Developing a massively parallel yeast two-hybrid assay to characterize thousands to millions of pairwise protein-protein interactions in a single pot

Alexandr Baryshev

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Georg Seelig, Chair

Eric Klavins

Stanley Fields

Program Authorized to Offer Degree:

Electrical & Computer Engineering

University of Washington

Abstract

Developing a massively parallel yeast two-hybrid assay to characterize thousands to millions of pairwise protein-protein interactions in a single pot

Alexandr Baryshev

Chair of the Supervisory Committee:

Georg Seelig

Department of Electrical & Computer Engineering

The interaction between proteins drives nearly all important processes inside living cells. The mechanisms underlying protein-protein interactions remain not fully understood, which necessitates experimental verification of functional de novo designed binders for the purposes of synthetic biology and therapeutics. This, in turn, necessitates a technology to measure lots of protein-protein interactions in parallel. While existing massively parallel approaches have already been implemented to generate interactome-size datasets, they are still either labor intensive or too sophisticated and expensive to be implemented by most labs. We have developed and extensively characterized an easy-to-use high-throughput method to measure binary protein-protein interactions (PPIs). In our approach, we can measure all pairwise

interactions between a library of proteins in a single one-pot experiment. We make use of the conventional yeast two-hybrid assay's genetic selection principle in which the expression level of a growth-essential gene is determined by the strength of interaction between two proteins of interest; one protein is fused to the DNA binding domain and the other one is fused to the transcriptional activation domain of a split transcription factor driving the expression of growth-essential gene and thus a growth selection can be used to differentiate stronger interacting protein pairs from weaker ones. We construct a population of yeast cells where in each cell only one distinct pair of proteins from a protein library is expressed in the form of fusions to the binding and activation domains. This is achieved by expressing both fusion proteins from the same yeast centromere plasmid maintained at a single copy inside each yeast cell. This cell population is then subjected to the growth selection to differentiate cells with stronger interacting proteins from the weaker ones. By integrating unique DNA barcodes into distinct plasmids and comparing the abundance of each distinct barcode before and after the selection process by means of next-generation sequencing, a fitness score for every protein pair is calculated and used as a proxy for the interaction strength. We show that measured fitness scores exhibit log-linear correlation with independently measured dissociation constants reported in literature, ranging from R^2=0.55 for a set pro-survival BCL2 protein family and their de novo designed inhibitors to R^2=0.95 for a set of parallel heterodimeric alpha-helical coiled coils with variable length.

Table of Contents

Chapter 1. Literature review

Uncovering interactomes to improve understanding of underlying mechanisms of protein-protein interactions (PPIs) and advancing the design of de novo protein-protein interaction requires a method to detect and characterize PPIs in parallel.

Until the advent of next-generation sequencing (NGS), high-throughput characterization of PPIs relied on in-vitro mass spectrometry (Richards et al., 2021) and array based methods (Walhout & Vidal, 2001) which are often expensive, hard to use or require expensive and laborious proteins purification step.

Surface-based in vitro methods like yeast- or phage-display have embraced next-generation sequencing technologies to increase throughput but are limited to "several-versus-many" types of screening due to limiting resolution of existing fluorescent reporters and require purification of target proteins. Another high-throughput surface-based method that enabled "library-against-library" screening is AlphaSeq (Younger et al., 2017) a proprietary technology that is not readily available to any lab. The authors mention that the approach is ill-suited for screening of homodimer libraries. Proteins that might bear a nuclear localization signal might be poorly expressed on the yeast surface when using AlphaSeq.

In vivo two-hybrid based systems have been a powerful tool in discovering novel interactions in the last 30 years. The idea of the method is to fuse a pair of proteins of interest to the DNA binding- and transcriptional activation domains of a split transcription factor to reconstitute its transcriptional activity that is driving some reporter, usually a growth essential enzyme (Fields & Song, 1989). While there has been a number of high-throughput yeast and bacterial two-hybrid

methods, such as eY2H, Smart-pooling Y2H, Y2H-Seq, Stitch-Seq Y2H, BFG-Y2H, CrY2H-seq, RLL-Y2H, PPISeq, developed recently, all of them have some limitations that leave room for further improvement in terms of cost, labor and simplicity.

Earlier developed methods such as eY2H, Smart-pooling Y2H, Y2H-Seq, Stitch-Seq Y2H, BFG-Y2H, while expanding protein interaction screening capabilities, suffer from being expensive and labor intensive. For example, the most advanced of the above methods, BFG-Y2H (Yachie et al., 2016), requires individual treatment of positive colonies followed by PCR of each successful colony. Moreover, every protein requires individual barcoding which can get costly when the library size is big.

While CrY2H-seq (Trigg et al., 2017) is also a cre-recombination based high-throughput version of classical Y2H, unlike BFG-Y2H it doesn't barcode proteins. Instead, it uses recombination to integrate bait and prey cassettes into one plasmid. Fragment containing sequences of both proteins is then amplified in a PCR reaction, tagmented and sequenced using Illumina sequencing platforms. Due to tagmentation only 2.4% of all reads contain the identity of both proteins, which means that very deep sequencing is needed for large libraries.

RLL-Y2H (Yang et al., 2018) is similar to CrY2H-seq and uses recombination to integrate bait and prey cassettes into one plasmid followed by amplification of the fragment containing proteins' sequences via PCR. But instead of using tagmention, like in CrY2H-seq, RLL-Y2H uses type I restriction enzymes to cut the amplified fragment and produce an approximately 110nt long fragment that contains 20nt of 3' sequence of each protein thus allowing it to maintain interaction identity. Only this short fragment is sequenced afterwards using high-throughput sequencing avoiding the need for a deep sequencing inherent to CrY2H. While RLL-Y2H is probably the easiest method to use for most libraries, it comes short for certain types of libraries, in particular,

libraries of de novo designed synthetic proteins whose amino acid sequence differs in just a few positions (Younger et al., 2017) making it impossible to reliably identify proteins since only 20nt of the 3' end is sequenced.

AVA_Seq (Andrews et al., 2019) is a bacterial two-hybrid method in which both hybrid proteins are expressed from the same plasmid. To achieve the uptake of only one plasmid per cell one has to limit the amount of transformed plasmids, which might limit the maximal size of screened libraries.

In (Johnson et al., 2021) the transcriptome of input cells is converted into a library of RNA-barcoded proteins using a modified mRNA-display technology called SMART-display. In particular, mRNAs of input cells are first reverse transcribed to create single stranded complementary DNA strands (cDNA). Then, in a template switching reaction cDNA strands are turned into double stranded DNA (dsDNA) and a template switching oligo is added to the newly synthesized strands. In a subsequent polymerase chain reaction (PCR) a T7 promoter and a ribosome binding sequence (RBS) are added to the above dsDNA library. The formation of the dsDNA library is completed by adding an amino acid analog puromycin at the 3' end of the dsDNA strands. The resulting dsDNA library is subjected to in vitro transcription to create an intermediate mRNA library which is then translated into a library of proteins. During translation, when ribosome's site A encounters puromycin it prematurely terminates the translation so that protein's mRNA is left attached to the protein thus creating an RNA-barcoded protein. Since PROPER-seq is an in vitro technique it might not capture interactions between proteins that rely on post translational modifications or in vivo localization. Moreover, out of more than 200.000 reported previously uncharacterized interactions, only a handful of them were validated using separate techniques thus implying that other independent massively parallel techniques are welcome in the

field to cross-validate each other. Authors also mention that they have not investigated whether DNA-tags of proteins interfere with any protein-protein interactions. Additionally, they have not investigated how prone the method is to false positive interactions resulting from simple abundance of some proteins or protein-DNA interactions.

In recent years a significant progress has been made in using machine learning, and deep learning in particular, to predict proteins' structures resulting in algorithms like AlphaFold (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021). AlphaFold has recently been reported to predict the shape of nearly all of 200 million known proteins (Callaway, 2022). Despite these advances, these algorithms are not yet able to predict the effects of missense mutations on the three-dimensional structure of proteins, let alone on the protein-protein interactions (Buel & Walters, 2022). These algorithms could learn from vast amounts of experimental data characterizing the effects of missense mutations on protein-protein interactions if such data existed. In one of the first attempts to generate an experimental data set exploring the effect of missense mutation on PPIs was attempted in (Diss & Lehner, 2018) by developing a high-throughput NGS-based protein interaction assay called deepPCA to quantify how more than 120,000 pairs of point mutations affect the binding between the FOS and JUN proto-oncoproteins that forms the AP-1 transcription factor complex. The method is based on the two-hybrid principle called protein fragment complementation assay (PCA) in which two proteins of interest are fused to complementary halves of a methotrexate-resistant variant of murine dihydrofolate reductase and expressed in yeast to reconstitute the activity of the enzyme allowing growth in the presence of methotrexate. The method was further improved in (Faure et al., 2022) by using two separate selection assays that are both based on the PCA to globally map the energetic and allosteric landscapes of SH3 and PDZ protein interaction domains.

Chapter 2. Introduction to conventional yeast two-hybrid and our approach to scale it up.

Basics of the conventional yeast two-hybrid approach

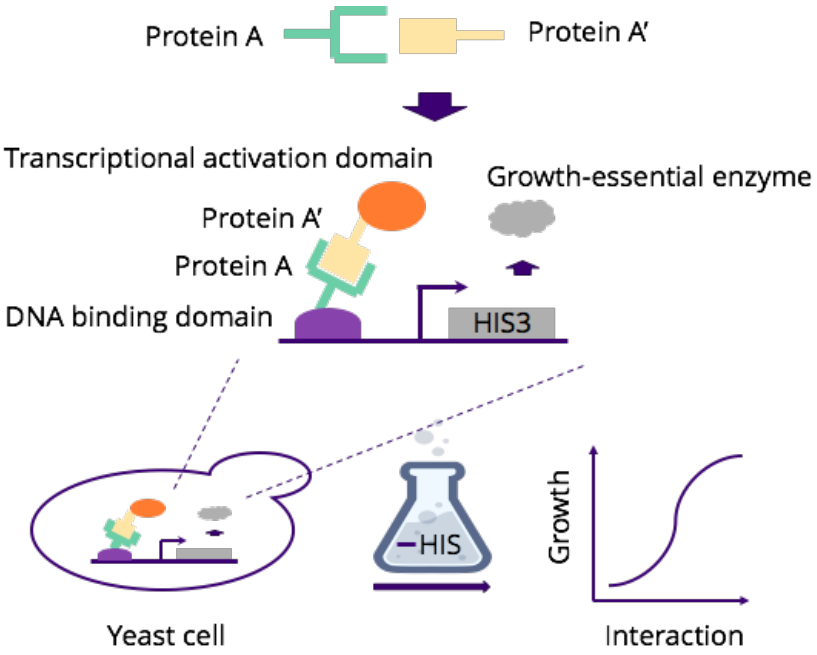

Figure 1. Conventional yeast two-hybrid. To screen a pair of proteins A and A' for interaction, they are fused to DNA binding and transcriptional activation domain of a split transcription factor (TF). The reconstituted transcriptional activity of the TF drives the expression of a growth essential enzyme HIS3. When yeast cells containing the circuit are grown in a growth media lacking histidine amino acid, the growth curve can be served as a proxy for interaction.

The yeast two-hybrid method (Y2H) is used as a main tool for detecting protein-protein interactions in this project. The Y2H method was invented almost 30 years ago (Fields & Song, 1989), but it's still extensively used to discover protein-protein interactions. The idea of the Y2H method is based on reconstituting the transcriptional activity of a split transcription factor (TF), by fusing a pair of interacting proteins A and A' to the DNA binding domain (BD) and transcriptional activation domain (AD) of the transcription factor. Thus formed fusion proteins

$BD - A$ and $AD - A'$, termed hybrid proteins, can physically bind each other and activate the expression of a target gene that is driven by a promoter that has an upstream activation sites (UAS) to which the binding domain binds (Figure 1).

When an enzyme essential for growth is used as the target protein, e.g. HIS3 enzyme, the interaction between a random pair of proteins A and A' can be detected by growing yeast cells, expressing both hybrid proteins $BD - A$ and $AD - A'$, in a growth selective media lacking an amino acid catalyzed the enzyme (histidine amino acid in the case when HIS3 gene is used as the target gene). The interaction between proteins A and A' results in the expression of the growth essential enzyme which catalyzes the production of the lacking amino acid. The stronger the interaction the more amino acid is produced which results in the faster cell culture growth (i.e. faster increase of the optical density of the cell culture). Thus, the maximal slope of the growth curve can be used as a proxy for the interaction strength (Figure 1).

Scaling up

Our approach to make the Y2H method high-throughput is to pair up the use of the low-copy centromeric plasmid (CEN) and the Next-generation sequencing (NGS). This will allow us to measure the pairwise interactions of a library of hybrid proteins in a massively parallel way. We encode each possible combinatorial pair of hybrid proteins $BD - A_i$ and $AD - A'_j$, $i = \underline{1, N}$, $j = \underline{1, M}$, on the same CEN plasmid. We assume a library-against-library screening where we have N proteins A_i, $i = \underline{1, N}$, in the first library, and M proteins A'_j, $j = \underline{1, M}$, in the second library. One possible configuration of promoters expressing both hybrid proteins is depicted in Figure 2. Since CEN plasmid is a single-copy plasmid, typically found at single copy in cells, only one distinct pair of proteins $BD - A_i$ and $AD - A'_j$ can interact in each cell.

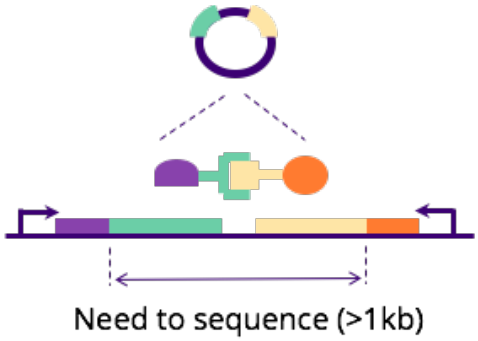

Figure 2. One possible orientation of promoters, or expression cassettes, on the centromeric plasmid to express both hybrid proteins from the same plasmid.

The library of CEN plasmids carrying all possible combinatorial pairs of hybrid proteins can then either be assembled directly in yeast cells (*Saccharomyces Cerevisiae*) by transforming linear double stranded DNA fragments encoding all the necessary sequences, or can first be assembled in Escherichia Coli (*E.coli*) cells, purified from *E.coli* and then transformed into yeast cells. Upon the transformation yeast cells are grown for at least 36 hours in a synthetic complete media lacking histidine (SC-TRP). Since TRP gene is knocked out from EBY100 yeast strain, the use of a TRP expressing cassette in the CEN-containing plasmid allows to only select those cells that have assembled or incorporated the plasmid after 36 hours of outgrowth. Moreover, although yeast cells can have a relatively high transformation rates, an average of just one CEN-containing plasmid is found after 36 hours of outgrown. After the 36h of outgrowth, the yeast cell culture is divided into two batches. The first batch, dubbed “pre-selection” batch, is subjected to plasmid extraction. The second batch, dubbed “post-selection” batch, is then diluted to approximately 1000 cells per 50uL and subjected to selective growth in a synthetic complete media lacking tryptophan and histidine amino acids for approximately 24-48 hours (depends on a particular kind of proteins being screened) to select for strongly interacting pairs. The selection process in stopped upon reaching mid-log phase, or approximately OD600=1.0, and the cells from this batch are then subjected to plasmid extraction. The pre-selection and post-selection

plasmid aliquots are then sequenced to determine relative population frequency of each distinct plasmid. Then, the post-selection frequency of every distinct plasmid is divided by the pre-selection one to obtain the enrichment fold change of the relative frequency, dubbed “HT-Y2H enrichment score”. A heatmap of enrichment score for all possible pairs in the library can then be plotted. The summary of the approach is shown in Figure 3.

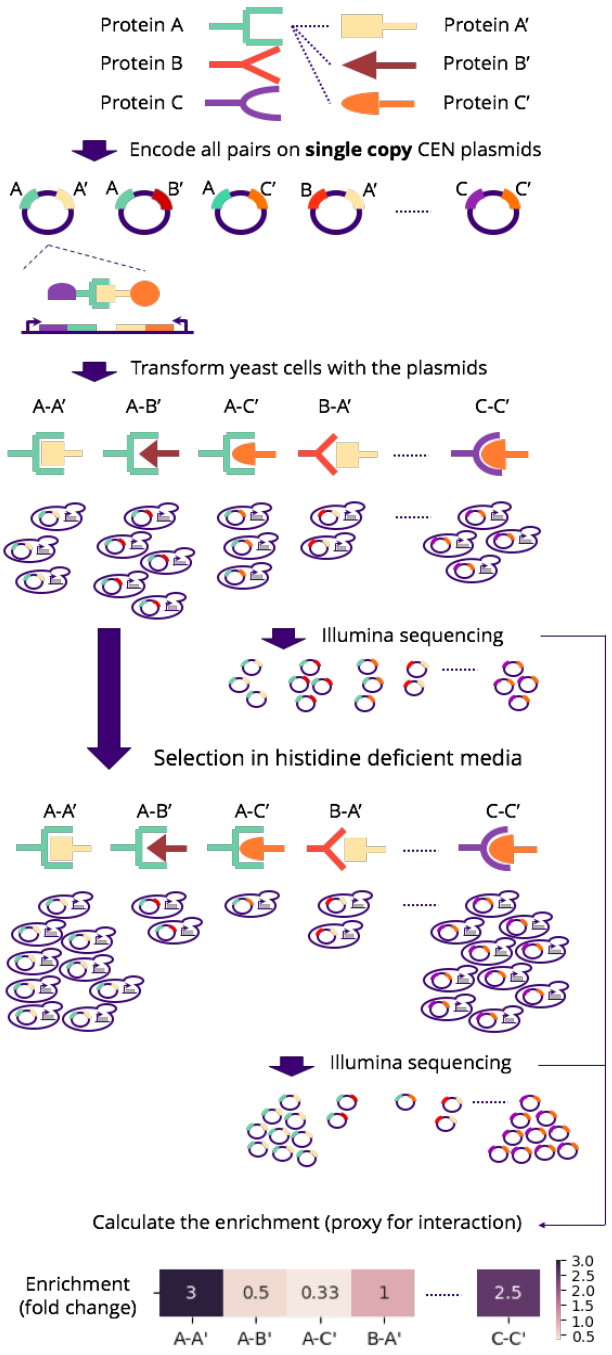

Figure 3. Scaling up of the Y2H method. All possible pairwise combinatorial combinations of hybrid proteins are expressed from the same single copy centromeric (CEN) plasmid so that only one specific pair of proteins can be expressed and can interact inside a single yeast cell. The fold change of relative population frequency of every distinct plasmid before and after the selection in histidine deficient media serves as a proxy for interaction.

Testing cassettes orientations

We hypothesize that for the Y2H to work, the expression levels of both hybrid proteins should be at around the same level. So, we decided to investigate the impact of promoter orientation and layout on protein production. To that end we created a set of plasmids where two fluorescent proteins (GFP and mCherry) were put under control of different promoters and arranged differently. Specifically, we created five different plasmids using a total of five different promoters (Figure 4). We tested arranging promoters back to back (Figure 4a), facing in the same direction (Figure 4b,c,d) and opposing one another (Figure 4Figure 5e). We then performed flow cytometry experiments to quantify expression of both fluorescent proteins for all five panels (Figure 5). Based on these results we chose to orient promoters facing each other in the following experiments. Moreover, we selected pCYC1 (red bar, Figure 5A) to drive GFP and pTEF1 (blue bar, Figure 5A) to drive mCherry. Although we did not test this particular combination initially, our data suggest that these promoters should result in similar expression levels of the two proteins, given that the fluorescence levels are similar when normalized to the control.

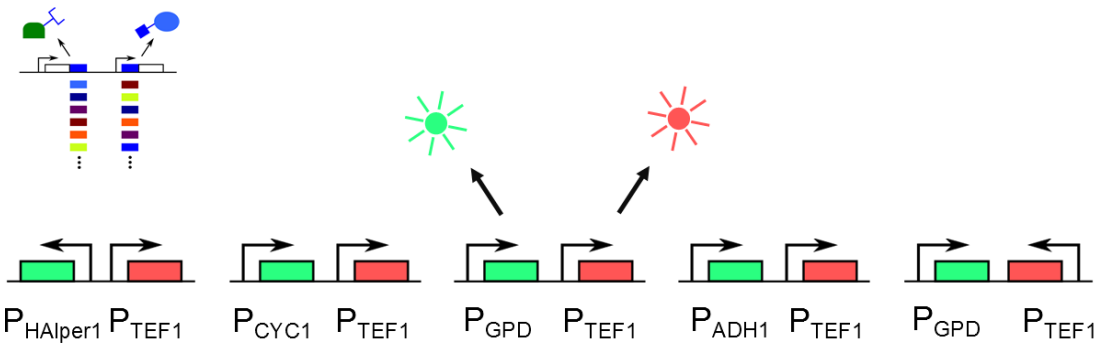

Figure 4. Testing different orientations and expression levels of two expression cassettes located on the same CEN plasmid. Green (GFP) and red (mCherry) fluorescent proteins are used to evaluate expression levels under different promoters in comparison to the base strain that lacks both fluorescent proteins. We hypothesized that the optimal expression conditions will be achieved when the change in fluorescent levels is approximately the same for both red and green channels when compared to the base strain.

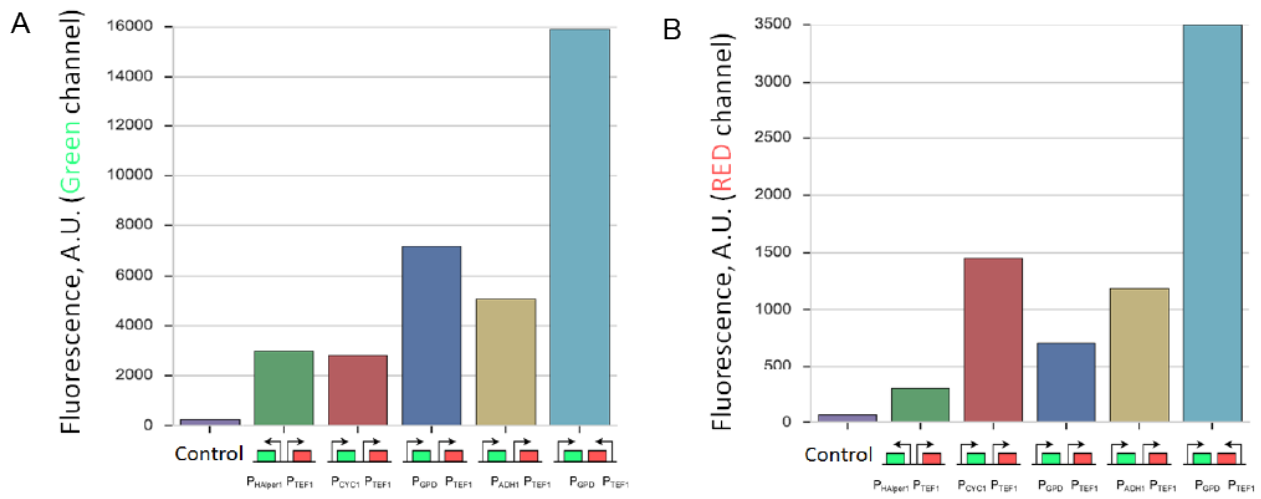

Figure 5. Green (A) and red (B) fluorescent channels for different orientations and combinations of promoters driving the expression of GFP and mCherry encoded on the same CEN plasmid. The left panel shows only the GFP levels for each distinct combination and orientation of promoters on the plasmid. Similarly, the right panel shows only the mCherry levels for those same combinations and orientation of promoters on the plasmid. "Control" corresponds to the strain without any plasmids expressing fluorescent proteins. The ratio of the GFP signal in the red bar (pCYC1 promoter) on the left panel to the GFP signal in the "control" strain is approximately the same as the ratio of mCherry signal in the cyan bar (pTEF1 promoter) on the right panel to the mCherry signal in the "control" strain. Hereby, pCYC1 and pTEF1 promoters facing each other were used to drive the expression of hybrid proteins BD-X_i and AD-Y_j in our HT-Y2H assay.

Selecting binding- and activation- domain for high-throughput yeast two-hybrid

The original yeast two-hybrid method uses the transcription factor (TF) Gal4p as an expression system to modulate the levels of the target fusion gene GAL1-lacZ (Fields & Song, 1989). Gal4p expression system is a nutritional perturbation system that affects cell physiology. GAL1-lacZ gene placed under the control of a promoter containing Gal4p upstream activation sites (UAS_{GAL4}) can be induced in the presence of galactose. Since Gal4p activity is attenuated in glucose, the expression system requires growing yeast cells on an alternative carbon source like raffinose or glycerol. Moreover, the DNA-binding domain of Gal4p has only 6bp of specificity and has many potential off-target sites in the yeast genome.

An expression system with no detectable residual off-target effects and inducible with an inert hormone beta-estradiol that is not affecting cellular physiology has been introduced in (McIsaac et al., 2013). This system uses an artificial transcription factor (ATF) that contains modular and designable Cys2His2 zinc-finger DNA-binding domain of mouse transcription factor Zif268, the human estrogen receptor and VP16 activation domain. Each of the three zinc-fingers comprising the binding domain of the mouse transcription factor Zif268 has a 3bp of specificity resulting in the total of 9bp of specificity for this binding domain.

We thus selected the binding domain of mouse transcription factor Zif268 and VP16 activation domain as the binding- and activation- domains for our high-throughput yeast two-hybrid assay to drive the expression of the growth essential enzyme HIS3.

Selecting transcriptional terminators

To express functional proteins, we also need to include transcriptional terminator sequences into the 3'UTR of the expressed proteins. In the optimization experiment described above, we used tCYC1 terminator for GFP and tSyn27 for mCherry characterized in (Curran et al., 2015). As these choices seemed to work reliably, we decided to move ahead with the same terminators for the HT-Y2H experiment.

Chapter 3. Pilot experiments.

Contributions: De novo protein sequences in L14, L38, L39 experiments were designed by Zibo Chen, Scott Boyken. De novo protein sequences in L43 were designed by Ajasja Ljubetic and Ryan Kibler. High-throughput experiments L14, L33, L39, L42, L43, L44, including data analysis and figures preparation for this chapter were made by Alexandr Baryshev.

Experiment L14: screening a small set of de novo designed homo multimers.

In our pilot experiment we decided to characterize all pairwise interactions in an all-against-all fashion between 10 de novo designed homo-multimers composed of modular arrays of hydrogen-bond containing networks designed by Scott Boyken and Zibo Chen in Professor David Baker's lab at the University of Washington (Boyken et al., 2016). Those interactions has been previously experimentally characterized by Post-Doctoral scholar Benjamin Groves using a low-throughput Y2H assay and thus provide a good reference. The low throughput assay used LexA binding and activation domains which are different from Z3EV and VP16 used in the HT-Y2H described here.

It was previously mentioned in Chapter 2 that plasmids extracted from pre- and post-selection samples are sequenced to calculate HT-Y2H enrichment score but no detailed yet have been provided what part of the plasmids is sequenced. To identify a pair of proteins expressed from the same plasmid it would be logical to sequence the region that contains both proteins' sequences (see Figure 6). Often, the length of this region is greater than 1kb and is not amenable to Illumina NGS technology that can reliably sequence short DNA strands up to 1kb. To avoid this obstacle, a unique random DNA barcode sequence needs to be introduced in each CEN

plasmid, in the middle of the sequence separating two expression cassettes (highlighted by different colors in Figure 6). This way, each possible combinatorial pair of hybrid proteins is uniquely linked to its unique barcode. Enrichment of the barcodes defines the strength of the interaction between distinct pairs of proteins. Pairs of interacting proteins are identified from the sequencing data by their unique barcodes.

After deciding to introduce a random unique DNA barcode on every plasmid, a library of plasmids encoding all possible combinations of the above homo-multimers was assembled directly in yeast cells using homologous recombination. In particular, the transformation mixture contained the following linear fragments:

- A linear backbone fragment bearing binding- and activation- domain sequences along with centromere signal sequence which ensures that only one plasmid is found per cell on average.
- Two sets of linear fragments containing homo-multimers' sequences and corresponding 5' and 3' sequences homologous to the backbone fragment. The first fragment contains binder 1 and is homologous to the end of the backbone containing the AD. The other fragment contains binder 2 and is homologous to the end of the backbone containing the DBD. Homology regions were chosen to be around 30nt.
- Linear fragment containing tCYC1 and tSynth27 terminators with a random 20nt sequence between the terminators that serve as a barcode and will be used in the data processing stage to identify which pair of proteins was expressed in each yeast cell (Figure 6). The fragment is generated by PCR from a plasmid containing the tCYC1 terminator using a primer that contains a random 20nt sequence as well as the tSynth27 short synthetic terminator.

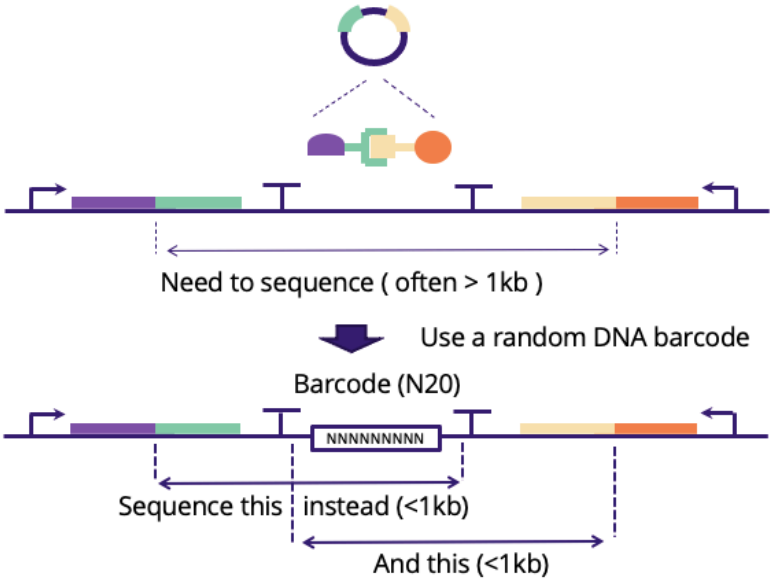

Figure 6. Since the amplicon containing both proteins' sequences is often too long (>1kb) for Illumina sequencing platforms, we can introduce a unique random barcode in the middle between the expression cassettes and sequence two shorter amplicons (<1kb) and keep the initial association using the barcode.

Briefly, we transformed fragments into cells, performed outgrowth of the cell culture for 36 hours in tryptophan deficient media to drop excessive copies of plasmids and ensure that each cell carries just one plasmid. DNA was then extracted from a fraction of the cell culture and prepared for sequencing. Another cell aliquot was subjected to selection in histidine and tryptophan deficient media to select for strongly interacting pairs. After approximately 24 hours of selection, cells were lysed and extracted plasmid DNA was prepared for sequencing as detailed below. To quantify fitness, barcode counts after selection were compared to those before selection.

Results of this experiment are shown in the left panel of Figure 7 and compared to previously published low-throughput Y2H data (the middle panel of Figure 7). The ideal expected interaction outcome is shown in the right most panel of Figure 7. Visual inspection of the experimental interaction heat maps obtained with the high- and low-throughput Y2H

approaches shows reasonable agreement. In particular, good agreement is found in the highlighted regions (upper left corner and diagonal).

Although promising, a major concern with this first experiment was a high rate of barcode collisions wherein a single barcode mapped to multiple protein pairs. We hypothesize that since barcodes were added during a PCR reaction these collisions are an artifact of PCR over-amplification. It is likely that collisions could be avoided if barcodes were instead added through a Klenow extension or directly during assembly in yeast.

