Evolution of a Eukaryotic Transcription Factor's co-TF Dependence Involves Multiple Intrinsically Disordered Regions Affecting Activation and Autoinhibition

Lindsey F. Snyder[1], Emily M. O'Brien[2], Jia Zhao[2,*], Jinye Liang[2], Yuning Zhang[3], Wei Zhu[4], Thomas J. Cassier[2], Nicholas J. Schnicker[5, 6], Xu Zhou[7], Raluca Gordân[3, 4, 8, 9], Bin Z. He[1,2]

[1] Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA; [2] Department of Biology, University of Iowa, Iowa City, IA; [3] Department of Biostatistics & Bioinformatics, Duke University, Durham, NC; [4] Department of Molecular Genetics & Microbiology, Duke University, Durham, NC; [5] Protein and Crystallography Facility, University of Iowa, Iowa City, IA; [6] Department of Molecular Physiology and Biophysics, University of Iowa, Iowa City, IA; [7] Department of Pediatrics, Division of Gastroenterology, Hepatology and Nutrition, Boston Children's Hospital and Harvard Medical School, Boston, MA; [8] Department of Computer Science, Duke University, Durham, NC [9] Department of Cell Biology, Duke University, Durham, NC

* Current address: Laboratory of Immunophysiology, The Ragon Institute of Mass General, MIT, and Harvard; Department of Biology, Massachusetts Institute of Technology. Cambridge, MA, USA.

ABSTRACT

Combinatorial control by multiple transcription factors (TFs) allows for increased specificity and information integration. Despite its prevalence in eukaryotes and crucial roles in gene regulation, the reason why eukaryotic TFs often depend on one another, and whether and how such codependency evolves, are not well understood. We exploit natural variation in co-TF dependence in the phosphate starvation response between related yeasts to address this

question. In *Saccharomyces cerevisiae*, a bHLH TF, Pho4, relies on the homeodomain co-TF Pho2 to regulate ~28 genes. By contrast, Pho4 in a related yeast pathogen, *Candida glabrata*, exhibits significantly reduced Pho2 dependence and an expanded target set of ~70 genes. Biochemical analyses showed *C. glabrata* Pho4 (CgPho4) binds to the same consensus motif with a 3-4 fold higher affinity than ScPho4. A machine-learning-based prediction and yeast one-hybrid assay identified two Intrinsically Disordered Regions (IDRs) in CgPho4 that can boost the activity of the main activation domain but showed little to no activity on their own. We also found evidence for autoinhibition as a mechanism for the co-TF dependence. We identified an IDR in ScPho4 next to its DNA binding domain that functions as a double-edged sword: it both allows for enhanced TF activity with Pho2, but also constrains the activity when the co-TF is absent. Together, our results provide a detailed molecular picture of how co-TF dependence is mediated and how evolution in functional domains, including in the IDR of the TF, can lead to changes in codependency and result in significant rewiring of the regulatory targets.

INTRODUCTION

Combinatorial regulation is the requirement of two or more transcription factors (TFs) to bind in a defined configuration in order for gene regulation to occur. It is prevalent and plays an important role in eukaryotic transcription, where it allows for increased specificity and integration of multiple upstream signals. Eukaryotic TFs recognize short degenerate motifs that are present hundreds to thousands of times in the genome. Requiring more than one TF to bind and regulate a gene increases the specificity by excluding genes whose promoters encode some but not all the TFs' motifs, or if they were in the wrong orientation. It also allows for the integration of information from different upstream sources, such as the internal circadian clock and external light intensity, or tissue selector and morphogen gradient (Guss et al. 2001; Markson et al. 2013; Puszynska and O'Shea 2017). Being such a crucial part of any gene regulatory network, TF co-TF combinations are often taken for granted. However, we have surprisingly limited knowledge on what makes TFs rely on one another -- whether there is a hard limitation on what

a TF can possibly do alone, or if natural selection favored TFs that are dependent on other TFs, so as to achieve the regulatory specificity and information integration. It's also unclear what the consequences of altering or disrupting such codependency would be.

Hints to why TFs rely on one another can be gleaned from biochemical and structural studies of TF cooperativity. One reason for TF codependency is the cooperative binding to DNA through either direct or indirect mechanisms. A direct mechanism involves physical interactions between TFs that enhances the binding affinity to DNA via conformational changes. Examples of this include the human bHLH TFs MAX and MYC, whose monomers cannot bind DNA on their own, but homo- or heterodimers of them can (Morgunova and Taipale 2017). Cooperativity can also be achieved indirectly, where binding of one TF causes local changes in the DNA shape such that binding by the second TF is favored. The ternary complex of the yeast mating type TF, alpha2, Mcm1, and the DNA showed that Mcm1 can bend the DNA to position it so that alpha2 can bind and make a physical interaction with Mcm1 (Tan and Richmond 1998). In another example, one of the cofactors of the TALE homeodomain protein, MEIS1, was shown to bind to the opposite side of the DNA in the solved structure, ruling out the possibility of a direct interaction (Jolma et al. 2015). Another indirect mechanism involves two TFs that bind within the span of a nucleosome competing against the same nucleosome and effectively cooperating (Morgunova and Taipale 2017). This evidence raises the possibility that enhancement of the DNA binding affinity could alter TF codependency.

Another less studied reason for TF codependency is the ability to recruit transcriptional coactivators, such as the mediator complex, chromatin remodelers, and general TFs. Many studies have detailed the composition of TF activation domains and how they vary greatly in strength and ability to recruit different coactivators (DelRosso et al. 2023; Staller et al. 2018; Sanborn et al. 2021). This difference in the ability and specificity to recruit important cofactors offers another possible reason for the need of multiple TFs to activate transcription. For example, Pho4 and Pho2 have both been shown to physically interact with NuA4, a histone

acetyltransferase (Nourani et al. 2004). Similarly, GATA4 and TBX5 were both required for the histone deacetylases to be recruited to endothelial genes, and absence of either results in downstream dysregulation and a heart disease phenotype (Ang et al. 2016). Lastly, a more recently proposed mechanism involves combinatorial regulation through the creation of condensates, or liquid-liquid phase separation, formed by proteins with large intrinsically disordered regions and enhanced by binding to DNA or interactions with other proteins (Wei et al. 2020). Together, these studies gave insight into how TFs cooperate to bind DNA and recruit coactivators but failed to address the question of why these TFs need to cooperate.

We take advantage of natural variation in co-TF dependency in the phosphate starvation (PHO) response network in yeast to answer some of the above questions. The PHO pathway has been studied extensively in *S. cerevisiae*. Two TFs are required for gene activation, a basic Helix-Loop-Helix protein Pho4 and a homeodomain TF Pho2 (Fascher, Schmitz, and Hörz 1990). Together they activate a dedicated regulon of ~28 genes to restore phosphate homeostasis (Ogawa, DeRisi, and Brown 2000). It has been shown that *S. cerevisiae* Pho4 (ScPho4) requires Pho2 to 1) bind cooperatively to their respective motifs (Barbaric et al. 1998; Zhou and O'Shea 2011), 2) help recruit the histone acetyltransferase (HAT) (Nourani et al. 2004), and 3) help recruit general transcription factors and the PolII complex (Magbanua et al. 1997). In the related commensal yeast, *C. glabrata,* Pho4 (CgPho4) is far less dependent on Pho2 and can bind and activate gene expression on its own (Kerwin and Wykoff 2009; He, Zhou, and O'Shea 2017). In turn, CgPho4 significantly expanded its targets (60-70 genes), reaching beyond phosphate homeostasis (He, Zhou, and O'Shea 2017) (Fig. 1A). This was true both in its own genome and when CgPho4 protein replaces the endogenous ScPho4 CDS and is expressed in *S. cerevisiae* (He, Zhou, and O'Shea 2017). Further testing with more Pho4 orthologs suggested that this reduced Pho2-dependence was likely due to derived changes in the Nakaseomyces genus, which includes *C. glabrata* and also two more opportunistic pathogens (He, Zhou, and O'Shea 2017). Here, we take advantage of this natural variation

between the orthologous Pho4's to elucidate the genetic basis and molecular mechanisms for co-TF dependence and its evolution.

We propose two models for the difference in Pho2-dependence between ScPho4 and CgPho4 (Fig. 1B). The first is an "enhanced activity" model, where CgPho4 is able to bind more tightly to DNA and recruit the general transcription factors and the Pol II complex, hence is less dependent on the co-TF, Pho2, than ScPho4 is. The second is an "autoinhibition model", where ScPho4 is proposed to be inhibited. This model has been suggested previously and involves intramolecular interactions between its DBD and the N-terminal AD, which is relieved upon interaction with Pho2 (Shao, Creasy, and Bergman 1996). Under this model, CgPho4 either lacks the self-inhibition or the effect is far weaker and hence doesn't depend on Pho2. It is worth noting that these two models are not mutually exclusive.

Here, we used a combination of *in vitro* binding assays and *in vivo* yeast one- and two-hybrid assays to characterize the functions of the DBD and AD in each Pho4 ortholog, and also genetically dissected the regions in the two Pho4 responsible for the co-TF dependence, with the help of chimeras of CgPho4 and ScPho4 assayed by a well-tested dual fluorescence reporter coupled with flow cytometry. Our findings give insight into how *trans* evolution can lead to reduced co-TF dependence and, in turn, why co-TFs are needed in the first place.

RESULTS

Domain organization and sequence divergence between ScPho4 and CgPho4

Based on genetic and biochemical studies, ScPho4 encodes the following functional regions from its N- to its C-terminus (Fig. 1C): a regulatory region (R1) interacting with the negative regulator Pho80 (Jayaraman, Hirst, and Goding 1994), the activation domain (AD) (N. Ogawa and Oshima 1990; McAndrew et al. 1998), a region encoding the nuclear export and import signals (NLS) (Komeili and O'Shea 1999; O'Neill et al. 1996), a protein-interaction domain with the front part contacting Pho80 and the second part interacting with Pho2 (P2ID) (Hirst et al. 1994; Jayaraman, Hirst, and Goding 1994), and finally the basic Helix-Loop-Helix (bHLH) DNA

binding domain (DBD) (N. Ogawa and Oshima 1990; Shimizu et al. 1997). Both ScPho4 and CgPho4 were predicted to be intrinsically disordered outside the DBD, with the exception for some low confidence helices in R1 and AD (Fig. S1). They and the DBD show a higher level of conservation between ScPho4 and CgPho4 (Fig. 1C). The NLS and P2ID regions contain four functionally important phosphorylation sites targeted by the cyclin-dependent kinase complex Pho80/85 (Kaffman et al. 1994; O'Neill et al. 1996). Those located in the NLS control Pho4's nuclear translocation, while phosphorylation of the one in P2ID disables ScPho4 from interacting with Pho2 (Komeili and O'Shea 1999). All four Pho80/85 targeted phosphorylation sites are clearly identifiable in the alignment (Fig. S2); in fact, they were previously found to be conserved in orthologs outside the Saccharomycotina subphylum (Gomes-Vieira et al. 2018).

CgPho4 binds the same DNA motif with 3-4 times higher affinity than ScPho4 does

To test the enhanced activity model, we first compared the DNA binding ability of the two Pho4 orthologs. First, we needed to determine if the two Pho4 orthologs recognize the same DNA sequence. We used our previous Chromatin-Immunoprecipitation (ChIP-seq) data for both Pho4 in the *S. cerevisiae* background to identify their respective motifs (He, Zhou, and O'Shea 2017) (Materials and Methods). The result showed that they bind the same E-box motif "CACGTG", with no obvious difference in the flanking nucleotides (Fig. 2A). The relatively small number of ChIP peaks (74 and 118) leave the possibility of small but significant differences in flanking nucleotide preference between the two Pho4's. To comprehensively characterize and compare the binding preferences of ScPho4 vs CgPho4, we turned to Protein Binding Microarrays (PBM, Materials and Methods). Two library designs were used. The first comprises all possible 9-mer sequences (universal PBM); the second library consists of probes centered on ChIP-identified binding sites for both Pho4's in their native genome, termed "genomic context PBM" or gcPBM (Gordân et al. 2013). Results from both PBM experiments revealed no evidence for binding specificity differences between ScPho4 and CgPho4 (Fig. 2B, C).

Having established that the two Pho4 orthologs bind the same sequence, we then tested the hypothesis that CgPho4 binds more tightly than ScPho4 does. To do so, we recombinantly purified the DBD region of both Pho4, and measured their binding affinity to a set of 17-bp oligos that contain either the consensus motif “CACGTG”, or the nine 1-bp-off half site variants of the palindromic motif (Maerkl and Quake 2007). Using Biolayer Interferometry (BLI), we found CgPho4 DBD binds the consensus motif >3 times more tightly than ScPho4 DBD does (K_D = 5.4 nM vs 1.4 nM for ScPho4 and CgPho4, respectively; raw Student’s t-test for log K_D difference P = 0.013, Holm-Bonferroni corrected P = 0.1). Among the suboptimal motifs tested, only “CCCGTG” showed a 2-fold or more difference in K_D ($K_{D,\ ScPho4}$ / $K_{D,\ CgPho4}$ = 2, raw P = 0.05, Holm-Bonferroni corrected P = 0.31, Fig. 2D). We confirmed the K_D difference for the consensus “CACGTG” sequence using an Electrophoretic Mobility Shift Assay (Fig. S3). We conclude that CgPho4 recognizes the same DNA sequences as ScPho4 but binds the consensus E-box motif and at least one suboptimal motif with a 2-fold or lower K_D than ScPho4.

CgPho4 encodes two additional Activation Enhancer Domains (AEDs)

We next tested the hypothesis that CgPho4 has increased transactivation potential compared to ScPho4 under the enhanced activity model. We first predicted regions with activation potential in both Pho4 orthologs using PADDLE, a Convoluted Neural Network trained on 150 activation domains (ADs) from 164 TFs in *S. cerevisiae* (Sanborn et al. 2021). PADDLE recovered the experimentally identified AD in ScPho4 between aa 60-102, which contained the 9aaTAD motif (Fig. 3A, orange triangle). No other regions with significant activation potential were predicted in ScPho4. By contrast, three regions in CgPho4 were predicted to have activation capabilities, including one corresponding to the AD in ScPho4, which was predicted with strong potential (Z-score > 6, Fig. 3B). Two additional regions, one overlapping R1 and the other spanning the NLS and P2ID, were predicted to have medium activation strength (Z-score > 4). which we will refer to as E1 and E2, respectively. Interestingly, the best match to the 9aaTAD motif pattern in CgPho4 was located in E2 rather than in the AD (Fig. 3B, orange triangle).

To test the function of all predicted regions with activation potential in both Pho4, we set up a yeast one-hybrid assay, in which the candidate region was fused to the Gal4 DBD and the activation potential was measured by a genome-integrated *GAL1*pr-mCherry reporter in *S. cerevisiae* (Materials and Methods). We confirmed that ScPho4's AD was able to activate the reporter above the background level (~7 fold, Holm-Bonferroni corrected $P < 0.01$). CgPho4's AD offered a similar level of activation (5.6 fold, corrected $P < 0.01$). ScPho2 AD provided the highest activation, at ~12 fold over the background, consistent with the hypothesis that its recruitment by ScPho4 enhances gene activation (Fig. 3C, corrected $P < 0.01$). We were able to further localize the required residues for activation in ScPho4's AD to a region of 32 aa centered on the 9aaTAD motif (Fig. 3A, C). Neither E1 nor E2 from either Pho4 activated the reporter on its own (Fig. 3C). This was surprising, given that CgPho4's E2 contains a match to the 9aaTAD as ScPho4's AD does (Fig. 3B). We hypothesized that E1 and E2 in CgPho4 could enhance the activity of the AD. We first tested ScE1 and CgE1 fused to their own AD. We found that only the latter resulted in a significant enhancement in the activation potential (Fig. 3D). Next, we fused CgE1 or CgE2 in both orientations to the minimal AD of ScPho4 (ScAD_9aa). We found that indeed, both regions were able to significantly enhance the activity of the ScAD_9aa (Fig. 3D).

In summary, we found that CgPho4 encodes two Activation Enhancer Domains (AEDs) in its Intrinsically Disordered Regions (IDRs). The AEDs have little activation capability on their own, but can significantly enhance the activation potential of the main AD from either Pho4. We hypothesize that they helped CgPho4 become less dependent on Pho2 as a result.

A dual fluorescence reporter coupled with flow cytometry accurately determines the activity of Pho4 chimeras with and without Pho2

So far, we compared DNA binding and activation domain capacity either *in vitro* or in a fusion protein context between CgPho4 and ScPho4. While they support the enhanced activity model, it remains unclear whether and how they contribute to divergence in Pho2-dependence in the native protein context; also, it is unknown if any other region also contributes to the difference.

To answer these questions, we divided ScPho4 and CgPho4 into five corresponding parts (Fig. 4A), with breakpoints chosen in well-aligned regions and based on predicted secondary structures and experimentally identified domains (Fig. S2). We created all 2^5=32 combinations as well as 20 additional ones with alternative breakpoints (Fig. 4A). All constructs were C-terminal tagged with mNeon to quantify the protein levels, and were expressed from the native *ScPHO4* promoter and UTRs (Fig. 4B). We also created two *S. cerevisiae* host strains, in which the *ScPHO5* CDS was replaced by an mCherry reporter, and *pho80* was knocked out so that all Pho4 chimeras were constitutively nuclear localized (Kaffman et al. 1994; O'Neill et al. 1996). One of the two host strains also had *pho2* knocked out. For each chimera, we measured its *PHO5pr*-mCherry and Pho4-mNeon levels using flow cytometry in both hosts. While mCherry and mNeon levels for the same Pho4 construct varied between experiments, the ratios between the two were consistent and characteristic of the specific construct (Fig. 4C, D). We therefore defined the activity of a Pho4 chimera as the ratio between the median fluorescence intensity (MFI) of mCherry and mNeon, which we will refer to as A_{PHO2} or $A_{pho2\Delta}$ from hereon.

