**Point by point responses to reviewers**

*Reviewer comments are in italic, with suggestions underlined.*

Our responses to the reviewers are in blue.

Text from the manuscript is in black, with changes in red.

**Referee #1:**

*The authors have made extensive changes to their manuscript to respond to the concerns raised by the reviewers. This includes new experiments and a very detailed analysis of enzyme kinetics and discussion of assumptions made during the analysis, supported by a new collaboration with David Lambright, an expert in GTPase enzymology.*

*In addition, the AP-MS data have been re-analyzed and changes were made to the way in which data are presented.*

*I commend the authors for the detail with which they have responded to concerns raised by the reviewers and their thoughtful response.*

*Taken together, the additional data presented in the revised manuscript and changes to the text and figures have significantly improved it and strengthened the conclusions drawn. All my concerns have been adequately addressed.*

We thank Reviewer 1 for the thoughtful comments as well as suggestions that helped us to improve the manuscript.

***Referee #2:***

*The authors have addressed most of my concerns satisfactorily. In particular, I am pleased to see the enzyme kinetics section, where many of the assumptions and methods have been clarified. The kinetics simulations add value to estimating the accuracy of the parameters obtained from fits. The paper is essentially ready for publication except for the one major point below.*

We thank Reviewer 2 for helping us improve the enzyme kinetics section, as well as other thoughtful comments that helped us improve the manuscript.

*Major point:*

*I have only one major point that is still unsatisfactory and that is point (1) of my original review regarding designing mutants to test the predictive power of the model. While I agree with the authors' explanation that rational design is not possible at this stage, the authors may have misunderstood the intent behind the comment.*

*I suggest the following experiment: take ANY three-four new mutants (can be reasonable guesses not found in literature, can be random, the authors can even use ease of protein expression as a guide) not included in the current analysis that are not wt-like from kcat/KM measurements. Use the ln(GAPreleff/GEFreleff) to predict which clusters of alleles will be affected by the mutation. Then, acquire GI profiles to test the functional effect and see if it falls in line from what their model predicts. Basically add three-four more data points to figure 4a but on new mutants not involved in the clustering analysis. This would be like leaving out 20% of the reflections when solving a crystal structure and using the structural model to predict the reflections.*

*From my standpoint, there seems to be no difficulty in doing this experiment other than the effort involved in adding 15-20% more data to Figure 4a, and I really think this will shed light on the predictive power of the model. Post-facto rationalization of the effects found in literature such as G21V and T24N are not nearly as satisfactory.*

As we discussed with the editor, adding more data by repeating the entire study with a small set of different mutants (in a paper that has so much data from several different approaches) would not improve the manuscript in any meaningful way, since:

(i) we have already nicely identified the range of *in vitro* activities and *in vivo* genetic profiles that we would encounter, and adding more mutants would almost certainly not identify more categories;

(ii) our mutants already map to all interfaces on the protein that are known to be involved in the different interactions and functions; therefore, choosing three-four new mutants randomly or for ease of purification, as the reviewer suggests, would most likely result in mutants with no detectable effects on the kinetics parameters.

As a more technical point, we respectfully disagree with the reviewer that obtaining the additional data would entail “no difficulty”. We would likely need to purify and assay many new mutants before finding 3-4 new examples of the kinetic categories we already identified. We would then need to perform independent biological replicates of EMAP screens with our entire library of 1444 strains, together with many of the mutants we already screened to allow for robust comparison; these experiments together would amount to repeating a considerable fraction of the current scope of our study.

We however agree that a “leave data out” analysis, as the reviewer suggests in analogy to solving a crystal structure, is a good suggestion (also inspired by ideas from Reviewer 4 on testing various perturbations to the data analysis). We hence included the computational analysis described below in **Supplementary File 1**, section: **Robustness of the analysis to leaving out data, 2. Withholding mutants:**

We performed a computational analysis where we withheld each of our mutants from the analysis one at a time, perform the clustering of genetic interaction profiles for the remaining data as in **Fig.4a**, and then assign the withheld mutant to the group whose centroid is most correlated with the mutant. The Figure below shows that in 21/22 cases, the withheld mutant had the highest correlation with the centroid of its original group (dark bars). This analysis confirms the robustness of our analysis and addresses the question whether our model would be capable of placing a new mutant not included in the analysis into the correct category.

