

RESEARCH

Open Access



Optimization of a deep mutational scanning workflow to improve quantification of mutation effects on protein–protein interactions

Alexandra M Bendel^{1,2}, Kristjana Skendo³, Dominique Klein¹, Kenji Shimada¹,
Kotryna Kauneckaitė-Griguolė^{1,2} and Guillaume Diss^{1*}

Abstract

Deep Mutational Scanning (DMS) assays are powerful tools to study sequence–function relationships by measuring the effects of thousands of sequence variants on protein function. During a DMS experiment, several technical artefacts might distort non-linearly the functional score obtained, potentially biasing the interpretation of the results. We therefore tested several technical parameters in the deepPCA workflow, a DMS assay for protein–protein interactions, in order to identify technical sources of non-linearities. We found that parameters common to many DMS assays such as amount of transformed DNA, timepoint of harvest and library composition can cause non-linearities in the data. Designing experiments in a way to minimize these non-linear effects will improve the quantification and interpretation of mutation effects.

Keywords Deep mutational scanning, Protein–protein interactions, High-throughput assay, Non-linearity, Method optimization

Background

How information encoded in DNA sequence is translated into molecular function is central to our understanding of the genotype–phenotype relationship. Deep Mutational Scanning (DMS) assays allow to evaluate the effect of thousands of sequence variants on a given molecular function in parallel and have therefore become powerful tools to address this question. In DMS assays, a variant library is created by mutagenizing a sequence of interest and then exposed to a selective assay that separates variants based on their activity. Because the assay is

performed in a pooled format, high-activity variants are enriched under the selective conditions while low-activity variants are depleted. Enrichment and depletion are quantified using deep sequencing [1, 2].

Proteins perform most biochemical functions within a cell and do so by interacting with one another. The specificity of these protein–protein interactions (PPIs), i.e. the information about which two proteins interact with one another, is encoded in the DNA sequence and is therefore central to the sequence–function relationship. Moreover, it has been shown that disease-associated variants are enriched at interaction interfaces, which is why perturbations of PPIs are of major interest in medical and pharmacological research [3].

The fast-progressing advances made in sequencing technology enable the discovery of more and more variants in the human population with potential effects on PPI affinity. The vast amount has long surpassed a number feasible to study individually. Recent publications

*Correspondence:

Guillaume Diss
guillaume.diss@fmi.ch

¹ Friedrich Miescher Institute for Biomedical Research (FMI), Basel, Switzerland

² University of Basel, Basel, Switzerland

³ University of Geneva, Geneva, Switzerland



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

have shown that DMS assays are very well suited for the study of PPIs or protein-ligand interactions at large scale [4–10].

Deep sequencing-based protein complementation assay (deepPCA) is one DMS approach to measure PPIs that proved to be very useful to study determinants of specificity in conserved protein interaction domains such as human basic leucine zippers [5, 11]. In this assay, a Dihydrofolate Reductase-based Protein-fragment Complementation Assay (DHFR-PCA) is combined with DMS by performing DHFR-PCA in a pooled format and measuring the effect of thousands of mutations in a protein of interest on its ability to interact with its partners. DHFR-PCA is based on the reconstitution of a methotrexate (MTX)-insensitive mouse DHFR variant (Fig. 1A).

Both halves of the enzyme (referred to as DH-tag for the N-terminal fragment and FR-tag for the C-terminal fragment) are used to tag each of the two proteins of interest and expressed in *Saccharomyces cerevisiae* cells. Interaction between these two proteins promotes complementation between the two DHFR fragments and sustains growth in presence of MTX, which inhibits the endogenous DHFR [12, 13].

DHFR-PCA is quantitative, i.e. the growth rate of the respective cells depends on the concentration of reconstituted DHFR, which is itself a function of protein abundance and the affinity of the interaction [14]. In deepPCA, this quantitative aspect is leveraged in a pooled setting, where variants of the interacting proteins are enriched or depleted depending on the strength of their PPI.

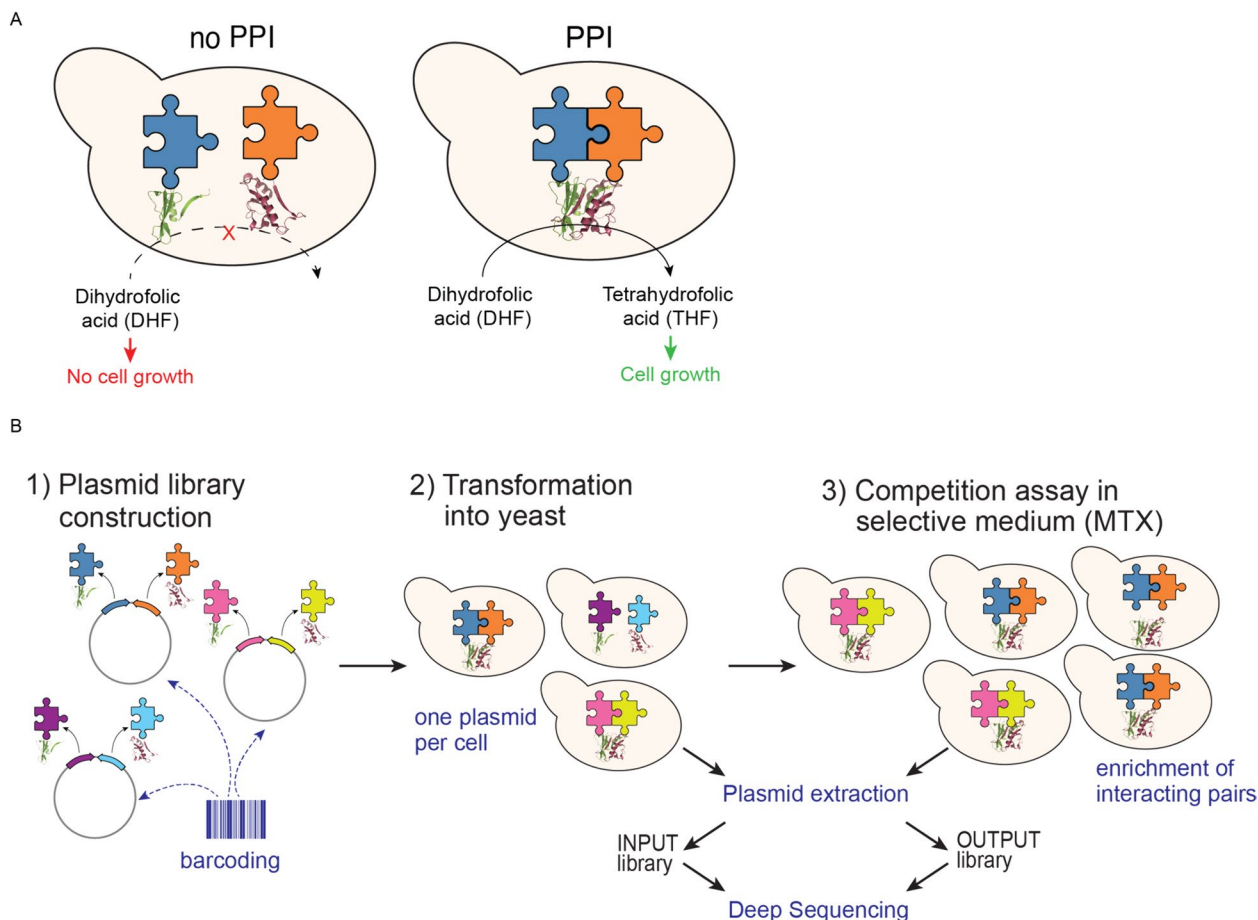


Fig. 1 The deepPCA general principle and workflow. **A** In DHFR-PCA, proteins of interest are tagged with the two opposite halves of a MTX-insensitive DHFR variant. Upon interaction of the tagged proteins, DHFR is reconstituted and functional. Yeast cells expressing the interacting pair can sustain cell growth in presence of MTX [13]. **B** deepPCA: pairs of proteins tagged with opposite DHFR termini are expressed from single plasmids. Molecular barcodes encoded on the plasmids allow identification of the protein pair. Plasmids are transformed into yeast cells so that each cell expresses one plasmid. Cells expressing interacting proteins are enriched in selective medium containing MTX. Enrichment and depletion of interacting and non-interacting pairs is quantified by barcode sequencing of plasmids extracted from input and output cultures [11]

deepPCA has been designed to allow library-on-library screening. First, intermediate libraries for DH- and FR-fusions are cloned with random DNA barcodes. The association between barcode and variant is determined for each library by deep sequencing and the two libraries are then combined on the same plasmid in a way that juxtaposes the barcodes. This plasmid library is then transformed into yeast cells at a low multiplicity of transformation, i.e., to minimize the number of cells with two or more plasmids. Competitive growth in MTX-containing medium then results in enrichment of interacting pairs and depletion of non-interactors. Input and output cultures (before and after competition) are harvested, and plasmids are extracted.

