***eLife’s* transparent reporting form**

We encourage authors to provide detailed information *within their submission* to facilitate the interpretation and replication of experiments. Authors can upload supporting documentation to indicate the use of appropriate reporting guidelines for health-related research (see [EQUATOR Network](http://www.equator-network.org/%20)), life science research (see the [BioSharing Information Resource](https://biosharing.org/%22%20%5Ct%20%22_blank)), or the [ARRIVE guidelines](http://www.plosbiology.org/article/info%3Adoi/10.1371/journal.pbio.1000412) for reporting work involving animal research. Where applicable, authors should refer to any relevant reporting standards documents in this form.

If you have any questions, please consult our Journal Policies and/or contact us: editorial@elifesciences.org.

**Sample-size estimation**

* You should state whether an appropriate sample size was computed when the study was being designed
* You should state the statistical method of sample size computation and any required assumptions
* If no explicit power analysis was used, you should describe how you decided what sample (replicate) size (number) to use

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

We do not have sample-size estimation per-se, the most closely related sections of our manuscript are: “Estimating fluorescence for each genotype from HiSeq sequencing of sorted populations” and “Genotype data filtering” in the Methods section.

**Replicates**

* You should report how often each experiment was performed
* You should include a definition of biological versus technical replication
* The data obtained should be provided and sufficient information should be provided to indicate the number of independent biological and/or technical replicates
* If you encountered any outliers, you should describe how these were handled
* Criteria for exclusion/inclusion of data should be clearly stated
* High-throughput sequence data should be uploaded before submission, with a private link for reviewers provided (these are available from both GEO and ArrayExpress)

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

The number of replicates is detailed in “Genotype data filtering” and “Urea sensitivity assays” sections of Methods and also in Table 1, Supplementary Data 2 and Supplementary Data 3

**Statistical reporting**

* Statistical analysis methods should be described and justified
* Raw data should be presented in figures whenever informative to do so (typically when N per group is less than 10)
* For each experiment, you should identify the statistical tests used, exact values of N, definitions of center, methods of multiple test correction, and dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals; and, for the major substantive results, a measure of effect size (e.g., Pearson's r, Cohen's d)
* Report exact p-values wherever possible alongside the summary statistics and 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05.

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

The information can be found in Figure Legend of SFigure 3, SFigure 9 and SFigure 11, and pages 17, 25 and 31.

(For large datasets, or papers with a very large number of statistical tests, you may upload a single table file with tests, Ns, etc., with reference to sections in the manuscript.)

**Group allocation**

* Indicate how samples were allocated into experimental groups (in the case of clinical studies, please specify allocation to treatment method); if randomization was used, please also state if restricted randomization was applied
* Indicate if masking was used during group allocation, data collection and/or data analysis

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

This information can be found in the “Modelling the fitness landscape of GFPs with neural networks” section.

**Additional data files (“source data”)**

* We encourage you to upload relevant additional data files, such as numerical data that are represented as a graph in a figure, or as a summary table
* Where provided, these should be in the most useful format, and they can be uploaded as “Source data” files linked to a main figure or table
* Include model definition files including the full list of parameters used
* Include code used for data analysis (e.g., R, MatLab)
* Avoid stating that data files are “available upon request”

Please indicate the figures or tables for which source data files have been provided:

Source data and raw data are available for all main and supplementary figures. Furthermore, we include the source data for the gels and measurements in SFigure 8:

figure\_S8\_source\_data\_1: Coomassie staining of full lysates, pellet, and supernatant fractions of cells expressing non-fluorescent variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Raw data.

figure\_S8\_source\_data\_2: Coomassie staining of full lysates, pellet, and supernatant fractions of cells expressing non-fluorescent variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Uncropped, annotated gel image.

figure\_S8\_source\_data\_3: Coomassie staining of full lysates, pellet, and supernatant fractions of cells expressing bright variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Raw data.

figure\_S8\_source\_data\_4: Coomassie staining of full lysates, pellet, and supernatant fractions of cells expressing bright variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Uncropped, annotated gel image.

figure\_S8\_source\_data\_5: Anti-His-Tag western blotting of full lysates, pellet, and supernatant fractions of cells expressing non-fluorescent variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Raw data.

figure\_S8\_source\_data\_6: Anti-His-Tag western blotting of full lysates, pellet, and supernatant fractions of cells expressing non-fluorescent variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Uncropped, annotated blot image.

figure\_S8\_source\_data\_7: Anti-His-Tag western blotting of full lysates, pellet, and supernatant fractions of cells expressing bright variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Raw data.

figure\_S8\_source\_data\_8: Anti-His-Tag western blotting of full lysates, pellet, and supernatant fractions of cells expressing bright variants from amacGFP, cgreGFP, and ppluGFP2 libraries. Uncropped, annotated blot image.