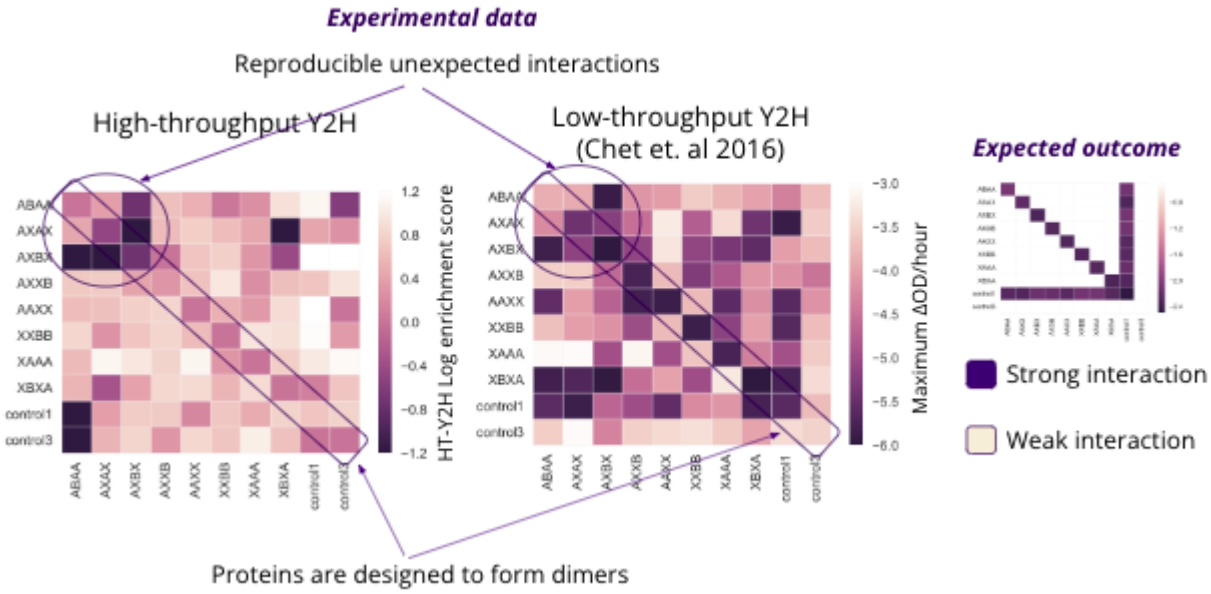

Figure 7. Comparing high- and low-throughput yeast two-hybrid methods. The heatmap of log enrichment scores of the HT-Y2H is plotted on the left panel after discarding colliding barcodes. The heatmap of the maximal change in cell culture optical density is plotted in the middle panel (data published in 2016 Science paper). Visual inspection of the interaction heat maps obtained with the high and low throughput Y2H approaches shows reasonable agreement. In particular, a good agreement is found in the highlighted regions (upper left corner and diagonal). The ideal expected outcome of the interaction between these de novo designed binders is plotted on the right panel. Darker color denotes stronger interaction.

Design update: Each binder is individually barcoded.

To address this concern, we decided to update our design and assembly strategy. Specifically, the single barcode design was abandoned in favor of an alternative approach where

each protein binder was assigned a predefined DNA barcode sequence during DNA synthesis. In each fragment, the barcodes is 3' of the binder and separated from the sequence encoding the binder by a constant sequence used for PCR (dubbed PCR handles H1 and H2, see Figure 8B). While this strategy is bit more costly and less flexible because the barcodes need to be synthesized on the same oligo that also encodes the binder, it ensures that barcode collision cannot occur and it also ensures that all barcodes have a sufficient edit distance.

With this new design plasmid assembly in yeast still requires four fragments as shown in Figure 8A. However, the fragment containing the terminators now has a fixed sequence and no barcode.

Importantly, the sequencing amplicon used to read out the protein pair identity now spans the two barcodes (Figure 8B). PCR with primers using the constant PCR handles H1 and H2 is used to amplify the barcode pairs.

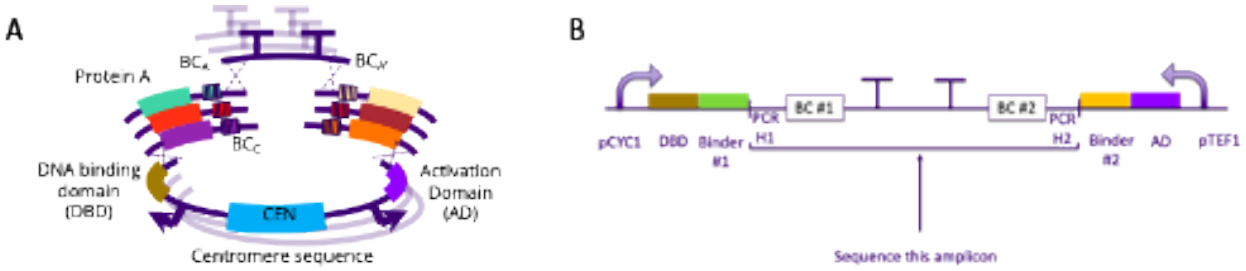

Figure 8. (A) Fragments used in the plasmid assembly. (B) Organization of the assembled plasmid highlighting the sequencing amplicon carrying the barcodes. PCR handles H1 and H2 are used to add Illumina sequencing primer regions, Illumina P5 and P7 flow cell adaptor sequences, and Illumina N7XX and N5XX indexes.

Experiment L33. Testing our HT-Y2H assay on previously characterized set of human pro-survival BCL2 family proteins and their de novo designed inhibitors.

To make our HT-Y2H assay semi-quantitative in the future and allow for quantitative characterization of protein-protein interactions in terms of dissociation constant Kd would

require a benchmark set of proteins which have already been characterized and for which dissociation constants Kd have been measured and published. Unfortunately, the previously used set of de novo designed homo-multimers have been characterized using conventional low-throughput yeast two-hybrid incapable of identifying dissociation constants. Our search for a such benchmark set ended by selecting a set of binders from (Berger et al., 2016). Specifically, in that work a set of de novo designed three-helix bundle protein inhibitors was designed to specifically bind to six human pro-survival BCL2 family proteins. The interactions between all designed binders and their targets were characterized quantitatively using biolayer interferometry (Berger et al., 2016) and later using a proprietary Alpha-seq technology (Younger et al., 2017).

Following (Berger et al., 2016), we selected six BCL2 homologues and nine synthetic binders for them resulting in a total of 15 distinct proteins. We decided to test each of these 15 proteins in either orientation (i.e. fused to the AD or fused to the BD) resulting in a total of 225 possible PPIs. We performed two replicates of this experiment over a month or so and found good agreement (R^2=0.58) with previously published results where applicable. Figure 9 shows interaction heatmaps for our method as well as Alpha-seq and correlation between the two approaches.

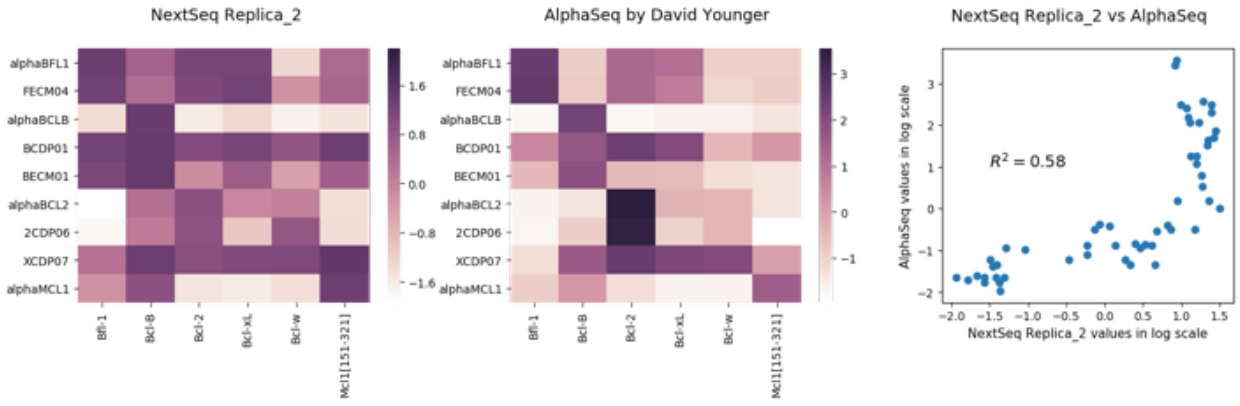

Figure 9. A heatmap of the log enrichment scores of our HT-Y2H assay's Replicate#2 (A) and of negative log values of dissociation constants Kd of Biolayer Interferometry (B) published in [(Berger et al., 2016)]. (C) We found good correlation between our results and the published data.

We also found good agreement between independent replicates (R^2=0.74, Figure 10). We note that one binder-AD pair failed to assemble in the first replicate and was excluded from the correlation analysis.

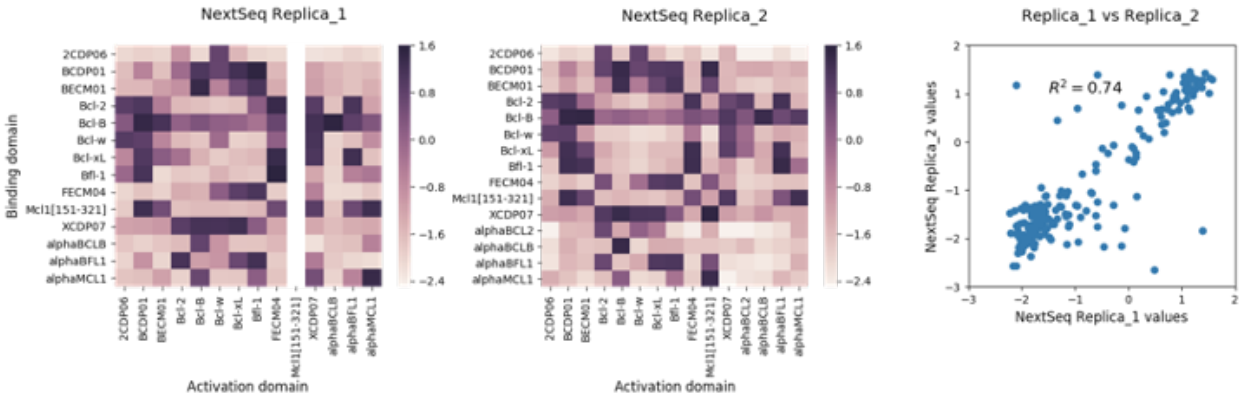

Figure 10. Excluding AD-MCL1[151-321] pairs that failed to assemble in the first replicate of the experiment, we found great correlation between the two replicates of the experiment (R^2=0.74). Log values of enrichments scores were compared.

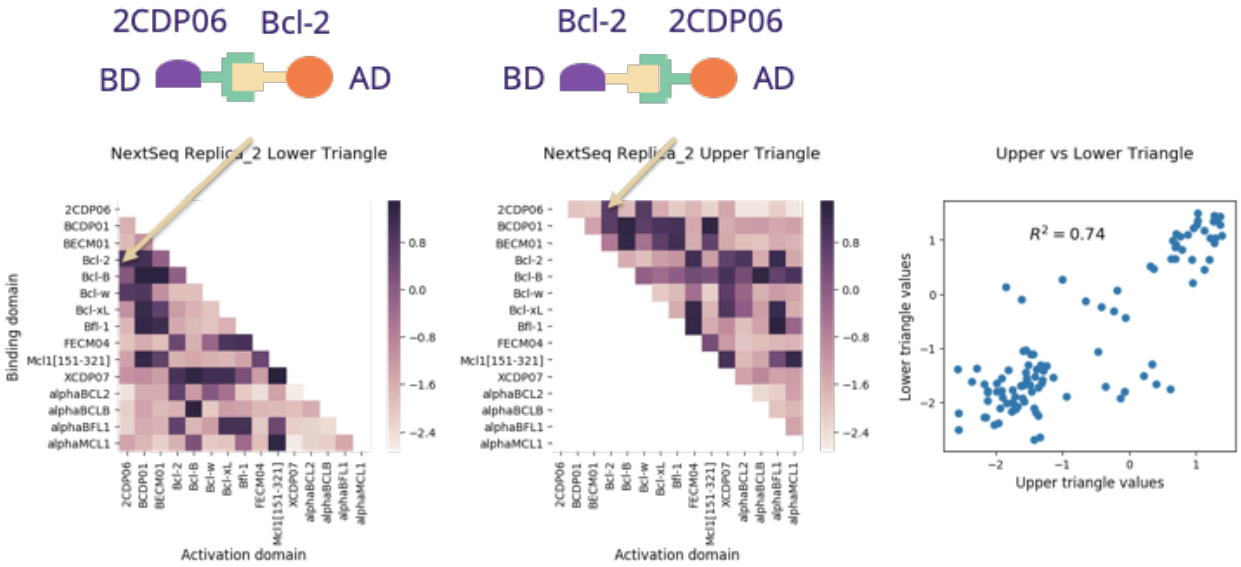

Figure 11. Correlation between log values of enrichment score of the main and reciprocal orientations was found to be quite good (R2=0.74). More specifically, in a all-against-all screening experiment like this one, every pair appears in the data twice since each binder in the pair can be fused to either the AD or DBD. E.g. 2CDP06 and Bcl-2 pair can appear as DBD-2CDP06/AD-Bcl-2 (left lower triangular enrichment score heatmap) or as DBD-Bcl-2/AD-2CDP06 (right upper triangular enrichment score heatmap). Good correlation between these main and reciprocal orientations indicate good consistency of our HT-Y2H assay.

Finally, we also investigated the correlation between orientations (Figure 11). That is, each pair occurs twice in the data because a given protein can be fused either to the AD or BD. Ideally, interactions would not be impacted by the orientation but in practice different protein fusions can result in different expression levels. We find that for this set of interaction domains the measured interactions largely agree between orientations (R^2=0.74).

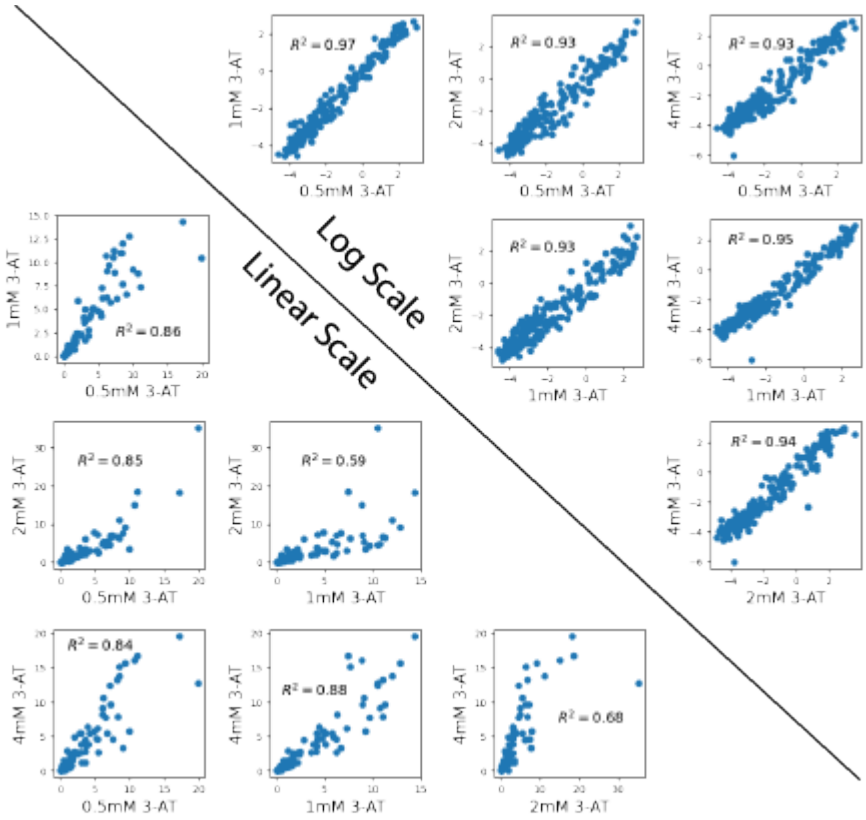

Figure 12. Comparing linear (under the diagonal line) and log (above the diagonal line) values of HT-Y2H enrichment scores of L33 library selected under different (0.5, 1, 2, 4 mM) concentrations of HIS-3 inhibitor 3-amino-1 2 4-triazole also called 3-AT.

To understand whether our results are sensitive to the strengths of the selective pressure we performed the experiment with the BCL binders in the presence of varying amounts of 3-AT (Figure 12). 3-AT binds His3 which increases the selective pressure because more His3 has to be produced to enable growth. We tested 0.5, 1, 2 and 4 mM 3-AT. Overall, we found that for this

range of 3-At concentrations, results were very similar. Still, we found that concentrations around 1-2mM 3-AT resulted in the best correlation with previously published data.

Experiment L38. Characterizing interaction between 52 pairs of de novo hydrogen bond network containing heterodimers.

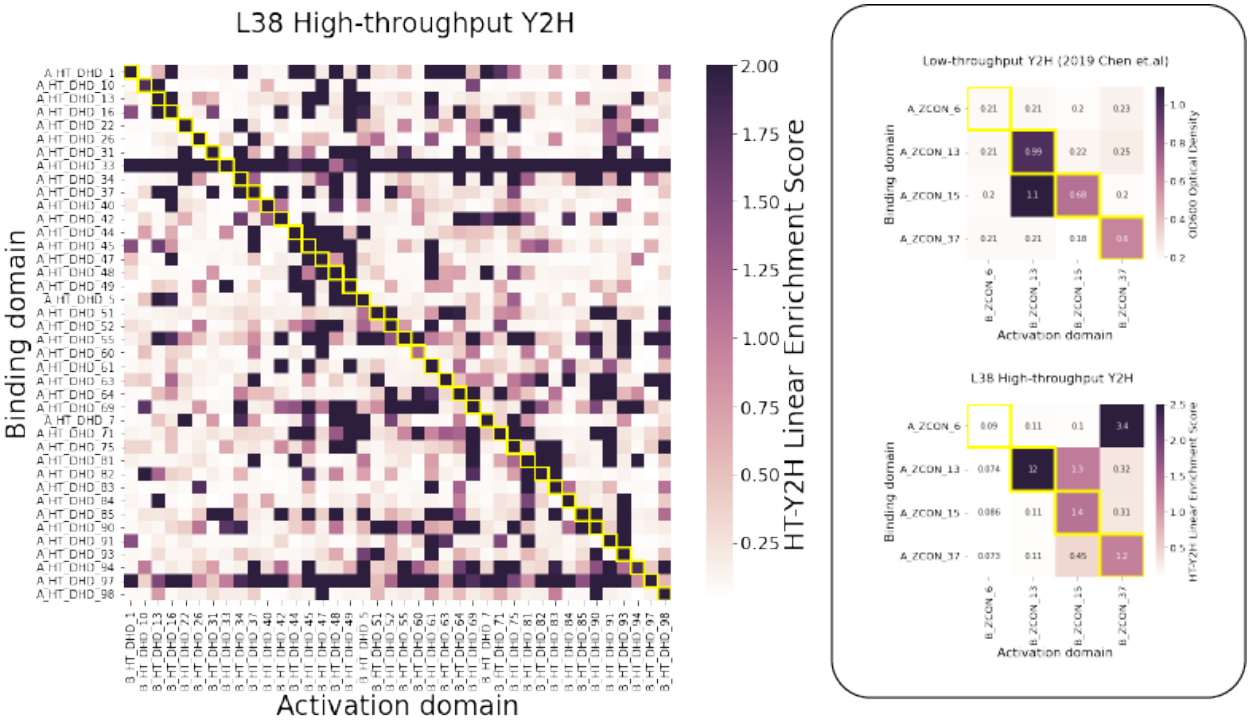

Figure 13. (A) HT-Y2H linear enrichment scores for newly designed hydrogen bond network heterodimers. In this experiment, one partner in each pair ("A binder") was fused to the BD and the other partner ("B" binder) was fused to the AD resulting in a total or 104x104=10,816 interactions tested in a single experiment. Only 40 out 100 newly designed pairs had linear enrichment score >1 and were deemed successful on-target designs. Those 40 pairs are highlighted by yellow squares in the heatmap (so called "diagonal interactions"). (B) "Control" heterodimers previously characterized in [(Z. Chen et al., 2019)] (upper heatmap) were added to the newly designed hetero-dimers and we found a good visual agreement between our assay's results for this binders and the published data.

After validating our HT-Y2H approach using published reference data, we next turned to characterizing interactions between a large set of de novo designed hydrogen bonding heterodimers. These heterodimers were computationally designed to have strong pairwise interactions, but we also note that the design approach cannot account for off-target interactions.

Therefore, heterodimer interactions are not necessarily orthogonal. As additional experimental controls we thus also added four heterodimer pairs that had been previously characterized and published.

In this experiment, one partner in each pair ("A binder") was fused to the DBD and the other partner ("B" binder) was fused to the AD resulting in a total or 104x104=10,816 interactions tested in a single experiment. It is worth noting that we identified 99.5% of all possible combinatorial pairs in the input sample implying that nearly all possible pairs have assembled during the transformation.

Overall, we found good agreement between the four control pairs in our experiment and previously reported data (Figure 13B). In particular, the results agreed for all on-target interactions (diagonal entries). However, each assay found one strong off-target interaction not observed with the other assay.

Among the newly designed 100 pairs, we detected strong on-target interactions (linear enrichment score >1) for 40 of them (Figure 13A) representing a 40% design success rate. However, we also note a large number of off-target interactions. In the entire pool of 10,000 possible interactions we observe 13% to have an enrichment score >1. One binder, A_HT_DHD_33 fused to the BD, either binds strongly to all possible partners or activates gene expression constitutively through an unknown mechanism.

Experiment L39. Characterizing interaction between 52 pairs of de novo hydrogen bond network containing heterodimers in all-against-all manner.

Having successfully assembled 99.5% of all possible pairs in the previous L38 experiment we decided to increase the library size and test how many pairs we can assemble in a single

experiment. In particular, in this experiment we were still working with 100 newly designed heterodimers characterized in L38 experiment. But unlike in L38 experiment, where one partner in each pair ("A binder") was fused to the BD and the other partner ("B" binder) was fused to the AD resulting in a total or 104x104=10,816 interactions tested in a single experiment, in this experiment all binders were fused to both the AD and BD resulting in an all-against-all screening. The number of possible heterodimers in this experiment is thus 200x200=40,000. Moreover, we added 5 more previously characterized hetero-dimer pairs in addition to the 4 control pairs already used in L38, resulting in 218x218=47,524 possible pairs.

Once again, we were able to identify almost 99.5% of all possible pairs in the assembled library before subjecting it to a selection process. The transformation efficiency was about 1M transformants.

We performed three replicates of this experiment, each separated by about a month in time. Each time the replicate started with a fresh transformation but using the same stocks of transformed fragments.

Out of 40 successful on-target interactions revealed in L38 library 34, 33, and 34 had a linear enrichment score greater than 1 in the replicates #1, #2 and #3 of L39 library, which amounts to 85, 82.5 and 85% implying high degree of reproducibility (Figure 14).

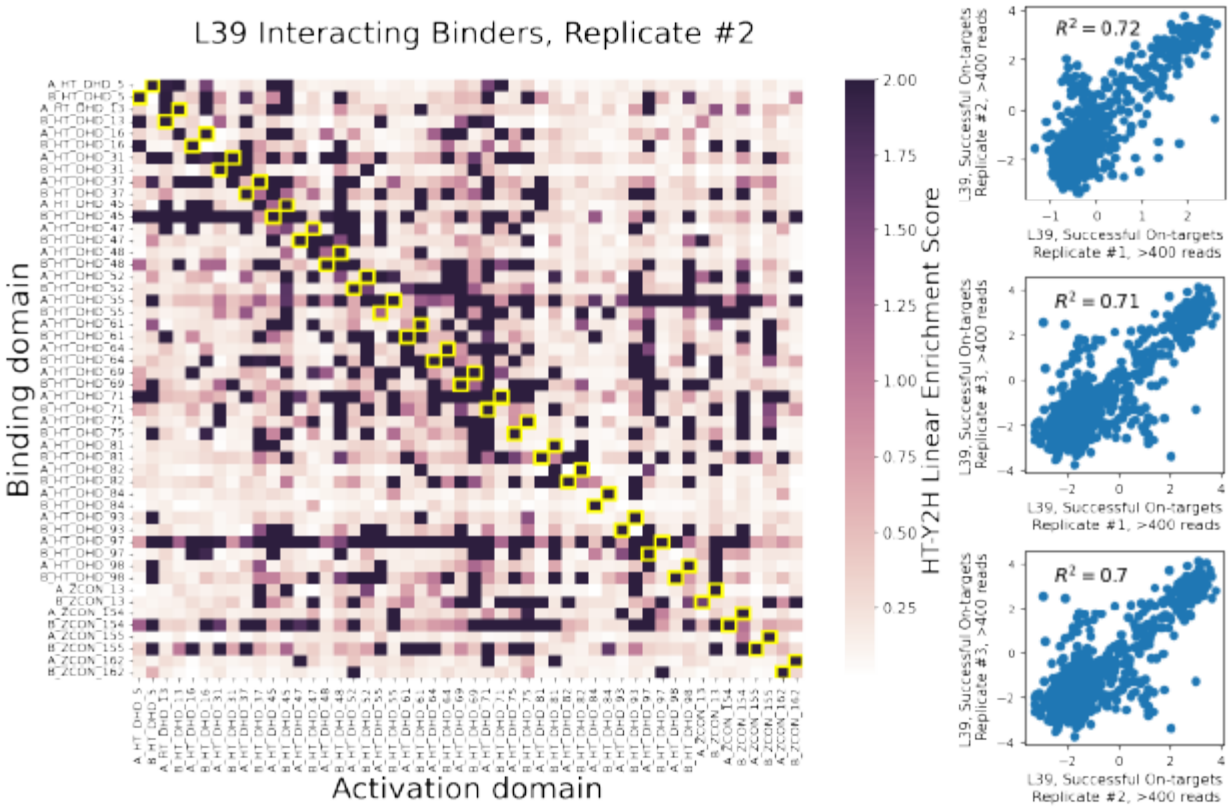

Figure 14. 33 out of 40 pairs that were deemed successful on-target interactions in L38 experiment were found to interact in both orientations in replicate #2 of L39 experiment. Log values of enrichment scores of "high-quality pairs" that had at least 400 reads in the "input" sample, were found to have a high degree of correlation of about R^2=0.7 in three replicates of the experiment.

We found high degree of correlation between the linear (R^2=0.39-0.53) and log (R^2=0.63-0.83) enrichment values of the control ZCON heterodimers in the three replicates of the experiment (Figure 15). There was also good agreement between our data and the data characterizing the interactions between these binders using low-throughput Y2H published in (Z. Chen et al., 2019).

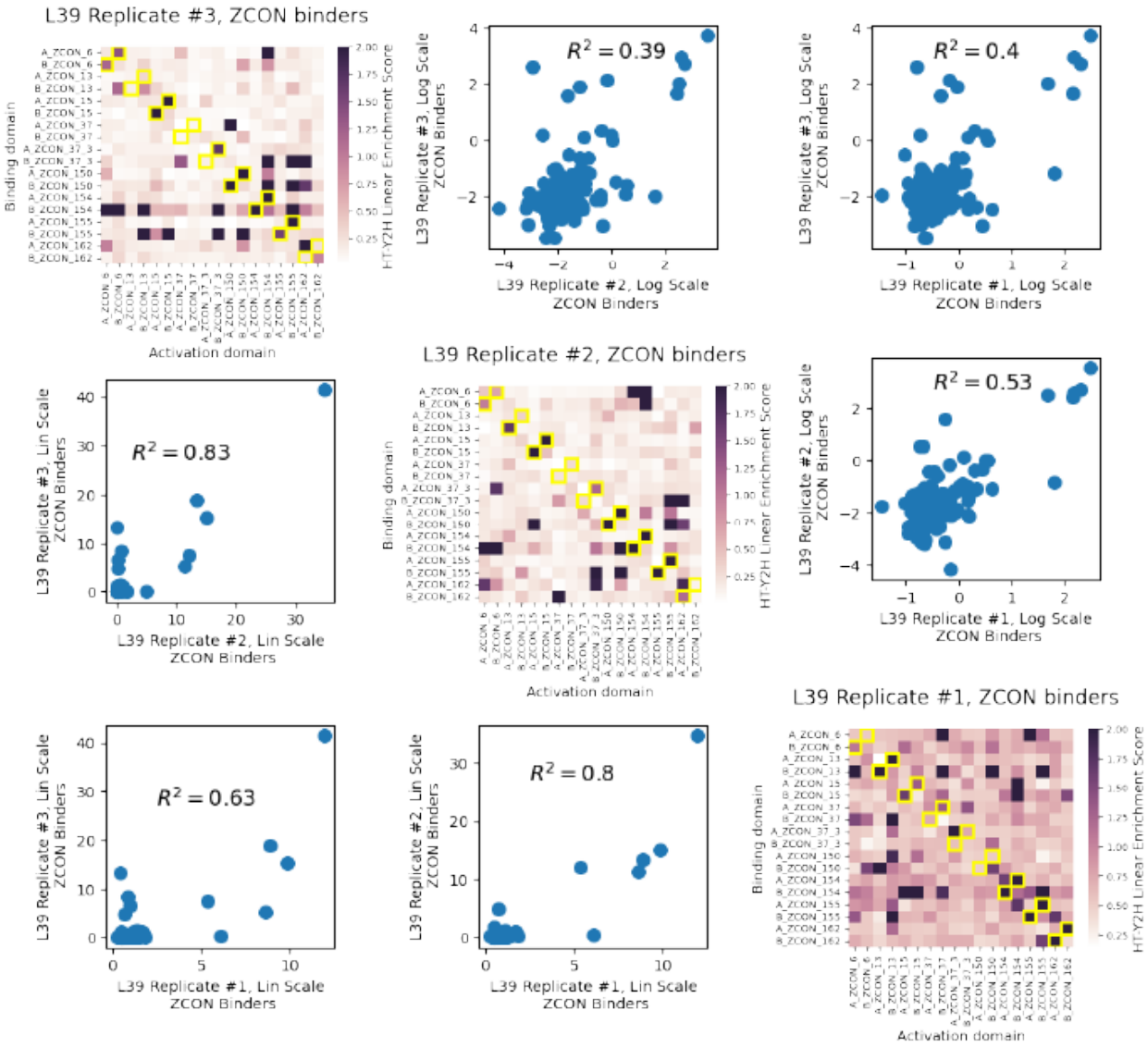

Figure 15. We found high degree of correlation between the linear and log enrichment values of the control ZCON heterodimers in the three replicates of the experiment. There was also good agreement between our data and the data characterizing the interactions between these binders using low-throughput Y2H published in (Chen et al., 2019). The three left lower scatter plots compare linear enrichment values while the three upper right corner scatter plots compare log enrichment values.

Experiment L42. Testing our assay using previously characterized ZCON heterodimers.

Albeit good overall agreement between L38 and L39 library screens, and in particular good correlations between the enrichment values for the “control” ZCON hetero-dimer binders in the three replicates of the L39 experiment, we decided to screen that control set of just 9 ZCON heterodimer pairs in a separate experiment. We hoped that smaller library size would diminish potential noise in the data and allow us to get a much higher number of reads per each pair in a single Illumina sequencing run, thus ensuring the achievement of the “true” interaction picture between the binders.

The resulting interaction strength heatmap is shown in Figure 16. As can be seen in the figure, our high-throughput data coincide with the previously published low-throughput Y2H data in correctly identifying 3 out of 7 interacting pairs (ZCON13, ZCON154, ZCON155) and one non-interacting one (ZCON6). The remaining 4 pairs identified as interacting by the low-throughput assay (ZCON37, ZCON150, ZCON37_3, ZCON15) were not confirmed by our high-throughput screening.