Enrichment and depletion of individual pairs are quantified by deep sequencing of the molecular barcodes encoded on the plasmids. Frequencies of each pair in input and output populations are determined and used to calculate the growth rate (generations per hour) of cells expressing each pair of proteins as a proxy for the concentration of the complex (Fig. 1B).

deepPCA's readout, like many other DMS assays, maps non-linearly to the latent trait where mutations are expected to have additive effects [5, 15]. Indeed, the readout is directly proportional to the concentration of the complementation complex, which is itself non-linearly related to the underlying energetic dimension where mutations that are not energetically coupled affect the free energy of the complex additively. Mutations that have additive effects at the energetic level will thus show non-additive effects at the level of protein complex concentration, which can be mis-interpreted as the two mutations having some structural or functional relationship. This phenomenon has been referred to as global, phantom or non-specific epistasis [5, 16–18]. Regressing out these non-linearities is therefore required to access mutational effects on the latent additive trait. This is typically achieved by employing an appropriate mechanistic model describing the system, such as two- or three-states thermodynamic models for PPIs [5, 6, 11, 17, 19, 20].

The experimental pipeline can also introduce additional non-linearities when technical factors affect different variants in different ways depending on their position within the dynamic range of the assay. These non-linearities can sometimes be accounted for even when not explicitly included in the thermodynamic model, for instance by being incorporated in some global parameters determining the shape of other non-linearities. However, they can still have negative effects by restricting the dynamic range of the assay. It is therefore important to identify potential sources of non-linearity in the workflow to improve the quantification of mutational effects.

The purpose of this study is thus to perform optimization experiments at different critical steps of the experimental pipeline to identify sources of non-linearity in the deepPCA workflow. The parameters tested were 1) amount of transformed library DNA, 2) MTX concentration, 3) timepoint of harvest, and 4) library composition. This improved our understanding of both the biological/physical and technical non-linearities and will allow us to account and correct for these sources of error in our experimental design. The improved data quality will enable a more accurate thermodynamic modelling of the underlying biophysical processes, benefiting a deeper understanding of evolutionary, epistatic, and pathogenic effects of mutations. Currently, many DMS assays follow a similar experimental design [21], with read-outs that can be affected by the same technical parameters, especially 1), 3) and 4). Therefore, our findings for deepPCA are not limited to this particular assay but can also be applied to a variety of other DMS applications.

Results and discussion

A yeast transformation protocol with minimal double transformants and optimized library coverage

If a cell contains two plasmids expressing pairs of proteins with different interaction strengths, the cell's growth rate is mostly determined by the strong interaction pair and the change in frequency of the weak pair will be overestimated, which can be a source of non-linearity. Therefore, to preserve the quantitative nature of DHFR-PCA, the fraction of cells carrying two or more different plasmids must be minimized.

To determine the proportion of cells co-transformed by more than one plasmid in our large scale, high-efficiency transformation protocol, we transformed different amounts (between 250 ng and 1 µg) of an equimolar mixture of two centromeric plasmids (pRS415 and pDL00266) carrying each a different auxotrophic selection marker. We then plated on selective plates missing one or both corresponding metabolites to count the total number of transformed cells and the number of double transformants and calculated the percentage of cells carrying more than one plasmid (see Methods). Transformants are selected in two consecutive rounds for many generations before the start of the competition in the MTX medium, in order to overgrow the large majority of non-transformed cells. This should allow for cells carrying two different plasmids to lose one and thus decrease the percentage of double transformants [22]. To measure the rate of loss, we compared the percentage of cells carrying more than one plasmid directly after transformation and at the end of the selection process, i.e., just before inoculation of the MTX medium (Supplementary Table 1).

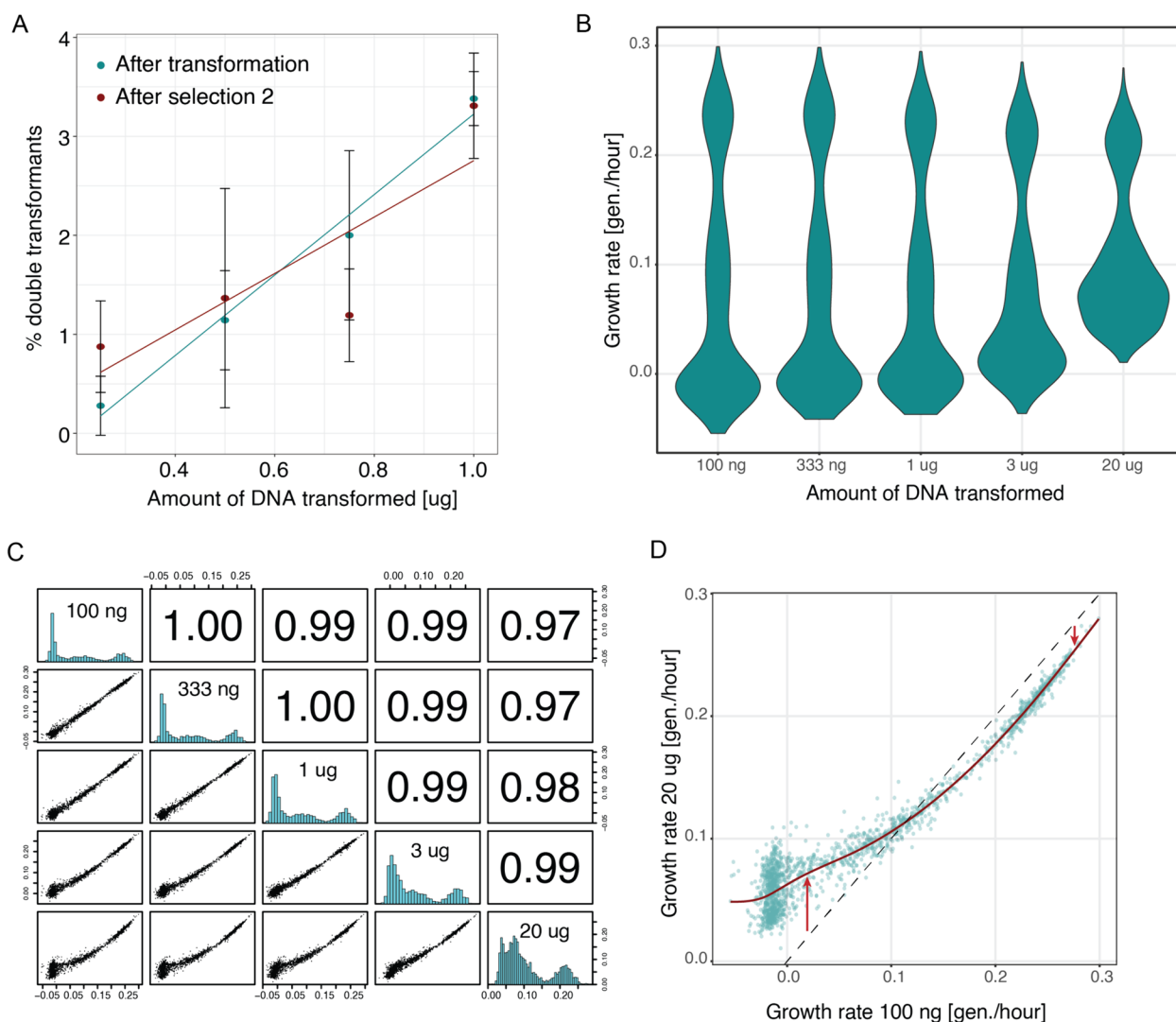


Fig. 2 Effect of transforming with different amounts of DNA. **A** Fraction of double transformants per cell for different amounts of DNA transformed. Cells were transformed with DNA amounts from 250 ng to 1 μ g. The amount of double transformants was quantified after initial transformant selection and after selection round 2 (see Methods). **B** Distributions of growth rates obtained from deepPCAs performed with cells transformed with differing amounts of DNA between 100 ng and 20 μ g. **C** Pairwise correlations of growth rates between samples from deepPCAs performed with cells transformed with differing amounts of DNA. Scatter plots are shown in the bottom left panels, histograms are shown on the diagonal, and Pearson correlation coefficients are shown in the upper right panels. **D** Scatterplot between samples transformed with 100 ng DNA and 20 μ g DNA

The percentage of double transformants directly after transformation and before MTX selection increased in a linear fashion with increasing DNA amount (Fig. 2A). The total number of transformants also increases in a linear fashion when transforming with higher amounts of DNA (Supplementary Fig. 1A). Moreover, it remained stable indicating that the cells maintain both plasmids over the multiple generations of growth.

