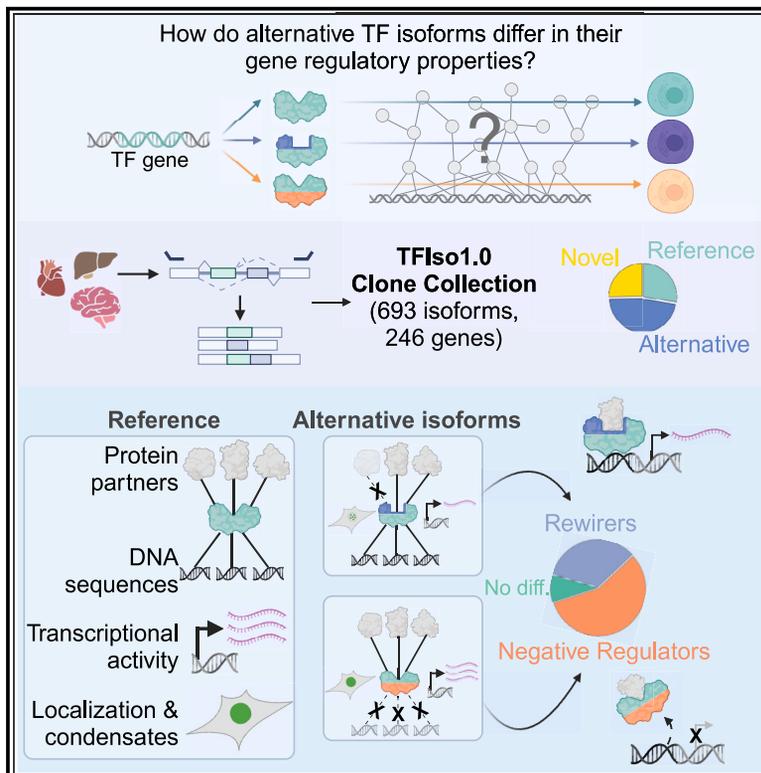


Molecular Cell

Widespread variation in molecular interactions and regulatory properties among transcription factor isoforms

Graphical abstract



Highlights

- Profiled DNA/protein binding, activation, and localization of 100s of TF isoforms
- Two-thirds of alternative TF isoforms differ from reference in molecular functions
- Sequence differences outside of TF domains frequently impact molecular functions
- Two main categories of alternative TF isoforms: “negative regulators” and “rewirers”

Authors

Luke Lambourne, Kaia Mattioli, Clarissa Santoso, ..., Marc Vidal, Martha L. Bulyk, Juan I. Fuxman Bass

Correspondence

kaia.mattioli@gmail.com (K.M.), nidhi.sahni.2025@gmail.com (N.S.), marc_vidal@dfci.harvard.edu (M.V.), mlbulyk@genetics.med.harvard.edu (M.L.B.), fuxman@bu.edu (J.I.F.B.)

In brief

Lambourne, Mattioli, Santoso, et al. compare protein isoforms of the same transcription factor genes through high-throughput profiling of DNA-binding, protein-binding, activation, localization, and condensate formation. Differences between isoforms are widespread, often unpredictable from sequence differences, and inform a disease-associated categorization of alternative isoforms into rewirers or negative regulators.



Resource

Widespread variation in molecular interactions and regulatory properties among transcription factor isoforms

Luke Lambourne,^{1,2,3,22} Kaia Mattioli,^{4,22,*} Clarissa Santoso,^{5,6,22} Gloria Sheynkman,^{1,2,3,23} Sachi Inukai,^{4,23} Babita Kaundal,^{7,23} Anna Berenson,⁸ Kerstin Spirohn-Fitzgerald,^{1,2,3} Anukana Bhattacharjee,^{9,10} Elisabeth Rothman,⁴ Shaleen Shrestha,⁵ Florent Laval,^{1,2,3,11,12} Brent S. Carroll,⁴ Stephen P. Plassmeyer,^{13,14} Ryan J. Emenecker,^{13,14} Zhipeng Yang,^{1,2,3} Deepa Bisht,⁷ Jared A. Sewell,⁵ Guangyuan Li,^{9,10} Anisa Prasad,^{4,15} Sabrina Phanor,⁴ Ryan Lane,⁵ Devlin C. Moyer,⁶ Toby Hunt,¹⁶ Dawit Balcha,^{1,2,3} Marinella Gebbia,^{1,17,18,19} Jean-Claude Twizere,^{1,11,12} Tong Hao,^{1,2,3} Alex S. Holehouse,^{13,14} Adam Frankish,¹⁶ Josh A. Riback,²⁰ Nathan Salomonis,^{9,10} Michael A. Calderwood,^{1,2,3} David E. Hill,^{1,2,3} Nidhi Sahni,^{7,*} Marc Vidal,^{1,2,3,*} Martha L. Bulyk,^{1,4,21,*} and Juan I. Fuxman Bass^{1,5,6,8,24,*}

¹Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA 02215, USA

²Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA

³Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

⁴Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

⁵Department of Biology, Boston University, Boston, MA 02215, USA

⁶Bioinformatics Program, Boston University, Boston, MA 02215, USA

⁷Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁸Molecular Biology, Cell Biology & Biochemistry Program, Boston University, Boston, MA 02215, USA

⁹Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA

¹⁰Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

¹¹TERRA Teaching and Research Centre, University of Liège, Gembloux 5030, Belgium

¹²Laboratory of Viral Interactomes, GIGA Institute, University of Liège, Liège 4000, Belgium

¹³Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA

¹⁴Center for Biomolecular Condensates, Washington University in St. Louis, St. Louis, MO 63110, USA

¹⁵Harvard College, Cambridge, MA 02138, USA

¹⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

¹⁷The Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada

¹⁸Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada

¹⁹Lunenfeld-Tanenbaum Research Institute (LTRI), Sinai Health System, Toronto, ON M5G 1X5, Canada

²⁰Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

²¹Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

²²These authors contributed equally

²³These authors contributed equally

²⁴Lead contact

*Correspondence: kaia.mattioli@gmail.com (K.M.), nidhi.sahni.2025@gmail.com (N.S.), marc_vidal@dfci.harvard.edu (M.V.), mlbulyk@genetics.med.harvard.edu (M.L.B.), fuxman@bu.edu (J.I.F.B.)
<https://doi.org/10.1016/j.molcel.2025.03.004>

SUMMARY

Most human transcription factor (TF) genes encode multiple protein isoforms differing in DNA-binding domains, effector domains, or other protein regions. The global extent to which this results in functional differences between isoforms remains unknown. Here, we systematically compared 693 isoforms of 246 TF genes, assessing DNA binding, protein binding, transcriptional activation, subcellular localization, and condensate formation. Relative to reference isoforms, two-thirds of alternative TF isoforms exhibit differences in one or more molecular activities, which often could not be predicted from sequence. We observed two primary categories of alternative TF isoforms: “rewirers” and “negative regulators,” both of which were associated with differentiation and cancer. Our results support a model wherein the relative expression levels of, and interactions involving, TF isoforms add an understudied layer of complexity to gene regulatory networks, demonstrating the importance of isoform-aware characterization of TF functions and providing a rich resource for further studies.



INTRODUCTION

Gene regulatory programs are major drivers of cellular phenotypes in development and disease and are controlled by sequence-specific transcription factors (TFs).¹ The last four decades have seen an explosion in studies and throughput to determine TF DNA-binding specificities,^{2–5} transcriptional activities,^{6–8} and protein-protein interactions (PPIs).^{9–11} These efforts have focused on generating profiles for the wild type, reference isoforms, or individual domains. However, few studies consider the multiple TF proteoforms resulting from alternative promoter, splice site/junction, and/or terminal exon usage (Figure 1A).^{12–14}

Recent studies have investigated how TF coding variants affect functions such as DNA binding and transcriptional activity, ranging from no detectable effect to complete loss or even gain of functions.^{5,15,16} However, different TF isoforms remain far less studied, despite being widespread. Indeed, TFs are among the most frequently spliced classes of genes^{17,18} and most human TFs are present as multiple isoforms¹⁹ (~4,100 isoforms across ~1,600 TF genes).^{20,21} This is likely a substantial underestimate, as novel disease- and condition-specific isoforms continue to be detected by long-read RNA sequencing (RNA-seq) technologies.^{21–25} Importantly, mass spectrometry studies suggest that most frame-preserving isoforms are translated, highlighting the importance of studying their functional activities.²⁶

Though most TF isoforms remain uncharacterized, some are known to exhibit drastically different functions,^{27,28} differentially binding to DNA, cofactors, or chromatin-associated proteins.^{18,27,29–31} Notably, two isoforms of *FOXP1* exhibit different DNA-binding specificities and consequently drive opposing phenotypes in differentiation.³² TF isoforms also play distinct roles in disease. For example, altered expression of an alternative isoform of *WT1* causes Frasier syndrome.³³ The alternative and reference isoforms differ by only 3 amino acids (aa) (–KTS) but diverge substantially in DNA-binding specificity.³⁴ Moreover, TF isoforms can be dysregulated in cancer.³⁵ Several oncogene or tumor suppressor TFs encode dominant-negative isoforms that compete with reference isoform activity, including *STAT3*,^{36,37} *ESR1*,^{38,39} and *TP53*.⁴⁰

These few cases are striking, prompting the question of whether they represent a more general phenomenon of alternative isoforms diversifying TF functions. Previously, we reported that protein isoforms often exhibit functional differences in their PPIs, supporting the latter hypothesis.⁴¹ However, few studies have interrogated isoform-resolved TF functions, primarily due to technical limitations. For example, chromatin immunoprecipitation (ChIP)-seq studies use antibodies that rarely distinguish between isoforms.⁴² Most large-scale studies of human TF-DNA binding consider only the reference isoforms or only DNA-binding domains (DBDs).^{2,3,43,44} Additionally, high-throughput transcriptional activity studies mostly test short peptides, which can miss synergistic or antagonistic effects between domains within full-length isoforms.^{6,7,45} Overall, there is a need for high-throughput, integrative, experimental approaches to dissect the mechanisms by which alternative isoform usage alters TF regulatory functions.

Here, we present an in-depth, experimentally driven investigation into the functional differences between 693 isoforms of 246

TFs. The results reveal system-scale relationships between TF sequence and functional diversity, including DNA binding, transcriptional activation, PPIs, localization, and condensate formation. In this study, we present evidence that most alternative isoforms diversify TF functions, provide a quantitative survey of the mechanisms involved, and propose that this rewiring of molecular functions through alternative isoforms constitutes an often overlooked but important layer of complexity in gene regulation in development and disease.