A screen shot of a computer screen

Description automatically generated with low confidence

**Leave-one-out analysis of Fig. 4a**: Each bar graph shows the Pearson correlation values between the indicated withheld mutant and the centroid of each of the three groups identified by hierarchical clustering of the remaining 21 mutants. We grouped the bar graphs according to the original group to which each mutant was assigned in **Fig. 4a**. For each of the withheld mutants, the dark bar represents the expected group (group I left bar, group II middle bar, group III right bar). With the exception of G80A, which is slightly more correlated with the group I centroid (Pearson correlation = 0.38) than the centroid of its original group, group II (Pearson correlation = 0.28), all other mutants have the highest correlation with their original groups.

*Minor points:*

*1) Line 333: should be "point size as in b"?*

We corrected the figure caption.

*2) On p. 11 of the rebuttal letter, the equation for w, I get ((S0-kcat\*[E]t\*time)/KM) in the exponent, KM dividing S0 also. Can the authors please check?*

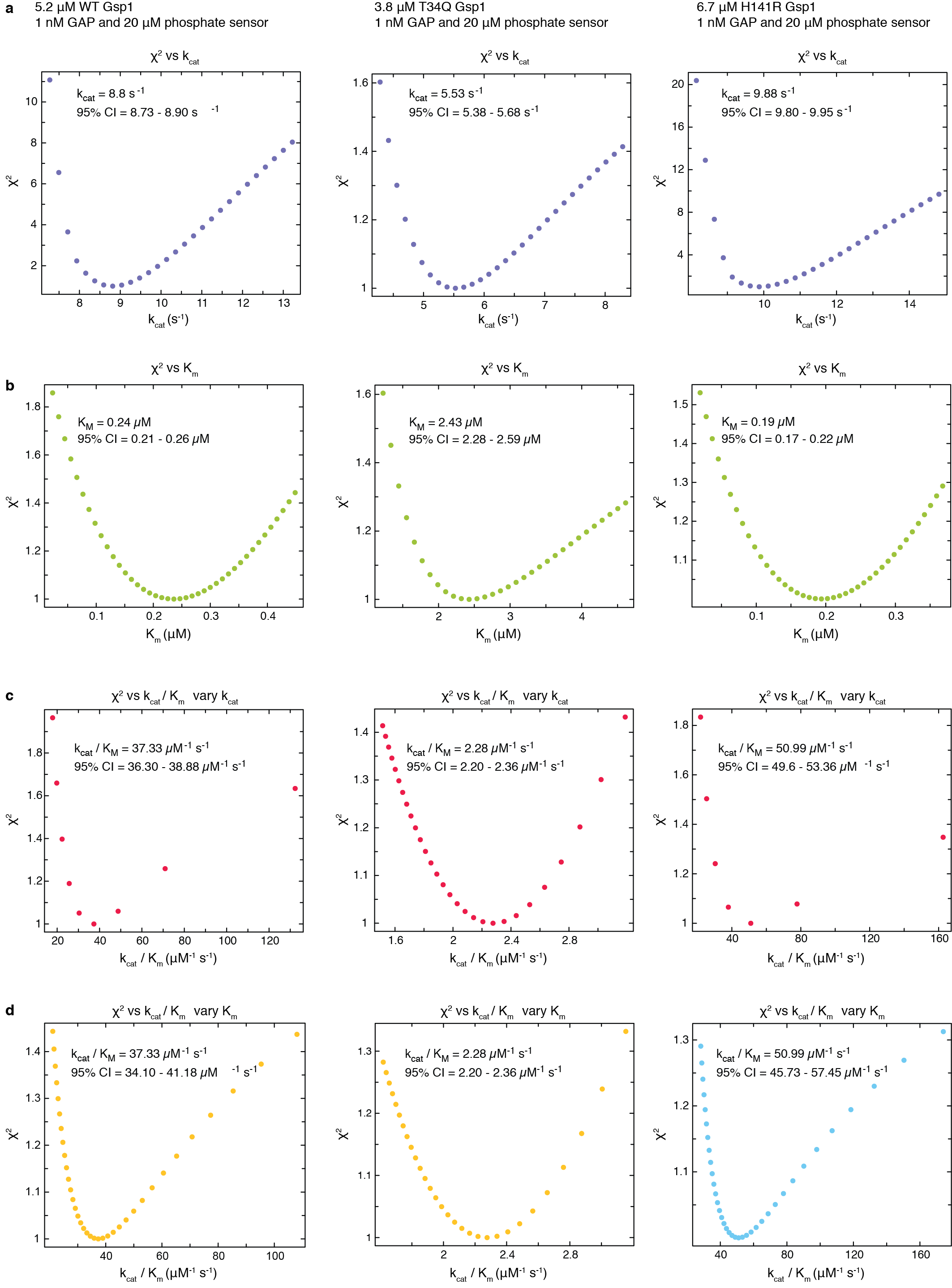
We would like to thank the reviewer for catching this typographical error. We have used the correct equation (with the Km dividing the S0 too) in our calculations and have now also corrected the equation in the manuscript.

