# Paired yeast one-hybrid assays to detect DNA-binding cooperativity and antagonism across transcription factors

Anna Berenson[1], Ryan Lane[1], Luis Soto[2], Mahir Patel[3], Cosmin Ciausu[3], Zhaorong Li[1], Yilin Chen[1], Sakshi Shah[1], Clarissa Santoso[1], Xing Liu[1], Kerstin Spirohn[4,5,6], Tong Hao[4,5,6], David E. Hill[4,5,6], Marc Vidal[4,5,6], Juan I. Fuxman Bass[1,4*]

[1]Department of Biology, Boston University, Boston, MA 02215

[2]Tri-Institutional Program in Computational Biology and Medicine, New York, New York

[3]Department of Computer Science, Boston University, Boston, MA 02215

[4]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, 02215

[5]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, 02215

[6]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115

* Correspondance to: J.I.F.B. (fuxman@bu.edu).

## ABSTRACT

**Cooperativity and antagonism between transcription factors (TFs) can drastically modify their binding to regulatory DNA elements. While mapping these relationships between TFs is important for understanding their context-specific functions, existing approaches either rely on DNA binding motif predictions, interrogate one TF at a time, or study individual TFs in parallel. Here, we introduce paired yeast one-hybrid (pY1H) assays to detect cooperativity and antagonism across hundreds of TF-pairs at DNA regions of interest. We provide evidence that a wide variety of TFs are subject to modulation by other TFs in a DNA sequence-specific manner. We also demonstrate that TF-TF**

**relationships are often affected by alternative isoform usage, and identify cooperativity and antagonism between human TFs and viral proteins. pY1H assays provide a broadly applicable framework to study how different functional relationships affect protein occupancy at regulatory DNA regions.**

## MAIN

Gene expression is controlled by the binding of transcription factors (TFs) to regulatory DNA elements to direct the recruitment of cofactors and the transcriptional machinery. The logic of transcriptional regulation by TFs is complex as TFs can positively or negatively affect one another's ability to bind DNA[1-3]. This results in the binding of different combinations of TFs at promoters and enhancers, fine-tuning transcriptional output[4]. Some TFs bind DNA cooperatively, either via mutual cooperativity (e.g., as heterodimers or by indirect cooperativity) or when a DNA-bound TF recruits a second TF. Other TFs antagonize one another by sequestration via protein-protein interactions or by competing for binding at specific DNA sites. As a result of these functional relationships, individual TFs are often limited to binding DNA under certain conditions, such as in the presence of a cooperator or the absence of an antagonist.

Understanding these functional relationships between TFs at regulatory DNA regions is essential for mapping their roles in different contexts, but has thus far been difficult to achieve experimentally. DNA binding predictions based on motif analysis often identify many more potential binding events than are observed *in vivo*[5]. Predictions are generally more challenging for TF heterodimers, exacerbated by the fact that binding motifs have not been determined for most heterodimers due to challenges in producing and purifying protein complexes *in vitro*[6,7]. Single-molecule footprinting can be used to narrow down potential sites of co-binding of most TFs genome-wide; however, this approach still relies on the quality and availability of known

DNA binding motifs, as well as their ability to predict TF dimer binding[8,9]. Other genome-wide experimental methods such as ChIP-seq[10] and CUT&RUN[11] profile one TF at a time. Therefore, cooperativity between TF-pairs is often inferred from correlation in binding profiles or determined using genetic perturbations (e.g., TF overexpression, knockout, or knockdown)[3,12,13]. Additionally, genome-wide experiments are limited to detecting interactions occurring in the cell types and conditions studied which could be influenced by local chromatin states and co-expression of multiple other TFs, obscuring functional relationships between TF-pairs of interest. Furthermore, these approaches typically focus on cooperative DNA binding but do not account for antagonistic relationships.

Enhanced yeast one-hybrid (eY1H) assays provide a complementary approach by mapping protein-DNA interactions (PDIs) on a TF-wide scale using a reporter-based readout[14-17]. eY1H assays evaluate interactions between an array of hundreds of TFs and different DNA regions of interest (e.g., promoters and enhancers) which are integrated into specific loci in the yeast genome. This allows the identification of the repertoire of possible PDIs at these DNA regions rather than binding events occurring in a specific condition or cell type. However, as each arrayed yeast strain only expresses one TF, eY1H assays typically cannot identify heterodimer-DNA interactions or other cooperative or antagonistic relationships between TFs[18].

Here, we introduce paired yeast one-hybrid (pY1H) assays, an adaptation of eY1H assays using TF-pair yeast arrays to detect cooperative binding and antagonism between hundreds of TF-pairs at DNA regions of interest. This approach reveals that these functional relationships occur across well-known and lesser-known TF-pairs in a DNA sequence-specific manner. Cooperative TF-pairs have significant evidence of *in vivo* co-binding in ChIP-seq experiments and often involve one ubiquitously expressed TF and one tissue-specific TF, while antagonistic pairs frequently occur between two ubiquitous TFs. We also observe that different TF isoforms have varying functional relationships with other TFs, further expanding the TF landscape. Furthermore, we show that viral proteins are able to antagonize the binding of

human TFs to their DNA targets or direct them to new targets, providing mechanistic insight into host transcriptional reprogramming by viruses. Overall, pY1H assays constitute a robust and versatile approach to study functional relationships that modulate DNA targeting by TFs.

## RESULTS

### pY1H assay design

eY1H assays utilize a DNA-bait yeast strain containing a DNA region of interest integrated into the yeast genome upstream of two reporter genes (*HIS3* and *lacZ*) and a TF-prey strain expressing a TF fused to the Gal4 activation domain (AD). The DNA-bait and TF-prey yeast strains are mated pairwise using a robotic platform[14,19]. In the event of TF-DNA binding, the AD promotes the expression of both *HIS3* (allowing yeast to overcome inhibition by the His3p competitive inhibitor 3-amino-1,2,4-triazole) and *lacZ* (producing a blue compound in the presence of X-gal). In pY1H assays, each "TF-pair" yeast strain expresses two TFs of interest, one or both of which are fused to an AD. The two TFs are cloned into different expression vectors (pAD2μ-TRP1 and pGADT7-GW-LEU2) to allow for selection using both the TRP1 and LEU2 markers (**Fig. 1a**). These vectors both have a 2μ origin of replication and use the ADH1 promoters to express both TFs at similar levels, as evidenced by similar reporter activities for the same TF when expressed from each vector (**Supplementary Fig. 1a**). Reporter signal from the TF-pair yeast is compared to that from two corresponding single-TF control strains to detect reporter activation that is synergistic (i.e., the activity of the TF-pair is much stronger than either single-TF) or antagonistic (i.e., the activity of the TF-pair is much weaker than the activity of one of the single-TFs) (**Fig. 1b**). To analyze the pY1H data, we developed DISHA (Detection of Interactions Software for High-throughput Analyses), a computational pipeline and visual analysis tool for assessing reporter intensity and comparing yeast strains (**Supplementary Figs. 2** and **3**). By integrating DISHA analysis with manual curation, we identified cooperative and antagonistic events with a high level of reproducibility (**Supplementary Fig. 1b**).

