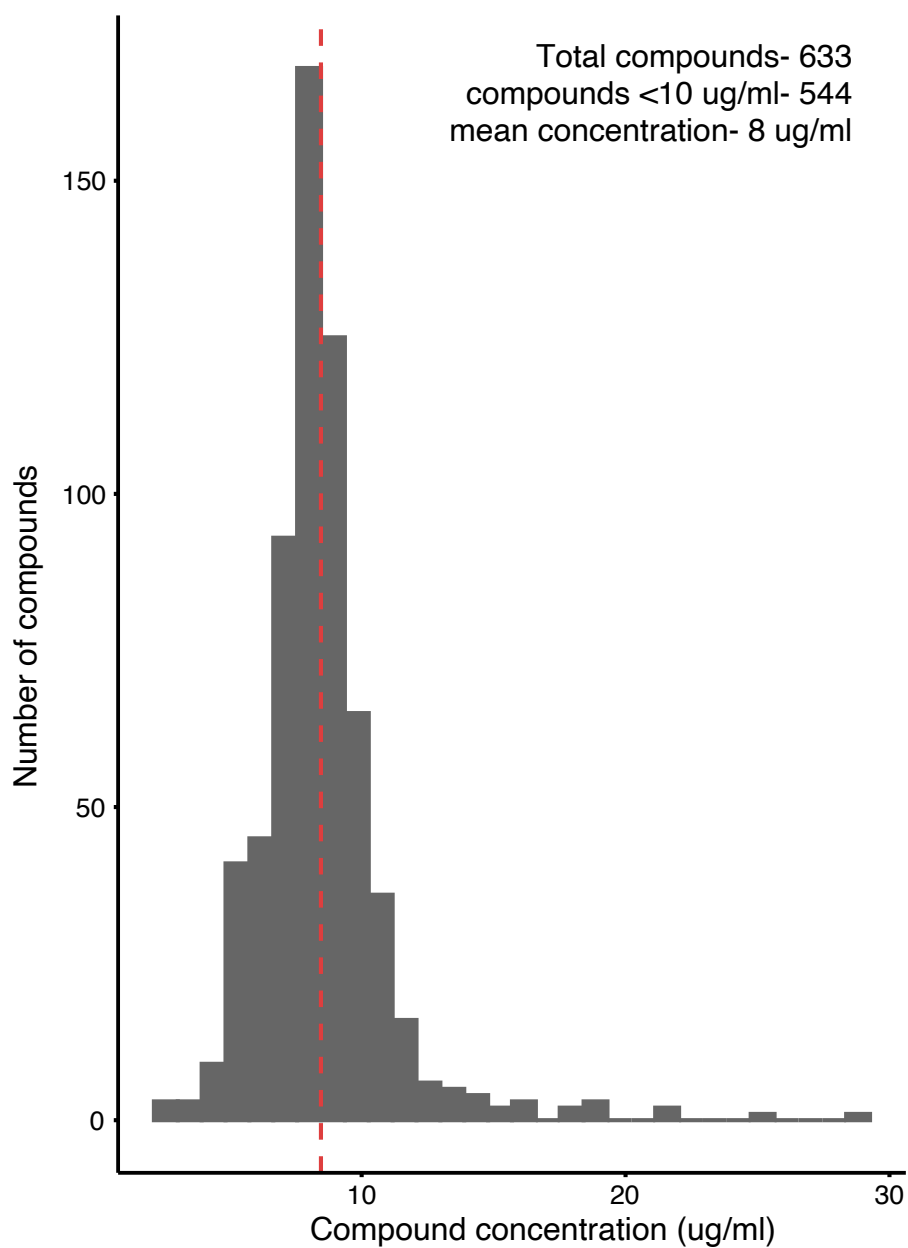


Figure S6. Compound concentration distribution in the barcoded collection screen. Histogram showing the concentration distribution of the 633 compounds screened in this study (mean = 8 $\mu\text{g/mL}$). Of these, 544 compounds ($\approx 80\%$) were assayed at concentrations below 10 $\mu\text{g/mL}$. Piotrowski et al. 2017, screened at a uniform concentration of 10 $\mu\text{g/mL}$.