The two Activation Enhancer Domains (AEDs) in CgPho4 increased the activity of the chimeric TFs but were insufficient to make them Pho2-independent

To investigate the contribution of individual regions to Pho2-dependence and their pairwise interactions, we analyzed the 2^5 chimeras as two sets based on whether they were majority ScPho4 or CgPho4. In each set, we compared the chimeras' activity with or without Pho2 to the endogenous Pho4 in that set. The resulting estimates of 1- and 2- region swaps on either Pho4 backgrounds were plotted as four triangular heatmaps (Fig. 5A). For example, a chimera with its NLS from CgPho4 (NLS:Cg) and the rest from ScPho4 led to estimates of the activity difference between NLS:Cg and NLS:Sc on ScPho4 background with or without Pho2. The two heatmaps largely mirrored each other, suggesting that differences between the corresponding regions in the two Pho4 are mostly context independent.

Based on our previous results, we expected R1, NLS, and DBD in CgPho4 to enhance the activity of the chimera with and without Pho2: the first two encode CgE1 and a majority of CgE2_9aa, which we showed enhance the activity of the main AD (Fig. 4), while DBD:Cg is expected to increase the binding ability of the chimera. While we indeed found R1:Cg and NLS:Cg to enhance A_{PHO2} on ScPho4's background (Fig. 5A top left, 5B, 5C), contrary to our expectation, DBD:Cg reduced the chimera's activity with Pho2. Also unexpectedly, R1:Cg and NLS:Cg swaps had a much smaller effect on $A_{pho2\Delta}$ (Fig. 5A, lower left), suggesting that despite their ability to increase A_{PHO2}, the chimeras were still dependent on Pho2. In fact, none of the 1- or 2-region swaps led to large increases in $A_{pho2\Delta}$. This argues against a classic lock-and-key model for autoinhibition, which we expect to be broken by swapping one of the regions involved.

Since the NLS:Cg region encodes both the NLS/NES and a key part of the second AED (CgE2_9aa), we asked which of these two mechanisms was responsible for its positive effect on A_{PHO2}. To test the possibility that the enhanced A_{PHO2} was due to more Pho4 chimera being inside the nucleus, we performed fluorescence microscopy to quantify the fraction of nuclear-localized Pho4 proteins in six constructs that bear either NLS:Sc or NLS:Cg. No significant difference between the two groups were found (Fig. S4, F-test $P = 0.13$). We thus conclude that the positive effects of R1 and NLS from CgPho4 on A_{PHO2} were mainly due to their ability to boost activation.

To further examine the effects of swapping regions 1-3 from CgPho4, including whether they contribute non-additively, we plotted their A_{PHO2} and $A_{pho2\Delta}$ in comparison with ScPho4 and CgPho4 (Fig. 5B). To test for epistasis, we fit a linear model, $Y = X_0 + R1 + AD + NLS + R1{:}AD + R1{:}NLS + AD{:}NLS + R1{:}AD{:}NLS$, where the first term represents the activity of ScPho4, the next three terms represent the main effect of each CgPho4 region, and the rest for epistasis. We found that R1 and NLS had significant, positive effects on both A_{PHO2} and $A_{pho2\Delta}$, although the magnitude were much smaller for $A_{pho2\Delta}$ (Fig. 5C, R1 and NLS Holm-Bonferroni corrected $P < 0.05$). Moreover, both show significant positive epistasis with AD for A_{PHO2} (corrected $P < 0.05$

for R1:AD and AD:NLS), and trend in the same direction for $A_{pho2\Delta}$. In light of our activation enhancing results (Fig. 3D), we suggest that the two AEDs in CgPho4 work more effectively with AD:Cg than with AD:Sc. Curiously, R1 and NLS show negative epistasis for A_{PHO2} but positive epistasis for $A_{pho2\Delta}$. Lastly, the three way interaction terms are negative for both A_{PHO2} and $A_{pho2\Delta}$.

We also examined the effect of swapping CgPho4's DBD. On its own, swapping DBD:Cg decreased both A_{PHO2} and $A_{pho2\Delta}$ compared to ScPho4 (Fig. 5B), suggesting that DBD:Cg may be incompatible with one or multiple regions in ScPho4. In summary, we found that R1:Cg and NLS:Cg increased the activity of the chimera with Pho2, but were insufficient to remove the Pho2 dependence alone or in combination. In the presence of Pho2, R1:Cg and NLS:Cg both showed positive epistasis with AD:Cg on ScPho4's background. Lastly, no 1- or 2-region swaps from CgPho4 into ScPho4 increased $A_{pho2\Delta}$ to the level of CgPho4, suggesting multiple mechanisms at play. Together, we conclude that CgPho4's two AEDs can increase the activity of the chimera in the presence of Pho2, although they fail to restore the activity without Pho2 to CgPho4's level, suggesting that additional mechanisms are at play. CgPho4's DBD, either alone or in combination with R1 and the NLS region, did not contribute to increased activity.

A double-edged sword: Pho2 Interaction Domain (P2ID) in ScPho4 allows for enhanced activity with Pho2 but restricts its activity without Pho2

In the 1- and 2- region swap heatmaps (Fig. 5A), region 4, aka P2ID, showed a puzzling pattern: swapping CgPho4's P2ID (P2ID:Cg) into ScPho4 background had a dominant negative effect on A_{PHO2}; it also offered little increase in $A_{pho2\Delta}$. We interpreted this as evidence for ScPho4 requiring both P2ID and Pho2 to function. In the meantime, swapping ScPho4's P2ID (P2ID:Sc) into CgPho4 increased A_{PHO2} beyond that of CgPho4, especially when combined with DBD:Sc, suggesting that P2ID:Sc allowed Pho4 from either species to take advantage of Pho2's assistance. To investigate the unique property of the P2ID further, we plotted all chimeras based on their A_{PHO2} and $A_{pho2\Delta}$ values (Fig. 6A). Strikingly, the chimeras fell into three distinct

groups based on the identity of their P2IDs. The first group had P2ID from CgPho4, and their activities were not dependent on Pho2, falling along the diagonal line. The second group had P2ID from ScPho4. These had strong activities with Pho2 and low to no activity without, similar to ScPho4. The third group included additional chimeras with mixed P2ID and they filled the intermediate space between the first two groups. Notably, many chimeras in the second group showed higher A_{PHO2} than ScPho4. Many of these chimeras contained regions 1 and 3 from CgPho4, reflecting our results above showing that these two regions enhance the activity of ScPho4-majority chimeras in the presence of Pho2 (Fig. 5B).

The evidence above suggests that P2ID:Sc has two effects: allowing for collaborative regulation with Pho2 and restricting Pho4's activity without Pho2. Can these two functions be separated? To answer this question, we compared the mixed P2ID chimeras to those with whole P2IDs from either Pho4. We found that swapping P2ID:Sc into CgPho4 resulted in a 1.2-fold increase in A_{PHO2} and a 71% reduction in $A_{pho2\Delta}$ (Fig. 6B, row 2 vs 3). Interestingly, swapping just the second half of P2ID:Sc (aa 205-242) into CgPho4 resulted in a ~40% reduction in A_{PHO2} and 75% reduction in $A_{pho2\Delta}$. By contrast, just swapping the first half of P2ID:Sc (aa 177-204) resulted in 1.76-fold increase in A_{PHO2} and maintained the same level of $A_{pho2\Delta}$ (1.1-fold). From this, we deduced that the first half of P2ID:Sc mainly functions to get Pho2's help while the second half appears to limit Pho4's activity without Pho2. The same trend was observed in another set of chimeras with lower activities (Fig. 6B, rows 6-9).

When examining the scatterplot in Fig. 6A, we noted that a number of chimeras had very low activities even with Pho2 ($A_{PHo2} < 3.6$, or 20% of A_{PHO2} for ScPho4). Most of these chimeras have P2ID:Cg and DBD:Sc (Fig. 6C, rows 3-7). We hypothesized that DBD:Sc requires P2ID:Sc and Pho2 to function in the context of the full length Pho4, which was surprising since DBD from both Pho4 were able to bind DNA *in vitro* (Fig. 2). But this hypothesis would also explain the dominant negative effect of P2ID:Cg swaps in the ScPho4 background (Fig. 5A). Under this hypothesis, we predicted that if we put P2ID:Sc back by inserting it in between P2ID:Cg and

DBD:Sc, we would rescue the non-functional chimeras in the presence of Pho2. Indeed, they did (Fig. 6C, rows 8-12), with some even exceeding the activities of ScPho4 and CgPho4. However, all of them still required Pho2. This supports our previous conclusion that P2ID:Sc, likely its second half, has a suppressive effect on the entire Pho4.

Given that chimeras with P2ID:Cg have equal activities with and without Pho2, we asked if CgPho4 still physically interacts with Pho2 in *S. cerevisiae*. Using the yeast two-hybrid assay, we were unable to detect an interaction between CgPho4ΔbHLH (aa 2-463) and a region of ScPho2 known to interact with multiple TFs, including ScPho4 (Bhoite et al. 2002) (Fig. 6D). By contrast, ScPho4ΔbHLH (aa 2-250) was able to interact with the same region of ScPho2 as previously found (Fig. 6D). This is consistent with our observation above, where chimeras with P2ID:Cg and lacking the region from ScPho4 mediating Pho2-interaction cannot be enhanced by Pho2 (Fig. 6A).

In summary, our chimera dissection revealed three key regions behind the difference in co-TF dependence. Notably, all three are IDRs. Among them, a region adjacent to the bHLH in ScPho4 (P2ID:Sc) functions as a double-edged sword: it both allows Pho4 to gain activity with Pho2's help and restricts it when Pho2 is absent. We showed that these two functions are encoded by physically separate parts, offering a path for determining their respective mechanisms of actions. CgPho4 lacks the ability to interact with and use Pho2's help. Instead, the two AEDs we identified earlier - both IDRs - conferred higher activity independent of Pho2. This, combined with the lack of the autoinhibition by P2ID:Sc, makes CgPho4 as active as ScPho4 and not dependent on the co-TF.

DISCUSSION

In this work, we investigated the genetic and mechanistic basis for natural variation in co-TF dependence between two Pho4 orthologs. We found that CgPho4's DBD and two AEDs conferred enhanced DNA binding and activation capability, supporting the enhanced activity

model. In the meantime, its P2ID provides no or very weak interactions with Pho2 (Fig. 6D). For ScPho4, we found that its P2ID both allows for efficient Pho2 interaction and hence enhanced activity with Pho2, and limits Pho4's activity when Pho2 is absent. We showed that these two functions can be physically separated (Fig. 6B). This result supports the autoinhibition model, although a classic lock-and-key model is unlikely to be correct (Fig. 5). Combining these results, we come to the conclusion that ScPho4 needs Pho2 for cooperative binding to its DNA motif and to enhance its transactivation potential. CgPho4 has enhanced capabilities at both tasks, and lacks the autoinhibitory interaction in ScPho4, explaining its lower Pho2-dependence.

While the DNA binding affinities of the orthologous Pho4s to the consensus DNA were different, the actual magnitude of the change was not dramatic (Fig. 2D). Whether such a difference has a significant effect *in vivo* would depend on the nuclear concentration of the TF – when the concentration is much higher than both Pho4's K_D, the difference would have little effect on its own. But if the concentration is close to K_D, a 3-4 fold difference could make a significant difference. Since the chimeras with the CgPho4 DBD failed to show increased activity, we conclude that in our experimental setup – due to the lack of *pho80*, all chimeras are constitutively inside the nucleus – the difference in DBD's binding affinity alone doesn't have a significant effect on Pho2-dependence. The contribution of the additional activation domains in CgPho4 is more profound and evident in the chimera data (Fig. 5A, B). However, we still don't know how these AEDs work. It's possible that the ScPho4 AD and CgPho4 AD are mainly responsible for recruiting the general TFs and chromatin remodelers, while CgPho4 E1 and E2, although unable to recruit those cofactors alone, can stabilize the main AD or increase the strength of these interactions, hence increasing the overall activation potential. The method of recruitment and/or coactivators recruited probably also differs between the Pho4 orthologs since we were unable to narrow down a smaller region of residues required for activation in CgPho4 like we did with ScPho4 (Fig. 3C). The important residues in the CgPho4 AD are likely distributed throughout the AD region.

The role of the P2ID is also a critical and surprising finding. We know from previous literature that this region in ScPho4 is responsible for interaction with Pho2 (Hirst et al. 1994). We assumed that the Pho2 interaction was conserved in the two orthologs because in the published ChIP-seq data there is evidence for CgPho4 and Pho2 co-binding to promoters (He, Zhou, and O'Shea 2017). Since the CgPho4 construct we tested was unable to interact with Pho2 by yeast two-hybrid (Fig. 6D), we have a couple of possible explanations that require further testing. First, there could be a change in the type of cooperativity present between CgPho4 and Pho2. ScPho4 needs to have a strong physical interaction but it's possible that CgPho4 and Pho2 interact weakly through a DNA mediated mechanism. Alternatively, CgPho4's bHLH is participating in the interaction with Pho2 and so our construct, which excludes this region, was unable to interact.

The presence of an autoinhibitory intramolecular interaction also still needs to be further examined. We ruled out the possibility of a classic lock-and-key interaction, which we predicted would be broken by swapping any one region on the ScPho4 background, thus enhancing the activity without Pho2, which we didn't see (Fig. 5A). However, we did find evidence for a three-part module with ScPho4's P2ID, DBD, and Pho2, where if any one part was missing there was no activity (Fig 6B, C). This finding also goes against the typical view of TFs being modular, including the DBD, which is assumed to function on its own. In fact, both *in vitro* and recent *in vivo* studies support the ScPho4 DBD binding DNA when expressed without the rest of the TF (Donovan et al. 2023). However, a key difference between that result and ours is that the readout in our assay requires the full chimera to not only bind but also activate. So, biochemical binding may be insufficient to induce gene expression due to short dwell times or epistasis with the other parts of the TF.

In conclusion, our results provided a detailed molecular picture of how co-TF dependence is mediated and how evolution in functional domains, including in the IDR of the TF, can lead to changes in codependency and result in significant rewiring of the regulatory targets. Because

combinatorial control by multiple TFs is such a widespread and important mode of regulation in eukaryotes, understanding the mechanistic basis and the evolution of TF codependency is a critical problem that merits further investigations.

MATERIALS AND METHODS

All plasmids and strains are listed in Tables 1 and 2. Computational and statistical analyses performed in this study are available at https://github.com/binhe-lab/E013-Pho4-evolution, which will be archived using Zenodo and minted with a DOI at the time of publication.

Bioinformatic analyses of Pho4 orthologs

Pho4 ortholog sequences were from the Yeast Gene Order Browser (http://ygob.ucd.ie/), and were aligned using ProbCons (Do et al. 2005) via JalView's Web Service (Waterhouse et al. 2009; Troshin et al. 2018). The alignment was manually edited to align the five Pho80/85 motifs (O'Neill et al. 1996). Secondary structures for ScPho4 and CgPho4 were predicted using the PSIPRED 4.0 (Buchan and Jones 2019). 9aaTAD motifs were predicted using the webapp https://www.med.muni.cz/9aaTAD/. For ScPho4, one match was found using the moderately stringent pattern, located at aa 75-83; for CgPho4, one match was identified using the most stringent pattern, at aa 270-278, with three more using the moderately stringent patterns (aa 20-28, 24-32, 282-290). Among the latter three, aa 24-32 had a lower % match at 67% vs 83% for the others.

Identifying Pho4 binding motifs from ChIP-seq data

Chromatin Immunoprecipitation (ChIP-seq) for ScPho4 and CgPho4 were previously performed in *S. cerevisiae*, with both Pho4 expressed from the same endogenous ScPho4 locus with native regulatory sequences (He, Zhou, and O'Shea 2017). ChIP identified peaks for both Pho4 were downloaded from the supplementary files of the above publication. The sequences under each peak were extracted from *S. cerevisiae* genome (sacCer3, NCBI refseq assembly GCF_000146045.2), which were submitted to the "peak-motifs" tool without control sequences

on the RSAT Fungi server (https://rsat.france-bioinformatique.fr/fungi/) (Thomas-Chollier et al. 2011). The top motif was reported for each Pho4 ortholog.