We next measured the effects of transforming different DNA amounts and hence different percentages of double transformants on deepPCA interaction scores.

We constructed a small library based on a JUN variant library recently assayed [11]. We combined the 608 single amino acid substitutions of the leucine zipper domain of human JUN (FR-tagged) with three out of 54 wild-type human basic leucine zippers (bZIPs; DH-tagged). The three wildtype bZIPs were selected to span a range of interaction strengths between JUN and its wild-type partners. FOS was used as a strong interactor, ATF7 as a weak interactor and NFE2 as a non-interactor. This library was used in most experiments presented in this paper and will be referred to as the balanced JUN library.

We performed deepPCA after transforming yeast cells with five different amounts of library DNA, expanding the range used above to better detect effects on interaction scores (100 ng, 333 ng, 1 μ g, 3 μ g, and 20 μ g). We first compared the correlations between variant frequencies in input and output samples across all conditions. Overall, the correlations were very similar, but it was apparent that the output samples from the cells transformed with 20 μ g of library DNA correlated stronger with all input samples than the other quantities transformed, indicating that library variants are less enriched or depleted when transformed with this high DNA amount (Supplementary Fig. 1B; Supplementary Table 2). To further investigate this, we calculated the per variant growth rate (generations per hour) for each sample and observed a narrower distribution in the 20 μ g sample while the distribution in the remaining samples were very similar (Fig. 2B; Supplementary Table 3). When looking at the correlations between growth rates, the correlation coefficients were very high but decreasing with increasing difference in DNA amount (Fig. 2C). This is once more especially apparent for the 20 μ g sample for which we observe a clear non-linear trend with growth rates of other samples. This can be explained by the expected high average number of plasmids per cell in that sample. For example, when directly comparing growth rates of the sample transformed with 100 ng DNA and the one transformed with 20 μ g DNA, slow growers are most severely affected and grow faster in the high DNA sample (Fig. 2D). This is because slow growers will likely get co-transformed with a plasmid that expresses a pair that interacts more strongly, and thus increase the growth rate of the cell in MTX beyond the level expected for the weakly interacting variant alone. There is only a slight decrease in growth rates for the strong interactors, which might be due to interference from the weaker co-transformed plasmid on the expression from the stronger plasmid. Importantly, these effects are not stochastic and affect the trend, as further demonstrated by the high correlation between output replicates of the 20 μ g sample (Supplementary Fig. 1C). Because of the large number of transformants per variant pair obtained (~1000) and the high co-transformation percentage expected with 20 μ g of DNA, each variant pair is expected to be co-transformed with a large fraction of all other variant pairs. The average effect thus becomes deterministic. On the contrary, if the average number of transformants per variant was low, a variant pair in one replicate might be co-transformed with strong variant pairs while in another replicate it might be co-transformed with weak variant pairs.

The non-linear trend and the narrowing of the distribution of interaction scores is still apparent at 3 μ g and negligible at 1 μ g and lower amounts. We thus choose 1 μ g as an optimal trade-off between the high number of total transformants required when working with large libraries and a low percentage of double transformants with minimal effects on the distribution of interaction scores.

Increasing concentration of MTX results in higher but linear selection strength

Next, we tested the effect of MTX concentration during competitive growth. Previous deepPCA studies used a standard concentration of 200 μ g/mL MTX [5, 6, 11, 5, 6], which we here refer to as 1X MTX. This concentration was first determined for performing DHFR-PCA on agar plates [13, 23], and we therefore decided to verify that it is also well-suited for deepPCA in liquid culture. We evaluated the effect of 1/20X, 1/4X, 1/2X, 1X, and 2X MTX concentration and a DMSO only control, controlling for the effects that some variants can have on actual cellular fitness, for instance by interfering with other cellular components. In addition, we also compared the results from different MTX vendors (using 1X MTX in the assay; Supplementary Tables 4 & 5).

We did not observe any effect from the supplier (Supplementary Fig. 2A). We also did not observe any fitness effect in DMSO (Pearson correlation coefficient of 1.00 between input counts and output counts from cells grown in DMSO; Figure 3A, upper left). On the contrary, in presence of as low as 1/20X MTX, enrichments of some variants can be observed which is reflected in the lower correlation coefficients to input counts (Fig. 3A). The correlation coefficients decrease together with increasing MTX concentration, indicating stronger selection at higher MTX concentrations. The correlation between output counts of the 1/4, 1/2, 1 and 2X samples is 0.99, further indicating little differences in enrichments at different MTX concentrations.

Increasing MTX concentration stretches the growth rate distributions (Fig. 3B), further indicating stronger selection regimes. The lower mode of the distribution decreases (lower modes of 0.0984, 0.05, 0.0347, 0.0164, and 0.00532 for 1/20X, 1/4X, 1/2X, 1X, and 2X MTX concentration, respectively) while the fast-growing variants are little affected (upper modes are 0.2448, 0.238, 0.2272, 0.2322, and 0.225 for 1/20X, 1/4X, 1/2X, 1X, and 2X MTX concentration, respectively), indicating a higher background growth at lower MTX concentration, probably resulting from an incomplete inhibition of the endogenous DHFR. Since the growth rates show a very high correlation at different MTX concentrations and do not appear non-linearly related (Supplementary Fig. 2B), increasing it from 1 to 2X does not confer significant