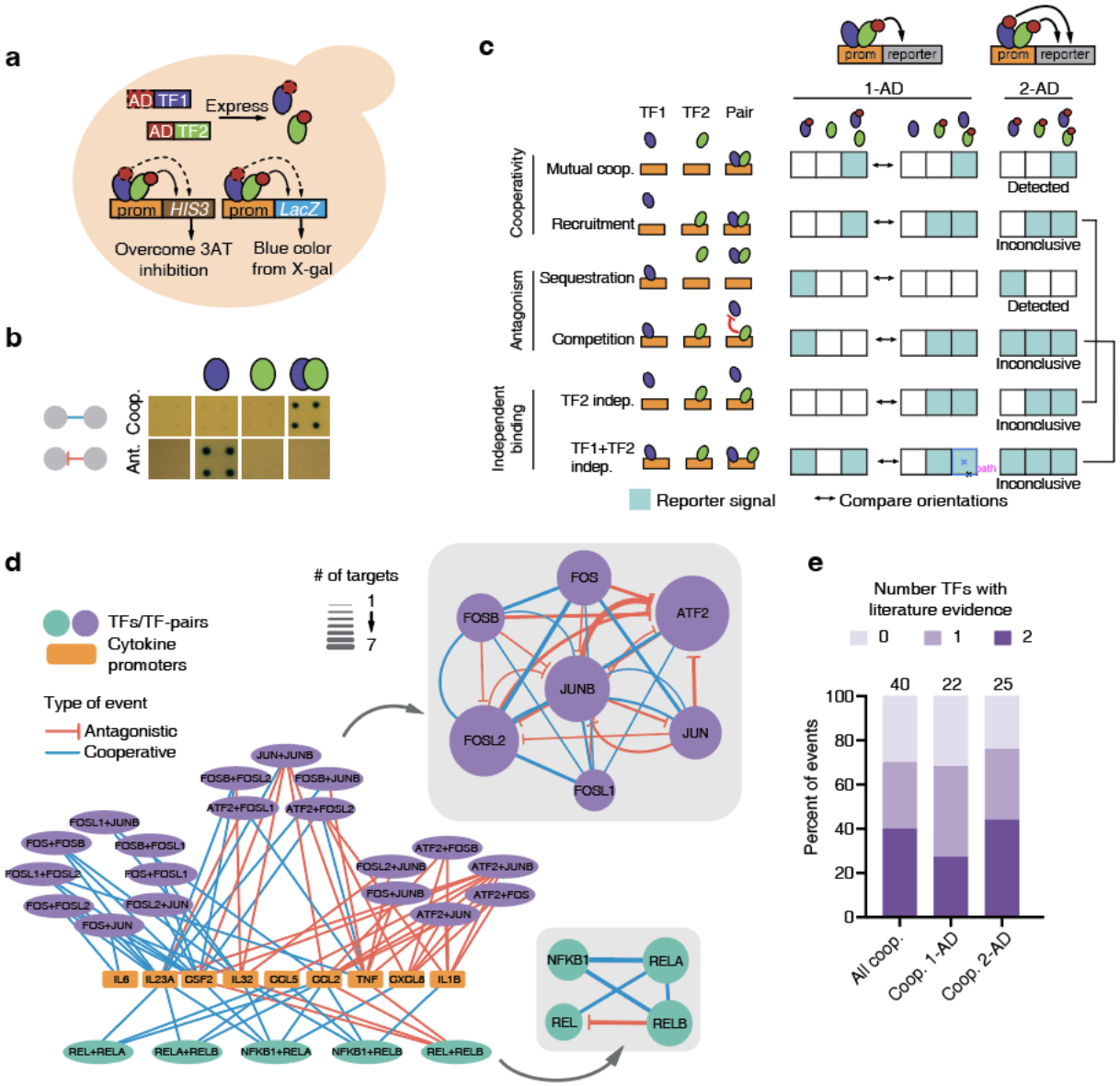


**Figure 1. Paired yeast one-hybrid (pY1H) assays.** (**a**) Schematic of pY1H assays. A DNA-bait yeast strain with a DNA sequence of interest (e.g., a promoter) cloned upstream of the *HIS3* and *lacZ* reporter genes is mated with a TF-pair prey strain expressing two TFs fused or not to the Gal4 activation domain (AD). If an AD-containing TF binds the DNA of interest, reporter expression will allow the yeast to grow in media lacking histidine and in the presence of the His3p inhibitor 3-amino-1,2,4-triazole (3AT), and turn blue in the presence of X-gal. (**b**) pY1H assays detect cooperative and antagonistic interactions by comparing single-TF and TF-pair yeast strains. (**c**) Comparison between 1-AD and 2-AD screen designs for different cooperative (mutual cooperativity and recruitment), antagonistic (sequestration and competition), and independent DNA binding modalities. Teal boxes indicate cases where reporter activity is expected. While the 1-AD can distinguish between the six indicated binding modalities if reciprocal AD orientations are tested, the 2-AD design can only detect mutual cooperativity and sequestration. (**d**) Results of pY1H screen between NF-κB and AP-1 TF-pairs and cytokine gene promoters. Main network shows connections between TF-pairs and cytokine promoters. Insets show cooperative and antagonistic relationships between TFs. Node size indicates the number of events. (**e**) Overlap of NF-κB and AP-1 pY1H interactions with the literature.

We focused on two possible pY1H assay designs, the "1-AD" design in which only one TF in each TF-pair is fused to an AD and the "2-AD" design in which both TFs are fused to an AD. These assay designs can be applied to identify different types of functional relationships **(Fig. 1c)**. By testing both possible AD orientations for each TF-pair (TF1-AD + TF2, TF1 + TF2-AD), the 1-AD design can be used to differentiate between two classes of cooperativity - mutual cooperativity and recruitment of one TF by another - and between two classes of antagonism - sequestration and competition. The 2-AD design can detect mutual cooperativity and sequestration using only one yeast strain per TF-pair, but cannot differentiate recruitment and competition from independent TF binding **(Fig. 1c)**.

**Mapping cooperative and antagonistic relationships between NF-κB and AP-1 TF-pairs**

NF-κB and AP-1 TFs often bind DNA as heterodimers, constituting a well-established model to benchmark pY1H assays and compare the 1-AD and 2-AD designs[20,21]. We evaluated the binding of 6 NF-κB and 21 AP-1 TF-pairs to the promoters of 18 cytokine genes, each known to be regulated by at least one NF-κB and one AP-1 subunit[22] **(Supplementary Tables 1-3)**. By assessing results from the 1-AD design, we observed examples of mutual cooperativity, recruitment, sequestration, and competition, while the 2-AD design showed more robust evidence of mutual cooperativity and sequestration, confirming the expected divergent uses of the two assay designs **(Supplementary Fig. 4a)**. Interestingly, though sequestration is generally expected to cause global loss of binding of the sequestered TF, some sequestering relationships such as that between REL and RELB were DNA sequence-specific, as RELB did not prevent REL binding at all promoters tested (**Supplementary Fig. 4b**). This suggests a mechanism in which TF dimerization forms a complex that retains DNA binding ability but has altered sequence specificity, as has been previously reported[23-25].

For further analysis, we considered the union of all cooperative events (including mutual cooperativity and recruitment) and antagonistic events (including sequestration and competition)

observed using either assay design (**Supplementary Table 4)**. Overall, we detected 40 cooperative binding events between 17 TF-pairs and 9 cytokine promoters (**Fig. 1d**). For 70% of these events, one or both TFs were known to regulate that cytokine[22] (**Fig. 1e**), suggesting that pY1H can recapitulate known PDIs while revealing previously undetected interactions that require cooperativity. Cooperative events identified using the two assay designs showed similar overlap with existing literature. We also observed 32 antagonistic events between 12 TF-pairs at 8 cytokine promoters (**Fig. 1d**). This includes antagonism of REL by RELB at 4 cytokine promoters, consistent with findings that RELB/RELB and REL/RELB dimers display reduced DNA binding compared to other NF-κB dimers[26,27], as well as novel antagonistic AP-1 TF-pairs. Overall, this screen detected novel instances of sequence-specific cooperativity and antagonism between highly-studied NF-κB and AP-1 TFs. This demonstrates the utility of pY1H assays to map these functional relationships and provides new information about h ow NF-κB and AP-1 subunits combine to enhance or inhibit targeting of certain promoters. Additionally, we observed the expected differences between the 1-AD and 2-AD assay designs, confirming their applicability to study different types of cooperative and antagonistic events.

## pY1H screen using a large-scale TF-pair array

We expanded the scope of pY1H assays by generating a large-scale TF-pair yeast array. We compiled a list of 868 TF-pairs based on reported protein-protein interactions or homology with interacting pairs (pTF1.0)[28,29] **(Supplementary Table 5)**. We used TF-encoding ORF clones[30-32] (**Supplementary Table 6**) to generate TF-prey yeast strains and sequence confirmed a final array of 297 TF-pairs (**Fig. 2a** and **Supplementary Table 7**), which has a similar distribution of TF families as pTF1.0 (**Fig. 2b** and **2c**). Given that the TF-pairs in our array are known or suspected to function as heterodimers, we selected the 2-AD assay design to robustly detect mutual cooperativity (hereafter “cooperativity”) and sequestration (hereafter “antagonism”) using a minimal number of yeast strains. We conducted a pY1H screen between these 297 TF-pairs

and 18 cytokine promoters (**Supplementary Table 1**) and detected 180 cooperative binding events and 257 instances of binding antagonism across 15 cytokine promoters

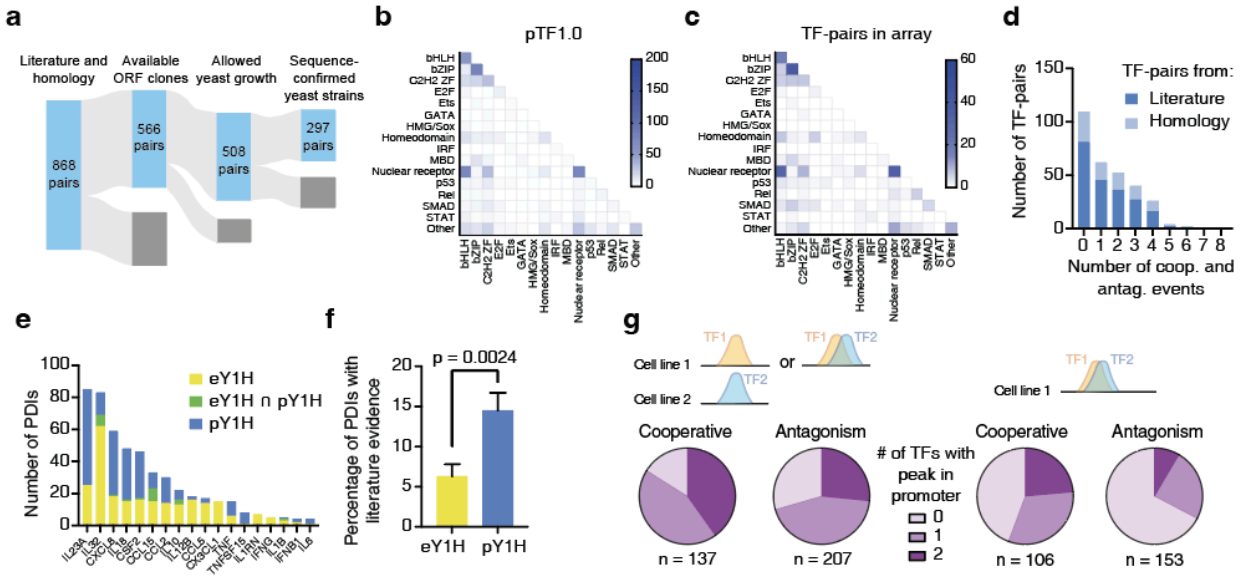


**Figure 2. Large-scale pY1H screen and validation.** (**a**) Generation of a large-scale TF-pair array for pY1H screening composed of 297 sequence-confirmed TF-pairs and their corresponding single-TF strains. (**b, c**) Number of TF-pairs for each TF family-pair in pTF1.0 (**b**) and in the TF-pair array (**c**). (**d**) Distribution of cooperative and antagonistic events detected for TF-pairs in our array. (**e**) Comparison between eY1H protein-DNA interactions (PDIs) and cooperative PDIs by pY1H assays. (**f**) Percentage of eY1H and pY1H PDIs with literature evidence. Significance by proportion comparison test. (**g**) Comparison of pY1H results with ChIP-seq data from GTRD. For pY1H interactions, we indicate whether ChIP-seq peaks for one or both TFs have been reported in any cell line (left) and in the same cell line (right).