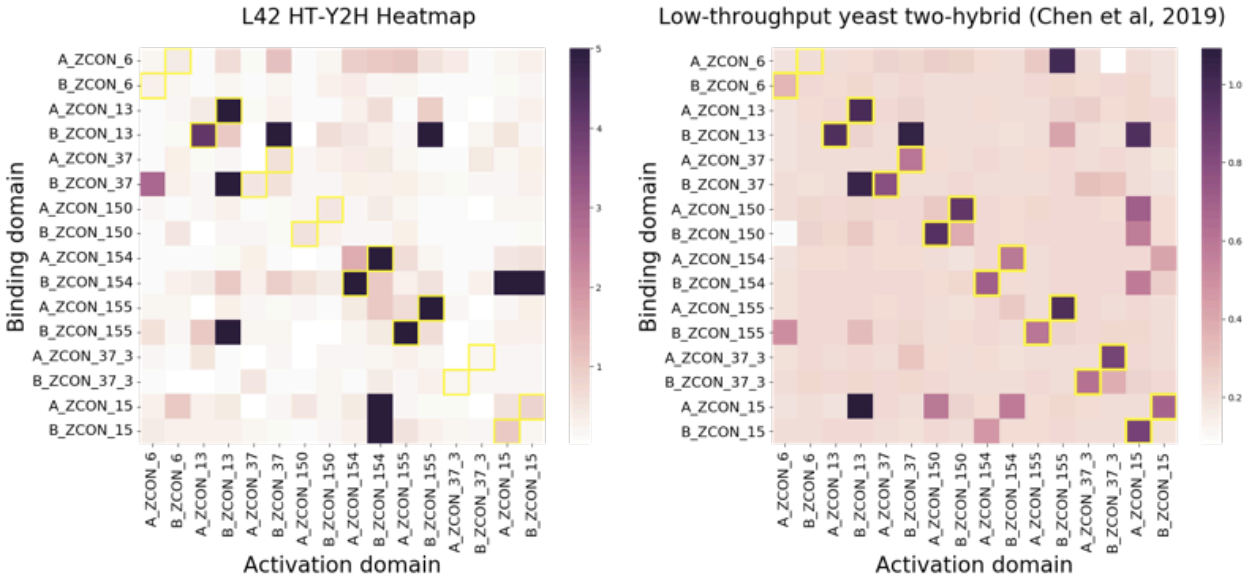

Figure 16. Interaction strength heatmap obtained using our HT-Y2H assay (left panel) and previously published low-throughput Y2H data (right panel). While there is a good agreement between the assays for ZCON6/13/154/155 pairs, our assay could not confirm four pairs ZCON15/37/37_3/150 as interacting ones.

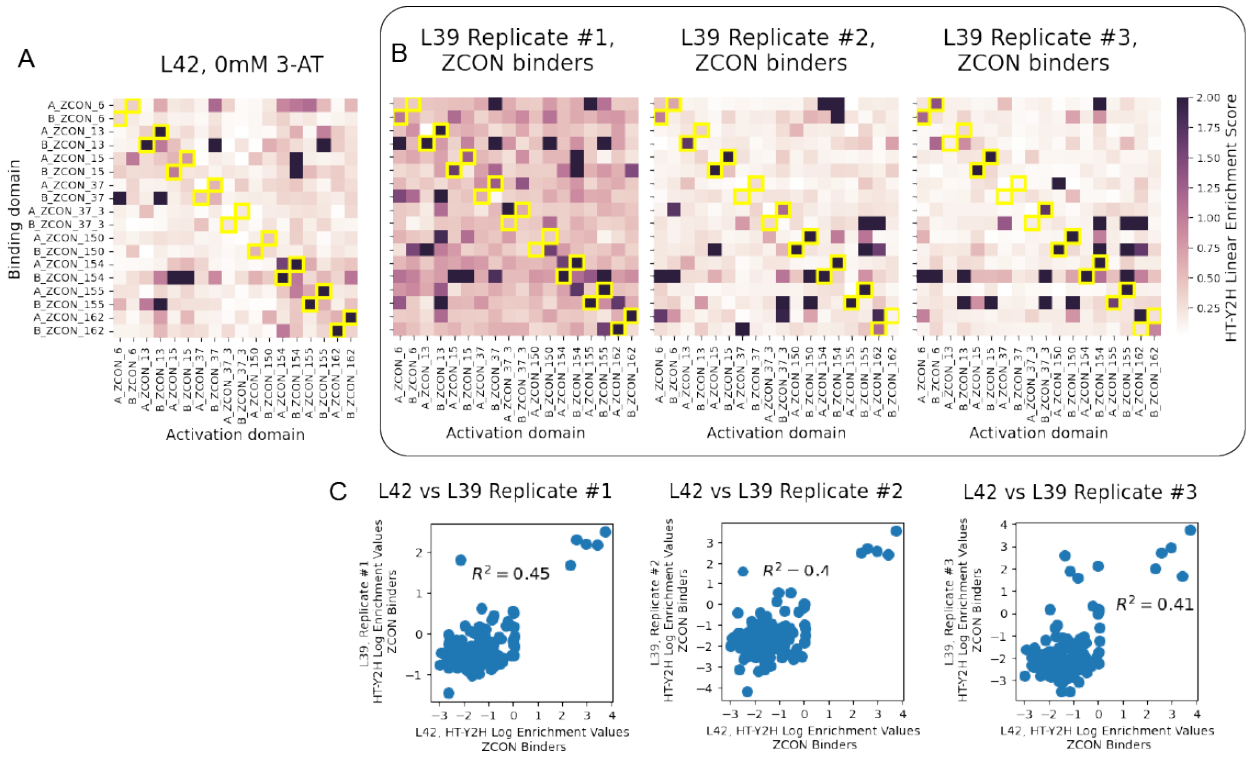

Figure 17. A good visual agreement can be seen between the interaction strength heatmap of separately screened L42 library of ZCON binders (A) and the interaction heatmap for those same binders obtained in either replicate of larger L39 library (B). Although the Pearson correlation coefficients (C) are moderately good, our HT-Y2H assay correctly picks the same on-target interactions in all the above experiments separating them from non-interacting binders.

Since the selection time in this experiment was relatively short at about 18 hours potentially revealing only the strongly interacting pairs, we decided to repeat the experiment using different amounts of HIS3 inhibitor called 3-AT (3-Amino-1,2,4-triazole) to extend the dynamics and selection time range of the experiment. We previously found that 1mM to 2mM of 3-AT was the ideal concentration of 3-AT to add to the selection media to achieve best correlation with already published results for BCL2 family proteins. But since the paper in which ZCON proteins have been characterized uses 100mM 3-AT added to the selection media, we decided to check whether that high concentration would reveal missing interactions (ZCON37, ZCON150, ZCON37_3, ZCON15), so we used 10mM, 20mM and 80mM of 3-AT in the next experiment. As can be seen in the Figure 18, significantly increasing the concentration of 3-AT did not reveal missing interactions but rather worsened the interaction picture by introducing more off-target interactions with increasing concentration of 3-AT. It is important to note that at 80mM 3-AT there were fewer on-target interactions than at 0mM whereas the amount of off-target interactions increased significantly. Moreover, from the Figure 18 it can be seen that even 10mM of 3-AT already significantly distorts the interaction picture. The results of this experiment conclude that previously identified optimal concentration of 1mM to 2mM of 3-AT is ideal for our assay. The discrepancy between 3-AT concentrations ideal for our assay (1-2mM 3-AT) and that in the published low-throughput Y2H (100mM 3-AT) can be explained by using different yeast strains and different expression systems used to drive the target HIS3 gene, including different constitutive promoters used to express hybrid proteins.

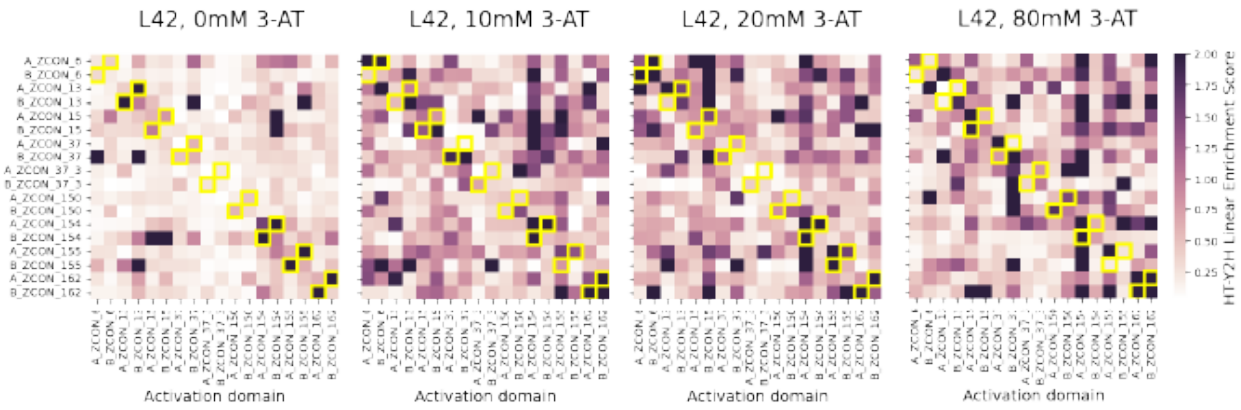

Figure 18. Testing whether increasing the amount of HIS3 inhibitor called 3-AT (3-Amino-1,2,4-amino triazole) in the selection media which increases the selection pressure, would reveal more on-target interactions among ZCON heterodimers. In particular, 80mM of 3-AT is close to 100mM of 3-AT used in the paper (Chen et al., 2019) where these binders were characterized. As can be seen from the heatmaps, higher concentrations of 3-AT introduce many off-target interactions confirming our previous finding that 1-2mM 3-AT concentration is optimal for our HT-Y2H assay. Almost an order of magnitude discrepancy between optimal 3-AT conditions in our assay (1-2mM 3-AT) and low-throughput Y2H (100mM 3-AT) can be explained by using different expression systems and yeast strains.

Experiment L43. 104x104 library of additionally re-designed hydrogen bond network containing binders.

Since we observed high degree of reproducibility applying our method to characterize de novo designed heterodimers, we collaborated with Postdoctoral scholar Ajasja Ljubetic from the Baker lab to characterize a new set of heterodimers with shorter redesigned backbone. The library contained 52 pairs of on-target binders. 9 pairs of previously characterized ZCON binders were added to this library resulting in the overall library size of 122x122=14,884 possible combinatorial pairs.

We performed two technical replicates of the experiment where the selection for interaction in the histidine deficient media was performed in two different flasks using two aliquots of cells taken from the same transformed library, i.e. from the same “input” sample. Input (pre-selection) and output (post-selection) samples of Replicate #1 received approximately 22 and 18 million

reads. Replicate's #2 input and output samples had approximately 25 and 20 million reads. So, each sample had more than 1000 reads per pair considered to be deep enough sequencing coverage.

The number of pairs with linear enrichment score greater than 1 in Replicate #1 equals 1677 corresponding to 11.3% of all interactions being positive. Excluding 9 control ZCON pairs, or 18 proteins in total, the library is 104x104, or 52 pairs. Out of 52 pairs 14 pairs interact in both orientations, or 27% success rate (see Figure 19).

We also found good correlation between the enrichment scores of control ZCON binders in both replicates of this experiment and the scores obtained in the previous L42 experiment (see Figure 20).

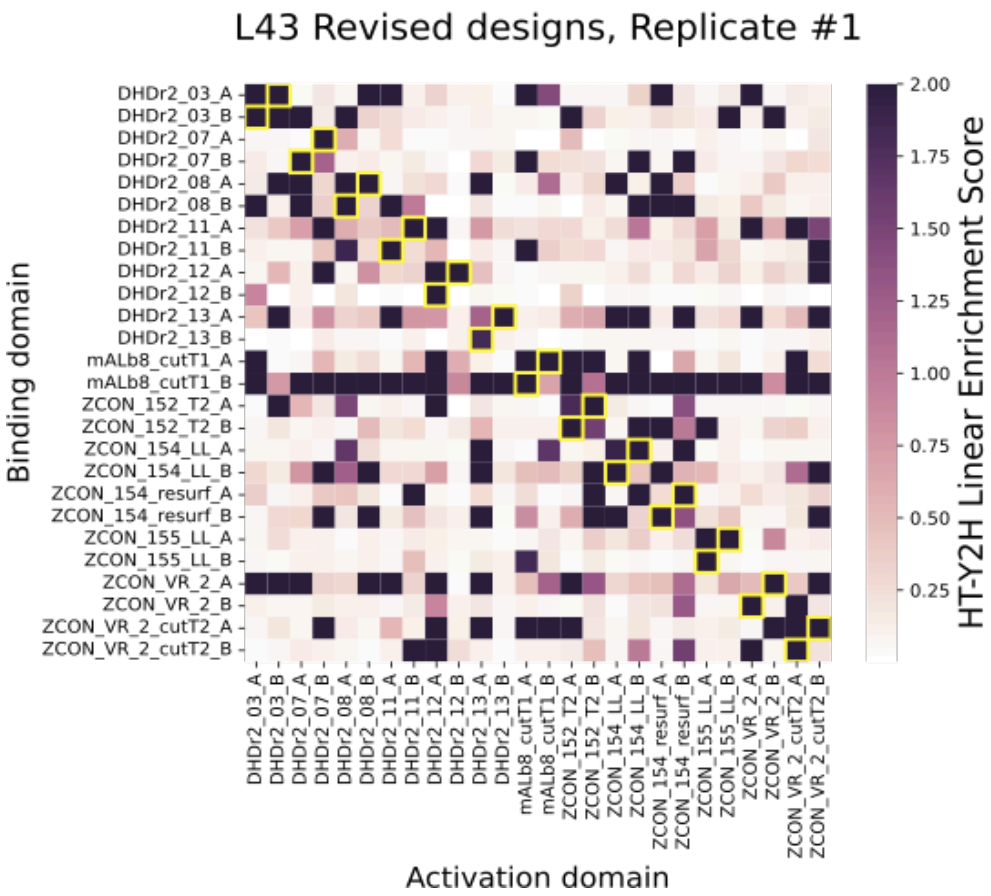

Figure 19. Characterizing interactions between redesigned heterodimers. Shown are only successful on-target interactions that had approximately 27% success rate (14 pairs out of 52 on-target pairs were found to interact) whereas 11.3% of all pairs were found to interact in the experiment, which includes off-target interactions.

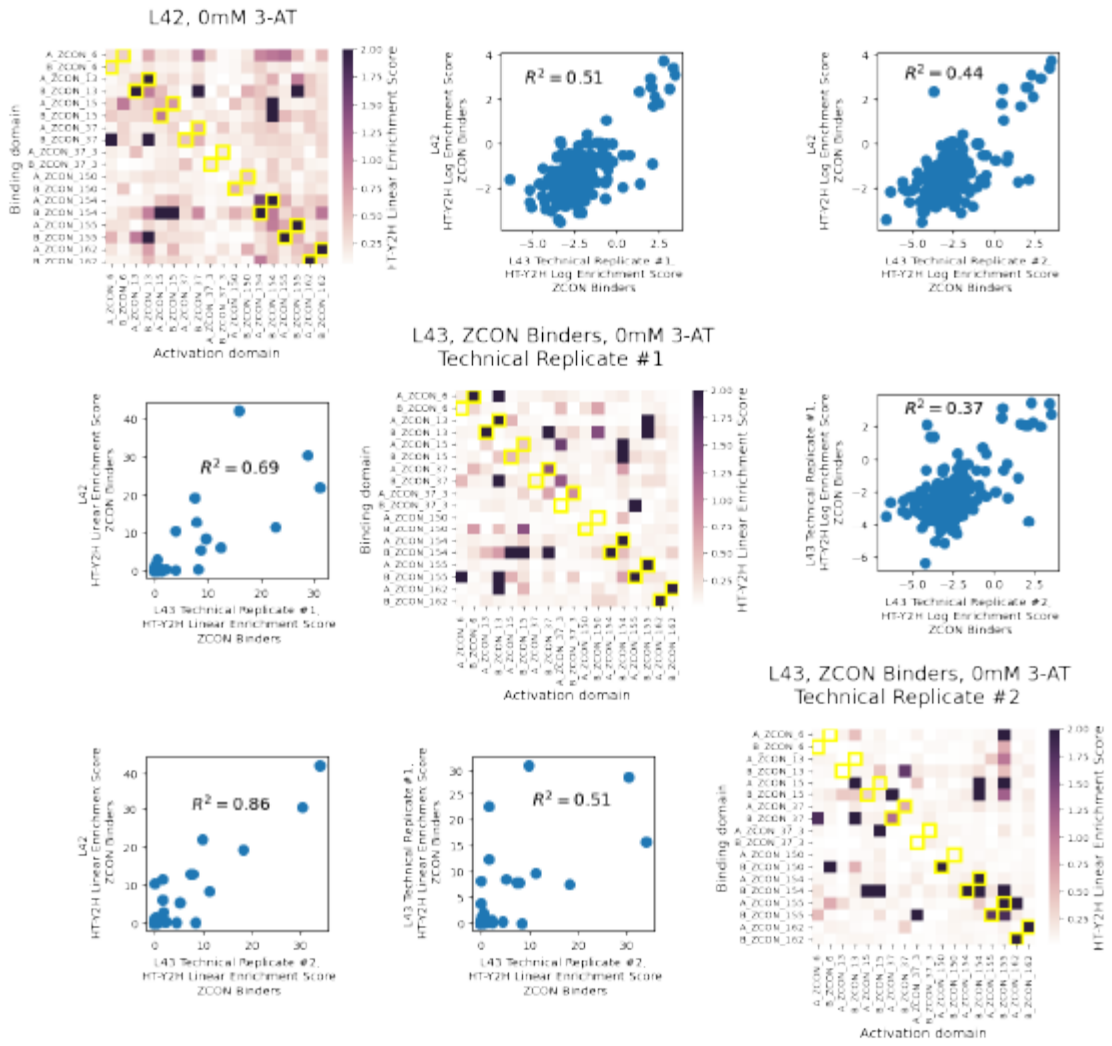

Figure 20. Comparing the interaction strength heatmaps for ZCON heterodimers in L42 and L43 experiments. A good visual and quantitative agreement can be seen in the figures. Lower left corner scatter plots exhibit good linear correlation between both the replicates of L43 experiment as well as between both replicates and L42 experiment. The upper right corner plots the same correlations in the log scale.

Experiment L44. Combining binders from libraries L33, L39 and L43 together and screening them in an all-against-all manner in a single experiment.

Following our success in the L39 experiment assembling nearly 99.5% of all possible protein pairs during the transformation, we decided to investigate whether we can successfully assemble even larger libraries. For that purpose, we combined all previously made stocks of

double stranded DNA fragments encoding proteins from L33, L39 and L43 libraries, mixed them with three linear fragments composing the backbone of the plasmid and transformed them into our base y777 yeast strain. The resulting library size is 337*337=113569 protein pairs. The transformation efficiency was 1 million transformants when the standard electroporation protocol was used. The library was subjected to the standard selection process and the “input” and “output” sequencing samples were prepared using the standard procedure. The analysis of the sequencing data revealed that nearly 99% of all the pairs assembled successfully.

Most of the successful on-target interactions found in all three replicates in the previous L39 experiment were found to interact in the current experiment in at least one orientation confirming consistency of our assay (see Figure 21). Log enrichment scores of binders in this subset of successful designs calculated in this L44 experiment have a good Pearson correlation of R^2=0.54-0.56 with the values in either of the three replicates of L39 experiment.

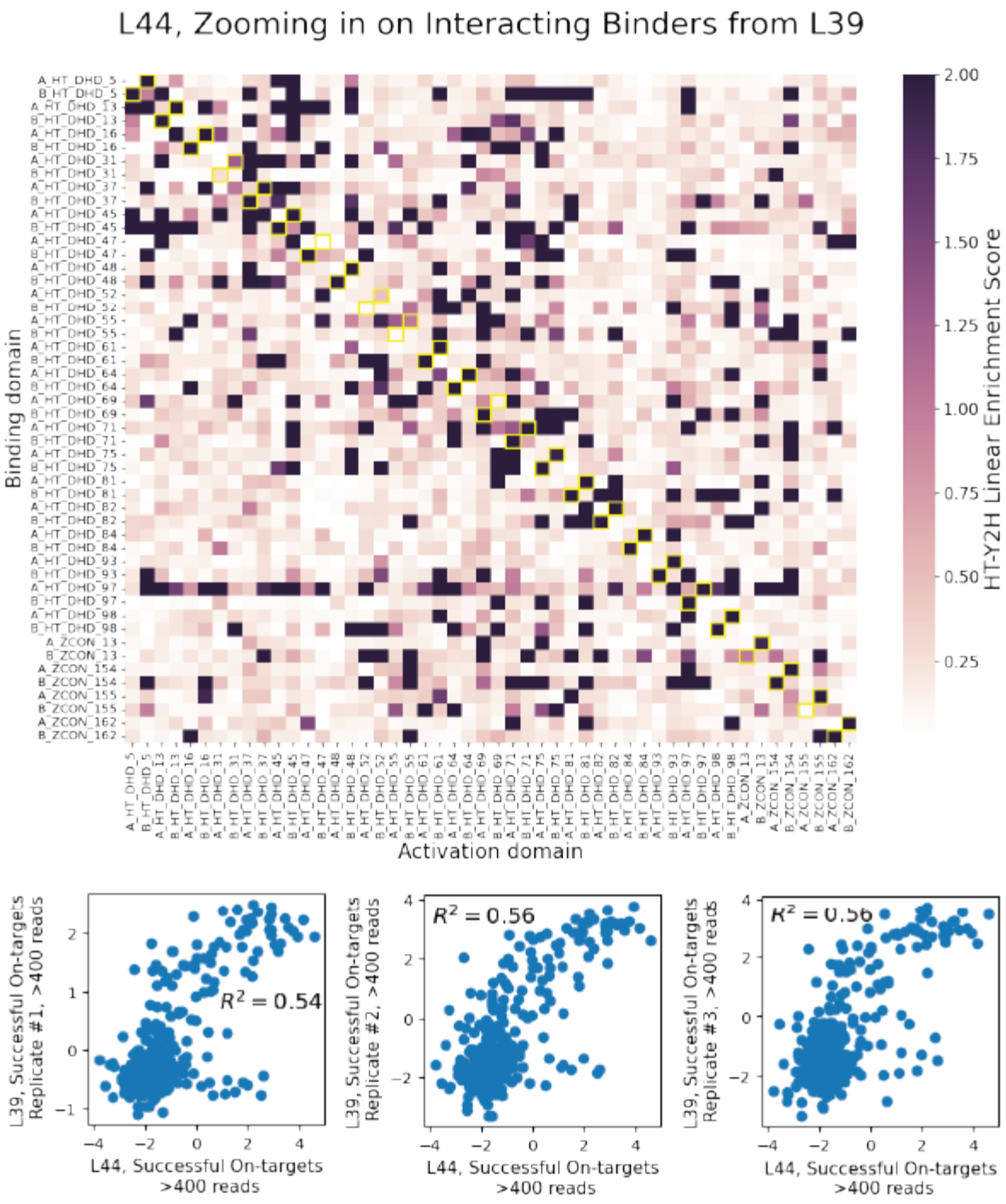

Figure 21. On-target designs that were found to interact in all three replicates of L39 experiment are shown in this interaction strength heatmap. Scatterplots show good correlation between the log values of enrichment score of the L44 experiment and those of either of the three replicates of the L39 experiment for high-quality pairs that had at least 400 reads in the "input" sample (before the selection).

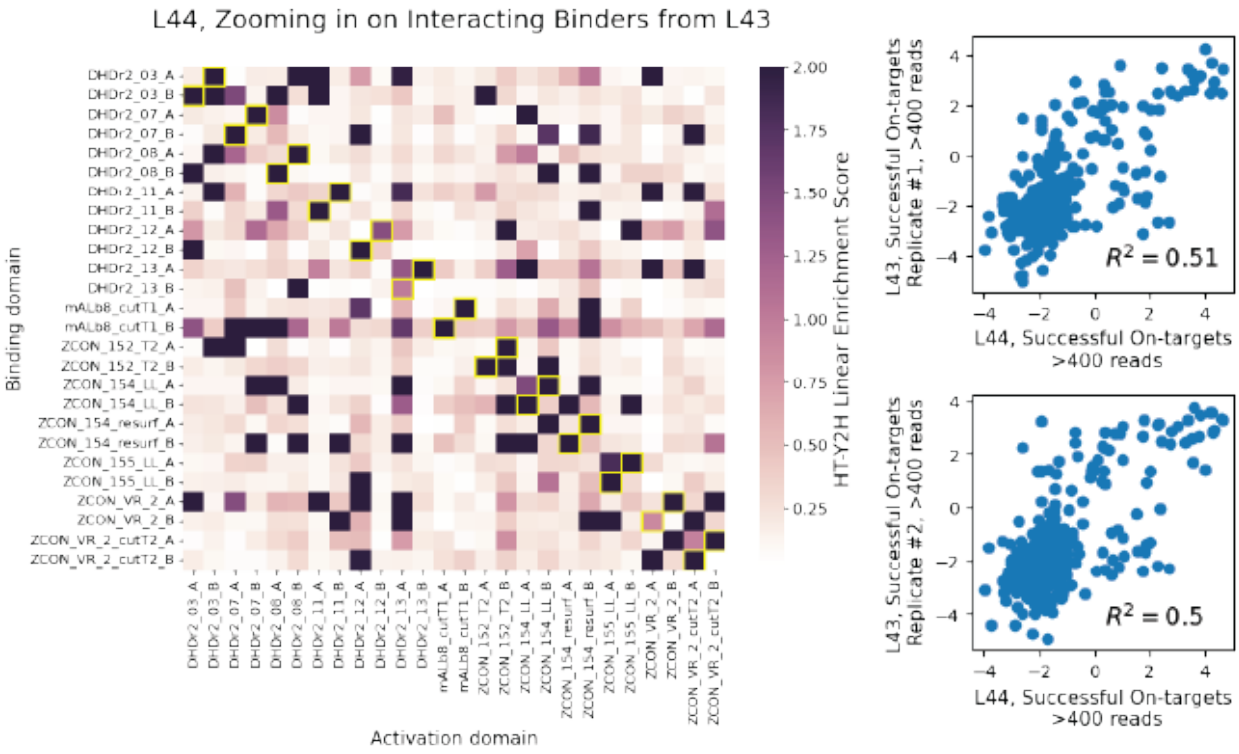

Figure 22. On-target designs that were found to be successful in both replicates of the L43 experiment were found to interact in the L44 experiment, as shown in this heatmap. Scatterplots show good correlation between the log values of enrichment score of the L44 experiment and those of either of the two replicates of the L43 experiment for high-quality pairs that had at least 400 reads in the "input" sample (before the selection).

Experiment L45. Characterizing interaction between BCL2 homologs and their de novo designed inhibitors under different conditions to investigate consistency of our HT-Y2H assay.

To test the consistency of the plasmid extraction protocol and the robustness and reproducibility of the HT-Y2H protocol, the following experiment was performed. A library of six BCL-2 homologous and their 9 de novo designed inhibitors, previously tested in L33 experiment, was transformed into the base yeast strain and outgrown for 36 hours in SC-TRP media (Flask#1) to achieve a single plasmid in each yeast cell. At this point, when the OD of the cell culture has reached 1.0, a fraction of the cells was frozen in ZymoResearch Solution1 buffer

for subsequent plasmid extraction and NGS sequencing (Sample#1) and the remaining cells were used to inoculate another SC-TRP flask (Flask#2) and SC-HIS-TRP flask (Flask#3). Flask#3 was grown for 24 hours till reaching OD=0.5 at which point a fraction of cells was collected and frozen (Sample#3) and 250uL of the remaining cells were inoculated in a fresh 50mL of SC-HIS-TRP media (Flask#4) to continue the selection for interaction process for another 16 hours until reaching OD=1.0 at which point all the cells were collected, split into two aliquots of 100 million cells each and frozen (Samples #4 and #5).

Meanwhile, Flask#2 (SC-TRP) was grown for 12 hours until reaching OD=1.0 at which point a fraction of cells (Sample#2) was frozen and 125uL and 250uL from the remaining cell culture were used to inoculate another SC-HIS-TRP Flask#5 and Flask#6. Upon reaching OD=1.0 cells from Flasks #5 and #6 were split into two aliquots and frozen (Samples #6 and #7, and #8 and #9 correspondingly).

The experiment layout can is depicted in Figure 23.

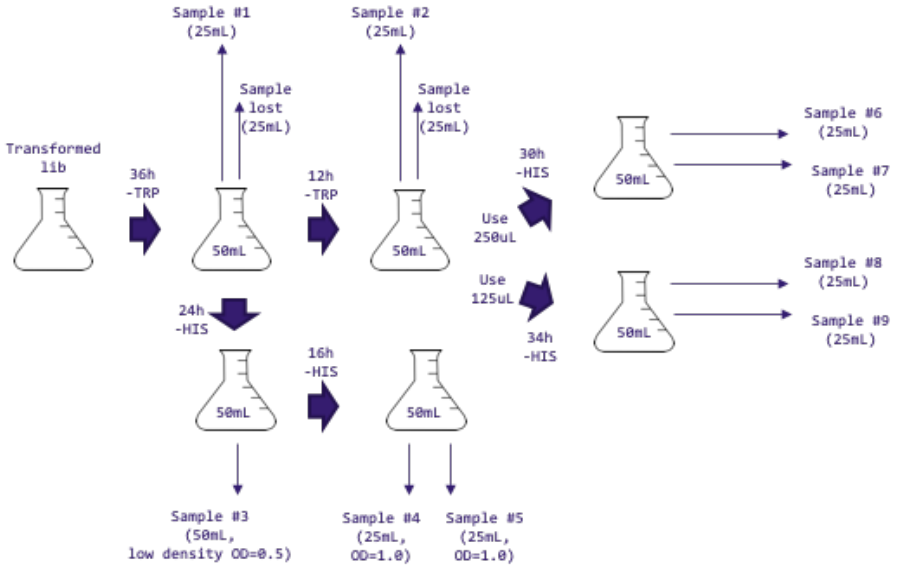

Figure 23. L45 experiment layout. A backbone lacking SH3 domain and both fluorescent proteins is used in this experiment in hope to eliminate any potential misfolding of proteins to achieve better agreement with the published data.

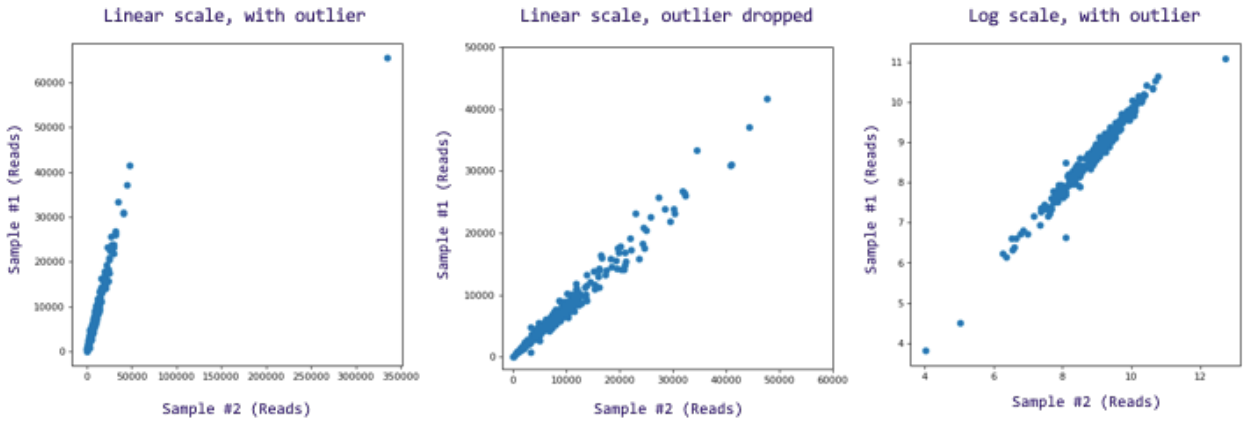

Figure 24. Investigating whether the number of reads per each distinct pair would change if the library would be selected for additional 12 hours following regular 36 hours upon the transformation. We found that passaging cells for 36 and 48 hours upon electroporating the cells does not affect the number of read counts per pair, i.e. the population phenotype stayed the same. One outlier has been detected in the sample #2. The high correlation coefficient also implies consistency of performing plasmids extractions and qPCR.