*3) Same p. 11 of rebuttal, line below the equation for w should be "are shown in".*

We have corrected the typo.

*4) An alternate way of checking whether parameters can be reliably obtained from their Michaelis-Menten fits apart from simulations is to run chi2 surfaces, independently as a function of kcat, KM and kcat/KM (keep the parameter fixed at a value and evaluate the chi2, then change the value systematically, keeping it fixed all the time). If the change in chi2 is steep, this means that parameters cannot correlate with each other and compensate for each other to give a good fit (low chi2). I leave it up to the authors to ascertain whether this additional test is necessary.*

We have done this analysis and have now included it as a **Supplementary File 1 Figure 16**.



**Supplementary Figure 16 Estimated error around the maximum likelihood estimated values of the Michaelis-Menten parameters.** Plotted is the change in χ2 statistics as each of the parameters was fixed in gradual increments around the maximum likelihood value. The χ2 values are relative to the maximum likelihood values.

Error estimate analysis is shown for three of the Gsp1 variants: wild type Gsp1, the low efficiency Gsp1 T34Q mutant, and the high efficiency Gsp1 H141R mutant. 95% CI is the estimated 95% confidence interval for each value, based on the χ2 surface. **a**, Change of χ2 statistics as the kcat value is varied around the maximum likelihood value. **b**, Change of χ2 statistics as the Km value is varied around the maximum likelihood value. **c**, Change of χ2 statistics as the kcat/Km value is varied around the maximum likelihood value and the Km is kept fixed at the maximum likelihood value (kcat is varied). **d**, Change of χ2 statistics as the kcat/Km value is varied around the maximum likelihood value and the kcat is kept fixed at the maximum likelihood value (Km is varied).

**Referee #3:**

*The authors have made substantial changes to the revised manuscript, including the results of new experiments, as well as improved interpretation of previous results. These changes have improved the focus of the manuscript and have addressed my main concerns regarding data quality. I recommend publication of this revised manuscript.*

We thank Reviewer 3 for thoughtful suggestions that helped us improve the manuscript.

**Referee #4:**

*My comments below mainly concern the statistical aspects of the study.*

*1. For the two hierarchical clustering analyses performed in the paper, one used average linkage and the other used ward’s linkage. Can the authors justify why they chose to use two different types of linkages in their studies? Also, how stable are the clustering results against data noise? For example, if small perturbations were introduced to the data, e.g., randomly sample a subset of mutants/alleles (vector elements), would similar results be obtained using the same analysis procedure? How about data outliers?*

We included the following analysis in **Supplementary File 1:**

**Linkages used for hierarchical clustering analyses**:

For clustering of Gsp1 mutants and E-MAP library genes (in **Figs. 1d** and **EDF 3ab**) we used average linkage to be consistent with how we and others have clustered and represented genetic interaction (GI) data in previously published E-MAP datasets (as detailed in: Braberg *et al*, Nature Protocols, 2016). Even though our data is based on screens of point mutants of a single protein, different from most previous studies that screen knockouts of many different genes, we show that average linkage remained an appropriate criterion for clustering our E-MAP matrix based on the recovery of known groups of functionally related genes within the dendrogram of library genes (**Fig. 1d** and **EDF 3b**).

The clustering analysis in **Fig. 4a** had the goal to assign the Gsp1 mutants by functional similarity to classes in an unbiased manner, and assess whether the classes of mutants matched the grouping defined by the *in vitro* kinetics and NMR data. To quantify functional similarity, we adopted the widespread approach of computing correlation coefficients between GI profiles. Most studies have represented these data as networks and used existing annotations (typically Gene Ontology) annotations to assert functional groupings (Costanzo *et al*. Science, 2016), but we sought to use an unsupervised clustering approach instead. To do so, we used Ward’s linkage criterion, since it was designed to build hierarchies by selecting joining operations that minimize within-group dispersion (Murtagh and Legendre, *Journal of classification*, 2014) to find compact, spherical clusters. Indeed, we found Ward’s linkage to result in rounder clusters reflecting known biological functions and that were less sensitive to sparsely populated outliers. In contrast, we found the average linkage criterion to be more sensitive to a few sparsely populated outliers (resulting in a variety of group sizes).

Nonetheless, to show that the linkage method used does not alter our primary conclusions regarding the grouping of mutants, we now compare the clustering from average linkage Ward’s method in dendrograms shown in the Figure below. Although both methods identify the three main classes of mutants (I, II, and III below, I, III, and IV above), average linkage is more sensitive to the sparsely populated vectors and outliers, resulting in a wider variety of cluster sizes, and so Ward’s method was used for the main figure.

