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

[^] Current addresses: Department of Genetics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

[*] correspondence should be addressed to bin-he@uiowa.edu

Running title: Evolution of a eukaryotic TF's dependence on co-TF

ABSTRACT

Combinatorial control by multiple transcription factors (TFs) is a hallmark of eukaryotic gene regulation. Despite its prevalence and crucial roles in enhancing specificity and integrating information, the mechanisms behind why eukaryotic TFs depend on one another, and whether such interdependence evolves, are not well understood. We exploit natural variation in co-TF dependence in the yeast phosphate starvation (PHO) response to address this question. In the model yeast *Saccharomyces cerevisiae*, the main TF, Pho4, relies on the co-TF Pho2 to regulate ~28 genes. In a related yeast pathogen, *Candida glabrata*, its Pho4 exhibits significantly reduced Pho2 dependence and has an expanded target set of ~70 genes. Biochemical analyses showed *C. glabrata* Pho4 (CgPho4) binds to the same consensus motif with 3-4-fold higher affinity than ScPho4 does. A machine-learning-based prediction and yeast one-hybrid assay identified two Intrinsically Disordered Regions (IDRs) in CgPho4 that boost the activity of the main activation domain but showed little to no activity on their own. We also found evidence for autoinhibition behind the co-TF dependence in ScPho4. An IDR in ScPho4 next to its DNA binding domain was found to act as a double-edged sword: it both allows for enhanced activity with Pho2, and inhibits Pho4's activity without Pho2. This study provides a detailed molecular picture of how co-TF dependence is mediated and how its evolution, mainly driven by IDR divergence, can lead to significant rewiring of the regulatory network.

41 **INTRODUCTION**

42 Transcription factors (TFs) are the cornerstone of gene regulatory networks. In eukaryotes, TFs
43 often work collaboratively to regulate gene expression. This combinatorial control is crucial for
44 enhancing specificity, because most eukaryotic TFs recognize short and degenerate motifs –
45 typically less than 10 bps – that appear hundreds to tens of thousands of times in the genome
46 (Stewart et al., 2012). TFs with the same family of DNA binding domains (DBDs) also recognize
47 highly similar motifs (Berger et al., 2008). Despite these, TFs often bind only a fraction of their
48 motifs *in vivo* and regulate an even smaller subset of the genes they bind to (Zhou and O'Shea,
49 2011); paralogous TFs regulate distinct sets of genes (Merhej et al., 2016; Sánchez-Higueras et
50 al., 2019). The key to achieving this specific regulation is the requirement of two or more TFs to
51 jointly regulate the target genes (Avsec et al., 2021; Slattery et al., 2011; Todeschini et al.,
52 2014). Another important function of combinatorial control and TF interdependence is to allow
53 cells to integrate multiple upstream signals. In fly development, for example, a combination of
54 tissue-specific "selector" and morphogen gradient – both are TFs – precisely determine the
55 expression pattern and define the cell fate (Guss et al., 2001).

56 The importance of combinatorial control and the molecular mechanisms behind TF
57 interdependence are traditionally studied using mutations that disrupt TF interactions. For
58 instance, a missense mutation that disrupts the interaction between two cardiac TFs, GATA4
59 and TBX5, was shown to cause dysregulation of cardiac genes and lead to malformation of
60 heart tissues (Ang et al., 2016). Similarly, studies of combinatorial control evolution often focus
61 on losses (and gains) of TF interactions, like the MADS-box TF, Mcm1, and its co-TFs, MATa2
62 and MATα2 in the yeast mating type pathway (Baker et al., 2012; Britton et al., 2020). A far less
63 explored type of change in the combinatorial control is a change in the TF interdependence
64 itself, that is, when TFs involved in cooperative regulation evolved to be more or less dependent
65 on the co-TF(s). Such changes are expected to dramatically rewire the network by altering the
66 specificity and signals required for gene regulation, which could have profound implications in

67 disease and evolution. In relation to this possibility, not all eukaryotic TFs require other TFs to
68 function. The yeast TF, Gal4, regulates its target genes on its own (Traven et al., 2006). This
69 raises several intriguing and unanswered questions: if TF interdependence and combinatorial
70 control evolve, what genetic changes underlie the divergence, what aspects of the TF activities
71 are impacted, and what are the consequences on the regulatory output?

72 Intriguingly, natural variation in co-TF dependence exists in the yeast phosphate
73 starvation (PHO) response network. In the model yeast *Saccharomyces cerevisiae*, the main TF
74 of the PHO response, Pho4 (hereinafter as ScPho4), strongly depends on the co-TF, Pho2, to
75 induce 27/28 of its target genes (Zhou and O'Shea, 2011). In a related human yeast pathogen,
76 *Candida glabrata*, its Pho4 (hereinafter as CgPho4) is far less dependent on Pho2 and induces
77 twice as many genes (Fig. 1A) (He et al., 2017; Kerwin and Wykoff, 2009). The level of Pho2-
78 dependence varies quantitatively among Pho4 orthologs in other yeasts and is correlated with
79 the number of genes induced in a common genome background (He et al., 2017). This latter
80 observation is consistent with the role of combinatorial control in enhancing specificity. What
81 remains unknown is what genetic differences between Pho4 orthologs underlie the divergence
82 in co-TF dependence, and what TF activities are impacted by those variations.

83 Here, we propose two non-mutually exclusive models for the difference in Pho2-
84 dependence between ScPho4 and CgPho4 (Fig. 1B). The first "enhanced activity" model is
85 based on studies showing that ScPho4 requires Pho2 to **(1)** bind cooperatively to the target
86 gene promoters (Barbaric et al., 1998), **(2)** recruit the histone acetyltransferase (HAT) complex
87 (Nourani et al., 2004) and **(3)** help recruit general TFs and the PolII complex (Magbanua et al.,
88 1997). Under this model, we hypothesize that CgPho4 binds more tightly to DNA than ScPho4,
89 and is more capable of recruiting general transcription factors (TFs) and the PolII complex,
90 thereby becoming less Pho2-dependent. The second "autoinhibition" model is based on a study
91 suggesting that ScPho4 is auto-inhibited, and that interaction with Pho2 unmasks its activation

92 domain and allows it to function (Shao et al., 1996). This model predicts that CgPho4 either
93 lacks or has far weaker effects of the auto-inhibition and hence doesn't depend on Pho2.
94 In this study, we tested these two models by comparing the DNA binding and activation
95 abilities of the two Pho4 orthologs, and systematically swapping regions between the two Pho4s
96 in a series of 50 chimeric TFs, then quantifying their activities with and without Pho2. Our results
97 support both models as contributing to the Pho2-dependence variation. To our surprise, while
98 CgPho4 DBD binds to the same consensus motif with a 3-4-fold higher affinity than ScPho4
99 DBD, swapping DBD alone failed to yield the expected results. Instead, the differences in Pho2-
100 dependence originated primarily from differences in the Intrinsically Disordered Regions (IDRs)
101 in the two TFs that modulate both their Activation Domain (AD) and DNA binding domain (DBD)
102 activities. Therefore, our results reveal that evolution in a eukaryotic TF protein, particularly
103 through changes in the IDR, can lead to divergence in co-TF dependence, which in turn results
104 in a more than two-fold change in the size of the target network.

105 **RESULTS**

106 **Domain organization and sequence divergence between ScPho4 and CgPho4**

107 Based on genetic and biochemical studies, ScPho4 encodes the following functional domains
108 from its N- to its C-terminus (Fig. 1C): a regulatory region (R1) interacting with the negative
109 regulator Pho80 (Jayaraman et al., 1994), the activation domain (AD) (McAndrew et al., 1998;
110 Ogawa and Oshima, 1990), a region encoding the nuclear export and import signals (NLS)
111 (Komeili and O'Shea, 1999; O'Neill et al., 1996), a protein-interaction domain with the
112 interacting with both Pho80 and the co-TF, Pho2 (referred together as P2ID) (Hirst et al., 1994;
113 Jayaraman et al., 1994), and finally the bHLH DNA binding domain (DBD) (Ogawa and Oshima,
114 1990; Shimizu et al., 1997). Both ScPho4 and CgPho4 were predicted to be mostly intrinsically
115 disordered outside the DBD, with some low confidence helices in R1 and AD (Fig. S1). CgPho4
116 is significantly longer than ScPho4 (533 vs 312 amino acids, Fig. 1C). The NLS and P2ID
117 regions contain four functionally important phosphorylation sites targeted by the cyclin-

118 dependent kinase complex Pho80/85 (Kaffman et al., 1994; O'Neill et al., 1996). Those located
119 in the NLS control Pho4's nuclear translocation, while phosphorylation of the one in P2ID
120 disables ScPho4 from interacting with Pho2 (Komeili and O'Shea, 1999). All five Pho80/85
121 targeted phosphorylation sites are clearly identifiable in the alignment (Fig. S2); in fact, they
122 were previously found to be conserved in orthologs outside the Saccharomycotina subphylum
123 (Gomes-Vieira et al., 2018). Thus, while the protein length and sequence diverged significantly
124 between the two Pho4 orthologs, the domain architecture and post-translational modification
125 motifs appear to be highly conserved in evolution.

126 **CgPho4 binds the same consensus motif with a higher affinity compared to ScPho4**

127 To test the enhanced activity model, we first compared the DNA binding ability of the two Pho4
128 orthologs. First, we needed to determine if the two Pho4 orthologs recognize the same DNA
129 sequence. We used our published Chromatin-Immunoprecipitation (ChIP-seq) data for both
130 Pho4 orthologs expressed in the *S. cerevisiae* background to identify their respective motifs (He
131 et al., 2017) (Materials and Methods). The result showed that they bind the same E-box motif
132 "CACGTG", with no obvious difference in the flanking nucleotides (Fig. 2A). The relatively small
133 number of ChIP peaks (74 for ScPho4 and 118 for CgPho4) leave the possibility of small but
134 significant differences in the flanking nucleotide preference between the two Pho4s. To
135 comprehensively characterize and compare the binding preferences of ScPho4 vs CgPho4, we
136 turned to Protein Binding Microarrays (PBM, Materials and Methods). Two library designs were
137 used: the first one comprises all possible 9-mer sequences to unbiasedly scan the sequence
138 space; the second library contains a subset of 36-bp sequences centered on ChIP-identified
139 binding sites for both Pho4's in their native genome. The latter is designed to interrogate
140 extended flanking positions ("genomic context PBM" or gcPBM, (Gordân et al., 2013)). Results
141 from both PBM experiments revealed no evidence for binding specificity differences between
142 ScPho4 and CgPho4 (Fig. 2B, C). We conclude that ScPho4 and CgPho4 have the same
143 sequence specificity.

144 Next, we tested the hypothesis that CgPho4 binds more tightly than ScPho4 does. To do
145 so, we purified the DBD region of both Pho4s, and measured their binding affinity to a set of 17-
146 bp oligos that contain either the consensus motif “CACGTG”, or the nine 1-bp-off half site
147 variants of the palindromic motif (Maerkl and Quake, 2007). Using Biolayer Interferometry (BLI),
148 we found that CgPho4 DBD binds the consensus motif >3 times more tightly than ScPho4 DBD
149 does (K_D = 5.4 nM vs 1.4 nM for ScPho4 and CgPho4, respectively; raw Student’s t-test for log
150 K_D difference P = 0.013, Holm-Bonferroni corrected P = 0.1). Among the suboptimal motifs
151 tested, only “CCCGTG” showed a 2-fold or more difference in K_D ($K_{D,\ ScPho4}$ / $K_{D,\ CgPho4}$ = 2, raw P
152 = 0.05, Holm-Bonferroni corrected P = 0.31, Fig. 2D). We confirmed the K_D difference for the
153 consensus “CACGTG” sequence using an Electrophoretic Mobility Shift Assay (Fig. S3). We
154 conclude that CgPho4 recognizes the same DNA sequence as ScPho4 but binds to the
155 consensus E-box motif and at least one suboptimal motif with a 2-fold or lower K_D than ScPho4.

156 **CgPho4 encodes two additional Activation Enhancer Domains (AEDs)**

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

169 To test the function of all predicted regions with activation potential in both Pho4s, we
170 set up a yeast one-hybrid assay, in which the candidate region was fused to the Gal4 DBD and
171 the activation potential was measured by a genome-integrated *GAL1*pr-mCherry reporter in *S.*
172 *cerevisiae* (Fig. 3C, Materials and Methods). We confirmed that ScPho4's AD was able to
173 activate the reporter above the background level (7-fold, Holm-Bonferroni corrected $P < 0.01$).
174 CgPho4's AD offered a similar level of activation (5.6-fold, corrected $P < 0.01$). We also tested
175 ScPho2's AD and found it to have the highest activity, consistent with the hypothesis that its
176 recruitment by ScPho4 enhances gene activation (12-fold, corrected $P < 0.01$). We were able to
177 further localize the required residues for activation in ScPho4's AD to a region of 32 aa centered
178 on the 9aaTAD motif (Fig. 3A, C). Neither E1 nor E2 from either Pho4 activated the reporter on
179 its own (Fig. 3C). This was surprising, given that CgPho4's E2 contains a match to the 9aaTAD
180 as ScPho4's AD does (Fig. 3B). We hypothesized that E1 and E2 in CgPho4 could enhance the
181 activity of the AD. We first tested ScE1 and CgE1 fused to their own AD, and found that only the
182 latter resulted in a significant enhancement in the activation potential (Fig. 3D). Next, we fused
183 CgE1 or CgE2 in both orientations to the minimal AD of ScPho4 (ScAD_9aa). We found that
184 indeed, both regions were able to significantly enhance the activity of the ScAD_9aa (Fig. 3D).
185 In summary, we found that CgPho4 encodes two Activation Enhancer Domains (AEDs)
186 in its Intrinsically Disordered Region (IDR). The AEDs have little activation capability on their
187 own but can significantly enhance the activation potential of the AD of either of the Pho4
188 orthologs. Since these two AEDs are present in CgPho4 but not in ScPho4, we hypothesize that
189 they contribute to CgPho4 being less dependence on Pho2.

190 **A dual fluorescence reporter assay accurately measures the activity of Pho4 chimeras**
191 **with and without Pho2 for dissecting the divergence in co-TF dependence**

192 So far, we compared DNA binding and activation potential between CgPho4 and ScPho4 by
193 isolating individual regions and testing them either *in vitro* or in a synthetic *in vivo* system. While
194 they support the enhanced activity model, it remains unclear whether and how they contribute to

195 divergence in Pho2-dependence in the native Pho4 protein context. They also do not test or rule
196 out the autoinhibition model, which necessitates intramolecular interactions. To answer these
197 questions, we divided ScPho4 and CgPho4 into five corresponding parts (Fig. 4A), with
198 breakpoints chosen in well-aligned regions and considering predicted secondary structures and
199 experimentally identified domains (Fig. S2). We created all 2^5=32 combinations as well as 28
200 additional ones with alternative breakpoints (Fig. 4A). All constructs were C-terminally tagged
201 with mNeon to quantify the protein levels, and were expressed from the native *ScPHO4*
202 promoter and UTRs (Fig. 4B). We also created two *S. cerevisiae* host strains, in which the
203 *ScPHO5* CDS was replaced by an mCherry reporter, and *pho80* was knocked out so that all
204 Pho4 chimeras were constitutively nuclear localized (Kaffman et al., 1994; O'Neill et al., 1996).
205 One of the two host strains had *pho2* knocked out. For each chimera, we measured its *PHO5pr*-
206 mCherry and Pho4-mNeon levels using flow cytometry in both hosts. While mCherry and
207 mNeon levels for the same Pho4 construct varied between experiments, the ratios between the
208 two were consistent and characteristic of the specific construct (Fig. 4C, D). We therefore
209 defined the activity of a Pho4 chimera as the ratio between the median fluorescence intensity
210 (MFI) of mCherry and mNeon, which we will refer to as A_{PHO2} or $A_{pho2\Delta}$ from hereon.