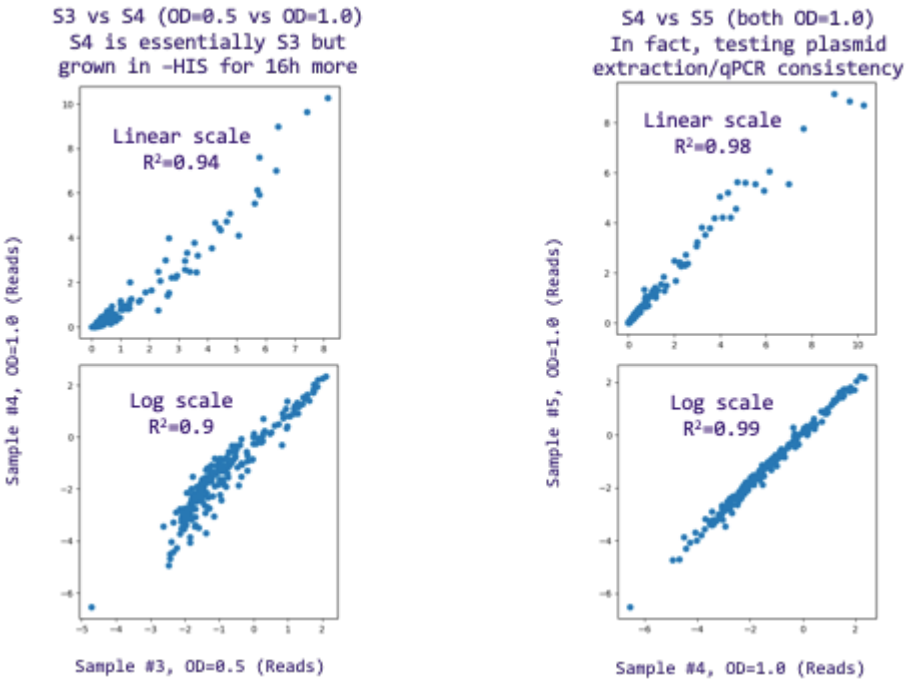

Figure 25. Comparison of samples S3 and S4 investigates whether subjecting cells to additional selection time by diluting down the sample S3 would affect the interaction pictures. The high degree of correlation between enrichment scores of samples S3 and S4 confirms that additional selection time is not necessary. Moreover, this also implies consistency of performing plasmid extractions and qPCR. The high degree of correlation between the enrichment scores of samples S4 and S5 that were made from the same batch of cells additionally confirms consistency of performing plasmid extractions and qPCR.

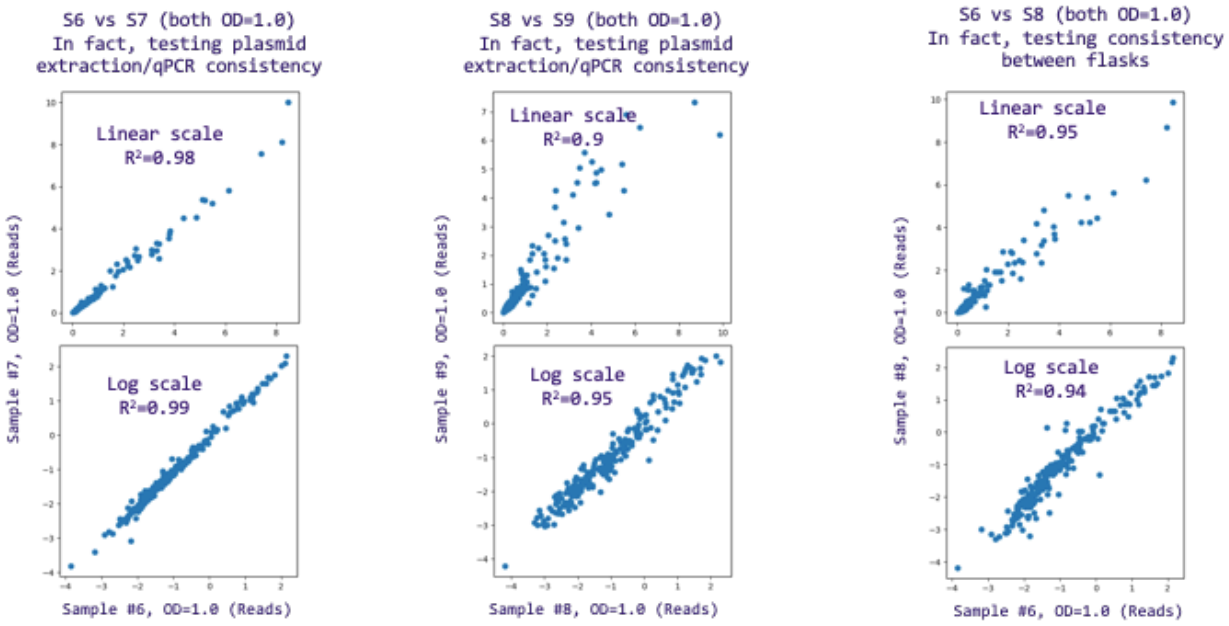

Figure 26. The high degree of correlation between the enrichment scores of samples S6 and S7 that were made from the same batch of cells additionally confirms consistency of performing plasmid extractions and qPCR. The R^2=0.9 and R^2=0.95 Pearson correlation between linear and log enrichment scores of samples S8 and S9 is an example of slight variability in the plasmid extraction process and subsequent qPCR. The high degree of correlation between the enrichment scores of samples S6 and S8 shows that using different amount of cells to start the selection process does not affect the interaction outcome too much.

Experiment L48. Using truncating plasmid backbone to characterizing interaction between BCL2 homologs and their de novo designed inhibitors.

Still trying to figure out whether higher agreement with already published data for ZCON hetero-dimers and BCL2 family binders can be achieved using our assay, we further hypothesized that truncating the hybrid proteins sequences by removing the fluorescent and SH3 domain proteins sequences from them might eliminate potential misfolding of hybrid proteins and improve the results. We decided to test this hypothesis on the BCL2 family of proteins tested previously. After removing the above-mentioned sequences, the hybrid proteins' constructs

looked like the following: DBD-linker1-SV40NLS-linker2-Protein1 and SV40NLS-AD-linker3-Protein2.

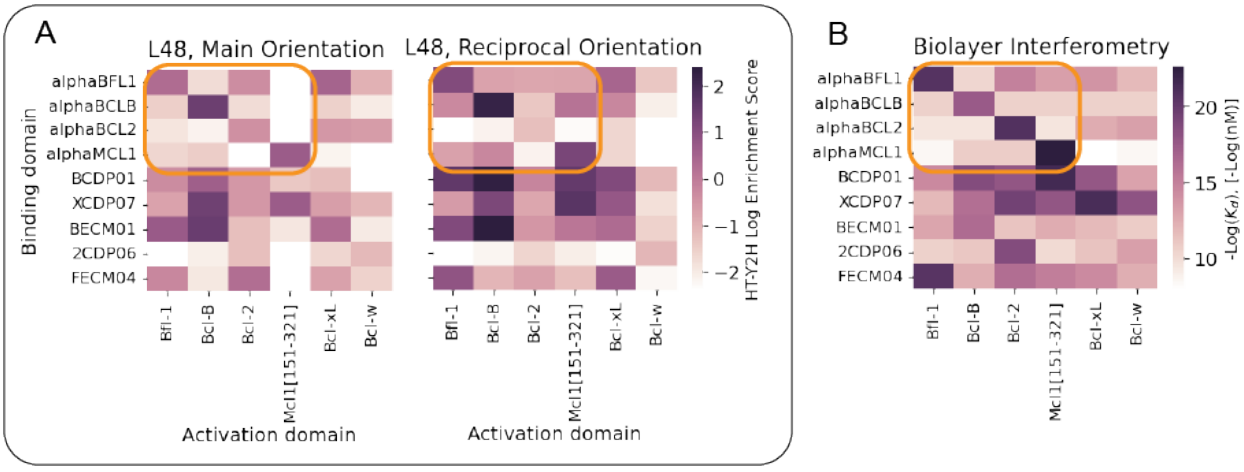

Figure 27. Comparing the main and reciprocal orientations' interaction strength heatmap obtained using truncated backbone in our HT-Y2H assay (A) with the data obtained using Biolayer Interferometry (B) in (Berger et al., 2016). A modest visual correlation can be observed. Highlighted orange region encircles interactions from the main figure published in (Berger et al., 2016) containing de novo designed inhibitors that bind their targets the most specifically.

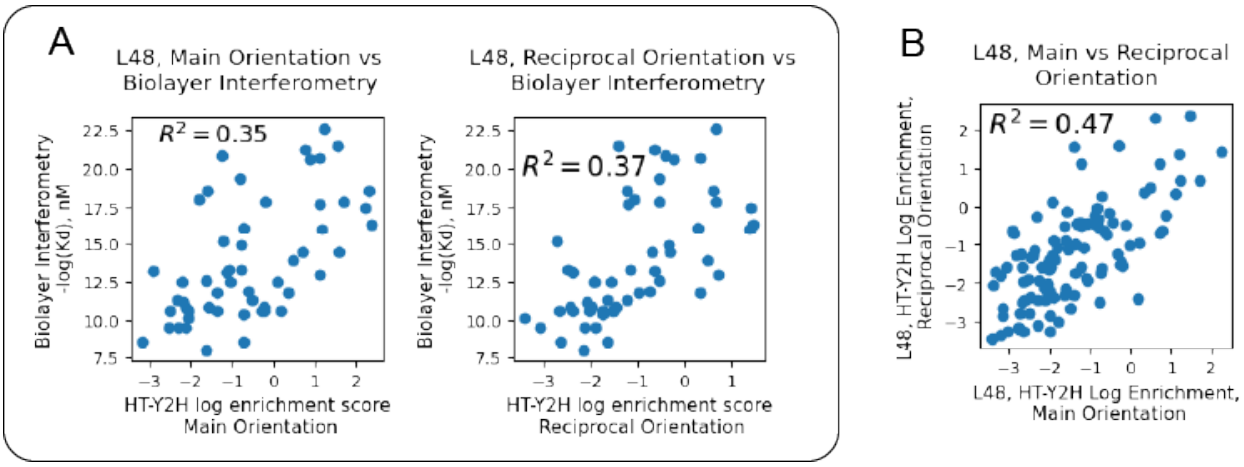

Figure 28. Modest correlation between our HT-Y2H assay's results and Biolayer Interferometry data can be observed (A). (B) Consistency of our HT-Y2H assay is confirmed by good correlation between the main and reciprocal orientations' enrichment score.

Experiment L49. Screening 337x337 library with truncated backbone design.

The run is under sequenced with 4M Reads for TRP.

The library includes all the hydrogen bonds containing proteins A_DHD_XXX and B_DHD_XXX and 15 proteins previously screened by David Younger and Stephanie Berger.

Chapter 4. Achieving success in optimizing our assay.

Contributions: Proteins in experiments L50, L52, L55, L65 were designed by Ph.D. student Johannes Linder (Professor Georg Seelig's lab); Proteins in experiment L67 were designed by postdoctoral scholars Ajasja Ljubetic and Basil Wicky. Proteins in experiment L70 were designed by postdoctoral scholars Ajasja Ljubetic. All low- and high-throughput Y2H experiments in this chapter, data analysis, figure preparations were performed by Alexandr Baryshev.

Swapping barcodes and terminators revealed previously missing interactions.

In RLL-Y2H assay both hybrid proteins are encoded on the same plasmid with two expression cassettes facing each other similar to our system with two terminators between the two coding regions. The RLL-Y2H assay uses mMeI restriction enzyme site between the stop codons and the terminators allowing to cut a double stranded fragment containing the last 18nt of each protein on its ends. Those 18nt of proteins' sequences remaining on the fragment after restriction allow to identify interacting proteins provided that those last 18nt are diverse enough.

Inspired by the paper, we were interested in whether replacing barcodes and their PCR handles in our system with 6nt long MmeI restriction enzyme site would lead to the same results. In particular, we were interested how the system would behave when applied to the control ZCON heterodimers, in particular to ZCON37 pair which has been confirmed to interact using cryo-EM imaging.

We decided to perform a low-throughput chemical transformation of our base yeast strain using approximately the same amounts of transformed linear fragments DNA. The transformation mixture was plated on yeast SC-HIS agar media that allowed selection for the

interaction. To our surprise, two days after the transformation there were significantly more transformed colonies on the plate corresponding to the design without barcodes and PCR handles compared to the design previously used in all our high-throughput experiments. This discovery led us to believe that a PCR handle and a barcode put right after the stop codon and before the terminator was disrupting the termination process in some cases, like in the plasmid encoding the ZCON37 with PCR handles and barcodes. It is important to note that this design apparently still worked for other protein pairs, like for example (ZCON13/154/155).

We then decided to introduce this change, and put terminators before the PCR handles and barcodes in our future experiments.

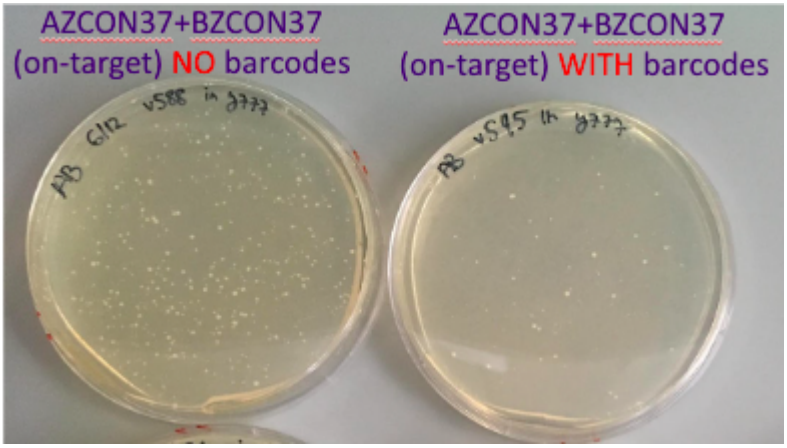

Figure 29. Low-throughput chemical transformation of two strains. The strain on the left plate has no barcode and its corresponding PCR handle, instead the stop codon is followed by a GTCGGA sequence which is a reverse complement sequence to MmeI restriction enzyme's recognition sequence TCCGAC. The straight on the right plate has a PCR handle and a 20nt barcode between the stop codon and the terminator. The truncated version of the backbone is used in both strains and carries no fluorescent proteins or SH3 domain.

We further successfully used this design lacking a barcode and its PCR handle to screen one interacting (P3-P4) and one non-interacting (P1-P4) pair of coiled coils from (Lebar et al., 2020). As can be seen in the Figure 30, the on-target pair P3-P4 had a significantly higher amount of yeast colonies on the histidine lacking plate that selects for interaction than the not-on-target pair P1-P4.

This pilot experiment with P1-P4 coiled coils encouraged us to consider screening the remaining coiled coils P5-P12 in the next experiments in hope to achieve agreement between our HT-Y2H assay's results and the published data (Lebar et al., 2020).

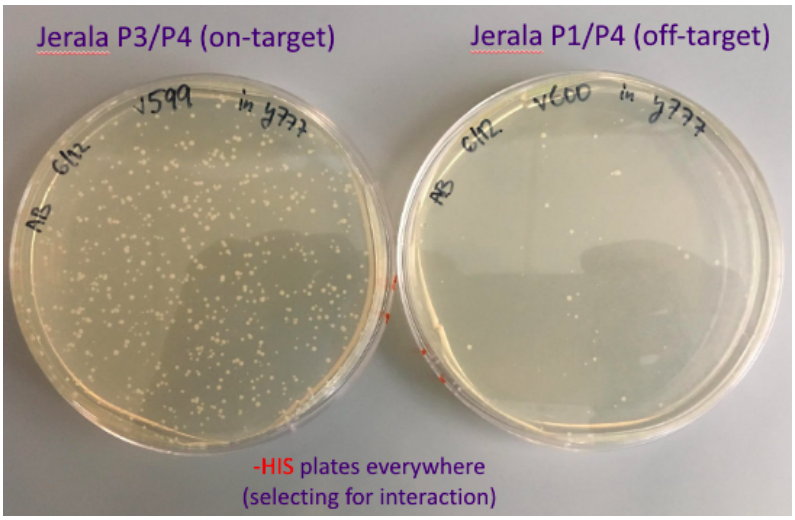

Figure 30. Low-throughput transformation of two strains, one of which encodes on-target pair of coiled coils P3-P4 (left plate) and the other encodes a non-interacting pair of coiled coils P1-P4 (right plate). Approximately the same amount of transformed DNA fragments was used for the both strains which allows us to state that assay correctly pick up interactions between just these two cases of interacting and non-interacting coiled coils.

Revised Plasmid Design Part

A new plasmid design is used in the subsequent libraries. Each binder's sequence is followed by a stop codon which in turn is followed by a synthetic terminator from (Curran et al., 2015). Tsynth23 and Tsynth27 are used correspondingly to terminate the transcription of DBD and AD fusion proteins. The terminators are followed by PCR handles named RL3 and RL4 which are followed by predetermined 20nt long DNA barcode sequences that are usually synthesized along with binder sequences. 22nt long insulation sequence dubbed TS separates the barcode sequences. TS sequence is used as a homology sequence during the plasmids assembly in yeast.

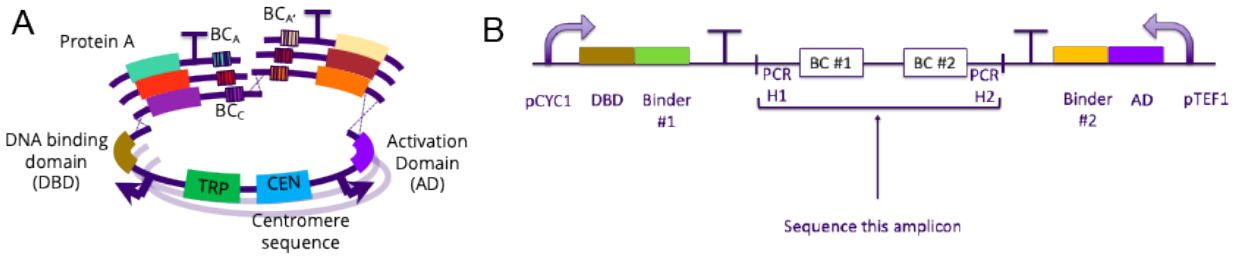

Figure 31. (A) Revised plasmid assembly. Barcodes are following terminators. (B) Sequencing amplicon contains just the PCR handles H1 and H2 and pre-defined barcode sequences BC1 and BC2.

Choosing P1-P12 coiled coils as another control set.

A coiled coil is a structural motif found in some proteins and it usually consists of 2-7 alpha helices that are coiled together. Coiled coils are usually part of transcriptional factors and thus are involved in various biological processes such as transcriptional regulation. While the general problem of predicting the folded structure of a protein given its amino acid sequence has not been solved, coiled coil is one of the few motifs for which the relationship between the sequence and final folded structure is relatively well understood. Coiled coils usually contain a repeated pattern of 7 amino acids, called heptad repeat, *hxxhcxc* where *h* is a hydrophobic and *c* is a charged amino-acid residue. The composition of a coiled coil is sometimes represented as a 7-register sequence *abcdefg* where positions *a* and *d* are occupied by a hydrophobic amino-acid residue (usually leucine, iso-leucine or valine).

In (Lebar et al., 2020) Lebar et al. have successfully designed a set of orthogonal 4-heptad long coiled coils that we thought could be used as another benchmark set of proteins that we can use to optimize our assay. The interaction between these peptides was characterized using split luciferase assay in the original paper (see Figure 32).

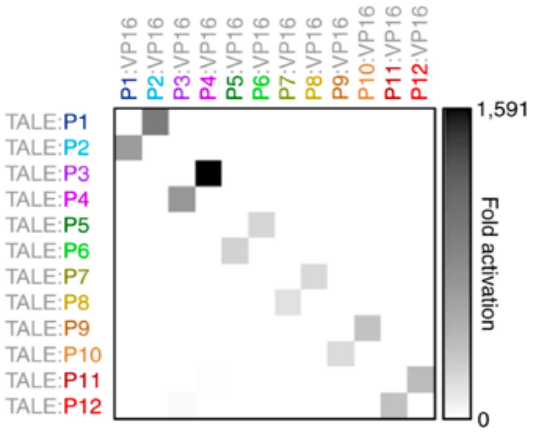

Figure 32. An orthogonal set of 4-heptad long coiled coils characterized using split luciferase assay. The gray bar scale indicates fold change in luminescence.

Experiment L50. Characterizing interactions between 10-heptad long coiled coils designed by a deep exploration network (DEN) computational algorithm.

Large amounts of collected experimental knowledge about the interaction between coiled coils allowed (Potapov et al., 2015) and (Fong et al., 2004) to come up with machine learning algorithms to predict the interaction outcome for any two provided coiled coil pairs.

In collaboration with another PhD student Johannes Linder from the Seelig lab, we designed 4 pairs of 8-heptad long and 4 pairs of 10-heptad long coiled coils, that we designed to maximize their interaction strength with just one other coiled coil in the set and minimize the interaction strength with all the remaining coils, thus forming an orthogonal 8-pairs coiled coil system.

The sequences were ordered using the new “barcode following terminator” design and ordered in a double stranded DNA gene fragment from Twist Bioscience. The order also included coiled coils P1-P4 were added to the transformation mixture for the control purposes to see how P1-P4 coiled coils interact in the library context.

Although, as can be seen in Figure 33, a few on-target designs exhibited desired interaction in at least one orientation, a significant number of strong unanticipated off-target interactions can be seen in the interaction heatmap, mainly among 10-heptad long coiled coils fused to the activation domain and other shorter coils fused to the binding domain. From the discussion with Postdoctoral scholar Ajasja Ljubetic from the Baker lab it was hypothesized that the longer coils are sticker and thus lead to stronger interactions. It can also be seen from the figure that although the log enrichment score for on-target pairs P1-P2 and P3-P4 is negative, which corresponds to overall decrease in relative frequency representation in the population, the interaction between these on-target pairs is still visually much stronger than between other not-on-target combinations of P1-P4 binders, which gives that that had P1-P4 peptides been screened separately by themselves they would clearly have positive log enrichment scores.

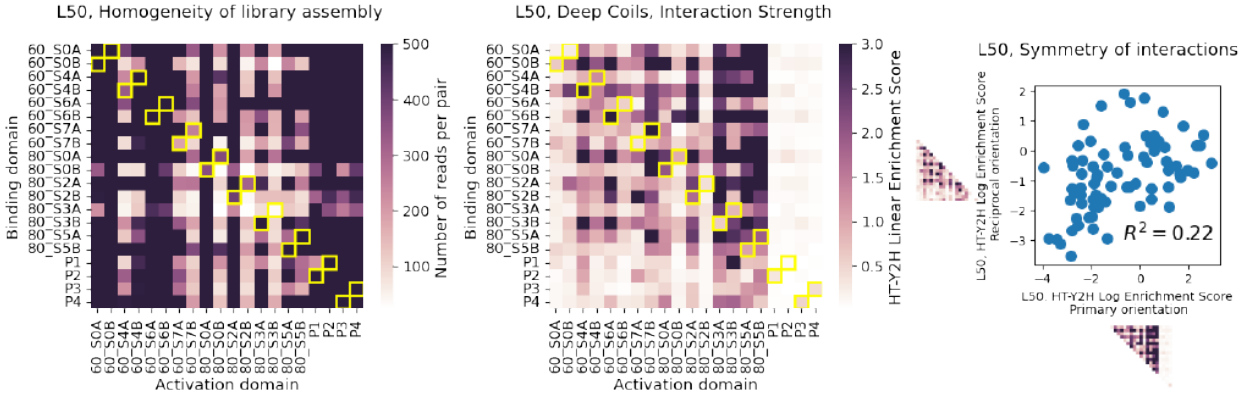

Figure 33. (A) The heatmap of the number of reads per pair for the "input" sample (i.e. before the selection for interaction) characterizing the homogeneity of the library assembly. As can be seen from the figure some binders fused to activation domain and binding domain assembled with lower efficiency which is indicated by lighter color in the figure. (B) From this subfigure showing the interaction strength heatmap it can first be seen that DEN coils were interacting with each other and P1-P4 coiled coils. Second, the interaction strength between short coiled coils P1-P4 and almost any other binder is noticeably weaker than the interaction between DEN coiled coils. It can also be seen that the interactions between 10-heptad long coils, in particular S3A, S3B, S5A, S5B and other binders in the library are stronger than between shorter 8-heptad long coils and especially 4-heptad coils P1-P4, which might be attributed to longer coils being "stickier" than the shorter ones. (C) Symmetry of interactions is moderate with the Pearson correlation coefficient between the upper and lower triangular heatmaps equal to R^2=0.22.

Experiment L51. Characterizing interactions between orthogonal coiled coil peptides P1-P12 from (Lebar et al., 2020) using IDT DNA Technologies as a DNA supplier.

Since previously our assay had successfully picked up the on-target interaction between P3-P4 coiled coils and didn't identify interaction between not-on-target P1-P4 pair, and since the interaction between the on-target pairs P1-P2 and P3-P4 was clearly stronger than the not-on-target interactions between other combinations of P1-P4 binders, we decided to screen the full library of P1-P12 peptides. Thus, a library of P1-P12 orthogonal coiled coil peptides was ordered in a form of a pool of single-stranded oligonucleotides synthesized by IDT DNA. An 80% GC-rich sequence used as a linker separating DNA-binding domain on one end of the plasmid backbone from one of the binders required us to use an additional PCR handle to amplify binder1 gene fragment. That PCR handle has no homology to the plasmid backbone and as a result is chewed away by yeast internal machinery (Figure 34A).

In the resulting interaction strength heatmap (Figure 34B) it can be seen our assay correctly identifies on-target interactions that have a linear enrichment score greater than 1.0 in all six pairs in at least one orientation, although significant amount of unanticipated interactions is present in the heatmap, which makes the whole heatmap look not as neat as the luciferase heatmap from the original paper (Lebar et al., 2020).

We then performed another independent replicate of the experiment starting from the yeast transformation. After the transformed library was passaged in a flask for 36 hours to drop redundant plasmids from yeast cell and maintain only one plasmid per yeast cell on average, two aliquots from this flask were inoculated in two separate flasks with media lacking histidine amino acid to start the protein-protein interaction selection. Thus we performed two technical

replicates of this experiment this time. The resulting interaction heatmaps are shown in Figure 35 and labeled as Replicate #2 and #3. Very high correlation between linear and log enrichment values of these technical replicates can be observed in Figure 35. Moderate correlation was observed between either of these two technical replicates and the first independent replicate of the experiment.

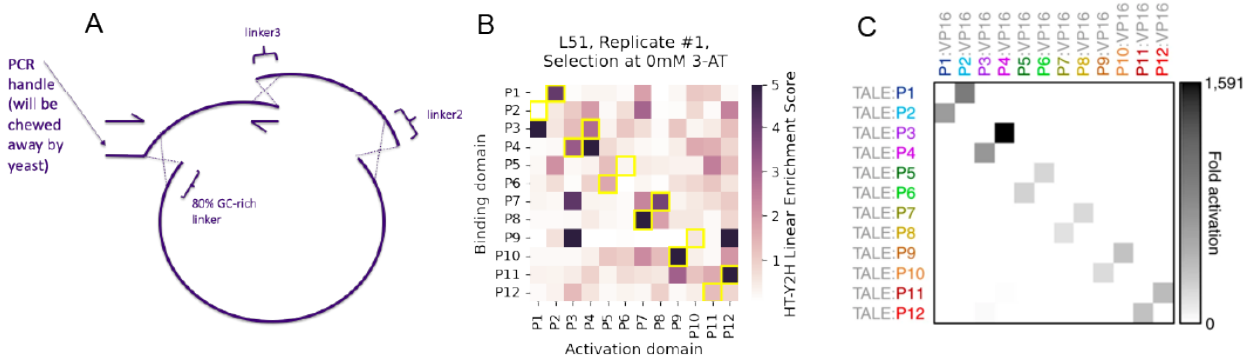

Figure 34. (A) Since the truncated backbone of the plasmid ends with an 80% GC-rich sequence that serves as a homology region during plasmid assembly and since this sequence is poorly amenable to PCR, an extra PCR handle needs to be added 5' of this GC-rich sequence on all oligos in the pool that encode P1-P12 peptides that will be fused to the binding domain. This PCR handle has no homology to any sequence on the backbone and thus will be chewed away by yeast machinery during homologous recombination. (B) Our HT-Y2H assay's interaction strength heatmap shows that our assay correctly identifies on-target interactions that have a linear enrichment score greater than 1.0 in all six pairs in at least one orientation, although significant amount of unanticipated interactions is present in the heatmap, which makes the whole heatmap look not as neat as the luciferase heatmap from the original paper (C)

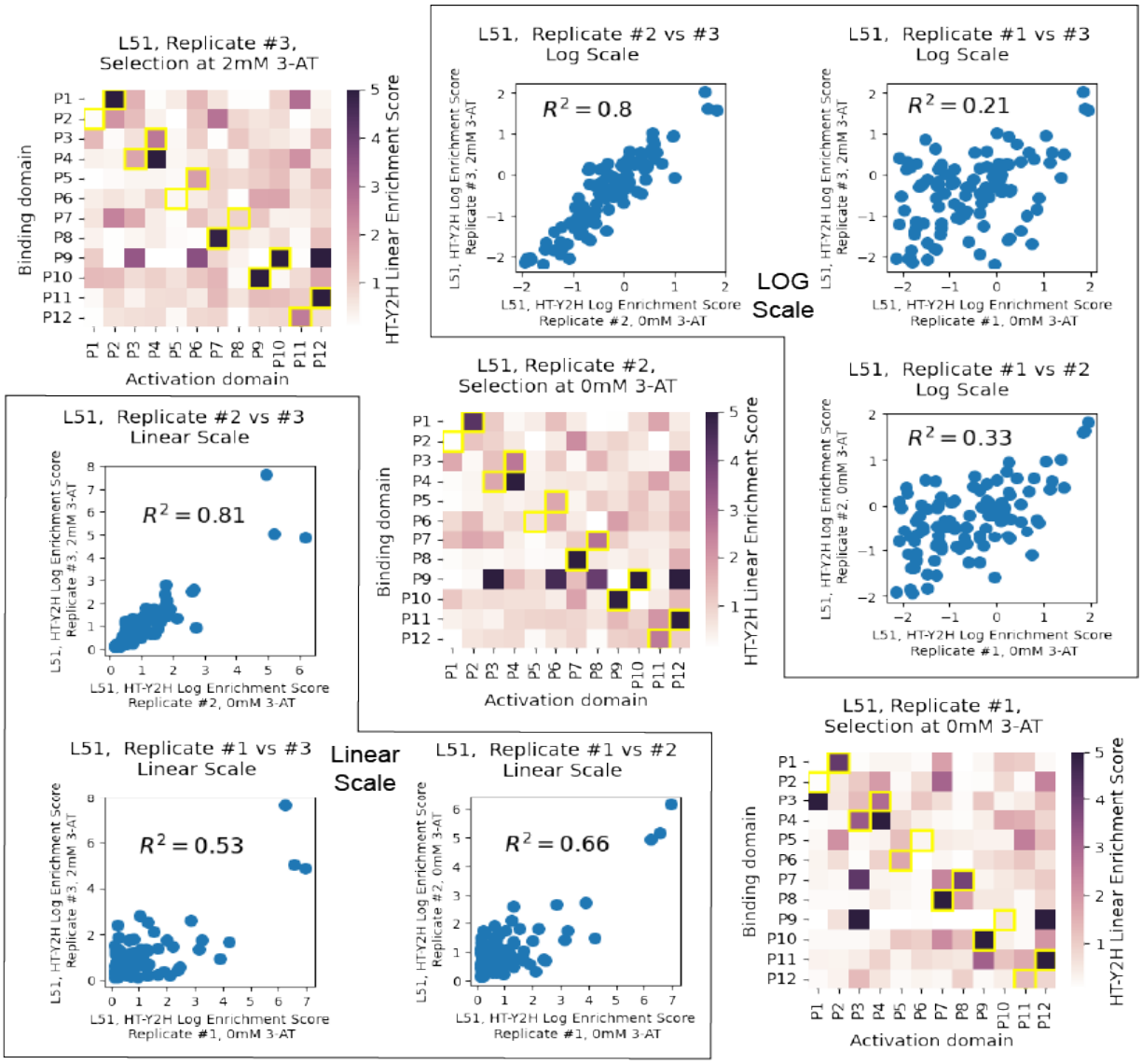

Figure 35. Interaction strength heatmap of the first independent replicate is shown in the left lower corner (labeled as Replicate #1). The second independent replicate is in turn split into two technical replicates labeled as Replicate#2 and #3 which are shown in the center and in the right upper corner. Very high correlation between linear and log enrichment values of these technical replicates can be observed in the corresponding scatterplots. Moderate correlation was observed between either of these two technical replicates and the first independent replicate of the experiment.