Protein expression and purification for DNA binding assays

All recombinant proteins were expressed using pET-11a (Sigma #69436-3) based vectors in BL21(DE3) *E. coli* cells. Vector maps are available upon request. The DNA binding domain (DBD) constructs included ScPho4 DBD-6xHIS (aa 236-312) and CgPho4 DBD-6xHIS (aa 452-533) cloned downstream of the T7 promoter in pET-11a. The transformed bacterial strains were grown overnight in LB + 100 ug/mL ampicillin. The overnight culture was diluted to OD 0.1 and the cells were grown to OD 0.6 and induced with 1mM IPTG for 2 hours. Induced cells were collected by centrifugation at 6000 rpm at 4ºC for 35 minutes, and were snap frozen in liquid nitrogen, then stored at -80ºC until purification. Our preliminary experiments showed that the protein of interest was largely in the insoluble fraction. Therefore, we performed a refold protocol on the Ni-NTA column. Briefly, frozen pellets were resuspended in 1xPBS, pH7.4 with cOmplete EDTA-free protease inhibitor (Sigma #11836170001) and sonicated for 45 cycles of 1 sec on / 2 sec off repeated 3 times at 50% power to lyse the cells. Sonicated samples were centrifuged at 35k rpm for 30 minutes. The supernatant was discarded, and the pellet was resuspended in the solubilization buffer (20 mM TrisBase, 0.5 M NaCl, 5 mM imidazole, 5.5 M Guanidine Hydrochloride, 1 mM 2-mercaptoethanol pH 8.0) and stirred at ~25ºC for an hour. The solubilized pellet was centrifuged at 35k rpm for 30 minutes and the supernatant was filtered and loaded onto a 5mL Ni-NTA column. The column was washed with a urea buffer (20 mM TrisBase, 0.5 M NaCl, 20 mM imidazole, 5.5 M Urea, 1 mM 2-mercaptoethanol pH 8.0). Then, the protein was refolded with a reverse urea gradient from buffer A (20 mM TrisBase, 0.5 M NaCl, 20 mM imidazole, 5.5 M Urea, 1 mM 2-mercaptoethanol pH 8.0) to buffer B (20 mM TrisBase, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0) on a BioRad FPLC. The refolded protein was eluted off of the Ni-NTA column using a gradient from buffer C (20 mM TrisBase, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0) to buffer D (20 mM

TrisBase, 0.5 M NaCl, 500 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0). Fractions containing the protein of interest were identified by gel electrophoresis, pooled, and diluted with a no salt buffer (25 mM Na_2HPO_4 pH 7.0, 0.5 mM THP) and run on a Heparin column equilibrated with low salt buffer (25 mM Na_2HPO_4 pH 7.0, 0.15 M NaCl, 0.5 mM THP). Protein was eluted with a gradient from low to high salt buffer (25 mM Na_2HPO_4 pH 7.0, 1.5 M NaCl, 0.5 mM THP). Protein-containing fractions were concentrated using a 3 kDa cutoff Amicon centrifugal filter (Sigma UFC8003) and loaded onto a Superdex 75 size exclusion column equilibrated with a storage buffer (25 mM Na_2HPO_4 pH 7.0, 0.5 M NaCl, 0.5 mM THP). Fractions containing the expected size products were collected, analyzed by gel electrophoresis, and stored in the storage buffer at 4ºC.

N-GST-CgPho4 full length was constructed for the Protein Binding Microarray assay. The corresponding N-GST-ScPho4 purification was described before (Zhang et al. 2021). Both proteins were grown and lysed as described in (Zhang et al. 2021). Briefly, BL21 *E. coli* cells containing the constructs were induced at OD600 0.8 with 1mM IPTG and collected by centrifugation after 3 hours of induction at 30ºC. Pellets were snap frozen in liquid nitrogen and stored at -80ºC until purification. Cells were lysed with rLysozyme (Millipore 71110) for 20 minutes at room temp in the lysis buffer (1xPBS, pH7.4) with the cOmplete protease inhibitor tablet and 1 mM PMSF. The protein was run on a GST column equilibrated with the lysis buffer and eluted with a gradient to buffer B (50mM Tris, 10mM glutathione, pH 8). GST-CgPho4 fractions containing the protein were pooled and loaded onto a heparin column equilibrated with low salt buffer (25mM Na_2HPO_4 pH 7.4, 150mM NaCl, 0.5mM THP) and eluted with a gradient to high salt buffer (25mM sodium phosphate dibasic pH 7.4, 1.5M NaCl, 0.5mM THP). Fractions containing the protein were pooled and concentrated with a 30 kDa cutoff Amicon centrifugal filter (Sigma UFC9030) and loaded onto a Superdex 200 size exclusion column equilibrated with the storage buffer (25mM HEPES pH 7.4, 500mM NaCl, 0.5mM THP). Protein was run on an

SDS-PAGE and pure fractions were pooled and concentrated. 10% glycerol was added before snap freezing and storage at -80ºC.

Universal Protein Binding Microarray (uPBM)

The uPBM was performed following the PBM protocol as described in (Berger and Bulyk 2009). GST-ScPho4 and GST-CgPho4 (full length) were prepared as described above. Both proteins were assayed at a final concentration of 1 uM as determined by the optical absorbance. An 8x15k array was used to assay all possible 9-mers, from which a robust 7-mer enrichment score is derived. The non-parametric enrichment score, or E-score, is invariant to differences in the concentration of the proteins used in the assay, and thus are suitable for comparisons of relative affinities between arrays. E-score ranges from -0.5 (lowest enrichment) to +0.5 (highest enrichment). Scores greater than 0.35 corresponds to specific TF-DNA binding (Berger and Bulyk 2006; Gordân et al. 2013). Data analysis was performed using custom Perl scripts as described in (Berger and Bulyk 2009).

Genomic Context Protein Binding Microarray (gcPBM)

All probes in the DNA library are 60 bp in length with 24 bp complementary to the primer used for double stranding and 36 bp of genomic region centered on the E-box motif or its variant. The library contains 1) 5711 *S. cerevisiae* genomic regions with putative Pho4 binding sites; 2) 150 negative controls from the *S. cerevisiae* genome not specifically bound by Pho4 (Zhang et al. 2021); 3) 4000 DNA sequences used in previous MITOMI experiments to calibrate the binding affinities (Maerkl and Quake 2007); 4) 150 genomic regions from *C. glabrata* that contain Pho4 consensus and nonconsensus binding sites (He, Zhou, and O'Shea 2017); 5) 100 probes from the library of sequences we tested with Biolayer Interferometry; and 6) 150 negative controls from the *C. glabrata* genome not bound by CgPho4. We used the NNNNGTG, CACNNN, and GTGNNN libraries from Maerkl and Quake (2007) in our gcPBM design. The MITOMI and BLI probes required the addition of random flanks to maintain the 36bp length;10 different flanking sequences were generated for each sequence. Each probe is represented by six replicates,

including three replicates in each orientation and were distributed randomly across the array. The custom 8x60k (8 chambers, 60,000 DNA spots per chamber) was synthesized by Agilent Technologies. The gcPBM was performed following the PBM protocol as described in (Zhang et al. 2021; Gordân et al. 2013). The log transformed median intensity of the 6 replicate probes were used for comparisons.

Biolayer interferometry (BLI)

A library of 17-bp dsDNA was used for the assay. The consensus probe had the sequence "CTAGTCC**CACGTG**TGAG", with the E-box motif bolded, and was identical to the DNA used in the crystal structure of ScPho4's bHLH domain (Shimizu et al. 1997). Nine half-site variants, including "AACGTG, TACGTG, … CAAGTG, ACGGTG, CATGTG" were constructed on this background. For each probe, the complementary ssDNA oligos were synthesized by Integrated DNA Technologies (IDT). One of the two probes was biotinylated on the 5' end. To anneal them into dsDNA, 1 pmol (5 uL of 200 uM) of the biotinylated oligo was mixed with 2 pmol (10 uL of 200 uM) of the complementary, unmodified strand in the Nuclease-Free Duplex Buffer (30 mM HEPES, pH 7.5; 100 mM potassium acetate). The mixture was heated to 95ºC and then left at room temperature for it to cool down. Annealed probes were stored at -20ºC until use.

ScPho4 DBD and CgPho4 DBD were purified and stored as described above. Protein quality was checked weekly for signs of degradation using Dynamic Light Scattering (DLS) on a DynaPro NanoStar instrument. Before each experiment, the concentration of the protein prep was measured in triplicates using a NanoDrop instrument. The mean concentration was used to prepare the protein dilutions and calculate K_D.

BLI experiments were performed on an Octet RED 96 instrument at 30ºC with 1000 rpm shaking. For each protein-probe pair, eight streptavidin (SA) biosensors were hydrated for 15-20 min at room temperature in the 1x kinetics buffer (1x PBS pH 7.4, 0.01% BSA, 0.002% Tween-20, 0.005% Sodium azide). A black 96 well flat-bottom plate was loaded with experimental components also diluted in the 1x kinetics buffer. To begin the experiment, biosensors were

equilibrated in a 1x kinetics buffer for 60 seconds to reach the baseline. Seven biosensors were then submerged in 35 nM biotinylated annealed DNA for 30 seconds, while one biosensor was submerged in the 1x kinetics buffer as a no-DNA control. Biosensors were then submerged in 1 ug/ml biocytin for 60 seconds to block any empty streptavidin pocket on their surface, before being dipped back into a 1x kinetics buffer for 60 seconds for a baseline measurement. Loaded and blocked biosensors were then submerged into a gradient of protein concentrations calculated for each probe based on the K_D. To obtain reliable measurements, we use a concentration range that spans 10x to 1/10 of the K_D values. Biosensors stayed in the protein solution for 900-1000 seconds or until equilibrium was reached.

Data analysis was performed in the ForteBio Data analysis v11 software. After subtracting the background and aligning the y-axis, the processed data were subjected to either a steady state analysis or a kinetic curve fitting. For steady state analysis, the equilibrium-level signal from each biosensor was plotted against the protein concentration, from which K_D was calculated. Kinetic curve fitting was done using a one-site specific binding with hill slope model as implemented in the Data analysis v11 software. The latter more effectively fit the BLI data and thus were utilized for most of the analysis. For the consensus sequence, we found the kinetic curve-based estimates for both ScPho4 and CgPho4 to have higher variance, likely due to the fast kinetics not adequately captured by the model (Weeramange et al. 2020). Steady-state analysis gave consistent mean K_D as the kinetic analysis did, but resulted in lower variance, and was used instead. Two to four replicates were performed for each 17bp DNA library sequence, using at least two independent protein preps.

Electrophoretic Mobility Shift Assay (EMSA)

IR700 labeled 17bp consensus DNA (same as the consensus DNA for BLI above, Integrated DNA technologies) was diluted to 0.1 nM and mixed 1:1 with a 2x dilution series of the DBD of interest starting at 32nM in binding buffer (20mM Tris-HCL, pH 8, 150mM KCl, 10% glycerol, 5mM MgCl2, 1mM EDTA, and 1mM DTT). The mixture was incubated for 1 hour at 4ºC. A 10%

native PAGE gel in 1x Tris Glycine buffer (2.5 mM Tris pH 8.3, 19.2 mM glycine) was prerun at 200V for 20 min. The DNA/protein mixture was then loaded onto the gel and ran for 25-35 minutes. The gel was imaged using the Odyssey FC imager in the 700 nm channel. To estimate K_D, the unbound band in each protein-containing lane was quantified using the ImageStudio software (with background subtraction), and divided by the no protein control. This value was then subtracted from 1 and plotted against the protein concentrations. A nonlinear curve fitting was performed in Prism v10.2.1 using the one site specific binding model. K_D estimates and the standard deviations were reported.

Yeast media and growth

Yeast cells were grown in Yeast extract-Peptone Dextrose (YPD) medium or Synthetic Complete (SC) medium, using Yeast Nitrogen Base without amino acids (Sigma Y0626) supplemented with 2% glucose and amino acid mix. Phosphate starvation medium was made using Yeast Nitrogen Base with ammonium sulfate, without phosphates, without sodium chloride (MP Biomedicals, 114027812) and supplemented to a final concentration of 2% glucose, 1.5 mg/ml potassium chloride, 0.1 mg/ml sodium chloride and amino acids, as described previously (He, Zhou, and O'Shea 2017). Phosphate concentration in the medium was measured using a Malachite Green Phosphate Assay kit (Sigma, MAK307).

Yeast strain and plasmid construction

The hosts containing the endogenous *PHO5*pr-mCherry reporter were constructed by replacing the endogenous coding sequence with mCherry using CRISPR/Cas9. Briefly, guide RNAs targeting *PHO5* were designed using Benchling and cloned into bRA89 (AddGene 100950), a plasmid that contains the Cas9 protein and gRNA scaffold (Anand, Memisoglu, and Haber 2017). Homology arms to the *PHO5* 5'UTR and 3' UTR were added onto an mCherry donor DNA using PCR. The donor DNA and plasmid were co-transformed into the host using the standard LiAc transformation protocol. Transformants were selected for the CRISPR plasmid using hygromycin resistance and screened with PCR. Positive clones were validated using

Sanger sequencing and fluorescence microscopy. Pho2 was then knocked out using a HIS3 cassette with homology arms to the *PHO2* 5'UTR and 3'UTR.

The chimeric Pho4 plasmid library was constructed using Golden Gate. The fragments of CgPho4 and ScPho4 were PCR amplified using Phusion polymerase with unique overhangs for Golden Gate. They were then assembled and inserted into a pRS315-based backbone that contains the Sc*PHO4* promoter, a C-terminal in-frame mNeon tag followed by the Sc*PHO4* 3' UTR and terminator. Primers were designed using the NEBridge Golden Gate assembly tool. The inserts were verified with PCR and Sanger sequencing and then transformed into the yeast hosts using either the standard LiAc protocol (Gietz and Schiestl 2007) or the Zymo yeast transformation kit (Zymo research T2001) and plated onto SD-leu media.

Flow cytometry

Cells were inoculated into a 96 deep-well plate (Fisher Scientific 07-200-700) and grown overnight in SD-Leu or SC medium supplemented to final concentrations of 0.13 mg/ml adenine and 0.1 mg/ml tryptophan to reduce autofluorescence (Lam, Steger, and O'Shea 2008). Cultures were diluted in the morning to an OD600 of 0.15 and grown to an OD600 of 0.6. These cells were directly subjected to flow cytometry at a flow rate of 25 uL/min on an Attune Nxt flow cytometer fitted with an autosampler. Data were collected for FSC, SSC and the appropriate fluorescence channels. Calibration beads (Spherotech RCP305A) were run routinely to ensure that experiments from different days were comparable. Pho4-mNeon was measured in the BL1 channel with 488 nm excitation and 510±10 nm emission; mCherry was measured in the YL2 channel using 561 nm excitation and 615±25 nm emission. Voltages for each channel were set by the brightest sample and negative control so the sample signals were between 10^2-10^5. Events were gated based on FSC / SSC to remove non-cells, then on the FSC-W / FSC-H to isolate singlets. At least 10,000 gated events were collected per sample. Each strain was measured at least three times, and on two different dates. A nonfluorescent strain was included in every experiment for subtracting the autofluorescence. Further gating and analyses were

performed in R using the FlowClust and FlowCore packages. Detailed analysis scripts are available in the project github repository.

Fluorescent microscopy for Pho4 nuclear localization

Six Pho4 constructs were chosen, with three bearing NLS:Sc (SSSSS, CCSCC, SSSCC) and three bearing NLS:Cg (SSCSS, CCCCC, SSCCS). They were transformed into either a *PHO2* wild type or *pho2*Δ background. Yeast cells were grown to mid-log phase in SD-Leu and fixed using 4% paraformaldehyde for 10 minutes, washed twice with 1xPBS before DAPI staining. 4',6-diamidino-2-phenylindole (DAPI, Sigma, D9542) was diluted in the respective medium and added to the fixed cells at a final concentration of 10 ug/mL. Cells were incubated with DAPI in the dark for 30 minutes before DAPI was removed and cells were washed twice with 1xPBS. Fixed cells were mounted on a 1.2% agarose pad on a depression slide for image. Fluorescent imaging was performed on a Leica epifluorescence microscope. A 405 nm laser was used for DAPI excitation and 430-550 nm for emission. For Pho4-mNeON, 488nm was used for excitation and 530 nm for emission. A bright field image was also taken to show the cell boundaries. A total of three images were thus recorded for each field of view.

Microscopy analysis was performed in ImageJ. To quantify the nuclear fraction of Pho4 proteins, ten cells were randomly selected for each construct in each host background, and the experiment was repeated two times, resulting in a total of 15 or 20 cells quantified in *PHO2* and *pho2*Δ backgrounds, respectively. Cell boundaries were manually traced using the bright field channel, and nucleus using the DAPI channel. Both were added to the ROI manager, and the raw integrated density (sum of pixel values) was quantified for both areas. The ratio of nuclear vs whole cell integrated density was used for plotting and statistical analysis.

Yeast one-hybrid and yeast two-hybrid

Indicated regions from CgPho4, ScPho4, and ScPho2 were PCR amplified and cloned into pGBD-C3 (James, Halladay, and Craig 1996) using Gibson Assembly. Plasmids were verified with PCR and Sanger sequencing. The *GAL1*pr-mCherry reporter was created using a pRS306

based integrative plasmid digested with StuI and inserted into the *ura3* locus of a *gal4Δ gal80Δ S. cerevisiae* strain. The reporter was tested using flow cytometry according to the protocol above using only the red (YL2) channel. Plasmids were then transformed into this yeast strain using the standard LiAc method (Gietz and Schiestl 2007) or the Zymo yeast transformation kit (Zymo research T2001) and selected for on SD-trp. For the yeast two-hybrid, the AD and DBD fusion plasmids were constructed by PCR amplifying the indicated regions and cloning into pGBD-C3 and pGAD-C3 (James, Halladay, and Craig 1996) using Gibson Assembly. The plasmids were co-transformed and selected for on SD -leu -trp. Positive colonies were patched onto fresh plates and grown at 30ºC for 24 hrs before replica plating onto SD -leu -trp -ade -his to test for interactions. Plates were imaged after 45-50 hours.