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

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

221 Based on our previous results, we expected R1, NLS, and DBD in CgPho4 to enhance
222 the activity of the chimera with and without Pho2: the first two encode CgE1 and a majority of
223 CgE2_9aa, which we showed enhance the activity of the main AD (Fig. 4); DBD:Cg is expected
224 to increase the binding ability of the chimera. We indeed found R1:Cg and NLS:Cg to enhance
225 A_{PHO2} on ScPho4's background (Fig. 5A top left, 5B, 5C); contrary to our expectation, however,
226 DBD:Cg reduced the chimera's activity with Pho2. Also unexpectedly, R1:Cg and NLS:Cg
227 swaps had a much smaller effect on $A_{pho2\Delta}$ (Fig. 5A, lower left), suggesting that despite their
228 ability to increase A_{PHO2}, the chimeras were still dependent on Pho2. In fact, none of the 1- or 2-
229 region swaps led to large increases in $A_{pho2\Delta}$. This argues against a classic lock-and-key model
230 for autoinhibition, which we expect to be broken by swapping one of the regions involved.
231 Since the NLS:Cg region encodes both the NLS/NES and a key part of the second AED
232 (CgE2_9aa), we asked which of these two mechanisms was responsible for its positive effect on
233 A_{PHO2}. To test the possibility that the enhanced A_{PHO2} was due to a stronger nuclear localization
234 activity of NLS:Cg, we performed fluorescence microscopy to quantify the fraction of nuclear-
235 localized Pho4 proteins in six constructs that bear either NLS:Sc or NLS:Cg. No significant
236 difference between the two groups was found (Fig. S4, F-test $P = 0.13$). We thus conclude that
237 the positive effect of NLS from CgPho4 on A_{PHO2} was mainly due to its ability to boost activation.
238 To further examine the effects of swapping the R1, AD, and NLS regions from CgPho4,
239 including whether they contribute non-additively, we individually examined their A_{PHO2} and $A_{pho2\Delta}$
240 in comparison with ScPho4 and CgPho4 (Fig. 5B). To estimate their main and epistatic effects,
241 we fit a linear model to the data, i.e., $Y = X_0 + R1 + AD + NLS + R1{:}AD + R1{:}NLS + AD{:}NLS +$
242 $R1{:}AD{:}NLS$, where the first term represents the activity of ScPho4, the next three terms the
243 main effect of each CgPho4 region (":Cg" omitted for brevity), and the rest are epistasis terms.
244 We found that R1:Cg and NLS:Cg had significant, positive effects on both A_{PHO2} and $A_{pho2\Delta}$,
245 although the magnitude was much smaller for $A_{pho2\Delta}$ (Fig. 5C, R1 and NLS Holm-Bonferroni
246 corrected $P < 0.05$). Moreover, both show significant, positive epistasis with AD:Cg for A_{PHO2}

247 (corrected $P < 0.05$ for R1:AD and AD:NLS), and trend in the same direction for $A_{pho2\Delta}$. This
248 suggests that the two AEDs in CgPho4 work more efficiently with AD:Cg than with AD:Sc.
249 Curiously, R1:Cg and NLS:Cg show negative epistasis for A_{PHO2} but positive epistasis for $A_{pho2\Delta}$.
250 Lastly, the three-way interaction terms are negative for both A_{PHO2} and $A_{pho2\Delta}$.
251 We also examined the effect of swapping CgPho4's DBD. On its own, swapping DBD:Cg
252 decreased both A_{PHO2} and $A_{pho2\Delta}$ compared to ScPho4 (Fig. 5B), suggesting that DBD:Cg may
253 be incompatible with one or multiple regions in ScPho4.
254 In summary, we found that R1:Cg and NLS:Cg increased the activity of the chimera with
255 Pho2 but were insufficient to remove the Pho2 dependence alone or in combination. In the
256 presence of Pho2, R1:Cg and NLS:Cg both showed positive epistasis with AD:Cg on ScPho4's
257 background. Lastly, no 1- or 2-region swaps from CgPho4 into ScPho4 increased $A_{pho2\Delta}$ to the
258 level of CgPho4. Together, we conclude that CgPho4's two AEDs can increase the activity of
259 the chimera in the presence of Pho2, although they fail to restore the activity without Pho2 to
260 CgPho4's level, suggesting that additional mechanisms are at play. CgPho4's DBD, alone or in
261 combination with R1 and the NLS region, did not contribute to increased activity.

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

264 In the 1- and 2- region swap heatmaps (Fig. 5A), the P2ID showed a puzzling pattern: swapping
265 CgPho4's P2ID (P2ID:Cg) into the ScPho4 background had a dominant negative effect on A_{PHO2}
266 while offering little increase in $A_{pho2\Delta}$. This suggests that P2ID:Sc is essential for ScPho4's
267 function. Conversely, swapping ScPho4's P2ID (P2ID:Sc) into CgPho4 increased A_{PHO2} beyond
268 that of CgPho4; the effect is even stronger when combined with DBD:Sc (Fig. 5B). These
269 suggest that P2ID:Sc allows Pho4 from either species to take advantage of Pho2's assistance.
270 To investigate the unique property of the P2ID further, we plotted all chimeras based on their
271 A_{PHO2} and $A_{pho2\Delta}$ values (Fig. 6A). Strikingly, the chimeras fell into three distinct groups based on
272 the identity of their P2IDs. The first group had P2ID from CgPho4, and their activities were not

273 dependent on Pho2, falling along the diagonal line. The second group had P2ID from ScPho4.
274 These had strong activity with Pho2 and low to no activity without Pho2, like ScPho4 does. The
275 third group included additional chimeras with mixed P2ID regions, and they filled the
276 intermediate space between the first two groups. Notably, many chimeras in the second group
277 showed higher A_{PHO2} than ScPho4. Most of them contained the R1 and NLS regions from
278 CgPho4, consistent with our results above showing that these two regions enhanced the activity
279 of the Pho4 chimeras in the presence of Pho2 (Fig. 5A, B).
280 The evidence above suggests that P2ID:Sc has two effects: allowing for collaborative
281 regulation with Pho2 and restricting Pho4's activity without Pho2. Can these two functions be
282 separated? To answer this question, we compared the mixed P2ID chimeras to those with
283 whole P2IDs from either Pho4. We found that swapping P2ID:Sc into CgPho4 resulted in a 1.2-
284 fold increase in A_{PHO2} and a 71% reduction in $A_{pho2\Delta}$ (Fig. 6B, row 2 vs 3). Interestingly,
285 swapping just the second half of P2ID:Sc (aa 205-242) into CgPho4 resulted in a ~40%
286 reduction in A_{PHO2} and 75% reduction in $A_{pho2\Delta}$. By contrast, just swapping the first half of
287 P2ID:Sc (aa 177-204) resulted in 1.76-fold increase in A_{PHO2} and maintained the same level of
288 $A_{pho2\Delta}$ (1.1-fold). From this, we deduced that the first half of P2ID:Sc mainly functions to interact
289 with Pho2 while the second half appears to limit Pho4's activity without Pho2. The same trend
290 was observed in another set of chimeras with lower activities (Fig. 6B, rows 6-9).
291 Several chimeras had very low activities even with Pho2 (A_{PHo2} < 3.6, or 20% of A_{PHO2} for
292 ScPho4). Most of these nonfunctional chimeras have P2ID:Cg and DBD:Sc (Fig. 6C, rows 3-7).
293 We hypothesize that DBD:Sc requires P2ID:Sc and Pho2 to function in the context of the full
294 length Pho4. This is contrary to the conventional view that DBDs can function on its own, which
295 is supported by our own *in vitro* results (Fig. 2). Nonetheless, we reasoned that if the above
296 hypothesis is correct, putting P2ID:Sc back by inserting it in between P2ID:Cg and DBD:Sc
297 should rescue the non-functional chimeras in the presence of Pho2. That is what we observed
298 (Fig. 6C, rows 8-12), with some of the chimeras even exceeding the A_{PHO2} of ScPho4 and

299 CgPho4. However, all of them still required Pho2. These results support the above hypothesis,
300 showing that the dual-functional P2ID is essential for ScPho4 to function.

301 Given that chimeras with P2ID:Cg have equal activities with and without Pho2 (Fig. 6A), we
302 asked if CgPho4 still physically interacts with Pho2 in *S. cerevisiae*. Using the yeast two-hybrid
303 assay, we were unable to detect an interaction between CgPho4ΔDBD (aa 2-463) and a region
304 of ScPho2 known to interact with multiple TFs, including ScPho4 (Bhoite et al., 2002) (Fig. 6D).
305 By contrast, ScPho4ΔDBD (aa 2-250) was able to interact with the same region of ScPho2 as
306 previously found (Fig. 6D). This is consistent with our observation above, where chimeras with
307 P2ID:Cg and lacking the region from ScPho4 mediating Pho2-interaction cannot be enhanced
308 by Pho2 (Fig. 6A).

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

318 **DISCUSSION**

319 Combinatorial control plays crucial roles in eukaryotic gene regulation. Mutations disrupting TF
320 interactions can cause dysregulation and disease (Ang et al., 2016). However, how mutations
321 can alter TF interdependence itself, whether in disease or evolution, is less understood. In this
322 study, we investigated the molecular basis for natural variation in co-TF dependence in the
323 yeast phosphate starvation (PHO) response. We found three key differences between two

324 orthologous Pho4s with varying dependence on the co-TF Pho2: **(1)** DNA Binding Affinity:
325 CgPho4 binds the same consensus DNA motif with 3-4-fold higher affinity than ScPho4; **(2)**
326 Activation potential**:** CgPho4 has two unique activation enhancing domains (AEDs) that
327 increase the activation potential of both Pho4s. **(3)** Autoinhibition: ScPho4 contains an IDR next
328 to its DBD that both allows it to interact with Pho2 to gain enhanced activity, and inhibits its
329 activity in the absence of Pho2. Therefore, our results support both the enhanced activity model
330 and the autoinhibition model for the difference in Pho2-dependence between Pho4 orthologs.
331 Among the three differences, the contribution of CgPho4's two AEDs to the TF's activity
332 and dependence on Pho2 is well supported by the yeast one-hybrid data and chimeric Pho4
333 results (Fig. 3, 5). By contrast, the significance of the binding affinity differences between the
334 two Pho4s remains unclear (Fig. 2D). Binding kinetics predicts that if the nuclear concentrations
335 of both Pho4 are much higher than their K_D, a 3-4-fold difference in affinity would have little
336 impact. Conversely, if the nuclear concentration of Pho4 is close to its K_D, the same difference
337 could significantly affect gene induction. Existence of the second scenario is supported by a
338 study showing that nuclear ScPho4 levels are much lower at intermediate phosphate
339 concentrations than in extreme starvation conditions, leading to differential binding and
340 induction of ScPho4's targets (Ming Yip et al., 2023). In our chimeric Pho4 experiments,
341 replacing ScPho4's DBD with that from CgPho4 reduced Pho4's activity rather than increasing it
342 both with and without Pho2 (Fig. 5). It seems to suggest that the binding affinity difference has
343 no functional impact *in vivo*. However, it is worth noting that we assayed the activity of chimeric
344 Pho4s in the *pho80*Δ background where all Pho4 proteins are constitutively inside the nucleus
345 at near-maximal levels, hence not allowing us to test the effect of the affinity difference at a
346 lower level of nuclear Pho4. Future studies will need to measure Pho4 activities at varying
347 nuclear concentrations to test the above hypothesis.
348 Surprisingly, we found that ScPho4's DBD requires both ScPho4's P2ID and Pho2 for
349 full activity, challenging the view that TFs are modular, where DBD are expected to function

350 independently (Fig. 6C). Although our and others' studies confirm that ScPho4 DBD can bind
351 DNA *in vitro* (Fig. 2) and even *in vivo* when overexpressed on its own (Donovan et al., 2023),
352 they may not fully reflect its physiological activity, which depends on the nuclear concentration,
353 nucleosomal context, and interactions with other TF regions and cofactors. Further experiments
354 are needed to resolve the paradox and re-examine the assumption of TF modularity.

355 Only one of the three functional differences identified between the two Pho4 orthologs is
356 in a structured region. The two AEDs and the dual-functional P2ID:Sc are both within IDRs.
357 IDRs are abundant in eukaryotic TFs, with over 80% having at least one such region (Liu et al.,
358 2006). Approximately 75% of ScPho4 and CgPho4 were predicted to be IDR (Fig. S1). While
359 most studies of TF function and evolution focus on the structured regions like the DBD, our
360 results highlight the potential significance of IDR divergence in TF evolution (Hsu et al., 2021;
361 Lynch et al., 2011). On one hand, IDRs play crucial roles in nearly all aspects of a TF's function,
362 including recruiting cofactors, interacting with co-TFs, contributing to binding specificity and
363 forming molecular condensates (Brodsky et al., 2020; Ferrie et al., 2023, 2022; Ji et al., 2024;
364 Krois et al., 2018; Mindel et al., 2023). On the other hand, IDRs evolve much faster than
365 structured regions, potentially offering more raw variation for natural selection to act on. Our
366 current understanding of how IDR evolves and affects TF function is limited by challenges in
367 aligning their sequence and studying their functions. Progress in experimental techniques like
368 phase separation (Ji et al., 2024) and large language models trained on protein sequences offer
369 new opportunities to investigate TF IDR's function and evolution (Rives et al., 2021).

370 Lastly, our study illustrates a novel mechanism for the evolution of combinatorial control
371 in eukaryotes. Unlike previous research that focused on the gain and loss of TF interactions in
372 gene regulatory networks (Tuch et al., 2008; Wapinski et al., 2010), we show that TF's
373 dependence on co-TF itself can evolve. This divergence can lead to significant rewiring of the
374 regulatory network, as seen with Pho4, where reduced co-TF dependence correlates with an
375 expanded target gene network (He et al., 2017). This contrasts with the dominant pattern seen

376 in the literature on combinatorial control evolution: the Johnson lab, for example, showed that
377 changes in TF combinations in the yeast mating type pathway altered the mode of regulation
378 but maintained the overall network output (Baker et al., 2012; Britton et al., 2020; Nocedal and
379 Johnson, 2015; Tsong et al., 2006). While cases of TF interdependence evolution are still rare,
380 it is interesting to compare our example with Gain-of-Function mutations in kinases: in the proto-
381 oncogene ABL, a fusion with another gene called BCL makes the merged protein independent
382 of the upstream and downstream activators, which drives excessive cell proliferation and leads
383 to Chronic Myeloid Leukemia (Sherbenou and Druker, 2007). Whether similar changes in TF
384 dependence on co-TFs lead to misregulation and in turn causes diseases or novel phenotypes
385 is an extremely interesting question for future studies.

386 In summary, our study provides a detailed molecular picture of how co-TF dependence
387 is mediated and how it evolves, particularly through IDR changes. Further exploration of both
388 questions is essential for understanding gene regulation and regulatory evolution in eukaryotes.