(**Supplementary Table 8**). Of the TF-pairs tested, 63% showed at least one cooperative or antagonistic interaction, including 60 of the 88 TF-pairs selected based on homology (**Fig. 2d** and **Supplementary Figs. 6a,b**). These involve TFs from a variety of families and include both intra- and inter-family TF-pairs (**Supplementary Fig. 5c-f**), suggesting that cooperative binding and antagonism are prevalent for a wide range of TF-pairs. From our cooperative binding events, pY1H revealed an additional 234 individual PDIs not previously detected by eY1H assays at the cytokine promoters tested (**Fig. 2e**). More importantly, pY1H-derived PDIs

showed a greater overlap with the literature than eY1H PDIs (**Fig. 2f**), demonstrating that pY1H assays can recover known and novel PDIs not detectable by eY1H assays.

pY1H cooperative events significantly overlapped with motif predictions and ChIP-seq data (**Fig. 2g** and **Supplementary Figs. 6c-g**). For 40% of cooperative interactions with available data, both TFs have ChIP-seq peaks in the promoter in at least one cell line, a significantly greater overlap than expected for a randomized network (**Supplementary Fig. 6f**). Furthermore, for cell lines with ChIP-seq data for both TFs, 24% of cooperative interactions had ChIP-seq peaks for both TFs in the same cell line (**Supplementary Fig. 6g**). This provides strong evidence for *in vivo* co-binding of our cooperative TF-pairs at the target promoters identified. ChIP-seq overlap for antagonistic TF-pairs was not significant. This was expected, as we hypothesize that our antagonistic events represent sequestration rather than competitive binding of both TFs.

**TF-TF relationships are sequence-specific and connect ubiquitous and tissue-specific TFs**

While 83 TFs participated exclusively in either cooperativity or antagonism across the cytokine promoters tested, 54 TFs, including FOS and others typically considered to be mainly cooperative, participated in both event types, suggesting that individual TFs have distinct functional relationships with different TF partners. (**Figs. 3a-c**). Interestingly, 21 TF-pairs were cooperative or antagonistic depending on the promoter sequence (**Fig. 3c**), likely due to motif presence, spacing, and orientation. For example, MXI1 antagonized MAX at the IL18 and CCL15 promoters which have MAX motifs but no MXI1 motifs, while both TFs cooperated at the CCL5 promoter that has overlapping MAX/MXI1 motifs at 2 locations (**Supplementary Fig. 7a**). The observed differences in functional relationships with TF partners even extend to paralogous TFs . While some sets of highly similar TF paralogs showed identical relationships with TF partners, others showed major differences in both their TF-TF relationships and DNA targets

(**Fig. 3d and Supplementary Fig. 7b**). This suggests partner and target neofunctionalization and subfunctionalization between paralogs, and may explain the lack of specificity observed for DNA binding predictions that rely on very similar motif preferences between paralogs.

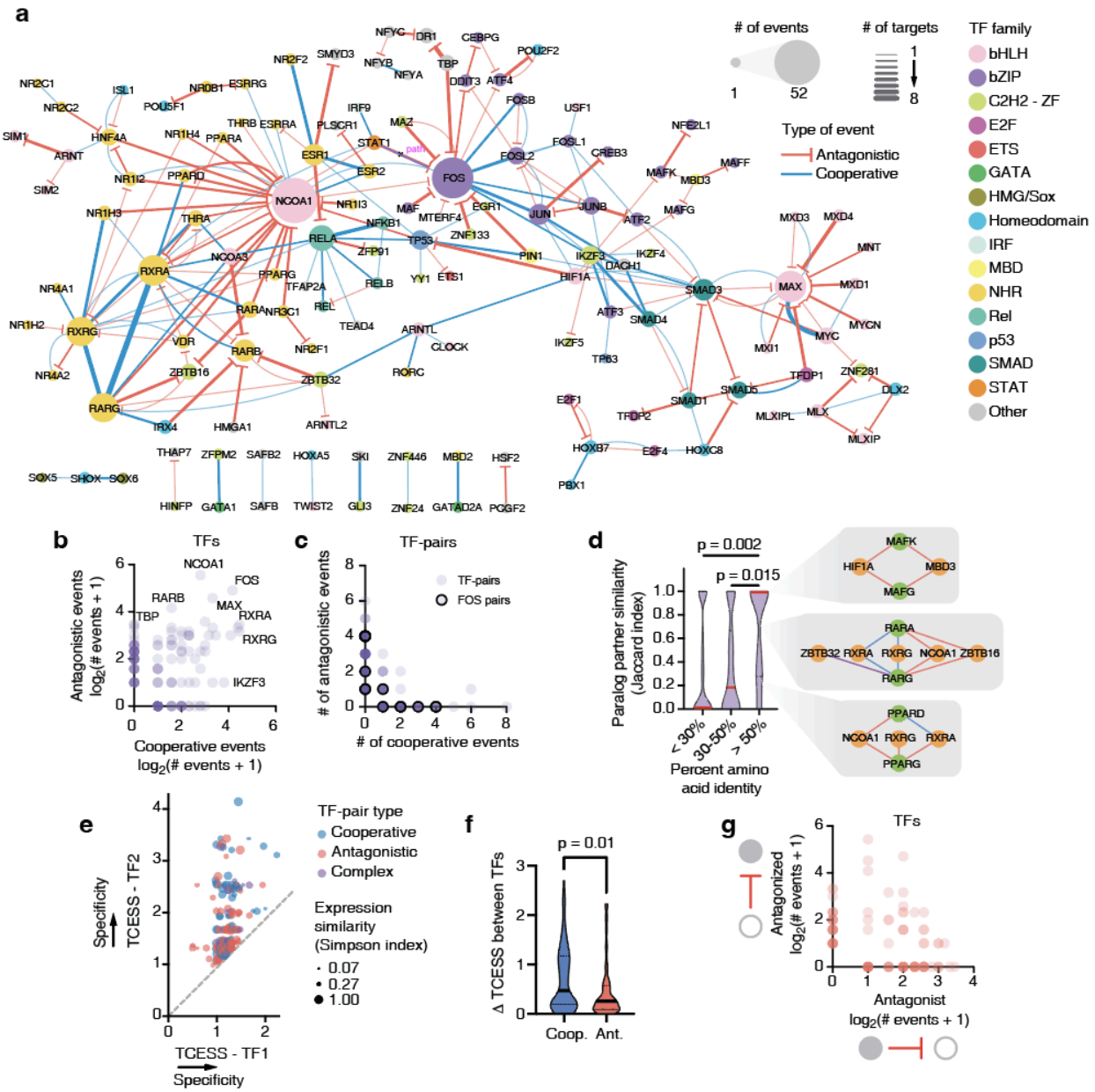


**Figure 3. pY1H assays map cooperative and antagonistic relationships between TFs.** (**a**) Network of cooperative and antagonistic relationships between TFs at cytokine promoters screened. (**b**) Number of cooperative and antagonistic events observed for individual TFs. (**c**) Number of cooperative and antagonistic events observed for TF-pairs. FOS-containing pairs are outlined in black. (**d**) Similarity in cooperative and antagonistic relationships with shared TF partners (Jaccard index) between paralogs. Significance determined by Mann-Whitney's U test. Insets show relationships between paralog-pairs (green) with partners (orange). Edges in red, blue, and purple indicate antagonistic, cooperative, and

complex relationships, respectively. (**e**) Tissue/cell-type expression specificity score (TCESS) for TFs in pairs showing cooperativity, antagonism, or both (complex). For each TF-pair, the larger TCESS value was plotted on the y-axis. Dot size indicates the Simpson co-expression similarity. (**f**) Difference in TCESS between TFs in cooperative and antagonistic pairs. Significance by Mann-Whitney's U test. (**g**) Number of antagonistic events in which each TF acted as the "antagonist TF" or "antagonized TF".

Cooperativity and antagonism may be mechanisms by which tissue- and cell type-specific TFs modulate the function of more ubiquitous TFs. Using single-cell RNA-seq data from the Tabula Sapiens atlas[33], we calculated a tissue/cell type expression specificity score (TCESS) for TFs in pairs demonstrating cooperativity and/or antagonism, where TFs with TCESS ~ 1 are ubiquitously expressed and higher values indicate greater tissue specificity (**Supplementary Tables 9** and**10**). We observed that these functional relationships often occur between ubiquitous-ubiquitous and ubiquitous-specific TF-pairs (**Fig. 3e**). Even for ubiquitous-specific TF-pairs, TFs were expressed in overlapping sets of tissues, with 97% of all TF-pairs coexpressed in at least one tissue or cell type, indicating potential venues for cooperative and antagonistic interactions to occur *in vivo*. Interestingly, TFs in cooperative pairs had a significantly greater difference in TCESS than TFs in antagonistic pairs, while the expression overlap was similar for both types of TF-pairs (**Fig. 3f**). This suggests that cooperativity is the preferred mechanism for modulation of ubiquitous TFs by tissue-specific TFs, as cooperative events more commonly occur between ubiquitous-specific pairs, while antagonism may constitute a broader mechanism whereby pairs of ubiquitous TFs limit one another's DNA binding across a wide range of tissues and cell types.

**pY1H assays identify highly cooperative and frequently antagonized TFs**

Cooperative binding events were observed between 95 TF-pairs from diverse TF families (**Supplementary Fig. 5c,d**). About 2/3 of these events indicated obligate cooperative binding, while about 1/3 showed enhanced binding of one or both TFs. This includes known heterodimers such as bHLH, nuclear hormone receptor, bZIP, and Rel pairs (**Supplementary**

**Fig. 5c,d**). Interestingly, we observed many TFs that participated in a disproportionate number of cooperative binding events (e.g., TP53, RXRA, RELA, and IKZF3) many of which, to our knowledge, have not been reported. This confirms the utility of pY1H assays to detect novel cooperative events in an unbiased manner.

Extensive antagonism was also observed between 114 TF-pairs (**Supplementary Fig. 5e,f**). Some TFs such as NCOA1, FOS, MAX, and RARB were frequently antagonized (**Fig. 3a**), suggesting that these TFs are highly influenced by the repertoire of co-expressed TFs. While most TFs functioned exclusively as antagonists or antagonized TFs in our screen, 27 TFs participated in each role at different promoters, suggesting that the role of a given TF depends on its TF partner as well as the target DNA sequence (**Fig. 3g**). This is likely due to differences in specificity between the individual TFs.