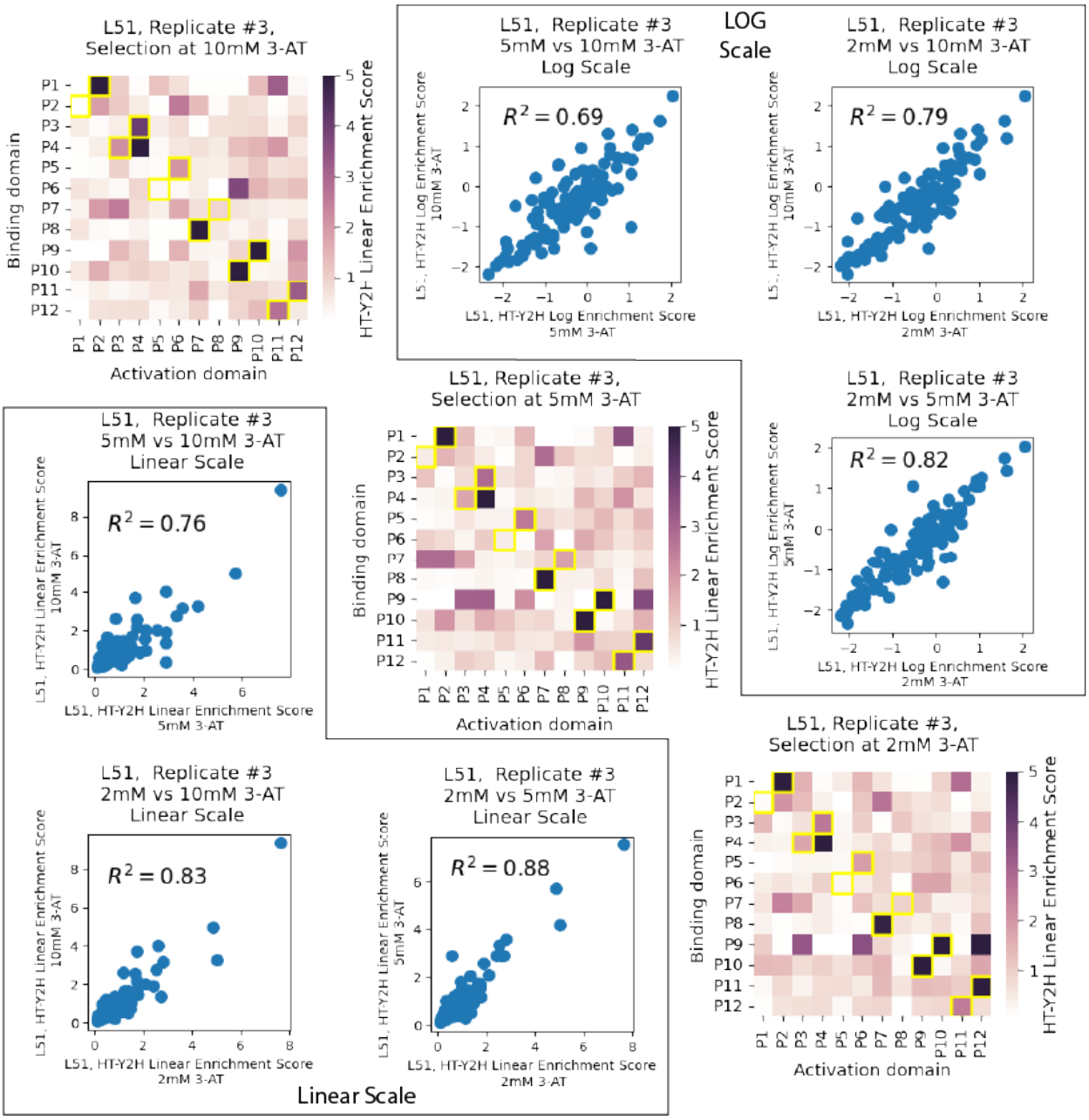

Figure 36. The interaction strength heatmaps for the third technical replicate of the experiment that was started from the glycerol stock made previously. The experiment was conducted at three different concentrations of 3-AT: 2mM, 5mM and 10mM of 3-AT. Both qualitative and quantitative comparisons show great agreement between different concentrations of 3-AT. It can be seen that our HT-Y2H assay correctly picks up almost on-target interactions in at least one orientation. It can also be noticed that, with one exception, only the on-target interactions have higher enrichment score than non-target interactions, although admittedly the number of unanticipated off-target interactions is present in the heatmap.

We hypothesized that the off-target interactions might have appeared due to some artifacts in IDT's oligo pool synthesis process. To investigate this hypothesis we decided to use a few

remaining aliquots of cells obtained before and after the selection process, and prepare and sequence amplicons that contained the full sequence of the coiled coils. Thanks to the short length of 33 amino acids of the coiled coils this longer sequencing amplicon was well shorter than 1000nt and thus amenable to Illumina sequencing.

Experiment L52. Characterizing interactions between shorter 4-heptad long DEN-designed coiled coils.

After discovering in L50 experiment that 10-heptad long coiled coils exhibited much stronger interaction than the shorter coils with reduced degree of specificity, Johannes Linder's DEN algorithms was adjusted to design a new set of 16 4-heptad long coiled coils, that we designed to maximize their interaction strength with just one other coiled coil in the set and minimize the interaction strength with all the remaining coils, thus forming an orthogonal 8-pairs coiled coil system.

Similarly to L51 experiment, to reduce the cost of the experiment the DEN designed coiled coils' sequences were ordered in the form of oligo pools from IDT Biosciences with approximately 30% accuracy rate for every sequence in the pool. Similarly to L51 experiment, an extra PCR handle is added 5' of the 80% GC-rich sequence on all oligos in one of the pools. Double stranded stock of L51 P1-P12 binders was added to the transformation mixture for the control purposes to have another look at P1-P12 interaction.

The resulting interaction strength heatmap for P1-P12 coiled coils and DEN designed coiled coils is shown in Figure 37. Similarly to L51 experiment, while our assay picks up all on-target interaction between P1-P12 coiled coils in at least one orientation, a significant number of

unanticipated interactions is present in the heatmap. Out of all DEN signed coiled coils only one pair S2A-S2B exhibited interaction.

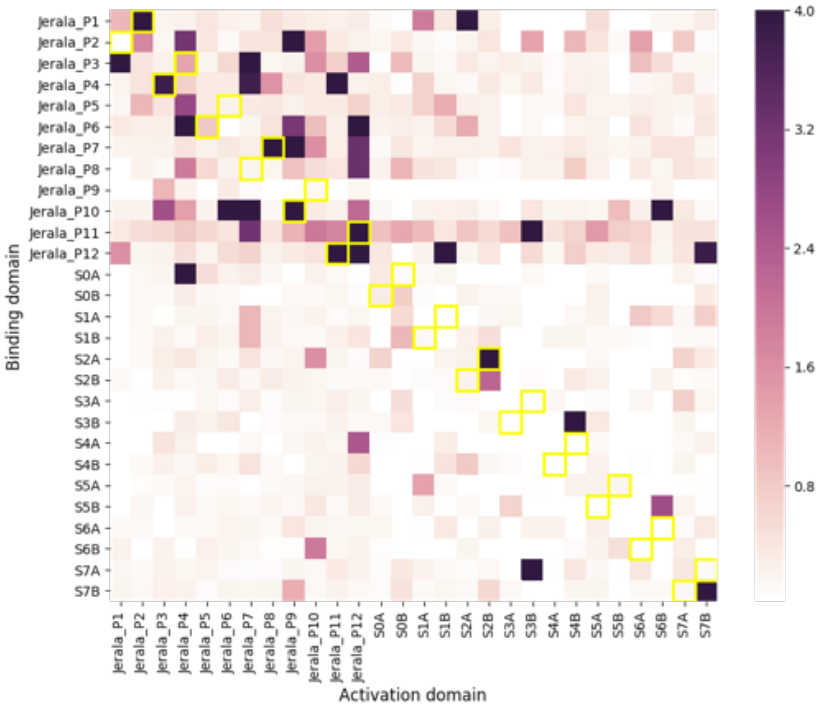

Figure 37. Interaction strength heatmap for P1-P12 coiled coils and DEN designed coiled coils. Similarly to L51 experiment, while our assay picks up all on-target interaction between P1-P12 coiled coils in at least one orientation, a significant number of unanticipated interactions is present in the heatmap. Out of all DEN signed coiled coils only one pair S2A-S2B exhibited interaction.

Experiment L54. Revising a protein linker between the binding domain and its fused protein, and using this revised design to screen P1-P12 coiled coils again.

The “input” sample in this experiment was sequenced with a good sequencing depth, i.e. it received approximately 400000 reads overall which translates to approximately 2800 reads per pair at library size of 144 pairs.

To simplify the previous design and remove the extra PCR handle, we used different codons in the 80% GC-rich protein linker reducing the GC content to 63% which was more amenable to PCR (Figure 38).

Although the resulting interaction heatmap (Figure 39A) shows that our assay still correctly identifies on-target interactions in at least one orientation in the current L54 experiment after codon-optimizing the 80% GC-rich linker, a significant amount of unanticipated interactions is observed. Low correlation with L51 experiment seen in the scatterplots is mainly due to different off-target interactions appearing in this L54 experiment compared to L51 experiment.

We also performed three technical replicates of L54 experiment at three (2/10/20mM) concentrations of 3-AT. Consistent with previous findings, our method's results exhibit very high correlation between different technical replicates of the experiment (Figure 40).

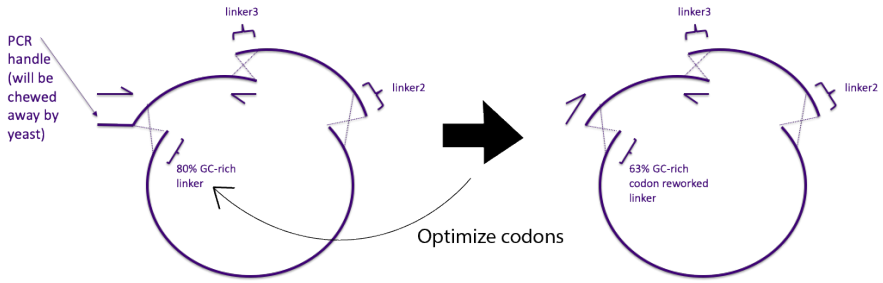

Figure 38. Optimizing codons of the 80% GC-rich sequence transformed this sequence to 63% GC-rich which was amenable to PCR and thus allowed to drop the use of the extra PCR handle.

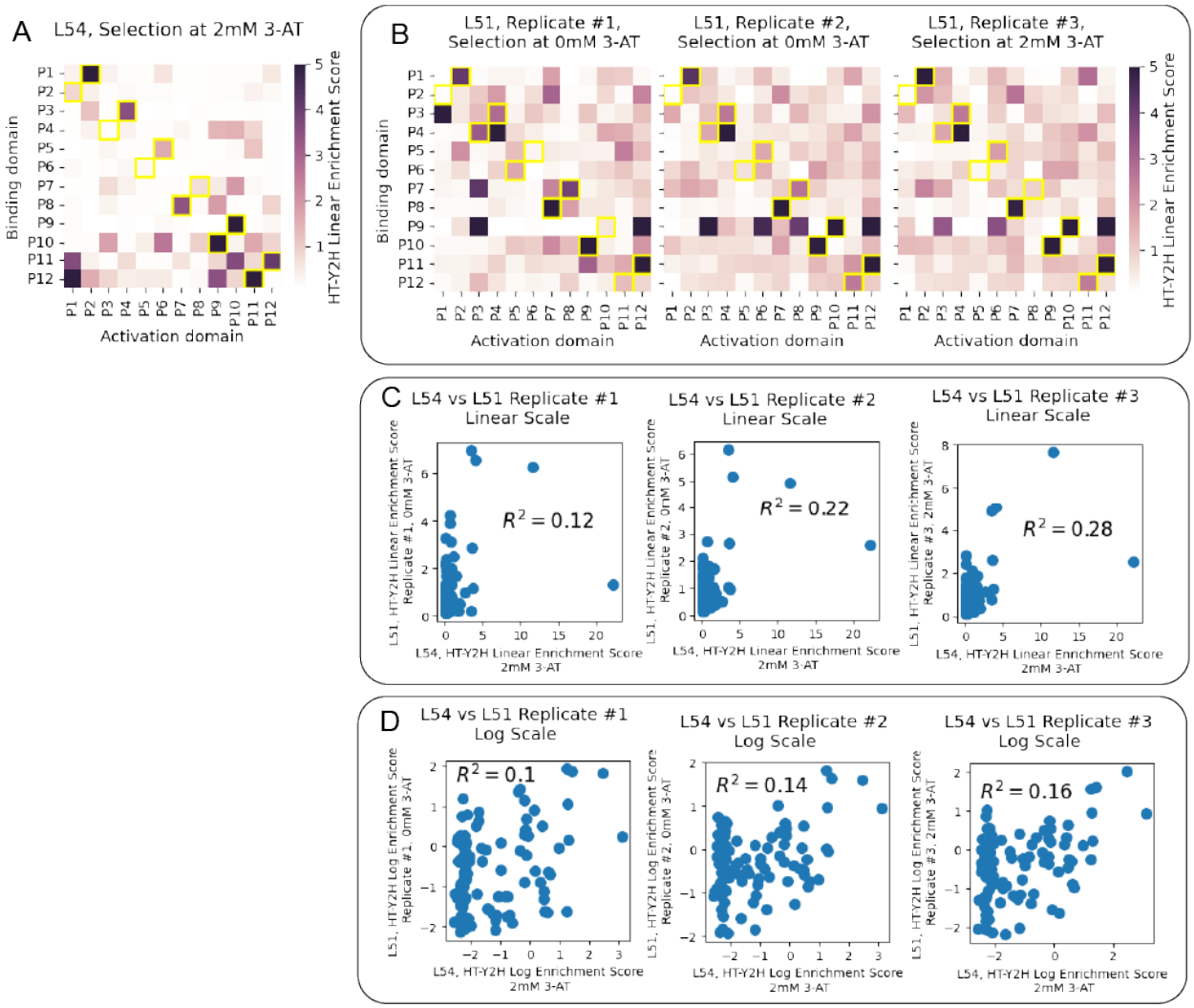

Figure 39. Comparing the results of L54 experiment that used codon optimized linker with the results of L51 experiment that used an extra PCR handle during plasmid assembly. Although in this current L54 experiment our assay correctly identifies on-target interactions in at least one orientation, a significant amount of unanticipated interactions is observed. Low correlation with L51 experiment seen in the scatterplots is mainly due to different off-target interactions appearing in this L54 experiment compared to L51 experiment.

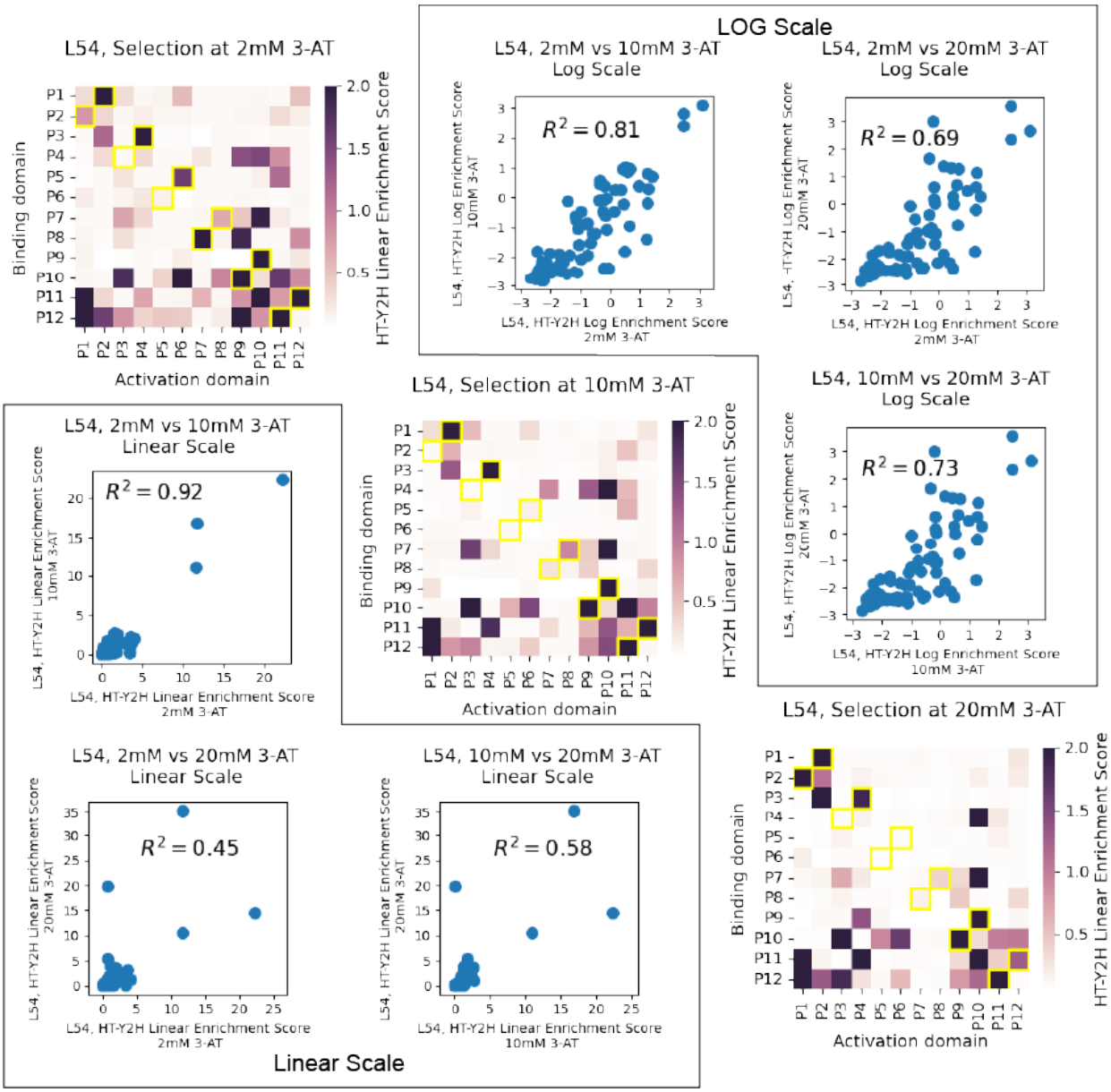

Figure 40. Technical replicates of L54 experiment performed at three (2/10/20mM) concentrations of 3-AT. Consistent with previous findings, our method's results correlate really great between different technical replicates of the experiment.

Experiment L55. Characterizing interactions between 4-heptad long DEN-designed coiled coils from experiment L52 using revised protein linker between the binding domain and its fused protein.

In a separate experiment we decided to use the codon-optimized linker to again screen the library DEN generated coiled coils. We again used IDT DNA synthesized oligo pools to prepare double stranded DNA fragments encoding the peptides. We sequenced the "input" and the "output" samples with a good sequencing depth of approximate 3000 reads per pair in each sample (772K reads for the "input" sample at library size of 16*16=256 combinatorial pairs).

As can be seen in the resulting interaction heatmap in Figure 41, a few on-target pairs have been confirmed by the experiment, including S2A-S2B pair previously identified as interacting in L52 experiment in the same orientation.

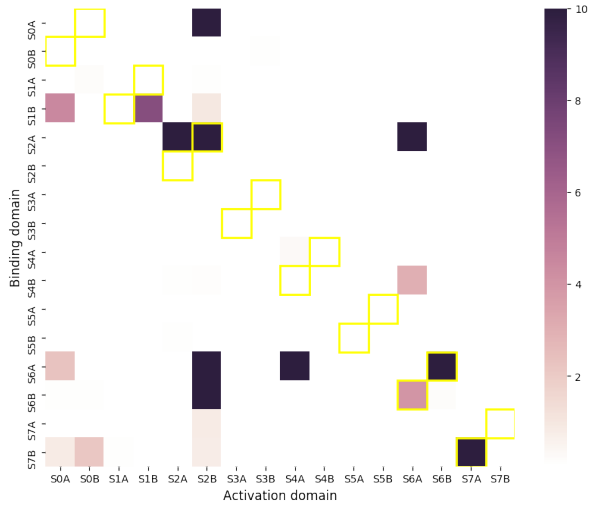

Figure 41. Interaction strength heatmap for 4-heptad long DEN generated coiled coils from experiment L52. The colorbar indicates the linear enrichment score.

L61. Characterizing interactions between orthogonal coiled coils P1-P12 using the revised protein linker and using Twist Biosciences as a supplier for double stranded DNA gene fragments encoding P1-P12 sequences.

Hypothesizing that the high error ratio during the IDT's synthesis process of single-stranded oligonucleotide pool could be responsible for the unanticipated off-target interaction appearing in the L51, L52, L54 experiments, we decided to order every coiled coil sequence (P1-P12) from a different vendor (Twist Bioscience) in the form of a double stranded DNA fragment with more >98% of synthesized sequences promised to be correct. The fragments were then individually amplified in separate tubes in PCR reactions which were subsequently analyzed on a Qiagen QIAxcel capillary electrophoresis fragment analyzer to ensure correct length of the PCR product. The resulting PCR aliquots were then pooled together in a single 1.5mL Eppendorf tube and were cleaned using Kapa Magnetic Beads. The purified mixture was then combined with the three linear fragments composing the backbone and transformed into the base y777 yeast strain. The standard selection process was applied to the transformed library. The resulting interaction strength heatmap is shown in Figure 42 along with the published data. As can be seen in the figure, after revisiting the design by placing the barcode and its PCR handle after the terminator and ordering accurately synthesized gene fragments for P1-P12 binders, our assay achieves nearly ideal agreement with the published data.

Figure 43 and Figure 44 show how significantly reads counts change for the on-target pairs.

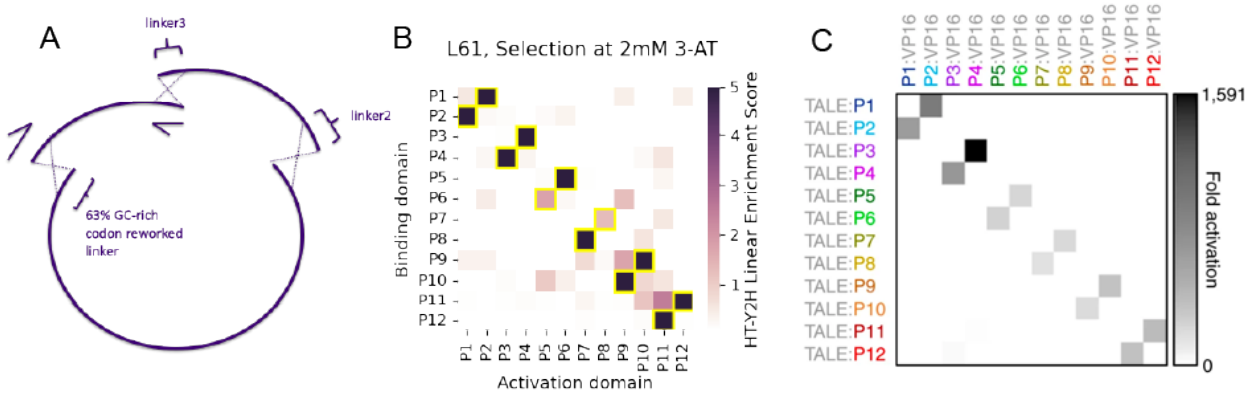

Figure 42. (A)this L61 library uses codon-optimized linker that doesn't require an extra PCR handle like. (B) There is now very good visual agreement between Interaction strength heatmaps obtained using our assay and using luciferase assay in (Lebar et al., 2020).

```
In [11]: trp_48h_df
```

Out[11]:

Binder1	Jerala_P1	Jerala_P2	Jerala_P3	Jerala_P4	Jerala_P5	Jerala_P6	Jerala_P7	Jerala_P8	Jerala_P9	Jerala_P10	Jerala_P11	Jerala_P12
Jerala_P1	14	1356	3068	1644	2895	4762	3556	2271	2007	4838	2468	2808
Jerala_P2	571	48	3980	2330	5504	3296	4359	4395	1814	3133	3345	3622
Jerala_P3	1615	3416	1540	269	2565	1728	2176	1716	2965	3781	1641	1691
Jerala_P4	2322	2577	1484	9	4121	4917	3106	2262	4152	4647	2266	3052
Jerala_P5	3203	4294	3212	2329	1320	1203	3083	2649	2917	2038	2945	3141
Jerala_P6	2453	2679	2999	1449	1663	23	3166	2849	1546	2524	2376	1919
Jerala_P7	1688	2073	1745	1048	3027	2176	1564	2019	2028	1334	2102	2541
Jerala_P8	3013	3599	2486	2153	3406	2744	2071	3	3133	1600	2190	3965
Jerala_P9	2816	2271	4845	1896	7351	3432	2664	4258	7	1339	2204	3540
Jerala_P10	4259	5282	6396	1905	5595	5601	5186	3338	1561	8	2959	7001
Jerala_P11	1736	3909	2653	1747	2224	2725	1566	2400	2218	1520	75	1073
Jerala_P12	3432	4263	3173	1707	3629	3169	3227	3097	2249	3112	1503	3174

Figure 43. Number of sequencing reads mapping to each pair in the "input" sample (before the selection). Highlighted numbers indicate on-target interactions. As can be seen, the read counts are uniform implying consistent library assembly. 8 out of 12 homo-dimers along the diagonal show very low counts indicating issues with the assembly of the plasmids encoding the same protein fused both to the AD and DBD. The homo-dimer plasmid assembly issue is attributed to increased degree of homology of the two linear fragments encoding the same protein.

In [9]: `his_2mM_3AT_df`

Out[9]:

Binder1	Jerala_P1	Jerala_P2	Jerala_P3	Jerala_P4	Jerala_P5	Jerala_P6	Jerala_P7	Jerala_P8	Jerala_P9	Jerala_P10	Jerala_P11	Jerala_P12
Jerala_P1	9.0	62934.0	215.0	129.0	251.0	381.0	281.0	171.0	951.0	250.0	144.0	1334.0
Jerala_P2	31951.0	10.0	599.0	211.0	1073.0	1184.0	341.0	363.0	138.0	253.0	247.0	349.0
Jerala_P3	117.0	241.0	158.0	11454.0	192.0	110.0	141.0	120.0	175.0	256.0	95.0	116.0
Jerala_P4	174.0	713.0	79476.0	3.0	341.0	347.0	236.0	193.0	360.0	966.0	1524.0	268.0
Jerala_P5	221.0	280.0	230.0	191.0	204.0	21659.0	347.0	196.0	257.0	135.0	871.0	283.0
Jerala_P6	166.0	1363.0	230.0	119.0	3625.0	6.0	244.0	193.0	2437.0	159.0	196.0	140.0
Jerala_P7	122.0	117.0	135.0	80.0	221.0	157.0	144.0	3266.0	141.0	124.0	1348.0	201.0
Jerala_P8	238.0	276.0	189.0	171.0	243.0	259.0	29996.0	NaN	260.0	1096.0	176.0	370.0
Jerala_P9	1140.0	866.0	308.0	102.0	450.0	196.0	2462.0	906.0	15.0	69496.0	137.0	219.0
Jerala_P10	326.0	415.0	1021.0	160.0	7314.0	2544.0	418.0	280.0	59826.0	8.0	2000.0	604.0
Jerala_P11	190.0	383.0	291.0	262.0	258.0	259.0	478.0	273.0	278.0	2162.0	231.0	8068.0
Jerala_P12	264.0	289.0	284.0	113.0	255.0	220.0	461.0	217.0	144.0	496.0	35130.0	353.0

Figure 44. Number of sequencing reads mapping to each pair in the "output" sample (after the completion of selection). Highlighted numbers indicate on-target interactions and are noticeably higher other numbers corresponding to non-interacting pairs.

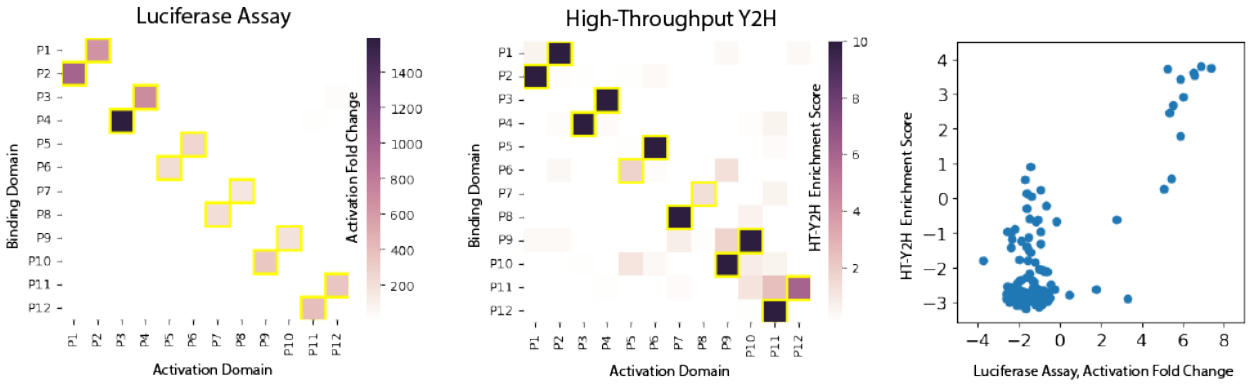

Figure 45. (A) Heatmap of activation fold change values obtained using luciferase assay, linear scale, data taken from (Lebar et al., 2020). (B) Heatmap of HT-Y2H enrichment scores, linear scale. (C) Pearson correlation between luciferase assay's activation fold change values and HT-Y2H enrichment values, log scale.

Since L61 is similar to L54 library and uses the same codon optimized glycine-serine linker but is synthesized in the form of a double stranded DNA gene fragment by Twist Bioscience whereas L54 is synthesized as an oligo pool by IDT DNA Technologies, it would be interesting to compare how two experiments correlate with each other. In the Figure 46 It can be seen that there's moderate correlation (R^2=*0.29* in the log scale) between the two experiments although L61 has significantly less unanticipated interactions.

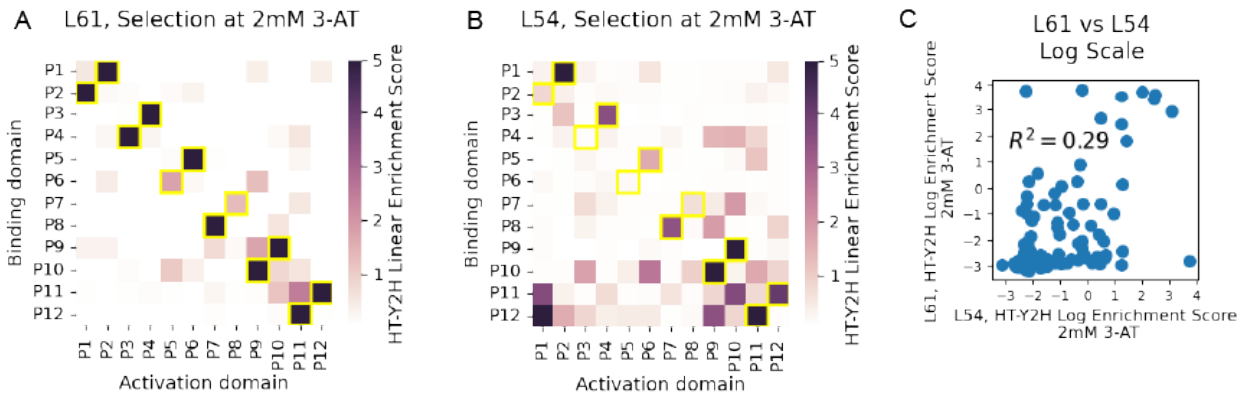

Figure 46. Comparing interaction heatmaps of L61 (A) and L54 (B) libraries that use the same codon-optimized linker but different types and suppliers of DNA to prepare the fragments encoding the proteins in the library (Twist Bioscience's dounble stranded gene fragments vs IDT DNA's single stranded oligo pools). (C) There is still moderate correlation between log enrichment values of these libraries.