389 **MATERIALS AND METHODS**

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

393 **Bioinformatic analyses of Pho4 orthologs**

394 Pho4 ortholog sequences were from the Yeast Gene Order Browser (http://ygob.ucd.ie/), and
395 were aligned using ProbCons (Do et al., 2005) via JalView's Web Service (Troshin et al., 2018;
396 Waterhouse et al., 2009). The alignment was manually edited to align the five Pho80/85 motifs
397 (O'Neill et al., 1996). Secondary structures for ScPho4 and CgPho4 were predicted using the
398 PSIPRED 4.0 (Buchan and Jones, 2019). 9aaTAD motifs were predicted using the webapp
399 https://www.med.muni.cz/9aaTAD/ (Piskacek et al., 2007). For ScPho4, one match was found
400 using the moderately stringent pattern, located at aa 75-83; for CgPho4, one match was

401 identified using the most stringent pattern, at aa 270-278, with three more using the moderately
402 stringent patterns (aa 20-28, 24-32, 282-290). Among the latter three, aa 24-32 had a lower %
403 match at 67% vs 83% for the others.

404 **Identifying Pho4 binding motifs from ChIP-seq data**

405 Chromatin Immunoprecipitation (ChIP-seq) for ScPho4 and CgPho4 were previously performed
406 in *S. cerevisiae*, with both Pho4 expressed from the same endogenous ScPho4 locus with
407 native regulatory sequences (He et al., 2017). ChIP identified peaks for both Pho4s were
408 downloaded from the supplementary files of the above publication. The sequences under each
409 peak were extracted from *S. cerevisiae* genome (sacCer3, NCBI refseq assembly
410 GCF_000146045.2), which were submitted to the "peak-motifs" tool without control sequences
411 on the RSAT Fungi server (https://rsat.france-bioinformatique.fr/fungi/) (Thomas-Chollier et al.,
412 2011). The top motif was reported for each Pho4 ortholog.

413 **Protein expression and purification for DNA binding assays**

414 All recombinant proteins were expressed using pET-11a (Sigma #69436-3) based vectors in
415 BL21(DE3) *E. coli* cells. Vector maps are available upon request. The DNA binding domain
416 (DBD) constructs included ScPho4 DBD-6xHIS (aa 236-312) and CgPho4 DBD-6xHIS (aa 452-
417 533) cloned downstream of the T7 promoter in pET-11a. The transformed bacterial strains were
418 grown overnight in LB + 100 ug/mL ampicillin. The overnight culture was diluted to OD 0.1 and
419 the cells were grown to OD 0.6 and induced with 1mM IPTG for 2 hours. Induced cells were
420 collected by centrifugation at 6000 rpm at 4ºC for 35 minutes, and were snap frozen in liquid
421 nitrogen, then stored at -80ºC until purification. Our preliminary experiments showed that the
422 protein of interest was largely in the insoluble fraction. Therefore, we performed a refold protocol
423 on the Ni-NTA column. Briefly, frozen pellets were resuspended in 1xPBS, pH7.4 with cOmplete
424 EDTA-free protease inhibitor (Sigma #11836170001) and sonicated for 45 cycles of 1 sec on / 2
425 sec off repeated 3 times at 50% power to lyse the cells. Sonicated samples were centrifuged at

426 35k rpm for 30 minutes. The supernatant was discarded, and the pellet was resuspended in the
427 solubilization buffer (20 mM TrisBase, 0.5 M NaCl, 5 mM imidazole, 5.5 M Guanidine
428 Hydrochloride, 1 mM 2-mercaptoethanol pH 8.0) and stirred at ~25ºC for an hour. The
429 solubilized pellet was centrifuged at 35k rpm for 30 minutes and the supernatant was filtered
430 and loaded onto a 5mL Ni-NTA column. The column was washed with a urea buffer (20 mM
431 TrisBase, 0.5 M NaCl, 20 mM imidazole, 5.5 M Urea, 1 mM 2-mercaptoethanol pH 8.0). Then,
432 the protein was refolded with a reverse urea gradient from buffer A (20 mM TrisBase, 0.5 M
433 NaCl, 20 mM imidazole, 5.5 M Urea, 1 mM 2-mercaptoethanol pH 8.0) to buffer B (20 mM
434 TrisBase, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0) on a BioRad FPLC.
435 The refolded protein was eluted off of the Ni-NTA column using a gradient from buffer C (20 mM
436 TrisBase, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0) to buffer D (20 mM
437 TrisBase, 0.5 M NaCl, 500 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0). Fractions
438 containing the protein of interest were identified by gel electrophoresis, pooled, and diluted with
439 a no salt buffer (25 mM Na_2HPO_4 pH 7.0, 0.5 mM THP) and run on a Heparin column
440 equilibrated with low salt buffer (25 mM Na_2HPO_4 pH 7.0, 0.15 M NaCl, 0.5 mM THP). Protein
441 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
442 mM THP). Protein-containing fractions were concentrated using a 3 kDa cutoff Amicon
443 centrifugal filter (Sigma UFC8003) and loaded onto a Superdex 75 size exclusion column
444 equilibrated with a storage buffer (25 mM Na_2HPO_4 pH 7.0, 0.5 M NaCl, 0.5 mM THP).
445 Fractions containing the expected size products were collected, analyzed by gel
446 electrophoresis, and stored in the storage buffer at 4ºC.
447 N-GST-CgPho4 full length was constructed for the Protein Binding Microarray assay.
448 The corresponding N-GST-ScPho4 purification has been described before (Zhang et al., 2021).
449 Both proteins were grown and lysed as described in (Zhang et al., 2021). Briefly, BL21 *E. coli*
450 cells containing the constructs were induced at OD600 0.8 with 1mM IPTG and collected by
451 centrifugation after 3 hours of induction at 30ºC. Pellets were snap frozen in liquid nitrogen and

452 stored at -80ºC until purification. Cells were lysed with rLysozyme (Millipore 71110) for 20
453 minutes at room temp in the lysis buffer (1xPBS, pH7.4) with the cOmplete protease inhibitor
454 tablet and 1 mM PMSF. The protein was run on a GST column equilibrated with the lysis buffer
455 and eluted with a gradient to buffer B (50mM Tris, 10mM glutathione, pH 8). GST-CgPho4
456 fractions containing the protein were pooled and loaded onto a heparin column equilibrated with
457 low salt buffer (25mM Na_2HPO_4 pH 7.4, 150mM NaCl, 0.5mM THP) and eluted with a gradient
458 to high salt buffer (25mM sodium phosphate dibasic pH 7.4, 1.5M NaCl, 0.5mM THP). Fractions
459 containing the protein were pooled and concentrated with a 30 kDa cutoff Amicon centrifugal
460 filter (Sigma UFC9030) and loaded onto a Superdex 200 size exclusion column equilibrated with
461 the storage buffer (25mM HEPES pH 7.4, 500mM NaCl, 0.5mM THP). Protein was run on an
462 SDS-PAGE and pure fractions were pooled and concentrated. 10% glycerol was added before
463 snap freezing and storage at -80ºC.

464 **Universal Protein Binding Microarray (uPBM)**

465 The uPBM was performed following the PBM protocol as described in (Berger and Bulyk, 2009).
466 GST-ScPho4 and GST-CgPho4 (full length) were prepared as described above. Both proteins
467 were assayed at a final concentration of 1 uM as determined by the optical absorbance. An
468 8x15k array was used to assay all possible 9-mers, from which a robust 7-mer enrichment score
469 is derived. The non-parametric enrichment score, or E-score, is invariant to differences in the
470 concentration of the proteins used in the assay, and thus are suitable for comparisons of relative
471 affinities between arrays. E-score ranges from -0.5 (lowest enrichment) to +0.5 (highest
472 enrichment). Scores greater than 0.35 correspond to specific TF-DNA binding (Berger and
473 Bulyk, 2006; Gordân et al., 2013). Data analysis was performed using custom Perl scripts as
474 described in (Berger and Bulyk, 2009).

475 **Genomic Context Protein Binding Microarray (gcPBM)**

476 All probes in the DNA library are 60 bp in length with 24 bp complementary to the primer used
477 for double stranding and 36 bp of genomic region centered on the E-box motif or its variant. The

478 library contains 1) 5711 *S. cerevisiae* genomic regions with putative Pho4 binding sites; 2) 150
479 negative controls from the *S. cerevisiae* genome not specifically bound by Pho4 (Zhang et al.,
480 2021); 3) 4000 DNA sequences used in previous MITOMI experiments to calibrate the binding
481 affinities (Maerkl and Quake, 2007); 4) 150 genomic regions from *C. glabrata* that contain Pho4
482 consensus and nonconsensus binding sites (He et al., 2017); 5) 100 probes from the library of
483 sequences we tested with Biolayer Interferometry; and 6) 150 negative controls from the *C.*
484 *glabrata* genome not bound by CgPho4. We used the NNNNGTG, CACNNN, and GTGNNN
485 libraries from Maerkl and Quake (2007) in our gcPBM design. The MITOMI and BLI probes
486 required the addition of random flanks to maintain the 36bp length;10 different flanking
487 sequences were generated for each sequence. Each probe is represented by six replicates,
488 including three replicates in each orientation and were distributed randomly across the array.
489 The custom 8x60k (8 chambers, 60,000 DNA spots per chamber) was synthesized by Agilent
490 Technologies. The gcPBM was performed following the PBM protocol as described in (Gordân
491 et al., 2013; Zhang et al., 2021). The log transformed median intensity of the 6 replicate probes
492 were used for comparisons.

493 **Biolayer interferometry (BLI)**

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

504 ScPho4 DBD and CgPho4 DBD were purified and stored as described above. Protein
505 quality was checked weekly for signs of degradation using Dynamic Light Scattering (DLS) on a
506 DynaPro NanoStar instrument. Before each experiment, the concentration of the protein prep
507 was measured in triplicates using a NanoDrop instrument. The mean concentration was used to
508 prepare the protein dilutions and calculate K_D.
509 BLI experiments were performed on an Octet RED 96 instrument at 30ºC with 1000 rpm
510 shaking. For each protein-probe pair, eight streptavidin (SA) biosensors were hydrated for 15-20
511 min at room temperature in the 1x kinetics buffer (1x PBS pH 7.4, 0.01% BSA, 0.002% Tween-
512 20, 0.005% Sodium azide). A black 96 well flat-bottom plate was loaded with experimental
513 components also diluted in the 1x kinetics buffer. To begin the experiment, biosensors were
514 equilibrated in a 1x kinetics buffer for 60 seconds to reach the baseline. Seven biosensors were
515 then submerged in 35 nM biotinylated annealed DNA for 30 seconds, while one biosensor was
516 submerged in the 1x kinetics buffer as a no-DNA control. Biosensors were then submerged in 1
517 ug/ml biocytin for 60 seconds to block any empty streptavidin pocket on their surface, before
518 being dipped back into a 1x kinetics buffer for 60 seconds for a baseline measurement. Loaded
519 and blocked biosensors were then submerged into a gradient of protein concentrations
520 calculated for each probe based on the K_D. To obtain reliable measurements, we use a
521 concentration range that spans 10x to 1/10 of the K_D values. Biosensors stayed in the protein
522 solution for 900-1000 seconds or until equilibrium was reached.
523 Data analysis was performed in the ForteBio Data analysis v11 software. After
524 subtracting the background and aligning the y-axis, the processed data were subjected to either
525 a steady state analysis or a kinetic curve fitting. For steady state analysis, the equilibrium-level
526 signal from each biosensor was plotted against the protein concentration, from which K_D was
527 calculated. Kinetic curve fitting was done using a one-site specific binding with Hill slope model
528 as implemented in the Data analysis v11 software. The latter more effectively fit the BLI data
529 and thus were utilized for most of the analysis. For the consensus sequence, we found the

530 kinetic curve-based estimates for both ScPho4 and CgPho4 to have higher variance, likely due
531 to the fast kinetics not adequately captured by the model (Weeramange et al., 2020). Steady-
532 state analysis gave consistent mean K_D as the kinetic analysis did, but resulted in lower
533 variance, and was used instead. Two to four replicates were performed for each 17bp DNA
534 library sequence, using at least two independent protein preps.

535 **Electrophoretic Mobility Shift Assay (EMSA)**

536 IR700 labeled 17bp consensus DNA (same as the consensus DNA for BLI above, Integrated
537 DNA technologies) was diluted to 0.1 nM and mixed 1:1 with a 2x dilution series of the DBD of
538 interest starting at 32nM in binding buffer (20mM Tris-HCL, pH 8, 150mM KCl, 10% glycerol,
539 5mM MgCl2, 1mM EDTA, and 1mM DTT). The mixture was incubated for 1 hour at 4ºC. A 10%
540 native PAGE gel in 1x Tris Glycine buffer (2.5 mM Tris pH 8.3, 19.2 mM glycine) was prerun at
541 200V for 20 min. The DNA/protein mixture was then loaded onto the gel and ran for 25-35
542 minutes. The gel was imaged using the Odyssey FC imager in the 700 nm channel. To estimate
543 K_D, the unbound band in each protein-containing lane was quantified using the ImageStudio
544 software (with background subtraction) and divided by the no protein control. This value was
545 then subtracted from 1 and plotted against the protein concentrations. A nonlinear curve fitting
546 was performed in Prism v10.2.1 using the one site specific binding model. K_D estimates and the
547 standard deviations were reported.

548 **Yeast media and growth**

549 Yeast cells were grown in Yeast extract-Peptone Dextrose (YPD) medium or Synthetic
550 Complete (SC) medium, using Yeast Nitrogen Base without amino acids (Sigma Y0626)
551 supplemented with 2% glucose and amino acid mix. Phosphate starvation medium was made
552 using Yeast Nitrogen Base with ammonium sulfate, without phosphates, without sodium chloride
553 (MP Biomedicals, 114027812) and supplemented to a final concentration of 2% glucose, 1.5
554 mg/ml potassium chloride, 0.1 mg/ml sodium chloride and amino acids, as described previously

555 (He et al., 2017). Phosphate concentration in the medium was measured using a Malachite
556 Green Phosphate Assay kit (Sigma, MAK307).

557 **Yeast strain and plasmid construction**

558 The hosts containing the endogenous *PHO5*pr-mCherry reporter were constructed by replacing
559 the endogenous coding sequence with mCherry using CRISPR/Cas9. Briefly, guide RNAs
560 targeting *PHO5* were designed in Benchling (https://www.benchling.com) and cloned into
561 bRA89 (AddGene 100950), a plasmid that contains the Cas9 protein and gRNA scaffold (Anand
562 et al., 2017). Homology arms to the *PHO5* 5'UTR and 3' UTR were added onto an mCherry
563 donor DNA using PCR. The donor DNA and plasmid were co-transformed into the host using
564 the standard LiAc transformation protocol. Transformants were selected for the CRISPR
565 plasmid using hygromycin resistance and screened with PCR. Positive clones were validated
566 using Sanger sequencing and fluorescence microscopy. Pho2 was then knocked out using a
567 *HIS3* cassette with homology arms to the *PHO2* 5'UTR and 3'UTR.