**Alternative isoform usage alters TF-TF relationships**

Most human TFs are expressed as multiple isoforms, expanding the number of functionally distinct TFs[34,35]. We used pY1H assays to determine whether alternative isoforms of a given TF differ in their functional relationships with other TFs. We screened 37 TF isoform-pairs involving immune-related TFs for binding to 102 cytokine gene promoters (**Fig. 4a**) (**Supplementary Tables 1, 11,** and **12**). Alternative isoforms often differed in binding modalities, in many cases switching between dependent binding types (cooperative and antagonistic) **(Fig. 4a,b, Supplementary Table 13)**. For example, while the STAT1-202 isoform showed cooperative binding with IRF9, the STAT1-201 isoform antagonized IRF9 binding **(Fig. 4c)**. In other cases, alternative isoforms had varying levels of dependence on other TFs, switching between dependent and independent binding. For example, DNA binding of the MAX-205 isoform was typically independent of MNT, while binding of the MAX-202 isoform was always antagonized by MNT (**Fig. 4d**).

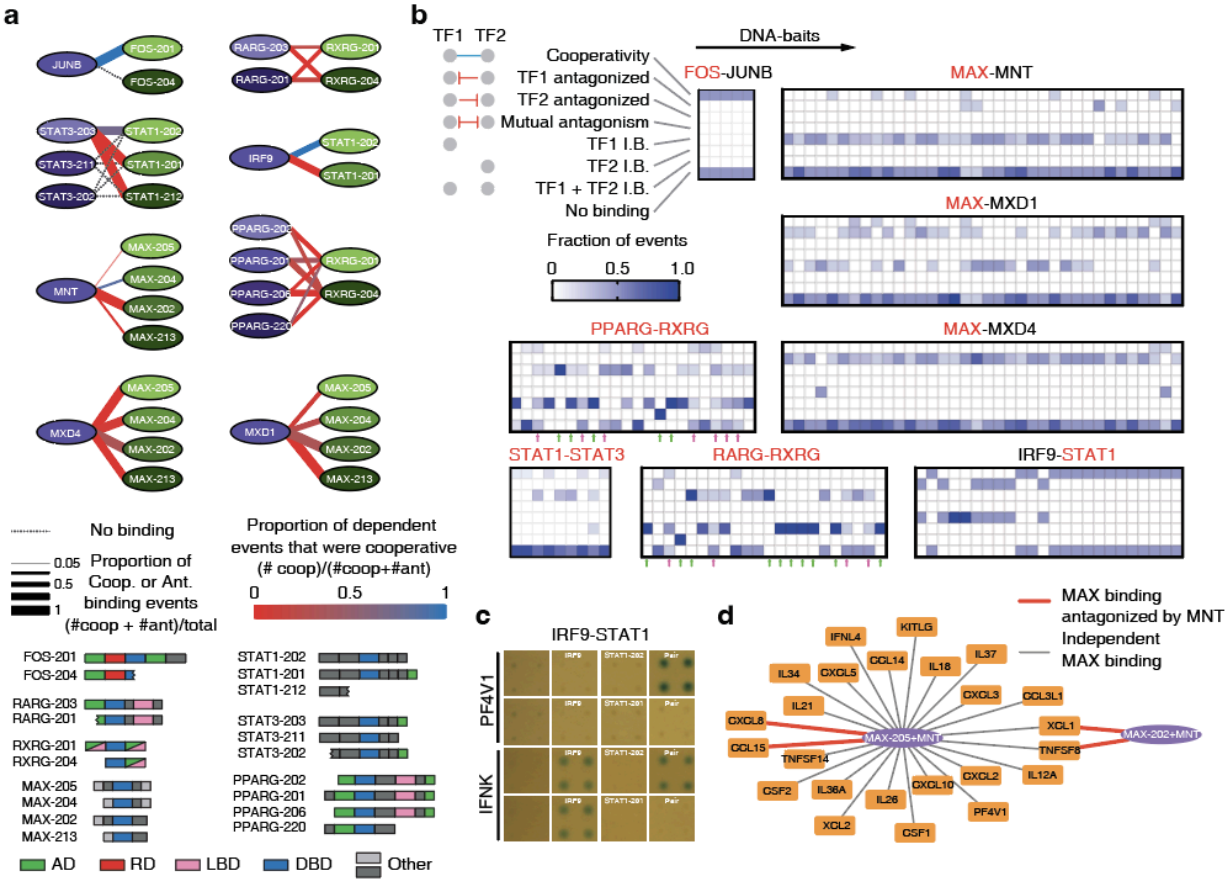


**Figure 4. Application of pY1H to study TF isoforms.** (**a**) Relationships between TF isoform-pairs observed by pY1H assays. Domain-based schematic of TF isoforms studied are indicated. (**b**) For each TF-pair, the proportion of TF isoform-pairs that show each type of binding modality (cooperativity, TF1 or TF2 antagonized, mutual antagonism, TF1 and/or TF2 independent binding, or no binding) across DNA-baits. Names in red indicate TFs for which alternative isoforms were studied. Green arrows indicate DNA-baits where all TF isoform-pairs for PPARG-RXRG or RARG-RXRG show identical binding modalities; magenta arrows indicate DNA-baits where different TF isoform-pairs for PPARG-RXRG or RARG-RXRG show at least three different binding modalities. (**c**) Relationship between alternative STAT1 isoforms and IRF9 at the PF4V1 and IFNK promoters. (**d**) Interactions between MNT-MAX dimers and cytokine promoters. Gray lines indicate independent MAX binding to the cytokine promoter, whereas red lines indicate that MAX binding was antagonized by MNT.

Although the binding modalities were often similar across DNA targets for specific isoform-pairs, in other cases the effect of isoform usage differed between promoters. For PPARG/RXRG and RARG/RXRG, alternative isoforms showed identical binding modalities at some promoters (**Fig. 4b** green arrows) and divergent modalities at other promoters (**Fig. 4b** magenta arrows).

As alternative TF isoforms can differ in both DNA binding and protein-protein interactions due to gain or loss of different protein domains, we suspect that alternative isoform usage can affect DNA binding modalities by multiple different mechanisms. For example, STAT3-203 shows mostly cooperative binding with STAT1-202 but is antagonized by STAT1-212, a truncated isoform missing its DNA binding domain, suggesting that the STAT3/STAT1-212 dimer has reduced DNA binding affinity (**Fig. 4a**). However, STAT3 binding is also antagonized by the STAT1-201 isoform, which retains its DNA binding domain but has an additional protein binding/activation domain, suggesting an alternative mechanism where altered affinity between STAT3 and STAT1 affects the equilibrium between STAT3/STAT3 homodimers and STAT3/STAT1 heterodimers. Altogether, these findings suggest that alternative isoforms may affect DNA targeting by forming complexes with altered DNA binding specificity and/or due to differences in protein-protein interactions.

**Viral proteins alter DNA targeting of host genes by human TFs**

Viruses express viral transcriptional regulators (vTRs) that can modulate host gene expression, altering immune responses, apoptosis, differentiation, and cell cycle dynamics[36]. vTRs participate in extensive interactions with human proteins[36-38], but less is known about the functional outcomes of these interactions. We leveraged pY1H assays to investigate mechanisms by which vTRs affect binding of human TFs to gene promoters **(Fig 5a)**. We generated a pY1H array of 113 protein pairs containing one human TF and one vTR **(Fig. 5b)** and screened for interactions with 83 promoters of cancer-related genes (**Supplementary Tables 1,14,15**). We observed both cooperativity and antagonism between 11 vTRs and 11 human TFs **(Fig. 5c, Supplementary Table 16)**. Interestingly, the HBZ protein from human T-lymphotropic virus 1 (HTLV-1) cooperated with human DDIT3 to bind two promoters, but antagonized the binding of CEBPG to four promoters, although both DDIT3 and CEBPG are bZIP TFs. This indicates that a given vTR can have different effects on human TFs, even within

the same TF family. Distinct vTRs from a virus can also have different effects on the binding of a human TF. For example, Epstein-Barr virus proteins EBNA3B and EBNA3C cooperated with and antagonized RBPJ, respectively, providing a potential mechanism for observations that EBNA3 proteins alter the expression of distinct sets of host genes via interactions with RBPJ[39-41]. Most of the functional relationships we found between vTRs and human TFs were novel and therefore provide evidence suggesting that different viruses can rewire host gene regulatory networks by altering host TF targets.

