

A large-scale cancer-specific protein-DNA interaction network

Yunwei Lu¹†, Anna Berenson^{1,2}†, Ryan Lane¹†, Isabelle Guelin¹, Zhaorong Li³, Yilin Chen¹, Sakshi Shah¹, Meimei Yin¹, Luis Fernando Soto-Ugaldí⁴, Ana Fiszbein^{1,2,3}, Juan Ignacio Fuxman Bass^{1,2,3,*}

¹ Biology Department, Boston University, Boston, MA, 02215, USA

² Molecular Biology, Cellular Biology and Biochemistry Program, Boston University, Boston, MA, 02215, USA

³ Bioinformatics Program, Boston University, Boston, MA, 02215, USA

⁴ Tri-Institutional Program in Computational Biology and Medicine, New York, NY, 10021, USA

† These authors contributed equally to this work

* To whom correspondence should be addressed. Tel: 617-353-2448; Email: fuxman@bu.edu

Abstract

Cancer development and progression are generally associated with dysregulation of gene expression, often resulting from changes in transcription factor (TF) sequence or expression. Identifying key TFs involved in cancer gene regulation provides a framework for potential new therapeutics. This study presents a large-scale cancer gene TF-DNA interaction network as well as an extensive promoter clone resource for future studies. Most highly connected TFs do not show a preference for binding to promoters of genes associated with either good or poor cancer prognosis, suggesting that emerging strategies aimed at shifting gene expression balance between these two prognostic groups may be inherently complex. However, we identified potential for oncogene targeted therapeutics, with half of the tested oncogenes being potentially repressed by influencing specific activator or bifunctional TFs. Finally, we investigate the role of intrinsically disordered regions within the key cancer-related TF estrogen receptor α (ESR1) on DNA binding and transcriptional activity, and found that these regions can have complex trade-offs in TF function. Altogether, our study not only broadens our knowledge of TFs involved in the cancer gene regulatory network but also provides a valuable resource for future studies, laying a foundation for potential therapeutic strategies targeting TFs in cancer.

Introduction

Gene expression is often dysregulated in cancer due to changes in copy number, mutation or epigenetic changes in promoter and enhancer regions, or changes in the expression or activity of transcription factors (TFs) and chromatin modifying enzymes (1,2). Among the affected genes are those involved in cell differentiation, proliferation, apoptosis, DNA repair, immune regulation, and general biological processes such as translation and RNA processing, ultimately contributing to cancer development, progression, and metastasis (3,4).

Higher expression of certain genes has been associated with good or poor cancer prognosis (5). Some of these genes are associated with prognosis only in specific cancers, while others have the same or opposing associations in different cancer types. For instance, elevated expression of *GNAS* has been found to promote cell proliferation in breast cancer (6), while reduced expression of *CAMTA1* has been linked to adverse outcomes in neuroblastoma patients (7). Therefore, a promising potential cancer therapeutic strategy could consist of shifting the balance in expression between poor and good prognosis genes, which may eventually lead to increased cancer survival. The rational design of this strategy involves identifying TFs that preferentially regulate the expression of either poor or good prognosis genes. This requires the delineation of large-scale gene regulatory networks that evaluate the binding of hundreds of TFs to the regulatory elements of cancer-related genes.

Multiple experimental methods have been developed to identify TF-DNA interactions. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) and CUT&RUN are widely used to identify the genomic DNA regions that a TF binds *in vivo*. While these methods have provided extensive datasets, in particular by large consortia such as the ENCODE Project, lowly-expressed TFs and TFs for which ChIP-grade antibodies are not available remain understudied (8). Enhanced yeast one-hybrid (eY1H) assays provide a high-throughput complementary gene-centered method to identify the repertoire TFs that bind to DNA regions of interest by

testing >1,000 TFs simultaneously (9-11). In eY1H assays, each TF is fused to the yeast Gal4 activation domain and expressed in a separate yeast strain; binding of the TF to a DNA region of interest induces expression of two reporter genes, *HIS3* and *LacZ*, allowing yeast to grow and turn blue on readout plates. Given that eY1H assays involve expressing exogenous human TFs in yeast, they can detect interactions involving TFs that have low endogenous expression or that lack suitable antibodies. Although eY1H assays cannot test binding of heterodimeric TFs, we have recently addressed this limitation by developing paired yeast one-hybrid (pY1H) assays. This method evaluates pairs of TFs to detect cooperative binding and antagonism at DNA regions of interest (12).

In this study, we generated a clone resource of 700 of cancer-related gene promoters and used both eY1H and pY1H platforms to examine binding of monomeric/homodimeric and heterodimeric TFs to the promoters of 136 cancer-related genes. We identified 1,350 interactions between 265 TFs and the promoters of 108 cancer genes, and leveraged our promoter library to study disordered regions in the breast cancer-related TF estrogen receptor α (ESR1). Overall, our work provides new insights into the study of the regulation of cancer genes and provides a clone and data resource for the scientific community.

Material & Methods

Generation of entry clones and integrant yeast strains for cancer-related gene promoters

Entry clones and yeast strains for promoters of cancer-related genes were generated following established procedures (11,13). 700 promoters corresponding to 556 genes known to be linked with cancer, encompassing approximately 2 kb upstream of the transcription start site, were amplified from human genomic DNA (Clontech) using primers flanked with Gateway tails (Supplementary Table 1). These genes were selected from the Cancer Gene Census, as well as additional genes whose expression is often dysregulated in cancer (14). Alternative promoters were also included in cases where they show high levels of the activating histone mark H3K27ac.

Promoters were cloned into the pDONR-P4P1R vector using BP Clonase (ThermoFisher #11789100), resulting in a collection of Gateway entry clones whose sequences were verified through Sanger sequencing (Sequegen). Subsequently, each promoter was transferred to the pMW#2 (Addgene #13349) and pMW#3 (Addgene #13350) destination vectors using LR Clonase (ThermoFisher #11791100), positioning them upstream of the HIS3 and LacZ reporter genes, respectively. Destination vectors were linearized using single-cutter restriction enzymes (New England Biolabs R0520L, R0146L, R3127S, R0581S, R0193L, R0114S, R0187S, R0519L).

The pWM#2 and pWM#3 plasmids for each cancer promoter were integrated simultaneously into the Y1Has2 yeast genome, as previously outlined (13,15) and as described below. Yeast were cultured in 1 L liquid YAPD media at 30°C with shaking at 200 rpm until reaching OD600 = 0.5, followed by washing with sterile water and 1X TE + 0.1 M lithium acetate (TE/LiAc). Yeast were resuspended in TE/LiAc with salmon sperm DNA (ThermoFisher 15632011) at a dilution of 1:10, and 2 µg of each digested plasmid (pWM#2 and pWM#3) were added. Six volumes of TE/LiAc + 40% polyethylene glycol were added and gently mixed ten times, followed by a first incubation at 30°C for 30 minutes and a second incubation at 42°C for 20 min. The yeast were then resuspended in sterile water and plated on selective media lacking histidine and uracil to select for double integrants.

Sequence confirmation of cancer gene promoter yeast strains

Cancer gene promoter yeast strains were sequence-confirmed using the SWIM-seq protocol (16). In brief, yeast were treated with zymolyase (0.2 KU/mL) (United States Biological Z1004) for 30 min at 37°C followed by 10 min at 95°C to disrupt cell walls and release DNA. Promoter sequences were PCR-amplified in 96-well format using forward primers with well-specific barcodes. See primer design below:

Forward primer (pMW#2):

5'—AGACGTGTGCTCTTCCGATCT[barcode]GGCCGCCGACTAGTGATA—3'

Reverse primer (pMW#2):

5'—GGGACCACCCTTTAAAGAGA—3'

Forward primer (pMW#3):

5'—AGACGTGTGCTCTTCCGATCT[barcode]GCCAGTGTGCTGGAATTCG—3'

Reverse primer (pMW#3):

5'—ATCTGCCAGTTTGAGGGGAC—3'

PCR reactions were conducted using DreamTaq Polymerase (ThermoFisher EP0705) under the following conditions: 95°C for 3 min; 35 cycles of: 95°C for 30 s, 56°C for 30 s, 72°C for 4 min; final extension at 72°C for 7 min. Amplicons from each 96-well plate were pooled and purified using the PCR Purification Kit (ThermoFisher K310002). Each pooled sample was prepared as a single sequencing library by the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute; DNA was sheared using an ultrasonicator (Covaris) prior to tagmentation. Libraries were sequenced using a NovaSeq with ~10 million reads (paired-end, 150 bp) per library. For a promoter yeast strain to be confirmed, we required at least 25% of sequencing reads for the pMW#3 vector to align to the expected promoter sequence. Sequencing data can be found at the NCBI Sequence Read Archive at accession number PRJNA1015222.

eY1H screening

We performed eY1H assays using a human TF yeast array (15) as previously described and as follows using a high-density array ROTOR robot (Singer Instruments). The three-plate human TF

yeast array and promoter yeast strains were mated pairwise on permissive media agar plates and incubated at 30°C for 1 day. Mated yeast were then transferred to selective media agar plates lacking uracil and tryptophan to select for successfully mated yeast and incubated at 30°C for 2 days. Diploid yeast were finally transferred to selective media agar plates lacking uracil, tryptophan, and histidine, with 5 mM 3AT and 320 mg/L X-gal. Readout plates were imaged 2, 3, 4, and 7 days after final plating. Results are reported in Supplementary Table 2.

pY1H screening

We performed pY1H assays using a previously generated TF-pair array (12). Screening of TF-pairs and cancer gene promoters was performed similarly to eY1H screens as previously described and as follows using a high-density array ROTOR robot (Singer Instruments). The five-plate TF-pair yeast array and promoter yeast strains were mated pairwise on permissive media agar plates and incubated at 30°C for 1 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 2 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 5 mM 3AT and 320 mg/L X-gal. Readout plates were imaged 2, 3, 4, and 7 days after final plating. Results are reported in Supplementary Tables 2 and 3.