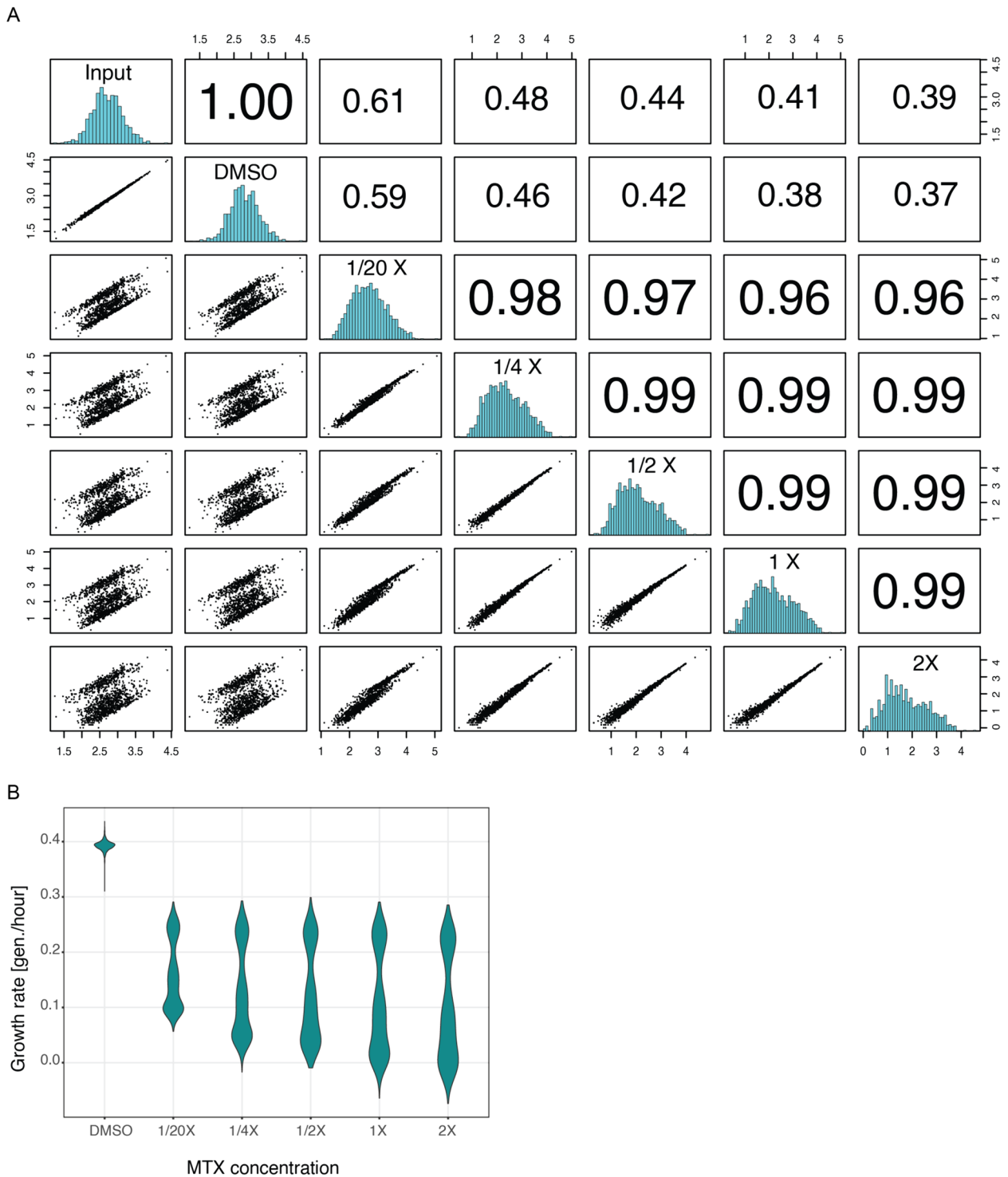


Fig. 3 Different MTX concentrations do not result in non-linearities, but selection pressure is higher in higher concentration. **A** Pairwise correlations of average counts between input and output samples of deepPCAs selected in DMSO and different concentrations of MTX. Scatter plots are shown in the bottom left panels, histograms are shown on the diagonal, and Pearson correlation coefficients are shown in the upper right panels. **B** Distributions of growth rates of deepPCAs selected in DMSO and different concentrations of MTX

benefit and would not justify the higher cost. However, decreasing MTX concentration 2X or 4X would have the advantage of a lower cost and might therefore be considered for very large and costly experiments. The slightly higher background growth can be accounted for by the global parameters of the thermodynamic model that linearly rescales the growth rate to the fraction bound [6]. We thus conclude that 1X is an optimal trade-off between cost and selection strength.

The lag-phase at the start of the competition non-linearly affects growth rates

We next set out to determine the ideal harvesting strategy to maximize the size and spread of the measured effects, while keeping the error small. To this end, different harvesting time points for the input and output samples were considered. Previously, the input cultures were harvested before inoculating the MTX-containing selection medium. A portion of these cells was used to inoculate the competition culture which was grown for five more generations before harvesting the output sample. At the beginning of the growth in the MTX medium, cells have enough stores of tetrahydrofolate (THF), the product of DHFR, and downstream metabolites to sustain growth until relying solely on the complemented DHFR activity. Therefore, cells expressing different pairs of proteins are expected to grow at the same rate for a few hours. This is clearly apparent in screens on agar plates, where cells at all array positions were able to grow to form small colonies and only the positions at which cells expressed interacting proteins would grow further. This lag before the selection kicks in can thus potentially non-linearly affect the estimated growth rate such that the difference between true and estimated growth rate would be larger for weak interactors since the MTX-dependent growth time is overestimated. Strong interactors, on the other hand, might not see a strong decrease in growth rate upon MTX selection, and their estimated growth rate will be very close to the true growth rate (Fig. 4A).

To investigate this effect, we used the same JUN balanced library and compared growth rates obtained following the normal protocol to those obtained by first pre-incubated cells in the MTX medium for 2.5 hours

before collecting the input sample and inoculating fresh MTX medium. Both output cultures were harvested at a cell density corresponding to a population growth of five generations. As expected, we do see an apparent slight non-linear relationship between growth rates with a stronger effect on weak interactors, i.e., pairs with a lower growth rate that show a disproportionately stronger increase in generation time than the strong interactors (Fig. 4B; Supplementary Tables 4 & 6).

Pre-incubation in MTX is therefore recommended to minimize non-linear effects. However, the extra steps make the experimental procedure more complex and costly and might add to the technical noise of the data while the non-linear effects can be accounted for by the global parameters of the thermodynamic model that control the curvature of the non-linear transformation [6]. The decision of whether or not to perform the pre-incubation is thus left to the experimenter depending on the use of the data.

The choice of harvesting timepoint is another source of non-linearities

The number of generations in competitive growth might also result in differences between estimated and true growth rate (Fig. 4C). We therefore performed deepPCA by harvesting the input samples before MTX selection and harvesting the output samples after 2, 3, 4, 4.5 and 5 generations of total population growth (Supplementary Tables 7 & 8). For this experiment, we used the full JUN variant library combined with the 54 wt bZIPs from [11]. We first compared the distributions of generation times across the different samples. Though similar, the harvest after 2 generations shows a slight shift towards higher average growth rates indicating that the selective pressure was not applied long enough for strong depletion of the slow growers (Supplementary Fig. 3A). Additionally, a non-linear relationship between the samples harvested after different generation times that seems to affect slow-growers disproportionately stronger is observed. The higher the difference in generations at harvest the more pronounced is this effect, being the strongest between the samples harvested at 2 and at 5 generations (Fig. 4D, Supplementary Fig. 3B). This can also be explained by the

(See figure on next page.)

Fig. 4 Growth time and pre-selection phase non-linearly affect growth rates. **A** Harvesting input samples before or after a pre-competition phase in which all yeast cells still grow at equal growth rates can result in differences in measured growth rate. Harvesting before pre-competition phase can result in an overestimation of growth rate of slow growers because the assumed duration of selection is longer than the actual duration. **B** Scatter plot comparing growth rates of deepPCAs for which the input sample has been harvested before or after the pre-competition phase (after 2.5 hours in MTX). **C** Harvesting after shorter growth time can result in further overestimation of growth rate for slow growers because the pre-competition phase represents a larger fraction of the overall selection time and more strongly influences the estimated growth rate than after a long growth time. **D** Scatter plot comparing samples of deepPCAs grown for 2 generations and 5 generations. **E** Standard errors of growth rate as a function of growth rate for samples harvested after different generations

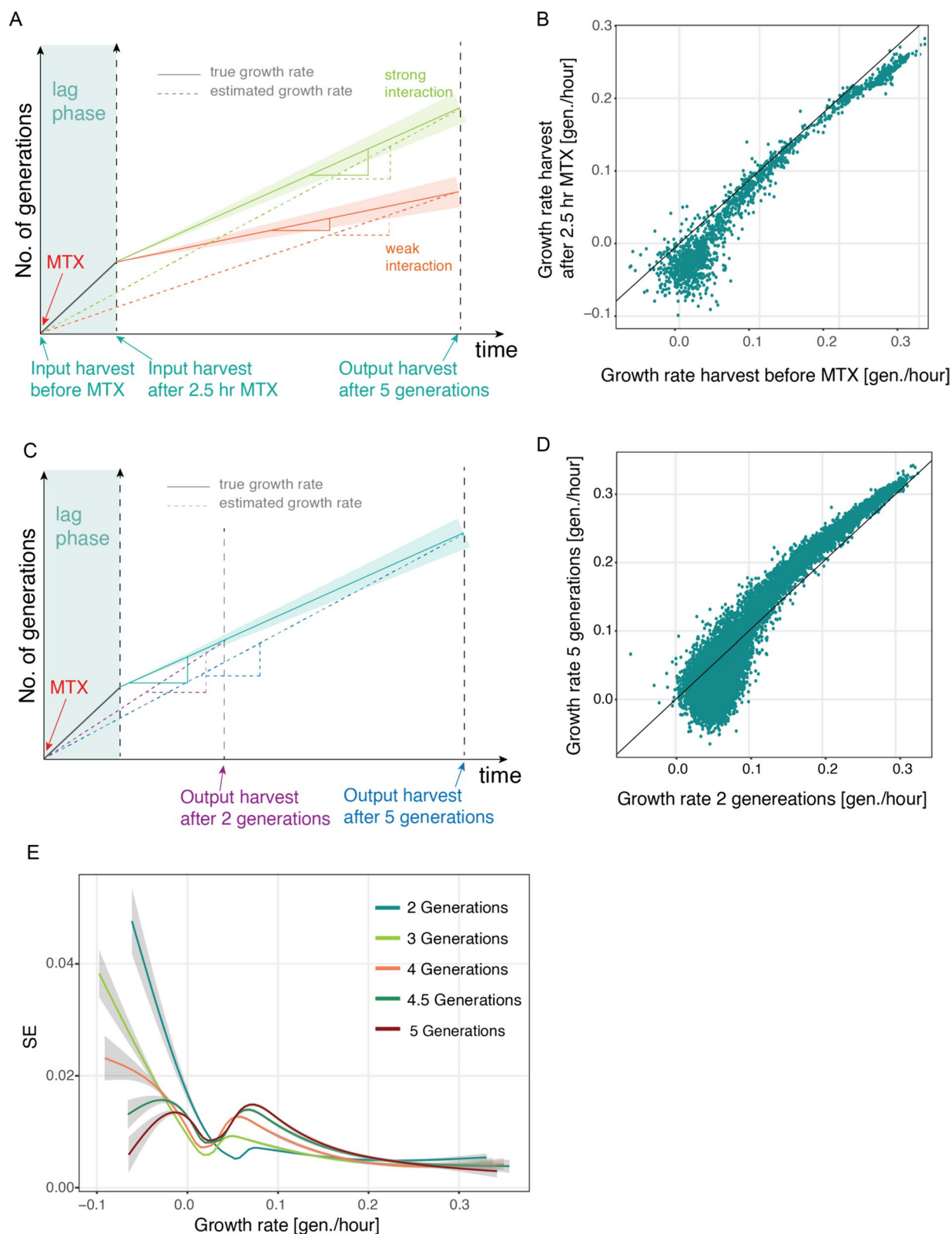


Fig. 4 (See legend on previous page.)