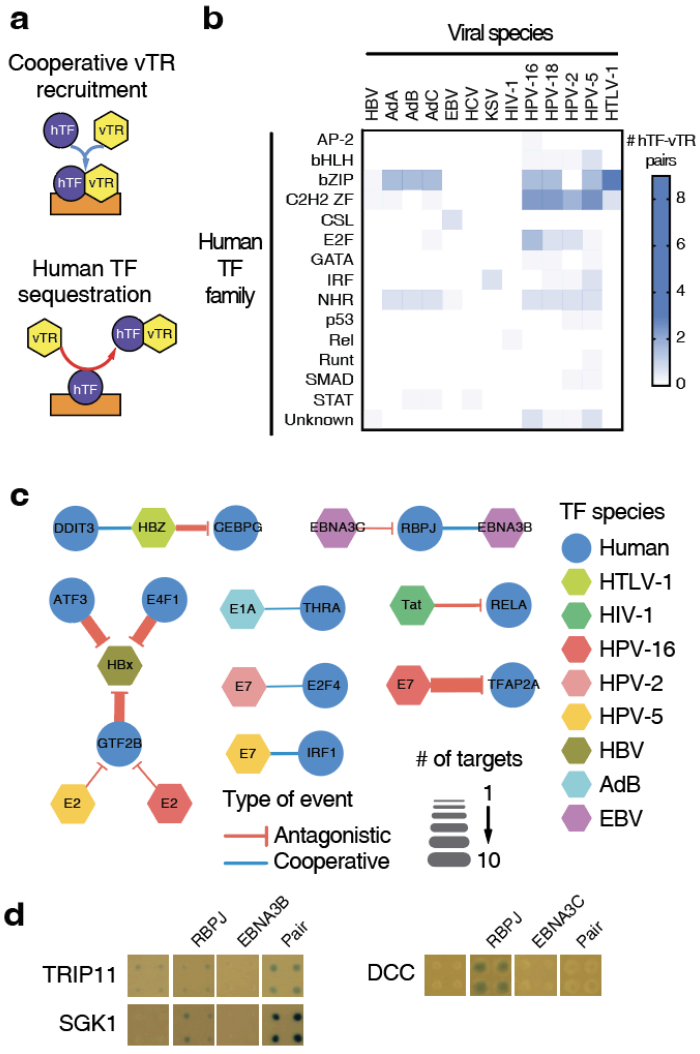


**Figure 5. Application of pY1H assays to study viral transcriptional regulators (vTRs).** (**a**) Examples of models by which vTRs can affect human TF (hTF) binding. vTRs can cooperate with hTFs to bind to DNA elements or a vTR can sequester an hTF, preventing its binding to DNA. (**b**) Number of hTF-vTR pairs tested for binding to 83 cancer gene promoters. hTFs are classified by TF families, while vTRs are classified based on the virus of origin. (**c**) Network of relationships between human TFs and vTRs at 83 cancer gene promoters. (**d**) Examples of RBPJ-EBNA3B cooperative binding to the TRIP11 and SGK1 promoters, and of EBNA3C antagonism of RBPJ binding to the DCC promoter.

## DISCUSSION

In this study, we introduce pY1H assays to identify DNA-binding cooperativity and antagonism across broad arrays of proteins, circumventing limitations often encountered by other approaches such as reliance on known DNA binding motifs, dependence on endogenous protein expression, and chromatin-related confounders. Studies of TF-TF relationships have primarily focused on cooperativity, namely in the context of heterodimer-DNA binding[12,42,43]. However, our work shows that DNA binding antagonism between TFs is equally common and may play an equivalent role in conveying regulatory specificity. Additionally, we observed that both cooperativity and antagonism extend to a wide range of TFs, many of which were not previously thought to function as heterodimers, highlighting the need for TF-wide approaches to identify these types of functional relationships.

Our results also show that DNA binding of a TF depends heavily on the repertoire of TFs and other proteins in the nucleus. While numerous studies have explored the effect of chromatin states on TF binding[44-46], our findings suggest that TF-TF relationships may also contribute to the drastic differences in genome-wide binding patterns of TFs observed across tissues and cell types, and help explain the limited expression correlation often observed between TFs and their target genes[47]. Additionally, we found that isoform variants and viral proteins drastically alter DNA targeting by TFs, which may contribute to differences in TF function across tissues and in certain disease states (e.g., in cancers that alter splicing patterns or during viral infection). Integrating TF-TF relationships observed by pY1H assays with genome-wide mapping of TF-DNA binding in different cellular contexts may better inform machine learning efforts to predict enhancer and promoter activity based on sequence and provide mechanistic insights into gene dysregulation in disease.

pY1H assays identify cooperative and antagonistic interactions in a heterologous context by expressing two TFs at a time. Therefore, orthogonal experiments may be required to determine the specific contexts in which these events occur. However, using a heterologous

assay has the advantage of interrogating the direct effects of DNA sequence on binding patterns of TF-pairs in the absence of other TFs from the same species that could have confounding interactions with the TFs evaluated.

pY1H assays can be used for diverse applications, leveraging both the 1-AD and the 2-AD designs. An immediate advance for this approach would involve expanding the human TF-pair array to incorporate all known and predicted TF-pairs. Pairs of isoforms or mutants of the same TF can also be studied to detect potential functional switches or dominant negative effects between them. pY1H assays can also be applied to study the binding and functional relationships between TFs from non-human species, leveraging existing Gateway-compatible TF clone resources from *Caenorhabditis elegans*[14], *Drosophila melanogaster*[16], *Mus musculus*[48], and *Arabidopsis thaliana*[17]. Additionally, pY1H assays can be used to study interactions involving other proteins within the nucleus, including cofactor or scaffold protein recruitment by TFs, as well as expanded arrays of viral/human and viral/viral protein pairs. In summary, pY1H assays provide widespread evidence of complex functional relationships between TFs and constitute a broadly applicable method for studying occupancy of protein pairs at DNA regions of interest.

**Authors contributions**

A.B. and J.I.F.B. conceived the project. A.B. and R.L. performed the pY1H screens with contributions from Y.C., S.S., C.S., X.L.. A.B., J.I.F.B., L.S., and Z.L. performed data analyses. M.P. and C.C. developed DISHA. K.S., T.H., M.V., D.E.H. provided human TF and isoform ORFs. A.B. and J.I.F.B. wrote the manuscript with contributions from L.S., Z.L., M.P., and C.C. All authors read and approved the manuscript.

## Acknowledgments

This work was funded by the National Institutes of Health grants R35 GM128625 awarded to J.I.F.B and U01 CA232161 awarded to J.I.F.B and M.V.. We thank Dr. Trevor Siggers for critically reading and commenting on the manuscript.

## METHODS

### TF-pair and DNA-bait selection

For our initial pY1H screen, we selected all 6 possible pairs of available NF-κB clones (NFKB1, REL, RELA, and RELB) and all 21 possible pairs of available AP-1 clones (FOS, FOSB, FOSL1, FOSL2, JUN, JUNB, ATF2). Of these 27 pairs, 24 were tested using both the 1-AD and 2-AD screen designs, and 3 were tested only in the 1-AD design (**Supplementary Table 3**). Using the CytReg2.0 database[22], we selected 18 cytokines that have been shown to be regulated by at least one NF-κB subunit and at least one AP-1 subunit (**Supplementary Table 1**). Yeast DNA-bait strains corresponding to the promoters of these cytokines (which were previously generated[22]) were screened against the collection of NF-κB and AP-1 TF-pairs and single-TFs.

For the large-scale TF-pair array, we selected all 429 TF-pairs with protein-protein interactions (PPIs) reported in the LitBM database[28]. We then added all 252 additional TF-pairs with more than two pieces of PPI evidence in the BioGRID database[29]. Finally, we added 187 pairs based on amino acid identity with selected pairs (**See "Predicting possible TF-TF interactions based on homology" below**). This resulted in an initial list of 868 TF-pairs, which we named pTF1.0 (**Supplementary Table 5**). After cloning, yeast transformations, and sequence confirmation, we obtained a final array of 297 TF-pairs for screening (**Supplementary Table 7**). We selected the same 18 cytokine promoters tested in the initial screen to use as DNA-baits (**Supplementary Table 1**).

To study alternative isoforms, we selected TFs with known immune regulatory functions: FOS, MAX, STAT1, STAT3, PPARG, RARG, and RXRG. We studied isoforms for these TFs available from the TFIso1.0 collection from the Center for Cancer Systems Biology (CCSB) at the Dana-Farber Cancer Institute and included a subset of TF partners for these TFs from the TF-pair array. This resulted in a final array of 37 TF isoform-pairs for screening (**Supplementary Table 12**) against 119 cytokine promoters for which DNA-bait yeast strains were previously generated[22] (**Supplementary Table 1**).

To determine cooperativity and antagonism between viral transcriptional regulators (vTRs) and human TFs, we used VirHostNet[37], Uniprot, and primary literature to select pairs of vTRs and human TFs which have been shown to interact via PPIs. We supplemented these with additional vTR-TF pairs based on homology with known pairs to include similar proteins across viruses (e.g., E7 from HPV-2 and E7 from HPV-5). Once filtered for available ORF clones, this resulted in an initial list of 353 protein pairs. After cloning, yeast transformations, and sequence confirmation, we generated a final array of 113 vTR-TF pairs for screening (**Supplementary Table 15**). For DNA-baits, we selected 83 promoters of genes associated with cancer (**Supplementary Table 1**).

**Predicting possible TF-TF interactions based on homology**

PPIs involving human TFs were downloaded from the LitBM database[32]. For all analyses, we considered all 1,639 human TFs reported in the Lambert list[49]. To identify possible TF-TF interactions, we used the following approach:

1. If two TFs ($TF_x$ and $TF_y$) were reported to interact in LitBM; then, each $TF_a$ highly similar to $TF_x$, and each $TF_b$ highly similar to $TF_y$ was considered as new possible pairs of interactors ($TF_x$ and $TF_b$, $TF_a$ and $TF_y$, and $TF_a$ and $TF_b$).
2. To determine the amino acid sequence similarity between TFs, the percent identity was determined using multiple alignments performed using Clustal 2.1[50]. A cutoff of 68.83%

462 was used to identify highly similar TFs, as this corresponds to the 99.9$^{th}$ percentile in the
463 percent identity matrix.
464
465 **Generation of TF-pair prey background yeast strain**
466 pY1H assays require transformation with two TF-prey plasmids. We selected the TRP1 and LEU2
467 as selection markers for these plasmids. Given that the Yα1867 yeast strain used for eY1H assay
468 is *TRP1*- but *LEU2*+, we disrupted the endogenous *LEU2* gene in Yα1867 yeast using the M3926
469 leu2::KanMX3 disruptor converter plasmid with G418 resistance (Addgene). M3926 was digested
470 with BamHI (New England Biolabs) and ethanol precipitated. Yα1867 yeast were transformed
471 with 2 µg digested plasmid as previously described and plated on YAPD-agar with 100 µg/mL
472 G418. We confirmed that Yα1867Δ*leu2* yeast were unable to grow on media lacking leucine.
473
474 **Generation of TF-pair ORF collections and yeast strains**
475 Most human TF ORFs were obtained from ORFeome 8 and 9 collections from the CCSB[28,30-32],
476 while the remaining TF ORFs were obtained from the eY1H human TF ORF collection[51]
477 **(Supplementary Tables 2,6)**. Alternative TF isoform clones were obtained from the TFIso1.0
478 collection from the CCSB (**Supplementary table 11**). vTR ORF clones were synthesized by
479 GeneArt (**Supplementary table 14**). All clones were obtained as Gateway Cloning-compatible
480 entry clones and transferred to the corresponding destination vectors by LR cloning.
481 TF ORFs were cloned into yeast expression vectors using LR Gateway Cloning
482 (ThermoFisher). For each TF-pair, one TF was cloned into the pAD2µ-TRP1 (Walhout lab)
483 plasmid and the other TF was cloned into the pGADT7-GW-LEU2 plasmid (Addgene #61702).
484 Cloned TF-pairs (~250 ng for each TF) were transformed into Yα1867Δleu2 yeast simultaneously,
485 as previously described[51]. Transformed yeast were plated on selective media lacking tryptophan
486 and leucine to select for double transformants.
487