Experiment L62. Characterizing interactions between orthogonal coiled coils P1-P12 using the previous version of the protein linker between the binding domain and its fused protein, and using Twist Biosciences as a supplier for double stranded DNA gene fragments encoding P1-P12 sequences.

Having achieved great success using high accuracy double stranded DNA gene fragments synthesized by Twist Biosciences in combination with codon-optimized linker, we decided to test whether we would still achieve great agreement with the published data when we use double stranded gene fragments from Twist Biosciences and the older 80% GC-rich linker that requires an additional PCR handle.

For this experiment we ordered a set of double stranded gene fragments from Twist Biosciences encoding proteins that will be fused to the 80% GC-rich linker, containing that linker (Figure 47A) and performed the experiment starting from the transformation stage.

The resulting heatmap revealed anticipated on-target interactions in at least one orientation (Figure 47B) and still exhibited good visual agreement with the published luciferase assay data. Compared to L61 experiment there were stronger unanticipated interactions present, albeit compared to the IDT DNA based experiment L51, there were fewer such interactions.

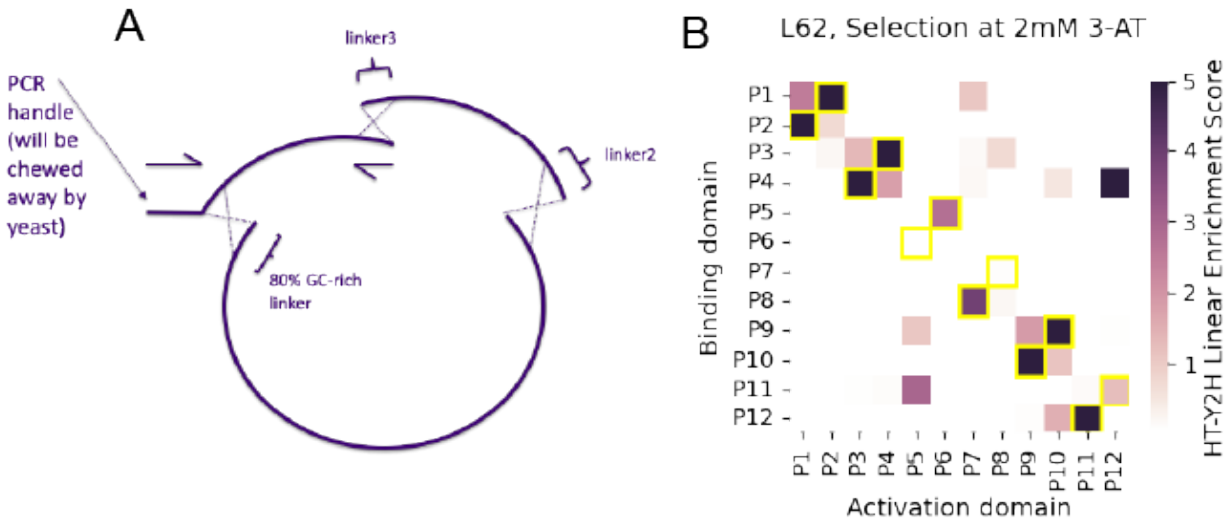

Figure 47. (A)this L62 library uses the 80% GC-rich linker that requires an extra PCR handle like in L51 libraries. (B) There is still very good visual agreement between Interaction strength heatmaps obtained using our assay and using luciferase assay in (Lebar et al., 2020).

Going forward, we decided to use codon-optimized 63% GC-rich linker that doesn't require additional PCR handles in all the subsequent experiments. We also use Twist Biosciences double stranded gene fragments in all the subsequent experiments due to better results achieved using their DNA compared to IDT DNA products.

Experiment L63. Screening intraviral SARS-CoV2 PPIs.

SARS-CoV2 virus has upended and ravaged our society in the last two years. Tremendous amounts of effort were devoted to the characterization of the mechanism of action of the virus including characterization of protein-protein interactions between the viral proteins. Several

papers have been published applying different methods such as low-throughput Y2H, co-immuno precipitation to study the intraviral PPIs (Springstein et al., 2021), (Jiang et al., 2021), (Liu et al., 2021), (Gordon et al., 2020), (M. Chen et al., 2021), (Xu et al., 2021). Since some of the later published paper were revealing previously unknown interactions, it is really important to characterize the intraviral interactions using a few different methods independently, that's why we decided to select a few SARS-CoV2 proteins for which interactions have been characterized using low-throughput Y2H and co-immunoprecipitation in (Li et al., 2021) and use our assay to characterize the interactions between those proteins in a all-against-all manner. The resulting interaction heatmap can be seen in Figure 48 does not quite agree with the published data besides in the case of ORF9 homodimer. Since some of the interaction revealed in the original paper using LT-Y2H method have not been confirmed by CO-IP method and vice versa, additional characterization of those interactions might be needed using other methods. Another explanation for the disagreement between our assay's results and that in the paper might be due to the fact that that interaction that HT-Y2H reveals might be too strong and just eclipse other interactions that are deemed as non-interacting due to having linear enrichment score lower than 1.0, but in screened using low-throughput version of our assay they might be exhibiting good interaction. The latter issue might be resolved by adding a set of control proteins with a different range of dissociation constants Kd which will be done in the subsequent L68 library.

Performing another technical replicate of the experiment at 10mM of 3-AT (Figure 49) confirmed the same interactions revealed by the first technical replicate at 2mM 3-AT, assuring that there were no artifacts in the first technical replicate of the experiment.

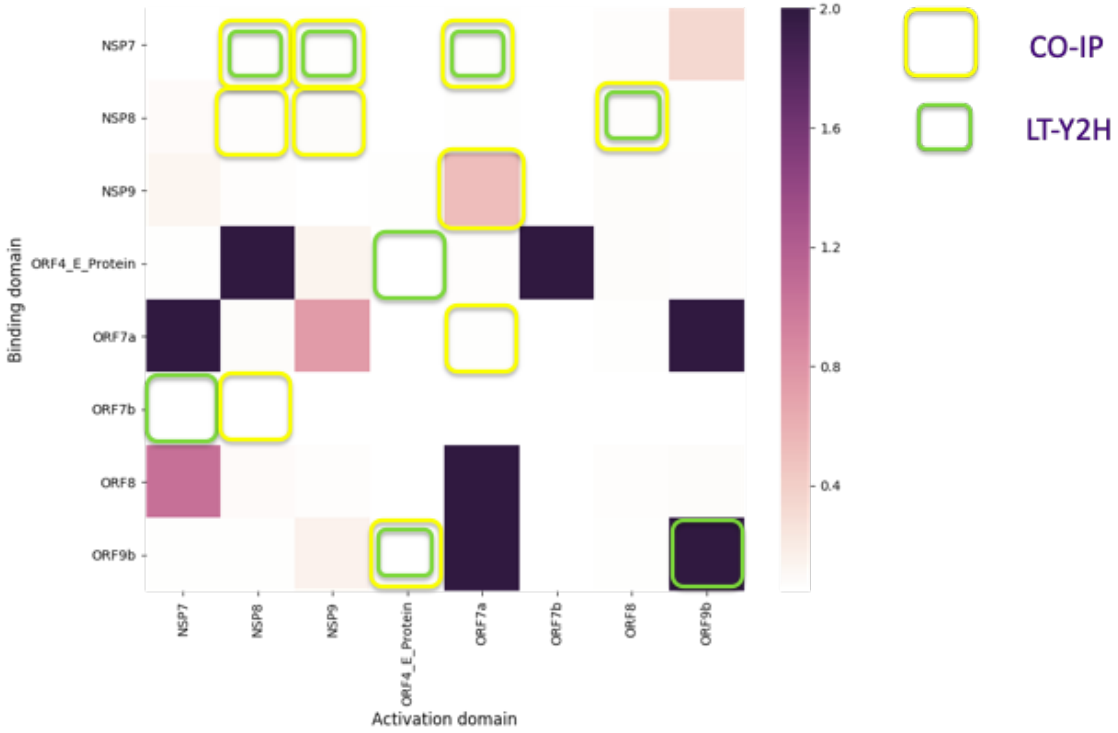

Figure 48. Interaction strength heatmap obtained using our HT-Y2H assay. Green and yellow highlighted squares indicate interactions revealed using either low-throughput Y2H or co-immunoprecipitation methods in (Li et al., 2021).

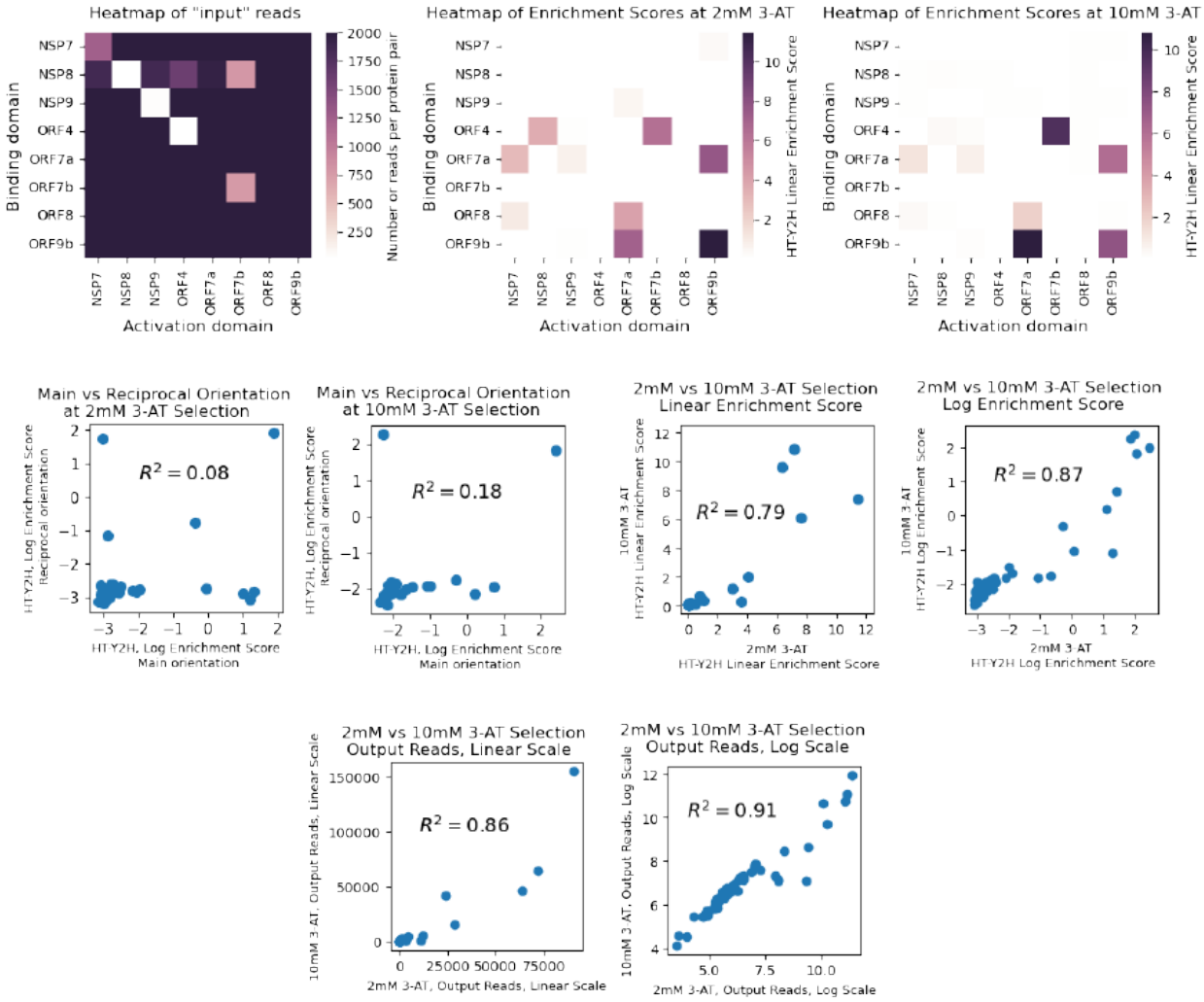

Figure 49. Heatmap of the depth of sequencing shows that besides poorly assembled homo-dimers almost all other pairs have successfully assembled during the transformation and had more than 1000 reads per pair in the sequencing run which deems them "high-quality" pairs. The selection was performed at two different concentrations of 3-AT (2mM and 10mM), and good visual agreement is seen between the two resulting interaction heatmaps which implies that our assay was performed consistently.

Experiment L64. Screening previously characterized control ZCON heterodimers using the revised design and TwistBioscience's double stranded DNA gene fragments.

Having achieved nearly ideal agreement between the results of screening of coiled coils P1-P12 designed in the Jerala lab, we decided to go back and use the revised "barcode following terminator" design to screen previously characterized and published 9 pairs of control ZCON hydrogen bond network hetero-dimers. This set of proteins has already been screened before using "barcode preceding terminator" plasmid design (L42) but our assay failed to confirm four pairs ZCON15/37/37_3/150 as interacting ones.

In this current L64 experiment binders' sequences with pre-defined 20nt long DNA barcodes were ordered from Twist Bioscience in the form of gene fragments. They were then individually amplified in separate tubes in PCR reactions which were subsequently analyzed on a Qiagen QIAxcel capillary electrophoresis fragment analyzer to ensure correct length of the PCR product. The resulting PCR aliquots were then pooled together in a single 1.5mL Eppendorf tube and were cleaned using Kapa Magnetic Beads. The purified mixture was then combined with the three linear fragments composing the backbone and transformed into the base y777 yeast strain. The standard selection process was applied to the transformed library. The resulting interaction strength heatmap is shown in Figure 50 and Figure 51 along with the previously obtained results in L42 experiment using "barcode preceding terminator" design and the published low-throughput Y2H data. As can be seen in the figures, after revisiting the design and placing the barcode and its PCR handle after the terminator our assay achieves nearly ideal agreement with the published data predicting the same interacting pairs in at least one orientation (ZCON15/37/150/155/) or both orientations (ZCON13/37_3/154/162).

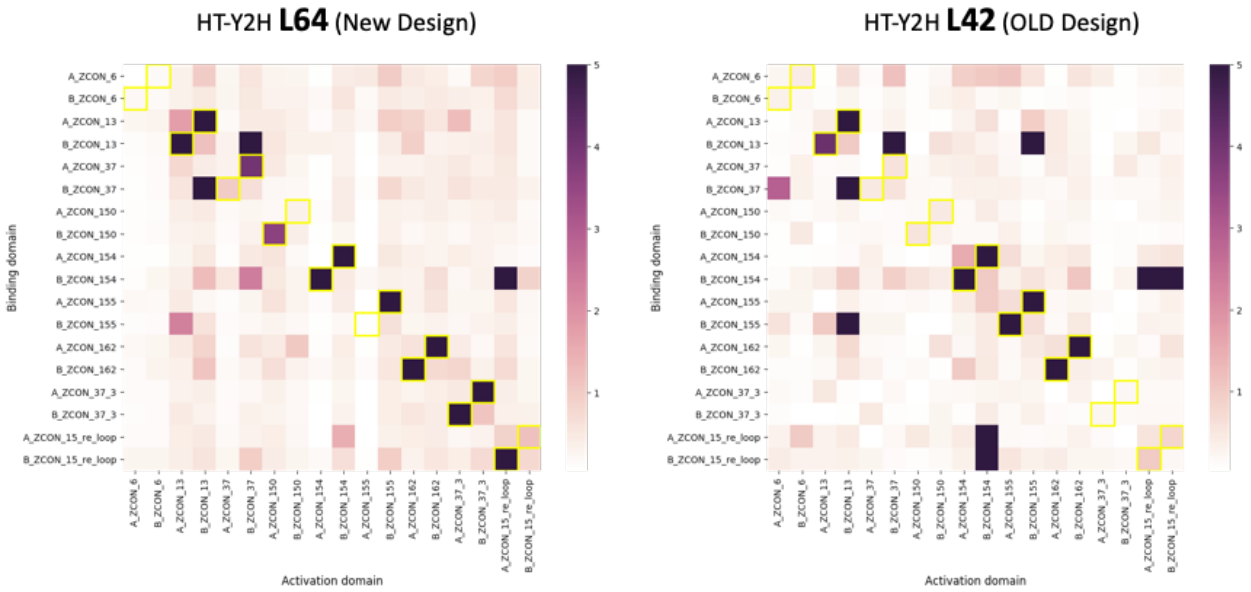

Figure 50. Comparing the new "barcode following terminator" design and the old "barcode preceding terminator" design. As can be seen from the L64 heatmap (left) the revised design allows to correctly identify all on-target ZCON interactions in at least one orientation significantly improving the performance of our assay compared with the previous "barcode preceding terminator" design used in L42 experiment (right).

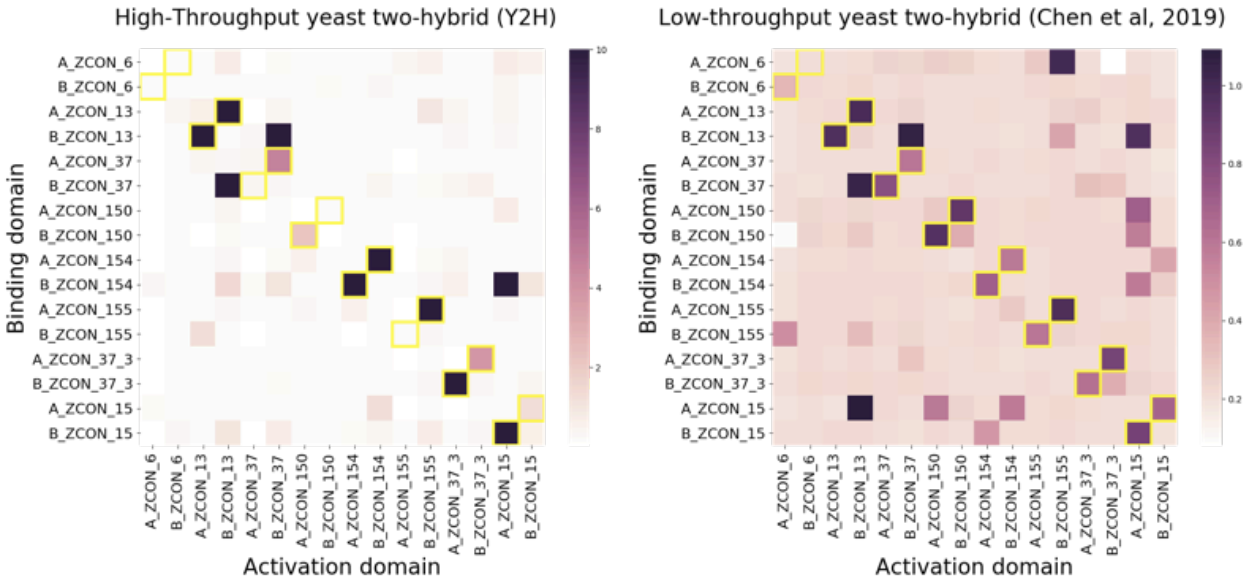

Figure 51. Comparing interaction strength heatmaps obtained using the revised "barcode following terminator" design of our HT-Y2H and using low-throughput Y2H assay from (Z. Chen et al., 2019). Nearly ideal agreement between these two methods can be observed.

Experiment L65. Characterizing interactions between 4-heptad long DEN-designed coiled coils from L52 experiment using the revised design and TwistBioscience's double stranded DNA gene fragments.

This library was sequenced with a good sequencing depth with each of the "input" and "output" samples received about 450K reads, which translates to approximately 1800 reads per pair with the library size of 16*16=256 possible combinatorial pairs.

Having achieved much better agreement between our HT-Y2H assay's results and the published data for coiled coils P1-P12 after using double stranded DNA gene fragments from Twist Bioscience instead of single stranded oligo pools from IDT for the library assembly, we wondered whether the same success can be achieved for DEN designed 4-heptad long coiled coils from L55 library.

The result of the library screening using Twist Bioscience's gene fragments can be seen in the Figure 52. As Figure 52A shows the library assembly was uniform and, besides the known issues of poor assembly of homodimers expressing plasmids, all other pairs received good sequencing coverage. Figure 52B and Figure 53A show good qualitative and quantitative agreement between the two technical replicates of the experiment.

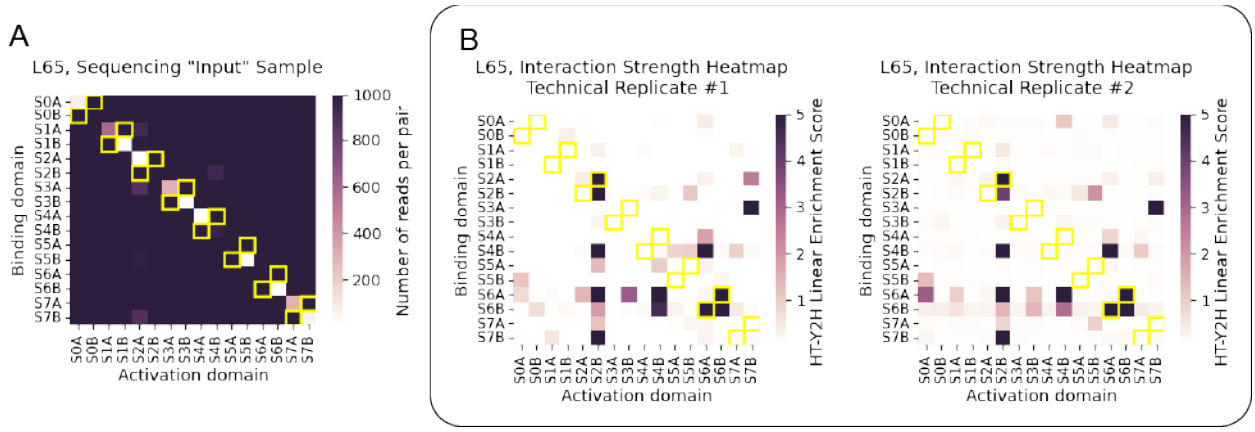

Figure 52. (A) A heatmap of reads per pair for the "input" sample representing the proxy to the homogeneity of the library assembly. Similarly to previous experiments, the heatmap reveals poor assembly of the plasmids encoding pairs of homo-dimers. Heterodimers encoding plasmids assembled uniformly and received good sequencing coverage. (B) There is a good visual agreement between the interaction strength heatmaps of both technical replicates of the experiment.

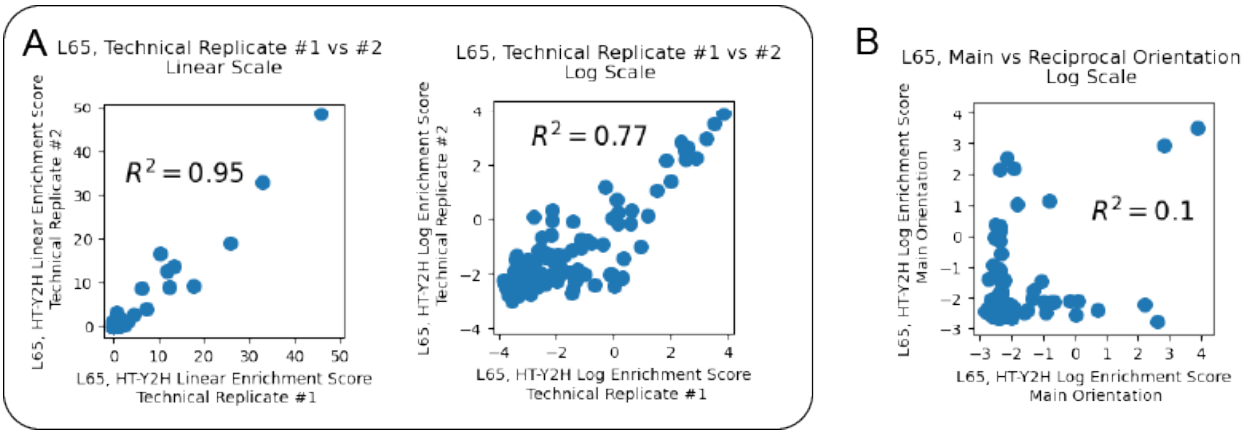

Figure 53. (A) Great degree of correlation both in linear and log scale was observed between the both technical replicates started from the same glycerol stock of the transformed library. The replicates were selected using separate flasks with all other steps in the protocol being identical (B) A low correlation between the log enrichment score of the main and the reciprocal orientations means that most of the interacting pairs were observed in just one orientation.

Figure 54 below qualitatively and quantitatively compares the interaction strength heatmaps of the L65 library that used Twist Bioscience double stranded gene fragments in the library assembly and that of the L55 library that used IDT DNA single stranded oligo pool. Both libraries received good sequencing coverage. In both cases, our assay picks up the same on-target interactions: S2A-S2B (in one orientation) and S6A-S6B (in both orientation).

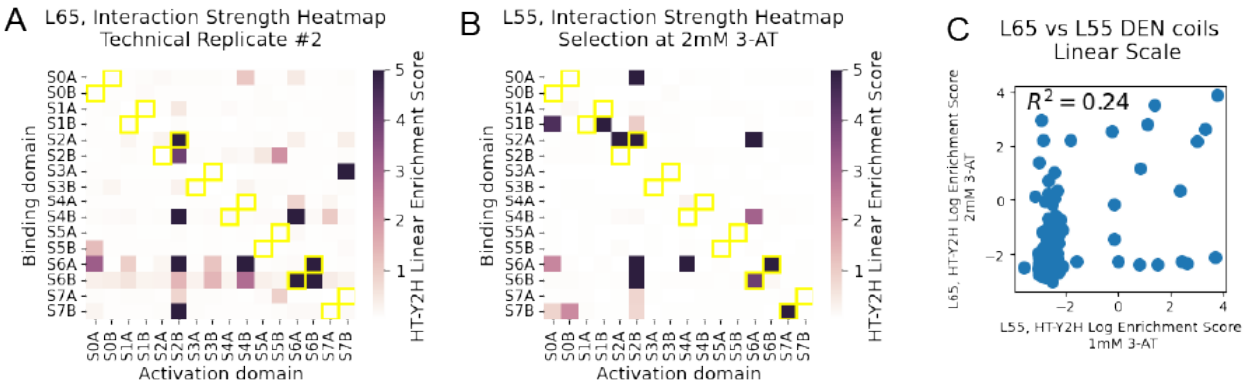

Figure 54. Qualitative and quantitative comparison of the results of screening the L65 (A) and L55 (B) libraries of DEN generated coiled coils. The L65 library was assembled using Twist Bioscience double stranded gene fragments and L55 was assembled using IDT DNA single stranded oligo pool. Modest correlation between the results can be observed, although in both cases the same three on-target pairs S2A-S2B, S6A-S6B, S6B-S6A were picked up by our assay.

Experiment L66. Characterizing interactions between interactions between BCL2 family proteins and their de novo designed inhibitors using the revised design and TwistBioscience's double stranded DNA gene fragments.

Having achieved nearly ideal agreement between our HT-Y2H data characterizing the interactions between coiled coils P1-P12 and hydrogen bond network containing hetero-dimers ZCON and the previously published data for those PPIs, we decided to additionally characterize the interactions between BCL2 homologs and their de novo designed inhibitors using the new "barcode following terminator" design and using high-quality gene fragments synthesized by Twist Bioscience. Similarly to how it was done in L61, L62, L63, L64 and L65 experiments, the gene fragments for the above binders along with their terminators, PCR handles and barcodes were ordered from Twist Bioscience, individually amplified in a PCR reaction, pooled together, fragment analyzed, purified using magnetic beads, and combined with the backbone fragments

and transformed into the base y777 yeast strain. The resulting transformed yeast library was subjected to standard selection protocol at 1mM 3-AT.

As can be seen in the Figure 55, the resulting heatmaps for either the main or reciprocal orientations show a good agreement with the published biolayer interferometry data (Figure 56) but a significant amount of strong unanticipated interactions was present in the heatmaps.

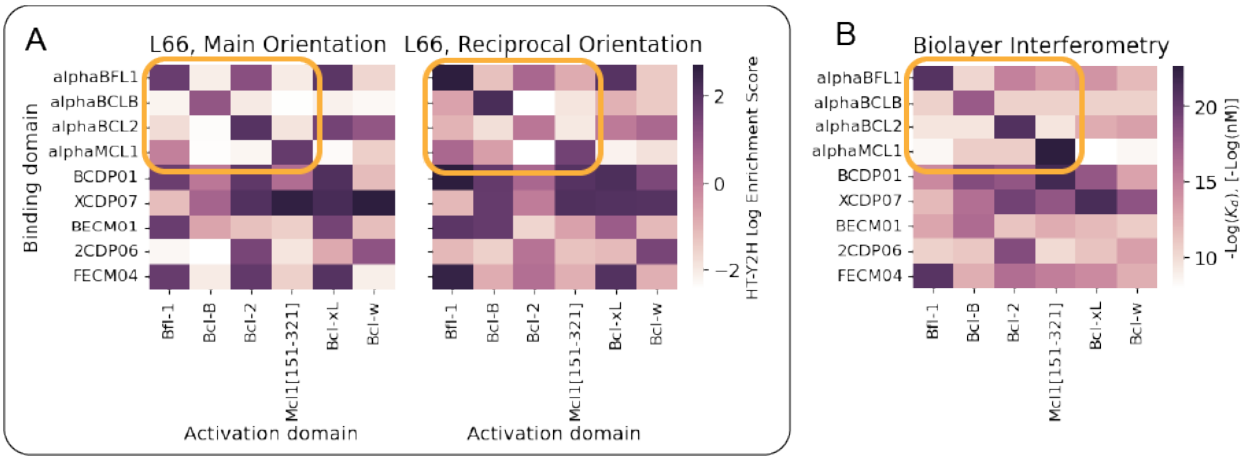

Figure 55. (A) Heatmap of HT-Y2H log enrichment scores of the main and reciprocal orientations. The highlighted region spotlights the set of interactions from the main figure in (Berger et al., 2016) that characterizes the interaction between BCL2 homologs and the their most specific de novo designed inhibitors.

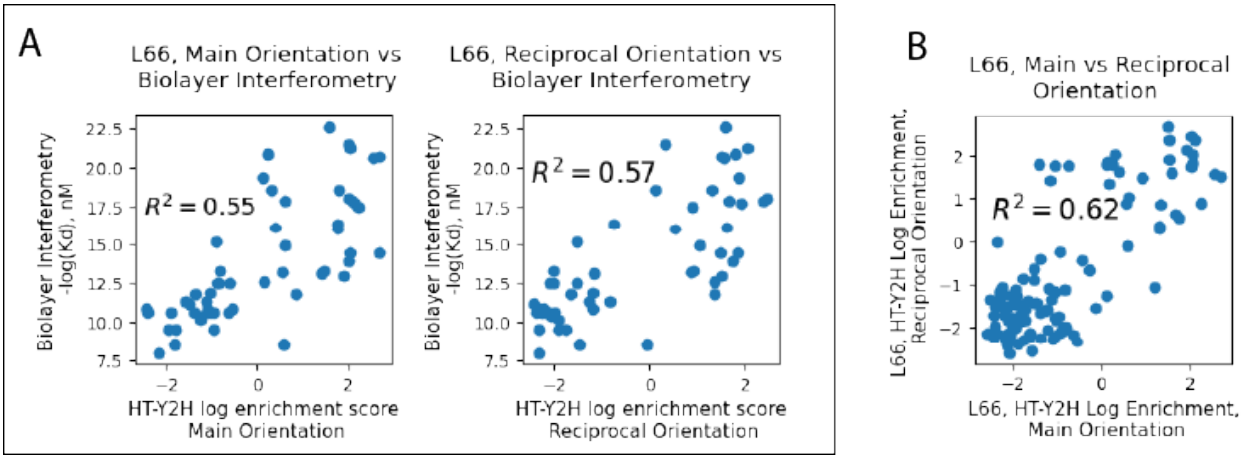

Figure 56. (A) We found a good correlation between the HT-Y2H main and reciprocal orientations' log enrichment values and the negative log values of the dissociation constant Kd obtained using biolayer interferometry in(Berger et al., 2016). (B) The correlation between the main and reciprocal orientations' log enrichment values of R^2=0.62 indicate good consistency of the HT-Y2H.