Graphical user interface, application, Teams

Description automatically generated

**Clustering of *S. cerevisiae* alleles and strong Gsp1 point mutants by the p-value of Pearson correlations using the average linkage criterion.** Hierarchical clustering of 22 strong Gsp1 point mutants by the p-value of Pearson correlations of their GI profiles and those of 276 *S. cerevisiae* alleles, using both the average linkage and Ward’s method as linkage criterion. Although both methods identify the three main classes of mutants (I, II, and III below, I, III, and IV above), average linkage is more sensitive to the sparsely populated vectors, resulting in a wider variety of cluster sizes.

We have also added a short note in the Methods section entitled *Gi profile correlation measurements* which notes on this observation and the approach we took:

We used Ward’s method rather than the average linkage criterion as we found the latter resulted in groups more sensitive to a few sparsely populated outliers. Using Ward’s methods resulted in rounder clusters, allowing us to identify biologically interpretable functional groups of mutants and alleles.

For an **analysis of data noise by sub-sampling**, please see our response to comment #3.

*2. For the statistical significance analysis, the authors derived approximate p-values for the computed Pearson correlations. However, for the vectors involved in the correlation estimation, are there dependencies between the vector elements and have such dependencies been considered? E.g., for the matrix of 22 mutants vs. 278 alleles, when computing the correlation between two mutants based on the 278 alleles, are there dependencies between the alleles? Note that serious dependencies between alleles would affect the estimation of correlations as well as the assessment of statistical significance (e.g., the current way to derive approximate p-values may not be appropriate anymore). The same arguments can be applied to the calculation of relationships between alleles if there are serious dependencies between mutants.* Have the authors considered these issues*?*

We thank the Reviewer for bringing up a key point regarding the underlying assumption of independence between alleles when computing correlations for our genetic interaction (GI) profiling analysis. We have now included the analyses described below in **Supplementary File 1**, section **Potential dependencies between alleles when computing correlations for our genetic interaction (GI) profiling analysis:**

We use Pearson correlations in three cases: (1) as a distance metric for clustering the Gsp1 E-MAP matrix (**Fig 1d** and **EDF 3ab**), (2) for quantifying the functional similarity of GI profiles of Gsp1 point mutants and *S. cerevisiae* alleles (**Fig 1g**, **Fig 4abd**, and **EDF 9abc**), and (3) as a distance metric for clustering the vectors of Gsp1 mutant correlations (**Fig 4a**). In all three cases, there are certainly dependencies between some *S. cerevisiae* alleles, as evidenced by their own clustering into groups according to their biological function (**Fig 1d**): mRNA export genes cluster together, meaning that if one mRNA export gene has a large negative S-score with a Gsp1 mutant, other mRNA export genes are likely to as well (relevant for case 1). Likewise, if the GI profile of a gene is significantly correlated with a Gsp1 mutant, other genes in the same pathway are likely to have correlated profiles as well (relevant for cases 2 and 3). This dependency is expected and is indeed a main benefit of a GI profiling approach, as the S-scores allow us to infer functional relationships between genes and ascribe likely functions to unknown genes.

With regards to Case 1, we note that assessing similarity of GI profiles using Pearson correlations without further correction for dependencies between alleles is a standard analysis (Collins 2007, Costanzo 2016).

Case 2 is the only analysis for which we compute statistical significance when using correlations. We accounted for the dependencies between alleles by adjusting our p-values to control the False Discovery Rate, which has been shown to be valid when this form of dependency (positive regression dependency) exists between test statistics (Benjamini and Yekutieli, *Annals of statistics*, 2001). In the manuscript figures we use these corrected p-values instead of correlation values for simplicity, because, as can be seen from the plot, only positive correlations of above 0.1 have significant p-values.

A picture containing chart

Description automatically generated

With regards to Case 3, most GI studies have used the correlations between GI profiles to define edge attributes for graphical representations of GI networks. We elected to keep the data in matrix form and cluster it to identify functionally similar groups of mutants and *S. cerevisiae* alleles in an unbiased fashion. To cluster the Gsp1 vectors of p-values (columns), we used Pearson correlations as a distance metric. To cluster the *S. cerevisiae* alleles (rows), we used the Euclidean distance instead of the Pearson correlation (as stated in the *GI profile correlation measurements* section of the Methods) because the vectors were only 22 entries long and many were sparse, making them especially sensitive to outliers when using Pearson correlation as the distance metric. To test whether the use of Pearson correlations for the clustering of mutant vectors significantly changes our clustering, we re-clustered the matrix in **Fig. 4a** using the Spearman correlation or the Euclidean distance as distance metrics instead. While there are slight differences in the ordering of mutants using these different distance metrics, the grouping of mutants is very similar to the original heatmap in **Fig 4a** in that it identifies a GAP-perturbed group of mutants, a GEF-perturbed group of mutants, and an intermediate group (**Figure** below). Thus, we believe this analysis correctly identifies three functional classes of Gsp1 mutants regardless of any effect that dependencies between the *S. cerevisiae* alleles has on the Pearson correlations.