568 The chimeric Pho4 plasmid library was constructed using Golden Gate. The fragments
569 of CgPho4 and ScPho4 were PCR amplified using Phusion Flash polymerase (Thermo
570 Scientific F548S) with unique overhangs for Golden Gate. They were then assembled and
571 inserted into a pRS315-based backbone that contains the Sc*PHO4* promoter, a C-terminal in-
572 frame mNeon tag followed by the Sc*PHO4* 3' UTR and terminator. Primers were designed using
573 the NEBridge Golden Gate assembly tool. The inserts were verified with PCR and Sanger
574 sequencing and then transformed into the yeast hosts using either the standard LiAc protocol
575 (Gietz and Schiestl, 2007) or the Zymo yeast transformation kit (Zymo research T2001) and
576 plated onto SD-leu media.

577 **Flow cytometry**

578 Cells were inoculated into a 96 deep-well plate (Fisher Scientific 07-200-700) and grown
579 overnight in SD-Leu or SC medium supplemented to final concentrations of 0.13 mg/ml adenine
580 and 0.1 mg/ml tryptophan to reduce autofluorescence (Lam et al., 2008). Cultures were diluted

581 in the morning to an OD600 of 0.15 and grown to an OD600 of 0.6. These cells were directly
582 subjected to flow cytometry at a flow rate of 25 uL/min on an Attune NxT flow cytometer
583 (ThermoFisher) fitted with an autosampler. Data were collected for FSC, SSC and the
584 appropriate fluorescence channels. Calibration beads (Spherotech RCP305A) were run
585 routinely to ensure that experiments from different days were comparable. Pho4-mNeon was
586 measured in the BL1 channel with 488 nm excitation and 510±10 nm emission; mCherry was
587 measured in the YL2 channel using 561 nm excitation and 615±25 nm emission. Voltages for
588 each channel were set by the brightest sample and negative control so the sample signals were
589 between 10^2-10^5. Events were gated based on FSC-H / SSC-H to remove non-cells, then on the
590 FSC-W / FSC-H to isolate singlets. At least 10,000 gated events were collected per sample.
591 Each strain was measured at least three times, and on two different dates. A nonfluorescent
592 strain was included in every experiment for subtracting the autofluorescence. Further gating and
593 analyses were performed in R using the FlowClust and FlowCore packages. Detailed analysis
594 scripts are available in the project github repository.

595 **Fluorescent microscopy for Pho4 nuclear localization**

596 Six Pho4 constructs were chosen, with three bearing NLS:Sc (SSSSS, CCSCC, SSSCC) and
597 three bearing NLS:Cg (SSCSS, CCCCC, SSCCS). They were transformed into either a *PHO2*
598 wild type or *pho2*Δ background. Yeast cells were grown to mid-log phase in SD-Leu and fixed
599 using 4% paraformaldehyde for 10 minutes, washed twice with 1xPBS before DAPI staining.
600 4',6-diamidino-2-phenylindole (DAPI, Sigma, D9542) was diluted in the respective medium and
601 added to the fixed cells at a final concentration of 10 ug/mL. Cells were incubated with DAPI in
602 the dark for 30 minutes before DAPI was removed and cells were washed twice with 1xPBS.
603 Fixed cells were mounted on a 1.2% agarose pad on a depression slide for image. Fluorescent
604 imaging was performed on a Leica epifluorescence microscope. A 405 nm laser was used for
605 DAPI excitation and 430-550 nm for emission. For Pho4-mNeON, 488nm was used for

606 excitation and 530 nm for emission. A bright field image was also taken to show the cell
607 boundaries. A total of three images were recorded for each field of view.
608 Microscopy analysis was performed in ImageJ. To quantify the nuclear fraction of Pho4
609 proteins, ten cells were randomly selected for each construct in each host background, and the
610 experiment was repeated two times, resulting in a total of 15 or 20 cells quantified in *PHO2* and
611 *pho2*Δ backgrounds, respectively. Cell boundaries were manually traced using the bright field
612 channel, and nucleus using the DAPI channel. Both were added to the ROI manager, and the
613 raw integrated density (sum of pixel values) was quantified for both areas. The ratio of nuclear
614 vs whole cell integrated density was used for plotting and statistical analysis.
615 **Yeast one-hybrid and yeast two-hybrid**
616 Indicated regions from CgPho4, ScPho4, and ScPho2 were PCR amplified and cloned into
617 pGBD-C3 (James et al., 1996) containing the Gal4 DBD using Gibson Assembly. Plasmids were
618 verified with PCR and Sanger sequencing. The *GAL1*pr-mCherry reporter was created using a
619 pRS306 based integrative plasmid digested with StuI and inserted into the *ura3* locus of a *gal4Δ*
620 *gal80Δ S. cerevisiae* strain. The reporter was tested using flow cytometry according to the
621 protocol above using only the red (YL2) channel. Plasmids were then transformed into this yeast
622 strain using the standard LiAc method (Gietz and Schiestl, 2007) or the Zymo yeast
623 transformation kit (Zymo research T2001) and selected for on SD-trp. For the yeast two-hybrid,
624 the Gal4 activation domain (AD) and Gal4 DNA binding domain (DBD) fusion plasmids were
625 constructed by PCR amplifying the indicated regions and cloning into pGBD-C3 and pGAD-C3
626 (James et al., 1996) using Gibson Assembly. The plasmids were co-transformed and selected
627 for on SD -leu -trp. Positive colonies were patched onto fresh plates and grown at 30ºC for 24
628 hours before replica plating onto SD -leu -trp -ade -his to test for interactions. Plates were
629 imaged after 45-50 hours.

630 **Statistical analyses**

631 *Binding affinity comparison by BLI* (Fig. 2): K_D estimates from either steady state or kinetic curve
632 fitting analyses were log10 transformed and a Student's t-test was used to compare ScPho4 vs
633 CgPho4 DBDs against the same 17-bp DNA. Raw two-sided *P*-values and Holm-Bonferroni
634 corrected *P* values for the 10 tests were reported; *Activation* (Fig. 3): median fluorescence
635 intensity (MFI) for the *GAL1pr*-mCherry reporter was recorded from flow cytometry for each
636 sample. For Fig. 3C, each Gal4 DBD fusion construct was compared to the background level.
637 Because the host (with *GAL1pr*-mCherry but no Gal4 DBD plasmid) and Gal4 DBD alone
638 showed a similar level of low MFI, we combined them as the reference group. A linear model
639 was fit in R with the following command `lm(MFI ~ Genotype)`, where "Genotype" is a factor
640 representing the constructs. This model estimates a common standard deviation for all
641 constructs and tests the significance of each construct against the reference group with a two-
642 sided t-test. The raw *P*-values were corrected for multiple testing using the Holm-Bonferroni
643 procedure. Constructs with a corrected *P*-value <= 0.05 were considered significant. For Fig.
644 3D, the first two groups were tested using a two-sample t-test, while the third involved multiple
645 levels, and were tested as described above for Fig. 3C. A Holm-Bonferroni correction was
646 applied to the raw *P*-values for all six tests together. *Epistasis between regions R1, AD, and*
647 *NLS of CgPho4* (Fig. 5B): to determine the main and interaction terms in these swaps from
648 CgPho4 on ScPho4's background, we fit a linear model, $Y = X_0 + R1 + AD + NLS + R1{:}AD +$
649 $R1{:}NLS + AD{:}NLS + R1{:}AD{:}NLS$ in R, using the command `lm(A ~ R1 * AD * NLS)`, where "A"
650 is either A_{PHO2} or $A_{pho2\Delta}$, and the independent variables are coded as ScPho4 = 0, CgPho4 = 1.
651 The raw *P*-values were adjusted for multiple testing using the Holm-Bonferroni procedure. *Split*
652 *P2ID* (Fig. 6B): The same procedure as applied to the activation region test above (Fig. 3) was
653 applied to the split P2ID swaps. Two sets of four chimeras were chosen, each with a reference
654 construct having P2ID:Cg, one with the entire P2ID swapped for the ScPho4 version, and two
655 with the first or second half of P2ID swapped for the ScPho4 version. The latter three were

656 compared to the reference using a linear model `lm(MFI ~ Genotype)`, for A_{PHO2} and $A_{pho2\Delta}$
657 separately. Holm-Bonferroni correction was applied to the two sets combined (6 tests in total),
658 again separately for A_{PHO2} and $A_{pho2\Delta}$. A corrected $P < 0.05$ was considered significant. *Nuclear*
659 *fraction of Pho4 chimeras* (Fig. S4): The ratio of the sum of pixel values in the GFP channel
660 inside the nucleus vs the whole cell was treated as the response variable, and the identity of the
661 NLS region ("Cg" vs "Sc") as the predictor. First, a two-way ANOVA was performed with the
662 formula `lm(Nuc_frac ~ NLS + Host)`, where Host is either *PHO2* or *pho2*Δ. Second, given that
663 ratios are often non-normally distributed, we performed the non-parametric Kruskal-Wallis test
664 with the command `kruskal.test(Nuc_frac ~ NLS)` separately in the *PHO2* and *pho2*Δ hosts. The
665 ANOVA F-test *P*-value (0.13) was reported in the results. The two *P*-values for the second test
666 were 0.42 (*PHO2*) and 0.36 (*pho2*Δ).

667 **ACKNOWLEDGEMENT**

668 We would like to thank Dr. Miles Pufall's lab for teaching us EMSA. We thank Dr. Jan Fassler
669 for sharing many yeast plasmids for the yeast one- and two-hybrids. Dr. Kim Harold generously
670 shared the TET-on inducible promoter plasmid with us. Drs. Jan Fassler, Miles Pufall, Craig
671 Ellermeier and Todd Washington all critically read the thesis chapter by LFS, on which this
672 manuscript is based. We thank Kyle Malcolm for helping with the early development of the BLI
673 assay. Christian Weinrich discovered the P2ID's effect on Pho2-dependence during his rotation.
674 We thank Dr. Yann Vanrobaeys for helping with an initial analysis using PADDLE. We would
675 like to acknowledge the use of resources at the Protein and Crystallography Facility within the
676 Carver College of Medicine at the University of Iowa and thank Lokesh Gakhar, Zhen Xu, and
677 Devin Reusch for assistance with protein purification and BLI assays. LFS is supported by the
678 NIH Predoctoral Training grant T32GM008629. BZH is supported by NIH R35GM137831 and a
679 startup fund from the University of Iowa.

680 **FIGURE LEGENDS**

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

695

696 **Fig. 2 CgPho4 DBD recognizes the same DNA motif as ScPho4 DBD, but binds ~4 fold**
697 **tighter to the consensus.** A) Motifs generated from ChIP-seq binding peaks for ScPho4 and
698 CgPho4 in the *S. cerevisiae* genome, showing a conserved CACGTG core motif. B) Universal
699 Protein Binding Microarray (uPBM) comparing the binding of the two Pho4's. A rank-order
700 based E-score is plotted for all 7-mers, where a score greater than 0.35 indicates specific TF-
701 DNA binding. C) Genomic context PBM (gcPBM) measuring binding to ChIP-identified genomic
702 regions with or without a consensus motif (Cg=*C. glabrata*, Sc=*S. cerevisiae*). Data plotted as
703 the natural log of signal intensities for both TFs. D) Dissociation constants (K_D) determined by
704 Biolayer interferometry (BLI) for 17-bp oligos containing either the consensus or nine one-bp-off
705 half-site variants. The red bar indicates the mean of >=2 biological replicates. Asterisks next to

706 the sequence indicate a significant difference by a two-sided Student's t-test at a 0.05 level after
707 Holm-Bonferroni correction.
708
709 **Fig. 3 CgPho4 encodes two Activation Enhancer Domains (AEDs) that can enhance the**
710 **activation potential of the main AD.** (A, B) PADDLE prediction of activation potential in
711 ScPho4 and CgPho4. Z-score > 4 and 6 indicate regions with medium and strong activation
712 potential. The five functional regions are labeled as in Fig. 1. Orange triangles mark the 9aaTAD
713 motif matches. Below the plots are diagrams of the Gal4 DBD fusion constructs used to test
714 activation, with their boundaries labeled on the right. C) Activation potential of each region was
715 tested in the context of a Gal4 DBD fusion driving a *GAL1pr*-mCherry reporter. Shown are the
716 mean and standard deviation of >3 biological replicates. "host" has the mCherry reporter but not
717 the Gal4 DBD fusion plasmid; "Gal4_DBD" has the reporter and the Gal4 DBD alone. ANOVA
718 was used to compare the Gal4 DBD fusions to the two negative controls combined; an asterisk
719 indicates significant difference at a 0.05 level after Holm-Bonferroni correction. D) CgE1 and
720 CgE2_9aa can enhance the activation capability of ScAD_9aa in an orientation-independent
721 manner. An asterisk indicates significance difference at a 0.05 level by a t-test between the
722 indicated groups after Holm-Bonferroni correction.
723
724 **Fig. 4 A dual fluorescence reporter system for accurately quantifying the activity of Pho4**
725 **chimeras with and without Pho2.** (A) ScPho4 and CgPho4 were divided into five
726 corresponding parts. Names of each region were based on the annotations for ScPho4.
727 Predicted alpha helices were drawn as cartoons outside of the box diagrams. All $2^5 = 32$
728 combinations (2 were the original proteins) were made in the main chimera set, and an
729 additional set of chimeras were constructed to test specific hypotheses. (B) A diagram showing
730 the design of the dual fluorescence reporter system. The chimeric Pho4 fragments were cloned
731 downstream of the endogenous *ScPHO4* promoter and were tagged at the C-terminus with an

732 in-frame mNeon fluorescent protein, followed by the endogenous *ScPHO4* 3'UTR and
733 terminator. The plasmid contains a CEN/ARS element, resulting in stable inheritance and low
734 biological variation. (C) The *PHO5pr-mCherry* levels correlate with the ScPho4-mNeon levels
735 between biological replicates. The linear regression line is shown in blue and a 95% confidence
736 band as a gray ribbon. (D) A_{PHO2} and $A_{pho2\Delta}$ values for ScPho4 and CgPho4 across 6 different
737 days of experiments, with multiple biological replicates per day, showing consistent behaviors.
738
739 **Fig. 5. Effects of 1- or 2-region swaps reveal that R1 and NLS from CgPho4 enhance the**
740 **activity, but the chimeras remain Pho2-dependent.** (A) Heatmaps showing the effects of 1-
741 or 2-region swaps from CgPho4 into ScPho4 (left) or vice versa (right). The top triangle shows
742 the effects with Pho2. The triangles at the base show the 1-region swap effects, while the
743 squares on the top show 2-region swap effects (see a and b in the key). The bottom half shows
744 the corresponding effects measured in the *pho2*Δ host. The color indicates activity difference,
745 where red is increased activity compared to the endogenous Pho4 and blue is decreased
746 activity. (B) A subset of the chimeras was shown with their A_{PHO2} and $A_{pho2\Delta}$ values normalized
747 to those of ScPho4 (Rel. A_{PHO2} or $A_{pho2\Delta}$). The error bars represent the standard deviation of the
748 mean for A_{PHO2} and $A_{pho2\Delta}$. The set in rows 3-9 show the effect of swapping R1, AD, and NLS
749 from CgPho4 into ScPho4, both with and without Pho2. Rows 10-12 show the effects of
750 swapping CgPho4's DBD alone or in combination with other regions. The last row shows a
751 chimera with an alternative breakpoint, where the 4th region included CgPho4 P2ID and the last
752 30 aa of the NLS region, which coincided with the second activation boosting domain. C)
753 Epistasis analysis of CgPho4 R1, AD, and NLS. Linear regression with a full interaction model
754 was fit to the data. Shown are estimates of the individual and interaction terms (bar and point)
755 and standard error (line), with dark gray indicating a Holm-Bonferroni corrected $P < 0.05$. The y-
756 axis shows the effects calculated using ScPho4 as the reference.
757