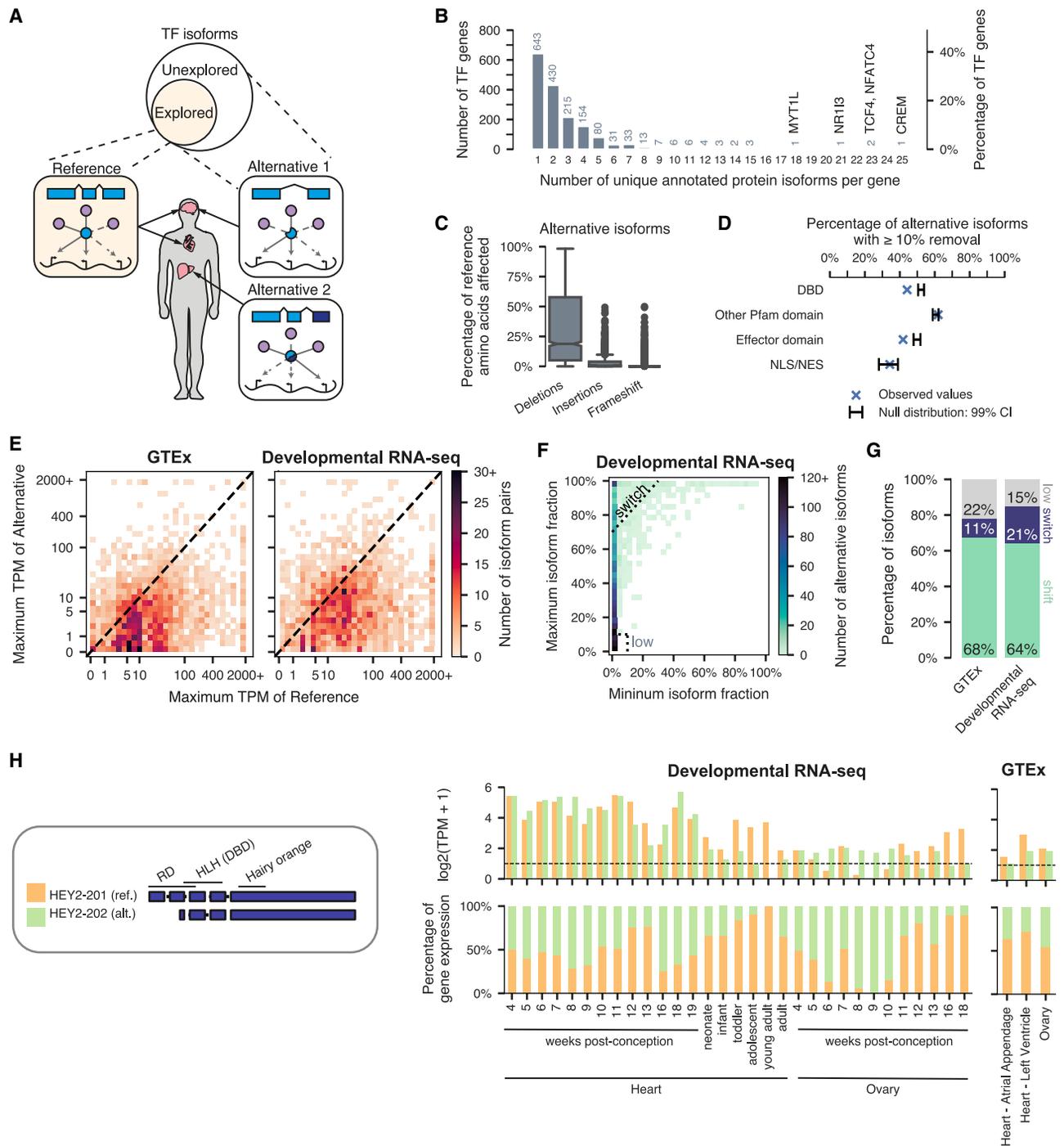
RESULTS

TF isoforms are prevalent and frequently affect functional domains

To investigate the prevalence of TF isoforms that may affect gene regulatory networks (GRNs), we cataloged annotated protein-coding isoforms of TF genes. GENCODE²¹ annotates 4,144 protein-coding isoforms for the 1,635 human TF genes,²⁰ with 992 TF genes (61%) encoding multiple isoforms (Figures 1B and S1A). For each gene, we defined a “reference” isoform using the MANE select representative transcript annotation set⁴⁶ and compared pairwise all “alternative” isoforms of each gene to their cognate reference isoforms. aa sequence differences between alternative and reference isoforms arise from alternative N-terminal regions, C-terminal regions, and/or alternatively spliced internal exons (Figure S1B). Across alternative TF isoforms, a median of 18.8% of aa are deleted (Figure 1C). Although insertions and frameshifts are rare, 195 (8.5%) and 68 (3%) isoforms contain insertions or frameshifts, respectively, affecting >10% of their total length (Figure 1C).

Domains such as DBDs and effector domains mediate specific biophysical interactions vital to TF functions. We mapped three key domain types to TF isoforms: (1) conserved structural domains (Pfam), separated into DBDs and other domains (e.g., ligand-binding domains); (2) effector domains shown to either activate or repress transcription^{6–8}; and (3) nuclear localization/export signals (NLS/NES). Overall, 1,728 alternative TF isoforms (75%) differed by ≥ 1 aa in one of these domains. Despite the frequency of affected domains, however, DBDs and effector domains are affected significantly less than expected by chance, whereas NLS/NES motifs and other Pfam domains are not (Figures 1D, S1C, and S1D). This supports reports that splicing boundaries often reside outside of domains, perhaps reflecting selection pressure to avoid deleterious splicing variants⁴⁷ or evolutionary selection through which entire exons are gained or lost.⁴⁸

Next, we examined the expression patterns of TF isoforms across GTEx,⁴⁹ which comprises primarily healthy human adult tissues, and a time-course series of human development across seven organs (“developmental” RNA-seq).⁵⁰ To correct for imbalance between datasets, we re-sampled GTEx to the equivalent size of the developmental RNA-seq (Figure S1E; STAR Methods). As expected, reference isoforms generally had higher maximum expression across tissues and developmental stages than did alternative isoforms; however, alternative isoforms had higher expression than their cognate reference isoforms in 522 (23%) and 551 (24%) cases in GTEx and developmental RNA-seq, respectively (Figure 1E).



(legend continued on next page)

Some TFs, including *FOXP1*,³² *REST*,⁵¹ and *GRHL1*,⁵² dramatically “switch” from expressing one particular isoform to another at key developmental stages. To determine the prevalence of switch events, we calculated the percentage of total TF gene expression for each isoform (“fractional isoform expression”). We considered an alternative isoform to exhibit a switch event if it changed its fractional expression by at least 70% between any two conditions. Isoforms with <10% fractional expression in any condition were classified as “low” fractional expression (22% in GTEx and 15% in developmental RNA-seq), whereas other isoforms had more subtle “shifts” across conditions (Figures 1F and S1F). Most TF isoforms (68% in GTEx and 64% in developmental RNA-seq) exhibited shifts rather than dramatic switching events (Figure 1G). Interestingly, the fraction of alternative isoforms that showed switches is higher in the developmental RNA-seq data than in GTEx data (21% versus 11%, respectively), suggesting that many alternative TF isoforms may affect gene regulation in early development.

Overall, most TF genes with multiple isoforms have at least one alternative isoform exhibiting switch or shift events in GTEx (94%) or the developmental RNA-seq data (96%). For example, an alternative isoform of *HEY2*, a cardiac transcriptional repressor,⁵³ lacks the N-terminal repression domain (RD) and is more abundantly expressed in developing ovaries relative to the reference (Figure 1H). In summary, most annotated alternative TF isoforms show differences in protein domains and variable expression across tissues, particularly in development. Our results suggest that alternative TF isoforms serve distinct roles in GRNs and thus underscore the need to functionally characterize TF isoforms.

Systematic characterization of TF isoforms reveals differences in molecular interactions and regulatory activity

Given that TF isoforms exhibit differences in primary sequence, structural domains, and expression patterns, we hypothesized that alternative TF isoforms likely exhibit widespread functional divergence. To investigate this hypothesis, we systematically assayed molecular functional differences across a large collection of TF isoforms. This collection (TFIso1.0) was generated using PCR from fetal and adult brain, heart, and liver—tissues with well-documented differences in isoform expression⁵⁴ (Figure 2A). TFIso1.0 comprises 693 isoforms, corresponding to 246 genes spanning most TF classes (Figure S2A; Data S1).

Comparing TFIso1.0 to GENCODE,²¹ 510 isoforms match known transcripts, whereas 183 (26%) are novel. These novel isoforms were manually curated using GENCODE standards to ensure high quality (STAR Methods). Because our cloning strategy used annotated N- and C-terminal regions for primer design, we likely missed unannotated alternative transcription start and polyadenylation sites, and, consequently, novel TF isoforms

were more likely to differ in internal exons and less likely to differ at the N- and C-terminals compared with annotated alternative TF isoforms (Figure 2B). Although the average expression of novel isoforms was generally lower than for annotated alternative isoforms, their maximum expression values across conditions were similar, indicating that novel isoforms may be more tissue- or developmental-stage-restricted (Figures 2C and S2B–S2D).

Experimentally solved 3D structures of alternative isoforms are rare; therefore, to observe differences in 3D structure between isoforms, we generated AlphaFold2 predictions.⁵⁵ TFs are enriched for intrinsically disordered regions (IDRs),⁵⁶ which have recently been implicated in phase separation^{57,58} and DNA binding.⁵⁹ Alternative isoforms had more residues in predicted IDRs than did reference isoforms (Figure S2E).

We assessed protein-DNA interactions (PDIs), transcriptional regulatory activities, and PPIs for isoforms in TFIso1.0 (Figure 2D; STAR Methods). We tested each function individually to see how they vary independently, which is rarely possible in endogenous contexts. We assessed TF-DNA binding using enhanced yeast one-hybrid (eY1H) assays,^{4,60} where each TF isoform was tested against a collection of 330 DNA-baits consisting of developmental enhancer or promoter elements (Data S2). The resulting PDI profiles include 186 DNA-baits for which PDIs were detected with at least one isoform. We assessed transcriptional regulatory activities using a modified mammalian one-hybrid (M1H) assay in HEK293T cells (Data S3), where full-length TF isoforms were tethered to a Gal4 DBD and transcriptional activity was measured by activation of the Gal4 upstream activation sequences upstream of the firefly luciferase gene (Figure S2F). We assessed PPIs using yeast two-hybrid (Y2H) assays in which each TF isoform was systematically screened against the human ORFeome v9.1, comprising 17,408 protein-coding genes,⁶¹ followed by pairwise testing of each TF isoform with all interaction partners for that TF. The resulting PPI profiles involved 253 isoforms of 87 TF genes, tested against 538 different protein partners (Data S4), with all major TF families being well-represented (Figure S2G). Binary PPI and PDI calls validated when random samples were re-tested in orthogonal assays (Figures S2H–S2K; Data S2 and S4; STAR Methods), and M1H activities were highly reproducible across biological replicates (Figure S2L).

Compared with their cognate reference isoforms, 84% of alternative isoforms differed in PPIs, 81% differed in PDIs, and 41% differed in transcriptional activity; although rarer than loss-of-function isoforms, we also observed gain-of-function isoforms in all three molecular activities (Figure 2E). Novel isoforms showed evidence of functionality in the assays at levels similar to annotated alternative isoforms (Figures 2F and S2M). Novel isoforms also showed evidence of translation in published ribosome profiling followed by sequencing (Ribo-seq) data.⁶² Among isoforms with at least 1 unique exon-exon junction, we

(F) Maximum and minimum alternative TF isoform fraction of gene expression in developmental RNA-seq data. Dashed lines define “switching” events and lowly expressed isoforms. Only isoforms for TF genes with ≥ 1 transcript per million (TPM) in ≥ 1 tissue time point are shown.

(G) Percentage of alternative isoforms that exhibit switch events, shift events, or are lowly expressed in each dataset.

(H) Left: exon diagram of *HEY2* isoforms. RD, repression domain; HLH, helix-loop-helix. Right: \log_2 TPM values (top) and isoform fractions (bottom) for *HEY2* isoforms in heart and ovary.