Statistical analyses

Binding affinity comparison by BLI (Fig. 2): K_D estimates from either steady state or kinetic curve fitting analyses were log10 transformed and a Student's t-test was used to compare ScPho4 vs CgPho4 DBDs against the same 17-bp DNA. Raw two-sided *P*-values and Holm-Bonferroni corrected *P* values for the 10 tests were reported; *Activation* (Fig. 3): median fluorescence intensity (MFI) for the *GAL1pr*-mCherry reporter was recorded from flow cytometry for each sample. For Fig. 3C, each Gal4 DBD fusion construct was compared to the background level. Because the host (with *GAL1pr*-mCherry but no Gal4 DBD plasmid) and Gal4 DBD alone (no region fused to it) showed a similar level of low MFI, we combined them as the reference group. A linear model was fit in R with the following command `lm(MFI ~ Genotype)`, where "Genotype" is a factor representing the constructs. This model estimates a common standard deviation for all constructs, and tests the significance of each construct against the reference group with a two-sided t-test. The raw *P*-values were corrected for multiple testing using the Holm-Bonferroni procedure. Constructs with a corrected *P*-value <= 0.05 were considered significant. For Fig. 3D, the first two groups were tested using a two-sample t-test, while the third involved multiple levels, and were tested as described above for Fig. 3C. A Holm-Bonferroni correction was

applied to the raw *P*-values for all six tests together. *Epistasis between regions 1-3 of CgPho4* (Fig. 5B): to determine the main and interaction terms in the region 1-3 swaps from CgPho4 on ScPho4's background, we fit a linear model, $Y = X_0 + R1 + AD + NLS + R1{:}AD + R1{:}NLS + AD{:}NLS + R1{:}AD{:}NLS$ in R, using the command `lm(A ~ R1 * AD * NLS)`, where "A" is either A_{PHO2} or $A_{pho2\Delta}$, and the independent variables are coded as ScPho4 = 0, CgPho4 = 1. The raw *P*-values were adjusted for multiple testing using the Holm-Bonferroni procedure. *Split P2ID* (Fig. 6B): The same procedure as applied to the activation region test above (Fig. 3) were applied to the split P2ID swaps. Two sets of four chimeras were chosen, each with a reference construct having P2ID:Cg, one with the entire P2ID swapped for the ScPho4 version, and two with the first or second half of P2ID swapped for the ScPho4 version. The latter three were compared to the reference using a linear model `lm(MFI ~ Genotype)`, for A_{PHO2} and $A_{pho2\Delta}$ separately. Holm-Bonferroni correction was applied to the two sets combined (6 tests in total), again separately for A_{PHO2} and $A_{pho2\Delta}$. A corrected $P < 0.05$ was considered significant. *Nuclear fraction of Pho4 chimeras* (Fig. S4): The ratio of the sum of pixel values in the GFP channel inside the nucleus vs the whole cell was treated as the response variable, and the identity of the NLS region ("Cg" vs "Sc") as the predictor. First, a two-way ANOVA was performed with the formula `lm(Nuc_frac ~ NLS + Host)`, where Host is either *PHO2* or *pho2*Δ. Second, given that ratios are often non-normally distributed, we performed the non-parametric Kruskal-Wallis test with the command `kruskal.test(Nuc_frac ~ NLS)` separately in the *PHO2* and *pho2*Δ hosts. The ANOVA F-test *P*-value (0.13) was reported in the results. The two *P*-values for the second test were 0.42 (*PHO2*) and 0.36 (*pho2*Δ).

ACKNOWLEDGEMENT

We would like to thank Dr. Miles Pufall's lab for teaching us EMSA. Dr. Jan Fassler shared many yeast plasmids and techniques with us, and critically read the thesis chapter, on which this manuscript is based. We would like to acknowledge the use of resources at the Protein and Crystallography Facility within the Carver College of Medicine at the University of Iowa and

thank Lokesh Gakhar, Zhen Xu, and Devin Reusch for assistance with protein purification and BLI assays.

FIGURE LEGENDS

Fig. 1 Different co-TF (Pho2) dependence between orthologous TF (Pho4) in yeast species. (A) Cartoon summary of previous work: *S. cerevisiae* Pho4 (ScPho4) depends on the co-TF, Pho2, for DNA binding and activation (top row), while Pho4 from *C. glabrata* (CgPho4) can bind and activate without the co-TF. See text for references. (B) Two *non-mutually exclusive* models to explain the divergence in Pho2-dependence. (C) The tiles in the middle show the conservation scores at aligned positions of the two Pho4 orthologs; the box diagrams represent ScPho4 and CgPho4, with gray lines connecting matching residues to the alignment tiles. Five regions were delineated based on knowledge about ScPho4: R1 = regulatory, interact with Pho80; AD = activation domain; NLS = nuclear import and export signals; P2ID = Pho2-interaction domain (also includes a region contacting Pho80); bHLH = basic Helix-Loop-Helix DNA binding domain. The lines above and below the boxes are a 53 aa tiling window prediction for transactivation potentials. The cartoon on the outside shows predicted alpha helices in each protein, with lighter color corresponding to lower prediction confidence.

Fig. 2 CgPho4 DBD recognizes the same DNA motif as ScPho4 DBD, but binds ~4 fold tighter to the consensus. A) Motifs generated from ChIP-seq binding peaks for ScPho4 and CgPho4 in the *S. cerevisiae* genome, showing a conserved CACGTG core motif. B) Universal Protein Binding Microarray (uPBM) comparing the binding of the two Pho4's. A rank-order based E-score is plotted for all 7-mers, where a score greater than 0.35 indicates specific TF-DNA binding. C) Genomic context PBM (gcPBM) measuring binding to ChIP-identified genomic regions with or without a consensus motif (Cg=*C. glabrata*, Sc=*S. cerevisiae*). Data plotted as the natural log of signal intensities for both TFs. D) Dissociation constants (K_D) determined by

Biolayer interferometry (BLI) for 17-bp oligos containing either the consensus or nine one-bp-off half-site variants. The red bar indicates the mean of >=2 biological replicates. Asterisks next to the sequence indicate a significant difference by a two-sided Student's t-test at a 0.05 level after Holm-Bonferroni correction.

Fig. 3 CgPho4 encodes two Activation Enhancer Domains (AEDs) that can enhance the activation potential of the main AD. (A, B) PADDLE prediction of activation potential in ScPho4 and CgPho4. Z-score > 4 and 6 indicate regions with medium and strong activation potential. The five functional regions are labeled as in Fig. 1. Orange triangles mark the 9aaTAD motif matches. Below the plots are diagrams of the Gal4 DBD fusion constructs used to test activation, with their boundaries labeled on the right. C) Activation potential of each region was tested in the context of a Gal4 DBD fusion driving a *GAL1pr*-mCherry reporter. Shown are the mean and standard deviation of >3 biological replicates. "host" has the mCherry reporter but not the Gal4 DBD fusion plasmid; "Gal4_DBD" has the reporter and the Gal4 DBD alone. ANOVA was used to compare the Gal4 DBD fusions to the two negative controls combined; an asterisk indicates significant difference at a 0.05 level after Holm-Bonferroni correction. D) CgE1 and CgE2_9aa can enhance the activation capability of ScAD_9aa in an orientation-independent manner. An asterisk indicates significance difference at a 0.05 level by a t-test between the indicated groups after Holm-Bonferroni correction.

Fig. 4 A dual fluorescence reporter system for accurately quantifying the activity of Pho4 chimeras with and without Pho2. (A) ScPho4 and CgPho4 were divided into five corresponding parts. Names of each region were based on the annotations for ScPho4. Predicted alpha helices were drawn as cartoons outside of the box diagrams. All $2^5 = 32$ combinations (2 were the original proteins) were made in the main chimera set, and an additional set of chimeras were constructed to test specific hypotheses. (B) A diagram showing

the design of the dual fluorescence reporter system. The chimeric Pho4 fragments were cloned downstream of the endogenous *ScPHO4* promoter and were tagged at the C-terminus with an in-frame mNeon fluorescent protein, followed by the endogenous *ScPHO4* 3'UTR and terminator. The plasmid contains a CEN/ARS element, resulting in stable inheritance and low biological variation. (C) The *PHO5pr-mCherry* levels correlate with the ScPho4-mNeon levels between biological replicates. The linear regression line is shown in blue and a 95% confidence band as a gray ribbon. (D) A_{PHO2} and $A_{pho2\Delta}$ values for ScPho4 and CgPho4 across 6 different days of experiments, with multiple biological replicates per day, showing consistent behaviors.

Fig. 5. Effects of 1- or 2-region swaps reveal that R1 and NLS from CgPho4 enhance the activity, but the chimeras remain Pho2-dependent. (A) Heatmaps showing the effects of 1- or 2-region swaps from CgPho4 into ScPho4 (left) or vice versa (right). The top triangle shows the effects with Pho2. The triangles at the base show the 1-region swap effects, while the squares on the top show 2-region swap effects (see a and b in the key). The bottom triangle shows the corresponding effects measured in the *pho2*Δ host. (B) A subset of the chimeras was shown with their A_{PHO2} and $A_{pho2\Delta}$ values normalized to those of ScPho4 (Rel. A_{PHO2} or $A_{pho2\Delta}$). The error bars represent the standard deviation of the mean for A_{PHO2} and $A_{pho2\Delta}$. The set in rows 3-9 show the effect of swapping region 1-3 from CgPho4 into ScPho4, both with and without Pho2. Rows 10-12 show the effects of swapping CgPho4's DBD alone or in combination with other regions. The last row shows a chimera with an alternative breakpoint, where the 4th region included CgPho4 P2ID and the last 30 aa of the NLS region, which coincided with the second activation boosting domain. C) Epistasis analysis of CgPho4 regions 1-3. Linear regression with a full interaction model was fit to the data. Shown are estimates of the individual and interaction terms (bar and point) and standard error (line), with dark gray indicating a Holm-Bonferroni corrected $P < 0.05$. All effects are calculated using ScPho4 as the reference.

Fig. 6 P2ID in ScPho4 allows for enhanced activation with Pho2 via physical interaction with the co-TF. (A) Scatter plot of all chimeric Pho4 constructs based on their A_{PHO2} and $A_{pho2\Delta}$ values, colored by the identity of their P2ID region. Mixed P2ID uses an additional breakpoint that separates that region into two halves (Fig. S2). (B) Comparing the effects of swapping the whole P2ID from ScPho4 into CgPho4 (gray shaded) vs individual halves (light green shaded). Two sets were tested (rows 2-5 and 6-9). Two-sided t-tests were performed as indicated by the bars and lines to the right. Holm-Bonferroni correction was applied to the raw *P*-values for both sets, separately for A_{PHO2} and $A_{pho2\Delta}$. An asterisk indicates corrected $P < 0.05$. (C) Testing the hypothesis that ScPho4 DBD requires the P2ID:Sc and Pho2 to be fully functional. Compared with the chimeras with P2ID:Cg + P2ID:Sc, which are all non-functional even with Pho2, adding back P2ID:Sc (”4+5”) rescued the activity with Pho2, but not without. (D) Pho4-Pho2 physical interaction was tested with a yeast two-hybrid assay. The bait and target regions were fused to the activation domain (AD) and DNA binding domain (DBD) of the yeast Gal4 protein. Interaction between them reconstitutes a functional Gal4 and allows cells to grow on the dropout media (right column), while all strains are able to grow on the rich media (left column). “Vector” indicates a Gal4 DBD or Gal4 AD only plasmid without a bait or target fused to it, and is included as a control. Two independent transformants are shown.

Fig. S1 ScPho4 and CgPho4 structural prediction shows both proteins are largely disordered outside the DBD. (A, B) Secondary structure predictions for ScPho4 and CgPho4 using the PSIPRED 4.0 server. (C, D) AlphaFold2 predicted monomer 3D structure for ScPho4 and CgPho4. Color represents model confidence. It has been shown that low and very low confidence regions overlap significantly and are predictive of protein disorder.

Fig. S2 ScPho4 and CgPho4 alignment and annotation. Protein sequence alignment was generated using ProbCons with 2 passes of consistency transformation and 100 iterative

refinements, and was manually edited to align the known Pho85 recognition motifs (red boxes). The shading behind the amino acid symbols represents level of conservation based on BLOSUM62 scores. The functional annotation on the top were based on ScPho4 (Uniprot #P07270). The orange boxes and texts refer to the five breakpoints used for constructing chimeric Pho4 constructs (Fig. 5, 6).

Fig. S3 Validation of difference in binding affinity to consensus motif. (A) EMSA with ScPho4 DBD and IR700 labeled 17bp consensus DNA. Lanes 1-7 contain 2x diluted protein; lane 8 is a no protein (NP) control. (B) Fraction of bound DNA was calculated from the unbound DNA band normalized to the NP lane. Fitting was done with a one site specific binding model in GraphPad Prism v10. C) EMSA with CgPho4 DBD and IR700 labeled 17bp consensus sequence. D) Same as (B) for CgPho4 DBD.

Fig. S4 NLS from ScPho4 and CgPho4 conferred similar levels of nuclear localization. Six Pho4 constructs with either NLS:Sc or NLS:Cg were quantified for their nuclear fraction in either *PHO2* or *pho2*Δ host background. Live fluorescence imaging was performed on mid-log phase cells. DAPI staining was used to identify the nucleus. Endogenous or chimeric Pho4 protein levels were quantified by the C-terminal mNeon tag. Integrated signal intensity inside the nucleus was divided by total integrated intensity across the entire cell for 35 cells in each construct. ANOVA with the identity of the NLS ("Cg" or "Sc") and Host (*PHO2* or *pho2*Δ) found the first factor to be not significant (F-test raw $P = 0.13$), i.e., no evidence for significant differences in nuclear fraction between constructs with NLS:Sc vs NLS:Cg. There was a significant difference for the second factor (F-test raw $P = 0.0003$), showing higher nuclear fraction in the *pho2*Δ background than in the *PHO2* background.

TABLES

Table 1. Plasmids used in this study

ID	Plasmid	Derived	Source	Comment
pH225	pRS306-Pmyo2-rtTA(S2)	pRS306	PMID: 18849996	Inducible Pho4
pH226	pTET07-ScPho4		PMID: 18849996	Inducible Pho4
pH238	TET07pr, CEN/ARS, LEU2	pH173	this study	Inducible Pho4
pH242	pTET07-ScPho4-mNeon	pH238	this study	Inducible Pho4
pH248	pTET07-CgPho4-mNeon	pH238	this study	Inducible Pho4
pH269	pTET07-Pho4 chimera (pH217)-mNeon	pH238	this study	Inducible Pho4
pH246	pGAD-C3		PMID:8978031, gift from Fassler lab	Yeast two-hybrid
pH247	pGBD-C3		PMID:8978031, gift from Fassler lab	Yeast one-hybrid
pH343	pRS306-GAL1pr-mCherry	pRS306	this study	Yeast one-hybrid
pH346	pRS314-GAL4pr-Gal4	pRS314	this study	Yeast one-hybrid
pH349	pGBD-Gal4DBD-ScE1	pH247	this study	Yeast one-hybrid
pH350	pGBD-Gal4DBD-CgE1	pH247	this study	Yeast one-hybrid
pH351	pGBD-Gal4DBD-ScE1-ScAD	pH247	this study	Yeast one-hybrid
pH352	pGBD-Gal4DBD-CgE1-CgAD	pH247	this study	Yeast one-hybrid
pH362	pGBD-Gal4DBD-CgAD	pH247	this study	Yeast one-hybrid
pH363	pGBD-Gal4DBD-ScAD	pH247	this study	Yeast one-hybrid
pH377	pGBD-Gal4DBD-CgE2	pH247	this study	Yeast one-hybrid
pH378	pGBD-Gal4DBD-ScE2	pH247	this study	Yeast one-hybrid
pH379	pGBD-Gal4DBD-ScAD.9aa	pH247	this study	Yeast one-hybrid
pH380	pGBD-Gal4DBD-CgAD.2	pH247	this study	Yeast one-hybrid
pH386	pGBD-Gal4DBD-CgE2.9aa	pH247	this study	Yeast one-hybrid
pH387	pGBD-Gal4DBD-CgAD.1	pH247	this study	Yeast one-hybrid
pH388	pGBD-Gal4DBD-CgE1-ScAD.9aa	pH247	this study	Yeast one-hybrid
pH389	pGBD-Gal4DBD-ScAD.9aa-CgE1	pH247	this study	Yeast one-hybrid
pH390	pGBD-Gal4DBD-ScAD.9aa-CgE2.9aa	pH247	this study	Yeast one-hybrid
pH391	pGBD-Gal4DBD-CgE2.9aa-ScAD.9aa	pH247	this study	Yeast one-hybrid
pH434	pGBD-Gal4DBD-ScPho2AD	pH247	this study	Yeast one-hybrid
pH394	pGAD-ScPho4ΔbHLH	pH246	this study	Yeast two-hybrid
pH395	pGBD-Pho2Pho4int	pH247	this study	Yeast two-hybrid
pH396	pGAD-CgPho4ΔbHLH	pH246	this study	Yeast two-hybrid
pH399	pGBD-CgPho2Pho4int	pH247	this study	Yeast two-hybrid
pH050	pET-11a-ScPho4 DBD-6xHis	pET11a	this study	BLI
pH051	pET-11a-CgPho4 DBD-6xHis	pET11a	this study	BLI

bH404	pET-11a-GST-CgPho4	pET11a	this study	PBM
pH073	bRA89, Cas9, CEN/ARS, HygR		PMID: 28405019, gift from Malkova lab	CRISPR
pH173	ScPHO4pr, mNeon, CEN/ARS, LEU2	pRS315	this study	Chimera backbone
pH188	Cg(1-533)	pH173	this study	Chimera
pH194	Sc(1-312)	pH173	this study	Chimera
pH209	Cg(1-112) Sc(100-176) Cg(283-533)	pH173	this study	Chimera
pH210	Cg(1-458) Sc(243-312)	pH173	this study	Chimera
pH211	Cg(1-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH212	Sc(1-42) Cg(45-533)	pH173	this study	Chimera
pH213	Cg(1-44) Sc(43-99) Cg(113-533)	pH173	this study	Chimera
pH215	Sc(1-99) Cg(113-533)	pH173	this study	Chimera
pH216	Sc(1-42) Cg(45-458) Sc(243-312)	pH173	this study	Chimera
pH217	Cg(1-44) Sc(43-176) Cg(283-533)	pH173	this study	Chimera
pH218	Cg(1-112) Sc(100-242) Cg(459-533)	pH173	this study	Chimera
pH219	Cg(1-282) Sc(177-312)	pH173	this study	Chimera
pH220	Sc(1-42) Cg(45-112) Sc(100-176) Cg(283-533)	pH173	this study	Chimera
pH221	Cg(1-44) Sc(43-99) Cg(113-282) Sc(177-312)	pH173	this study	Chimera
pH222	Sc(1-42) Cg(45-282) Sc(177-312)	pH173	this study	Chimera
pH223	Cg(1-44) Sc(43-99) Cg(113-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH224	Sc(1-42) Cg(45-112) Sc(100-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH227	Sc(1-176) Cg(283-533)	pH173	this study	Chimera
pH229	Sc(1-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH230	Sc(1-99) Cg(113-282) Sc(177-312)	pH173	this study	Chimera
pH231	Cg(1-44) Sc(43-312)	pH173	this study	Chimera
pH232	Sc(1-42) Cg(45-112) Sc(100-312)	pH173	this study	Chimera
pH233	Sc(1-242) Cg(459-533)	pH173	this study	Chimera
pH234	Cg(1-112) Sc(100-312)	pH173	this study	Chimera
pH235	Sc(1-99) Cg(113-458) Sc(243-312)	pH173	this study	Chimera
pH236	Cg(1-76) Sc(71-99) Cg(113-	pH173	this study	Chimera