758 **Fig. 6 P2ID in ScPho4 allows for enhanced activation with Pho2 via physical interaction**
759 **with the co-TF.** (A) Scatter plot of all chimeric Pho4 constructs based on their A_{PHO2} and $A_{pho2\Delta}$
760 values, colored by the identity of their P2ID region. Mixed P2ID uses an additional breakpoint
761 that separates that region into two halves (Fig. S2). (B) Comparing the effects of swapping the
762 whole P2ID from ScPho4 into CgPho4 vs individual halves. Two sets were tested (rows 2-5 and
763 6-9). Two-sided t-tests were performed as indicated by the bars and lines to the right. Holm-
764 Bonferroni correction was applied to the raw *P*-values for both sets, separately for A_{PHO2} and
765 $A_{pho2\Delta}$. An asterisk indicates corrected $P < 0.05$. (C) Testing the hypothesis that ScPho4 DBD
766 requires the P2ID:Sc and Pho2 to be fully functional. Compared with the chimeras with P2ID:Cg
767 + P2ID:Sc, which are all non-functional even with Pho2, adding back P2ID:Sc (”4+5”) rescued
768 the activity with Pho2, but not without. (D) Pho4-Pho2 physical interaction was tested with a
769 yeast two-hybrid assay. The bait and target regions were fused to the activation domain (AD)
770 and DNA binding domain (DBD) of the yeast Gal4 protein. Interaction between them
771 reconstitutes a functional Gal4 and allows cells to grow on the dropout media (right column),
772 while all strains are able to grow on the rich media (left column). “Vector” indicates a Gal4 DBD
773 or Gal4 AD only plasmid without a bait or target fused to it and is included as a control. Two
774 independent transformants are shown.
775
776 **Fig. S1 ScPho4 and CgPho4 structural prediction shows both proteins are largely**
777 **disordered outside the DBD.** (A, B) Secondary structure predictions for ScPho4 and CgPho4
778 using the PSIPRED 4.0 server. (C, D) AlphaFold2 predicted monomer 3D structure for ScPho4
779 and CgPho4. Color represents model confidence. It has been shown that low and very low
780 confidence regions overlap significantly and are predictive of protein disorder.
781
782 **Fig. S2 ScPho4 and CgPho4 alignment and annotation.** Protein sequence alignment was
783 generated using ProbCons with 2 passes of consistency transformation and 100 iterative

784 refinements and was manually edited to align the known Pho85 recognition motifs (red boxes).
785 The shading behind the amino acid symbols represents the level of conservation based on
786 BLOSUM62 scores. The functional annotation on the top were based on ScPho4 (Uniprot
787 #P07270). The orange boxes and texts refer to the five breakpoints used for constructing
788 chimeric Pho4 constructs (Fig. 5, 6).
789
790 **Fig. S3 Validation of difference in binding affinity to consensus motif. (**A) EMSA with
791 ScPho4 DBD and IR700 labeled 17bp consensus DNA. Lanes 1-7 contain 2x diluted protein;
792 lane 8 is a no protein (NP) control. (B) Fraction of bound DNA was calculated from the unbound
793 DNA band normalized to the NP lane. The fitting was done with a one site specific binding
794 model in GraphPad Prism v10. C) EMSA with CgPho4 DBD and IR700 labeled 17bp consensus
795 sequence. D) Same as (B) for CgPho4 DBD.
796
797 **Fig. S4 NLS from ScPho4 and CgPho4 conferred similar levels of nuclear localization.** Six
798 Pho4 constructs with either NLS:Sc or NLS:Cg were quantified for their nuclear fraction in either
799 *PHO2* or *pho2*Δ host background. Live fluorescence imaging was performed on mid-log phase
800 cells. DAPI staining was used to identify the nucleus. Endogenous or chimeric Pho4 protein
801 levels were quantified by the C-terminal mNeon tag. Integrated signal intensity inside the
802 nucleus was divided by total integrated intensity across the entire cell for 35 cells in each
803 construct. ANOVA with the identity of the NLS ("Cg" or "Sc") and Host (*PHO2* or *pho2*Δ) found
804 the first factor to be not significant (F-test raw $P = 0.13$), i.e., no evidence for significant
805 differences in nuclear fraction between constructs with NLS:Sc vs NLS:Cg. There was a
806 significant difference for the second factor (F-test raw $P = 0.0003$), showing higher nuclear
807 fraction in the *pho2*Δ background than in the *PHO2* background.
808 **TABLES**
809 **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-ScPho4ΔDBD	pH246	this study	Yeast two-hybrid
pH395	pGBD-Pho2Pho4int	pH247	this study	Yeast two-hybrid
pH396	pGAD-CgPho4ΔDBD	pH246	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(113-282) Sc(177-312)	pH173	this study	Chimera
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 R, Memisoglu G, Haber J. 2017. Cas9-mediated gene editing in Saccharomyces cerevisiae.

Ang Y-S, Rivas RN, Ribeiro AJS, Srivas R, Rivera J, Stone NR, Pratt K, Mohamed TMA, Fu J-D, Spencer CI, Tippens ND, Li M, Narasimha A, Radzinsky E, Moon-Grady AJ, Yu H, Pruitt BL, Snyder MP, Srivastava D. 2016. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167**:1734-1749.e22. doi:10.1016/j.cell.2016.11.033

Avsec Ž, Weilert M, Shrikumar A, Krueger S, Alexandari A, Dalal K, Fropf R, McAnany C, Gagneur J, Kundaje A, Zeitlinger J. 2021. Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet*. doi:10.1038/s41588-021-00782-6

Baker CR, Booth LN, Sorrells TR, Johnson AD. 2012. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151**:80–95. doi:10.1016/j.cell.2012.08.018

Barbaric S, Münsterkötter M, Goding C, Hörz W. 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. *Mol Cell Biol* **18**:2629–2639.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**:1266–1276. doi:10.1016/j.cell.2008.05.024

Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc* **4**:393–411. doi:10.1038/nprot.2008.195

Berger MF, Bulyk ML. 2006. Protein Binding Microarrays (PBMs) for Rapid, High-Throughput Characterization of the Sequence Specificities of DNA Binding ProteinsGene Mapping, Discovery, and Expression. New Jersey: Humana Press. pp. 245–260. doi:10.1385/1-59745-097-9:245

Bhoite LT, Allen JM, Garcia E, Thomas LR, Gregory ID, Voth WP, Whelihan K, Rolfes RJ, Stillman DJ. 2002. Mutations in the Pho2 (Bas2) Transcription Factor That Differentially Affect Activation with Its Partner Proteins Bas1, Pho4, and Swi5. *J Biol Chem* **277**:37612–37618.

Britton CS, Sorrells TR, Johnson AD. 2020. Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit. *Science* **367**:96–100. doi:10.1126/science.aax5217

Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, Barkai N. 2020. Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell*. doi:10.1016/j.molcel.2020.05.032

Buchan DWA, Jones DT. 2019. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* **47**:W402–W407. doi:10.1093/nar/gkz297

Buchler NE, Gerland U, Hwa T. 2003. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci* **100**:5136–5141. doi:10.1073/pnas.0930314100

Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S. 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* **15**:330–340. doi:10.1101/gr.2821705

Donovan BT, Chen H, Eek P, Meng Z, Jipa C, Tan S, Bai L, Poirier MG. 2023. Basic helix-loop-helix pioneer factors interact with the histone octamer to invade nucleosomes and generate nucleosome-depleted regions. *Mol Cell* **83**:1251-1263.e6. doi:10.1016/j.molcel.2023.03.006

Ferrie JJ, Karr JP, Graham TGW, Dailey GM, Zhang G, Tjian R, Darzacq X. 2023. p300 is an obligate integrator of combinatorial transcription factor inputs. *Mol Cell* S1097-2765(23)01023–7. doi:10.1016/j.molcel.2023.12.004

Ferrie JJ, Karr JP, Tjian R, Darzacq X. 2022. “Structure”-function relationships in eukaryotic transcription factors: The role of intrinsically disordered regions in gene regulation. *Mol Cell* **0**. doi:10.1016/j.molcel.2022.09.021

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**:31–34. doi:10.1038/nprot.2007.13

Gomes-Vieira AL, Wideman JG, Paes-Vieira L, Gomes SL, Richards TA, Meyer-Fernandes JR. 2018. Evolutionary conservation of a core fungal phosphate homeostasis pathway coupled to development in Blastocladiella emersonii. *Fungal Genet Biol* **115**:20–32. doi:10.1016/j.fgb.2018.04.004

Gordân R, Shen N, Dror I, Zhou T, Horton J, Rohs R, Bulyk ML. 2013. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep* **3**:1093–1104. doi:10.1016/j.celrep.2013.03.014

Guss KA, Nelson CE, Hudson A, Kraus ME, Carroll SB. 2001. Control of a Genetic Regulatory Network by a Selector Gene. *Science* **292**:1164–1167. doi:10.1126/science.1058312

He BZ, Zhou X, O’Shea EK. 2017. Evolution of reduced co-activator dependence led to target expansion of a starvation response pathway. *eLife* **6**. doi:10.7554/eLife.25157

Hirst K, Fisher F, McAndrew PC, Goding CR. 1994. The transcription factor, the Cdk, its cyclin and their regulator: directing the transcriptional response to a nutritional signal. *EMBO J* **13**:5410–5420.

Hsu IS, Strome B, Lash E, Robbins N, Cowen LE, Moses AM. 2021. A functionally divergent intrinsically disordered region underlying the conservation of stochastic signaling. *PLOS Genet* **17**:e1009629. doi:10.1371/journal.pgen.1009629

James P, Halladay J, Craig EA. 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**:1425–1436. doi:10.1093/genetics/144.4.1425

Jayaraman PS, Hirst K, Goding CR. 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. *EMBO J* **13**:2192–2199.

Ji D, Shao C, Yu J, Hou Y, Gao X, Wu Y, Wang L, Chen P. 2024. FOXA1 forms biomolecular condensates that unpack condensed chromatin to function as a pioneer factor. *Mol Cell* **84**:244-260.e7. doi:10.1016/j.molcel.2023.11.020

Kaffman A, Herskowitz I, Tjian R, O’Shea EK. 1994. Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. *Science* **263**:1153–1156.

Kerwin CL, Wykoff DD. 2009. Candida glabrata PHO4 Is Necessary and Sufficient for Pho2-Independent Transcription of Phosphate Starvation Genes. *Genetics* **182**:471–479. doi:10.1534/genetics.109.101063

Komeili A, O’Shea EK. 1999. Roles of Phosphorylation Sites in Regulating Activity of the Transcription Factor Pho4. *Science* **284**:977–980. doi:10.1126/science.284.5416.977

Krois AS, Dyson HJ, Wright PE. 2018. Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci* **115**:E11302–E11310. doi:10.1073/pnas.1814051115

Lam FH, Steger DJ, O’Shea EK. 2008. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**:246–250.

910 Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic Disorder in
911 Transcription Factors. *Biochemistry* **45**:6873–6888. doi:10.1021/bi0602718
912 Lynch VJ, May G, Wagner GP. 2011. Regulatory evolution through divergence of a
913 phosphoswitch in the transcription factor CEBPB. *Nature* **480**:383–386.
914 doi:10.1038/nature10595
915 Maerkl SJ, Quake SR. 2007. A Systems Approach to Measuring the Binding Energy
916 Landscapes of Transcription Factors. *Science* **315**:233–237.
917 doi:10.1126/science.1131007
918 Magbanua JP, Ogawa N, Harashima S, Oshima Y. 1997. The transcriptional activators of the
919 PHO regulon, Pho4p and Pho2p, interact directly with each other and with components
920 of the basal transcription machinery in Saccharomyces cerevisiae. *J Biochem (Tokyo)*
921 **121**:1182–1189.
922 McAndrew PC, Svaren J, Martin SR, Hörz W, Goding CR. 1998. Requirements for Chromatin
923 Modulation and Transcription Activation by the Pho4 Acidic Activation Domain. *Mol Cell*
924 *Biol* **18**:5818–5827. doi:10.1128/MCB.18.10.5818
925 Merhej J, Thiebaut A, Blugeon C, Pouch J, Ali Chaouche MEA, Camadro J-M, Le Crom S,
926 Lelandais G, Devaux F. 2016. A Network of Paralogous Stress Response Transcription
927 Factors in the Human Pathogen Candida glabrata. *Front Microbiol* **7**.
928 doi:10.3389/fmicb.2016.00645
929 Mindel V, Brodsky S, Cohen A, Manadre W, Jonas F, Carmi M, Barkai N. 2023. Intrinsically
930 disordered regions of the Msn2 transcription factor encode multiple functions using
931 interwoven sequence grammars. *Nucleic Acids Res* gkad1191.
932 doi:10.1093/nar/gkad1191
933 Ming Yip H, Cheng S, Olson EJ, Crone M, Maerkl SJ. 2023. Perfect adaptation achieved by
934 transport limitations governs the inorganic phosphate response in S. cerevisiae. *Proc*
935 *Natl Acad Sci U S A* **120**:e2212151120. doi:10.1073/pnas.2212151120
936 Nocedal I, Johnson AD. 2015. How Transcription Networks Evolve and Produce Biological
937 Novelty. *Cold Spring Harb Symp Quant Biol* **80**:265–274.
938 doi:10.1101/sqb.2015.80.027557
939 Nourani A, Utley RT, Allard S, Côté J. 2004. Recruitment of the NuA4 complex poises the PHO5
940 promoter for chromatin remodeling and activation. *EMBO J* **23**:2597–2607.
941 doi:10.1038/sj.emboj.7600230
942 Ogawa N, Oshima Y. 1990. Functional domains of a positive regulatory protein, PHO4, for
943 transcriptional control of the phosphatase regulon in Saccharomyces cerevisiae. *Mol*
944 *Cell Biol* **10**:2224–2236. doi:10.1128/mcb.10.5.2224-2236.1990
945 O'Neill EM, Kaffman A, Jolly ER, O'Shea EK. 1996. Regulation of PHO4 nuclear localization by
946 the PHO80-PHO85 cyclin-CDK complex. *Science* **271**:209–212.
947 Piechura JR, Amarnath K, O'Shea EK. 2017. Natural changes in light interact with circadian
948 regulation at promoters to control gene expression in cyanobacteria. *eLife* **6**.
949 doi:10.7554/eLife.32032
950 Piskacek S, Gregor M, Nemethova M, Grabner M, Kovarik P, Piskacek M. 2007. Nine-amino-
951 acid transactivation domain: establishment and prediction utilities. *Genomics* **89**:756–
952 768. doi:10.1016/j.ygeno.2007.02.003
953 Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, Fergus R.
954 2021. Biological structure and function emerge from scaling unsupervised learning to
955 250 million protein sequences. *Proc Natl Acad Sci* **118**:e2016239118.
956 doi:10.1073/pnas.2016239118
957 Sanborn AL, Yeh BT, Feigerle JT, Hao CV, Townshend RJ, Lieberman Aiden E, Dror RO,
958 Kornberg RD. 2021. Simple biochemical features underlie transcriptional activation
959 domain diversity and dynamic, fuzzy binding to Mediator. *eLife* **10**:e68068.
960 doi:10.7554/eLife.68068