Identifying cooperative and antagonistic interactions

Yeast plate images were processed and visualized using the DISHA (Detection of Interactions Software for High-throughput Analyses) software as previously described (12). TF-pair strains were sorted based on each index (cooperativity, antagonism index 1, and antagonism index 2) separately. Images were then manually analyzed to identify 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 no or only 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 no or only weak reporter activity.

Literature and ChIP-seq evidence for interactions detected by eY1H and pY1H assays

Literature evidence for eY1H- and pY1H-derived interactions was determined by performing searches in the PubMed database. If there was at least one piece of experimental evidence indicating the binding or regulation of the TF to the cancer promoter or regulation of the cancer gene, then the TF-gene interaction was considered to be previously reported. Results are reported in Supplementary Table 2.

ChIP-Seq data were downloaded from GTRD Database (17) in MACS2 (18) peak calling results format. If a peak was called in ChIP-Seq data for a given TF and the center of the peak was within the corresponding promoter region, the TF was considered to bind the promoter. Results are reported in Supplementary Table 2. Code for this analysis is available at <https://zenodo.org/records/10558771>.

TF and cancer gene survival analysis

RNA-Seq data associated with clinical data from 33 tumor types was downloaded from TCGA and organized using TCGAbiolinks (19). Expression data was then normalized using the counts-per-million (CPM) method and log₂ transformed. To determine whether the expression levels of cancer-related genes and TFs in our eY1H-derived network were associated with good or poor

prognosis, survival analyses were conducted using the normalized RNA-Seq data for each tumor type. First, the Cox Proportional-Hazards Model was used to test whether the high/low expression level of the gene or TF will impact survival significantly (adjusted p value < 0.05); the hazard ratio from the Cox Proportional-Hazards Model indicates whether the high or low expression of the gene leads to good or poor prognosis. An ANOVA analysis was then used to regress out the confounding factors of age, gender, race, tumor size, tumor metastasis and tumor stage. All the survival analyses were performed using the survival package (20). Results are reported in Supplementary Table 4. Code for this analysis is available at <https://zenodo.org/records/10558771>.

Determining mutation incidence of TFs

The COSMIC database (21) was used to determine the number of cancer cases in which mutations have been observed for each TF. The mutation frequency was calculated as the total number of cases with mutations minus the number of synonymous mutations, divided by the total number of all cases. Information can be found in Supplementary Table 5.

ESR1 DNA constructs

The COSMIC database (21) was used to identify mutations in the ESR1 IDRs occurring in breast cancer patients as well as across cancers. Mutations occurring in at least two patients were selected to be tested by eY1H and mammalian one-hybrid (M1H) assays. Deletion constructs were selected to cover portions of the N- and C-terminal IDRs to better identify regions that affect DNA binding and transcriptional activity. The hinge region was replaced by a flexible linker consisting of 23 glycine-serine repeats to maintain the length and flexibility of this region. All ESR1 constructs were ordered from GenScript in pUC57 vectors. The ESR1 sequences were flanked by attB1 and attB2 sequences for Gateway cloning into the pDONR221 entry vector as well as

destination vectors for eY1H and M1H assays. Information and sequences for ESR1 constructs can be found in Supplementary Table 6.

Generation of DB-pEZY3 and 4xUAS-pGL4.23 vectors for M1H assays

The DB-pEZY3 and 4xUAS-pGL4.23 vectors were generated for M1H assays. To generate the DB-pEZY3 vector, the coding sequence for the yeast Gal4 DNA binding domain (DBD) was cloned into the pEZY3 mammalian expression vector upstream of the insert region. Proteins cloned into the insert region are therefore expressed with the Gal4 DBD fused to the N-terminus. To generate the 4xUAS-pGL4.23 vector, four copies of the yeast upstream activating sequence (UAS) site were cloned into the pGL4.23 vector upstream of the minimal promoter and firefly luciferase reporter gene. The UAS site is recognized by the Gal4 DBD, and therefore recruits any protein expressed as a fusion with the Gal4 DBD.

Cloning of ESR1 constructs

ESR1 constructs were cloned into the pDONR221 entry vector using BP Clonase (ThermoFisher #11789100) and verified by whole plasmid sequencing (Plasmidsaurus) and Sanger sequencing (Genewiz) to confirm the proper mutant insertion and discard clones with additional unwanted mutations. Confirmed entry vectors were cloned into DB-pEZY3 and pAD2 μ (Walhout lab) destination vectors using LR clonase (ThermoFisher #11791100). Plasmid samples were prepared using the Endotoxin-Free Miniprep Kit (101 BIO #W210650) following the supplier's protocol.

eY1H screening of ESR1 constructs

We performed eY1H assays as follows using a high-density array ROTOR robot (Singer Instruments). Yeast strains were arrayed such that all 18 ESR1 constructs and two empty control yeast strains were tested against 10 different promoters in each 1,536-colony agar plate. ESR1

construct yeast strains and all 508 cancer gene promoter yeast strains were mated pairwise on permissive media agar plates and incubated at 30°C for 1 day. Mated yeast were then transferred to selective media agar plates lacking uracil and tryptophan to select for successfully mated yeast and incubated at 30°C for 2 days. Diploid yeast were finally transferred to selective media agar plates lacking uracil, tryptophan, and histidine, with 5 mM 3AT and 320 mg/L X-gal. Readout plates were imaged 2, 3, 4, and 7 days after final plating. Results are reported in Supplementary Table 7.

Mammalian one-hybrid (M1H) assays of ESR1 constructs

M1H assays were conducted in HEK293T cells (ATCC #CRL-11268) to identify transcription activating or repressing functions of our ESR1 constructs. Cells were cultured in DMEM (Gibco #11965118) with 10% fetal bovine serum (Bio-Techne #S12450H) and 1% antibiotic-antimycotic (Gibco #15240062) at 37°C with 5% CO₂. Cells were plated at a density of ~10,000 cells/well in 96-well white opaque sterile plates (Falcon #25382-208) with growth media and incubated for 24 hours. Cells were transfected with Lipofectamine 3000 (Invitrogen #L3000001) following the manufacturer's protocol using 80ng DB-pEZY3 vector with a cloned ESR1 construct, 60ng 4xUAS-pGL4.23 vector, and 10ng renilla-pGL4.74 vector. Three biological replicates were performed for each construct, and an empty DB-pEZY3 vector with no cloned ESR1 construct was used as a negative control. Cells were incubated for 6 hours, treated with 100nM estradiol, and incubated for an additional 18 hours.

Luciferase assays were performed using the Dual-Glo Luciferase Assay System (Promega #E2940) following the manufacturer's protocol. Luminescence was measured on a Victor3 multilabel reader (PerkinElmer #1420) using renilla and firefly filters. Background signal from untransfected cells was subtracted from each renilla and firefly measurement. Firefly/renilla ratios for each sample were normalized to the average ratio for negative control samples transfected with the empty DB-pEZY3 vector.

Results

Generation of a comprehensive clone resource of cancer gene promoters

Systematic studies of TF-DNA binding and transcriptional activity often require large-scale clone resources of DNA elements that can be tested across functional assays. To study the regulation of cancer-related genes, identify TFs that bind to their promoters, and compare DNA binding profiles between TF variants, truncations and isoforms, we generated a large-scale resource of 700 promoter sequences (2 kb upstream of the transcription start sites) corresponding to 556 cancer-related genes, cloned into Gateway compatible vectors for easy transfer into different destination vectors that can be used in a variety of functional assays (e.g., eY1H, pY1H, and luciferase assays) (Figure 1A, Supplementary Table 1). These genes were selected from the Cancer Gene Census as well as additional genes whose expression is often dysregulated in cancer (14). To perform TF-DNA binding studies using eY1H and pY1H assays, we also transferred these clones into the appropriate destination vectors upstream of two reporter genes and successfully generated integrant yeast strains for 508 promoters corresponding to 426 cancer-related genes. Among these, 358 genes were classified as oncogenes, tumor suppressor genes, or genes involved in fusions, with a similar number of genes in each of these classes (Figure 1B), while the remaining 68 genes are not classified into these categories by the Cancer Gene Census. Our set of genes included both highly studied genes, with >1,000 publications in PubMed, as well as lowly studied genes with <10 publications, including genes with known associations to a variety of cancer types (Figure 1C-D). Regarding biological functions, our clone resource includes genes associated with transcription, immune regulation, cell cycle, cell death, DNA damage, and other cancer-related functions (Figure 1E). We did not observe any major bias between genes for which entry clones or yeast strains were successfully generated (Figures 1C-E).