effect of the lag-phase, which is proportionately stronger when the effective competition is shorter (Fig. 4C). For this reason, a harvest at later timepoint (i.e., after 5 generations) will result in estimated growth rates that are closer to the true growth rates than those estimated from samples harvested earlier.

On the other hand, with increasing time of growth, also the noise and therefore the error of measurement increases, especially for slow growers (Fig. 4E). This is likely due to stronger depletion of slow growers after prolonged growth and thus lower read counts, the main source of measurement error [24, 25]. Thus, choosing the timepoint of harvest is a tradeoff between how close to the true growth rates the estimated ones are and how high of an error and drop-out rate of slow growers is tolerable.

Library composition

In principle, measured growth rates for individual protein pairs should be independent of library compositions. To test this, we performed deepPCA using our pre-determined standard settings of 1 µg of DNA for transformation, harvest of input samples before the competition and harvest of output samples after 5 generations of growth. We screened three different libraries of identical sizes, the balanced JUN library, a slow-growing library in which we replaced the strong interactor FOS by the non-interactor ATF2, and a fast-growing library in which we replaced the non-interactor NFE2 by the strongest interactor JDP2. Importantly, ATF7 was present in all three libraries, allowing to compare growth rates and test if they are affected by library composition (Supplementary Tables 9 & 10). Growth rates for the ATF7 pairs were well correlated between the different libraries (Fig. 5A). However, their distributions differ, and the growth rates estimated for ATF7 pairs when competing against stronger pairs are lower (Fig. 5B). This is also true for other partners that overlap only in two libraries. For example, FOS shows overall higher growth rates in the balanced library where it forms the strongest interactions with JUN variants than compared to the fast-growing library where JDP2 forms the strongest interactions with JUN variants. This is more likely to be due to technical artefacts than actual differences in growth rates, such as inaccuracies in the estimation of frequencies from read counts. However, this has little impact on downstream analyses since modelling usually transforms growth rates to a different scale more relevant to a protein's molecular activity.

Conclusions

We have performed a multi-parameter optimization study of deepPCA to identify sources of non-linearity in the workflow. Parameters like amount of transformed

DNA and timepoint of harvest were identified to introduce non-linearities in the data. While these can easily be accounted for by thermodynamic modeling, identifying these sources of non-linearity can help improve the modelling process. Non-linearities can also affect the dynamic range of the assay, for instance by pushing variants into a saturation plateau. Different variants with different scores could thus appear similar because of the non-linearity in the read-out. Adapting the experimental procedure to minimize this non-linearity would also help improve quantification accuracy and ultimately draw correct conclusions about underlying biological processes.

In many cases, the choice of parameter is a trade-off between technical feasibility or cost and measurement accuracy. For most of our purposes, an optimal trade-off between total number of transformant obtained and frequency of double transformants, which affects measurement linearity, consists in transforming 1 µg of plasmid DNA (under the experimental conditions used here). The choice of MTX concentration is a trade-off between higher cost and a wider growth rate scale leading to more separation, i.e. larger differences, between the measured growth rates. We therefore recommend sticking to the concentration of 200 µg/mL that has been used so far. This can however be reduced to 50 µg/mL if cost becomes limiting without much impact on quantification. Concentrations of 10 µg/mL or lower should be avoided. We recommend a timepoint of harvest after 5 generations of growth as this prevents a strong influence of the initial lag phase on the measured growth rates. However, for libraries that consist of primarily slow-growing variants, a shorter growth time is advised as the prolonged growth time can result in noisy data.

We do note that all these experiments were performed using rather small libraries. Although our findings should be generally applicable, it is essential to thoroughly plan every DMS assay beforehand and take into consideration potential additional sources of non-linearity that might come from different library sizes or properties of the proteins screened.

Methods

Yeast strain

All experiments presented were performed in BY4742 (MAT α *his3* Δ 1 *leu2* Δ 0 *lys2* Δ 0 *ura3* Δ 0).

Media and buffer recipes

- LB: 25 g/L Luria-Broth-Base (Invitrogen, Waltham, MA, USA). Autoclaved 20 min at 121 °C.
- LB-agar with 2X Ampicillin (100 µg/mL): 25 g/L Luria-Broth-Base (Invitrogen, Waltham, MA, USA),

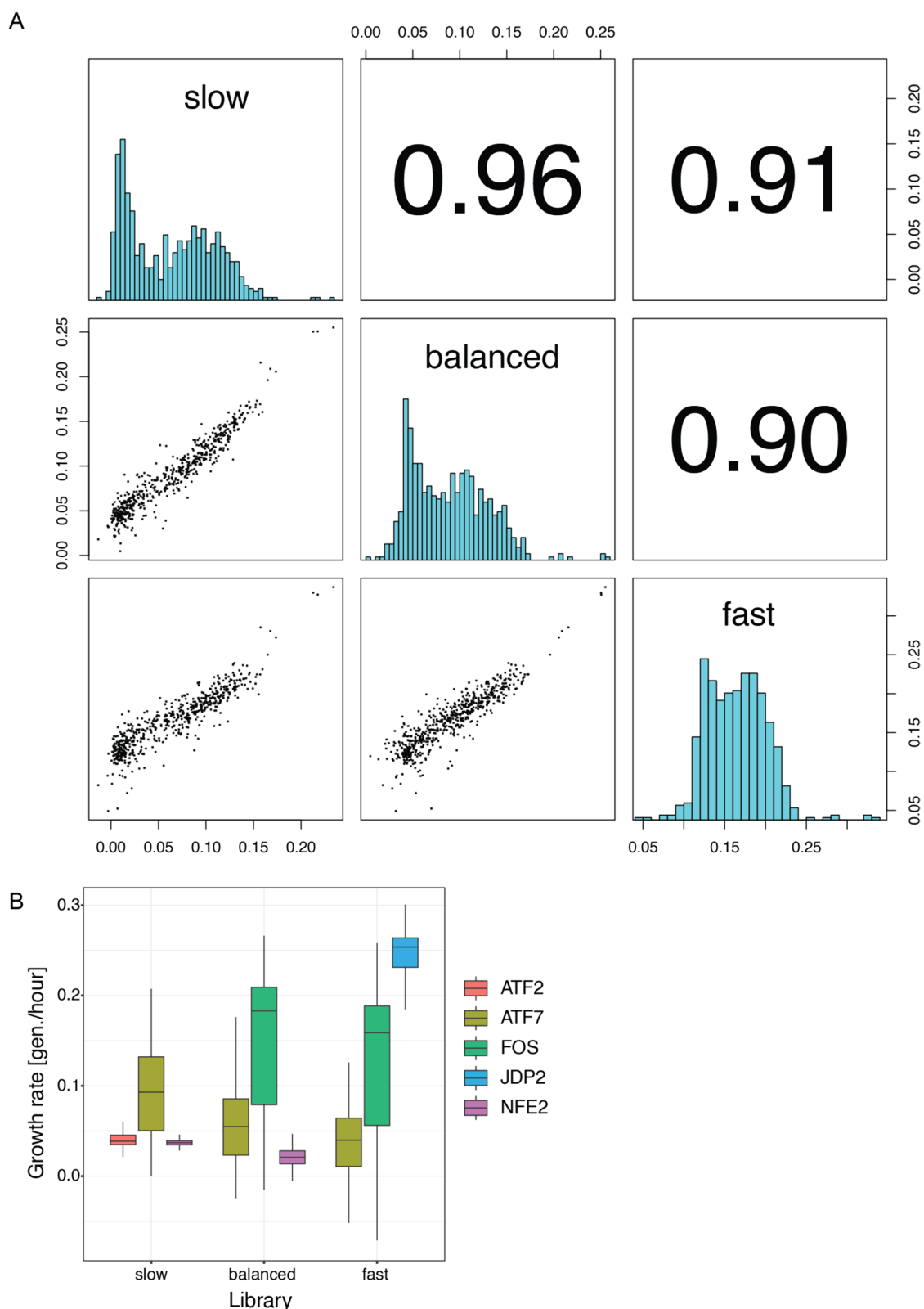


Fig. 5 Growth rates are not independent of library composition. **A** Pairwise correlations of overlapping components between growth rates obtained from deepPCAs performed with different libraries. Scatter plots are shown in the bottom left panels, histograms are shown on the diagonal, and Pearson correlation coefficients are shown in the upper right panels. **B** Distributions of growth rates for individual components obtained from deepPCAs of slow-growing, balanced and fast-growing libraries