**Generation of DNA-bait yeast strains**

DNA-bait yeast strains were generated as previously described[51] **(Supplementary Table 1).** Promoters of 83 genes with a known association with cancer, incorporating ~2kb upstream of the transcription start site, were amplified from human genomic DNA (Clonetech) using primers with Gateway tails (**Supplementary Table 1**). Promoters were first cloned into the pDONR-P4P1R vector using BP Clonase (ThermoFisher) to generate Gateway entry clones. Sequences were confirmed via Sanger sequencing. Each promoter was then cloned into the pMW#2 (Addgene #13349) and pMW#3 (Addgene #13350) destination vectors using LR Clonase (ThermoFisher), where they were inserted upstream of the HIS3 and LacZ reporter genes, respectively. Destination vectors were linearized with single-cutter restriction enzymes (New England Biolabs). The pWM#2 and pWM#3 plasmids for each promoter were integrated simultaneously into the Y1Has2 yeast genome as previously described[14].

**Sequence confirmation of TF-prey and DNA-bait yeast strains**

TF-pair prey and DNA-bait yeast strains were sequence-confirmed using the SWIM-seq protocol[32]. In brief, yeast were treated with zymolyase (0.2 KU/mL) for 30 min at 37°C followed by 10 min at 95°C to disrupt cell walls and release DNA. TF ORFs and DNA-baits were PCR-amplified in 96-well format using forward primers with well-specific barcodes. For TF-prey, one set of primers was designed so that they targeted both the pAD2μ-TRP1 and pGADT7-GW-LEU2 vectors. See primer design below:

Forward primer (TF-prey):

5’ - AGACGTGTGCTCTTCCGATCT[barcode]TAATACCACTACAATGGATGATGT – 3’

Reverse primer (TF-prey):

5’ – GGAGACTTGACCAAACCTCTGGCG – 3’

Forward primer (DNA-baits, pMW#2):

5’ - AGACGTGTGCTCTTCCGATCT[barcode]GGCCGCCGACTAGTGATA – 3’

514 Reverse primer (DNA-baits, pMW#2):

515 5' – GGGACCACCCTTTAAAGAGA – 3'

516 Forward primer (DNA-baits, pMW#3):

517 5' -AGACGTGTGCTCTTCCGATCT[barcode]GCCAGTGTGCTGGAATTCG – 3'

518 Reverse primer (DNA-baits, pMW#3):

519 5' – ATCTGCCAGTTTGAGGGGAC – 3'

520 PCR reactions were conducted using DreamTaq Polymerase (ThermoFisher) under the
521 following conditions: 95°C for 3 minutes; 35 cycles of: 95°C for 30 seconds, 56°C for 30 seconds,
522 72°C for 4 minutes; final extension at 72°C for 7 minutes.

523 Amplicons from each 96-well plate were pooled and purified using the PCR Purification
524 Kit (ThermoFisher). Each pooled sample was prepared as a single sequencing library by the
525 Molecular Biology Core Facilities at the Dana-Farber Cancer Institute; DNA was sheared using
526 an ultrasonicator (Covaris) prior to tagmentation. Libraries were sequenced using a NovaSeq with
527 ~10 million reads (paired-end, 150bp) per library.

528

529 **Bioinformatics analysis of TF-prey sequencing data**

530 The quality of FASTQ files were assessed using FastQC v.0.11 and MultiQC[52] software.
531 Demultiplexing and trimming of adapters, barcodes and primer sequences were carried out using
532 cutadapt 4.1[53] with the following parameters: -e 0.2 -pair-filter = both -O 10 for pAD2μ; and -e 0.2
533 -pair-filter = both -O 20 for pGADT7 vectors.

534 A FASTA file of the nucleotide sequences of expected TFs, including all possible isoforms,
535 was generated using the package BIOMART[54] in R. First, we obtained the isoform IDs considering
536 "ensembl" as dataset, 'ensembl_gene_id' as filter, and 'ensembl_trancript_id' as attributes. We
537 then used the getSequence() function to obtain the coding sequence for each isoform. The
538 resulting FASTA file was indexed using *bwa index*[55] and alignment was performed using *bwa mem*

with default parameters. Samtools 1.10[56] was used to sort, index, and convert from sam to bam files using parameters by default.

To quantify the number of reads aligned to the expected sequence in each well, we developed an in-house R script primarily based on Rsamtools functions. We considered only those reads that mapped a TF sequence with a primary alignment score greater or equal to 90 % of the trimmed read length, allowing for less than 5% of mismatches. We then determined the number of reads aligning to the expected sequence in each well, considering either the forward or reverse reads, and considered a correct match if the gene with the most aligned reads match the expected gene. Most wells had over 90% of reads aligned to the expected sequence. Additional positions in the arrays were verified by Sanger sequencing.

**pY1H screening**

Screening of TF-pairs and DNA-baits was performed similarly to eY1H screens as previously described[51] using a high-density array ROTOR robot (Singer Instruments). The five-plate TF-pair yeast array and DNA-baits were mated pairwise on permissive media agar plates and incubated at 30°C for one day. Mated yeast were then transferred to selective media agar plates lacking uracil, leucine, and tryptophan to select for successfully mated yeast and incubated at 30°C for two days. These selection plates were imaged and analyzed to identify array locations with failed yeast growth, which were then removed from further analysis. Diploid yeast were finally transferred to selective media agar plates lacking uracil, leucine, tryptophan, and histidine, with 5mM 3AT and 320 mg/L X-gal. Readout plates were imaged 2, 3, 4, and 7 days after final plating.

**Image processing**

To analyze the pY1H images we developed an open-source analyzer called DISHA (Detection of Interactions Software for High-throughput Analyses), in honor of Disha Patel who was very loved

and passed away too soon. DISHA uses classical computer vision algorithms and deep-learning approaches to accelerate the analysis of pY1H readout plates. The overall pipeline of DISHA (**Supplementary Fig. 2**) includes, in this processing order, boundary cropping, grid generation, and colony segmentation algorithms (https://github.com/mahir1010/D.I.S.H.A). The boundary cropping algorithm converts the input image to grayscale and rescales the image intensity (blue color due to β-galactosidase activity) to enhance the yeast colonies from the background. Then an approximate binary mask of the colonies is created using a fixed threshold value. The plate boundary cropping is performed by limiting the region of interest to the first and last white pixel encountered vertically and horizontally in the binary mask. This is followed by the grid generation algorithm to localize the yeast colonies further and assign coordinates to each set of quadruplicate colonies based on a 1,536 colony format (**Supplementary Fig. 2**). An approximate segmentation mask for the colonies is obtained through a sub-optimal subtraction of the plate background performed by a smoothing operation, followed by dynamic contrast stretching and convolving using edge detection kernels. The resulting mask is projected horizontally and vertically (**Supplementary Fig. 2**). The centers of the colonies are detected by zero-crossing analysis of the gradients of the projections (**Supplementary Fig. 2**). Given that equally-spaced pins are used for yeast transfer, we assumed that the colonies are equidistant from each other, and therefore, we can extrapolate the grids based on the centers. A UNet-based segmentation model[57] was trained on our curated yeast segmentation dataset. Briefly, a fixed-size patch was randomly selected from pY1H assay images and generated multiple segmentation maps by varying the parameters of our manual segmentation pipeline. This dataset was curated by manually discarding the incorrect segmentation maps.

The size and intensity of the colony can be considered a proxy for reporter activity and used to determine cooperativity or antagonism between TFs. The area is computed by counting the number of non-zero pixels in a region identified as a colony. The intensity is computed by removing the background pixels from the region of interest and adding all the remaining pixel

intensities. We further normalize this value by the area of the corresponding colony. Then a reporter signal score is calculated as follows that combines both area and intensity metrics of the TF pairs normalized by the metrics from empty-empty pair (neither vector expresses a TF).

$$RS_{TF1-TF2} = [(I - I_{min}) \times A]_{TF1-TF2} - AVG([(I - I_{min}) \times A]_{empty-empty})$$

Here, *I* is the intensity, $I_{min}$ is the minimum non-zero intensity, and A is the area of the colony.

Using this reporter signal we generate three indices: Cooperativity index, Antagonism Index 1, and Antagonism Index 2. They are defined as follows.

$$CooperativityIndex = RS_{TF1-TF2} - RS_{TF1-empty} - RS_{empty-TF2}$$

$$AntagonismIndex_1 = RS_{TF1-empty} - RS_{TF1-TF2}$$

$$AntagonismIndex_2 = RS_{empty-TF2} - RS_{TF1-TF2}$$

DISHA also incorporates a visualization tool to represent the data generated by the analyzer more intuitively (**Supplementary Fig. 3**). This includes a Plate view that shows a segmented plate image where colonies can be selected and filtered by single-TF o TF-pair; and a Table view that displays a colony image comparison for each TF-pair with the corresponding single-TFs as well as area and intensity metrics (https://github.com/mahir1010/D.I.S.H.A-viewer).

**Calling interactions**

TF-pair strains were sorted based on each index (cooperativity, antagonism index 1, and antagonism index 2) separately. Images were then manually analyzed to call cooperative and antagonistic interactions. To call an interaction, we required the following criteria:

1. TF-pair, TF1, and TF2 yeast strains all showed growth in the mating selection plates prior to transfer to readout plates.
2. On readout plates, ≥3 out of 4 quadruplicate colonies were uniform for TF-pair, TF1, and TF2 yeast strains.