Gene associations across compounds and clusters

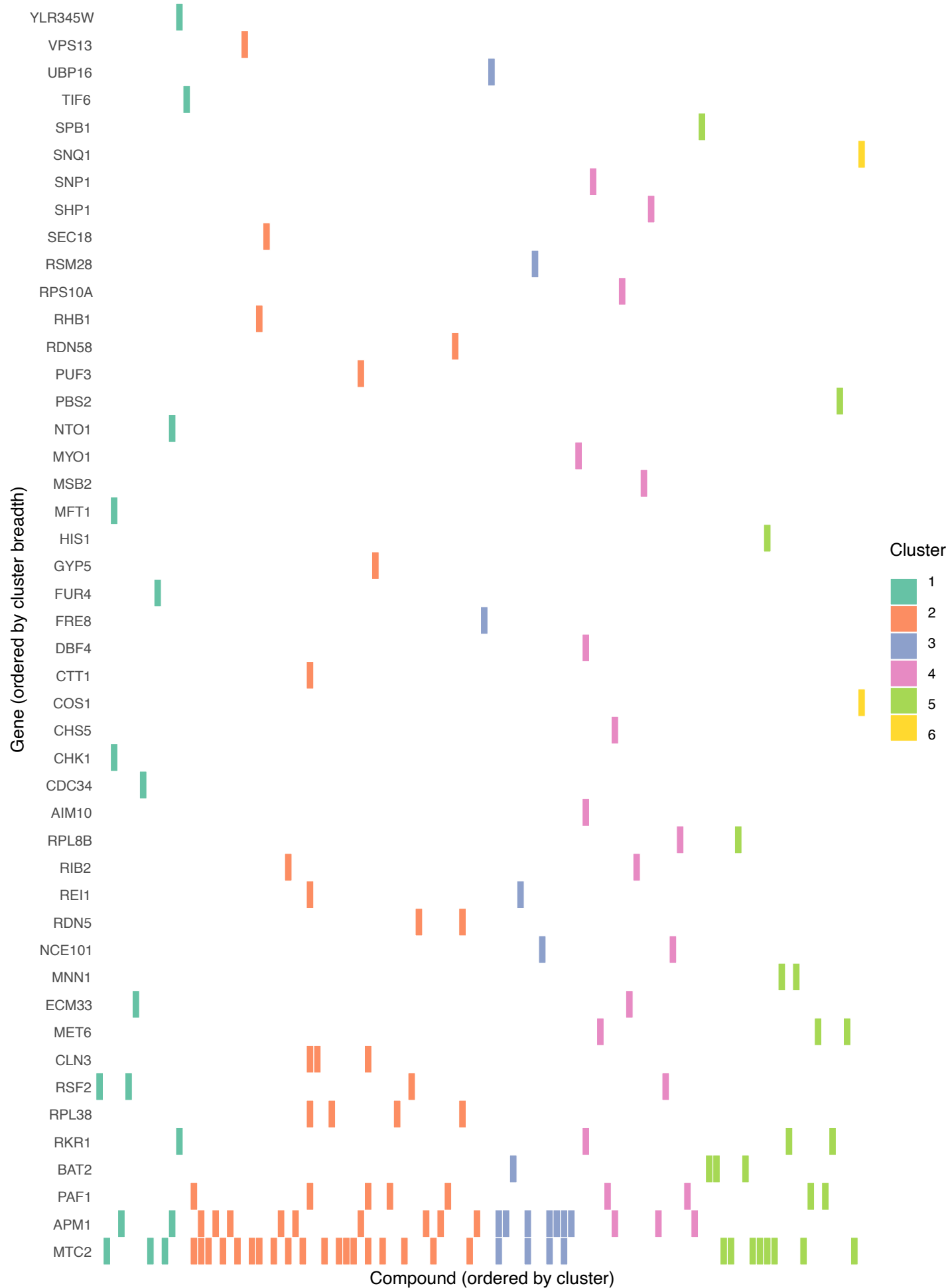


Figure S7. Genetic associations between SNPs and compounds. SNP-trait associations ($p < 1e-5$) across 106 compound conditions. In total, 52 associations mapped to 46 genes, spanning the six compound clusters.