7.5 g/L Agar, 1.2 g/L MgSO₄H₂O. Autoclaved 20 min at 121 °C. Cool-down to 45 °C. Addition of 100 mg/L Ampicillin.

- YPAD: 20 g/L Bacto-Peptone, 20 g/L Dextrose, 10 g/L Yeast extract, 25 mg/L Adenine. Filter-sterilized (Millipore Express[®] PLUS 0.22 µm PES, Merck, Darmstadt, Germany).
- SC-ura: 6.7 g/L Yeast nitrogen base without amino acids, 20 g/L glucose, 0.77 g/L complete supplement mixture drop-out without uracil. Filter-sterilized (Millipore Express[®] PLUS 0.22 µm PES, Merck, Darmstadt, Germany).
- SC-ura/ade/met: 6.7 g Yeast nitrogen base without amino acids and folic acid, 20 g/L glucose, 0.74 g/L complete supplement mixture drop-out without uracil, adenine and methionine. Filter-sterilized (Millipore Express[®] PLUS 0.22 µm PES, Merck, Darmstadt, Germany).
- SORB: 1 M sorbitol, 100 mM LiOAc, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0. Filter-sterilized (Millipore Express[®] PLUS 0.22 µm PES, Merck, Darmstadt, Germany).
- Plate mixture: 40 % PEG3350, 100 mM LiOAc, 10 mM Tris-HCL pH 8.0, 1 mM EDTA pH 8.0. Filter-sterilized (Millipore Express[®] PLUS 0.22 µm PES, Merck, Darmstadt, Germany).
- Recovery medium: YPAD + 0.5 M sorbitol. Filter-sterilized (Millipore Express[®] PLUS 0.22 µm PES, Merck, Darmstadt, Germany).
- Standard competition medium: SC-ura/ade/met + 200 µg/mL methotrexate (BioShop Canada Inc., Canada), 2 % DMSO.
- DTT buffer: 0.1 M EDTA-KOH pH7.5, 10 mM DTT
- Zymolyase buffer: 20 mM K-phosphate pH 7.2, 1.2 M sorbitol, 0.4 mg/mL Zymolyase 20T (amsbio, USbiological), 100 µg/mL RNase A
- Methotrexate (Sigma Aldrich/Merck, Darmstadt, Germany).

Plasmid

The libraries used in this study were derived from [11].

To study the effect of harvesting the cells after different growth generations, we used the exact same library as in [11]. For this, an intermediate library of all single amino acid variants of the JUN zipper region fused to the DHFR c-terminal half (referred to as FR-tag) and identifiable by 24 bp randomly generated molecular barcodes was created. Barcode and variants were associated as described in [11]. This library was combined with a pooled collection of plasmids carrying all 54 wt bZIPs fused to the DHFR N-terminal half (referred to as DH-tag) and identifiable by a Sanger sequencing-confirmed molecular barcode of 20 bp. The two intermediate libraries were

cloned together using high-efficiency cloning and standard restriction enzyme-mediated cloning with enzymes AvrII (New England Biolabs, Ipswich, MA) and HindIII-HF (New England Biolabs, Ipswich, MA) as described in [11]. For all barcodes, see Supplementary Tables 11 & 12.

The other experiments were performed with a sub-library of the above one. A standard balanced library was created combining the full JUN variant prey library with 3 wt bZIPs as bait. These were chosen based on their wt interaction strengths with wt JUN that were derived from [11]. To achieve a library of balanced growth rate, the JUN library was combined with ATF7, a medium-strength interactor of JUN, FOS, a strong interactor of JUN, and NFE2, a weak interactor of JUN, as bait using the same cloning strategy as above. This library was called balanced JUN library.

For testing different library compositions, two additional libraries were created. For a slow-growing library, the JUN variant library was combined with ATF7, NFE2 and one additional weak interactor of JUN, ATF2, as bait. This library will be referred to as slow JUN library. For a fast-growing library, the JUN variant library was combined with ATF7, FOS and one additional strong interactor of JUN, JDP2, as bait. Again, the same cloning strategy as above and as described in detail in [11] was applied. This library will be referred to as fast JUN library.

deepPCAs

Large-scale yeast transformation

Yeast cells of the strain BY4742 were grown to saturation overnight in YPAD. For individual replicates, single colonies were used to inoculate the culture. In the morning, different sizes of pre-culture were inoculated at an OD600 of 0.3 in 175 mL YPAD medium. The cells were grown for ca. 4 hours until they reached an OD600 between 1.2 and 1.4. (Because of the larger library size, all volumes at each step in the different generation times deepPCA were doubled. The exact protocol for transformation of this library can also be found in [11].) Next, the cells were harvested for 5 min at room temperature (RT) and 3,000 g. The pellets were washed first with 50 mL H₂O and then 50 mL SORB and finally resuspended in 7 mL SORB and incubated on a wheel at RT for 30 min. To each sample, 175 µl of 10 mg/mL salmon sperm DNA (Agilent Technologies, Santa Clara, CA) were added, and the tubes were mixed well. Afterwards, DNA was added to the samples according to Table 1 followed by thorough mixing by shaking the tube.

Afterwards, 35 mL of Plate Mixture were added to each sample and the samples were incubated at RT on a wheel for another 30 min. 3.5 mL of DMSO were added to each sample followed by a 20 min-heatshock in a 42 °C water bath. To ensure homogeneous distribution of

Table 1 Overview of DNA amounts and libraries used in individual experiments

Experiment	Amount of DNA transformed per 175 mL per-culture	Library used
Measurement of percentage of double transformants	250 ng, 500 ng, 750 ng, 1 µg (details see below)	Equimolar pool of similar sized plasmid backbones with URA3 and LEU2 auxotrophic selection markers
Transformation with different amounts of DNA deepPCA	100 ng, 333 ng, 1 µg, 3 µg, 20 µg	Balanced JUN library
Lag/no-lag phase deepPCA	1 µg	Balanced JUN library
Different MTX concentrations/brands, fitness effect in DMSO deepPCA	1 µg	Balanced JUN library
Different generation times deepPCA	3.5 µg	Full JUN vs. 54 wt bZIPs library [11]
Different library compositions deepPCA	1 µg	Balanced JUN library, slow JUN library, fast JUN library

the heat within the sample, the tubes are inverted a few times after 1, 2.5, 5, 7.5, 10 and 15 min. The cells were then spun down for 5 min at RT and 3,000 rpm. After thorough removal of the liquids, they were resuspended in recovery medium and incubated for 1 hour at 30 °C without agitation. The recovered cells were finally spun down for 5 minutes at RT and 3,000 g, the supernatant was removed, and the pellet was resuspended in 350 mL of SC-ura for selection 1. Additionally, cells were plated on SC – ura plates for counting the amount of transformants. For measuring the percentage of double transformants, cells were additionally plated of SC-leu and SC-ura-leu (see below). Selection 1 was grown for ca. 48

transformed according to the protocol above. To measure the number of transformants and double transformants, different volumes of cells that would result in countable amounts of colonies on the plates were plated after transformation according to Table 2 and grown for 2 days at 30 °C. Additionally, selection 1 cultures of 350 mL were set up for all samples. Each sample was performed in duplicate.