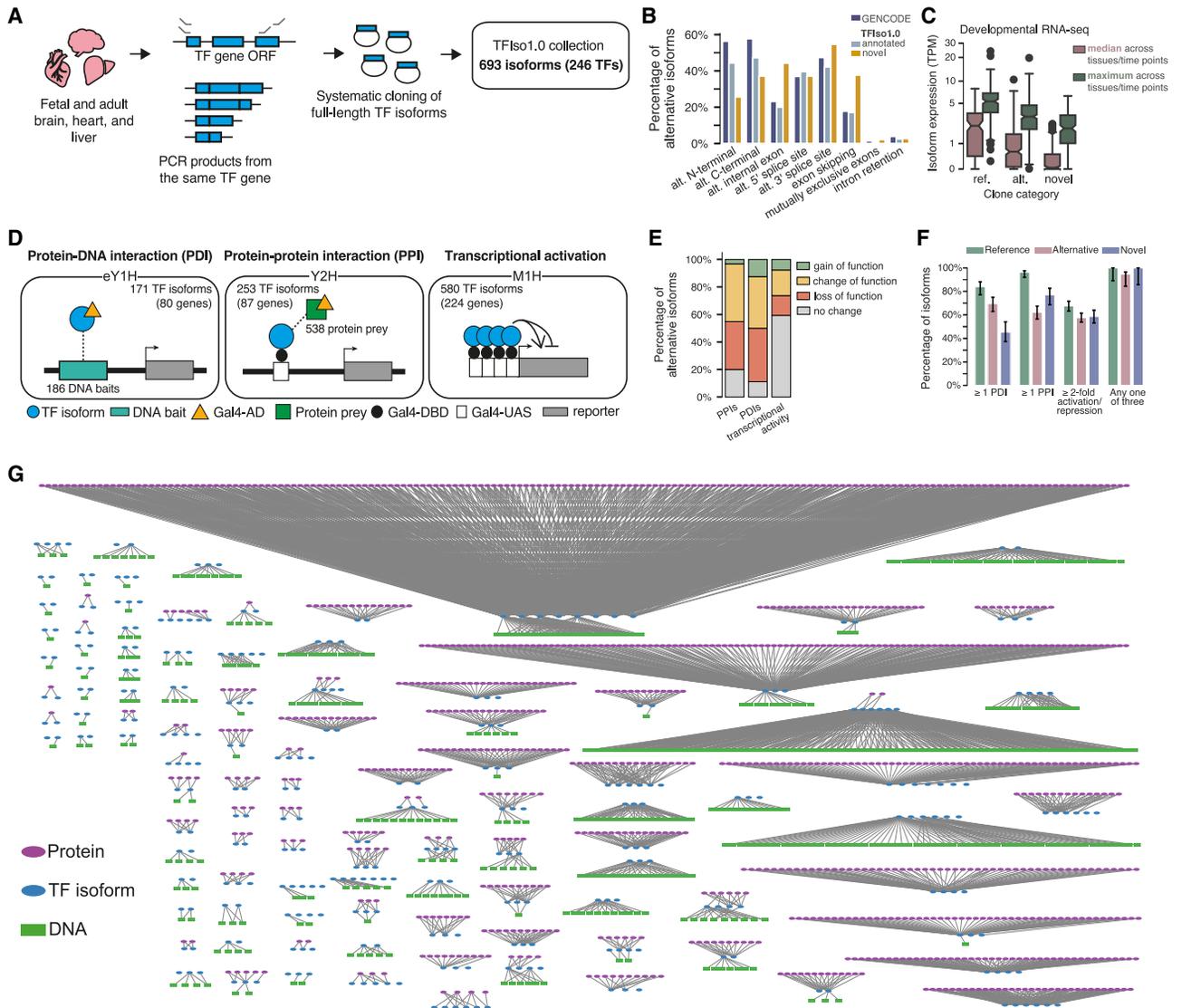
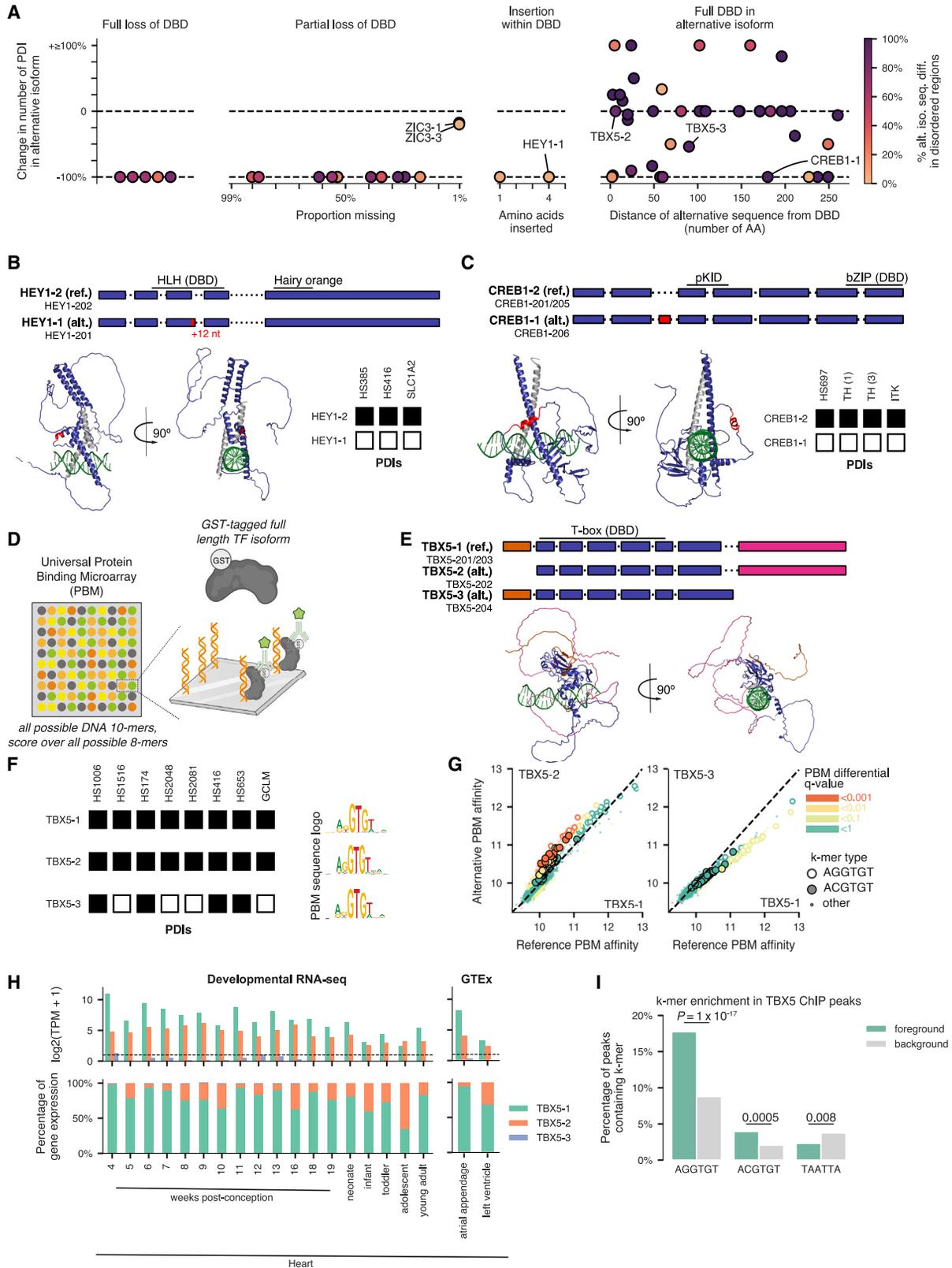


Figure 2. Overview of TFiso1.0 clone collection and TF molecular function assays

- (A) Generation of TFiso1.0.
 (B) Percentage of alternative isoforms exhibiting various sequence differences relative to reference, among all GENCODE-annotated isoforms, and TFiso1.0 annotated and novel alternative isoforms.
 (C) Median and maximum expression levels in developmental RNA-seq of reference, annotated alternative, and novel alternative isoforms in TFiso1.0.
 (D) Primary assays used in this study. eY1H, enhanced yeast one-hybrid; Y2H, yeast two-hybrid; M1H, mammalian one-hybrid; Gal4-AD, Gal4 activation domain; Gal4-DBD, Gal4 DNA-binding domain; Gal4-UAS, Gal4 upstream activation sequence.
 (E) Percent of alternative TF isoforms showing differences compared with their reference isoforms in each assay. Loss of function corresponds to loss of all PDIs or PPIs or complete loss of activation/repression in M1H; gain of function is defined reciprocally. No change corresponds to the same PDIs or PPIs or activation/repression fold-change of <2 .
 (F) Proportion of isoforms exhibiting ≥ 1 PDI, ≥ 1 PPI, ≥ 2 -fold activation/repression in M1H, or any one of the three across reference, annotated alternative, and novel alternative isoforms, normalized to the number of isoforms successfully tested in each assay. Error bars are 68.3% Bayesian CI.
 (G) PPI and PDI sub-networks profiling different TF isoforms.

found evidence of translation (≥ 5 mapped reads in at least 1 sample) of 72% (85/118) of reference isoforms, 34% (111/330) of annotated alternative isoforms and 10% (11/107) of novel isoforms (Figures S2N–S2P). This lower rate is expected as the validation rate scaled with overall RNA expression levels (Figure S2O) and because Ribo-seq data were obtained from

different cells/tissues than the ones from which novel isoforms were cloned. Indeed, we confirmed endogenous protein expression corresponding to novel isoform SP2-2 by western blotting in multiple cell lines selected based on that isoform's RNA-level expression (Figure S2Q), even though we did not detect this isoform in the Ribo-seq datasets.



(legend on next page)

Overall, we successfully assayed the PDIs, PPIs, and regulatory activities of 171, 253, and 580 TF isoforms of 80, 87, and 224 genes, respectively. Our isoform-specific PDI and PPI network shows the long-tailed degree distributions typical of biological networks,⁶³ with few isoforms binding to many interaction partners (Figure 2G). Altogether, our dataset comprises the most comprehensive characterization of TF isoforms' molecular interactions and regulatory properties reported to date.

DNA binding of alternative TF isoforms is influenced by differences both inside and outside the DBD

DNA binding is canonically achieved through structured DBDs. We therefore compared the PDI profiles of alternative TF isoforms to see how changes inside and outside the DBD affect DNA binding compared with their cognate reference TF isoforms (Data S2). Unsurprisingly, alternative TF isoforms missing the DBD completely lose the ability to bind DNA in eY1H assays (Figure 3A). Note that in eY1H assays, the isoforms are fused to an NLS; PDI loss is thus not due to mis-localization. In almost every case, alternative isoforms containing only a partial DBD also lost DNA binding. The exception is ZIC3, with two alternative isoforms that lose 3 aa of a five-zinc-finger array (Figure S3A).

Two alternative isoforms with PDI data have insertions within their DBD and both completely lose PDIs. One is an alternative isoform of HEY1 with a four aa insertion in the loop region of the basic-helix-loop-helix (bHLH) DBD, which fails to bind to any of the three DNA-baits that the reference isoform binds (Figure 3B). The loop region of MLX—a related TF—is important for stabilizing complexes of bHLH dimers.⁶⁵ Our results are consistent with previous findings that small insertions within DBDs can have strong effects on TF function.^{66,67}

Most assayed alternative TF isoforms, 42/63 (67%), contained complete, unaltered DBDs (Figure 3A, right). However, only 8/42 (19%) showed identical DNA-binding profiles to their reference isoforms, whereas 9/42 (21%) gained PDIs (Figure 3A). Sequence differences in regions close to DBDs are often associated with dramatic differences in DNA binding, consistent with evidence that flanking regions can play pivotal roles in TF-DNA binding⁶⁸; alternatively, this may suggest uncertainty in predictions of exact DBD boundaries.⁶⁹ Surprisingly, however, differ-

ences in regions far from the DBD and commonly in IDRs often affect DNA binding (Figure 3A). Indeed, 13/20 (65%) of alternative isoforms with sequence differences >100 aa from the DBD had differences in DNA binding and, of those, 69% were in disordered regions.