	282) Sc(177-312)			
pH237	Cg(1-44) Sc(43-70) Cg(77-282) Sc(177-312)	pH173	this study	Chimera
pH239	Sc(1-99) Cg(113-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH240	Cg(1-112) Sc(100-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH241	Cg(1-44) Sc(43-242) Cg(459-533)	pH173	this study	Chimera
pH250	Cg(1-44) Sc(43-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH251	Sc(1-42) Cg(45-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH252	Sc(1-42) Cg(45-112) Sc(100-242) Cg(459-533)	pH173	this study	Chimera
pH253	Cg(1-44) Sc(43-99) Cg(113-458) Sc(243-312)	pH173	this study	Chimera
pH254	Cg(1-44) Sc(43-176) Cg(283-327) Sc(205-242) Cg(459-533)	pH173	this study	Chimera
pH255	Cg(1-44) Sc(43-204) Cg(328-533)	pH173	this study	Chimera
pH256	Cg(1-44) Sc(43-220) Cg(377-533)	pH173	this study	Chimera
pH257	Cg(1-282) Sc(177-204) Cg(328-533)	pH173	this study	Chimera
pH258	Cg(1-327) Sc(205-242) Cg(459-533)	pH173	this study	Chimera
pH265	Sc(1-204) Cg(328-533)	pH173	this study	Chimera
pH266	Sc(1-153) Cg(250-533)	pH173	this study	Chimera
pH276	Sc(1-42) Cg(45-282) Sc(177-204) Cg(328-533)	pH173	this study	Chimera
pH277	Cg(1-44) Sc(43-99) Cg(113-282) Sc(177-204) Cg(328-533)	pH173	this study	Chimera
pH278	Sc(1-204) Cg(328-458) Sc(243-312)	pH173	this study	Chimera
pH279	Cg(1-44) Sc(43-204) Cg(328-458) Sc(243-312)	pH173	this study	Chimera
pH294	Cg(1-470) Sc(251-312)	pH173	this study	Chimera
pH301	Sc(1-247) Cg(464-533)	pH173	this study	Chimera
pH326	Sc(1-156) Cg(253-533)	pH173	this study	Chimera
pH327	Cg(1-252) Sc(157-312)	pH173	this study	Chimera
pH328	Sc(1-42) Cg(45-252) Sc(157-312)	pH173	this study	Chimera
pH329	Cg(1-44) Sc(43-247) Cg(464-533)	pH173	this study	Chimera
pH330	Cg(1-44) Sc(43-156) Cg(253-463) Sc(251-312)	pH173	this study	Chimera
pH331	Cg(1-44) Sc(43-156) Cg(253-533)	pH173	this study	Chimera

pH332	Sc(1-42) Cg(45-463) Sc(251-312)	pH173	this study	Chimera
pH334	Sc(1-42) Cg(45-252) Sc(157-247) Cg(464-533)	pH173	this study	Chimera
pH436	Cg(1-112) Sc(100-176) Cg(283-458) Sc(177-312)	pH173	this study	Chimera
pH438	Sc(1-176) Cg(283-458) Sc(177-312)	pH173	this study	Chimera
pH440	Cg(1-44) Sc(43-99) Cg(113-458) Sc(177-312)	pH173	this study	Chimera
pH441	Sc(1-42) Cg(45-458) Sc(177-312)	pH173	this study	Chimera
pH442	Sc(1-99) Cg(113-458) Sc(177-312)	pH173	this study	Chimera

Table 2. Yeast strains

ID	Genotype	Source
yH373	pho4::NAT; pho2::HIS3; pho80::TRP1; pho5::mCherry	this study
yH555	pho4::NAT; his3::pRS303-HIS3; pho80::TRP1; pho5::mCherry	this study
yH589	pho4::NAT; his3::pRS303-HIS3; pho80::TRP1; pho5::mCherry; ura3::pRS306-Pmyo2-rtTA(S2)-URA3	this study
yH549	gal4delta gal80delta LYS2::GAL1-HIS3 GAL2-ADE2 MET2::GAL7-lacZ	PMID: 8978031, gift from Fassler lab
yH696	gal4delta gal80delta LYS2::GAL1-HIS3; GAL2-ADE2; MET2::GAL7-lacZ; ura3::pRS306-*GAL1pr*-mCherry-URA3	this study
yH295	pho4::NAT; pho2::HIS3; pho80::TRP1; ura3::pRS306-*PHO5pr*-mCherry-URA3	this study
yH296	pho4::NAT; pho80::TRP1; ura3::pRS306-*PHO5pr*-mCherry-URA3	this study
yH156	pho80::TRP1	PMID: 28485712
various	chimeric Pho4 plasmids transformed into yH373 and yH555	this study
various	yeast two-hybrid plasmids transformed into yH696	this study
various	yeast one-hybrid plasmids transformed into yH696	this study
various	select chimeric Pho4 plasmids transformed into yH295 and yH296	this study
various	inducible Pho4 plasmids transformed into yH589	this study

REFERENCES

Anand, Ranjith, Gonen Memisoglu, and James Haber. 2017. “Cas9-Mediated Gene Editing in Saccharomyces Cerevisiae,” April. https://www.nature.com/protocolexchange/protocols/5791#/figures.

Ang, Yen-Sin, Renee N. Rivas, Alexandre J.S. Ribeiro, Rohith Srivas, Janell Rivera, Nicole R. Stone, Karishma Pratt, et al. 2016. “Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis.” *Cell* 167 (7): 1734-1749.e22. https://doi.org/10.1016/j.cell.2016.11.033.

Barbaric, S., M. Münsterkötter, C. Goding, and W. Hörz. 1998. “Cooperative Pho2-Pho4 Interactions at the PHO5 Promoter Are Critical for Binding of Pho4 to UASp1 and for Efficient Transactivation by Pho4 at UASp2.” *Molecular and Cellular Biology* 18 (5): 2629–39. https://doi.org/10.1128/mcb.18.5.2629.

Berger, Michael F., and Martha L. Bulyk. 2006. “Protein Binding Microarrays (PBMs) for Rapid, High-Throughput Characterization of the Sequence Specificities of DNA Binding Proteins.” In *Gene Mapping, Discovery, and Expression*, by Bina Minou, 338:245–60. New Jersey: Humana Press. https://doi.org/10.1385/1-59745-097-9:245.

Berger, Michael F, and Martha L Bulyk. 2009. “Universal Protein-Binding Microarrays for the Comprehensive Characterization of the DNA-Binding Specificities of Transcription Factors.” *Nature Protocols* 4 (3): 393–411. https://doi.org/10.1038/nprot.2008.195.

Bhoite, Leena T., Jason M. Allen, Emily Garcia, Lance R. Thomas, I. David Gregory, Warren P. Voth, Kristen Whelihan, Ronda J. Rolfes, and David J. Stillman. 2002. “Mutations in the Pho2 (Bas2) Transcription Factor That Differentially Affect Activation with Its Partner Proteins Bas1, Pho4, and Swi5.” *Journal of Biological Chemistry* 277 (40): 37612–18. http://www.jbc.org/content/277/40/37612.

Buchan, Daniel W A, and David T Jones. 2019. “The PSIPRED Protein Analysis Workbench: 20 Years On.” *Nucleic Acids Research* 47 (W1): W402–7. https://doi.org/10.1093/nar/gkz297.

DelRosso, Nicole, Josh Tycko, Peter Suzuki, Cecelia Andrews, null Aradhana, Adi Mukund, Ivan Liongson, et al. 2023. “Large-Scale Mapping and Mutagenesis of Human Transcriptional Effector Domains.” *Nature* 616 (7956): 365–72. https://doi.org/10.1038/s41586-023-05906-y.

Do, Chuong B., Mahathi S. P. Mahabhashyam, Michael Brudno, and Serafim Batzoglou. 2005. “ProbCons: Probabilistic Consistency-Based Multiple Sequence Alignment.” *Genome Research* 15 (2): 330–40. https://doi.org/10.1101/gr.2821705.

Donovan, Benjamin T., Hengye Chen, Priit Eek, Zhiyuan Meng, Caroline Jipa, Song Tan, Lu Bai, and Michael G. Poirier. 2023. “Basic Helix-Loop-Helix Pioneer Factors Interact with the Histone Octamer to Invade Nucleosomes and Generate Nucleosome-Depleted Regions.” *Molecular Cell* 83 (8): 1251-1263.e6. https://doi.org/10.1016/j.molcel.2023.03.006.

Fascher, K D, J Schmitz, and W Hörz. 1990. “Role of Trans-Activating Proteins in the Generation of Active Chromatin at the PHO5 Promoter in S. Cerevisiae.” *The EMBO Journal* 9 (8): 2523–28. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC552282/.

Gietz, R Daniel, and Robert H Schiestl. 2007. “High-Efficiency Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method.” *Nature Protocols* 2 (1): 31–34. https://doi.org/10.1038/nprot.2007.13.

Gomes-Vieira, André L., Jeremy G. Wideman, Lisvane Paes-Vieira, Suely L. Gomes, Thomas A. Richards, and José Roberto Meyer-Fernandes. 2018. “Evolutionary Conservation of a Core Fungal Phosphate Homeostasis Pathway Coupled to Development in Blastocladiella Emersonii.” *Fungal Genetics and Biology* 115 (June): 20–32. https://doi.org/10.1016/j.fgb.2018.04.004.

Gordân, Raluca, Ning Shen, Iris Dror, Tianyin Zhou, John Horton, Remo Rohs, and Martha L. Bulyk. 2013. “Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape.” *Cell Reports* 3 (4): 1093–1104. https://doi.org/10.1016/j.celrep.2013.03.014.

Guss, K. A., C. E. Nelson, A. Hudson, M. E. Kraus, and S. B. Carroll. 2001. “Control of a Genetic Regulatory Network by a Selector Gene.” *Science (New York, N.Y.)* 292 (5519): 1164–67. https://doi.org/10.1126/science.1058312.

He, Bin Z., Xu Zhou, and Erin K. O’Shea. 2017. “Evolution of Reduced Co-Activator Dependence Led to Target Expansion of a Starvation Response Pathway.” *eLife* 6. https://doi.org/10.7554/eLife.25157.

Hirst, K., F. Fisher, P. C. McAndrew, and C. R. Goding. 1994. “The Transcription Factor, the Cdk, Its Cyclin and Their Regulator: Directing the Transcriptional Response to a Nutritional Signal.” *The EMBO Journal* 13 (22): 5410–20.

James, P., J. Halladay, and E. A. Craig. 1996. “Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast.” *Genetics* 144 (4): 1425–36. https://doi.org/10.1093/genetics/144.4.1425.

Jayaraman, P. S., K. Hirst, and C. R. Goding. 1994. “The Activation Domain of a Basic Helix-Loop-Helix Protein Is Masked by Repressor Interaction with Domains Distinct from That Required for Transcription Regulation.” *The EMBO Journal* 13 (9): 2192–99.

Jolma, Arttu, Yimeng Yin, Kazuhiro R. Nitta, Kashyap Dave, Alexander Popov, Minna Taipale, Martin Enge, Teemu Kivioja, Ekaterina Morgunova, and Jussi Taipale. 2015. “DNA-Dependent Formation of Transcription Factor Pairs Alters Their Binding Specificity.” *Nature* 527 (7578): 384–88. https://doi.org/10.1038/nature15518.

Kaffman, A., I. Herskowitz, R. Tjian, and E. K. O’Shea. 1994. “Phosphorylation of the Transcription Factor PHO4 by a Cyclin-CDK Complex, PHO80-PHO85.” *Science (New York, N.Y.)* 263 (5150): 1153–56.

Kerwin, Christine L., and Dennis D. Wykoff. 2009. “Candida Glabrata PHO4 Is Necessary and Sufficient for Pho2-Independent Transcription of Phosphate Starvation Genes.” *Genetics* 182 (2): 471–79. https://doi.org/10.1534/genetics.109.101063.

Komeili, Arash, and Erin K. O’Shea. 1999. “Roles of Phosphorylation Sites in Regulating Activity of the Transcription Factor Pho4.” *Science* 284 (5416): 977–80. https://doi.org/10.1126/science.284.5416.977.

Lam, Felix H., David J. Steger, and Erin K. O’Shea. 2008. “Chromatin Decouples Promoter Threshold from Dynamic Range.” *Nature* 453: 246–50.

Maerkl, Sebastian J., and Stephen R. Quake. 2007. “A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors.” *Science* 315 (5809): 233–37. https://doi.org/10.1126/science.1131007.

Magbanua, J. P., N. Ogawa, S. Harashima, and Y. Oshima. 1997. “The Transcriptional Activators of the PHO Regulon, Pho4p and Pho2p, Interact Directly with Each Other and with Components of the Basal Transcription Machinery in Saccharomyces Cerevisiae.” *Journal of Biochemistry* 121 (6): 1182–89.

Markson, Joseph S., Joseph R. Piechura, Anna M. Puszynska, and Erin K. O’Shea. 2013. “Circadian Control of Global Gene Expression by the Cyanobacterial Master Regulator RpaA.” *Cell* 155 (6): 1396–1408. https://doi.org/10.1016/j.cell.2013.11.005.

McAndrew, P. C., J. Svaren, S. R. Martin, W. Hörz, and C. R. Goding. 1998. “Requirements for Chromatin Modulation and Transcription Activation by the Pho4 Acidic Activation Domain.” *Molecular and Cellular Biology* 18 (10): 5818–27. https://doi.org/10.1128/MCB.18.10.5818.

Morgunova, Ekaterina, and Jussi Taipale. 2017. “Structural Perspective of Cooperative Transcription Factor Binding.” *Current Opinion in Structural Biology* 47 (December): 1–8. https://doi.org/10.1016/j.sbi.2017.03.006.

Nourani, Amine, Rhea T Utley, Stéphane Allard, and Jacques Côté. 2004. “Recruitment of the NuA4 Complex Poises the PHO5 Promoter for Chromatin Remodeling and Activation.” *The EMBO Journal* 23 (13): 2597–2607. https://doi.org/10.1038/sj.emboj.7600230.

Ogawa, N., and Y. Oshima. 1990. “Functional Domains of a Positive Regulatory Protein, PHO4, for Transcriptional Control of the Phosphatase Regulon in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 10 (5): 2224–36. https://doi.org/10.1128/mcb.10.5.2224-2236.1990.

Ogawa, Nobuo, Joseph DeRisi, and Patrick O. Brown. 2000. “New Components of a System for Phosphate Accumulation and Polyphosphate Metabolism in Saccharomyces Cerevisiae Revealed by Genomic Expression Analysis.” *Molecular Biology of the Cell* 11 (12): 4309–21. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC15074/.

O’Neill, E. M., A. Kaffman, E. R. Jolly, and E. K. O’Shea. 1996. “Regulation of PHO4 Nuclear Localization by the PHO80-PHO85 Cyclin-CDK Complex.” *Science (New York, N.Y.)* 271 (5246): 209–12.