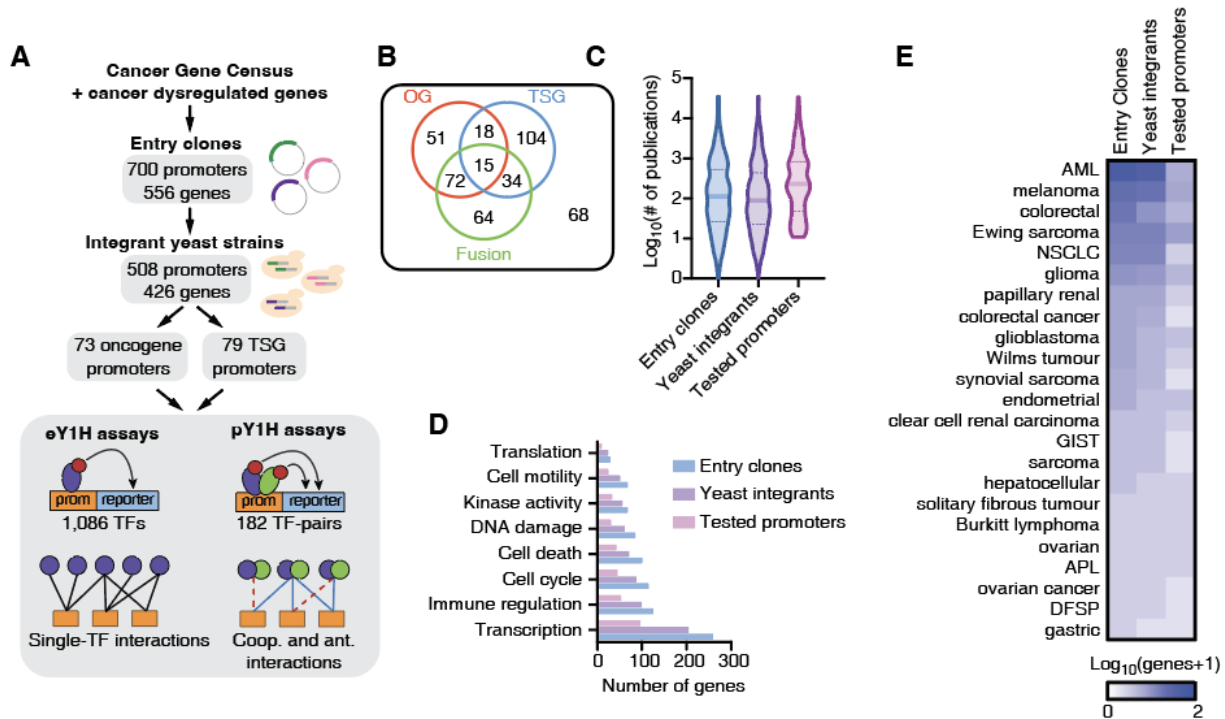


Figure 1: Generation of clone and yeast resource for cancer gene promoters.

(A) Schematic of the Gateway-compatible cancer gene promoter resource. Cancer genes were selected from the Cancer Gene Census as well as genes dysregulated in cancer. An entry clone resource of 700 promoters (556 genes) was generated as well as a yeast integrant resource corresponding to 508 promoters (426 genes). This yeast resource was tested in eY1H and pY1H assays for TF-DNA interactions. (B) Venn diagram of the number of oncogenes (OG), tumor suppressor genes (TSG) and genes involved in fusions for which yeast integrants were generated. (C) Violin plots correspond to the distribution of the number of publications per gene included in the entry clone resource, the yeast integrant collection, and the yeast integrants tested by eY1H/pY1H. (D) Number of genes associated with different biological functions for genes included in the entry clone resource, the yeast integrant collection, and the yeast integrants tested by eY1H/pY1H. (E) Number of genes associated with different cancer types among those in the set of entry clones, yeast integrants, and tested by eY1H assays.

A comprehensive cancer-associated TF-DNA network

Abnormal expression of cancer-related genes can lead to oncogenesis, cancer progression, and metastasis (22). Dysregulation of these genes can be caused by increased or decreased binding of certain TFs as a result of changes in TF expression, mutations in TFs or TF binding sites, or by alteration in TF activity due to dysregulation of upstream signaling pathways (23). To identify the TFs that bind to the promoters of cancer-related genes, we used eY1H and pY1H assays, which can identify the binding of hundreds of single or pairs of TFs to DNA elements of interest in

parallel. We prioritized 136 genes (152 promoter sequences), corresponding to 66 oncogenes and 70 tumor suppressor genes, including genes that were highly studied in the literature and promoters with high H3K27ac indicative of active usage as a regulatory DNA region (Figure 1C). These genes had a similar representation of biological process gene ontologies relative to the rest of the gene promoter clone resource ($p > 0.05$ by hypergeometric test) (Figure 1D), suggesting that our prioritized subset was functionally unbiased. We tested these DNA sequences against 1,086 TFs using eY1H assays (147,696 TF-DNA pairs tested in quadruplicate). In addition, we tested 123 of these promoter sequences (selected based on low levels of auto-activity seen in eY1H) against a collection of 182 TF-pairs and corresponding monomers using pY1H assays (22,386 TF-TF-DNA sets tested in quadruplicate). In total, we detected 1,350 TF-DNA interactions between 265 individual TFs (including 30 heterodimeric TFs) and the promoters of 108 genes (Figure 2A, Supplementary Table 2). Of these TFs, 84 are classified as transcriptional activators, 33 as repressors, and 42 as bifunctional based on their annotated effector domains (24).

Among the interactions detected by eY1H and pY1H assays, we found 559 interactions that were previously identified by ChIP-seq and 111 interactions that were reported in literature (47 interactions were reported by both ChIP-seq and literature) (Figure 2A). This illustrates the high quality of our cancer TF-DNA network. More importantly, we found 631 novel interactions, showing that our network also expands from previously reported interactions. This includes novel interactions involving TFs already known to bind to the promoters of some of the cancer genes tested. For example, GRHL3, a TF known to stimulate migration of endothelial cells and previously linked to different types of cancers (25), had 19 ChIP-seq and 2 literature interactions with our set of 152 promoters. Here, we found 15 additional interactions using eY1H assays (Figure 2B). This set of 36 genes displayed a significant enrichment in the cell differentiation gene ontology term, which is consistent with previous studies showing that GRHL3 is crucial for inducing genes within the epidermal differentiation complex, which supports terminal differentiation, suppressing hyper-proliferation (26,27). We also found interactions involving TFs

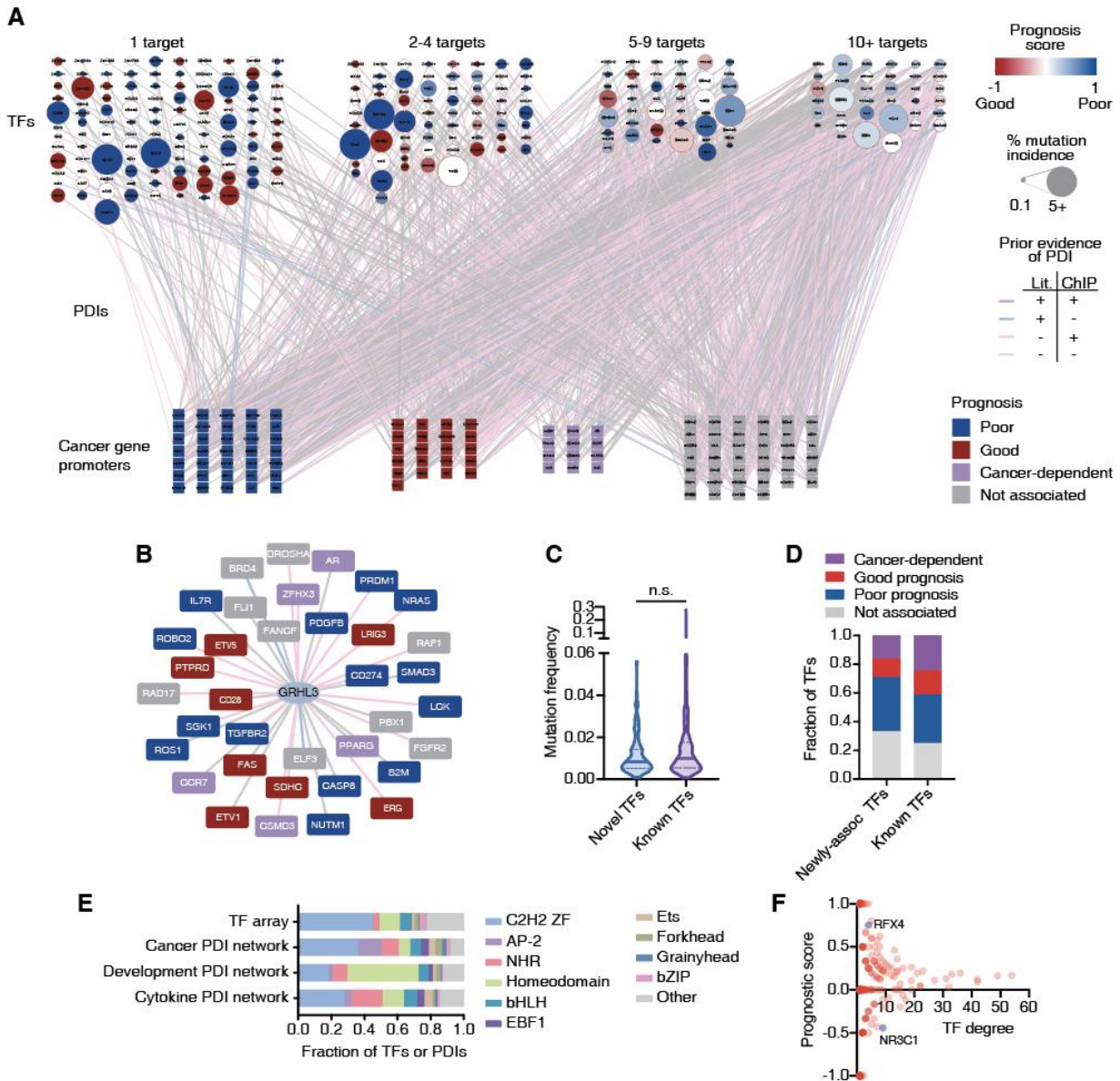


Figure 2: Large-scale cancer TF-DNA interaction network.