For example, an alternative isoform of CREB1 has an in-frame 14 aa exon inclusion, 165 aa N-terminal of the basic leucine zipper (bZIP) DBD, in a long disordered region (Figure 3C). We observed a complete loss of binding for this isoform across the 4 approximately 500- to 2,000-bp DNA sequences assayed (Figure 3C), whereas its transcriptional activity was retained, suggesting it is expressed and folded (Figure S3B). Reasoning that this loss of eY1H DNA binding might be due to differential DNA binding affinity or specificity between CREB1 isoforms, we performed *in vitro* universal protein-binding microarray (PBM) experiments using full-length CREB1 proteins (Figure 3D; Data S5; STAR Methods).^{64,70} Universal PBMs determine the relative binding affinity to all possible 8-bp sequences, allowing for higher resolution of sequence preferences.⁷¹ The alternative isoform of CREB1 showed subtly lower affinity for DNA than the reference (Figure S3C), suggesting that small differences in affinity may lead to marked changes in binding to longer DNA targets, resulting in binding signal below the sensitivity of eY1H assays. Therefore, the alternative CREB1 isoform may bind to other DNA targets not assayed here.

To investigate why some isoforms with unperturbed DBDs showed dramatic differences in DNA binding, we focused on two examples: DLX4 and PKNOX1. The reference isoforms, DLX4-1 and PKNOX1-1, were highly specific, yet alternative isoforms with truncations of large N-terminal IDRs, DLX4-2 and PKNOX1-3, showed broader binding (Figures S3D and S3E). In several TFs, IDR:DBD interaction modulates DNA-binding specificity,^{72–74} where the IDR acts as a tethered inhibitor of weaker interactions (Figure S3F). Therefore, we calculated mean-field-predicted interaction strengths between the isoform-specific IDRs and the DNA-binding residues in the homeodomains of PKNOX1 and DLX4 (STAR Methods).⁷⁵ Both reference isoforms showed strong putative intramolecular interactions between their IDRs and the DNA-binding residues, whereas the alternative isoforms showed comparatively weak interactions,

Figure 3. DNA-binding preferences of TF isoforms

(A) Change in PDIs compared with the reference isoform for alternative isoforms, categorized by the effect on the DBD. Color indicates the percentage of sequence differences between isoforms in predicted disordered protein regions.

(B) Top: exon diagrams of HEY1 isoforms with annotated Pfam domains. Red: 4 aa insertion. Bottom left: AlphaFold2 model of the HEY1 alternative isoform aligned to a structure of a homologous dimerized protein (gray) bound to DNA (green) (PDB: 4H10). Bottom right: PDIs for the 3 baits successfully assayed for both HEY1 isoforms; black box = binding, white box = no binding.

(C) Top: exon diagrams of CREB1 isoforms with annotated Pfam domains. pKID, phosphorylated kinase-inducible domain. Bottom left: AlphaFold2 model of the CREB1 alternative isoform aligned to a structure of a homodimer (gray) bound to DNA (green) (PDB: 1DH3). Bottom right: PDIs for the 4 baits successfully assayed for both CREB1 isoforms.

(D) Scores for all possible 8-mers are calculated from universal “all 10-mer” protein-binding microarrays (PBMs).

(E) Top: exon diagrams of TBX5 isoforms with annotated DBD. Bottom: AlphaFold2 model of the TBX5 reference isoform aligned to a DNA-bound structure (PDB: 5FLV).

(F) Left: PDIs for the 3 isoforms of TBX5, showing baits successfully tested against all 3 isoforms. Right: sequence logo derived from the top 50 8-mers in PBMs for each isoform.

(G) PBM affinity scores for alternative versus reference TBX5 for every 8-mer. Points are colored by differential affinity q value.⁶⁴ Open circles correspond to 8-mers containing the canonical TBX5 6-mer AGGTGT, filled circles correspond to 8-mers containing the altered 6-mer ACGTGT.

(H) Expression of TBX5 isoforms in developmental RNA-seq and GTEx in heart. Log₂ TPM values (top) and isoform fraction (bottom) for each TBX5 isoform.

(I) Enrichment of canonical TBX5 6-mer AGGTGT, altered 6-mer ACGTGT, or a negative control 6-mer TAATTA in TBX5 ChIP-seq peaks (foreground) compared with matched genomic negative control regions (background). p values from two-sided Fisher's exact tests.

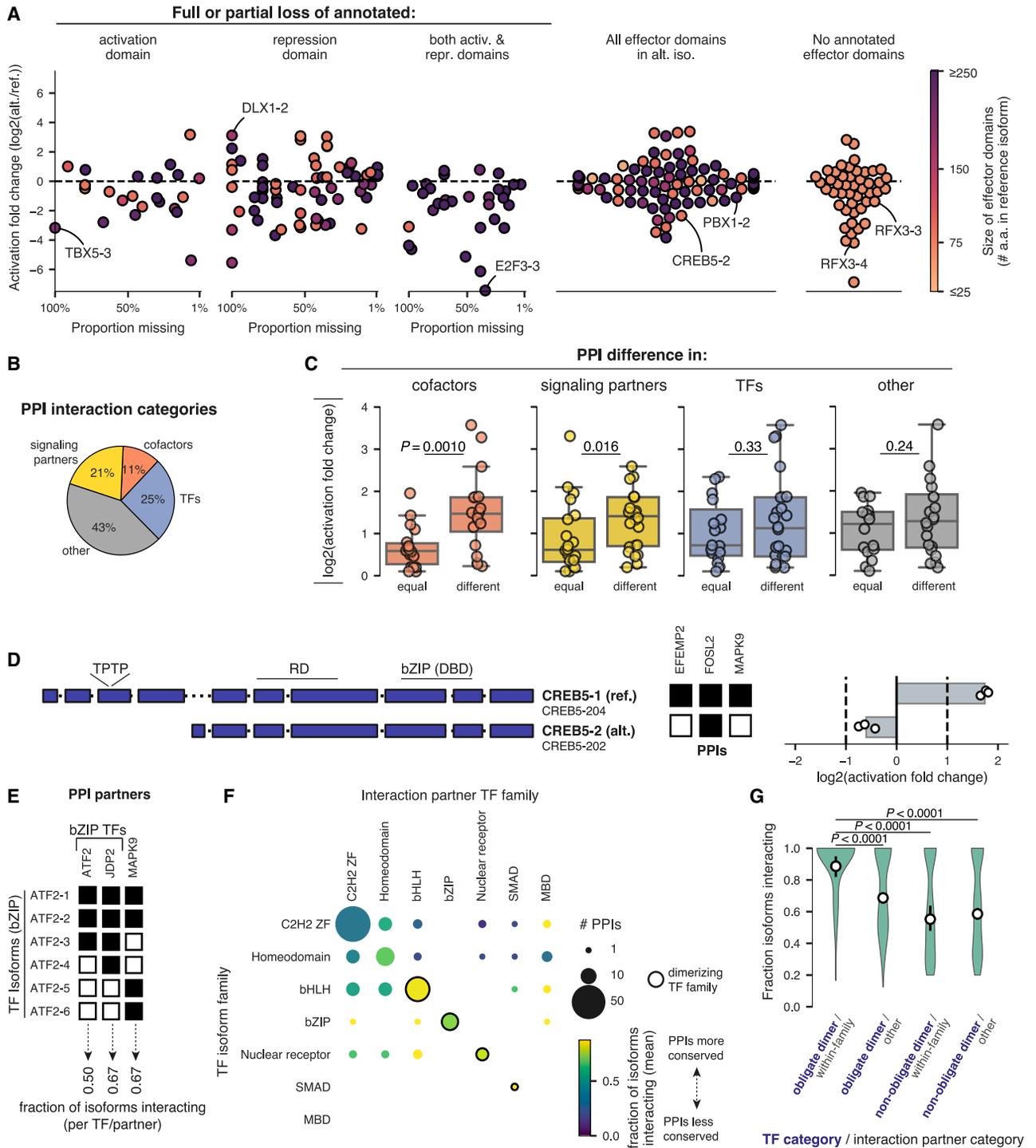


Figure 4. Transcriptional activity and protein-binding preferences of TF isoforms

(A) Change in transcriptional activity for alternative relative to reference isoform, categorized by their effect on effector domains. Points are colored by the size of annotated effector domains.

(B) Categories of PPI partners that interact with ≥ 1 TF isoform.

(C) Change in transcriptional activity associated with change in PPIs with various categories of partners. *p* values from one-sided Mann-Whitney tests.

(D) Left: exon diagrams of CREB5 isoforms. RD, repression domain. Middle: PPI results for the 2 CREB5 isoforms. Right: transcriptional activity of CREB5 isoforms.

(E) Schematic of the calculation of the fraction of isoforms interacting, using the PPI results for the partners successfully tested against all isoforms of ATF2.

(legend continued on next page)

consistent with the autoinhibitory model (Figures S3G–S3I). These results should be taken qualitatively, and many additional factors could also explain these DNA-binding differences (e.g., IDR interactions with DNA or cofactors). Nevertheless, these results provide a plausible and testable physical model for how changes outside the DBD could rewire TF DNA-binding specificity.

Another TF with differences in IDRs across isoforms is TBX5, a critical regulator of heart development,⁷⁶ with three annotated isoforms (Figure 3E). Alternative isoform TBX5-2 (TBX5e) differs from the reference isoform TBX5-1 (TBX5a)⁷⁷ in the N-terminal disordered region, adjacent to the T-box DBD. TBX5-3 differs instead in the disordered C terminus, affecting an activation domain (AD) distal to the DBD. TBX5-3 loses binding to half (4/8) of DNA-baits bound by reference TBX5-1, whereas TBX5-2 retains binding (8/8) (Figure 3F). We profiled these isoforms using universal PBMs (Figure 3G; Data S5). Consistent with eY1H assays, TBX5-3 had lower affinity for most 8-mers. Interestingly, although TBX5-2 and reference TBX5-1 have similar affinity for the highest affinity 8-mers that include the canonical TBX5 motif AGGTGT, TBX5-2 shows significantly higher affinity for a subset of moderate-affinity 8-mers. TBX5-2 is co-expressed with the reference TBX5-1 isoform throughout heart development (Figure 3H). Analysis of human TBX5 ChIP-seq data in cardiomyocytes⁷⁸ showed that these moderate-affinity 8-mers are enriched among TBX5 ChIP-seq peaks (Figures 3I and S3J), indicating that these 8-mers, preferentially bound by TBX5-2 *in vitro*, may play an important role *in vivo*. Altogether, our results support recent findings that TF IDRs modulate DNA binding, particularly in the context of a chromatinized genome,^{59,79–81} and highlight that changes in DNA binding are challenging to predict from sequence alone.

TF isoforms often differ in transcriptional activities due to changes in effector domains

Alternative isoforms can also rewire GRNs by altering transcriptional activity through changes in PPIs with cofactors, other TFs, or signaling proteins. We therefore measured differences in transcriptional activity between isoforms using M1H assays (Data S3). In M1H assays, isoforms are fused to Gal4-DBD, so transcriptional activity is decoupled from DNA binding, therefore removing confounding effects from other functions. As expected, ADs and RDs are enriched among TF isoforms with transcriptional activity above and below basal levels, respectively (Figures S4A and S4B).^{6–8} Overall, 125 alternative isoforms (49%) showed >2-fold difference in M1H activity compared with the reference isoform, with more alternative isoforms losing (89/254, 35%) rather than gaining (36/254, 14%) activity (Figure S4C). Moreover, we found that four TF genes (*FOXP3*, *MAX*, *MAZ*, and *ZNF544*) encode both activator and repressor isoforms (Figure S4D).