After 2 days, the colonies on the plates were counted, the total amount of transformants per 350 mL selection culture was calculated and for each individual sample, the percentage of double transformants was calculated using eq. 1.

$$\frac{2 \times \text{No. colonies on SC} - \text{ura} - \text{leu}}{\text{No. colonies on SC} - \text{ura} + \text{No. colonies on SC} - \text{leu} - \text{No. colonies on SC} - \text{ura} - \text{leu}} \times 100\% \tag{1}$$

hours until the cells reached saturation.

Measurement of percentage of double transformants

As plasmid with URA3 resistance, a plasmid derived from pRS416 [26] was used. As plasmid with LEU2 resistance, we used a lab-internal plasmid called pDL00266 that was derived from pRS415 [26] and was of similar size than pRS416 to facilitate equimolar mixing. An equimolar pool of the two plasmids was created and was

The averages of replicates and their standard deviations were calculated. The total number of transformants is represented by the denominator of equation 1 and corresponds to the sum of transformants on the -ura and -leu selections. Since these both also include double transformants that would hence be counted twice since present on the two plates, we then subtract the number of double transformants counted on the -ura/leu plates. The number of double transformants on the numerator of equation 1 is multiplied by two based on the assumption that

Table 2 Plating scheme for counting of double transformants directly after transformation

Sample no.	DNA amount	Plating SC - URA and SC - LEU	Plating on SC -URA/-LEU
1	250 ng	2 x 40 ul, 2 x 110 ul	2 x 400 ul, 2 x 1.1 mL
2	500 ng	2 x 35 ul, 2 x 100 ul	2 x 350 uL, 2 x 1 mL
3	750 ng	2 x 20 ul, 2 x 60 ul	2 x 200 ul, 2 x 600 ul
4	1 ug	2 x 10 ul, 2 x 50 ul	2 x 100 ul, 2 x 500 ul

25% of double transformants received to ura plasmids, 25% received two leu plasmids and 50% received one of each, which are the only ones that can be counted on the -ura/leu plates.

Next, the saturated cultures of selection 1 were measured and the number of cells required to inoculate 200 mL of selection 2 in SC-ura-ade-met at OD600 of 0.1 was harvested. The cultures were inoculated and grown for ca. 12 hours until an OD of 1.2 was reached. Since there was no selection for the maintenance of the -leu plasmid alone and it was expected to only observe pRS415 plasmids in double transformants together with the pRS416-derived plasmid, the second round of plating was only done on SC-ura and SC-ura-leu according to Table 3. The cells were again incubated at 30 °C for 2 days after which the colonies were counted and the total number of transformants was calculated.

For colony numbers see Supplementary Table 1.

The percentage of double transformants was now calculated using eq. 2.

$$\frac{2 \times \text{No. colonies on SC - ura - leu}}{\text{No. colonies on SC - ura} - \text{No. colonies on SC - ura - leu}} \times 100\% \tag{2}$$

The averages of replicates and their standard deviations were calculated.

Competition assay

All deepPCAs were started with selection 2: OD600 of the saturated cultures of selection 1 was measured and the number of cells required to inoculate 100 mL of SC-ura-ade-met at OD600 0.1 was harvested and re-suspended in the prepared medium. For the screen comparing the different harvesting timepoints that was performed with a larger library, 2 L of selection 2 were inoculated. The cells were grown at 30 °C and 200 rpm for ca. 12 hours until they reached an OD600 of 1.2. According to the condition tested, the competition assays were then performed slightly differently.

Transformation with different amounts of DNA and different library compositions

100 mL of competition medium were inoculated at OD600 0.05 and grown at 30 °C and 200 rpm until they reached an OD600 of about 1.6. The remaining cells from selection 2 were harvested at RT or 4°C for 5 min at 3,000 OD, the pellet was washed twice with sterile H2O and then frozen at -20 °C. When the competition culture had reached the desired OD600, the cells were harvested in the same way and frozen.

lag phase vs. no lag phase

For the + lag phase sample, 100 mL of competition medium were inoculated at OD600 0.05 and grown at 30 °C and 200 rpm until they reached an OD600 of about 1.6. The remaining cells from selection 2 were harvested at RT or 4°C for 5 min at 3,000 OD, the pellet was washed twice with sterile H₂O and then frozen at -20 °C. For the - lag phase sample, the entire selection 2 culture was used to inoculate 100 mL of competition

medium and was incubated at 30 °C and 200 rpm for 2.5 hours. From this pre-competition culture, another 100 mL of competition medium were inoculated at OD600 0.05 and grown at 30 °C and 200 rpm until they reached an OD600 of about 1.6. The remaining pre-competition culture was harvested as described above. All competition cultures that had reached the desired OD600 were harvested in the same way.

Different MTX concentration/brands/DMSO

The assay was performed as described in “Transformation with different amounts of DNA” above, but the composition of competition medium was slightly adjusted depending on the sample (Table 4). To make sure that the same amount of DMSO would be added to each sample

Table 3 Plating scheme for counting of double transformants after selection 2

Sample no.	DNA amount	Plating SC - URA	Plating on SC -URA/-LEU
1	250 ng	2 x 2 ul, 2 x 5 ul of 1/100 X diluted culture	2 x 200 ul of undiluted culture, 2 x 20 ul of undiluted culture, 2 x 2 ul of undiluted culture
2	500 ng	2 x 2 ul, 2 x 5 ul of 1/100 X diluted culture	2 x 200 ul of undiluted culture, 2 x 20 ul of undiluted culture, 2 x 2 ul of undiluted culture
3	750 ng	2 x 2 ul, 2 x 5 ul of 1/100 X diluted culture	2 x 200 ul of undiluted culture, 2 x 20 ul of undiluted culture, 2 x 2 ul of undiluted culture
4	1 ug	2 x 2 ul, 2 x 5 ul of 1/100 X diluted culture	2 x 200 ul of undiluted culture, 2 x 20 ul of undiluted culture, 2 x 2 ul of undiluted culture

Table 4 Concentration and brand of MTX used for different conditions in the deepCPA

Sample Name	Final concentration / brand MTX
DMSO	0 µg/mL
1/20 X MTX	10 µg/mL, BioShop Canada Inc., Canada
1/4 X MTX	50 µg/mL, BioShop Canada Inc., Canada
1/2 X MTX	100 µg/mL, BioShop Canada Inc., Canada
1 X MTX	200 µg/mL, BioShop Canada Inc., Canada
2 X MTX	400 µg/mL, BioShop Canada Inc., Canada
1 X MTX Sigma	200 µg/mL, Sigma-Aldrich, Merck, Darmstadt, Germany

(2 mL), different stock concentrations of DMSO had been prepared beforehand.

Different generation times

To harvest cells at low generation times and still get a sufficiently high number of cells for the subsequent DNA extraction and sequencing library preparation, the competition cultures were inoculated at different OD600s as listed in Table 5.

2 L of competition medium were inoculated at the indicated OD600s from selection 2 culture and grown until they reached the desired OD600. They were then harvested as described above.

DNA extraction

We recently introduced a novel technique to extract plasmid DNA from yeast cells that gives us about 10-times higher plasmid enrichment than the previous method as described in Diss & Lehner, 2018.

DNA from input and output samples was extracted by first spheroblasting the yeast cells followed by plasmid Mini- or Midiprep according to the culture size using Qiagen kits (QIAGEN, Hilden, Germany). To spheroblast the cells, pellets were thawed and incubated in 4 mL per 100 mL culture 0.1 M EDTA-KOH pH7.5, 10 mM DTT at 30 °C and 180 rpm shaking for 15 min. The cells were harvested at RT and 2,500 g for 5 min

Table 5 Inoculation and harvest OD600 for different numbers of generations

Generations to grow until harvest	Inoculation OD600	OD600 at harvest
2	0.2	0.8
3	0.1	0.8
4	0.05	0.8
4.5	0.05	1.2
5	0.05	1.6

and re-suspended in 4 mL per 100 mL culture of 0 mM K-phosphate pH7.2, 1.2 M sorbitol, 0.4 mg/ml Zymolyase 20T (amsbio, USbiological) and 100 µg/mL RNase A. Cells were incubated at 30 °C and 180 rpm shaking until spheroblasting was complete after approximately 2 hours. Spheroblasts were collected at RT and 2,500 g for 5 min and then re-suspended in 1.6 mL per 100 mL culture homemade buffer P1 (according to the manufacturers protocol, QIAGEN, Hilden, Germany). 1.6 mL per 100 mL culture of homemade buffer P2 (according to the manufacturers protocol, QIAGEN, Hilden, Germany) was added. The samples were mixed well by inversion. Finally, 1.6 mL per 100 mL culture of pre-cooled, commercial buffer P3 (QIAGEN, Hilden, Germany) were added and the samples were again mixed well.