961 Sánchez-Higueras C, Rastogi C, Voutev R, Bussemaker HJ, Mann RS, Hombría JC-G. 2019. In
962 vivo Hox binding specificity revealed by systematic changes to a single cis regulatory
963 module. *Nat Commun* **10**:3597. doi:10.1038/s41467-019-11416-1
964 Shao D, Creasy CL, Bergman LW. 1996. Interaction of Saccharomyces cerevisiae Pho2 with
965 Pho4 increases the accessibility of the activation domain of Pho4. *Mol Gen Genet MGG*
966 **251**:358–364. doi:10.1007/BF02172527
967 Sherbenou DW, Druker BJ. 2007. Applying the discovery of the Philadelphia chromosome. *J*
968 *Clin Invest* **117**:2067–2074. doi:10.1172/JCI31988
969 Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T.
970 1997. Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition.
971 *EMBO J* **16**:4689–4697. doi:10.1093/emboj/16.15.4689
972 Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B,
973 Bussemaker HJ, Mann RS. 2011. Cofactor binding evokes latent differences in DNA
974 binding specificity between Hox proteins. *Cell* **147**:1270–1282.
975 doi:10.1016/j.cell.2011.10.053
976 Stewart AJ, Hannenhalli S, Plotkin JB. 2012. Why Transcription Factor Binding Sites Are Ten
977 Nucleotides Long. *Genetics* **192**:973–985. doi:10.1534/genetics.112.143370
978 Tanay A, Regev A, Shamir R. 2005. Conservation and evolvability in regulatory networks: The
979 evolution of ribosomal regulation in yeast. *Proc Natl Acad Sci U S A* **102**:7203–7208.
980 doi:10.1073/pnas.0502521102
981 Thomas-Chollier M, Herrmann C, Defrance M, Sand O, Thieffry D, Helden J van. 2011. RSAT
982 peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res* gkr1104.
983 doi:10.1093/nar/gkr1104
984 Todeschini A-L, Georges A, Veitia RA. 2014. Transcription factors: specific DNA binding and
985 specific gene regulation. *Trends Genet* **30**:211–219. doi:10.1016/j.tig.2014.04.002
986 Traven A, Jelicic B, Sopta M. 2006. Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep*
987 **7**:496–499. doi:10.1038/sj.embor.7400679
988 Troshin PV, Procter JB, Sherstnev A, Barton DL, Madeira F, Barton GJ. 2018. JABAWS 2.2
989 distributed web services for Bioinformatics: protein disorder, conservation and RNA
990 secondary structure. *Bioinformatics* **34**:1939–1940. doi:10.1093/bioinformatics/bty045
991 Tsong AE, Tuch BB, Li H, Johnson AD. 2006. Evolution of alternative transcriptional circuits with
992 identical logic. *Nature* **443**:415–420. doi:10.1038/nature05099
993 Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD. 2008. The evolution of combinatorial
994 gene regulation in fungi. *PLoS Biol* **6**:e38. doi:10.1371/journal.pbio.0060038
995 Wapinski I, Pfiffner J, French C, Socha A, Thompson DA, Regev A. 2010. Gene duplication and
996 the evolution of ribosomal protein gene regulation in yeast. *Proc Natl Acad Sci U S A*
997 **107**:5505–5510. doi:10.1073/pnas.0911905107
998 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a
999 multiple sequence alignment editor and analysis workbench. *Bioinforma Oxf Engl*
1000 **25**:1189–1191. doi:10.1093/bioinformatics/btp033
1001 Weeramange CJ, Fairlamb MS, Singh D, Fenton AW, Swint-Kruse L. 2020. The strengths and
1002 limitations of using biolayer interferometry to monitor equilibrium titrations of
1003 biomolecules. *Protein Sci* pro.3827. doi:10.1002/pro.3827
1004 Zhang Y, Ho TD, Buchler NE, Gordân R. 2021. Competition for DNA binding between
1005 paralogous transcription factors determines their genomic occupancy and regulatory
1006 functions. *Genome Res* **31**:1216–1229. doi:10.1101/gr.275145.120
1007 Zhou X, O'Shea EK. 2011. Integrated approaches reveal determinants of genome-wide binding
1008 and function of the transcription factor Pho4. *Mol Cell* **42**:826–836.
1009 doi:10.1016/j.molcel.2011.05.025
1010
1011

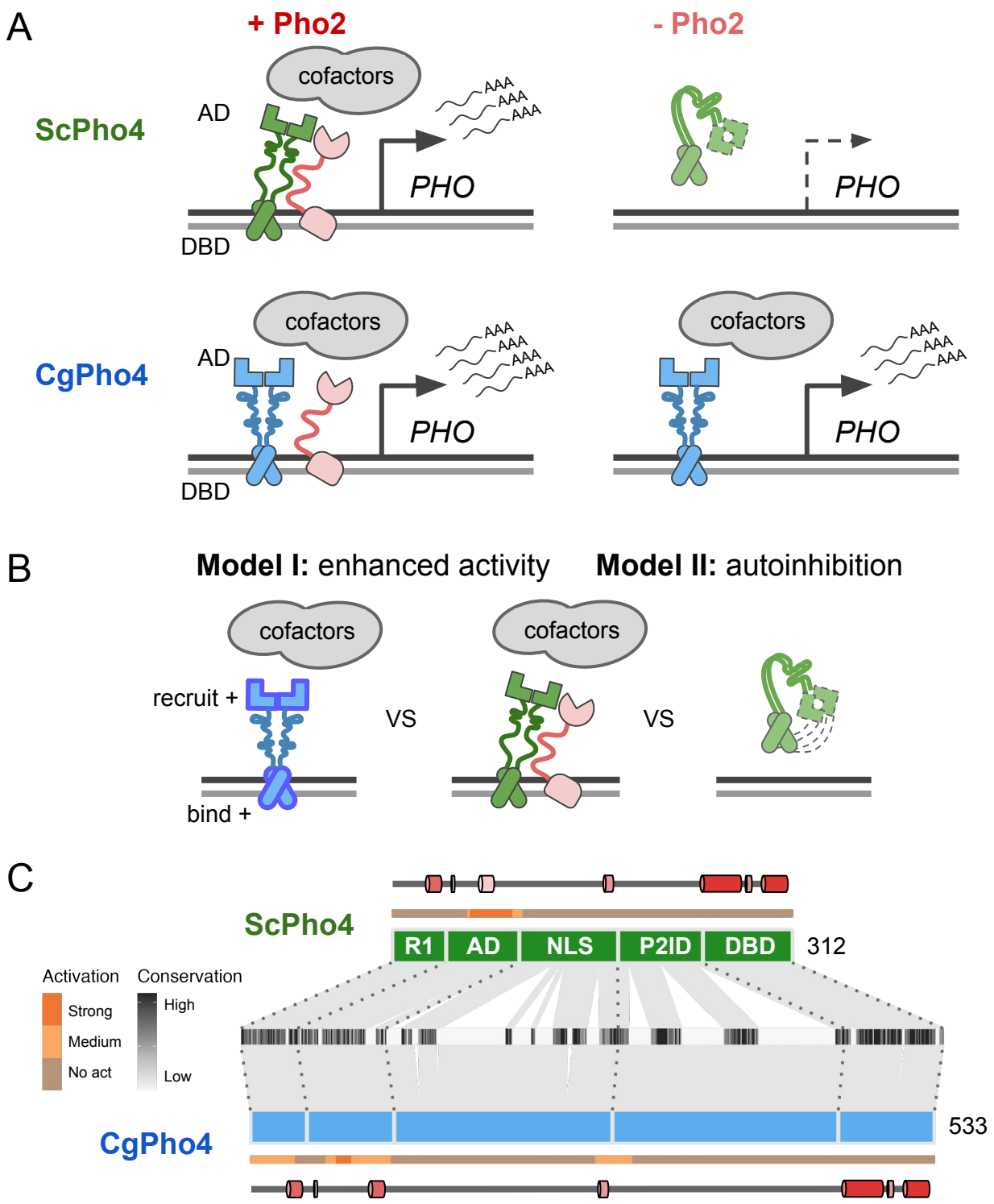


Fig. 1 Different co-TF (Pho2) dependence between orthologous TF (Pho4) in yeast species. (A) Cartoon summary of previous work: *S. cerevisiae* Pho4 (ScPho4) depends on the co-TF, Pho2, for DNA binding and activation (top row), while Pho4 from *C. glabrata* (CgPho4) can bind and activate without the co-TF. 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); DBD = 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.

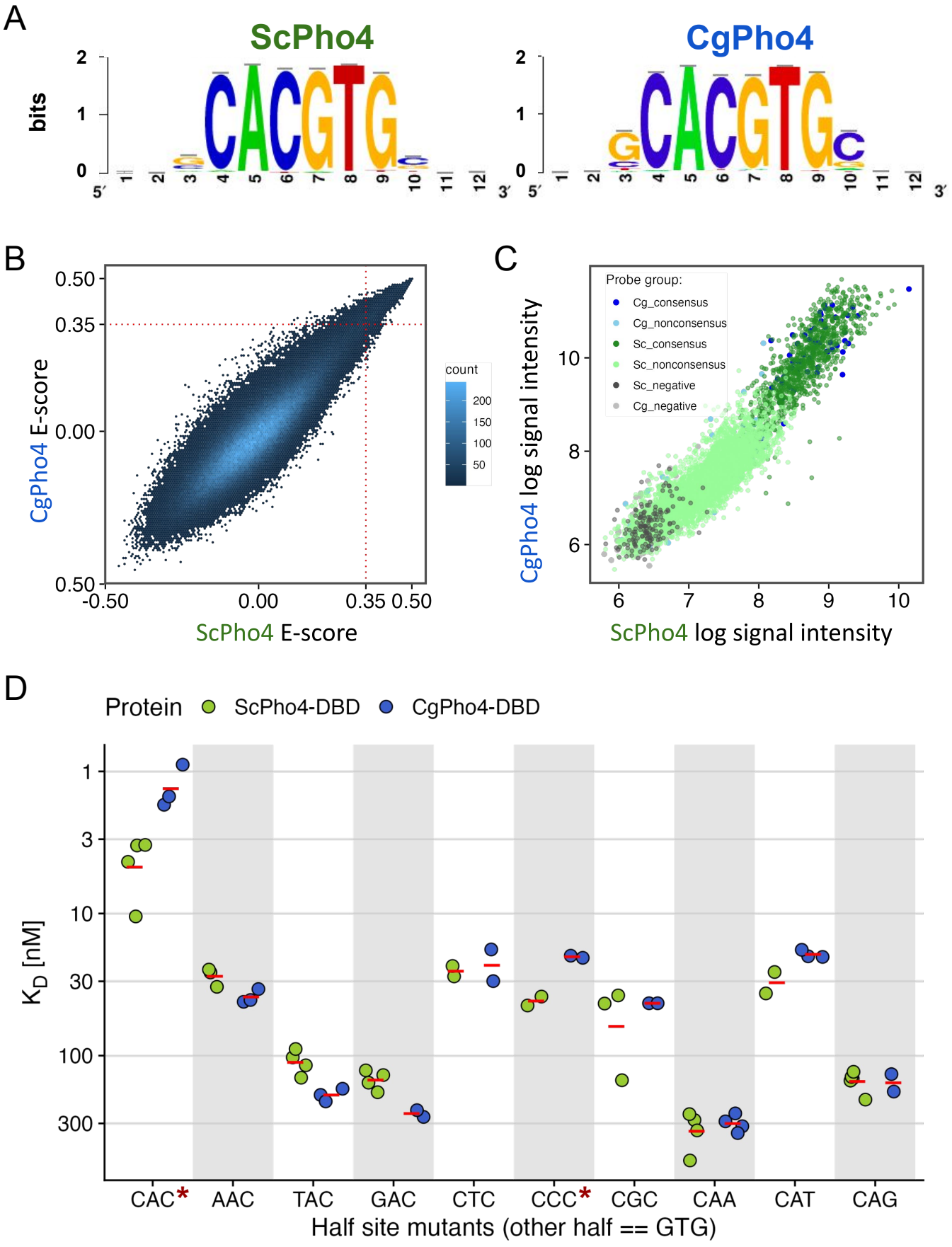
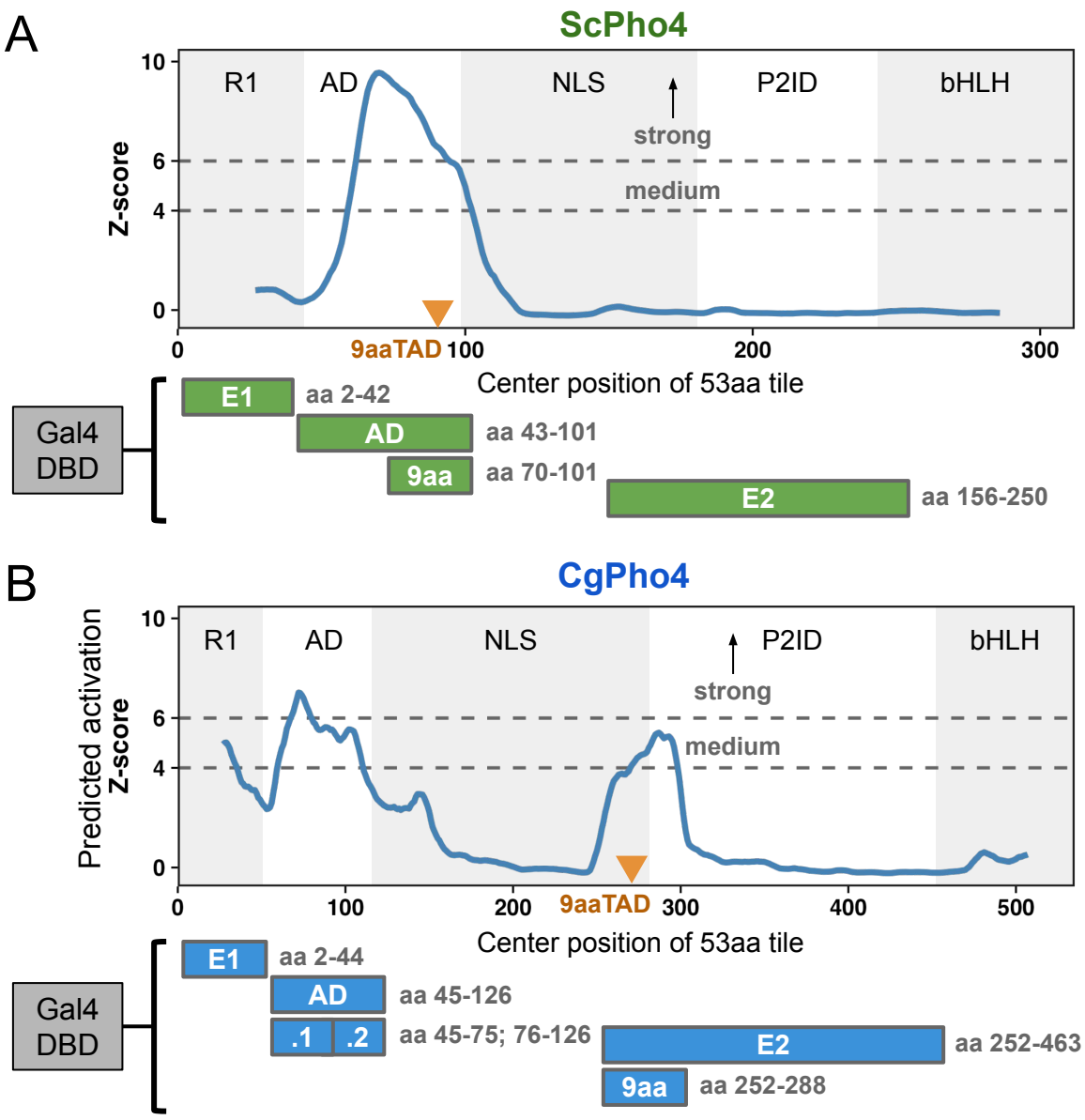
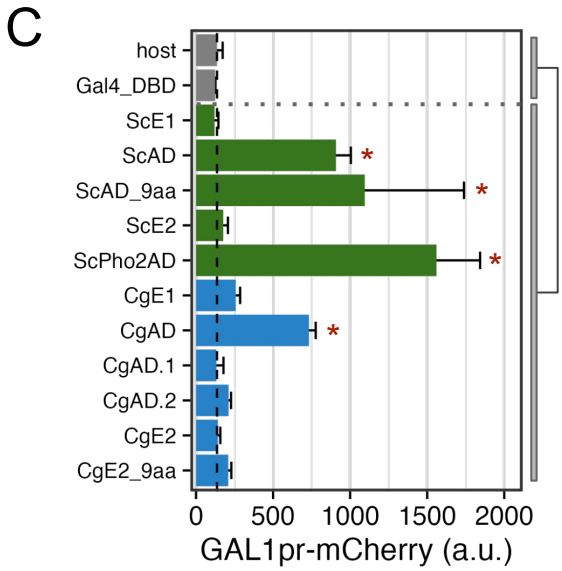
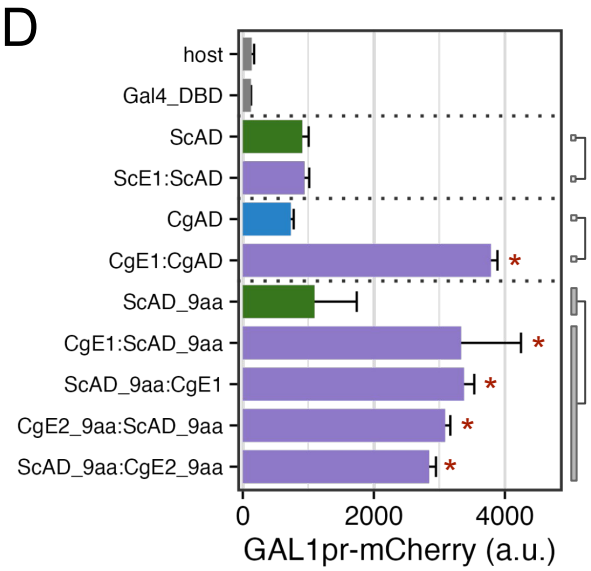