As expected, alternative TF isoforms with full or partial loss of ADs often showed reduced transcriptional activity ($p = 0.02$, paired two-sided Wilcoxon test; Figures 4A and S4E). In

contrast, the effect of RD loss was less clear (Figure 4A), potentially because of lower sensitivity to detect repression in M1H assays or cell-type-specific effects. Alternative isoforms that lose both ADs and RDs often lose transcriptional activity, suggesting a dominant effect of ADs (Figure 4A). For example, E2F3-2, which loses an entire AD, showed strongly decreased activity, whereas E2F3-4, which loses most of a RD, did not increase activity (Figure S4F).

Many isoforms that did not differ in any annotated effector domains showed differences in transcriptional activity (Figure 4A). For example, RFX3-4, which lacks the C-terminal domain, lost transcriptional activity, whereas RFX3-3, which largely retains this domain, also retained activity (Figure S4G). This suggests an AD in the C-terminal region of RFX3 not detected in previous tiling screens^{6,7} and highlights how profiling full-length TF isoforms can identify putative effector domains.

Changes in PPIs with cofactors and signaling proteins are associated with differences in activity between TF isoforms

Differences in transcriptional activity between isoforms likely result from PPI differences. We generated isoform-resolved PPI profiles, testing multiple isoforms of TF genes against a single isoform of their protein interaction partners (538 tested partners). We tested 3,509 isoform-resolved protein pairs for interactions, where, in each case, at least one isoform of the tested TF gene interacts with the partner, corresponding to 936 PPIs at the gene-gene level (Data S4). Of these, 684 (73%) varied across isoforms. We predicted the two interacting domains⁸² for 152 reference isoform PPIs (16%) and tested the association between domain disruption and PPI disruption in the alternative isoforms. Complete loss of the interaction domain resulted in loss of the corresponding PPIs and changes outside the domain often resulted in loss of PPIs (Figure S4H), similar to the effects of DBD disruption on PDIs (Figure 3A).

We next focused on 3 major classes of PPI partners likely to affect transcriptional activity: (1) transcriptional cofactors⁸³ (e.g., chromatin remodelers), (2) signaling partners (e.g., kinases) (STAR Methods), and (3) TFs (Figure 4B). Changes in cofactor and signaling partner binding between TF isoforms were associated with changes in transcriptional activity (Figure 4C). For example, CREB5-2 has strongly reduced activity compared with the reference isoform and loses 2 interaction partners, including the kinase MAPK9 (JNK2) (Figure 4D). Indeed, CREB5-2 is missing two conserved, phosphorylated⁸⁴ threonine-proline motifs that are substrates of JNK kinases,⁸⁵ suggesting that phosphorylation may be important for transcriptional activity of the CREB5 reference isoform, as with CREB1.⁸⁶

TFs that bind DNA as obligate dimers often maintain intra-family PPIs across isoforms

Many TFs bind DNA as dimers or multimers, such as bZIPs and bHLHs.²⁰ We therefore tested whether isoforms of these families

(F) Fraction of isoforms interacting for combinations of families of TF isoforms (y axis) and families of TF PPI partners (x axis). Black outline = families of obligate dimers. Circle size denotes number of PPIs, color denotes mean fraction of isoforms interacting. Only families with ≥ 3 TF partners are shown; for the full heatmap, see Figure S4I.

(G) Fraction of TF isoforms that retain interactions with various TF PPI partner types. p values from two-sided permutation tests.

more likely retain within-family dimerizing interactions compared with other predominantly monomeric families, such as C2H2 zinc fingers. For every TF gene and every PPI partner, we calculated the fraction of TF isoforms that interacted with the PPI partner (Figure 4E). Of the within-family TF-TF PPIs tested, 94% are heterodimers and only 6% are homodimers. On average, interactions between TFs of the same family were more often retained across isoforms for obligate dimers than other TF families (Figures 4F, 4G, and S4I). Obligate dimer TFs require these dimerizing PPIs to bind DNA; therefore, together with the observation that DBDs are often preserved across isoforms (Figure 1D), there is likely a selective pressure on alternative isoforms to retain DNA binding.

TF isoforms can be as distinct in molecular functions as TF paralogs

Gene duplication and alternative isoforms are two processes that produce novel proteins (Figure 5A). The interplay between these processes has been studied at the genome and transcriptome levels,^{87,88} but outside of a few examples,^{89–91} little is known about how these processes differentially affect molecular functions. We therefore evaluated how paralogous TFs (comparing the two reference isoforms) compare to TF isoforms (alternative versus cognate reference isoform). We found that PPI profiles, PDI profiles, and activation levels are more similar between isoforms of the same TF gene than between paralogous genes (Figures 5B–5D; Data S6). However, these observations are confounded by sequence similarity differences: paralogs vary more than isoforms (Figure 5E). After controlling for this, isoforms showed similar differences in molecular functions compared with paralogs (Figures 5F–5H and S5A–S5C).

For example, thyroid hormone receptor genes *THRA* and *THRB* evolved from an ancestral gene duplicated in vertebrates 500 million years ago.⁹² *THRA* and *THRB* share a 66.7% aa sequence identity and are particularly conserved within their DBDs and hormone receptor domains (Figure 5I). We have cloned alternative isoforms of both *THRA* and *THRB*, each with an altered effector domain (Figures 5J and 5K). We found that PDI differences were more subtle between the reference and alternative isoforms than between the paralogous reference isoforms (Figure 5L, left). In contrast, the alternative isoforms showed strong differences in transcriptional activity (Figure 5L, right). Thus, both gene duplication and alternative splicing have modulated the molecular functions of *THRA/THRB*, consistent with both mechanisms having affected this pathway in mammals.⁹³

Next, we investigated the largest family: C2H2 zinc fingers. In principle, the modular nature of zinc finger arrays could enable alternative isoforms to have different DNA-binding specificities by splicing individual zinc fingers.⁹⁴ However, alternative isoforms generally either completely preserved (67%) or completely removed (25%) the entire zinc finger array (Figures S5D and S5E). This contrasts to zinc finger TF paralogs, which show altered DNA binding due to differences in number and spacing of zinc fingers.⁶⁸

Altogether, our results support studies that found that, compared with gene duplication, alternative splicing results in sequence changes more concentrated within specific protein re-

gions and likely affect physico-chemical properties more dramatically.⁹⁵

Widespread differences in cellular localization and condensate formation between TF isoforms

In eukaryotes, gene regulation often involves nuclear condensates, which spatially sequester macromolecules into regions of local enrichment.^{96,97} This is aided by TF IDRs and is associated with increased gene activation⁵⁷ and pioneering activity.⁹⁸ We therefore tested whether TF isoforms differentially contribute to condensate formation. We expressed monomeric, enhanced green fluorescent protein (mEGFP)-tagged forms of 189 isoforms of 60 TF genes in HEK293T and U2OS cells and evaluated their subcellular localization and ability to form condensates using high-throughput confocal fluorescence microscopy,⁹⁹ focusing on TFs that showed isoform-level differences in PDIs, PPIs, or transcriptional activation (Figures 6A and 6B; Data S7).

Our observed localization of reference isoforms agreed with endogenous immunofluorescence data from the Human Protein Atlas¹⁰⁰ (Figure S6A). Additionally, there are no significant differences in endogenous expression levels between reference isoforms that form condensates and those that do not (Figure S6B), suggesting that differences in condensate formation are dependent on the protein and its cellular interactions rather than overexpression.

Half of alternative isoforms differed in condensate formation or localization in both cell lines (Figures 6C and S6C–S6G). Alternative isoforms that differed in condensate formation or localization showed differences in transcriptional activity; although they often showed differential PPIs, these differences were not significant, likely due to limited isoforms with PPI data (Figures S6H–S6J).

Overall, 19 alternative isoforms (15%) differed in condensate formation compared with their reference isoforms consistently across cell lines (Figure S6G). One example is PBX1b, a C-terminal truncation isoform of PBX1 associated with differentiation^{101–103} (Figures 6D–6H). Reference isoform PBX1a, which functions in cancer and development,^{104,105} forms nuclear condensates in both cell lines, whereas PBX1b does not (Figures 6G and S6K). To further characterize condensate formation in live cells expressing the PBX1 isoforms, we analyzed the relationship between total protein levels (determined by total GFP signal) compared with the protein level found only in the dilute phase, i.e., outside of condensates, across multiple cells displaying a range of TF-GFP expression levels.¹⁰⁶ Proteins forming condensates via phase separation have a critical threshold (saturation concentration [C_{sat}]) at which the protein becomes saturated, exhibited by the total concentration being substantially higher than the dilute concentration of protein. C_{sat} analyses confirmed that PBX1a phase separates into condensates (Figure 6H), whereas PBX1b does not. Moreover, the complement of the slope of the line above the C_{sat} , the “dominance,” is determined by whether the protein is sufficient to phase separate on its own, as indicated by a flat slope, or whether it requires other factors, as indicated by a steeper slope.¹⁰⁷ PBX1a has low dominance (Figure 6H), indicating phase separation through interactions with other factors rather than homotypic condensates. Consistent with this, PBX1a