(A) Cancer TF-DNA interaction network determined using eY1H and pY1H assays. Circular nodes represent TFs, while squares represent cancer gene promoters. Interactions are represented by edges colored based on whether there is evidence by ChIP-seq (pink), literature (blue), both (purple) or neither (gray). TF nodes are colored based on the prognostic score calculated as $(\# \text{poor prognosis targets} - \# \text{good prognosis targets}) / (\# \text{total number of targets})$. The borders of TF nodes are colored based on whether the TF is listed (red) in CGC. TF node size indicates the % of non-synonymous mutations across all cancers. Cancer gene promoters are colored based on whether their expression is associated with poor (blue), good (red), or cancer-dependent prognosis (purple). (B) Interaction network involving GRHL3. (C) Violin plot depicting the mutation frequency across cancers for TFs known to bind/regulate our set of cancer genes (purple), and novel TFs (blue). Statistical significance determined by two-tailed Mann-Whitney's U test. (D) Fraction of TFs whose expression levels are associated with poor or good cancer prognosis for TFs known to bind/regulate our set of cancer genes and novel TFs. (E) Fraction of

TFs and TF-DNA interactions corresponding to different TF families in the TF array and the cancer promoter, cytokine promoter, and developmental enhancer networks. (F) Scatter plot of the prognostic score versus degree (number of targets) for each TF.

that were previously not known to regulate any of the genes in our network. For example, ESRRG and CEBPE interact with 15 and 5 cancer gene promoters, respectively, but interactions with our set of promoters were not reported in the literature nor by ChIP-seq. Both TFs are known to have roles in different cancers such as myeloid leukemia (CEBPE) and gastric cancer and retinoblastoma (ESRRG) (28-30). Interestingly, this set of “novel” TFs has a similar mutation rate in cancer and a similar likelihood of having a significantly poor or good association with prognosis than TFs already “known” to regulate this set of cancer genes (Figures 2C-D). Altogether, the eY1H and pY1H interaction data provides direct binding evidence supporting existing literature and ChIP-seq data, while providing many novel interactions to delineate a more comprehensive cancer-related TF-DNA network.

TF family representation in the cancer-associated TF-DNA network

We observed interactions involving all major TF families including homeodomains, Cys2His2 zinc fingers (ZF-C2H2), nuclear hormone receptors (NHRs), basic helix–loop–helix (bHLH), and basic leucine zippers (bZIP). Compared to the proportion of TF families in the array, we found an over-representation of interactions involving the EBF1, grainyhead, NHR, and AP-2 families (Figure 2E), which are known to play important roles in tumor growth and progression via diverse mechanisms (4,31-33). Interestingly, we found that AP-2, in particular TFAP2B, is also enriched compared to previous screens against developmental enhancers and cytokine gene promoters (Figure 2E and Supplementary figures 1A-B), suggesting that this TF family may be more actively involved in cancer regulation. Indeed, AP-2 family members such as TFAP2A, TFAP2B, and TFAP2C have been shown to be involved in different cancer types such as glioblastoma, melanoma, acute myeloid leukemia, pancreatic cancer, and colorectal cancer (34,35).

Conversely, we observed a depletion of interactions involving homeodomain TFs, which were previously found to be enriched in the developmental enhancer network. This is consistent with their roles in the development of anatomical features during early embryogenesis (36), but less so in cancer. This illustrates that, although cancer is inherently a developmental/differentiation process, the underlying gene regulatory networks use different sets of TFs compared to developmental networks.

TFs show varying levels of bias for binding promoters of good or poor prognosis genes

TFs in our cancer TF-DNA network bind to a widely different number of promoters, ranging from 1 to 54 promoters. Half (117/265) of TFs bind to just one promoter, while 13.58% (36/265) bind to 10 or more (Figure 2A). This is consistent with a power-law distribution, which is frequently observed in gene regulatory networks (Supplementary figure 1C) (37). Compared to highly connected TFs (i.e., that bind 10 or more promoters), moderately connected TFs (i.e., that bind 2-9 promoters) tend to have a stronger bias for targeting either genes whose high expression is associated with good prognosis or genes whose high expression is associated with poor prognostics in at least one cancer type (Figure 2F, Supplementary Tables 4, 5, 8). For example, RFX4 was found to bind to four genes in our network, and high expression of three of them (BRCA1, ELF3, TGFBR2) was associated with poor cancer survival. Considering that RFX4 has a repression domain and is relatively highly mutated in cancer, it is likely that mutated RFX4 results in a loss of the restriction on the expression of these genes, eventually contributing to their association with poor survival outcomes. Conversely, NR3C1 was found to bind to nine genes in our network, of which five (ETV1, TEC, CD28, CBLB and FAS) were associated with good prognosis. Previous studies have shown that Diffuse Large B-Cell Lymphoma (DLBCL) patients with high NR3C1 expression had a better prognosis than those with low NR3C1 expression (38). Taking the fact that NR3C1 has an activation domain, it is possible that NR3C1 binds and activates the expression of these genes, resulting in a good prognosis.

TF hubs, characterized by their high connectivity within the network, have a substantial impact on gene regulatory networks and represent potentially valuable drug targets as they coordinate the expression of multiple genes. However, we found that none of the TFs with 10 or more interactions in our network display any significant bias toward binding to the promoters of genes associated with either good or poor prognosis. For example, EGR1, a TF that binds to 27 genes in our network, 7 known to be associated with poor and 4 associated with good cancer prognosis, has a complex role in cancer with both tumor-suppressing and -promoting activities (39,40). Similarly, RUNX2, another TF hub which has a dual transcriptional role (i.e., can act as an activator or a repressor) bound both to the promoters of genes associated with poor and good cancer prognosis. This complex and diverse set of binding targets associated with TF hubs suggests that targeting them could be challenging as the overall rewiring of the cancer gene regulatory network may be hard to anticipate. Indeed, TF hubs in our network do not have a higher mutation rate in cancer compared to moderately connected TFs, consistent with the lack of bias in binding good or poor prognosis genes.

Identification of TFs as potential targets to reduced oncogene expression

During cancer development and progression, oncogenes may undergo mutations, amplifications, or structural changes, leading to poor patient outcomes. After decades of efforts, drugs such as AMG510 (sotorasib) that targets KRAS have been developed to inhibit oncogene activity (41); however, numerous oncogenes still resist direct targeting. A potential alternative therapeutic approach involves reducing the expression of the oncogene using knockdowns or by modulating gene transcription by targeting TF activity. A barrier to this approach is the potential for concomitant upregulation of poor prognosis genes or the downregulation of good prognosis genes. To nominate TF candidates that could be targeted (e.g., by downregulation, targeted degradation, or small molecule inhibition) to decrease the expression of each oncogene tested, we established the following criteria: 1) the TF is an activator (or bifunctional), and 2) it

preferentially binds to the promoters of poor prognosis genes ($PS > 0.33$). For 25 of the 51 oncogenes tested, we found at least one TF that meets the above criteria (Figure 3). This includes druggable TFs such as PPARG and RARA as well as TFs for which drugs have not yet been developed. For instance, in our network, six TFs with an activation domain bound to the promoter of NUTM1, which is frequently rearranged or fused with other genes and whose overexpression is associated with poor prognosis in patients with NUT carcinoma (42), B-cell precursor acute lymphoblastic leukemia (43), oral squamous cell carcinoma (OSCC) (44), and thyroid carcinoma (45). Treatments especially for NUT carcinoma have been undergoing development during the last decades, however, these drugs mostly target genes fused to NUTM1 in cancer; for example, BET inhibitors target BRD4 in the BRD4-NUTM1 fusion protein. Our findings suggest that we could potentially target TF activators like ARNTL and SRF to tune down the expression of NUTM1.

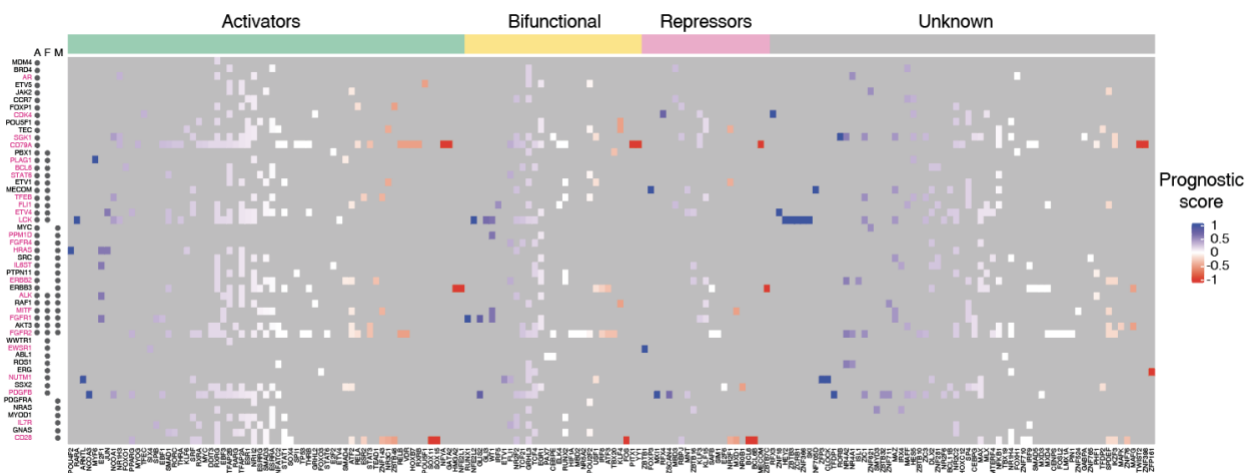


Figure 3: Heatmap of potentially targetable TFs to reduce oncogene expression.