A
ScPho4
CgPho4
bits
B
CgPho4 E-score
ScPho4 E-score
count
C
Probe group:
Cg_consensus
Cg_nonconsensus
Sc_consensus
Sc_nonconsensus
Sc_negative
Cg_negative
CgPho4 log signal intensity
ScPho4 log signal intensity
D
Protein
ScPho4-DBD
CgPho4-DBD
K_D [nM]
CAC*
AAC
TAC
GAC
CTC
CCC*
CGC
CAA
CAT
CAG
Half site 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
R1
AD
NLS
P2ID
bHLH
strong
medium
Z-score
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
strong
medium
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.)

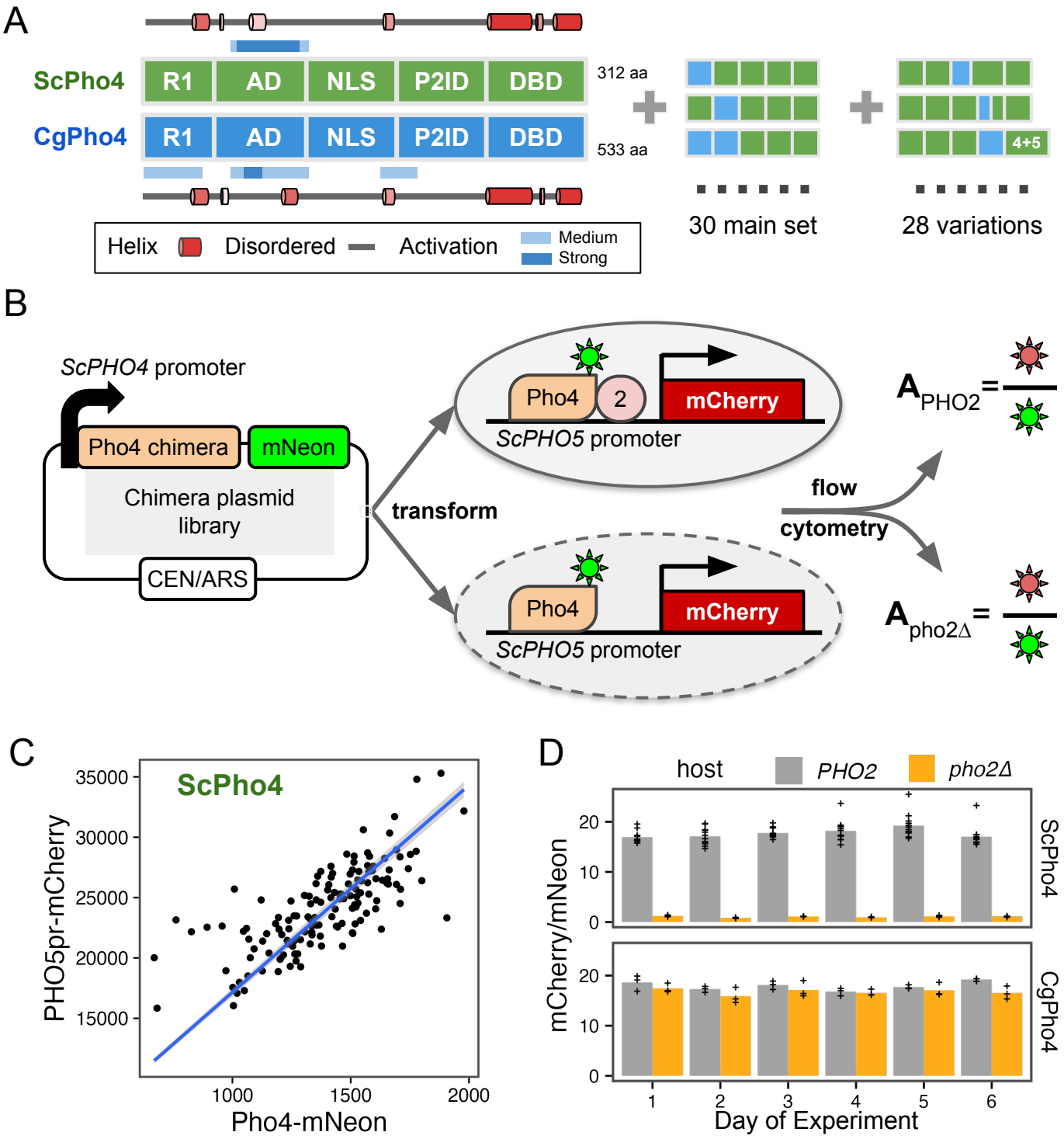


Fig. 4 A dual fluorescence reporter system for accurately quantifying the activity of Pho4 chimeras with and without Pho2. (A) ScPho4 and CgPho4 were divided into five corresponding parts. Names of each region were based on the annotations for ScPho4. Predicted alpha helices were drawn as cartoons outside 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.

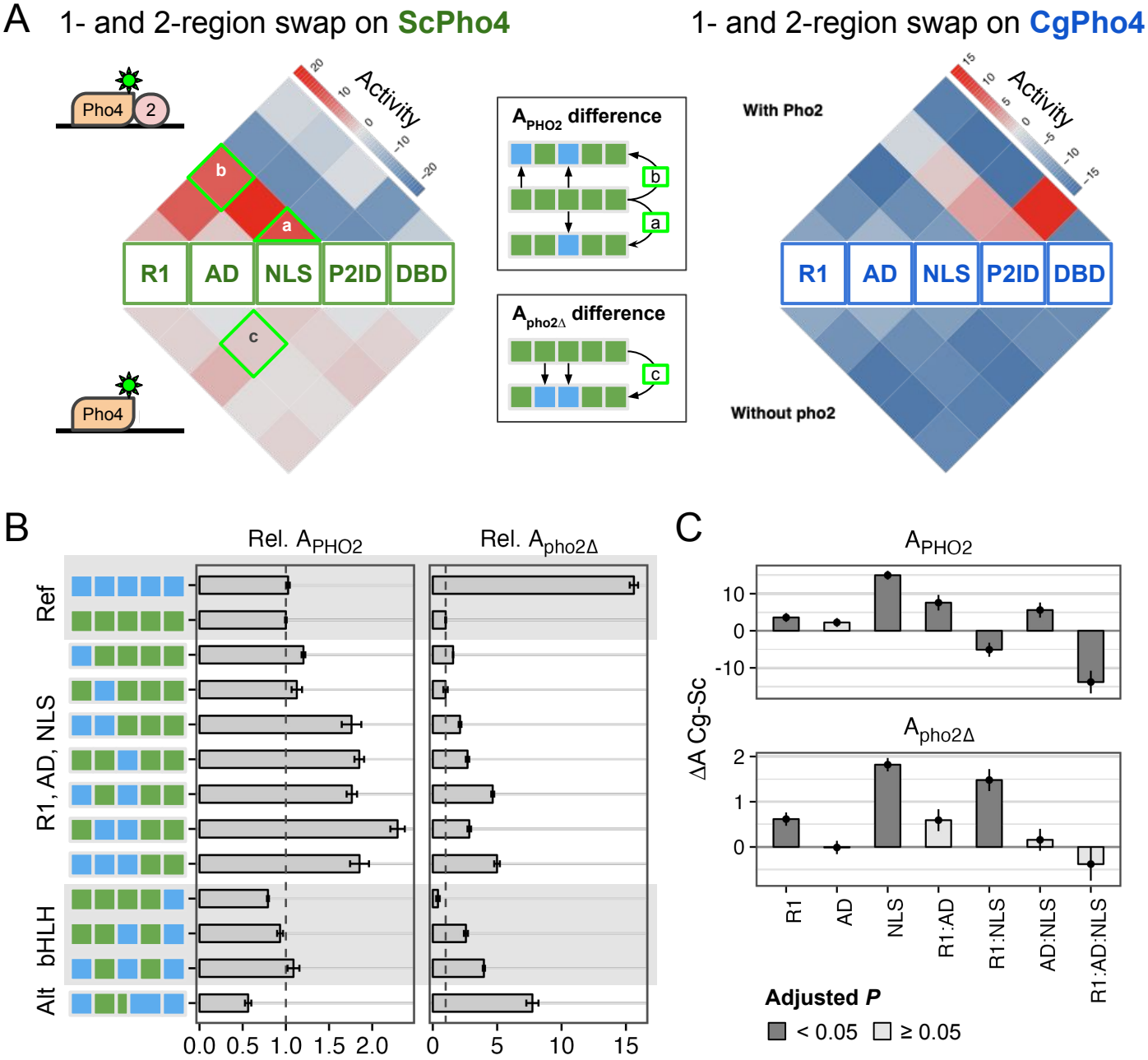


Fig. 5. Effects of 1- or 2-region swaps reveal that R1 and NLS from CgPho4 enhance the activity but the chimeras remain Pho2-dependent. (A) Heatmaps showing the effects of 1- or 2-region swaps from CgPho4 into ScPho4 (left) or vice versa (right). The top triangle shows the effects with Pho2. The triangles 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 half shows the corresponding effects measured in the *pho2*Δ host. The color indicates activity difference, where red is increased activity compared to the endogenous Pho4 and blue is decreased activity. (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 R1, AD and NLS 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 R1, AD and NLS. 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 $P < 0.05$. The y-axis shows the effects calculated using ScPho4 as the reference.

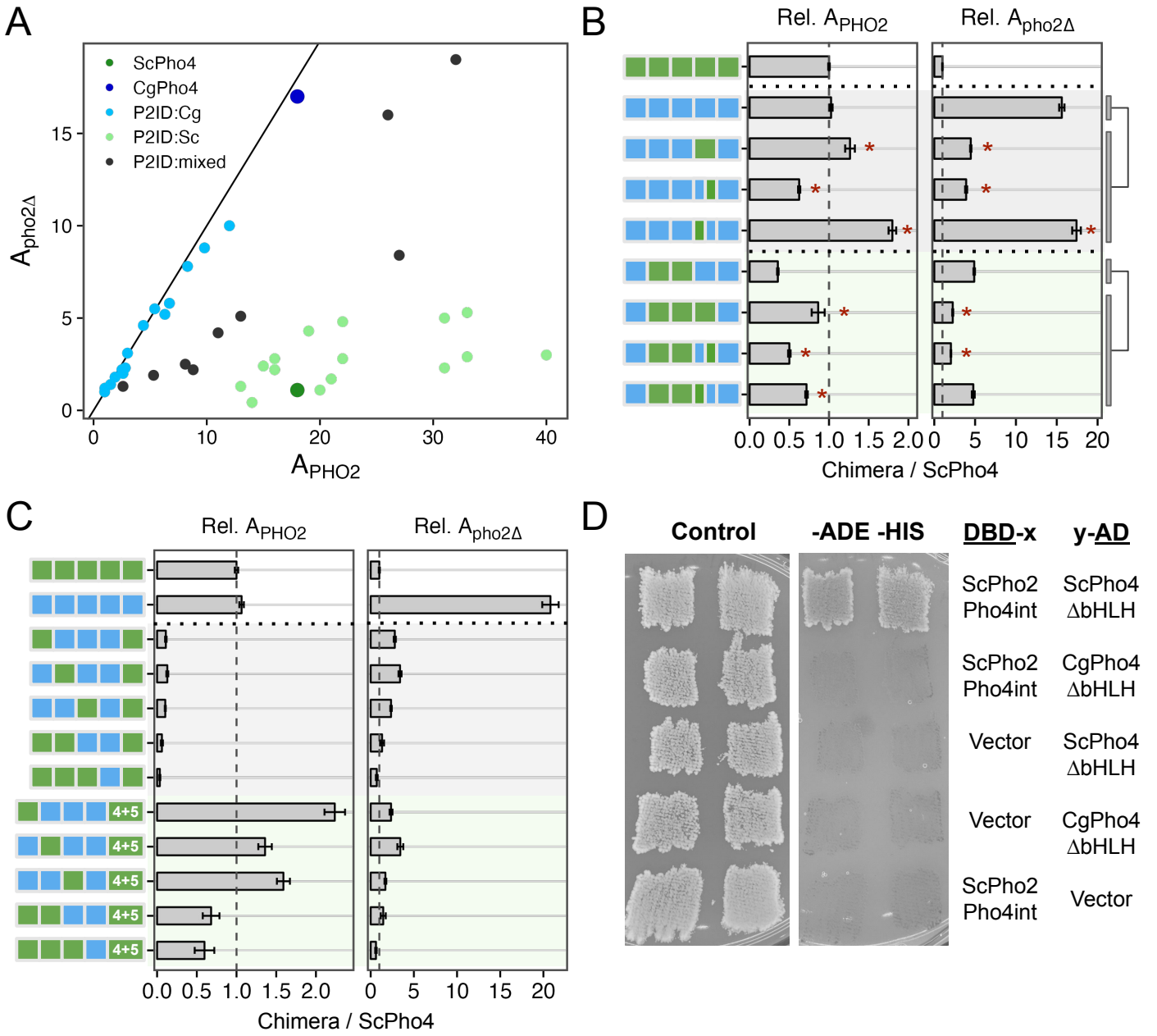


Fig. 6 P2ID in ScPho4 allows for enhanced activation with Pho2 via physical interaction with the co-TF. (A) Scatter plot of all chimeric Pho4 constructs based on their A_{PHO2} and $A_{pho2\Delta}$ values, colored by the identity of their P2ID region. Mixed P2ID uses an additional breakpoint that separates that region into two halves (Fig. S1). (B) Comparing the effects of swapping the whole P2ID from ScPho4 into CgPho4 vs individual halves. 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.