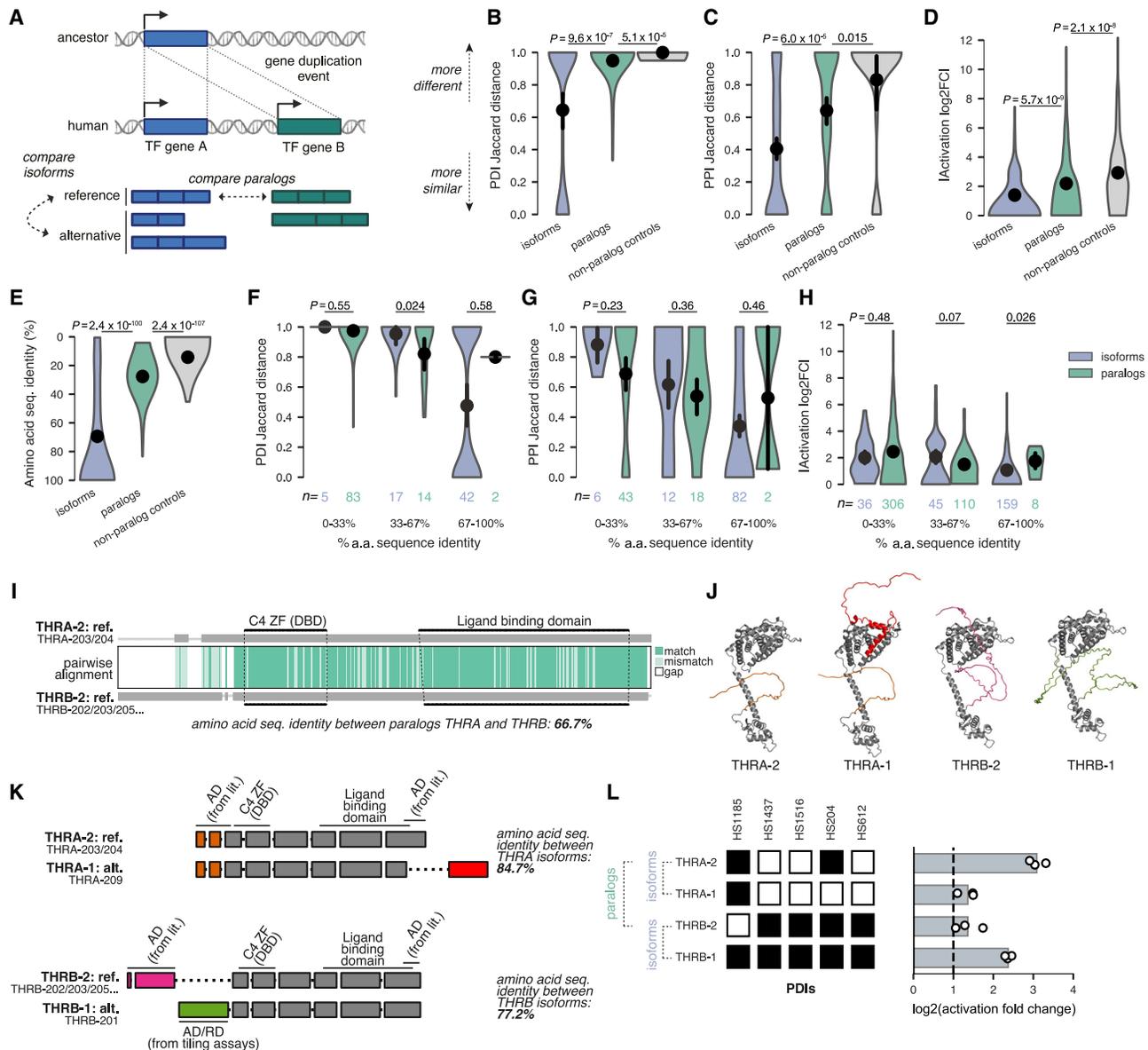
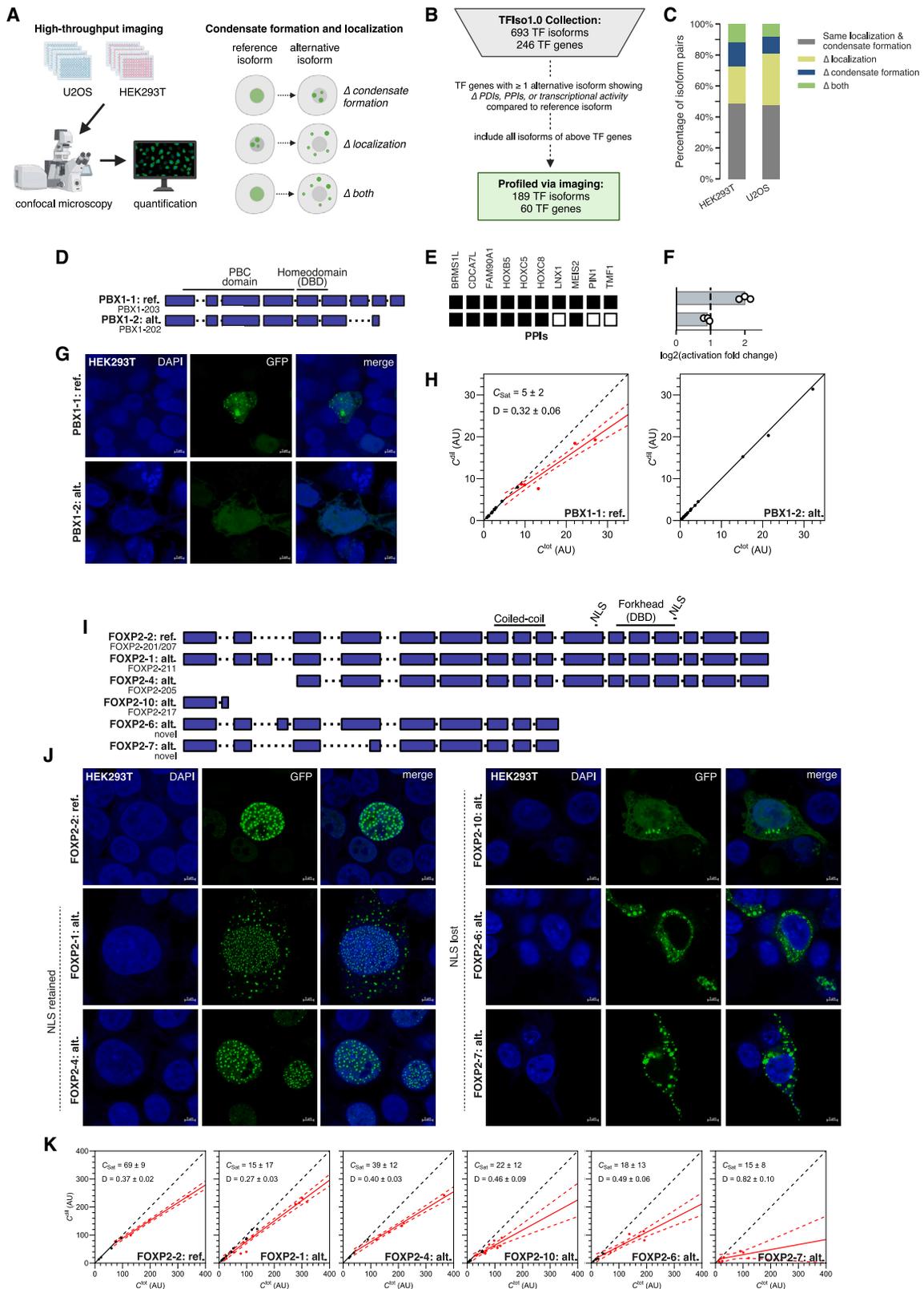


Figure 5. Functional differences between TF isoforms and TF paralogs

(A) TF paralogs compared with TF isoforms. (B and C) Jaccard distance in PDIs (B) and PPis (C) across reference/alternative isoform pairs, reference Paralog pairs, or non-paralog reference pairs as a negative control. (D) Absolute \log_2 fold-change in activation between isoforms, paralogs, and non-paralogs. (E) Amino acid sequence identity between isoforms, paralogs, and non-paralogs. Note: y axis inverted for consistency. (F–H) Analogous to (B)–(D) but with isoform and Paralog pairs binned by sequence identity. *n*, number of pairs in each bin. (I) Sequence alignment of THRA/THRB reference isoforms. (J) Predicted structures of THRA/THRB isoforms. (K) Exon diagrams of THRA/THRB isoforms. (L) Left: PDI results. Right: transcriptional activity of THRA/THRB isoforms. *p* values in (B)–(H) from two-sided Mann-Whitney tests and error bars on the mean are bootstrap calculated 95% CI.

showed interactions with three protein partners that PBX1b does not: LNX1, PIN1, and TMF1 (Figure 6E). Of these, TMF1 is a co-activator whose known partner, TRNP1, regulates nuclear condensates in neural differentiation.^{108–110} Additionally,

PBX1b shows lower transcriptional activity than PBX1a (Figure 6F). This suggests that PBX1a forms transcriptionally active nuclear condensates via interactions with protein partners not retained by PBX1b.



(legend on next page)

Alternative isoforms more often differed in localization than condensate formation (Figure 6C). For example, although all isoforms of FOXP2 formed non-homotypic condensates, their localization differed (Figures 6I–6K and S6L). The reference isoform contains two NLSs flanking its DBD.¹¹¹ Alternative isoforms missing the NLS formed cytoplasmic condensates, whereas those retaining the NLS formed nuclear condensates. However, we found few NLS annotations and no clear association between localization and NLS presence (Figure S6M).

Altogether, our data reinforce the need for expanded, isoform-aware characterization of TF condensate formation and subcellular localization, which are shaped by complex networks of macromolecular interactions.

Multi-dimensional characterization of TFs reveals two major classes of alternative isoforms: Negative regulators and rewirers

Several well-characterized alternative TF isoforms act as negative regulators of their cognate reference isoforms.^{27,35} For example, STAT3 β can compete with the oncogenic STAT3 α isoform for DNA binding but cannot activate transcription, functioning as a tumor suppressor.^{36,37,112} However, the extent to which negative regulators are representative of TF isoform function more globally is unknown.

We classified the 175 alternative isoforms with data in at least two assays (PDIs, PPIs, and transcriptional activity) into three categories: negative regulators, rewirers, and similar to reference (Figure 7A; Data S8). Negative regulators negatively affect the function of their reference isoforms (e.g., fail to bind key cofactors but bind to the same genomic targets, preventing the reference isoform from activating target genes). We therefore defined “negative regulators” as alternative isoforms that completely *lose* function in at least one assay while *retaining* function in another (Figure 7B). For example, the alternative isoform CREB1-1 fails to bind DNA but activates transcription and thus might interfere with the reference isoform by sequestering key cofactors (Figure 7C). We considered alternative isoforms with identical PDIs and PPIs and ≤ 2 -fold difference in activity to their cognate reference isoforms as “similar” and alternative isoforms otherwise different in PDI, PPI, or M1H profiles (without losing function in ≥ 1 assay) as rewirers (examples in Figures S7A and S7B). Only one isoform (PPARG-3) lost all tested functions (Figure S7C); we considered it “likely non-functional” and removed it from downstream analyses. We also considered

the subcellular localization of 126 alternative isoforms. We classified alternative isoforms whose localization changed to exclusively cytoplasmic as negative regulators; isoforms showing any other differences in localization were considered to be rewirers, unless considered negative regulators by other assays.

Of the classified alternative isoforms, 103 (59%) were negative regulators, 56 (32%) rewirers, and 15 (9%) similar to their reference isoforms (Figure 7D). Thus, most (91%) alternative isoforms differed substantially from their reference isoforms in at least 1 molecular property. Novel isoforms and annotated isoforms were distributed equally among both negative regulators and rewirers. Negative regulators lost any of the assayed TF functions (Figures 7E and S7D) and were found across all major families (Figure S7E). Intriguingly, TFs with only negative regulator alternative isoforms were more ubiquitously expressed than TFs with only rewirer alternative isoforms (Figures 7F and S7F). Overall, our results revealed that negative regulator isoforms are widespread among TFs, suggesting they may constitute an additional layer of transcriptional regulation.