Puszynska, Anna M, and Erin K O’Shea. 2017. “Switching of Metabolic Programs in Response to Light Availability Is an Essential Function of the Cyanobacterial Circadian Output Pathway.” *eLife* 6 (April): e23210. https://doi.org/10.7554/eLife.23210.

Sanborn, Adrian L, Benjamin T Yeh, Jordan T Feigerle, Cynthia V Hao, Raphael Jl Townshend, Erez Lieberman Aiden, Ron O Dror, and Roger D Kornberg. 2021. “Simple Biochemical Features Underlie Transcriptional Activation Domain Diversity and Dynamic, Fuzzy Binding to Mediator.” *eLife* 10 (April): e68068. https://doi.org/10.7554/eLife.68068.

Shao, D., C. L. Creasy, and L. W. Bergman. 1996. “Interaction of Saccharomyces Cerevisiae Pho2 with Pho4 Increases the Accessibility of the Activation Domain of Pho4.” *Molecular and General Genetics MGG* 251 (3): 358–64. https://doi.org/10.1007/BF02172527.

Shimizu, Toshiyuki, Atsuki Toumoto, Kentaro Ihara, Masato Shimizu, Yoshimasa Kyogoku, Nobuo Ogawa, Yasuji Oshima, and Toshio Hakoshima. 1997. “Crystal Structure of PHO4 bHLH Domain–DNA Complex: Flanking Base Recognition.” *The EMBO Journal* 16 (15): 4689–97. https://doi.org/10.1093/emboj/16.15.4689.

Staller, Max V., Alex S. Holehouse, Devjanee Swain-Lenz, Rahul K. Das, Rohit V. Pappu, and Barak A. Cohen. 2018. “A High-Throughput Mutational Scan of an Intrinsically

Disordered Acidic Transcriptional Activation Domain." *Cell Systems* 6 (4): 444-455.e6. https://doi.org/10.1016/j.cels.2018.01.015.

Tan, Song, and Timothy J. Richmond. 1998. "Crystal Structure of the Yeast MATα2/MCM1/DNA Ternary Complex." *Nature* 391 (6668): 660–66. https://doi.org/10.1038/35563.

Thomas-Chollier, Morgane, Carl Herrmann, Matthieu Defrance, Olivier Sand, Denis Thieffry, and Jacques van Helden. 2011. "RSAT Peak-Motifs: Motif Analysis in Full-Size ChIP-Seq Datasets." *Nucleic Acids Research*, December, gkr1104. https://doi.org/10.1093/nar/gkr1104.

Troshin, Peter V, James B Procter, Alexander Sherstnev, Daniel L Barton, Fábio Madeira, and Geoffrey J Barton. 2018. "JABAWS 2.2 Distributed Web Services for Bioinformatics: Protein Disorder, Conservation and RNA Secondary Structure." *Bioinformatics* 34 (11): 1939–40. https://doi.org/10.1093/bioinformatics/bty045.

Waterhouse, Andrew M., James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. 2009. "Jalview Version 2--a Multiple Sequence Alignment Editor and Analysis Workbench." *Bioinformatics (Oxford, England)* 25 (9): 1189–91. https://doi.org/10.1093/bioinformatics/btp033.

Weeramange, Chamitha J., Max S. Fairlamb, Dipika Singh, Aron W. Fenton, and Liskin Swint-Kruse. 2020. "The Strengths and Limitations of Using Biolayer Interferometry to Monitor Equilibrium Titrations of Biomolecules." *Protein Science*, January, pro.3827. https://doi.org/10.1002/pro.3827.

Wei, Ming-Tzo, Yi-Che Chang, Shunsuke F. Shimobayashi, Yongdae Shin, Amy R. Strom, and Clifford P. Brangwynne. 2020. "Nucleated Transcriptional Condensates Amplify Gene Expression." *Nature Cell Biology* 22 (10): 1187–96. https://doi.org/10.1038/s41556-020-00578-6.

Zhang, Yuning, Tiffany D. Ho, Nicolas E. Buchler, and Raluca Gordân. 2021. "Competition for DNA Binding between Paralogous Transcription Factors Determines Their Genomic Occupancy and Regulatory Functions." *Genome Research* 31 (7): 1216–29. https://doi.org/10.1101/gr.275145.120.

Zhou, Xu, and Erin K. O'Shea. 2011. "Integrated Approaches Reveal Determinants of Genome-Wide Binding and Function of the Transcription Factor Pho4." *Molecular Cell* 42 (6): 826–36. https://doi.org/10.1016/j.molcel.2011.05.025.

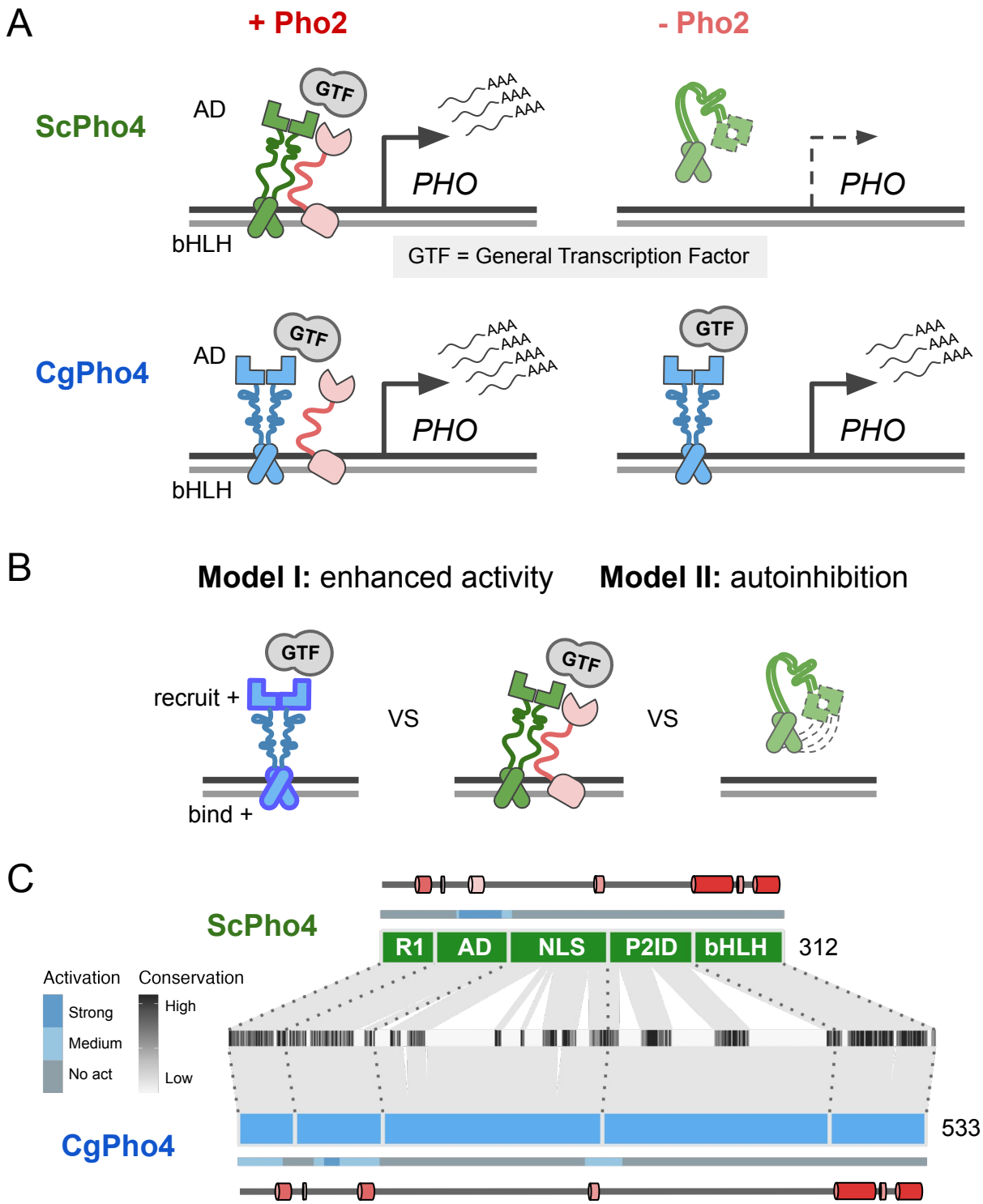


Fig. 1 Different co-TF (Pho2) dependence between orthologous TF (Pho4) in yeast species. (A) Cartoon summary of previous work: *S. cerevisiae* Pho4 (ScPho4) depends on the co-TF, Pho2, for DNA binding and activation (top row), while Pho4 from *C. glabrata* (CgPho4) can bind and activate without the co-TF. See text for references. (B) Two *non-mutually exclusive* models to explain the divergence in Pho2-dependence. (C) The tiles in the middle show the conservation scores at aligned positions of the two Pho4 orthologs; the box diagrams represent ScPho4 and CgPho4, with gray lines connecting matching residues to the alignment tiles. Five regions were delineated based on knowledge about ScPho4: R1 = regulatory, interact with Pho80; AD = activation domain; NLS = nuclear import and export signals; P2ID = Pho2-interaction domain (also includes a region contacting Pho80); bHLH = basic Helix-Loop-Helix DNA binding domain. The lines above and below the boxes are a 53 aa tiling window prediction for transactivation potentials. The cartoon on the outside shows predicted alpha helices in each protein, with lighter color corresponding to lower prediction confidence.

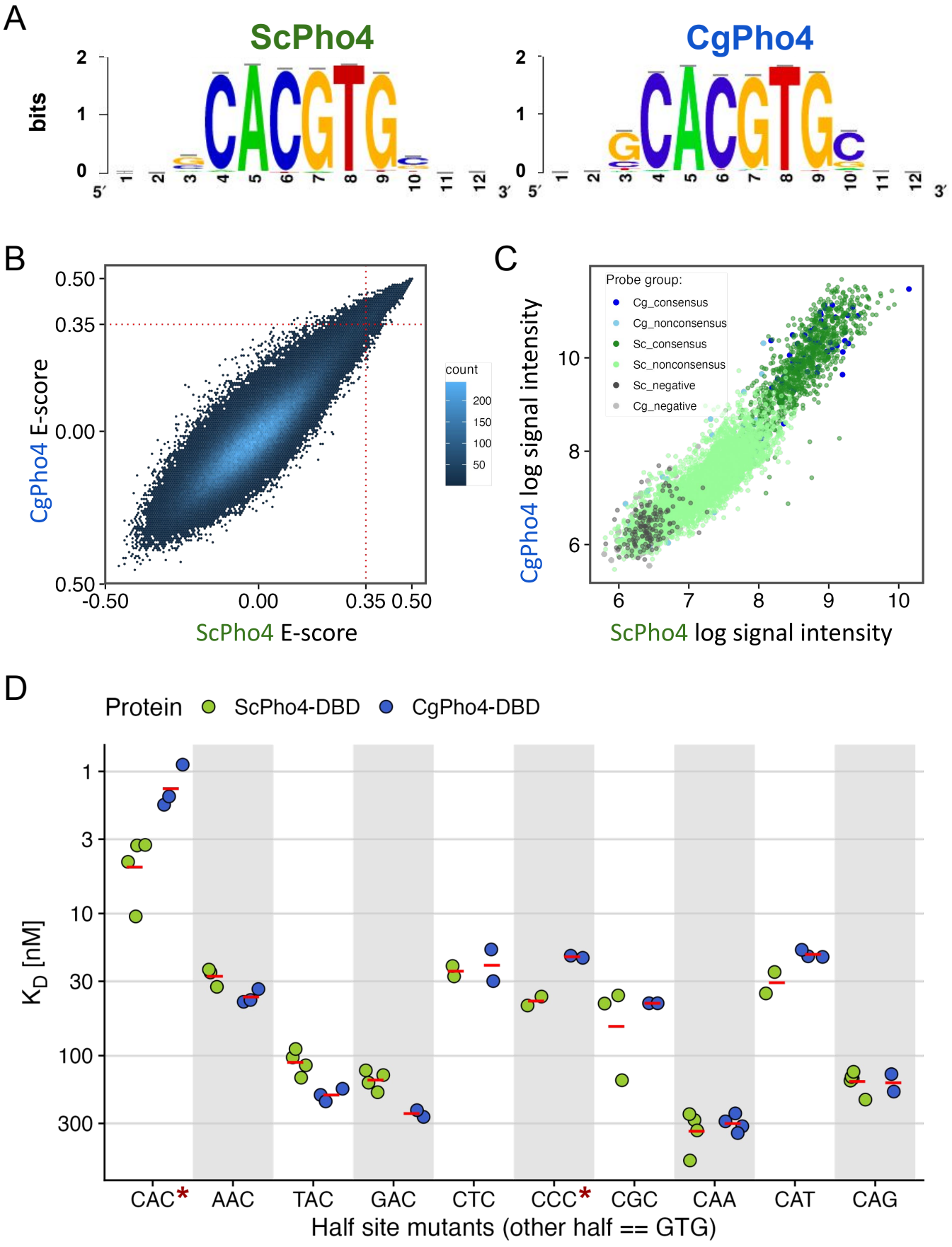
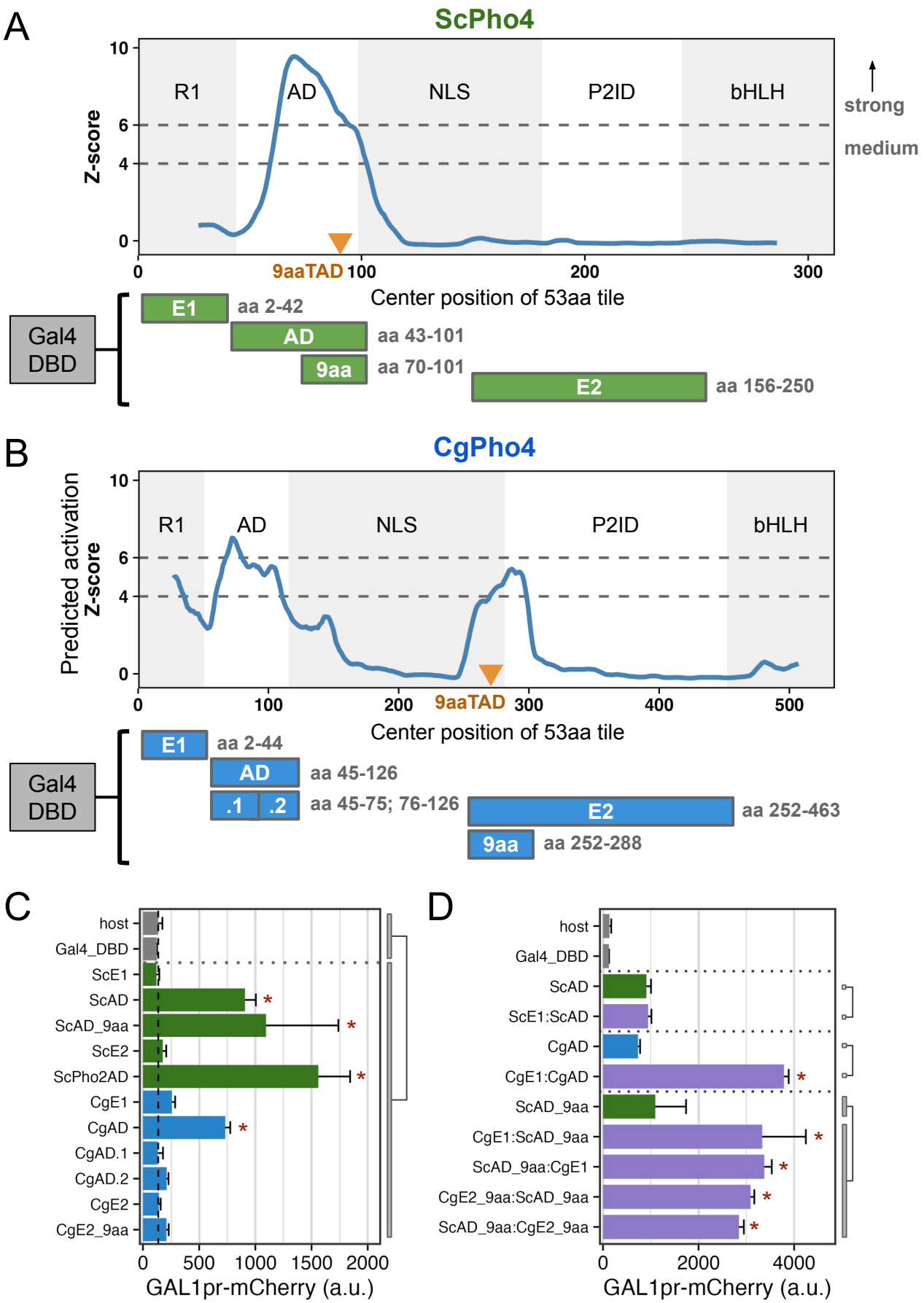

A
ScPho4
CgPho4
bits
B
CgPho4 E-score
ScPho4 E-score
count
C
Probe group:
Cg_consensus
Cg_nonconsensus
Sc_consensus
Sc_nonconsensus
Sc_negative
Cg_negative
CgPho4 log signal intensity
ScPho4 log signal intensity
D
Protein
ScPho4-DBD
CgPho4-DBD
K_D [nM]
CAC*
AAC
TAC
GAC
CTC
CCC*
CGC
CAA
CAT
CAG
Half site mutants (other half == GTG)

Fig. 2 CgPho4 DBD recognizes the same DNA motif as ScPho4 DBD, but binds ~4 fold tighter to the consensus. A) Motifs generated from ChIP-seq binding peaks for ScPho4 and CgPho4 in the *S. cerevisiae* genome, showing a conserved CACGTG core motif. B) Universal Protein Binding Microarray (uPBM) comparing the binding of the two Pho4's. A rank-order based E-score is plotted for all 7-mers, where a score greater than 0.35 indicates specific TF-DNA binding. C) Genomic context PBM (gcPBM) measuring binding to ChIP-identified genomic regions with or without a consensus motif (Cg=*C. glabrata*, Sc=*S. cerevisiae*). Data plotted as the natural log of signal intensities for both TFs. D) Dissociation constants (K_D) determined by Biolayer interferometry (BLI) for 17-bp oligos containing either the consensus or nine one-bp-off half-site variants. The red bar indicates the mean of >=2 biological replicates. Asterisks next to the sequence indicate a significant difference by a two-sided Student's t-test at a 0.05 level after Holm-Bonferroni correction.