3. For cooperative interactions, TF-pair yeast showed a strong or moderate reporter activity relative to the empty-empty strain. TF1 and TF2 yeast showed only weak or very weak reporter activity.
4. For antagonistic interactions, TF1 and/or TF2 yeast showed a strong or moderate reporter activity relative to the empty-empty strain. TF-pair yeast showed only weak or very weak reporter activity.

See **Supplementary Tables 4, 8, 13, and 16** for pY1H results.

**Literature overlap**

Overlap of pY1H interactions with existing literature was determined using the CytReg2.0 database[22]. If CytReg2.0 reported at least one piece of evidence for binding of a TF to a cytokine promoter or regulation of the cytokine by the TF, then the TF-cytokine interaction was considered to be previously reported. To compare with eY1H data, we determined whether the TF had been found to bind the same cytokine promoter DNA-bait sequence tested in both eY1H and pY1H assays. Results from eY1H and pY1H assays were both compared to CytReg2.0 data after removing eY1H interactions already reported in CytReg2.0.

**Overlap between ChIP-seq and pY1H interactions**

The ChIP-seq peaks mapping to the cytokine promoter sequences tested by pY1H assays were obtained from GTRD database[58] considering the following filters: peaks calling = MACS2, reference genome = hg38, format file = bigBeds. A TF was considered to be binding a cytokine promoter if the summit point of any significant peak ($p - value \leq 10^{-4}$) was located within the promoter's genomic coordinates. The output was a table showing the peak of the TF, its genomic coordinates, and the cell line used. TF-pairs detected by pY1H assays for which ChIP-seq data was available for both TFs were further considered. For each TF-pair interaction with a cytokine promoter, evidence for co-binding was considered when both TFs had ChIP-seq peaks within the

corresponding promoter, either in different or the same cell line, and the peak summits were within 50 bp of each other.

**Identification of binding sites of TF-pairs in cytokine promoters**

Position Weight matrix (PWM) motifs were downloaded from CISBP 2.0 database[59] for each TF. PWM motifs with all sites probabilities lower than 0.8 were removed to reduce low-specific motifs. To determine if a PWM motif was present within a promoter sequence, we calculated the sum of log odds for each position in each promoter using the following formula:

$$Score(s, PWM) = \sum_{t=0}^{|s|-k} \prod_{i=1}^{k} \left(\frac{PWM_i[s_{t+i}]}{p_i}\right)$$

Where i = 1,2,3,4 corresponding to {A,T,C,G}, $p_i$ is the background frequency of such nucleotide, which is 0.25. k = length of the PWM, |s| = length of the sequence. Each score was converted to a p-value using the TFMsc2pv function from the TFMPvalue package[60]. Motifs were filtered considering a $p - value \leq 10^{-4}$. As many motifs for the same TF were very similar, we merged all motifs for a TF that overlapped with each other using the following steps:

1. Consecutive motif for a TF within a DNA-bait sequence that shared 80% or more nucleotides were labeled into the same group.
2. For each group of overlapping ‘n’ motifs within a DNA-bait, we selected the sub-region corresponding to the intersection between all n motifs, only if this sub-region was four nucleotides or longer and named this as ‘core motif’.
3. If the intersection region was shorter than four nucleotides, we repeated the process by taking the intersection region shared by ‘n-1’ motifs.

This algorithm produces a set of non-overlapping core motifs of a TF within DNA-bait sequences. We manually reviewed the final list of core motifs to ensure that it was unique and did not overlap with others. To compare with pY1H interactions, a TF-pair was considered to

665 potentially binding a DNA-bait if a core motif for each single-TF was present in the DNA-bait within
666 10 nt of each other.
667
668 **Network randomization analysis**
669 The significance of overlap between TF-pairs determined by pY1H assays and those presenting
670 ChIP-seq peaks within the same promoter was evaluated by a network randomization analysis.
671 First, we built a directed network graph where the source node was ($TF_1$-$TF_2$), and the target
672 node was cytokine promoter used in the pY1H screen. Then, 10,000 networks were generated by
673 performing 20,000 edges-switches while maintaining the same degree for each node[61] using the
674 igraph package in R.
675 For the original pY1H network and each of the randomized networks, we determined the
676 number of edges overlapping with the ChIP-seq data. Based on the 10,000 random networks
677 generated, a Z distribution was used to obtain a Z-scores and p-values for the original pY1H
678 network. This analysis was performed considering: (1) ChIP-seq peaks found in the same cell
679 line, and (2) ChIP-seq peaks found in different cell lines.
680 A similar randomization analysis was performed to compare pY1H interactions with TF
681 motifs found in the corresponding cytokine promoters. We evaluated the significance of detecting
682 binding sites for both TFs anywhere in the promoters and within 10 bp from each other.
683
684 **Data visualization and statistical analyses**
685 Network visualizations were constructed using Cytoscape Version 3.9.1. Scatter plots, violin
686 plots, histograms, bar graphs, and heat maps were generated using GraphPad Prism Version 9.
687
688 **Paralog partner similarity**
689 TFs were classified based on their DBD family, as reported in Lambert *et al.*[49]. A pairwise
690 alignment was performed using the BLOSUM62 matrix from the package seqinr, and the amino

acid identity score was assigned to each pair of TFs from the same TF family. To determine if TFs with greater amino acid identity have similar functional relationships (antagonism and cooperativity) with their shared TF interactors tested by pY1H, we calculated the Jaccard similarity index as follows:

1. For a pair of TFs ($TF_a$, $TF_b$), we obtained the list of TF partners that were both tested by pY1H assays.
2. For each $TF_a$,, we generated a binary vector ($P_{1,c}$ , $P_{1,a}$ , $P_{2,c}$ , $P_{2,a}$ ,…), where $P_{i,c}$ indicates whether partner i has at least one cooperative interaction involving $TF_a$, (true = 1, false = 0), and where $P_{i,a}$ indicates whether partner i has at least one antagonistic interaction involving $TF_a$,.
3. Then the Jaccard index was determined as the number positions with 1 in both $TF_a$ and $TF_b$ vectors divided by the number of positions with a 1 in either $TF_a$ and $TF_b$ vectors.

The Jaccard score ranged from 0 to 1, where 1 indicate both TFs ($TF_a$, $TF_b$) have the same functional relationships with the same partners and 0 indicates both TFs have completely different functional relationships with their shared partners.

The percent amino acid identity was classified in three groups: Low identity (< 30%), Medium identity (30-50%) and high identity (>50%). A Mann-Whitney's U test was performed to evaluate significant differences between groups regarding paralog partner similarity based on the Jaccard index.

**TF expression analysis**

The Single Cell RNA-Seq data was obtained from the Tabula Sapiens atlas[33] **(Supplementary Table 9)**. To avoid technical confounding factors, only samples that were generated by 10X Genomics protocols were used. After obtaining the data, cells with no less than 500 genes, no more than 7,500 genes, no more than 10,000 UMIs, and no more than 25% mitochondrial

716 contents were kept for the downstream analyses. The normalized counts per cell were
717 generated by dividing the gene counts per cell by the total number of UMIs per cell and then
718 multiplied by 1,000,000, to determine the counts per million (cpm). After long normalizing the
719 cpms and conducting a principal component analysis, Harmony[62] was used to remove batch
720 effects. Then, the k-nearest neighbor graph was constructed between cells and the Louvain
721 community clustering was used to cluster cells based on the constructed graph. A total of 187
722 clusters across samples were identified. All the steps above were performed by Seurat in R
723 environment[63]. Differential expression analyses (Wilcoxon ranked sum test) were performed
724 between clusters to identify the genes that were significantly up regulated in each cluster. The
725 genes with false discovery rates smaller than 0.05 were used to compare with the gene markers
726 curated in the CellTypist[64] database to assign cell types to clusters in each sample.
727
728 **Tissue/cell type expression specificity scoring of genes**
729 To study the gene expression specificity among cell types and tissues, a tissue/cell type
730 expression specificity score (TCESS) was calculated for each TF adapting a previously entropy-
731 based approach to single-cell RNA-seq data[65] **(Supplementary Table 10)**. Briefly, given a
732 cluster C, which had n cells, the total expression of $TF_a$ was calculated using the following
733 formula:

734 $$Exp_{TF_a}^{C} = \left( \sum_{Cell \in C}^{Gene=TF_a} exp_{Gene}^{Cell} \right) + 1$$

735 Then the TCESS was calculated as:

736 $$TCESS = \sum_{TF_a}^{C \in dataset} \left( \frac{Exp_{TF_a}^{C}}{sum\left(Exp_{TF_a}^{C}\right)} \right) * log_2 \left( \frac{Exp_{TF_a}^{C} / sum\left(Exp_{TF_a}^{C}\right)}{mean\left(Exp_{TF_a}^{C} / sum\left(Exp_{TF_a}^{C}\right)\right)} \right)$$

737 The TCESS ranges from 0 when $TF_a$ expression is identical across all clusters to
738 log2(#clusters), in this case ~7.54, when $TF_a$ is expressed exclusively in one cluster.

739

740 **Transcription factors co-expression among tissue/cell types**

741 To study the co-expression patterns of pairs of TFs across cell types/tissues, a scoring system
742 based on the Simpson Index was developed[66]. In a given cell type/tissue cluster, if the cpms of
743 a given TF in the cluster was higher than 10% of the maximum cpms for the TF across all
744 clusters, the TF was considered 'expressed' in the given cluster. For example, if the $TF_a$ in a
745 cluster B is 1.2 cpms, and the maximum expression of $TF_a$ across all clusters is 10 cpms, then
746 $TF_a$ is considered to be expressed in cluster B. Then, for each $TF_a$, we generated a binary
747 vector indicating whether $TF_a$ was expressed in each of the 187 cell clusters. Finally, for every
748 pair of TFs we determined the co-expression score using the Simpson index, by dividing the
749 number of clusters expressing both TFs by the number of cluster where the most tissue specific
750 TF is expressed.