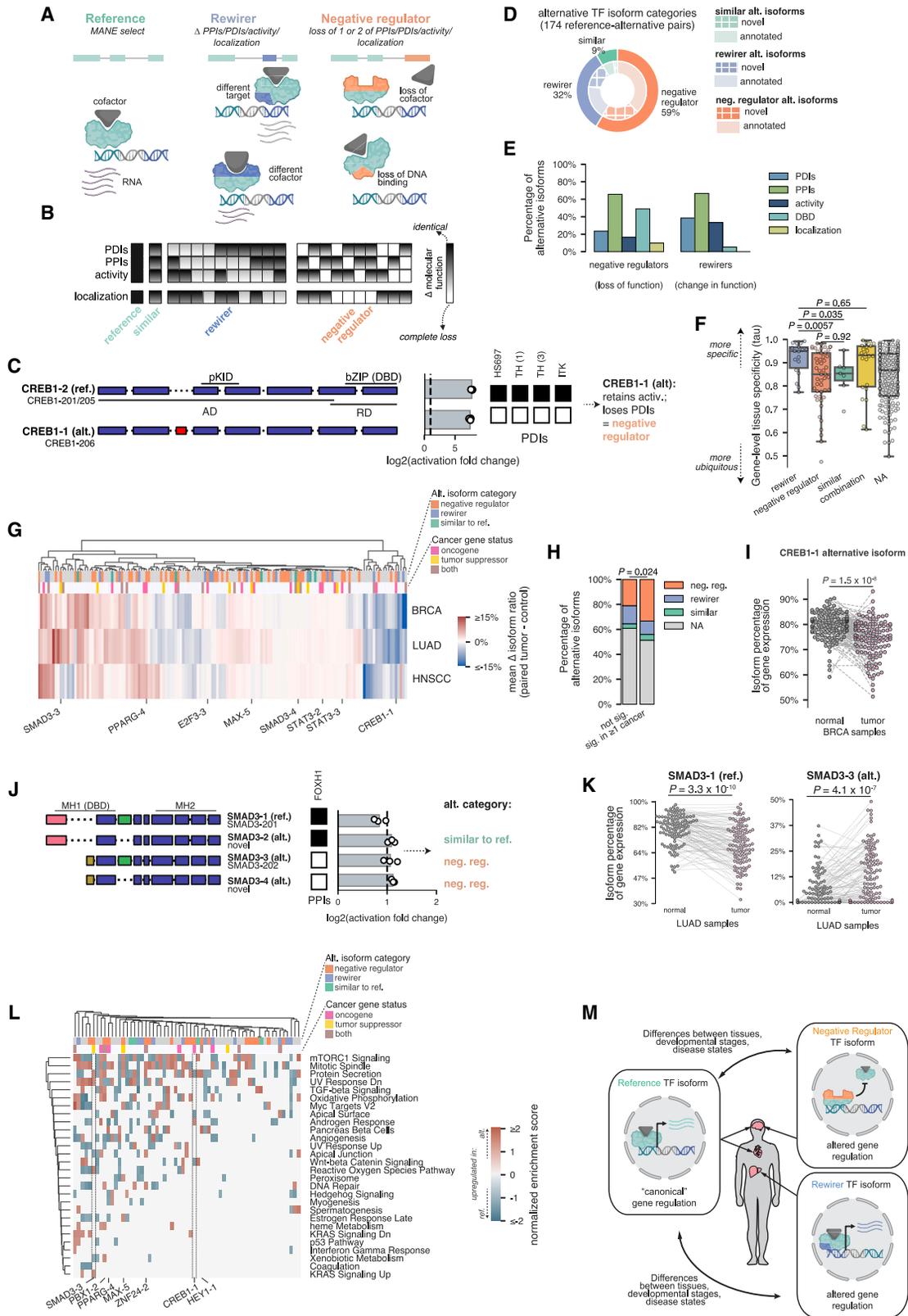
Alternative TF isoforms are associated with differentiation and cancer

TFs play important roles in development. Recently, Joung et al. performed a single-cell RNA-seq (scRNA-seq) over-expression screen of >3,000 annotated TF isoforms in human embryonic stem cells and found many that significantly affected differentiation.¹¹⁴ 220/246 (89%) of our reference and 183/447 (41%) of our alternative isoforms were included in their library. Because Joung et al. relied on gene annotations, most alternative isoforms missing in their library are our novel isoforms (168/264 [64%]). Isoforms from each category significantly affected differentiation, but the strongest effects corresponded to reference or rewirer isoforms (Figure S7G).

Because many well-characterized TF isoforms are negative regulators dysregulated in cancer,³⁵ we examined isoform expression in three cancers with paired tumor/normal samples in The Cancer Genome Atlas (TCGA): breast cancer (BRCA), lung adenocarcinoma (LUAD), and head and neck squamous cell cancer (HNSCC) (Data S9; STAR Methods). 314 isoforms in TFIs1.0 showed significant differential abundance between paired tumor and normal samples in at least one cancer type (Benjamini-Hochberg adjusted $p < 0.05$, two-sided paired Wilcoxon test, Figure 7G; Data S9); most were alternative isoforms (188/314 [60%]). Interestingly, negative regulator isoforms were

Figure 6. Condensate formation and subcellular localization differences between TF isoforms

- (A) High-throughput imaging pipeline.
- (B) Selection of tested isoforms.
- (C) Percent of alternative isoforms with differences in localization, condensate formation, both, or neither, compared with reference in HEK293T and U2OS cells.
- (D) Exon diagram of PBX1 isoforms in TFIs1.0.
- (E) PPI results for PBX1 isoforms.
- (F) Transcriptional activation of PBX1 isoforms.
- (G) Representative images of PBX1 isoform expression in HEK293T cells (63 \times magnification).
- (H) Saturation (C_{sat}) curve analysis of PBX1 isoforms. Dots represent individual cells, x axis shows total protein concentration from fluorescence (C^{tot}), y axis shows concentration in the dilute phase (C^{dil}). Arbitrary units (a.u.) are at reference settings. C_{sat} , saturation concentration; D, dominance.
- (I) Exon diagram of FOXP2 isoforms in TFIs1.0. NLS, nuclear localization sequence.
- (J) Representative images of FOXP2 isoform expression in HEK293T cells (63 \times magnification).
- (K) C_{sat} analysis of FOXP2 isoforms.



(legend on next page)

significantly enriched among isoforms differentially abundant in cancer (Figure 7H), suggesting that misregulation of negative regulator TF isoforms plays specific roles in aberrant cancer GRNs.

The alternative isoform of oncogene *CREB1* (*CREB1-1*)¹¹⁵ is a negative regulator significantly misregulated in cancer; *CREB1-1* strongly activated transcription but lost all PDIs (Figure 7C). Intriguingly, although the overall levels of *CREB1* gene expression were similar (Figures S7H and S7I), the alternative isoform was significantly downregulated in breast tumors compared with the reference (Figures 7I and S7J). Moreover, both *CREB1* isoforms were ubiquitously expressed (Figure S7K). This suggests that the reference isoform of *CREB1* acts as an oncogene, but the alternative isoform may have tumor-suppressive properties, potentially acting as a dominant negative.

More alternative isoforms were upregulated in tumors (101) than downregulated (74) (Figure 7G). One example is tumor suppressor *SMAD3*, for which the reference isoform is significantly downregulated in LUAD tumors, whereas the negative regulator isoform, which loses interaction with *SMAD3*'s known partner *FOXH1*¹¹⁶ and loses part of its DBD, is upregulated (Figures 7J, 7K, and S7L). This suggests that the alternative isoform of *SMAD3* may have oncogenic properties.

To evaluate the effects TF isoforms have on the transcriptome, we re-analyzed scRNA-seq profiles for 120 reference/alternative isoform pairs overlapping between TFiso1.0 and the Joung et al. over-expression screen, using a meta-cell aggregation strategy to control for variance in cell counts, and performed gene set enrichment analyses between isoforms (STAR Methods). Almost all of our cancer-associated alternative isoforms present in the over-expression dataset significantly affected at least one pathway when compared with their reference isoforms (64/66, 97%) (Figure 7L). For example, *KRAS* signaling is significantly upregulated by the alternative *SMAD3* isoform, which we predict is oncogenic. Conversely, *Wnt* signaling is significantly downregulated by the alternative *CREB1* isoform, which we predict is tu-

mor suppressive. Thus, our molecular profiling approach highlighted TF isoforms with varying functions and differential effects on downstream expression.

Altogether, our comprehensive, multi-dimensional characterization of hundreds of isoforms reveals that TF loci encode alternative proteoforms that fall into two primary categories: rewirers, which behave distinctly from their cognate reference isoforms, and negative regulators, which can act independently or in competition with their cognate reference isoforms, depending on their expression profiles (Figure 7M).

DISCUSSION

To understand the functional differences between TF isoforms, we generated a collection of 693 TF isoforms across 246 TF genes and systematically assayed their DNA binding, PPIs, transcriptional activity, localization, and condensate formation and integrated these data with public expression datasets. We provide these integrated results via tfisodb.org as a community resource.

We used exogenous assays to perform this survey, finding that two-thirds of alternative isoforms differed from their cognate reference isoforms in at least one molecular function. To achieve this isoform-level resolution at scale in endogenous contexts will require improved approaches for assaying isoform-specific TF occupancies *in vivo* and proteome-wide techniques for assaying PPIs. However, an advantage of our approach is the ability to disentangle how sequence differences affect TF functions that can be interconnected in the endogenous context. For example, a difference in genomic occupancy between isoforms by ChIP-seq may result from differences in DNA binding or differences in PPIs with required co-binding partners. Here, we isolated each variable at a time when assaying TF function (e.g., fusing TF isoforms to the Gal4-AD in eY1H assays, such that DNA-binding differences can be resolved separately from differences in cofactor binding).

Figure 7. Alternative TF isoforms can function as negative regulators

- (A) Example rewirer and negative regulator TF isoforms.
- (B) Illustrative heatmap of alternative isoform classification as either similar to the reference, rewirers, or negative regulators.
- (C) Example of a negative regulator (*CREB1-1*). Left: exon diagram. Middle: activation results. Right: PDI results.
- (D) Number of alternative isoforms per category (outer circle) with the numbers of annotated (solid colors) and novel (hatched colors) isoforms.
- (E) Percent of alternative isoforms that either show loss of function (negative regulators) or change in function (rewirers) in an assay, compared with their reference isoforms. For full details see Figure S7D.
- (F) Gene-level tissue specificities¹¹³ calculated from developmental RNA-seq among TF genes with alternative isoforms that are: all rewirers, all negative regulators, all similar to reference, a mixture of categories, or all unable to be classified (NA). *p* values from two-sided Mann-Whitney tests.
- (G) Differential abundance of TF isoforms in breast cancer (BRCA), lung adenocarcinoma (LUAD), and head and neck squamous cell cancer (HNSCC). Colors correspond to the mean difference in isoform abundance across samples, where red indicates increased abundance in tumors compared with controls. Top rows show alternative isoform categories (gray = NA) and the cancer status of the TF gene (white = none, both = annotated as both oncogene and tumor suppressor).
- (H) Percentage of alternative isoform categories that are significantly differentially abundant in at least one of the three profiled cancers compared with those that are not. *p* value from two-sided Fisher's exact test.
- (I) *CREB1* alternative isoform fraction in matched BRCA tumor and normal samples (paired from the same patient, denoted using dotted lines). *p* value from two-sided paired Wilcoxon test with multiple hypothesis correction.
- (J) *SMAD3* isoforms, with exon diagrams, categorizations, PPIs, and transcriptional activity.
- (K) Isoform fraction of reference *SMAD3-1* and alternative *SMAD3-3* in matched lung cancer tumor and normal samples. *p* value from two-sided paired Wilcoxon test with multiple hypothesis correction.
- (L) Pathways that are significantly different between alternative and reference isoforms in the Joung et al. over-expression dataset, colored by gene set enrichment analysis (GSEA)-normalized enrichment scores.
- (M) Example mechanism. Whereas rewirers lead to altered GRNs, negative regulators can lead to misregulation of canonical GRNs in the absence or presence of the reference isoform. Negative regulators that outcompete their reference isoforms in the same cell can be thought of as naturally occurring dominant negatives.

Molecular functions of TF isoforms can be difficult to predict from sequence and predicted structure alone. Indeed, differences in regions far from annotated DBDs (Figure 3A) and effector domains (Figure 4A) can affect TF functions. This is consistent with studies showing that IDRs can affect TF occupancy *in vivo*^{59,79} and underscores the importance of experiments using full-length proteins to characterize TF functions, complementing studying DBD-only constructs for DNA-binding^{3,44} and tiling-based peptide assays for transcriptional activity.^{6,7}

The extent to which natural dominant-negative isoforms exist within the context of the “TFome” has remained unclear. We present evidence that negative regulator isoforms are widespread among TFs and often misregulated in cancer (Figures 7D and 7H). Given that ubiquitously expressed TFs often have negative regulator isoforms (Figure 7F), we propose that many negative regulator isoforms will exert dominant-negative effects through diverse mechanisms. Sometimes, they may interfere with reference isoform function (e.g., by sequestering key binding partners). However, more complex scenarios are plausible, such as negative regulators playing distinct, non-TF roles in the cellular milieu. Future studies characterizing how negative regulators compete with their cognate reference isoforms may determine whether they act as true dominant negatives. However, our findings are consistent with decades-old ideas that negative regulators contribute to human disease.^{117,118}

Most human TFs have duplicated and diverged throughout evolution, expanding GRNs.^{119,120} Alternative isoforms also increase TF diversity but their effects on GRNs are less understood. We reveal that TF isoforms can behave as distinctly as paralogous TFs across all major molecular functions and thus likely expand GRN complexity alongside TF paralogs.