Fig. 3 CgPho4 encodes two Activation Enhancer Domains (AEDs) that can enhance the activation potential of the main AD. (A, B) PADDLE prediction of activation potential in ScPho4 and CgPho4. Z-score > 4 and 6 indicate regions with medium and strong activation potential. The five functional regions are labeled as in Fig. 1. Orange triangles mark the 9aaTAD motif matches. Below the plots are diagrams of the Gal4 DBD fusion constructs used to test activation, with their boundaries labeled on the right. C) Activation potential of each region was tested in the context of a Gal4 DBD fusion driving a *GAL1pr*-mCherry reporter. Shown are the mean and standard deviation of >3 biological replicates. "host" has the mCherry reporter but not the Gal4 DBD fusion plasmid; "Gal4_DBD" has the reporter and the Gal4 DBD alone. ANOVA was used to compare the Gal4 DBD fusions to the two negative controls combined; an asterisk indicates significant difference at a 0.05 level after Holm-Bonferroni correction. D) CgE1 and CgE2_9aa can enhance the activation capability of ScAD_9aa in an orientation-independent manner. An asterisk indicates significance difference at a 0.05 level by a t-test between the indicated groups after Holm-Bonferroni correction.


A
ScPho4
Z-score
R1
AD
NLS
P2ID
bHLH
strong
medium
9aaTAD
Center position of 53aa tile
Gal4 DBD
E1
aa 2-42
AD
aa 43-101
9aa
aa 70-101
E2
aa 156-250
B
CgPho4
Predicted activation Z-score
R1
AD
NLS
P2ID
bHLH
9aaTAD
Center position of 53aa tile
Gal4 DBD
E1
aa 2-44
AD
aa 45-126
.1
.2
aa 45-75; 76-126
E2
aa 252-463
9aa
aa 252-288
C
host
Gal4_DBD
ScE1
ScAD
ScAD_9aa
ScE2
ScPho2AD
CgE1
CgAD
CgAD.1
CgAD.2
CgE2
CgE2_9aa
GAL1pr-mCherry (a.u.)
D
host
Gal4_DBD
ScAD
ScE1:ScAD
CgAD
CgE1:CgAD
ScAD_9aa
CgE1:ScAD_9aa
ScAD_9aa:CgE1
CgE2_9aa:ScAD_9aa
ScAD_9aa:CgE2_9aa
GAL1pr-mCherry (a.u.)

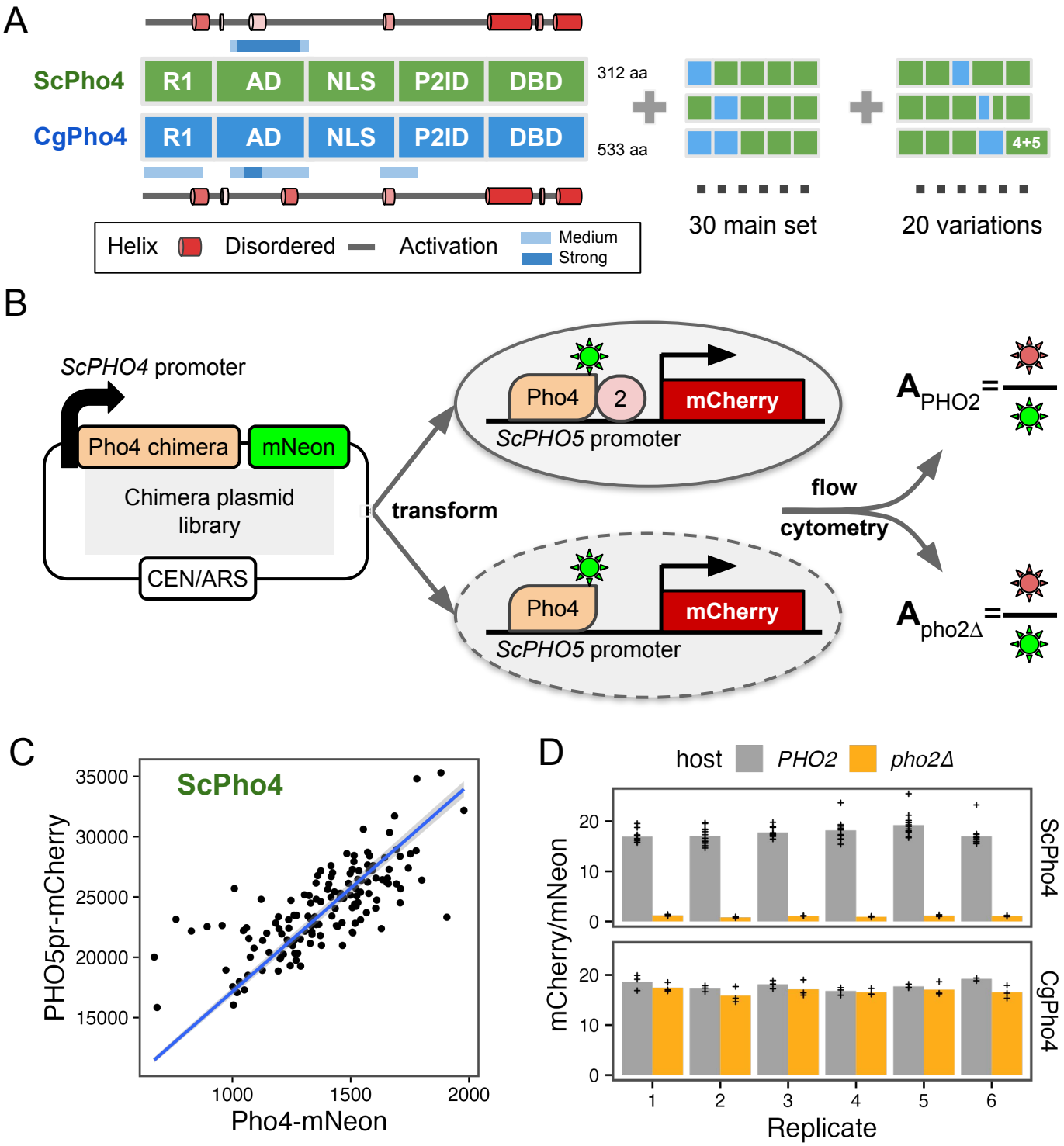


Fig. 4 A dual fluorescence reporter system for accurately quantifying the activity of Pho4 chimeras with and without Pho2. (A) ScPho4 and CgPho4 were divided into five corresponding parts. Names of each region were based on the annotations for ScPho4. Predicted alpha helices were drawn as cartoons outside of the box diagrams. All 2^5 = 32 combinations (2 were the original proteins) were made in the main chimera set, and an additional set of chimeras were constructed to test specific hypotheses. (B) A diagram showing the design of the dual fluorescence reporter system. The chimeric Pho4 fragments were cloned downstream of the endogenous *ScPHO4* promoter and were tagged at the C-terminus with an in-frame mNeon fluorescent protein, followed by the endogenous *ScPHO4* 3'UTR and terminator. The plasmid contains a CEN/ARS element, resulting in stable inheritance and low biological variation. (C) The *PHO5pr-mCherry* levels correlate with the ScPho4-mNeon levels between biological replicates. The linear regression line is shown in blue and a 95% confidence band as a gray ribbon. (D) A_{PHO2} and $A_{pho2\Delta}$ values for ScPho4 and CgPho4 across 6 different days of experiments, with multiple biological replicates per day, showing consistent behaviors.

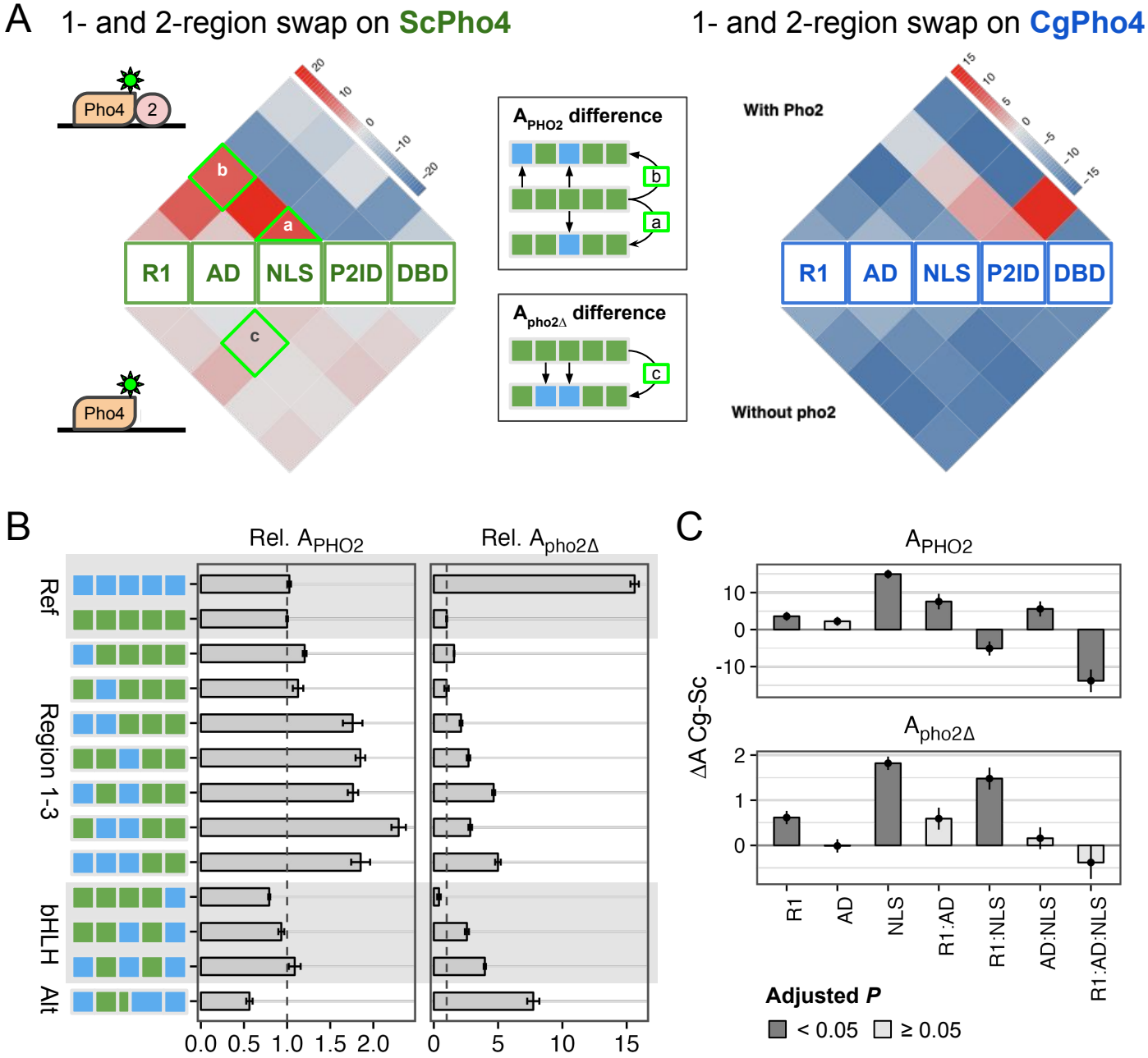


Fig. 5. Effects of 1- or 2-region swaps reveal that R1 and NLS from CgPho4 enhance the activity but the chimeras remain Pho2-dependent. (A) Heatmaps showing the effects of 1- or 2-region swaps from CgPho4 into ScPho4 (left) or vice versa (right). The top triangle shows the effects with Pho2. The triangles at the base show the 1-region swap effects, while the squares on the top show 2-region swap effects (see a and b in the key). The bottom triangle shows the corresponding effects measured in the *pho2*Δ host. (B) A subset of the chimeras were shown with their A_{PHO2} and $A_{pho2\Delta}$ values normalized to those of ScPho4 (Rel. A_{PHO2} or $A_{pho2\Delta}$). The error bars represent the standard deviation of the mean for A_{PHO2} and $A_{pho2\Delta}$. The set in rows 3-9 show the effect of swapping region 1-3 from CgPho4 into ScPho4, both with and without Pho2. Rows 10-12 show the effects of swapping CgPho4's DBD alone or in combination with other regions. The last row shows a chimera with an alternative breakpoint, where the 4th region included CgPho4 P2ID and the last 30 aa of the NLS region, which coincided with the second activation boosting domain. C) Epistasis analysis of CgPho4 regions 1-3. Linear regression with a full interaction model was fit to the data. Shown are estimates of the individual and interaction terms (bar and point) and standard error (line), with dark gray indicating a Holm-Bonferroni corrected $P < 0.05$. All effects are calculated using ScPho4 as the reference.

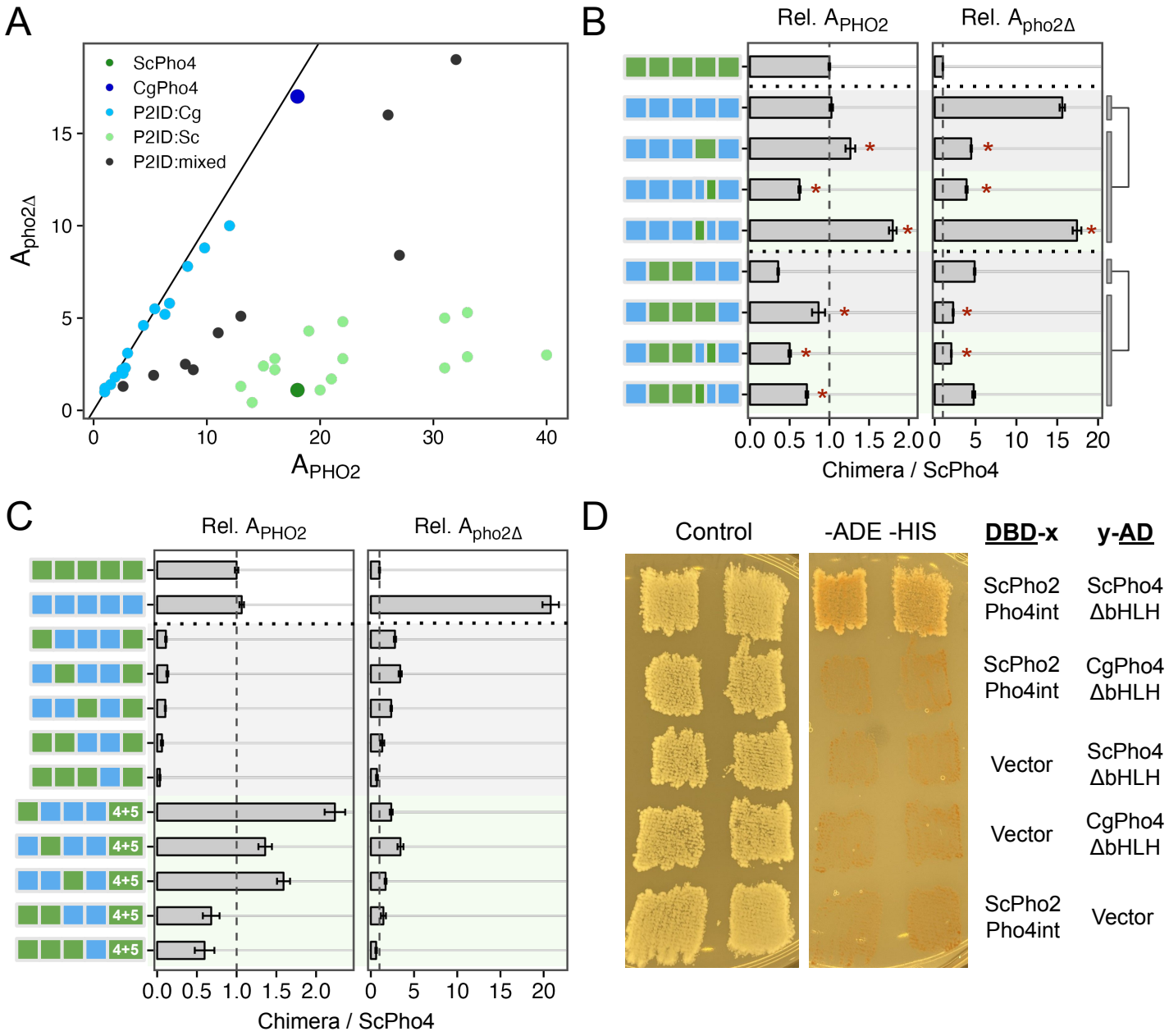


Fig. 6 P2ID in ScPho4 allows for enhanced activation with Pho2 via physical interaction with the co-TF. (A) Scatter plot of all chimeric Pho4 constructs based on their A_{PHO2} and $A_{pho2\Delta}$ values, colored by the identity of their P2ID region. Mixed P2ID uses an additional breakpoint that separates that region into two halves (Fig. S1). (B) Comparing the effects of swapping the whole P2ID from ScPho4 into CgPho4 (gray shaded) vs individual halves (light green shaded). Two sets were tested (rows 2-5 and 6-9). Two-sided t-tests were performed as indicated by the bars and lines to the right. Holm-Bonferroni correction was applied to the raw *P*-values for both sets, separately for A_{PHO2} and $A_{pho2\Delta}$. An asterisk indicates corrected $P < 0.05$. (C) Testing the hypothesis that ScPho4 DBD requires the P2ID:Sc and Pho2 to be fully functional. Compared with the chimeras with P2ID:Cg + P2ID:Sc, which are all non-functional even with Pho2, adding back P2ID:Sc ("4+5") rescued the activity with Pho2, but not without. (D) Pho4-Pho2 physical interaction was tested with a yeast two hybrid assay. The bait and target regions were fused to the activation domain (AD) and DNA binding domain (DBD) of the yeast Gal4 protein. Interaction between them reconstitutes a functional Gal4 and allows cells to grow on the dropout media (right column), while all strains are able to grow on the rich media (left column). "Vector" indicates a Gal4 DBD or Gal4 AD only plasmid without a bait or target fused to it, and is included as a control. Two independent transformants are shown.