Heatmap of prognostic scores for TFs that bind to oncogene promoters. TFs are classified as potential activators, bifunctional, or repressors based on annotated effector domains in TFRegDB. Oncogenes that contribute to cancer development through amplifications (A), fusions (F), or mutations (M) are indicated next to the gene name. Oncogenes indicated in magenta have at least one TF that is activator/bifunctional with a prognosis score > 0.33 .

Cooperative and antagonistic TF binding interactions to cancer gene promoters

In addition to the regulation by monomeric and homodimeric TFs, the expression of cancer-associated genes can also be modulated by higher order functional relationships between TFs at promoter sequences, including cooperativity and antagonism between TF pairs (46-48). To obtain a deeper understanding of how these relationships can contribute to specificity of cancer gene regulation, we used pY1H assays which test the binding of pairs of TFs to DNA regions of interest (12). We evaluated the binding of 182 TF-pairs to 123 cancer gene promoters and detected 90 cooperative and 136 antagonistic interactions, involving a total of 66 promoters and 67 TF-pairs (Figure 4A, Supplementary Table 3). We found 25 TFs to exclusively participate in cooperative binding events, 27 TFs exclusively participated in antagonistic binding events, and 21 TFs were observed to be involved in both cooperative and antagonistic interactions, depending on the TF partner or the gene promoter (Figure 4B). Of the 67 TF-pairs that showed at least one type of pY1H interactions, 23 exclusively participated in cooperativity (e.g., RXRG-NR1H3, SPDEF-ATF2, and TCF21-TCF4) and 37 of them exclusively participated in antagonism (e.g., MAX-MNT, DLX2-MLXIP, and HIF1A-RUNX2) (Figure 4A and 4C). Interestingly, 7 TF-pairs participated in both kinds of interactions (e.g., MAX-MYC and DLX2-ZNF281). Altogether, this suggests that individual TFs may regulate different target genes depending on their interacting TF partners.

We observed that TF-pairs from various families, including both intra- and inter-family pairs, exhibited cooperative and antagonistic binding (Figure 4A). We also noted that TFs from different families tend to show a preference for specific types of functional interactions (Figure 4D). For instance, NHRs predominantly engage in cooperative interactions, consistent with the well-known heterodimeric partnership in this TF family. For example, we observed that ESR1 and NR2F2, two TFs that are highly mutated in various types of cancer, cooperatively bind to 10 different promoters in our network. bHLHs are often involved in antagonistic relationships. This is mostly driven by the cancer-related TF MAX whose binding is antagonized by many other bHLH TFs such as MNT, MXD1, and MXD4.

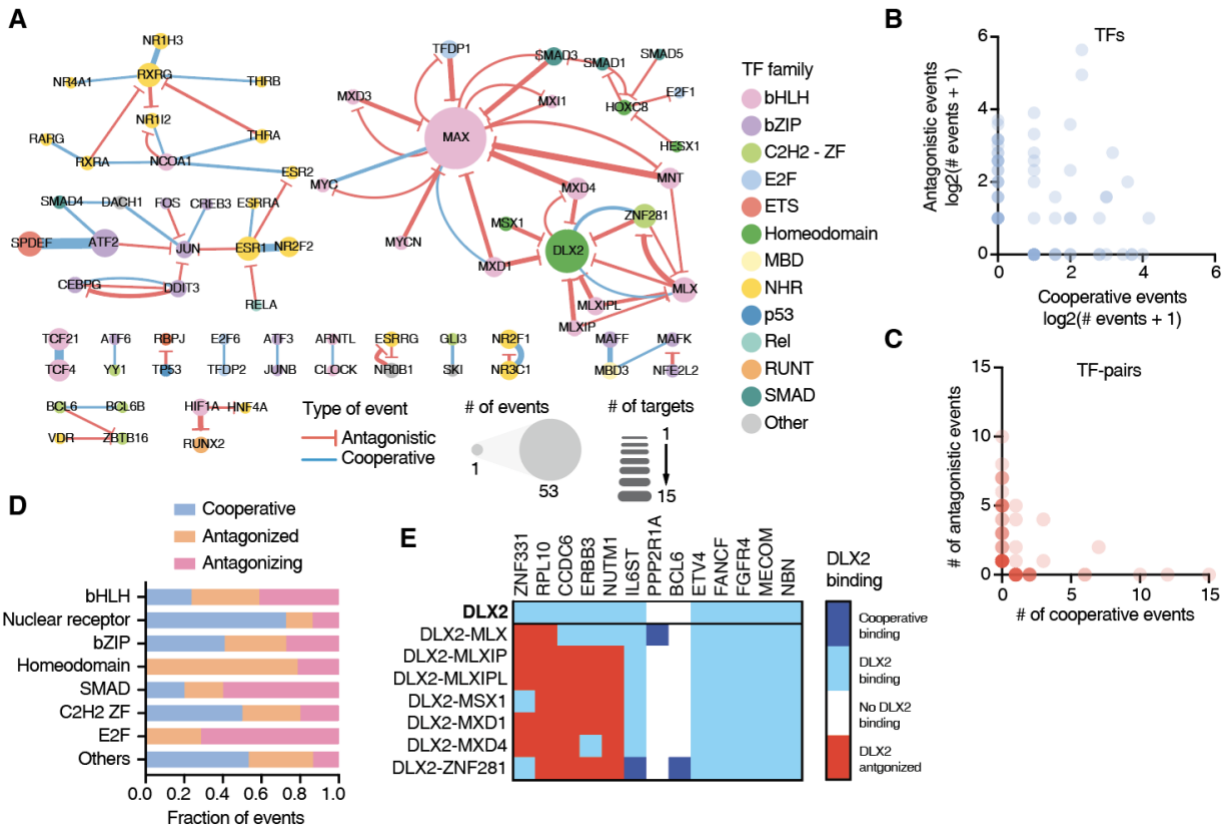


Figure 4: TF-binding cooperativity and antagonism at cancer gene promoters.

(A) Network of cooperative (blue) and antagonistic (red) relationships between TFs at the cancer gene promoters screened. Node size indicates the number of binding events for that TF. Edge width represents the number of cooperative or antagonistic events involving a specific TF-pair. (B) Number of cooperative and antagonistic events observed for individual TFs. (C) Number of cooperative and antagonistic events observed for TF-pairs. (D) Fraction of events where a TFs binds cooperatively, is antagonized by another TF, or antagonizes the binding of another TF for each TF family. (E) Heatmap of interactions involving DLX2, either on its own, or together with other TF partners at 13 cancer gene promoters. Dark blue – cooperative binding; light blue – indicates DLX2 binding not influenced by partner TF; white – no DLX2 binding; and red – DLX2 binding antagonized by TF partner.

For 21 TFs we observed different functional interactions depending on the TF partner. For example, DLX2, a TF known to be up-regulated during epithelial–mesenchymal transition and to promote cell survival (49), exhibited diverse binding relationships across TF partners at various DNA sequences. In our screen, DLX2 independently bound to the promoters of 11 genes, and at 5 of them, it showed no functional interaction with any partner TF (Figure 4E). However, at 5 promoter sequences, DLX2 was antagonized by other TFs such as MLXIP, MSX1, and MXD1.

Interestingly, DLX2 showed cooperativity or was antagonized by ZNF281 and MLX depending on the promoter sequence (Figure 4E). We had a similar observation for MYC, an essential regulator of cell growth overexpressed in many tumors, known to heterodimerize with MAX and cooperatively bind to DNA. We observed this cooperation between MYC and MAX at three promoters; however, we also noted that MYC was antagonized by MAX at the PDGFB promoter, known to be targeted by MYC (50). This aligns with a previous report indicating that MAX could antagonize MYC in a dose-dependent manner through the competition of MAX-MAX and MYC-MAX dimers for their common target DNA sites (51,52). Altogether, this highlights the complexity of higher order TF binding which can be heavily influenced by the partners of a TF but also by the target sequences involved, as we have previously observed for a small set of cytokine genes (12).

Intrinsically disordered regions affect ESR1 binding to DNA

In addition to broadening our knowledge of which TFs participate in the cancer gene regulatory network, we were also interested in conducting a more in-depth study of key cancer-related TFs. Estrogen receptor α (ESR1) is a TF that is frequently mutated or upregulated in breast cancer, and is therefore an important candidate for further functional study. ESR1 contains two structured domains – the DNA binding and the ligand binding domains – flanked by three intrinsically disordered regions (IDRs): the N-terminal region (amino acids 1-180) containing the transactivation function-1 (AF-1) domain, the hinge region (amino acids 254-305) connecting the DNA binding and ligand binding domains, and the C-terminal region (amino acids 553-595) (Figure 5A). IDRs in TFs have long been associated with roles in transcriptional activity, including most reported activation and repression domains (24). Recently, ChIP-seq studies have shown that IDRs can also modulate DNA binding across the genome, likely by affecting protein-protein interactions with other TFs and cofactors or by mediating condensate formation (53). We hypothesized that IDRs may also modulate DNA binding in a heterologous context, in the absence of other TFs and cofactors of the same species. We therefore set out to identify the contributions

of ESR1 IDRs to both DNA binding and transcriptional activity and to identify how known mutations in ESR1 IDRs might disrupt these functions.