We also decided to compare if there would be an agreement between the old "barcode preceding terminator" design used to characterize the interaction between BCL2 de novo

designed inhibitors and their targets in the L33, L45, L48, L49 libraries and new “barcode following terminator” design. As can be seen in the Figure 57, we found a reasonably good correlation between L66 log enrichment scores and those from each of the L33, L45, L48, L49 with the Pearson correlation ranging from 0.31 between L66 and L49, and up to 0.57 between L66 and L33. It is worth noting that since the L49 library was significantly undersequenced with just 50 reads per protein pair, the real correlation between L66 and L49 is likely higher.

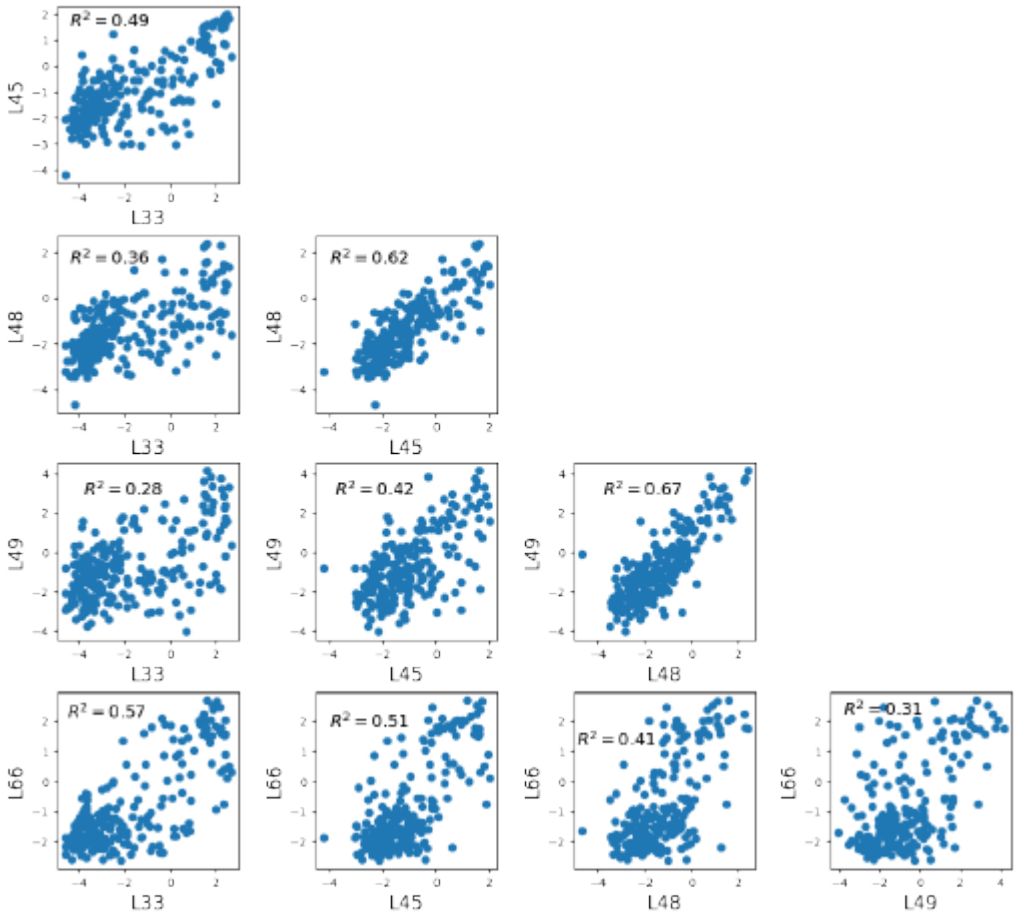

Figure 57. Comparing log enrichment values obtained in different experiments characterizing the interaction between BCL2 homologs and their de novo designed inhibitors. Interestingly, we found a good agreement between the old “barcode preceding terminator” design used in the L33, L45, L48, L49 experiments and the new “barcode following” terminator used in the L66 experiment, with the square of the Pearson correlation coefficient ranging from 0.31 between L66 and L49 and up to 0.57 between L66 and L33.

Experiment L67. Characterizing more coiled coils designed in the Baker lab using coiled coils P1-P12 as controls. Investigating whether using unique molecular identifiers (UMIs) affect the interaction picture.

In this experiment we are investigating whether using unique molecular identifiers (UMIs) allows to achieve even better agreement with the published data.

The protein library in this experiment is comprised of a few pairs of heterodimers previously designed in the Baker lab and characterized in (Rogers et al., 2014) combined with coiled coils P1-P12.

Sequencing amplicons generated in all the previous experiments were obtained as a result of 2 rounds of qPCR reaction. The first round of qPCR was performed till there was a sign of primer depletion which is indicated by an inflection point in the amplification curve, at which point the qPCR reaction would be stopped. Usually, about 15 cycles of qPCR was needed to reach this inflection point, although as low as 13 cycles and as high as 20 cycles were needed in some of the above experiments. To test whether such an approach introduces PCR bias and changes the interaction heatmap we decided to utilize unique molecular identifiers (UMIs) in the first stage of qPCR. In this approach 3 cycles of PCR in the first round were performed on plasmids extracted from both TRP and HIS samples to add 20nt long UMIs to PCR handles RL3 and RL4 followed by Illumina Read1 and Read2 sites. About 14 cycles of qPCR were performed in the second round and stopped at the inflection point.

Very high correlation between the “input” (TRP) sample’s read counts for L67 non-UMI and UMI samples in linear and log scale can be observed in the Figure 58 which implies that the use of UMIs will probably not affect the resulting interaction outcome.

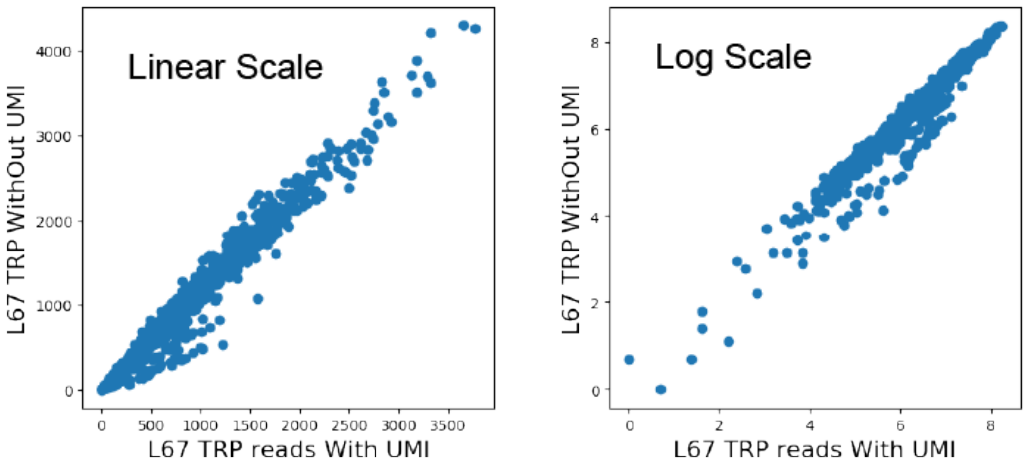

Figure 58. Comparing read counts for non-UMI and UMI containing "input" (TRP) samples.

Similarly, very high correlation between the "output" (HIS) sample's read counts for L67 non-UMI and UMI samples in linear and log scale can be observed in the Figure 59 which implies that the use of UMIs will probably not affect the resulting interaction outcome.

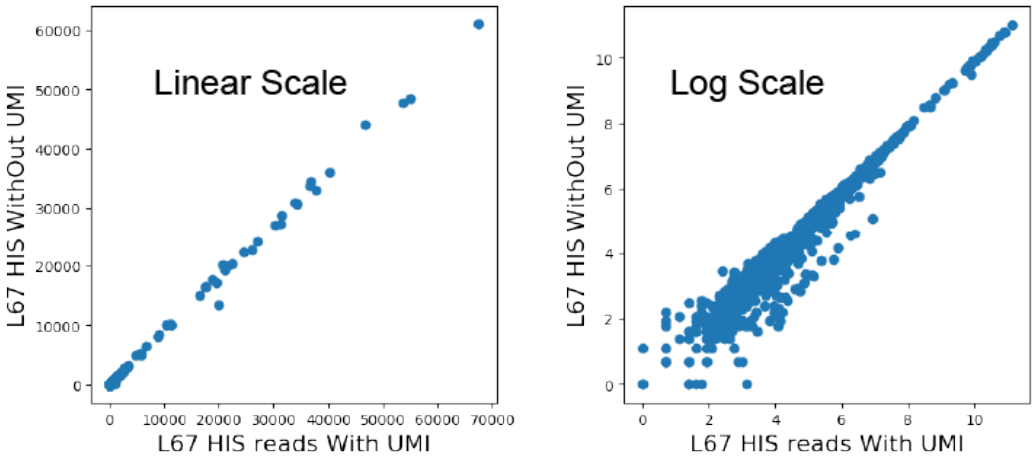

Figure 59. Comparing read counts for non-UMI and UMI containing "output" (HIS) samples.

As anticipated from visual inspection of Figure 58 and Figure 60 the resulting linear and log enrichment score of UMI and non-UMI samples had very high degree of correlation between each other (Figure 60). The corresponding heatmaps are shown in Figure 61.

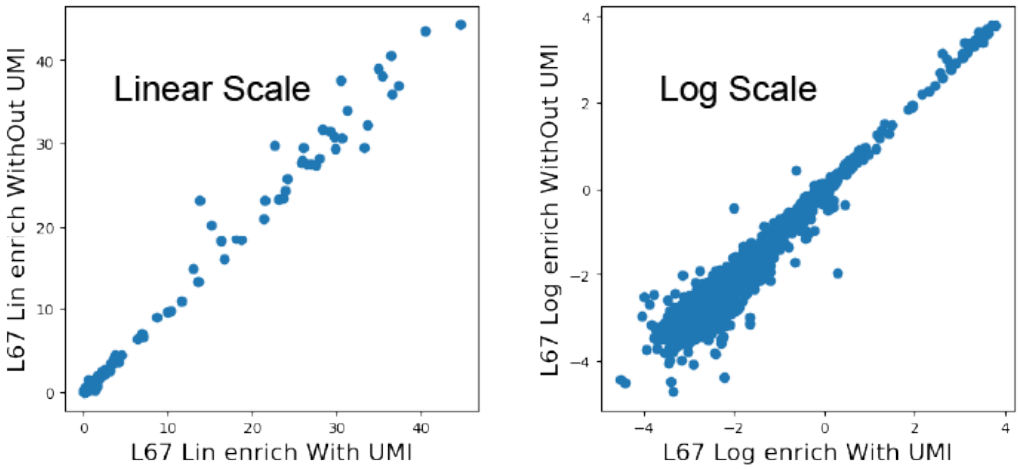

Figure 60. Comparing linear enrichment scores for non-UMI and UMI containing samples. These scores were obtained by diving "output" reads counts by the "input" ones (or by diving HIS counts by TRP counts).

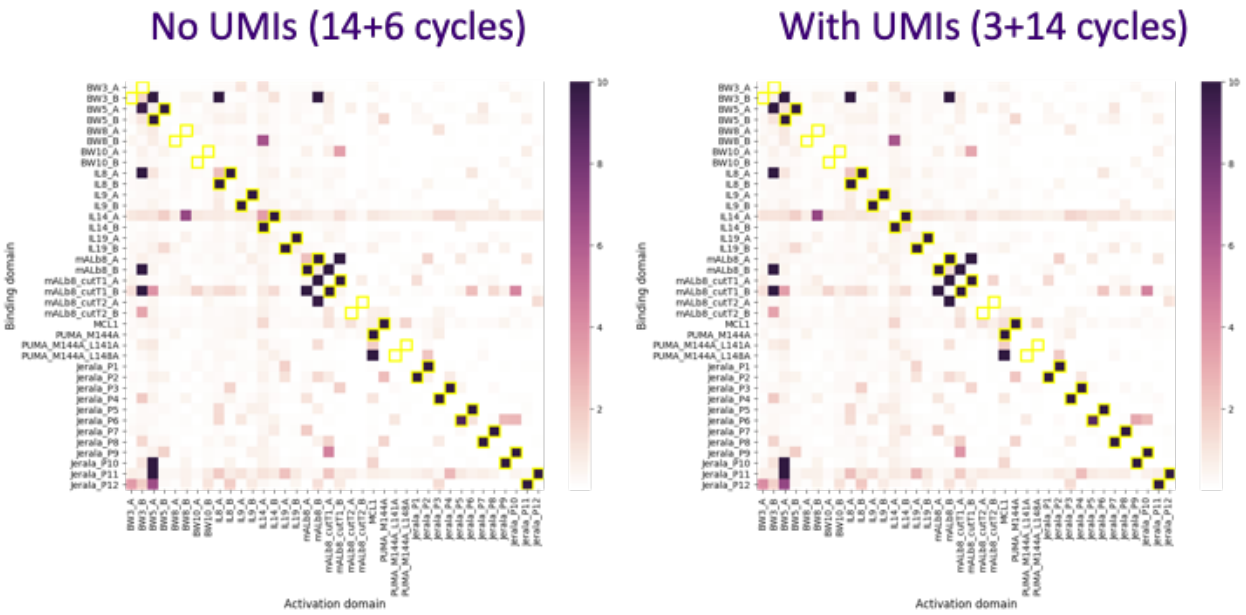

Figure 61. Interaction heatmaps obtained using UMIs in the qPCR process while preparing amplicons for sequencing (right), and not using UMIs in the qPCR (left).

According to our collaborator Ajasja Ljubetic some of the interactions in the L67 library were characterized by another method - NanoLuc Binary Technology (NanoBiT). The

comparison between HT-Y2H and NanoBiT for those interactions is summarized in the figure below.

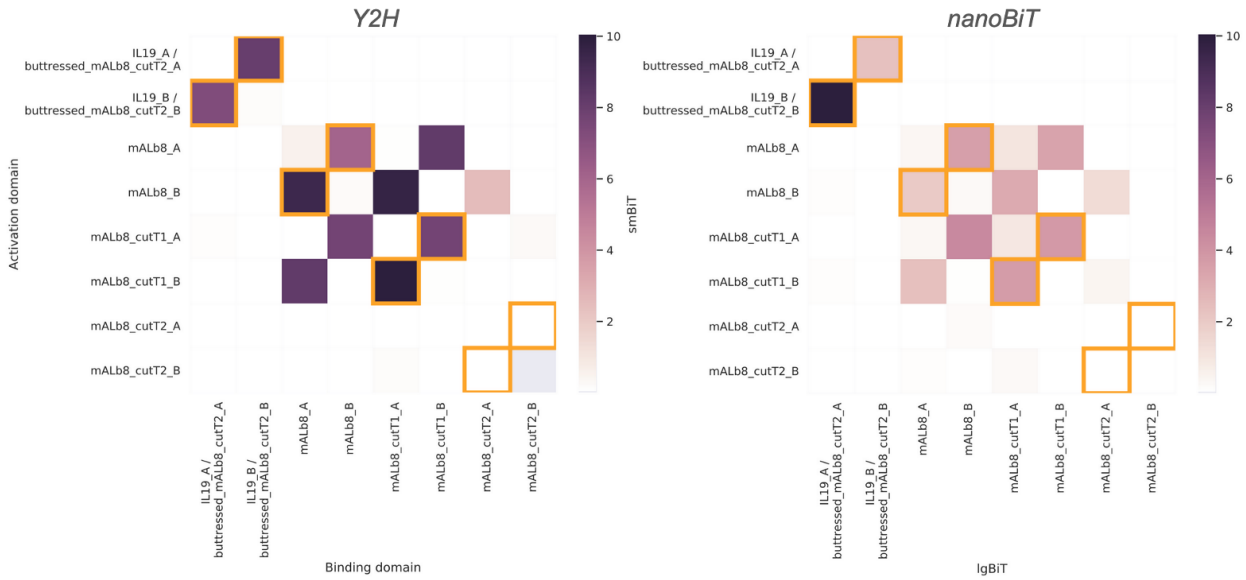

Figure 62. Visually comparing our HT-Y2H assay's linear enrichment scores with the data obtained in the Baker lab using nanoBiT assay (biolayer interferometry). An almost ideal visual agreement can be observed between these data.

Experiment L68. Characterizing variable length coiled coils with independently measured dissociation constants. Investigating qunatitative correlation between the dissociation constants and HT-Y2H fitness scores.

After our HT-Y2H assay's results for coiled coils P1-P12 coiled coils qualitatively matched published luciferase assay data, we envisioned the next logical step was to find coiled coils PPIs with measured dissociation constants Kd, and compare how HT-Y2H enrichment scores quantitatively correlate with the dissociation constants.

We identified a few sets of coiled coils from a few different papers for which dissociation constants Kd have been measured. Coiled coils of variable length with dissociation constants Kd

measured in (Thomas et al., 2013) (Figure 63) and 4-heptad long coiled coils with constants measured in (Plaper et al., 2021) were used in this L68 library. Additionally, previously characterized coiled coils P1-P12 whose dissociation constants have not been measured yet, have been added to the library to visual control purposes.

The interaction strength heatmap for variable length coils (Figure 63) obtained using HT-Y2H is shown in Figure 64. As can be seen in the Figure 64, HT-Y2H log enrichment scores correlation with measured Kd's with Pearson correlation coefficient R^2=0.95. We observed high self-consistency of HY-Y2H characterized by the symmetry of the heatmap, or the correlation between the upper and lower triangles of the interaction heatmap (R^2=0.9), that is also shown in the figure.

Figure 65A illustrates the homogeneity of library assembly. It can be seen that control coiled coils P1-P12 received much fewer reads than other pairs in the library which we attribute to using 1-year old stock of P1-P12 double stranded DNA fragments.

Figure 65B shows the resulting interaction strength heatmap of the experiment. It can be noticed that N1 exhibits strong auto-activation properties since it "interacts" with every other binder when fused to activation domain, whereas no such interactions are observed when N1 is fused to the binding domain. Again our method correctly identified on-target interactions between N and P binders characterized in (Plaper et al., 2021).

Despite receiving shallow sequencing coverage due to less homogeneous assembly, coiled coils P1-P12 still exhibited orthogonal interaction pattern observed in previous experiments (Figure 66C).

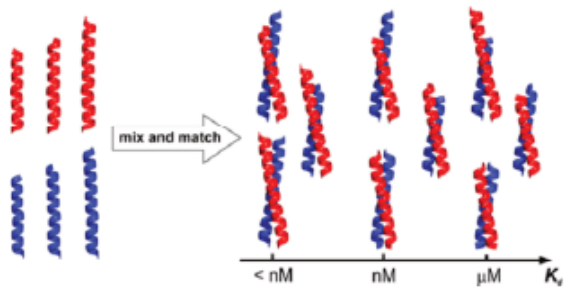

Table 1. Sequences for the Designed Heterodimeric Coiled Coils

	Sequence and heptad register[a]
	gabcdef gabcdef gabcdef gabcdef
CC-$A_X^3B_X^3$	G EIAALEK EXAALEW EIAALEQ GZ
	G KIAALKY KXAALKK KIAALKQ GZ
CC-$A_X^{3.5}B_X^{3.5}$	G LKQ EIAALEK EXAALEW EIAALEQ GZ
	G LKQ KIAALKY KXAALKK KIAALKQ GZ
CC-$A_X^4B_X^4$	GEIAALEQ EIAALEK EXAALEW EIAALEQ GZ
	GKIAALKQ KIAALKY KXAALKK KIAALKQ GZ

[a]Peptides were synthesized as C-terminal amides and were acetylated at the N-terminus. I-series, X = I, Z = S; N-series, X = N, Z = G.

Table 3. K_d Values (M) at 20 °C of the CC-A_NB_N Peptides

	CC-A_N^3	CC-$A_N^{3.5}$	CC-A_N^4
CC-B_N^3	$(3.10 \pm 0.58) \times 10^{-6}$	$(2.79 \pm 1.42) \times 10^{-7}$	$(7.47 \pm 3.30) \times 10^{-9}$
CC-$B_N^{3.5}$	$(6.40 \pm 2.77) \times 10^{-7}$	$(5.15 \pm 2.05) \times 10^{-9}$	$(2.08 \pm 0.73) \times 10^{-10}$
CC-B_N^4	$(3.87 \pm 1.45) \times 10^{-7}$	$(4.74 \pm 4.49) \times 10^{-10}$	$<1.0 \times 10^{-10}$

dx.doi.org/10.1021/ja312310g | J. Am. Chem. Soc. 2013, 135, 5161−5166

Figure 63. Adapted from paper (Thomas et al., 2013). Variable length coils experimentally characterized in (Thomas et al., 2013). (A) Two sets (A3, A3.5, A4) and (B3, B3.5, B4) of three to four heptads long coiled coils were screened in a mix-and-match experiment, where a pair of longer coils is anticipated to interact stronger than a pair of shorter coils and correspondingly to have a smaller dissociation constant in the nanoMolar region. (B) Amino acid sequences of the coiled coils. (C) Structural difference between I-Series and N-Series coils. (D) Measured dissociation constants in Molars for N-Series coils.

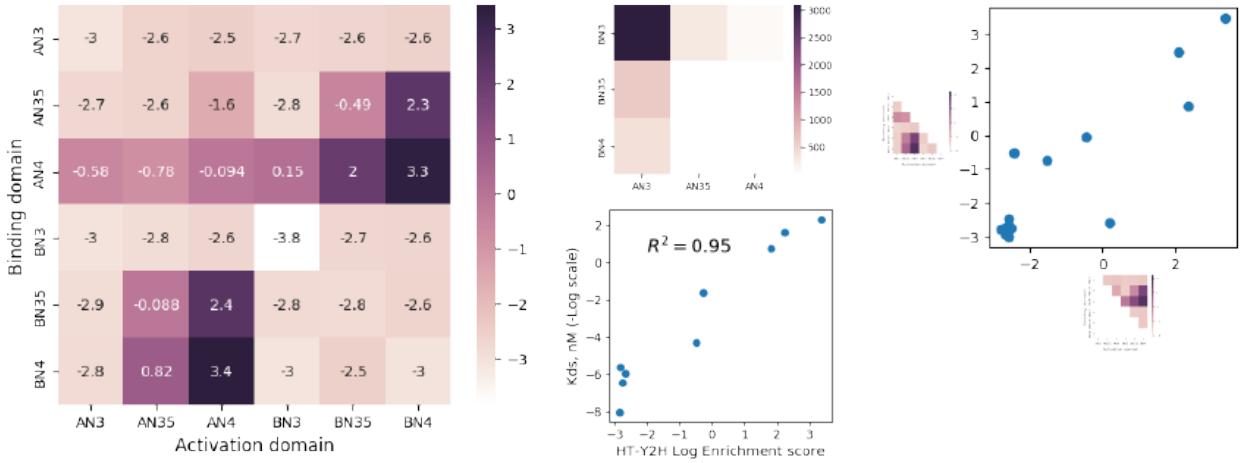

Figure 64. Variable length coils from (Thomas et al., 2013) screened by HT-Y2H. (A): Linear enrichment heatmap (B): "Output" reads (C): Pearson correlation between log values of HT-Y2H enrichment score and log values of dissociation constants Kd measure in (Thomas et al., 2013) was found to be R^2=0.95. (D): Symmetry of the heatmap (A): Pearson correlation between orientations R^2=0.95.

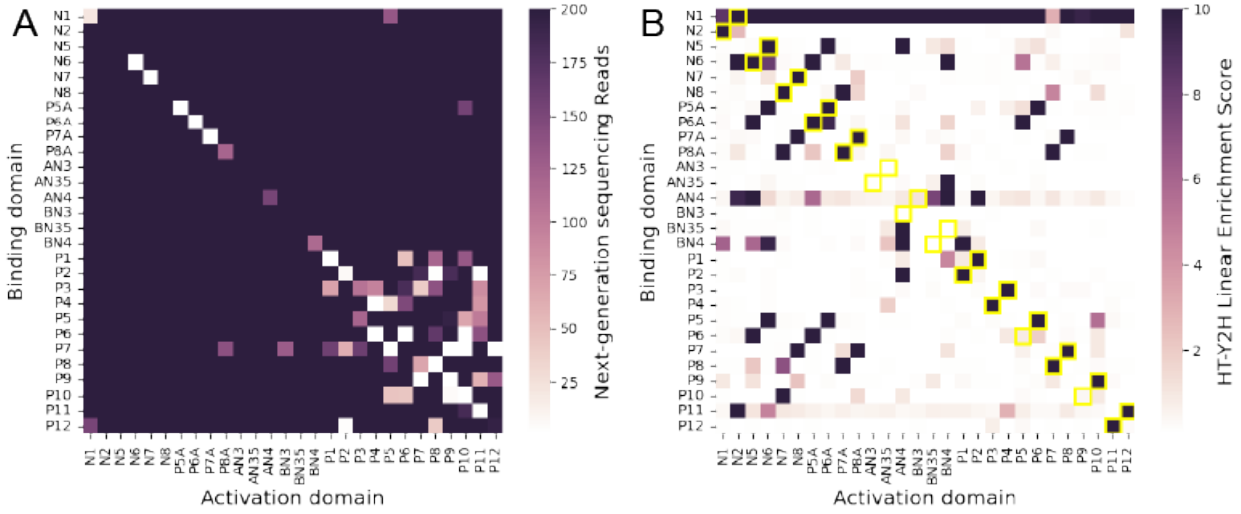

Figure 65. Heatmaps of "input" reads (A) and linear enrichment score (B) for library L68. (A) The color bar scale is limited to 200 reads to demonstrate sparseness of some areas of the heatmap. In particular, in addition to already mentioned sparseness along the diagonal attributed to the issue of homologous recombination in yeast, the subregion containing P1-P12 coils see to have very few reads implying that some pairs have likely not been assembled at all during the library transformation. This latter issue can be attributed to a one year old stock of double stranded DNA amplicons of P1 through P12 sequences used in the transformation compared to recently prepared N1, N2, N5-N8, P5A-P8A, ANs, BNs sequences. Still, in (B) one see almost ideal anticipated interaction picture between P1-P12 coils.

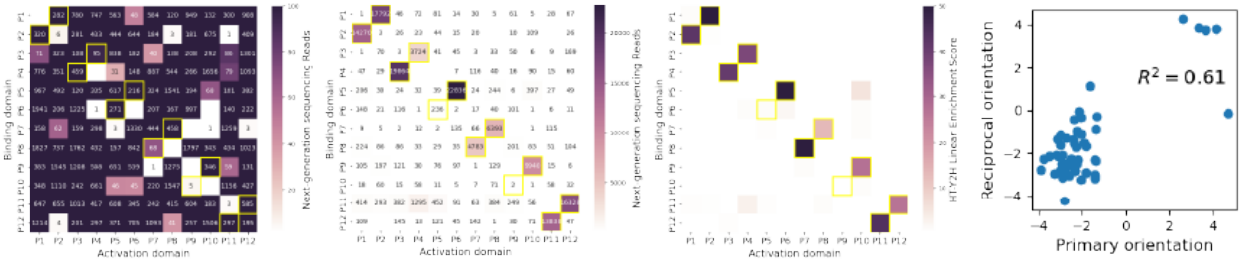

Figure 66. Interaction picture for P1 through P12 coiled coil from (Lebar et al., 2020). (A) Heatmap of "input" read counts. The color bar value is limited at 200 to demonstrate that some pairs had fewer than one hundred reads. Squares without any number inside them indicate that the corresponding pair was not identified in the sequencing data and thus likely was not assembled during the yeast transformation. Total of 8 pairs are completely missing in the sequencing data. (B) Heatmap of "output" reads. (C) Heatmap of linear enrichment scores. (D) Pearson correlation between log values of enrichment score between "primary" and "reciprocal" orientations for every unique protein-protein pair. In other words, this correlation indicates how symmetric the heatmap matrix in (C) is. Only those protein pairs were selected for which both primary and reciprocal input reads are greater than 50.

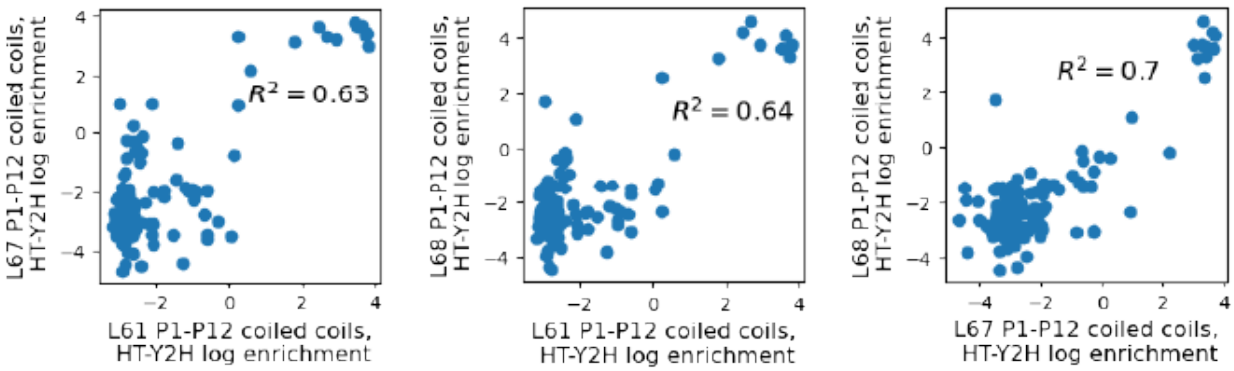

Figure 67. Comparing HT-Y2H enrichment values for P1 through P12 coiled coils obtained in library screens L61, L67, L68. (A) The Pearson correlation between log enrichment values from L61 and L67 library screens was found to be R^2=0.63. (B) The Pearson correlation between log enrichment values from L61 and L68 library screens was found to be R^2=0.64. (C) The Pearson correlation between log enrichment values from L67 and L68 library screens was found to be R^2=0.7.

Experiment L69. Using TwistBioscience's oligo pools to generate double stranded DNA fragments encoding binders sequences.

Currently, the screening of a library in an all-against-all manner requires that for every sequence in the library two double stranded DNA gene fragments need to be synthesized and ordered from Twist Bioscience due to the necessity to have different 5' and 3' sequences on those fragments (see Figure 68). The cost of the library synthesis gets prohibitively expensive when the library size gets in the double digits counts. Since potential applications of the method would deal with hundreds and even thousands of different sequences, e.g. mutagenized sequences or deep exploration networks predicted sequences, a more elegant, cheaper approach to the library synthesis and plasmid assembly needs to be invented.

One of the potential solutions to reduce the cost of DNA synthesis when increasing the library size is to use single stranded oligo pools. In this case, a single oligo in the pool will be used to make both double stranded fragments. To achieve this, that particular oligo needs to bear

just the most necessary sequence, in particular, a protein coding linker, a protein sequence, the protein's barcode sequence and a sequence that is shared by both terminators. Then, in two separate PCR reactions, using two different sets of primers, the remaining terminator specific sequences will be added along with two different protein coding linkers homologous to both ends of the linear backbone (see Figure 69).

This approach should allow screening of up to 80 amino acids long proteins (or equivalent of 240nt long), as the oligo pool scheme adds a total of 60 nucleotides at the 5' and 3' of the protein sequence.

We attempted this experiment once but the plasmid assembly wasn't successful which was confirmed by the absence of yeast colonies when 100uL of the transformed library was plated on an agar plate deficient of tryptophan amino acid that allows for the selection of plasmid assembly. We will troubleshoot the experiment and try it again in the future as there seem to be no technical hurdles that forbid this approach to work.