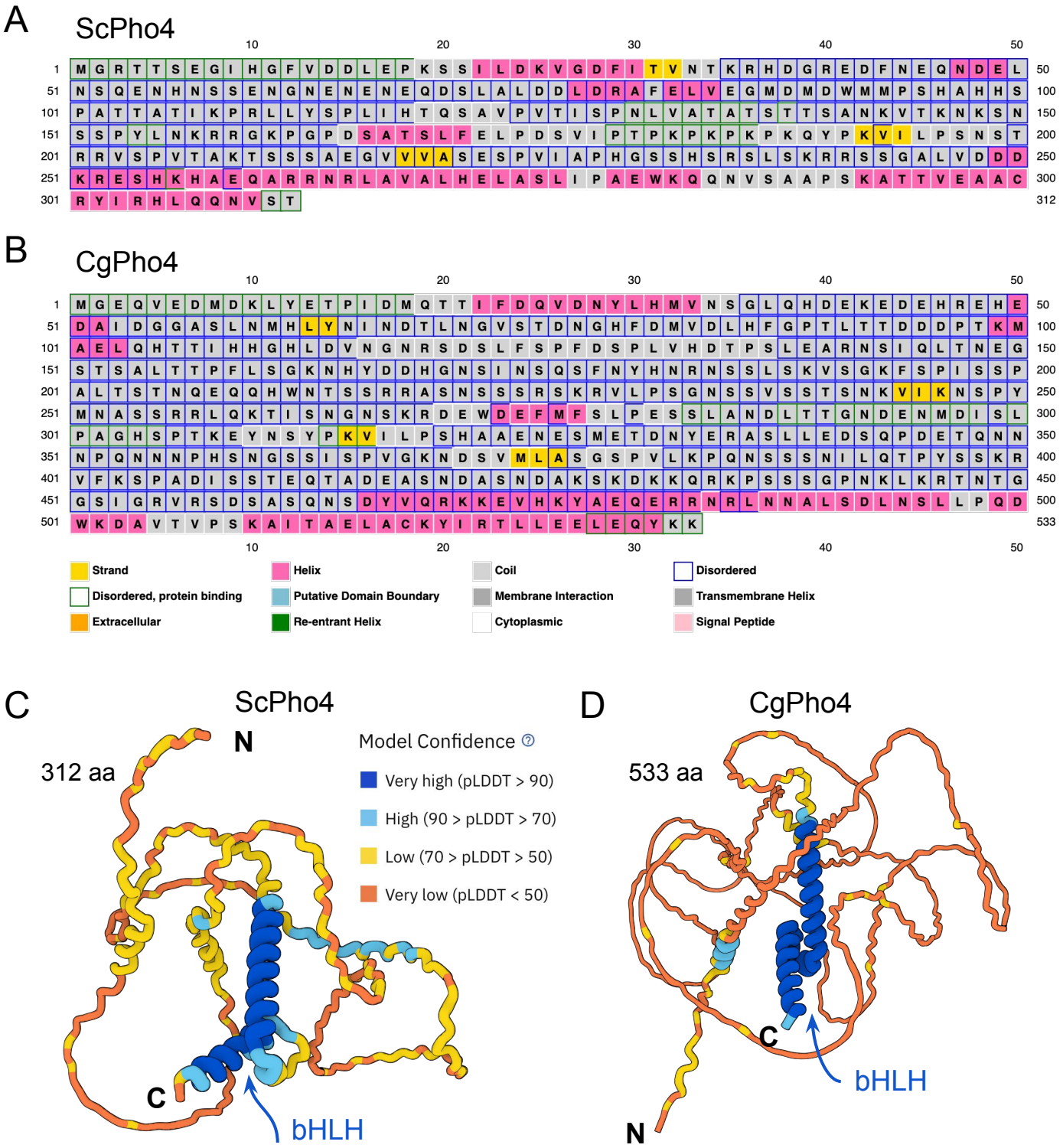


Fig. S1 ScPho4 and CgPho4 structural prediction shows both proteins are largely disordered outside the DBD. (A, B) Secondary structure predictions for ScPho4 and CgPho4 using the PSIPRED 4.0 server. (C, D) AlphaFold2 predicted monomer 3D structure for ScPho4 and CgPho4. Color represents model confidence. It has been shown that low and very low confidence regions overlap significantly and are predictive of protein disorder.

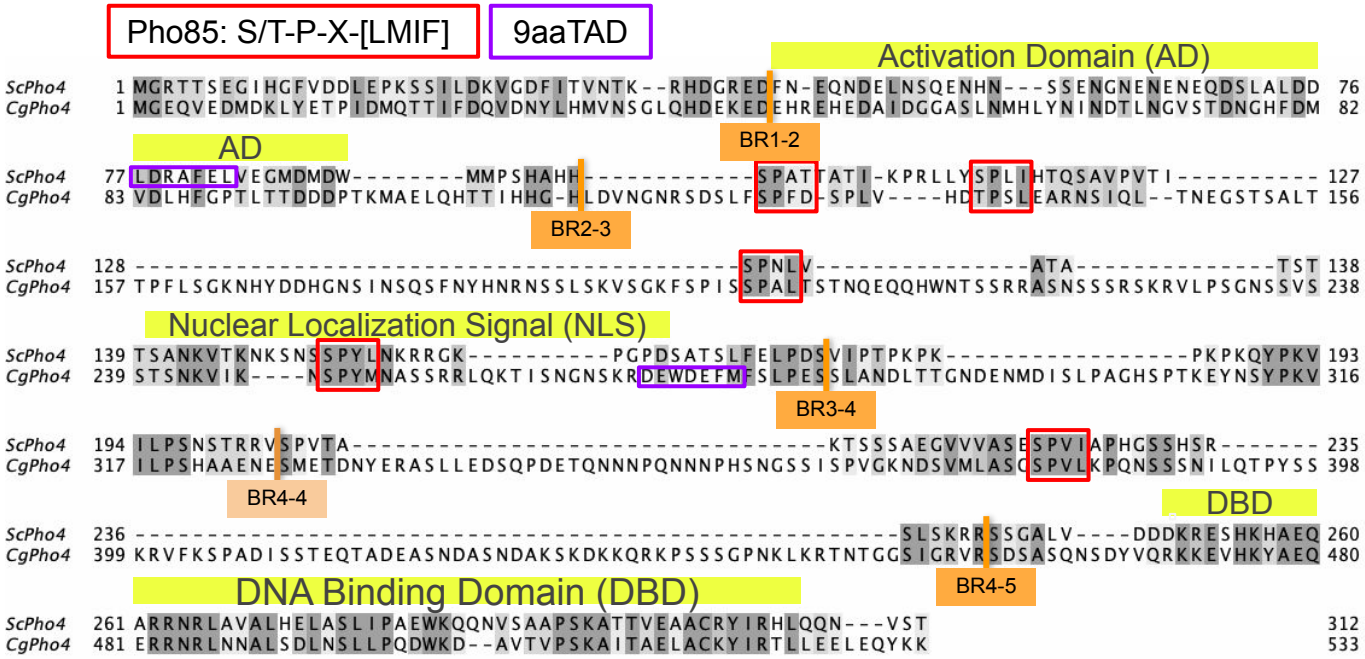


Fig. S2 ScPho4 and CgPho4 alignment and annotation. Protein sequence alignment was generated using ProbCons with 2 passes of consistency transformation and 100 iterative refinement, and was manually edited to align the known Pho85 recognition motifs (red boxes). The shading behind the amino acid symbols represent level of conservation based on BLOSUM62 scores. The functional annotation on the top were based on ScPho4 (Uniprot #P07270). The orange boxes and texts refer to the five breakpoints used for constructing chimeric Pho4 constructs (Fig. 5, 6).

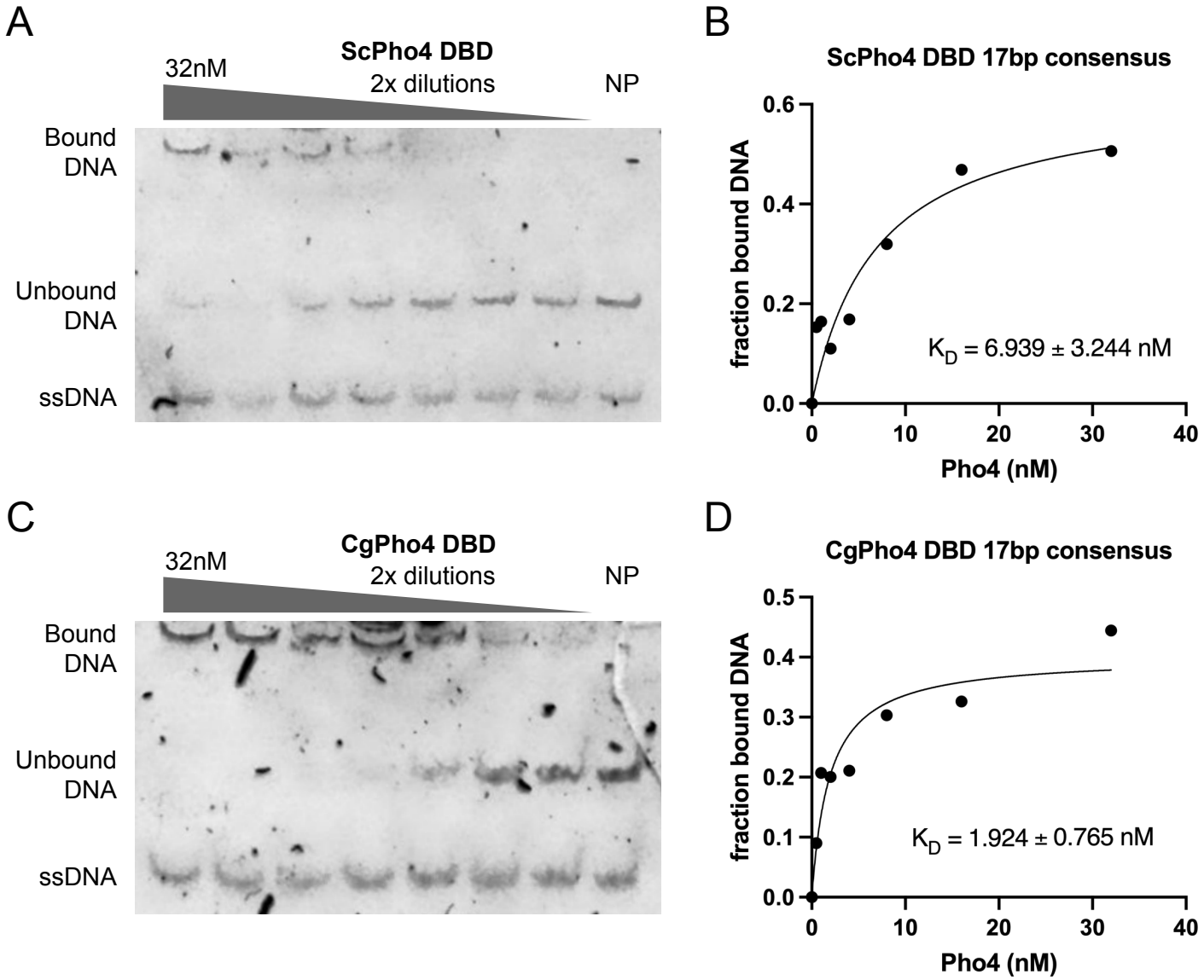


Fig. S3 Validation of difference in binding affinity to consensus motif. (A) EMSA with ScPho4 DBD and IR700 labeled 17bp consensus DNA. Lanes 1-7 contain 2x diluted protein; lane 8 is a no protein (NP) control. (B) Fraction of bound DNA was calculated from the unbound DNA band normalized to the NP lane. Fitting was done with a one site specific binding model in GraphPad Prism v10. C) EMSA with CgPho4 DBD and IR700 labeled 17bp consensus sequence. D) Same as (B) for CgPho4 DBD.

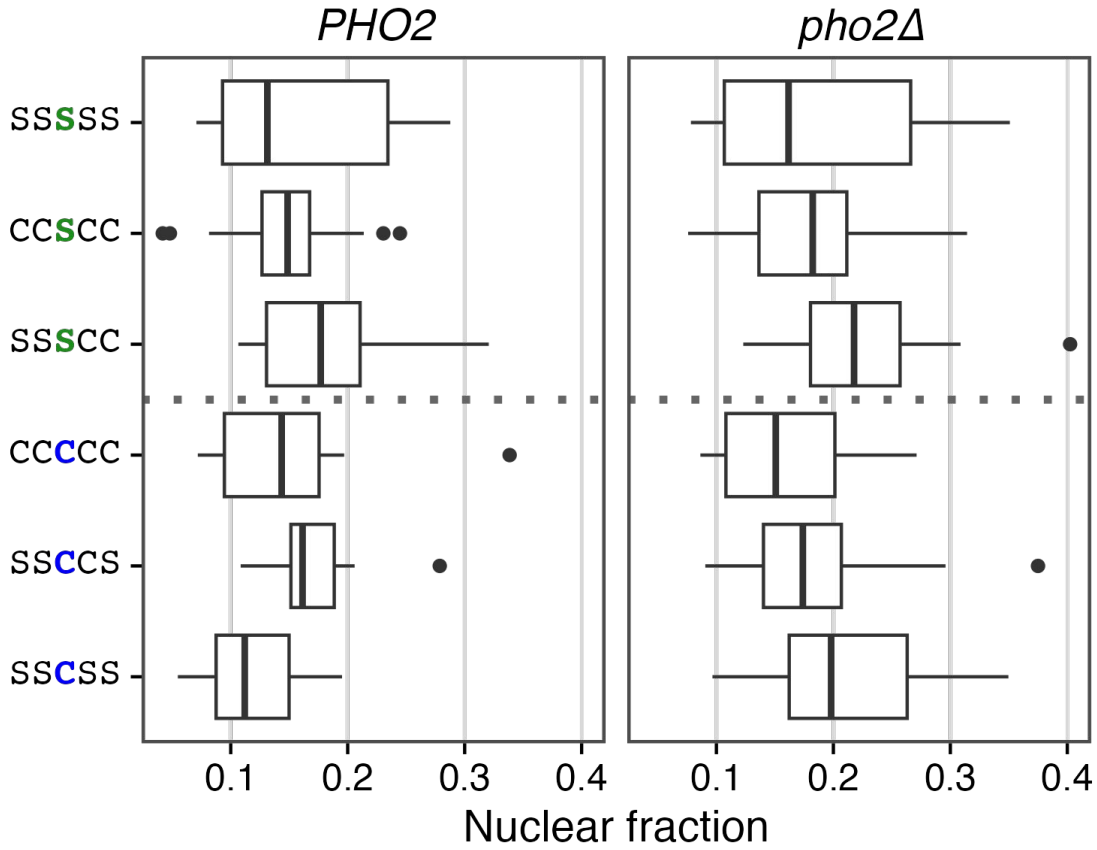


Fig. S4 NLS from ScPho4 and CgPho4 conferred similar levels of nuclear localization. Six Pho4 constructs with either NLS:Sc or NLS:Cg were quantified for their nuclear fraction in either *PHO2* or *pho2*Δ host background. Live fluorescence imaging was performed on mid-log phase cells. DAPI staining was used to identify the nucleus. Endogenous or chimeric Pho4 protein levels were quantified by the C-terminal mNeon tag. Integrated signal intensity inside the nucleus was divided by total integrated intensity across the entire cell for 35 cells in each construct. ANOVA with the identity of the NLS ("Cg" or "Sc") and Host (*PHO2* or *pho2*Δ) found the first factor to be not significant (F-test raw $P = 0.13$), i.e., no evidence for significant differences in nuclear fraction between constructs with NLS:Sc vs NLS:Cg. There was a significant difference for the second factor (F-test raw $P = 0.0003$), showing higher nuclear fraction in the *pho2*Δ background than in the *PHO2* background.