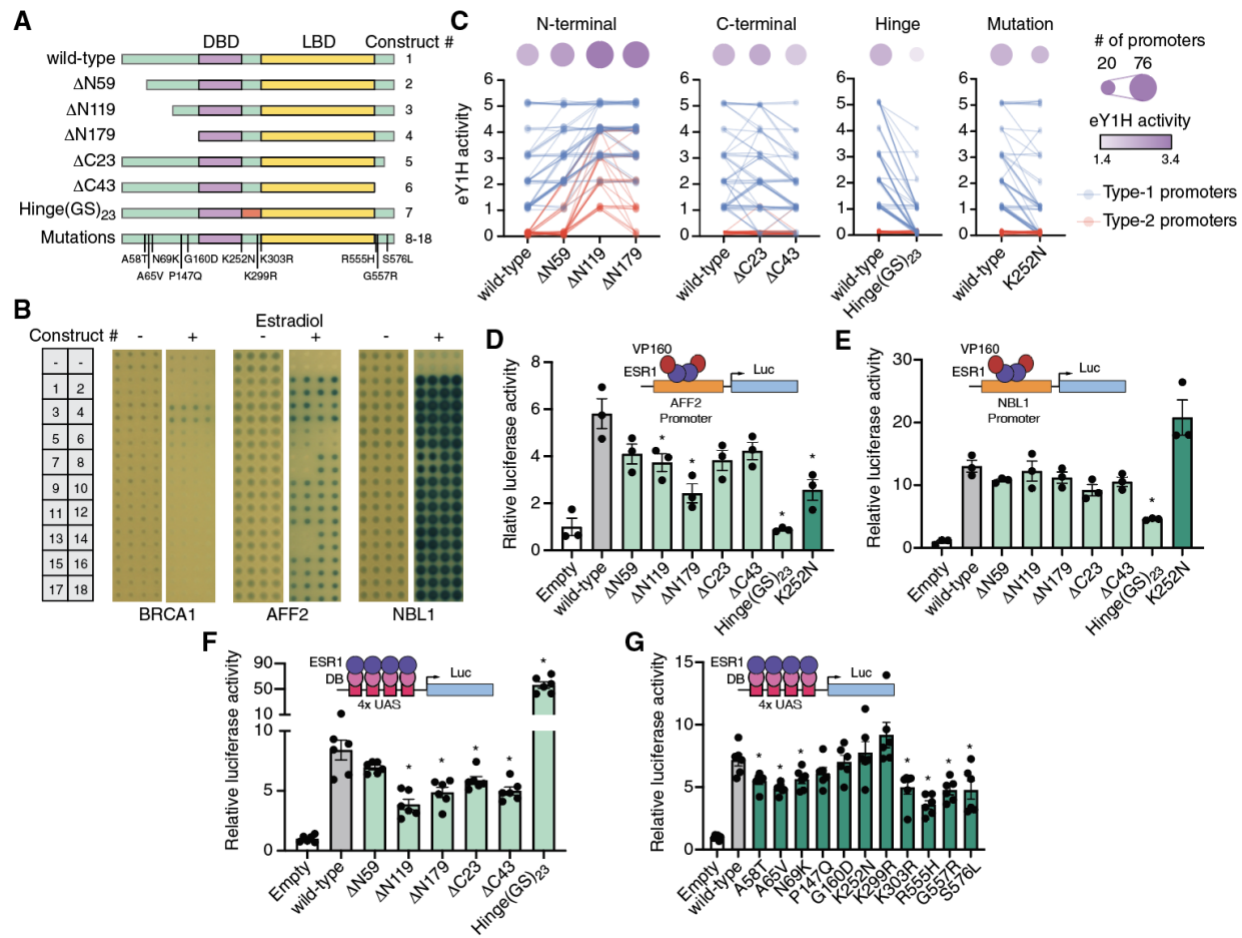


Figure 5: Role of ESR1 intrinsically disordered regions on DNA binding and transcriptional activity.

(A) Schematic of ESR1 constructs used. IDRs are indicated in green, DNA binding domain in purple, and ligand binding domain in yellow. (B) Examples of eY1H screens for binding of 18 different ESR1 constructs to the promoters of BRCA1, AFF2, and NBL1 in the presence or absence of 100 nM estradiol. (C) eY1H binding activity scored from 0 (no binding) to 5 (very strong binding) for different ESR1 constructs. Connected lines correspond to the same cancer gene promoter. Circle sizes indicate the number bound cancer promoters, whereas color intensity indicates the average eY1H activity across bound promoters. Type-1 promoters (blue) are those where wild type ESR1 binds; type-2 promoters (red) are those that wild type ESR1 does not bind. (D, E) Luciferase assays in HEK293T cells where the promoters of AFF2 (D) or NBL1 (E) are cloned upstream of firefly luciferase. ESR1 constructs are fused to 10 copies of the VP16 activator domain. Experiments were conducted in biological triplicates. * $p < 0.05$ Statistical significance determined by two-tailed Student's t-test. (F, G) Mammalian one-hybrid assays measuring the transcriptional activity of different ESR1 constructs. ESR1 fusions with the Gal4 DNA binding domain (DB) are recruited to 4 copies of the Gal4 binding site (UAS) cloned upstream of firefly luciferase. Experiments were conducted in biological sextuplicates. * $p < 0.05$ Statistical significance determined by two-tailed Student's t-test.

To determine whether these IDRs affect DNA binding we performed eY1H screens using wild-type ESR1, three truncations of the N-terminal IDR (Δ N59, Δ N119, and Δ N179), two truncations of the C-terminal IDR (Δ C23 and Δ C43), a replacement of the hinge region with 23 tandem copies of Gly-Ser (Hinge(GS)₂₃) to maintain flexibility of the linker region while removing the endogenous sequence, and 11 cancer-associated mutations in these IDRs reported in COSMIC (Figure 5A, Supplementary Table 6). Our entire collection of cancer promoters was tested against each ESR1 construct in the presence or absence of 100 nM estradiol. We found that estradiol was generally needed for DNA-binding activity, consistent with the need for estradiol-mediated dimerization for ESR1 binding to estrogen response elements (54) (Figure 5B). We observed that progressive N-terminal truncations led to increased binding strength for gene promoters that already bound the full-length wild-type ESR1 (type-1 promoters) and even led to novel DNA-binding events (type-2 promoters) (Figure 5C, Supplementary Table 7). This suggests that the N-terminal IDR, in particular amino acids 1-119, suppresses ESR1 binding to DNA. Truncations of the C-terminal region had a less clear effect, with a 23 amino acid truncation mildly increasing or decreasing DNA binding depending on the promoter sequence, while the 43 amino acid truncation mildly reducing DNA binding strength. Interestingly, we found that replacing the hinge region for a (GS)₂₃ flexible linker led to a strong reduction in DNA binding strength and the number of promoter sequences bound, suggesting that the hinge region is necessary for proper DNA binding. These observations were confirmed using reporter-based protein-DNA interaction assays in HEK293T cells treated with 100 nM estradiol (Figure 5D-E).

Most (10/11) cancer mutations tested did not affect ESR1 binding to the cancer promoters tested, suggesting that these mutations, if functional, likely affect other ESR1 molecular functions such as interactions with other TFs and cofactors. A notable exception was the K252N mutation, located at the N-terminal boundary of the hinge region, which reduced binding to promoters with weak/moderate wild-type ESR1 binding but did not affect binding to promoters with strong ESR1

binding (Figure 5C). This was confirmed using reporter assays in HEK293T cells. For example, the K252N mutation disrupted binding of ESR1 to the AFF2 promoter in both eY1H and reporter assays (Figures 5B and 5D), whereas this mutation had no effect on ESR1 binding to the NBL1 promoter (Figures 5B and 5E). Altogether, these results suggest that while the IDRs, in particular the N-terminal and hinge regions, have strong effect on DNA binding, point mutations in these regions generally have no or only mild effects.

ESR1 IDRs affect transcriptional activity

To evaluate whether the ESR1 mutations in the IDRs, and ESR1 IDRs in general, affect transcriptional activity, mammalian one-hybrid assays were performed in HEK293T cells. In these assays, cells were transfected with the different ESR1 constructs fused to the Gal4 DNA binding domain and a luciferase reporter vector driven by a minimal promoter and four copies of the Gal4 DNA binding site. Cells were then stimulated with 100 nM estradiol for 18 hours followed by measurement of luciferase activity. Progressive truncations of the N-terminal and C-terminal IDRs resulted in reduced transcriptional activity, consistent with these regions harboring the AF-1 and AF-2 activation domains (Figure 5F). Interestingly, the Hinge(GS)₂₃ replacement led to a 7-fold increase in transcriptional activity. These results show that there may be a trade-off, at least for the N-terminal IDR and hinge region, between the effect of ESR1 IDRs on DNA binding and transcriptional activity. While the N-terminal IDR suppresses DNA binding it contributes to transcriptional activation. Conversely, while the hinge region enhances DNA binding it reduces transcriptional activation.