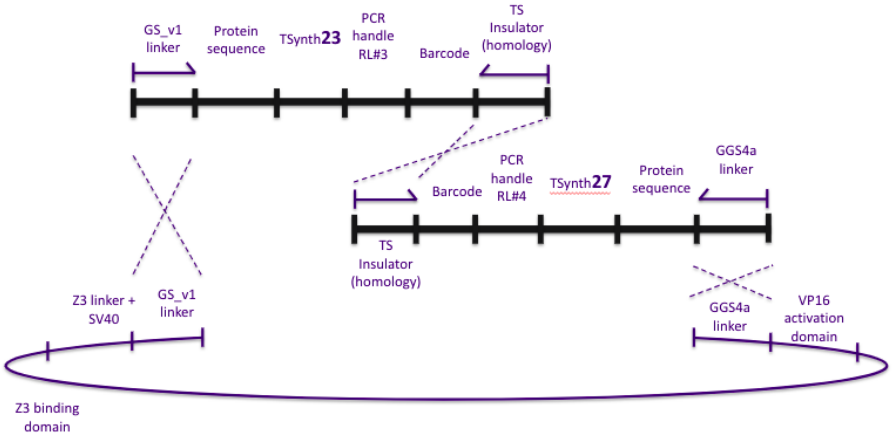

Figure 68. Current library assembly scheme uses double stranded DNA gene fragments synthesized by Twist Biosciences. The first fragment carries GS_v1 linker, first binder's sequence, TSynth23 synthetic terminator, PCR handle RL#3, first binder's pre-defined 20nt barcode sequence, and 22nt long TS insulator sequence used for homologous recombination assembly. The second fragment carries GGS4a linker homologous to the other end of the linear backbone, second binder's sequence, TSynth27 short synthetic terminator, a different PCR handle RL#4, second binder's pre-defined 20nt barcode sequence, and the TS homology sequence. Dashed lines represent homologous sequences needed to assemble the plasmid using yeast homologous recombination machinery.

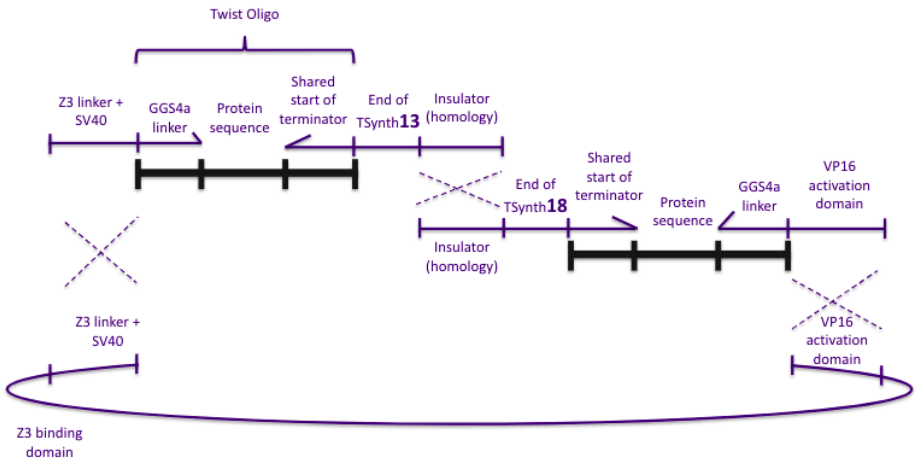

Figure 69. Proposed plasmid assembly scheme that uses 300nt long Twist Bioscience single stranded oligo pool to make two different double stranded fragments from the same oligo. The oligo sequence carries GGS4a linker, a binder's sequence and an 18nt long shared beginning of both Tsynth13 and Tsynth18 terminators. Then, in a PCR reaction a Z3EV linker and a nuclear localization signal SV40 is added to the fist fragment along with the remaining sequence of the TSynth13 synthetic terminator and an insulation sequence used in the homologous recombination assembly. Similarly, using different primers in another PCR reaction, a part of VP16 activation is added to the second fragment along with the remaining sequence of the TSynth18 synthetic terminator and the insulation sequence.

Experiment L70. Screening de novo designed "2+2" bundles of hydrogen bond network heterodimers.

In this experiment with collaborated with Postdoctoral scholar Ajasja Ljubetic who designed 37 pairs of "2+2" heterodimer bundles, where each protein is made up of two alpha helices joined by flexible linker (thus "2+2" bundles). Coiled coils P1-P12 were added for visual control purposes.

Unfortunately, the resulting heatmap did not exhibit anticipated orthogonal pattern, and another replicate of the experiment would be necessary to confirm it.

Figure 70 illustrates the homogeneity of P1-P12 coiled coils assembly. Similarly to L68 experiment, these control coiled coils received very few reads due to using 1-year old stock of

DNA used in to assemble the plasmids. Still, the resulting interaction heatmap for P1-P12 coiled coils exhibit orthogonal pattern (Figure 70C).

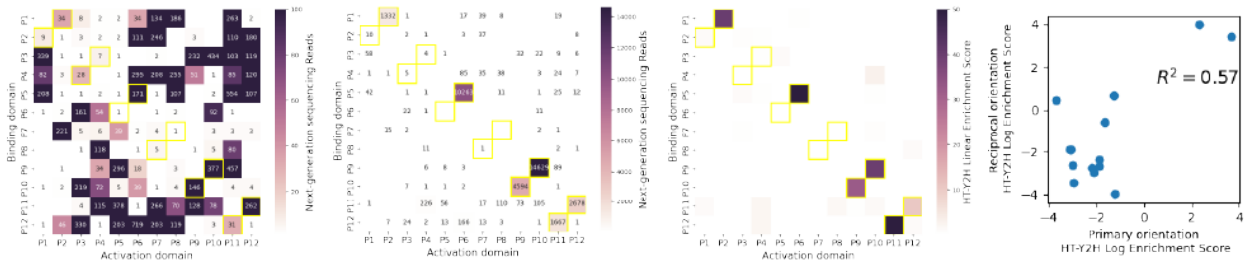

*Figure 70. Interaction picture between P1 through P12 coiled coils. (A) Heatmap of "input" read counts. The color bar value is limited at 200 to demonstrate that some pairs had fewer than one hundred reads. Squares without any number inside them indicate that the corresponding pair was not identified in the sequencing data and thus likely was not assembled during the yeast transformation. Total of 8 pairs are completely missing in the sequencing data. (B) Heatmap of "output" reads. (C) Heatmap of linear enrichment scores. (D) Pearson correlation between log values of enrichment score between "primary" and "reciprocal" orientations for every unique protein-protein pair. In other words, this correlation indicates how symmetric the heatmap matrix in (C) is. Only those protein pairs were selected for which both primary and reciprocal input reads are greater than 20, resulting in just 14 pairs out of 66 maximally possible ([12*12-12]/2=[144-12]/2=132/2=66).*

Conclusions

We have developed a massively parallel method for characterizing thousand to millions of pairwise protein-protein interactions in a single pot using Illumina next-generation sequencing experiment as a read-out. Our assay uses the conventional yeast two-hybrid method in its core and scales it up by expressing only one pair of proteins from the library in each yeast cell. In each two proteins of interest fused to the transcriptional activation and DNA-binding domains are expressed from the same plasmid that contains a centromere signal that allows a yeast cell to maintain just a single plasmid in the cell on average. If the two fusion proteins, when brought close together in the nucleus, can physically bind each other, a transcription of the growth essential enzyme HIS3 can occur. Thus, a simple growth selection in the histidine deficient media can be used as a proxy to determine the interaction strength. The relative frequency of

each plasmid before and after the selection process is used to determine the strength of the interaction between each protein pair in the library.

We have extensively characterized the assay. We first investigated how different orientations and combinations of promoters on the plasmid affect the expression levels of hybrid proteins by initially using green and red fluorescent proteins in place of hybrid proteins. We showed that pTEF1 and pCYC1 promoters led to similar change in expression levels of the green and red fluorescent proteins and selected them to drive the expression of hybrid proteins.

We showed that transforming yeast cells with linear double stranded DNA fragments encoding all the elements of the plasmid and using yeast homologous recombination to assemble plasmid libraries directly in yeast cells leads to robust and reproducible results. This approach greatly simplifies the protocol and reproduces the time and labor costs.

We showed that our assay achieves reproducible results by achieving a high degree of correlation between the replicates of the experiments.

We characterized our method using a few different sets of proteins, the interactions between which have been previously characterized by a few other methods in several papers. In particular, we used a family of human pro-apoptosis proteins BCL2 and their de novo designed inhibitors, three different sets of coiled coils peptides and de novo designed hydrogen bonds network containing heterodimers to validate our method. We achieved great qualitative and quantitative agreement between our assay's results and those from the published paper.

We identified one important aspect of plasmid design in the context of our method, in particular, we discovered that placing barcodes before transcriptional terminators in our plasmid design apparently disrupts the expression of some proteins regardless of the barcode sequence,

and that placing barcodes after the terminators leads to much better agreement with the published results.

We also discovered that using higher quality double stranded DNA gene fragments from Twist Bioscience to encode proteins leads to better agreement with published data and to more reproducible results.

Thus, the main finding is that using “barcode following terminator” plasmid design approach in combination with high quality double stranded DNA gene fragments synthesized by Twist Bioscience allows our HT-Y2H assay to achieve highest qualitative and quantitative agreement with the published data on a few different benchmark sets of proteins.

Method limitations

Inconsistency in assembling plasmids where the same protein is fused to the binding and activation domains, i.e. screening homodimers, is one of the main limitations of the method. We attribute this issue to the high degree of homology of linear fragments encoding the proteins and used in the plasmid assembly.

Interestingly, this inconsistency only arises in the current “barcode following terminator” design and was not appearing the previous “barcode preceding terminator” design.

Future directions

Troubleshoot L69 experiment using Twist Oligo Pools to prepare double stranded DNA fragments encoding proteins to reduce the cost and labor of our assay, as well as to allow

screening the interaction between thousands of different short peptides (up to 80 amino acids long) in a single pot which translates to millions of pairwise interactions.

Further experimental investigation is needed to address different potential bottlenecks and their influence on the outcome of the experiment, like what amount of cells vs library size to use when starting the selection process, what amount of cells to use for plasmid extraction, etc.

Summary table of all the experiments

Experiment N	Design	Binders	DNA Vendor
L14	Single Random Barcode	Homo-multimers XXXX	IDT Gblocks
L33	Barcode Preceeding Terminator	BCL2 family and their inhibitors	TwistBiosciences Gene fragments
L38	Barcode Preceeding Terminator	Designed Hetero-dimers	TwistBiosciences Gene fragments
L39	Barcode Preceeding Terminator	Designed Hetero-dimers	TwistBiosciences Gene fragments
L42	Barcode Preceeding Terminator	Control ZCON hetero-dimers	TwistBiosciences Gene fragments
L43	Barcode Preceeding Terminator	Designed Hetero-dimers V2	TwistBiosciences Gene fragments
L44	Barcode Preceeding Terminator	L33, L39, L43 combined	TwistBiosciences Gene fragments
L45	Barcode Preceeding Terminator	BCL2 and inhibitors	TwistBiosciences Gene fragments
L48	Barcode Preceeding Terminator	BCL2 and inhibitors	TwistBiosciences Gene fragments
L49	Barcode Preceeding Terminator	L33, L39, L43 combined	TwistBiosciences Gene fragments
L50	Barcode Following Terminator	10-heptad long DEN coiled coils	IDT OligoPools
L51	Barcode Following Terminator	Orthogonal coiled coils P1-P12	IDT OligoPools
L52	Barcode Following Terminator	4-heptad long DEN coiled coils	IDT OligoPools
L54	Barcode Following Terminator	Orthogonal coiled coils P1-P12	IDT OligoPools
L55	Barcode Following Terminator	4-heptad long DEN coiled coils	IDT OligoPools
L61	Barcode Following Terminator	Orthogonal coiled coils P1-P12	TwistBiosciences Gene fragments
L62	Barcode Following Terminator	Orthogonal coiled coils P1-P12	TwistBiosciences Gene fragments
L63	Barcode Following Terminator	SARS-CoV-2 intraviral PPIs	TwistBiosciences Gene fragments
L64	Barcode Following Terminator	Control ZCON hetero-dimers	TwistBiosciences Gene fragments
L65	Barcode Following Terminator	4-heptad long DEN coiled coils	TwistBiosciences Gene fragments
L66	Barcode Following Terminator	BCL2 family and their inhibitors	TwistBiosciences Gene fragments
L67	Barcode Following Terminator	More coiled coils, P1-P12 coils	TwistBiosciences Gene fragments
L68	Barcode Following Terminator	Coiled coils with measured Kds	TwistBiosciences Gene fragments
L70	Barcode Following Terminator	"2+2" bundles of heterodimers	TwistBiosciences Gene fragments

Yeast Electroporation Protocol

We are using a yeast transformation protocol from (Benatuil et al., 2010):

• Grow S. cerevisiae cells (EBY100) overnight to stationary phase (OD600 to or about 3) in YPD media (10 g/L yeast nitrogen base, 20 g/L Peptone and 20 g/L D-(+)-Glucose) on a platform shaker at 225 rpm and 30 °C.

• The next morning inoculate an aliquot of the overnight culture into 100 mL YPD media at an initiate 0.3 OD600.

• Continue to grow the inoculated cells on a platform shaker at 30 °C and 225 rpm until OD600 is approximately 1.6, usually after 5 hours.

• Collect yeast cells by centrifugation at 3000 rpm for 3 minutes and remove the media.

• Wash the cell pellet twice by 50 mL ice-cold water and once by 50 mL of ice-cold electroporation buffer (1 M Sorbitol / 1 mM CaCl2).

• Condition the yeast cells by re-suspending the cell pellet in 20 mL 0.1 M LiAc/10 mM DTT

and shaking at 225 rpm in a culture flask for 30 minutes at 30 °C.

• Collect the conditioned cells by centrifugation, wash once by 50 mL ice-cold electroporation buffer, and re-suspended the cell pellet in 100 to 200 µL electroporation buffer to reach a final volume of 1 mL.

• This corresponds to approximately 1.6×10^9 cells/mL and is sufficient for 2 electroporation

reactions of 400 µl each. The cell culture and preparation can be proportionally scaled up if more electrocompetent cells are needed to make larger or more libraries.

• The cells are kept on ice until electroporation.

DNA preparation and electroporation and plating:

• The DNA fragments for electroporation should be prepared in advance or in parallel to be

ready for transfection as soon as cells are ready.

• Combine 4 µg of digested vector backbone and 12 µg of DNA insert for each 400 µl electroporation reaction. The DNA mixture in water should be less than 50 µl. Reduce the volume by precipitation and resuspension in a smaller volume if necessary.

• Gently mix 400 µl electrocompetent cells and DNA and transfer to a pre-chilled BioRad GenePulser cuvette (0.2 cm electrode gap). Kept on ice for 5 minutes until electroporation.

• Electroporate the cells at 2.5 kV and 25 µF. Typical time constant ranges from 3.0 to 4.5 milliseconds.

• Transfer electroporated cells from each cuvette into 8 mL of 1:1 mix of 1 M sorbitol : YPD media. Incubate on a platform shaker at 225 rpm and 30°C for 1 hour.

• Collect cells by centrifugation and resuspend in SD-UT media (20 g/L glucose, 6.7 g/L yeast nitrogen base without amino acids, 5.4 g/L Na2HPO4, 8.6 g/L NaH2PO4·H2O and 5 g/L casamino acids [CSM-TRP-URA]). Use 250 mL SD-UT media for every 400 µl electroporated cells.

• Prepare 10-fold serially diluted cells from the cell suspension and plate 100 µl of $1/10^3$, $1/10^4$ and $1/10^5$ diluted cells onto selective SD plates and incubate at 30 °C incubator. Library size is determined from the colony counts after three days.

• Grow remaining cells in a baffled flask overnight on a shaker at 225 rpm and 30 °C. This is the transformed library.

Library Selection Protocol

Upon recovering cells in a 1:1 mis of YPAD:Sorbitol media, cells were resuspended in a baffled flask with 250mL of synthetic complete media lacking tryptophan amino acid (SC-TRP

media). Optical density value OD600 of that cell culture varied between 0.3 and 0.4 depending on the experiment. The flask was then put on an orbital shaker at 225rpm and 30C for 20 hours, at which point optical density was again measured and OD600 value was usually found to be around 1.0. Then cells were passaged for about 36 hours to allow yeast cells to drop redundant plasmid copies and achieve an average of a single plasmid per cell. In particular, the passage process was started by diluting 20-hour transformed library cell culture in a clean autoclaved baffled flask with fresh SC-TRP media at a 1:400 ratio and the cells were grown to approximately 1.0 OD600 value at which point the dilution procedure was repeated, multiple times if needed, with the same parameters until 36 hours have passed. At the end of the passage process a fraction of cells was inoculated in a baffled flask with SC-HIS-TRP media at a 1:400 ratio to start selecting for protein-protein interaction. The remaining SC-TRP cells were collected by centrifugation at 5000rpm (3000xg) for 5 minutes and resuspended in Zymo Research Solution1 buffer and stored in -80C freezer. The selection for interaction step usually lasted between 24 and 36 hours depending on the library and was concluded when cells in the SC-HIS-TRP flask reached OD600 value of 1.0, at which point cells were similarly collected by centrifugation, resuspended in Solution1 buffer and frozen at -80C.

Yeast Plasmid Extraction Protocol

The protocol adopted from Ph.D. student Cassie Bryan from Professor David Baker lab.

Zymoprep libraries:

Spin down cells (At least 100-fold over library size) – split into 2 aliquots if > 10^8 cells

Resuspend in 200 ul Solution 1

Freeze at -80C

Thaw and add 20 ul zymolyase

Inc at 37C for 4 hours, mixing once per hour

Freeze at -80C and thaw

Add 200 ul Soln 2 and invert to mix

Let sit 3-5 min

Add 400 ul Soln 3 and invert to mix

Centrifuge at max speed for 5 min

Transfer supernatant to new epp and spin again

Transfer supernatant to miniprep column

Spin 30 sec at 13000 rpm

Add 700 ul PB and spin 30 sec

Add 700 ul PE and spin 30 sec

Repeat PE wash

Empty and spin again for 1 min to dry column

Add 20 ul EB and spin 1 min

Reload column with eluate and spin again

Quantitative PCR (qPCR) protocol

Agilent real-time PCR machine is used

2 stages used. 1st stage is used to add 20nt long UMI and Illumina Read1 and Read2 primer sites. 2nd stage is usually 6-7 cycles to add sample indexes and P5/P7 flow cell adaptor sites.

Primers used for barcode-barcode amplicon:

Reaction proceeds till inflection point

98C 1min -> X cycles of: 98C 12sec / 67C 30sec / 72C 40sec -> 72C 5min final extension

Plasmid Map:

GGS4a Linker

VP16 Transcriptional Activation Domain

SV40 Nuclear Localization Signal (SV40 NLS)

pTEF1 promoter

Lac promoter

ColE1 E.Coli origin of replication

GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAG
TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACG
GGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGG
GCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTT
GCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTA
CATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATG
GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT
CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTA
TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGGGTCCTTTTCATCACGTGCTATAAAAATAATTATAATTTAAATTTTTTAATATAAATATATAAATTAAAAATAGAAAGTAAAAAAA
GAAATTAAAGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAGACTCTAGGGGGATCGCCAACAAATACTACCTTTTATCTTGCTCTTCCTGCTCTC
AGGTATTAATGCCGAATTGTTTCATCTTGTCTGTGTAGAAGACCACACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTATATC
TATTTAATCTGCTTTTCTTGTCTAATAAATATATATGTAAAGTACGCTTTTTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTTCTTCATTCCGTAACT
CTTCTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAAACATAAAAATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGATA
CGAGGCGCGTGTAAGTTACAGGCAAGCGATCCGTCCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACA
AGCCCGTCAGGGCGCGTCAGCGCGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGTTTA
AACAACGACATTACTATATATATAATATAGGAAGCATTTAATAGAACAGCATCGTAATATATGTGTACTTTGCAGTTATGACGCCAGATGGCAGTAGT
GGAAGATATTCTTTATTGAAAAATAGCTTGTCACCTTACGTACAATCTTGATCCGGAGCTTTTCTTTTTTTGCCGATTAAGAATTCGGTCGAAAAAAGA
AAAGGAGAGGGCCAAGAGGGAGGGCATTGGTGACTATTGAGCACGTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTATGTCTGTTATT
AATTTCACAGGTAGTTCTGGTCCATTGGTGAAAGTTTGCGGCTTGCAGAGCACAGAGGCCGCAGAATGTGCTCTAGATTCCGATGCTGACTTGCTGGG
TATTATATGTGTGCCCAATAGAAAGAGAACAATTGACCCGGTTATTGCAAGGAAAATTTCAAGTCTTGTAAAAGCATATAAAAATAGTTCAGGCACT
CCGAAATACTTGGTTGGCGTGTTTCGTAATCAACCTAAGGAGGATGTTTTGGCTCTGGTCAATGATTACGGCATTGATATCGTCCAACTGCATGGAGA
TGAGTCGTGGCAAGAATACCAAGAGTTCCTCGGTTTGCCAGTTATTAAAAGACTCGTATTTCCAAAAGACTGCAACATACTACTCAGTGCAGCTTCAC
AGAAACCTCATTCGTTTATTCCCTTGTTTGATTCAGAAGCAGGTGGGACAGGTGAACTTTTGGATTGGAACTCGATTTCTGACTGGGTTGGAAGGCAA
GAGAGCCCCGAAAGCTTACATTTTATGTTAGCTGGTGGACTGACGCCAGAAAATGTTGGTGATGCGCTTAGATTAAATGGCGTTATTGGTGTTGATGT
AAGCGGAGGTGTGGAGACAAATGGTGTAAAAGACTCTAACAAAATAGCAAATTTCGTCAAAAATGCTAAGAAATAGGTTATTACTGAGTAGTATTTA
TTTAAGTATTGTTTGTGCACTTGCCGCGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAG
TGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGA
CCACACCTCCCTTAACCAGATTCGAAAAGCGGCGGGTAACAGTATTGATGTAATCAATTTCCTACCTGAATCTAAAATTCCCGGGAGCAAGATCAAG
ATGTTTTCACCGATCTTTCCGGTCTCTTTGGCCGGGGTTTACGGACGATGGCAGAAGACCAAAGCGCCAGTTCATTTGGCGAGCGTTGGTTGGTGGAT
CAAGCCCACGCGTAGGCAATCCTCGAGCAGATCCGCCAGGCGTGTATATATAGCGTGGATGGCCAGGCAACTTTAGTGCTGACACATACAGGCATAT
ATATATGTGTGCGACGACACATGATCATATGGCATGCATGTGCTCTGTATGTATATAAAACTCTTGTTTTCTTCTTTTCTCTAAATATTCTTTCCTTATA
CATTAGGACCTTTGCAGCATAAATTACTATACTTCTATAGACACACAAACACAAATACACACACTAAATTAATACGGATTCTAGAACTAGTGGATCTA
CAAAATGGGAACCAGGCCATATGCTTGCCCAGTTGAATCTTGTGATAGAAGATTTTCACGTTCTGACGAATTGACAAGACATATTAGAATTCATACCG
GACAAAAGCCATTTCAATGTAGAATATGTATGAGAAACTTCTCCAGGTCAGATCACTTGACGACGCATATAAGAACACATACGGGCGAGAAGCCATT
CGCCTGTGACATATGTGGGAGAAAGTTTGCGAGAAGTGATGAACGTAAGCGTCATACCAAAATTCATACAGGCGGCGGAGGCACACCTGCAGCTGC

GTCGACTCTAGAGGATCCACCTAAAAAGAAGAGAAAGGTTGGTGGAGGAGGTTCTGGTGGAGGAGGTTCTAGTCCGGAAGATGAGATTCAAGCTTT
GGAGGAAGAAAACGCTCAACTAGAACAGGAGAACGCAGCGCTGGAGGAAGAGATTGCGCAACTAGAATATGGATGAGTGGTATGTTATATAACTGT
CTAGAAATAAAGAGTATCATCTTTCAAATAACGGAACGCTGTCTGGAGGNNNNNNNNNNNNNNNNNNNNATGTCGTAATAACCCCGCCCCGNNNNN
NNNNNNNNNNNNNNNGCAGTGGTCAGTGAGCTTTCGTTTGAAAGATGATACTCTTTATTTCTAGACAGTTATATATATATATATATATATATACCA
CCCATCATCCATATTTAAGGGCCTGGATTTTTTCCTTCAGTTGTTGGTTTTTCTCTTTTAGAGCAGCATTCTTCTCTTTAAGCTGTGCAATTTTGTCTTCC
GGACTACTTCCCCCGCTCCCTCCACTGCCTCCAGATCCACCCCCACCGTACTCGTCAATTCCAAGGGCATCGGTAAACATCTGCTCAAACTCGAAGTC
GGCCATATCCAGAGCGCCGTAGGGGGCGGAGTCGTGGGGGGTAAATCCCGGCCCCGGGGAATCCCCGTCCCCCAACATGTCCAGATCGAAATCGTCT
AGCGCGTCGGCATGCGCCATCGCCACGTCCTCGCCGTCTAAGTGGAGCTCAACCTTTCTCTTCTTTTTAGGCATTTTTTTCTAATGAGCAAGAAGTAAA
AAAAGTTGTAATAGAACAAGAAAAATGAAACTGAAACTTGAGAAATTGAAGACCGTTTATTAACTTAAATATCAATGGGAGGTCATCGAAAGAGAA
AAAAATCAAAAAAAAAAATTTTCAAGAAAAAGAAACGTGATAAAAATTTTTATTGCCTTTTTCGACGAAGAAAAAGAAACGAGGCGGTCTCGTTTAA
ACCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAGGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGGTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGC
GCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCA
CTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAA
AGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT
ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGA
AGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA

Bioinformatics pipeline:

Read1 and read2 carry barcode#1 and barcode#2.

Data processing begins by cutting 5'UTR and 3'UTR adapters RL#3 and RL#4 from fastq files obtained from Illumina NextSeq 550 or MiSeq sequenators. This is achieved by using cutadapt (Martin, 2011) software resulting in the extraction of the barcode#1 and barcode#2 corresponding to binder#1 and binder#2. Extracted 20nt long barcodes are then clustered using bartender software separately for binder#1 and binder#2.

Acknowledgements:

Devin Strickland, Postdoctoral scholar from Professor Eric Klavins's lab for pointing out

References:

Benatuil, L., Perez, J. M., Belk, J., & Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering, Design & Selection: PEDS*, *23*(4), 155–159. https://doi.org/10.1093/protein/gzq002

Berger, S., Procko, E., Margineantu, D., Lee, E. F., Shen, B. W., Zelter, A., Silva, D.-A., Chawla, K., Herold, M. J., Garnier, J.-M., Johnson, R., MacCoss, M. J., Lessene, G., Davis, T. N., Stayton, P. S., Stoddard, B. L., Fairlie, W. D., Hockenbery, D. M., & Baker, D. (2016). Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *ELife*, *5*. https://doi.org/10.7554/eLife.20352

Boyken, S. E., Chen, Z., Groves, B., Langan, R. A., Oberdorfer, G., Ford, A., Gilmore, J. M., Xu, C., DiMaio, F., Pereira, J. H., Sankaran, B., Seelig, G., Zwart, P. H., & Baker, D. (2016). De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science*, *352*(6286), 680–687. https://doi.org/10.1126/science.aad8865

Chen, M., Yan, C., Qin, F., Zheng, L., & Zhang, X.-E. (2021). The intraviral protein-protein interaction of SARS-CoV-2 reveals the key role of N protein in virus-like particle assembly. *International Journal of Biological Sciences*, *17*(14), 3889–3897. https://doi.org/10.7150/ijbs.64977

Chen, Z., Boyken, S. E., Jia, M., Busch, F., Flores-Solis, D., Bick, M. J., Lu, P., VanAernum, Z. L., Sahasrabuddhe, A., Langan, R. A., Bermeo, S., Brunette, T. J., Mulligan, V. K., Carter, L. P., DiMaio, F., Sgourakis, N. G., Wysocki, V. H., & Baker, D. (2019). Programmable design of orthogonal protein heterodimers. *Nature*, *565*(7737), 106–111. https://doi.org/10.1038/s41586-018-0802-y

Curran, K. A., Morse, N. J., Markham, K. A., Wagman, A. M., Gupta, A., & Alper, H. S. (2015). Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synthetic Biology*, *4*(7), 824–832. https://doi.org/10.1021/sb5003357

Fields, S., & Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, *340*(6230), 245–246. https://doi.org/10.1038/340245a0

Fong, J. H., Keating, A. E., & Singh, M. (2004). Predicting specificity in bZIP coiled-coil interactions. *Genome Biology*, *5*(2), R11. https://doi.org/10.1186/gb-2004-5-2-r11

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., … Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, *583*(7816), 459–468. https://doi.org/10.1038/s41586-020-2286-9

Jiang, Y., Tong, K., Yao, R., Zhou, Y., Lin, H., Du, L., Jin, Y., Cao, L., Tan, J., Zhang, X.-D., Guo, D., Pan, J.-A., & Peng, X. (2021). Genome-wide analysis of protein–protein interactions and involvement of viral proteins in SARS-CoV-2 replication. *Cell & Bioscience*, *11*(1), 140. https://doi.org/10.1186/s13578-021-00644-y

Lebar, T., Lainšček, D., Merljak, E., Aupič, J., & Jerala, R. (2020). A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nature Chemical Biology*, *16*(5), 513–519. https://doi.org/10.1038/s41589-019-0443-y

Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., Liu, C., Xiao, Z., Qu, Y., Yin, Y., Wang, C., Zhang, Y., Zhu, Z., Liu, Z., Peng, C., Zhu, T., & Liang, Q. (2021). Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing

SARS-CoV-2 Pathogenesis. *Med*, *2*(1), 99-112.e7. https://doi.org/10.1016/j.medj.2020.07.002

Liu et al. (2021). SARS-CoV-2–host proteome interactions for antiviral drug discovery. *Molecular Systems Biology*, *17*(11), e10396. https://doi.org/10.15252/msb.202110396

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

McIsaac, R. S., Oakes, B. L., Wang, X., Dummit, K. A., Botstein, D., & Noyes, M. B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research*, *41*(4), e57. https://doi.org/10.1093/nar/gks1313

Plaper, T., Aupič, J., Dekleva, P., Lapenta, F., Keber, M. M., Jerala, R., & Benčina, M. (2021). Coiled-coil heterodimers with increased stability for cellular regulation and sensing SARS-CoV-2 spike protein-mediated cell fusion. *Scientific Reports*, *11*(1), 9136. https://doi.org/10.1038/s41598-021-88315-3

Potapov, V., Kaplan, J. B., & Keating, A. E. (2015). Data-Driven Prediction and Design of bZIP Coiled-Coil Interactions. *PLOS Computational Biology*, *11*(2), e1004046. https://doi.org/10.1371/journal.pcbi.1004046

Rogers, J. M., Oleinikovas, V., Shammas, S. L., Wong, C. T., De Sancho, D., Baker, C. M., & Clarke, J. (2014). Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proceedings of the National Academy of Sciences*, *111*(43), 15420–15425. https://doi.org/10.1073/pnas.1409122111

Springstein, B. L., Deighan, P., Grabe, G. J., & Hochschild, A. (2021). A Bacterial Cell-Based Assay To Study SARS-CoV-2 Protein-Protein Interactions. *MBio*, *12*(6), e02936-21. https://doi.org/10.1128/mBio.02936-21

Thomas, F., Boyle, A. L., Burton, A. J., & Woolfson, D. N. (2013). A Set of de Novo Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. *Journal of the American Chemical Society*, *135*(13), 5161–5166. https://doi.org/10.1021/ja312310g

Xu et al. (2021, August 3). *Compartmentalization-aided interaction screening reveals extensive high-order complexes within the SARS-CoV-2 proteome—ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S2211124721009098

Younger, D., Berger, S., Baker, D., & Klavins, E. (2017). High-throughput characterization of protein-protein interactions by reprogramming yeast mating. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(46), 12166–12171. https://doi.org/10.1073/pnas.1705867114