Miniprep for cultures of 100 mL

The mixtures were transferred to 5 mL reaction tubes (Eppendorf, Hamburg, Germany) and spun at max. speed for 20 min in a tabletop centrifuge. The supernatant was recovered and plasmid DNA was purified following the standard QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany) protocol and eluted in 50 µl EB buffer.

Midiprep for cultures of 2L

The larger cultures were purified using a slightly adjusted Miniprep protocol as described in [11].

To quantify the amount of plasmid DNA in the extracts, qPCRs were performed using primers specific for the plasmid backbones (oGD241, oGD242). A plasmid of similar size to the library and known concentration was used to make a standard curve. The plasmid was pre-diluted to 0.4 ng/ul and a dilution series of 6 sequential 1/5 X dilution steps was performed in triplicate. qPCR was performed with the X SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA) according to the manufacturers protocol.

Sequencing library preparation

A single round of PCR was performed using NEBNext® Multiplex Oligos for Illumina® (New England Biolabs, Ipswich, MA) that anneal to the Illumina primer binding sites already present in the barcode region of the plasmid. Depending on the sequencing run type and the expected reads per sample, an at least 10-fold excess of plasmid molecules over expected reads was used as template for the PCRs. The libraries were amplified with Q5 polymerase (New England Biolabs, Ipswich, MA). 50 µl reactions were set up with the required number of plasmid molecules, 1 X Q5 reaction buffer, 0.25 µM primer mix, 200 µM dNTPs, and 0.5 µl Q5 polymerase. The PCR was run between 14 and 20 cycles, depending on the PCR efficiency of the individual sample, with an annealing

temperature of 63 °C and an extension at 72 °C for 30 seconds. Amplicons were gel-purified from a 1 % Agarose gel using the QIAquick Gel Extraction Kit (QIAGEN, Düren, Germany). After determining the concentration with Qubit (Invitrogen, Waltham, MA, USA), samples were pooled at equimolar ratios and the pool was purified once more using AMPure XP magnetic beads (Beckman Coulter, Indianapolis, IN) according to the manufacturers protocol using a 1:1 ratio of beads and sample. As a final quality control, qPCR was performed on the pool to confirm the concentration measured with Qubit using the KAPA Library Quantification Standards with Primer (Roche Sequencing Solutions Inc, Pleasanton, CA), and the fragment distributions were checked on a Bioanalyzer (Agilent Technologies, Santa Clara, CA). The final library was submitted for sequencing on an Illumina sequencer (different sequencers were used, namely NovaSeq, HiSeq, NextSet, and MiSeq; all from Illumina, San Diego, CA). 50 bp single-end sequencing was performed in all cases to cover the barcode region. To increase sequencing, PhiX phage library was spiked into the library at varying amounts between 10 and 25 %.

Data analysis

Raw read counts were processed and transformed into count tables using the mutscan R package [27]. Samples with an average Phred score below 20 and incorrect amplicon structure were discarded at this step. Afterwards, the sequences were matched with the barcode table and only perfect matches were kept. All barcodes belonging to the same variant were summed up and finally, all rows containing less than 10 reads in any input or less than 1 in any output were discarded. For each pair, growth rates were calculated using eq. 3.

$$\left(\log_2 \left(\frac{\text{countsOut} / \text{sum}(\text{countsOut})}{\text{countsIn} / \text{sum}(\text{countsIn})} \right) \right) + \log_2 \left(\frac{\text{ODOut}}{\text{ODIn}} \right) / \text{growth time}[\text{hours}] \quad (3)$$

Modes of the growth rates were calculated using a custom function in R that takes the y values of a density as input and identifies modes by comparing each element to its neighbors. If an element is greater than both its preceding and succeeding elements, it is considered a mode.

Figures were created in R Studio and Adobe Illustrator. All counts and average growth rates can be found in Supplementary Tables 2 to 10 as indicated in the main text.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10524-7>.

Supplementary Material 1.

Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.
Supplementary Material 13.
Supplementary Material 14.
Supplementary Material 15.
Supplementary Material 16.

Code availability

All custom scripts will be provided upon requests

Authors' contributions

AMB and GD conceptualized the project. KS and AMB performed all the experiments with support from DK, KS and KKG. AMB analyzed the data. AMB and GD interpreted the results. AMB and GD wrote the manuscript.

Funding

This work was supported by the Novartis Research Foundation and SNF Project grant 197593.

Availability of data and materials

Deep sequencing data is available at GEO with accession number GSE245485.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare they have no conflict of interest.

Received: 14 November 2023 Accepted: 14 June 2024

Published online: 24 June 2024

References

1. Fowler DM, Fields S. Deep mutational scanning: A new style of protein science. *Nat Methods*. 2014;11:801–7.
2. Fowler DM, Stephany JJ, Fields S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat Protoc*. 2014;9:2267–84.

3. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, et al. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell*. 2015;161:647–60.
4. Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, et al. Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics*. 2015;200:413–22.
5. Diss G, Lehner B. The genetic landscape of a physical interaction. *Elife*. 2018;7:1–31.
6. Faure AJ, Domingo J, Schmiedel JM, Hidalgo-Carcedo C, Diss G, Lehner B. Mapping the energetic and allosteric landscapes of protein binding domains. *Nature*. 2022;604:175–83.
7. Moesslacher CS, Woodsmith J, Auernig E, Feichtner A, Jany-Luig E, Jehle S, et al. Missense variant interaction scanning reveals a critical role of the FERM-F3 domain for tumor suppressor protein NF2 conformation and function. *Life Sci Alliance*. 2023;6:1–16.
8. Heyne M, Shirian J, Cohen I, Peleg Y, Radisky ES, Papo N, et al. Climbing up and down Binding Landscapes through Deep Mutational Scanning of Three Homologous Protein-Protein Complexes. *J Am Chem Soc*. 2021;143:17261–75.
9. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. 2020;182:1295–1310.e20.
10. Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, et al. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst*. 2010;6:1782–90.
11. Bendel A, Faure A, Klein D, Shimada K, Kempf G, Cavadini S, et al. The genetic architecture of protein interaction affinity and specificity. *bioRxiv*. 2023;2023.10.17.562688.
12. Pelletier JN, Campbell-Valois F-X, Michnick SW. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *PNAS*. 1998;95:12141–6.
13. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, et al. An in vivo map of the yeast protein interactome. *Science*. 2008;320:1465–70.
14. Remy I, Michnick S. Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *PNAS*. 1999;96:5394–9.
15. Tareen A, Kooshkbaghi M, Posfai A, Ireland WT, McCandlish DM, Kinney JB. MAVE-NN: learning genotype-phenotype maps from multiplex assays of variant effect. *Genome Biol*. 2022;23:98.
16. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol*. 2009;19:596–604.
17. Otwinowski J. Biophysical inference of epistasis and the effects of mutations on protein stability and function. *Mol Biol Evol*. 2018;35:2345–54.
18. Fraser HB, Plotkin JB. Using protein complexes to predict phenotypic effects of gene mutation. *Genome Biol* 8, R252 (2007). <https://doi.org/10.1186/gb-2007-8-11-r252>.
19. Kinney JB, Murugan A, Callan CG, Cox EC. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *PNAS*. 2010;107:9158–63.
20. Weng C, Faure AJ, Lehner B. The energetic and allosteric landscape for KRAS inhibition. *bioRxiv*. 2022;2022.12.06.519122.
21. Wei H, Li X. Deep mutational scanning: A versatile tool in systematically mapping genotypes to phenotypes. *Front Genet*. 2023;14:1–9.
22. Zhang Z, Moo-Young M, Christi Y. Plasmid Stability in Recombinant *Saccharomyces cerevisiae*. *Biotechnol Adv*. 1996;14:401–35.
23. Diss G, Gagnon-Arsenault I, Dion-Coté A-M, Vignaud H, Ascencio DI, Berger CM, et al. Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science*. 2017;355:630–4.
24. Faure AJ, Schmiedel JM, Baeza-Centurion P, Lehner B. DiMSum: An error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. *Genome Biol*. 2020;21:1–23.
25. Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, et al. A statistical framework for analyzing deep mutational scanning data. *Genome Biol*. 2017;18:1–15.
26. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*. 1989;122:19–27.
27. Sonesson C, Bendel AM, Diss G, Stadler MB. mutscan - a flexible R package for efficient end-to-end analysis of multiplexed assays of variant effect data. *Genome Biol*. 2023;24:1–22.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.