751 **REFERENCES**

752

753 1 Inukai, S., Kock, K. H. & Bulyk, M. L. Transcription factor-DNA binding: beyond binding
754 site motifs. *Curr Opin Genet Dev* **43**, 110-119, doi:10.1016/j.gde.2017.02.007 (2017).
755 2 Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif
756 syntax. *Nat Genet* **53**, 354-366, doi:10.1038/s41588-021-00782-6 (2021).
757 3 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from
758 ENCODE data. *Nature* **489**, 91-100, doi:10.1038/nature11245 (2012).
759 4 Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental
760 control. *Nat Rev Genet* **13**, 613-626, doi:10.1038/nrg3207 (2012).
761 5 Zia, A. & Moses, A. M. Towards a theoretical understanding of false positives in DNA
762 motif finding. *BMC Bioinformatics* **13**, 151, doi:10.1186/1471-2105-13-151 (2012).
763 6 Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding
764 specificity. *Nature* **527**, 384-388, doi:10.1038/nature15518 (2015).
765 7 Siggers, T. *et al.* Principles of dimer-specific gene regulation revealed by a
766 comprehensive characterization of NF-κB family DNA binding. *Nat Immunol* **13**, 95-102,
767 doi:10.1038/ni.2151 (2011).
768 8 Sönmezer, C. *et al.* Molecular Co-occupancy Identifies Transcription Factor Binding
769 Cooperativity In Vivo. *Mol Cell* **81**, 255-267.e256, doi:10.1016/j.molcel.2020.11.015
770 (2021).
771 9 Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S. & Krebs, A. R. Single-molecule
772 footprinting identifies context-dependent regulation of enhancers by DNA methylation.
773 *Mol Cell* **83**, 787-802.e789, doi:10.1016/j.molcel.2023.01.017 (2023).
774 10 Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev*
775 *Genet* **10**, 669-680, doi:10.1038/nrg2641 (2009).

11 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, doi:10.7554/eLife.21856 (2017).

12 Karczewski, K. J. *et al.* Cooperative transcription factor associations discovered using regulatory variation. *Proc Natl Acad Sci U S A* **108**, 13353-13358, doi:10.1073/pnas.1103105108 (2011).

13 Hu, Z., Killion, P. J. & Iyer, V. R. Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* **39**, 683-687, doi:10.1038/ng2012 (2007).

14 Reece-Hoyes, J. S. *et al.* Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat Methods* **8**, 1059-1064, doi:10.1038/nmeth.1748 (2011).

15 Fuxman Bass, J. I. *et al.* Human gene-centered transcription factor networks for enhancers and disease variants. *Cell* **161**, 661-673, doi:10.1016/j.cell.2015.03.003 (2015).

16 Hens, K. *et al.* Automated protein-DNA interaction screening of Drosophila regulatory elements. *Nat Methods* **8**, 1065-1070, doi:10.1038/nmeth.1763 (2011).

17 Gaudinier, A. *et al.* Enhanced Y1H assays for Arabidopsis. *Nat Methods* **8**, 1053-1055, doi:10.1038/nmeth.1750 (2011).

18 Sewell, J. A. & Fuxman Bass, J. I. Options and Considerations When Using a Yeast One-Hybrid System. *Methods Mol Biol* **1794**, 119-130, doi:10.1007/978-1-4939-7871-7_8 (2018).

19 Berenson, A. & Fuxman Bass, J. I. Enhanced Yeast One-Hybrid Assays to Study Protein-DNA Interactions. *Methods Mol Biol* **2599**, 11-20, doi:10.1007/978-1-0716-2847-8_2 (2023).

20 Oeckinghaus, A. & Ghosh, S. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harb Perspect Biol* **1**, a000034, doi:10.1101/cshperspect.a000034 (2009).

21 Karin, M., Liu, Z. & Zandi, E. AP-1 function and regulation. *Curr Opin Cell Biol* **9**, 240-246, doi:10.1016/s0955-0674(97)80068-3 (1997).

22 Santoso, C. S. *et al.* Comprehensive mapping of the human cytokine gene regulatory network. *Nucleic Acids Res* **48**, 12055-12073, doi:10.1093/nar/gkaa1055 (2020).

23 Funnell, A. P. & Crossley, M. Homo- and heterodimerization in transcriptional regulation. *Adv Exp Med Biol* **747**, 105-121, doi:10.1007/978-1-4614-3229-6_7 (2012).

24 Potoyan, D. A., Bueno, C., Zheng, W., Komives, E. A. & Wolynes, P. G. Resolving the NFκB Heterodimer Binding Paradox: Strain and Frustration Guide the Binding of Dimeric Transcription Factors. *J Am Chem Soc* **139**, 18558-18566, doi:10.1021/jacs.7b08741 (2017).

25 Rodríguez-Martínez, J. A., Reinke, A. W., Bhimsaria, D., Keating, A. E. & Ansari, A. Z. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *Elife* **6**, doi:10.7554/eLife.19272 (2017).

26 Hoffmann, A. & Baltimore, D. Circuitry of nuclear factor kappaB signaling. *Immunol Rev* **210**, 171-186, doi:10.1111/j.0105-2896.2006.00375.x (2006).

27 Moorthy, A. K., Huang, D. B., Wang, V. Y., Vu, D. & Ghosh, G. X-ray structure of a NF-kappaB p50/RelB/DNA complex reveals assembly of multiple dimers on tandem kappaB sites. *J Mol Biol* **373**, 723-734, doi:10.1016/j.jmb.2007.08.039 (2007).

28 Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212-1226, doi:10.1016/j.cell.2014.10.050 (2014).

29 Oughtred, R. *et al.* The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* **30**, 187-200, doi:10.1002/pro.3978 (2021).

30 Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat Methods* **8**, 659-661, doi:10.1038/nmeth.1638 (2011).

31 The ORFeome Collaboration: a genome-scale human ORF-clone resource. *Nat Methods* **13**, 191-192, doi:10.1038/nmeth.3776 (2016).

32 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402-408, doi:10.1038/s41586-020-2188-x (2020).

33 Jones, R. C. *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896, doi:10.1126/science.abl4896 (2022).

34 Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37-47, doi:10.1016/j.cell.2006.06.023 (2006).

35 Joung, J. *et al.* A transcription factor atlas of directed differentiation. *Cell* **186**, 209-229.e226, doi:10.1016/j.cell.2022.11.026 (2023).

36 Liu, X. *et al.* Human Virus Transcriptional Regulators. *Cell* **182**, 24-37, doi:10.1016/j.cell.2020.06.023 (2020).

37 Guirimand, T., Delmotte, S. & Navratil, V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res* **43**, D583-587, doi:10.1093/nar/gku1121 (2015).

38 Calderone, A., Licata, L. & Cesareni, G. VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res* **43**, D588-592, doi:10.1093/nar/gku830 (2015).

39 Robertson, E. S., Lin, J. & Kieff, E. The amino-terminal domains of Epstein-Barr virus nuclear proteins 3A, 3B, and 3C interact with RBPJ(kappa). *J Virol* **70**, 3068-3074, doi:10.1128/jvi.70.5.3068-3074.1996 (1996).

40 Wang, A. *et al.* Epstein-Barr Virus Nuclear Antigen 3 (EBNA3) Proteins Regulate EBNA2 Binding to Distinct RBPJ Genomic Sites. *J Virol* **90**, 2906-2919, doi:10.1128/jvi.02737-15 (2015).

41 Kalchschmidt, J. S. *et al.* EBNA3C Directs Recruitment of RBPJ (CBF1) to Chromatin during the Process of Gene Repression in EBV Infected B Cells. *PLoS Pathog* **12**, e1005383, doi:10.1371/journal.ppat.1005383 (2016).

42 Ibarra, I. L. *et al.* Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nat Commun* **11**, 124, doi:10.1038/s41467-019-13888-7 (2020).

43 Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-263, doi:10.1038/nrg2538 (2009).

44 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).

45 Coux, R. X., Owens, N. D. L. & Navarro, P. Chromatin accessibility and transcription factor binding through the perspective of mitosis. *Transcription* **11**, 236-240, doi:10.1080/21541264.2020.1825907 (2020).

46 Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76-81, doi:10.1038/s41586-018-0549-5 (2018).

47 Zaborowski, A. B. & Walther, D. Determinants of correlated expression of transcription factors and their target genes. *Nucleic Acids Res* **48**, 11347-11369, doi:10.1093/nar/gkaa927 (2020).

48 Gubelmann, C. *et al.* A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks. *Mol Syst Biol* **9**, 682, doi:10.1038/msb.2013.38 (2013).

49 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665, doi:10.1016/j.cell.2018.01.029 (2018).

50 Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145, doi:10.1002/pro.3290 (2018).

51 Reece-Hoyes, J. S. *et al.* Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. *Nat Methods* **8**, 1050-1052, doi:10.1038/nmeth.1764 (2011).

52 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).

53 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

54 Smedley, D. *et al.* BioMart--biological queries made easy. *BMC Genomics* **10**, 22, doi:10.1186/1471-2164-10-22 (2009).

55 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

56 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

57 Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. doi:https://doi.org/10.48550/arXiv.1505.04597 (2015).

58 Kolmykov, S. *et al.* GTRD: an integrated view of transcription regulation. *Nucleic Acids Res* **49**, D104-d111, doi:10.1093/nar/gkaa1057 (2021).

59 Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443, doi:10.1016/j.cell.2014.08.009 (2014).

60 Touzet, H. & Varré, J. S. Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms Mol Biol* **2**, 15, doi:10.1186/1748-7188-2-15 (2007).

61 Martinez, N. J. *et al.* A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22**, 2535-2549, doi:10.1101/gad.1678608 (2008).

62 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

63 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

64 Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197, doi:10.1126/science.abl5197 (2022).

65 Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744-752, doi:10.1016/j.cell.2010.01.044 (2010).

66 Fuxman Bass, J. I. *et al.* Using networks to measure similarity between genes: association index selection. *Nat Methods* **10**, 1169-1176, doi:10.1038/nmeth.2728 (2013).