A picture containing text

Description automatically generated

**Clustering of *S. cerevisiae* alleles and strong Gsp1 point mutants by the p-value of Pearson correlations using the alternative distance metrics.** Hierarchical clustering of 276 *S. cerevisiae* alleles and 22 strong Gsp1 point mutants by the p-value of Pearson correlations of their GI profiles compared to the relative efficiencies of GAP-mediated GTP hydrolysis and GEF-mediated nucleotide exchange as indicated (asterisks indicate not measured). The p-value is a false discovery rate adjusted one-sided (positive) p-value of the Pearson correlations (represented as gray scale). The underlying data is identical to that presented in **Fig. 4a**, but the column clustering was performed using the Spearman correlation or the Euclidean Distance rather than the Pearson correlation as a distance metric.

We have also added a note regarding the validity of FDR-controlled correction in the Methods, in the *GI profile correlation measurements* sections:

The FDR method of p-value correction has been shown to account for the positive dependency between test statistics, such as those arising from the underlying functional similarities between *S. cerevisiae* alleles (Benjamini and Yekutieli, 2001).

*3. As mentioned in the first comment, it would be helpful to include a systematic stability check of results against both data and method noise. For method noise, the authors could consider changing the analysis methods a bit and see how sensitive the analysis results are to some small changes in the analysis procedure. For example, if Spearman rank correlation instead of Pearson correlation was used, how would the results be affected?*

We believe that the recalculations with different linkage criteria and distance metrics we showed in the responses to comments #1 and #2 already address this point to a large extent, demonstrating robustness of the analysis to changes in methodology.

For testing data noise, we find it non-trivial to select a correct functional form of noise to add to the dataset (as one might add some small amount of noise drawn from a Gaussian distribution to simulate measurement error) since the values used in our study are test statistics: the S-score is a modified t-statistic quantifying the deviation from average *S. cerevisiae* colony sizes and correcting for batch effects, and the p-values for **Fig. 4** are from Pearson correlations of S-scores.

Instead, we evaluated to what extent the clustering procedures were robust to subsampling of the datasets and included the analysis below in **Supplementary File 1**, section **Sub-sampling EMAP data**.

We randomly subsampled the library genes in the Gsp1 E-MAP (**Fig. 1d**) and found that similar groupings of mutants were maintained down to 60% of the library (**Figure** below).

A picture containing text

Description automatically generated

**Subsampling of *S. cerevisiae* alleles maintains clustering of Gsp1 mutants based on their E-MAP profiles.** GI profiles of Gsp1 mutants. Negative S-score (blue) represents synthetic sick/lethal GIs, positive S-score (yellow) represents suppressive/epistatic GIs. Mutants and genes are hierarchically clustered by Pearson correlation. As in **Fig. 1d**, all 55 point mutants are included in the clustering of columns, but only the dendrogram branch containing the strong mutants is shown. The clustering of mutants is robust to subsampling, with similar ordering of mutants observed down to removal of at least 40% of library genes.

We also sub-sampled the *S. cerevisiae* alleles in the correlation p-value matrix (**Fig. 4a**) and found that the groupings of mutants were also maintained down to 60% (**Figure** below).

A picture containing window, indoor, light

Description automatically generated

**Random subsampling of *S. cerevisiae* alleles maintains clustering of Gsp1 mutants based on the by the p-value of Pearson correlations of their GI profiles.** The p-value is a false discovery rate adjusted one-sided (positive) p-value of the Pearson correlations (represented as gray scale). The grouping of mutants into the three observed groups is robust to subsampling, as the groups are maintained down to removal of at least 50% of alleles.

**References cited in the response letter**

Benjamini, Y., & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.

Braberg, H. *et al.* (2014) Quantitative analysis of triple-mutant genetic interactions. *Nat Protoc* 9, 1867–1881.

Collins, S. R. *et al.* (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*, *446*(7137), 806–810.

Costanzo, M. *et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353, aaf1420–aaf1420.

Murtagh, F., & Legendre, P. (2014) Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion?. *Journal of classification*, *31*(3), 274-295.