In an exponential growth model, the frequency of a double mutant strain in a given condition at a time represents its total growth from an initial number as a proportion of the total growth of all other strains in the pool:

Note: We add a pseudocount of to the count of every strain in our analysis to avoid a zero denominator in several calculations.

Here, is inversely related to the doubling time of strain and effectively represents the number of doublings of strain . A frequency at t=0 therefore represents:

To remove the unknown term, we define

We note that the term is the ratio between the initial and final number of cells in the pool, and can be calculated by the total number of generations of pool growth :

Therefore, can be calculated as:

To calculate , we take the mean of all neutral-neutral pairs:

We then obtain the relative growth rate of each strain compared to the wild type by dividing their number of doublings. In a constant exponential growth model, this metric is independent of time. In practice, represents the average growth rate over the measured time period.

To estimate the single mutant fitness and for a given pair, we use the mean estimate of or combined with neutral genes.

We then definite the genetic interaction score (GIS) as the difference between and the product of with :

Because there is uncertainty in , it is possible to estimate for , , or . Such values are assigned as when performing the calculation. Raw scores are reported in Table S1 for comparison to St. Onge et al.

To assign a threshold for positive and negative genetic interactions, several additional steps are performed. is converted to a standard score by calculating how many standard deviations is from given an estimate of uncertainty (

To calculate , we identify various sources of uncertainty. Another way to state is as such:

We then define an error model to calculate the standard error for each term used in this calculation:

: is directly dependent on and (the latter is taken as a constant and no error estimate is made). is calculated from and , which are in turn calculated from the count data and . The error in the count data is estimated using a Poisson distribution, where . Using the standard formula for propagation of uncertainty, and are converted to . We note that this model only captures the error in the count uncertainty arising from sequencing.

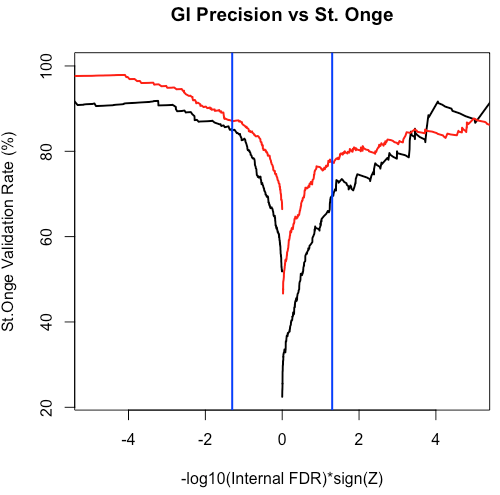
The mean of multiple strains are used to calculate each of these values, so the formula for the standard error for the mean for the respective calculations is used to determine their uncertainty.

The standard formula for uncertainty propagation is used to combine , and into .

To choose negative and positive cutoffs for for each condition (, we then analyze the distribution of in all unlinked neutral-neutral and neutral-DNA damage pairs (abbreviated to ), as few or no genetic interactions are expected to take place between these pairs. We use this distribution to estimate the number of false positives in the data.

To calculate and , we model a normal distribution for between neutral pairs. The normality of of neutral pairs was evaluated using a histogram and qqplot. We then use the estimated number of positives to estimate an ‘internal’ False Discovery rate for a given cutoff:

Cases where is greater than the observed interactions are assigned an FDR of 1. and are then chosen such that and . To validate these cutoffs against an external dataset, we calculated the validation rate in the St.Onge et al data as a function of internal estimated FDR. In this graph, negative values represent s for pairs with negative Z, positive values represent s for pairs with positive Z. Red lines correspond to the MMS condition, black lines correspond to the no drug condition.



At the chosen 5% internal cutoffs, we estimate a ~70-78% external validation rate for positive interactions, and ~85-87% external validation rate for negative interactions. We have included this score for each interaction, so that other cutoffs may be chosen.