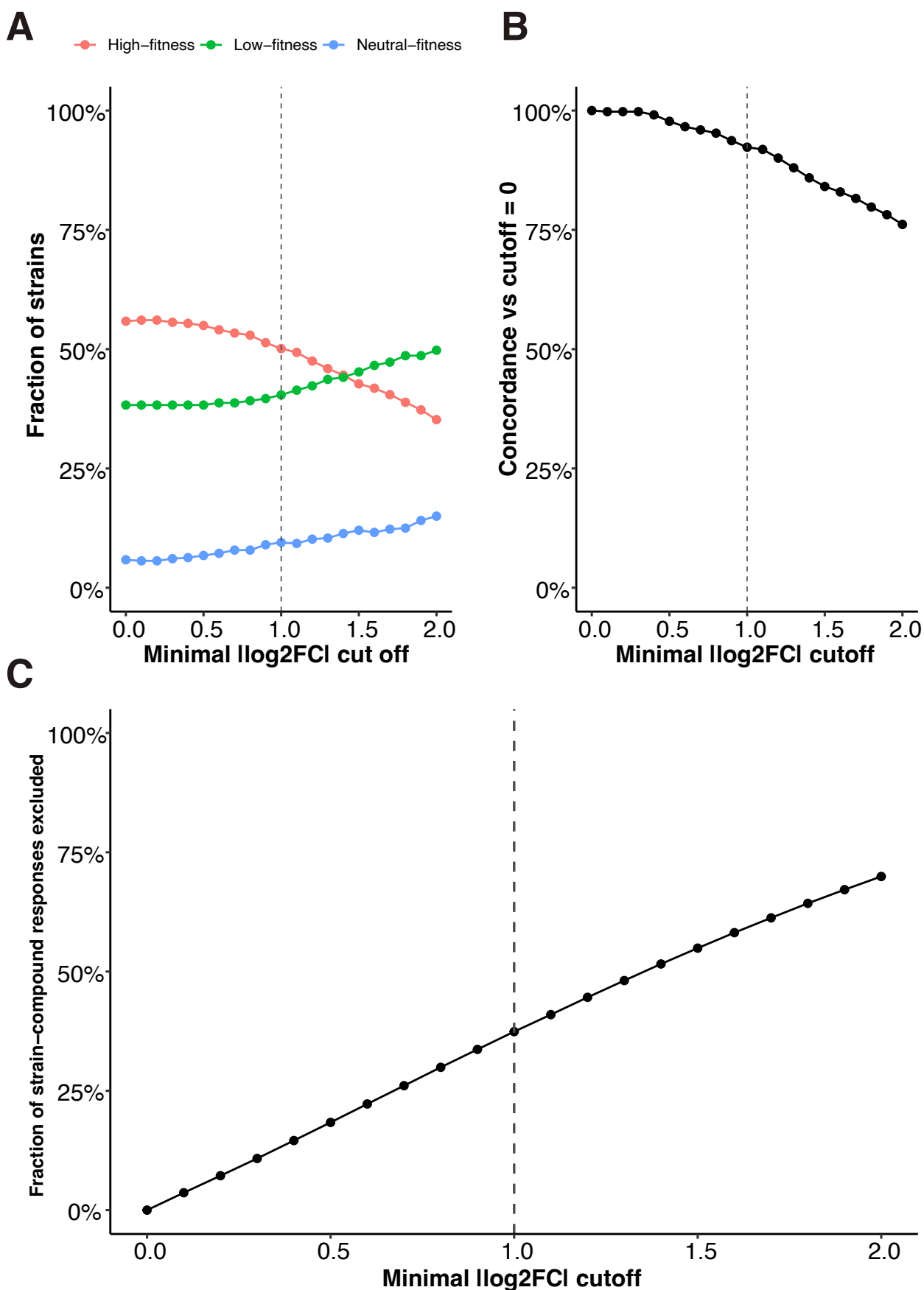


Figure S8. Robustness of strain fitness classification. Increasing minimal $|\log_2$ fold-change ($|\log_2\text{FC}|$) thresholds were applied prior to strain-level binomial testing by excluding strain-compound responses within a neutral band ($|\log_2\text{FC}| < \text{cutoff}$) and re-running the full classification at each cutoff. $|\log_2\text{FC}| = 1$ (vertical dashed line) corresponds to a two-fold change in strain abundance and $|\log_2\text{FC}| = 2$ to a four-fold change. (A) Fraction of strains classified as High-, Low-, or Neutral-fitness across cutoffs. (B) Concordance of strain categories relative to the no-cutoff (sign-based) classification across cutoffs; dashed line indicates $|\log_2\text{FC}| = 1$ (two-fold change). (C) Fraction of strain-compound responses excluded at each cutoff.