Next, we tested the effect of cancer-associated mutations in the IDRs on transcriptional activity using mammalian one-hybrid assays. Contrary to what we observed for DNA binding, multiple mutations (7/11) significantly reduced transactivation. These mutations resided in the N-terminal (3/5) and C-terminal IDRs (3/3), as well as a mutation in the hinge region (1/3). Interestingly, the K252N mutation that affected DNA binding did not affect transcriptional activity.

These results show that mutations in the IDRs mostly affected the transactivation function of ESR1, which is consistent with the direct involvement of IDRs in precise protein-protein interactions. The relative conservation of DNA binding function following the IDR point mutations tested suggests that larger changes in IDRs may be needed to affect DNA binding.

Discussion

In this study, we have delineated a large-scale cancer network involving 1,350 TF-DNA interactions between 265 TFs and the promoters of 108 genes. About half of the interactions detected were previously identified in ChIP-seq experiments or were reported in the literature, illustrating the high quality of our network, while also identifying novel interactions. In particular, our network expands our knowledge of cancer gene regulation by identifying interactions involving TFs not previously known to regulate cancer-related genes. These newly-associated TFs have a similar mutation rate in cancer to TFs known to regulate cancer genes, illustrating how our network can also nominate novel TFs involved in cancer. Further, for many of these TFs, such as TFEC, IRF5, and ERF, we found a significant association between TF expression and cancer prognosis using TCGA data (Supplementary Table 4), suggesting that the dysregulation of these TFs can also impact cancer outcomes.

Using the cancer TF-DNA interaction network, we identified TF hubs that bind to the promoters of multiple cancer-related genes. These hubs had a similar likelihood than non-hub TFs to be highly mutated in cancer, consistent with previous observations in protein-protein interaction networks that hubs are not enriched in disease-associated genes, but rather in essential genes (55,56). We also found that TF hubs generally bind to the promoters of both poor and good prognosis genes. Altogether, our findings suggest that TF hubs in the cancer network are unlikely to be suitable drug targets for cancer therapeutics both due to pleiotropy and to an unclear effect on prognosis. It also further suggests that drugs used to target some of these TFs

(e.g., agonists and antagonists of NHRs) in autoimmune and inflammatory conditions may also affect tumor cells in cancer patients receiving these treatments (57).

Most TFs that preferentially bind to the promoters of poor or good prognosis genes have low connectivity in our network. Therefore, it is possible that, if more DNA targets are tested, these TFs would have a more balanced distribution between good and poor prognosis genes, making it challenging to anticipate the outcome of targeting these TFs as we observed with TF hubs. This suggests that targeting TFs to shift the expression balance between good and poor prognosis genes may be generally challenging to attain. Nevertheless, targeting TFs could be a suitable strategy when the goal is to reduce the expression of a single (or few) mutated oncogene(s). This will require that targeting the TF has limited side effects and that it does not lead to an unfavorable expression balance between genes associated with good and poor prognosis. In this study, we identified 25 oncogenes whose expression could potentially be targeted by inhibiting an activator/bifunctional TF that has a positive prognosis score, and is therefore also likely to reduce the expression of poor prognosis genes. Whether this ultimately leads to the expected changes in expression and results in reduced proliferation, reduced migration, or increased cell death remains to be determined.

Our study reveals that cooperativity and antagonism between TFs may play an extensive role in the regulation of cancer-related genes. This could limit the efficacy of therapeutics involving the activation or overexpression of individual TFs, as the activation of one TF in a cooperative pair may not be sufficient to induce promoter targeting, while an antagonized TF could still be prevented from binding despite activation. Alternative approaches may involve combinatorial treatments targeting both TFs in key cooperative pairs or targeting of antagonistic TFs rather than TFs that directly bind the promoters of interest.

An intriguing aspect of our study is the evaluation of the role of TF IDRs, particularly focusing on ESR1, on DNA binding and transcriptional activity. Our findings show that IDRs, even the ones that are not in close proximity to a DNA binding domain, can modulate DNA binding both

increasing or decreasing the number and strength of interactions to target gene promoters. Our ESR1 results are consistent with a more general pattern showing that isoforms of different TFs, such as MAX, STAT1, and RXRG, with intact DNA binding domains can have different DNA targets and different functional relationships with interacting TF partners (12). Interestingly, IDRs that do not overlap with mapped effector domains can also impact transcriptional activity. In particular, replacing the hinge region of ESR1 for a flexible (GS)₂₃ linker increased activation by 7-fold while significantly reducing DNA binding. This suggests not only that IDRs modulate TF functions, but also that some IDRs may contribute to trade-offs between DNA binding and activity.

The use of yeast-based systems for analyzing TF-DNA interactions, while powerful, may not fully recapitulate the contexts in which they occur in cancer cells, but rather provides a repertoire of possible interactions for future study. Additionally, while our resource of 700 promoter clones is of great utility to the community studying cancer gene regulation, other distal elements such as enhancers and silencers may significantly contribute to the regulation of these cancer genes. Future studies could expand our resources and extend these analyses to encompass a wider range of regulatory elements. Overall, our work provides an experimental and informational resource that can facilitate and motivate future investigations of the role of TFs in cancer gene dysregulation. Such studies will open the door to explore the targeting of TFs as an avenue for cancer treatment.

Data availability

All data generated in this study are available within the manuscript and its supplementary information.

Supplementary Data statement

Supplementary Data are available at NAR online.

Authors contributions

J.I.F.B., Y.L., and A.B. conceived the project. A.B., M.Y., S.S., and Y.C. generated the plasmid and yeast clone collection. Y.L. and R.L. performed the eY1H and pY1H experiments. I.G. performed the M1H and luciferase assays. Y.L., L.S., and Z.L. performed data analyses. A.F. supervised Y.L. and edited the manuscript. Y.L., A.B., and J.I.F.B. wrote the manuscript with contributions from the other authors.

Acknowledgements

We thank Drs. Martha Bulyk and Haribabu Arthanari, and members of the Fuxman Bass laboratory for helpful discussions regarding experiments and data analysis.

Funding

This work was funded by the National Institutes of Health grants R35 GM128625 and U01 CA232161, awarded to J.I.F.B.

Conflict of Interest

The authors declare no conflicts of interest.

References

1. Djakiew, D. (2000) Dysregulated expression of growth factors and their receptors in the development of prostate cancer. *Prostate*, **42**, 150-160.
2. Vervoort, S.J., Devlin, J.R., Kwiatkowski, N., Teng, M., Gray, N.S. and Johnstone, R.W. (2022) Targeting transcription cycles in cancer. *Nat Rev Cancer*, **22**, 5-24.

3. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646-674.
4. Carrasco Pro, S., Hook, H., Bray, D., Berenzy, D., Moyer, D., Yin, M., Labadorf, A.T., Tewhey, R., Siggers, T. and Fuxman Bass, J.I. (2023) Widespread perturbation of ETS factor binding sites in cancer. *Nat Commun*, **14**, 913.
5. Sjostedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F. *et al.* (2020) An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, **367**.
6. Jin, X., Zhu, L., Cui, Z., Tang, J., Xie, M. and Ren, G. (2019) Elevated expression of GNAS promotes breast cancer cell proliferation and migration via the PI3K/AKT/Snail1/E-cadherin axis. *Clin Transl Oncol*, **21**, 1207-1219.
7. Henrich, K.O., Fischer, M., Mertens, D., Benner, A., Wiedemeyer, R., Brors, B., Oberthuer, A., Berthold, F., Wei, J.S., Khan, J. *et al.* (2006) Reduced expression of CAMTA1 correlates with adverse outcome in neuroblastoma patients. *Clin Cancer Res*, **12**, 131-138.
8. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K. *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res*, **48**, D882-D889.
9. Reece-Hoyes, J.S., Diallo, A., Lajoie, B., Kent, A., Shrestha, S., Kadreppa, S., Pesyna, C., Dekker, J., Myers, C.L. and Walhout, A.J. (2011) Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat Methods*, **8**, 1059-1064.
10. Berenson, A. and Fuxman Bass, J.I. (2023) Enhanced Yeast One-Hybrid Assays to Study Protein-DNA Interactions. *Methods Mol Biol*, **2599**, 11-20.
11. Fuxman Bass, J.I., Reece-Hoyes, J.S. and Walhout, A.J. (2016) Gene-Centered Yeast One-Hybrid Assays. *Cold Spring Harb Protoc*, **2016**, pdb top077669.
12. Berenson, A., Lane, R., Soto-Ugaldi, L.F., Patel, M., Ciausiu, C., Li, Z., Chen, Y., Shah, S., Santoso, C., Liu, X. *et al.* (2023) Paired yeast one-hybrid assays to detect DNA-binding cooperativity and antagonism across transcription factors. *Nat Commun*, **14**, 6570.
13. Fuxman Bass, J.I., Reece-Hoyes, J.S. and Walhout, A.J. (2016) Generating Bait Strains for Yeast One-Hybrid Assays. *Cold Spring Harb Protoc*, **2016**, pdb prot088948.
14. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E. *et al.* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*, **47**, D941-D947.
15. Reece-Hoyes, J.S., Barutcu, A.R., McCord, R.P., Jeong, J.S., Jiang, L., MacWilliams, A., Yang, X., Salehi-Ashtiani, K., Hill, D.E., Blackshaw, S. *et al.* (2011) Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. *Nat Methods*, **8**, 1050-1052.