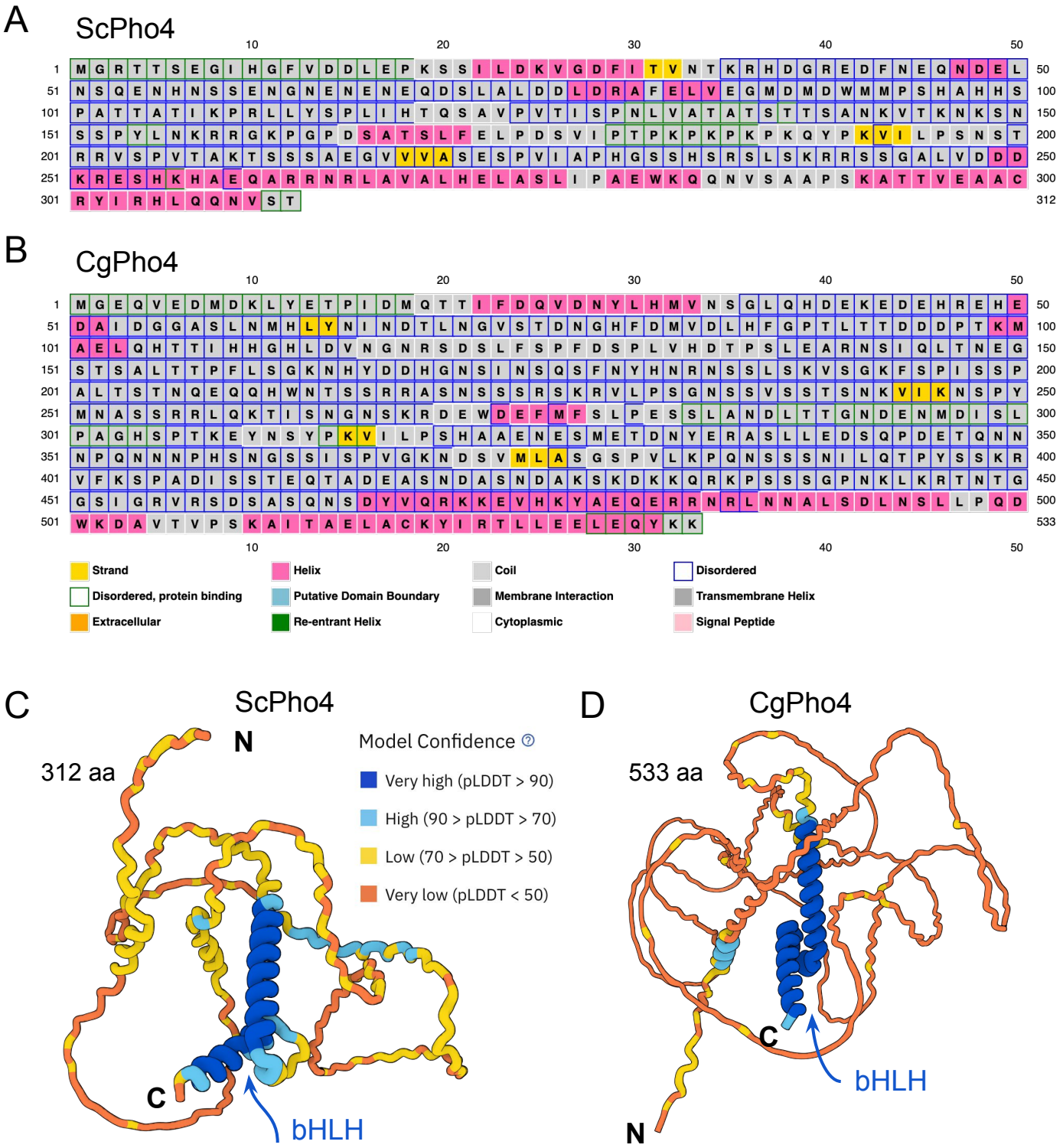


Fig. S1 ScPho4 and CgPho4 structural prediction shows both proteins are largely disordered outside the DBD. (A, B) Secondary structure predictions for ScPho4 and CgPho4 using the PSIPRED 4.0 server. (C, D) AlphaFold2 predicted monomer 3D structure for ScPho4 and CgPho4. Color represents model confidence.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

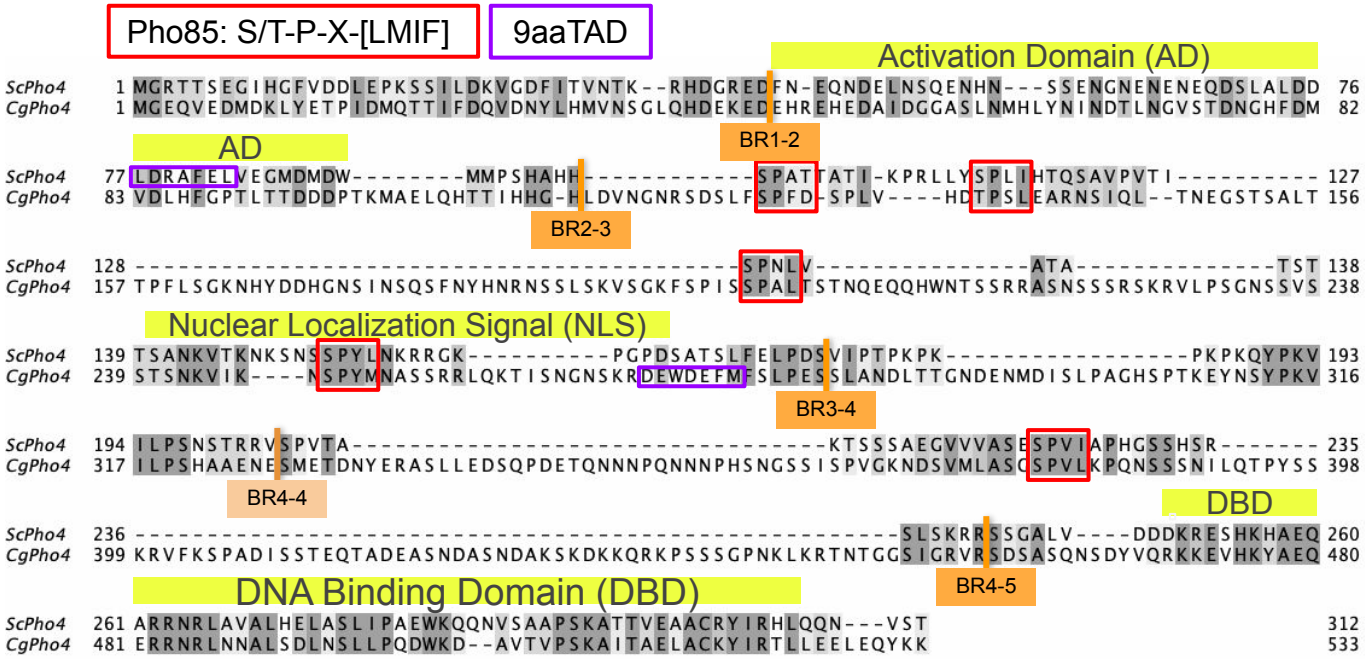


Fig. S2 ScPho4 and CgPho4 alignment and annotation. Protein sequence alignment was generated using ProbCons with 2 passes of consistency transformation and 100 iterative refinement, and was manually edited to align the known Pho85 recognition motifs (red boxes). The shading behind the amino acid symbols 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).

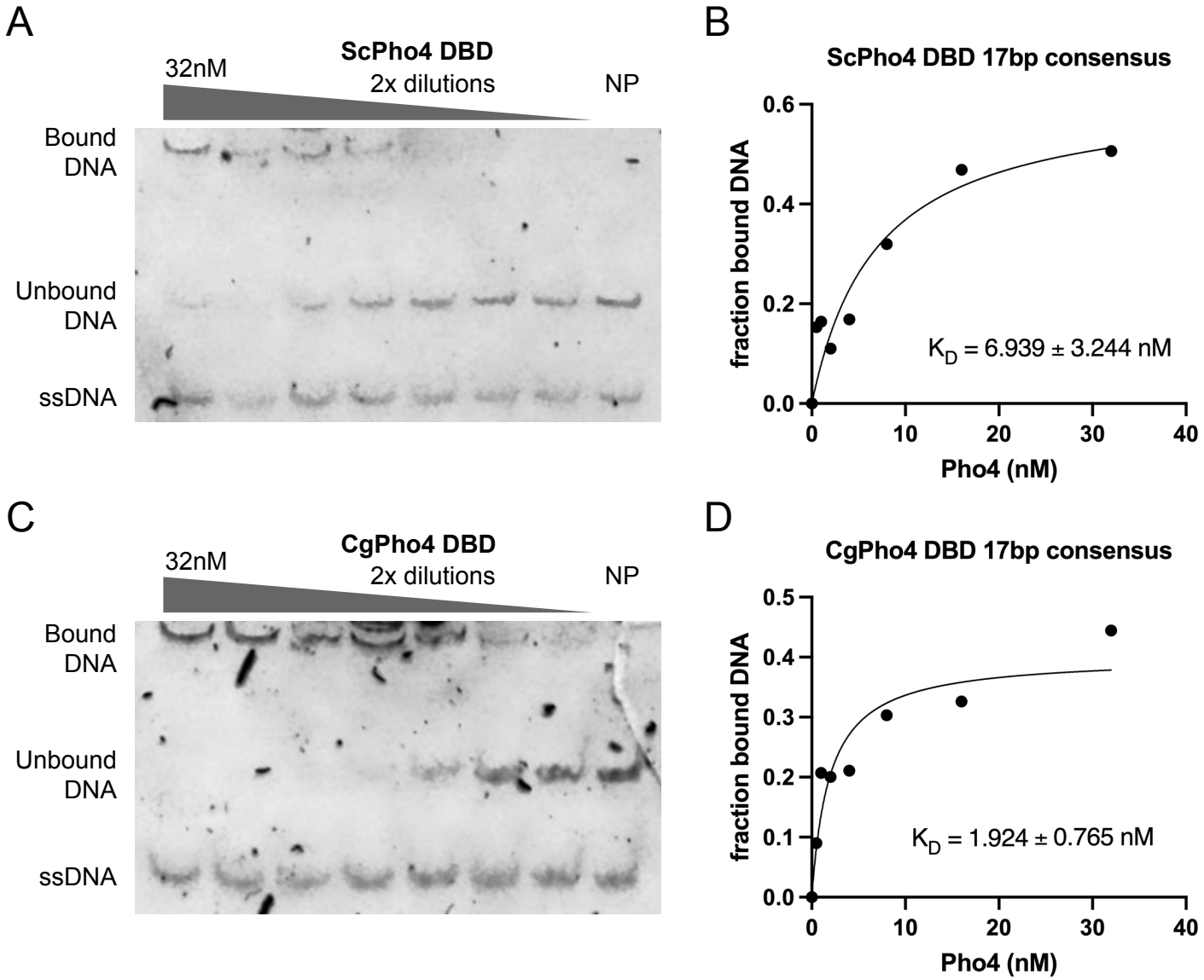


Fig. S3 Validation of difference in binding affinity to consensus motif. (A) EMSA with ScPho4 DBD and IR700 labeled 17bp consensus DNA. Lanes 1-7 contain 2x diluted protein; lane 8 is a no protein (NP) control. (B) Fraction of bound DNA was calculated from the unbound DNA band normalized to the NP lane. 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.

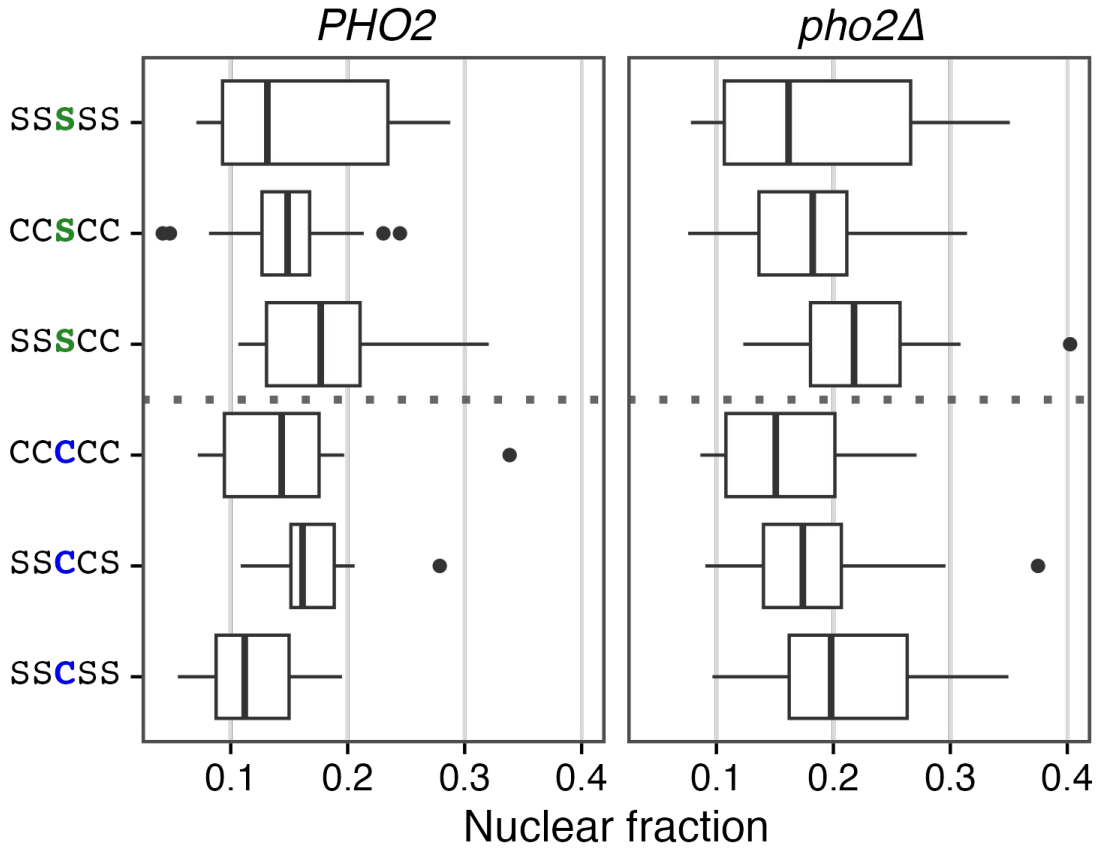


Fig. S4 NLS from ScPho4 and CgPho4 conferred similar levels of nuclear localization. Six Pho4 constructs with either NLS:Sc or NLS:Cg were quantified for their nuclear fraction in either *PHO2* or *pho2*Δ host background. Live fluorescence imaging was performed on mid-log phase cells. DAPI staining was 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.