16. Luck, K., Kim, D.K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B. *et al.* (2020) A reference map of the human binary protein interactome. *Nature*, **580**, 402-408.
17. Yevshin, I., Sharipov, R., Kolmykov, S., Kondrakhin, Y. and Kolpakov, F. (2019) GTRD: a database on gene transcription regulation-2019 update. *Nucleic Acids Res*, **47**, D100-D105.
18. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.
19. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I. *et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*, **44**, e71.
20. Therneau, T.M., Grambsch, P.M. and SpringerLink. (2000) *Modeling Survival Data: Extending the Cox Model*. 1st 2000. ed. Springer New York : Imprint: Springer, New York, NY.
21. Sondka, Z., Dhir, N.B., Carvalho-Silva, D., Jupe, S., Madhumita, McLaren, K., Starkey, M., Ward, S., Wilding, J., Ahmed, M. *et al.* (2024) COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res*, **52**, D1210-D1217.
22. Li, F., Wu, T., Xu, Y., Dong, Q., Xiao, J., Xu, Y., Li, Q., Zhang, C., Gao, J., Liu, L. *et al.* (2020) A comprehensive overview of oncogenic pathways in human cancer. *Brief Bioinform*, **21**, 957-969.
23. Gonda, T.J. and Ramsay, R.G. (2015) Directly targeting transcriptional dysregulation in cancer. *Nat Rev Cancer*, **15**, 686-694.
24. Soto, L.F., Li, Z., Santoso, C.S., Berenson, A., Ho, I., Shen, V.X., Yuan, S. and Fuxman Bass, J.I. (2022) Compendium of human transcription factor effector domains. *Mol Cell*, **82**, 514-526.
25. Wang, X.K., Zhou, F.F., Tao, H.R., Wang, X., Zhang, C., Su, F., Wang, S.P., Xu, L.H., Pan, X.K., Feng, M.H. *et al.* (2017) Knockdown of GRHL3 inhibits activities and induces cell cycle arrest and apoptosis of human colorectal cancer cells. *J Huazhong Univ Sci Technol Med Sci*, **37**, 880-885.
26. Huang, H., Liu, J., Li, M., Guo, H., Zhu, J., Zhu, L., Wu, S., Mo, K., Huang, Y., Tan, J. *et al.* (2022) Cis-regulatory chromatin loops analysis identifies GRHL3 as a master regulator of surface epithelium commitment. *Sci Adv*, **8**, eabo5668.
27. Lin, Z., Jin, S., Chen, J., Li, Z., Lin, Z., Tang, L., Nie, Q. and Andersen, B. (2020) Murine interfollicular epidermal differentiation is gradualistic with GRHL3 controlling progression from stem to transition cell states. *Nat Commun*, **11**, 5434.
28. Li, K., Du, Y., Wei, D.Q. and Zhang, F. (2019) CEBPE expression is an independent prognostic factor for acute myeloid leukemia. *J Transl Med*, **17**, 188.

29. Kang, M.H., Choi, H., Oshima, M., Cheong, J.H., Kim, S., Lee, J.H., Park, Y.S., Choi, H.S., Kweon, M.N., Pack, C.G. *et al.* (2018) Estrogen-related receptor gamma functions as a tumor suppressor in gastric cancer. *Nat Commun*, **9**, 1920.
30. Heckler, M.M., Thakor, H., Schafer, C.C. and Riggins, R.B. (2014) ERK/MAPK regulates ERRgamma expression, transcriptional activity and receptor-mediated tamoxifen resistance in ER+ breast cancer. *FEBS J*, **281**, 2431-2442.
31. Tsigelny, I.F., Kouznetsova, V.L., Pingle, S.C. and Kesari, S. (2014) bHLH Transcription factors inhibitors for cancer therapy: general features for in silico drug design. *Curr Med Chem*, **21**, 3227-3243.
32. Hu, H., Zhang, Q., Hu, F.F., Liu, C.J. and Guo, A.Y. (2021) A comprehensive survey for human transcription factors on expression, regulation, interaction, phenotype and cancer survival. *Brief Bioinform*, **22**.
33. Zhang, Y., Wang, Z., Ma, J., Huo, J., Li, Y., Wang, Y., Chen, H., Shan, L. and Ma, X. (2020) Bioinformatics Identification of the Expression and Clinical Significance of E2F Family in Endometrial Cancer. *Front Genet*, **11**, 557188.
34. Raap, M., Gierendt, L., Kreipe, H.H. and Christgen, M. (2021) Transcription factor AP-2beta in development, differentiation and tumorigenesis. *Int J Cancer*, **149**, 1221-1227.
35. Kolat, D., Kaluzinska, Z., Bednarek, A.K. and Pluciennik, E. (2019) The biological characteristics of transcription factors AP-2alpha and AP-2gamma and their importance in various types of cancers. *Biosci Rep*, **39**.
36. Banerjee-Basu, S. and Baxevanis, A.D. (2001) Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res*, **29**, 3258-3269.
37. Lima-Mendez, G. and van Helden, J. (2009) The powerful law of the power law and other myths in network biology. *Mol Biosyst*, **5**, 1482-1493.
38. Yan, C., He, X., Qi, R., Cao, L., Zheng, S., Huang, C., Yang, P., Wang, J., Zhu, M., Li, S. *et al.* (2023) Prediction and prognostic potential of NR3C1 gene expression level in DLBCL patients. *Hematology*, **28**, 2251199.
39. Wang, B., Guo, H., Yu, H., Chen, Y., Xu, H. and Zhao, G. (2021) The Role of the Transcription Factor EGR1 in Cancer. *Front Oncol*, **11**, 642547.
40. Adamson, E.D. and Mercola, D. (2002) Egr1 transcription factor: multiple roles in prostate tumor cell growth and survival. *Tumour Biol*, **23**, 93-102.
41. Huang, L., Guo, Z., Wang, F. and Fu, L. (2021) KRAS mutation: from undruggable to druggable in cancer. *Signal Transduct Target Ther*, **6**, 386.
42. French, C.A. (2018) NUT Carcinoma: Clinicopathologic features, pathogenesis, and treatment. *Pathol Int*, **68**, 583-595.
43. Hormann, F.M., Hoogkamer, A.Q., Beverloo, H.B., Boeree, A., Dingjan, I., Wattel, M.M., Stam, R.W., Escherich, G., Pieters, R., den Boer, M.L. *et al.* (2019) NUTM1 is a recurrent

- fusion gene partner in B-cell precursor acute lymphoblastic leukemia associated with increased expression of genes on chromosome band 10p12.31-12.2. *Haematologica*, **104**, e455-e459.
44. Riaz, A. and Khan, F.F. (2023) Comparative analysis of HPV-Human Protein Interaction Networks in Oropharyngeal and Oral Squamous cell carcinomas. *bioRxiv*, 2023.2012.2005.570064.
 45. Allison, D.B., Rueckert, J., Cornea, V., Lee, C.Y., Dueber, J. and Bocklage, T. (2022) Thyroid Carcinoma with NSD3::NUTM1 Fusion: a Case with Thyrocyte Differentiation and Colloid Production. *Endocr Pathol*, **33**, 315-326.
 46. Morgunova, E. and Taipale, J. (2017) Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol*, **47**, 1-8.
 47. Ibarra, I.L., Hollmann, N.M., Klaus, B., Augsten, S., Velten, B., Hennig, J. and Zaugg, J.B. (2020) Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nat Commun*, **11**, 124.
 48. Hu, S., Metcalf, E., Mahat, D.B., Chan, L., Sohal, N., Chakraborty, M., Hamilton, M., Singh, A., Singh, A., Lees, J.A. *et al.* (2022) Transcription factor antagonism regulates heterogeneity in embryonic stem cell states. *Mol Cell*, **82**, 4410-4427 e4412.
 49. Tiwari, N., Gheldof, A., Tatari, M. and Christofori, G. (2012) EMT as the ultimate survival mechanism of cancer cells. *Semin Cancer Biol*, **22**, 194-207.
 50. Winkler, M., Staniczek, T., Kurschner, S.W., Schmid, C.D., Schonhaber, H., Cordero, J., Kessler, L., Mathes, A., Sticht, C., Nessling, M. *et al.* (2021) Endothelial GATA4 controls liver fibrosis and regeneration by preventing a pathogenic switch in angiocrine signaling. *J Hepatol*, **74**, 380-393.
 51. Amati, B., Brooks, M.W., Levy, N., Littlewood, T.D., Evan, G.I. and Land, H. (1993) Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell*, **72**, 233-245.
 52. Amati, B. and Land, H. (1994) Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr Opin Genet Dev*, **4**, 102-108.
 53. Brodsky, S., Jana, T. and Barkai, N. (2021) Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Curr Opin Struct Biol*, **71**, 110-115.
 54. Yasar, P., Ayaz, G., User, S.D., Gupur, G. and Muyan, M. (2017) Molecular mechanism of estrogen-estrogen receptor signaling. *Reprod Med Biol*, **16**, 4-20.
 55. Barabasi, A.L., Gulbahce, N. and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet*, **12**, 56-68.
 56. Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M. and Barabasi, A.L. (2007) The human disease network. *Proc Natl Acad Sci U S A*, **104**, 8685-8690.

57. Patalano, S.D., Fuxman Bass, P. and Fuxman Bass, J.I. (2023) Transcription factors in the development and treatment of immune disorders. *Transcription*, 1-23.