In summary, our high-throughput exogenous assays shed light on the functional diversity within the human TFome due to alternative isoforms. Multiple consortia are focused on measuring functional effects at the gene level^{121,122}; we propose that characterizing the functions of alternative isoforms will be key to both understanding disease mechanisms and accurate interpretation of the effect of genomic variants. Altogether, our work highlights the importance of moving beyond gene-level resolution toward a more complex, proteoform-aware characterization of TF function.

Limitations of the study

Interactions were tested in non-physiological contexts, enabling isoform-level resolution in high throughput but entailing trade-offs, with some differences possibly being missed or not relevant in the endogenous context. For example, PPIs dependent on post-translational modifications could be missed. Our high-throughput microscopy cannot assess condensate material state. Methods to address this, such as microrheology and fluorescence recovery after photobleaching (FRAP), are difficult to scale.¹²³

The value of two-thirds of alternative TF isoforms differing from reference in at least one molecular function is likely an underestimate of the true fraction due to technical limitations of our assays. For example, further differences in DNA binding may be revealed by testing additional DNA-baits in eY1H assays, or

cell-type-specific differences in transcriptional activity might be uncovered by performing M1H assays in additional cell lines or conditions. Other TF molecular functions, such as ligand binding and RNA binding, were not explored. Finally, we tested protein isoforms, ignoring differences in UTRs, which can affect mRNA stability and protein expression.¹²⁴

RESOURCE AVAILABILITY

Lead contact

Requests for information should be directed to Juan I. Fuxman Bass (fuxman@bu.edu).

Materials availability

All unique/stable reagents generated in this study will be made available from the [lead contact](#) with a completed material transfer agreement.

Data and code availability

- Data are available at [tfisodb.org](https://www.ncbi.nlm.nih.gov/geo/). PBM data are available in GEO at accession GEO: GSE253638.
- Original code is available at github.com/CCSB-DFCI/TF_isoforms_paper and archived at Zenodo: <https://doi.org/10.5281/zenodo.14969075>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank Norman Davey for sharing expertise on disorder prediction, Yves Janin for providing the furimazine substrate, members of CCSB, and Bulyk and Fuxman Bass labs for helpful discussions. Schematics were created with [BioRender.com](https://www.bio-render.com/). Results are partially based upon data generated by TCGA Research Network: <https://www.cancer.gov/tcga>. This work was supported by NIH grants U01CA232161 (M.V., M.L.B., and J.I.F.B.), U24HG011451 (M.V., D.E.H., and A.F.), R35GM128625 (J.I.F.B.), R01CA226802 (N.S.), R35GM142647 (G.S.), R35GM137836 (N. Sahni), DP2-CA290639 (A.S.H.), F32HG012318-01 (K.M.), K99HG013345 (K.M.), and T32CA009361 (G.S. and N. Sahni). Additional awards include Charles A. King Trust Postdoctoral Research Fellowship (G.S.), Melanoma Research Foundation Career Development Award (G.S.), Belgian American Educational Foundation doctoral fellowship (F.L.), Wallonia-Brussels International (WBI)-World Excellence fellowship (F.L.), Belgian Télévie grant FC31747 (F.L. and J.-C.T.), the Fondation Léon Fredericq (F.L. and J.-C.T.), Fonds de la Recherche Scientifique (FRS – FNRS) Mobility and Congress funding (F.L.), Canadian Institutes for Health Research Foundation grant FDN-159926 (M.G.), and Deborah F. Allinger Fellowship (L.L.). N. Sahni is a CPRIT Scholar in Cancer Research (CPRIT New Investigator Grant RR160021) and is supported by the Andrew Sabin Family Foundation. J.A.R. is a CPRIT Scholar in Cancer Research (CPRIT New Investigator Grant RR210040). Confocal microscopy was performed at MDACC EMC Flow Cytometry and Cell Imaging Core (CPRIT core facility grant RP170628). M.V. is a Chercheur Qualifié Honoraire, and J.-C.T. is a Maître de Recherche from the Fonds de la Recherche Scientifique (FRS – FNRS, Wallonia-Brussels Federation, Belgium). HepG2 cells were a gift from Wolfram Goessling.

AUTHOR CONTRIBUTIONS

Conceptualization, L.L., K.M., C.S., G.S., S.I., N. Salomonis, D.E.H., M.V., M.L.B., and J.I.F.B.; methodology, L.L., K.M., C.S., G.S., M.A.C., D.E.H., N. Sahni, M.V., M.L.B., and J.I.F.B.; software, L.L., K.M., G.S., S.I., A. Bhattacharjee, and G.L.; validation, C.S., G.S., K.S.-F., F.L., D.C.M., T. Hunt, D. Balcha, and A.F.; formal analysis, L.L., K.M., S.I., S.P.P., R.J.E., A.S.H., J.A.R., N. Salomonis, and N. Sahni; investigation, C.S., G.S., S.I., B.K., A. Berenson, K.S.-F., E.R., S.S., F.L., B.S.C., Z.Y., D. Bisht, J.A.S., A.P., S.P., R.L., and M.G.; resources, N. Sahni, M.V., M.L.B., and J.I.F.B.; data curation, L.L., K.M., C.S.,

G.S., T. Hao, and N. Salomonis; writing – original draft, L.L., K.M., C.S., G.S., S.I., and J.I.F.B.; writing – review and editing, L.L., K.M., C.S., M.A.C., D.E.H., N. Sahni, M.V., M.L.B., and J.I.F.B.; visualization, L.L. and K.M.; supervision, K.M., J.-C.T., N. Salomonis, M.A.C., D.E.H., N. Sahni, M.V., M.L.B., and J.I.F.B.; project administration, L.L., K.M., D.E.H., N. Sahni, M.L.B., and J.I.F.B.; and funding acquisition, A.F., D.E.H., M.V., M.L.B., and J.I.F.B.

DECLARATION OF INTERESTS

M.L.B. is a co-inventor on US patents #6,548,021 and #8,530,638 on PBM technology and corresponding universal sequence designs, respectively. Universal PBM array designs used in this study are available via a materials transfer agreement with The Brigham & Women’s Hospital, Inc. M.L.B. is a member of the *Molecular Cell* advisory board.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Yeast strains
 - Human cell lines and cell culture
- [METHOD DETAILS](#)
 - TFIs0.1.0 clone collection generation and validation
 - TF isoform annotations
 - Western blots of novel TF isoforms
 - Detection of protein-DNA interactions using enhanced yeast one-hybrid assays
 - Validation of yeast one-hybrid protein-DNA interaction data using luciferase assays
 - Protein-DNA interaction assay using protein-binding microarrays
 - Protein-protein interaction assay with yeast two-hybrid (Y2H)
 - Protein-protein interaction validation using mammalian NanoLuc two-hybrid (mN2H)
 - Transcriptional activity using mammalian one-hybrid assays
 - Condensate formation assay
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
 - Gene Annotation
 - Protein domain annotation
 - Proportion of alternative isoforms with domain affected
 - AlphaFold structural prediction
 - Predicted disorder values
 - PPI partner classification
 - Domain-domain PPI annotation
 - RNA-seq analyses
 - Re-sampling GTEx data
 - Ribo-seq analyses
 - Pairwise sequence identity analyses
 - Paralogs definition
 - Functional assay quantification
 - Violin plots
 - PBM analysis
 - IDR:DBD inhibition analysis
 - TBX5 ChIP-seq analysis
 - Csat analysis
 - Human Protein Atlas localization validation
 - Classification of negative regulators and rewirers
 - TF Atlas mORF analyses
 - Paired tumor/normal TCGA analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2025.03.004>.

Received: March 18, 2024
Revised: December 6, 2024
Accepted: March 5, 2025
Published: March 26, 2025

REFERENCES

1. Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature* 335, 683–689. <https://doi.org/10.1038/335683a0>.
2. Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* 20, 861–873. <https://doi.org/10.1101/gr.100552.109>.
3. Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339. <https://doi.org/10.1016/j.cell.2012.12.009>.
4. Fuxman Bass, J.I., Sahni, N., Shrestha, S., Garcia-Gonzalez, A., Mori, A., Bhat, N., Yi, S., Hill, D.E., Vidal, M., and Walhout, A.J.M. (2015). Human gene-centered transcription factor networks for enhancers and disease variants. *Cell* 161, 661–673. <https://doi.org/10.1016/j.cell.2015.03.003>.
5. Barrera, L.A., Vedenko, A., Kurland, J.V., Rogers, J.M., Gisselbrecht, S.S., Rossin, E.J., Woodard, J., Mariani, L., Kock, K.H., Inukai, S., et al. (2016). Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science* 351, 1450–1454. <https://doi.org/10.1126/science.aad2257>.
6. Tycko, J., DelRosso, N., Hess, G.T., Aradhana, B., Banerjee, A., Mukund, A., Van, M.V., Ego, B.K., Yao, D., Spees, K., et al. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. *Cell* 183, 2020–2035.e16. <https://doi.org/10.1016/j.cell.2020.11.024>.
7. DelRosso, N., Tycko, J., Suzuki, P., Andrews, C., Aradhana, M., Mukund, A., Liongson, I., Ludwig, C., Spees, K., Fordyce, P., et al. (2023). Large-scale mapping and mutagenesis of human transcriptional effector domains. *Nature* 616, 365–372. <https://doi.org/10.1038/s41586-023-05906-y>.
8. 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. <https://doi.org/10.1016/j.molcel.2021.11.007>.
9. Grove, C.A., De Masi, F., Barrasa, M.I., Newburger, D.E., Alkema, M.J., Bulyk, M.L., and Walhout, A.J.M. (2009). A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell* 138, 314–327. <https://doi.org/10.1016/j.cell.2009.04.058>.
10. Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. *Cell* 140, 744–752. <https://doi.org/10.1016/j.cell.2010.01.044>.
11. Göös, H., Kinnunen, M., Salokas, K., Tan, Z., Liu, X., Yadav, L., Zhang, Q., Wei, G.-H., and Varjosalo, M. (2022). Human transcription factor protein interaction networks. *Nat. Commun.* 13, 766. <https://doi.org/10.1038/s41467-022-28341-5>.
12. Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. <https://doi.org/10.1038/nature08909>.
13. Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. *Cell* 126, 37–47. <https://doi.org/10.1016/j.cell.2006.06.023>.
14. Aebersold, R., Agar, J.N., Amster, I.J., Baker, M.S., Bertozzi, C.R., Boja, E.S., Costello, C.E., Cravatt, B.F., Fenselau, C., Garcia, B.A., et al. (2018). How many human proteoforms are there? *Nat. Chem. Biol.* 14, 206–214. <https://doi.org/10.1038/nchembio.2576>.

15. Cahan, P., Li, H., Morris, S.A., Lummerz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. *Cell* 158, 903–915. <https://doi.org/10.1016/j.cell.2014.07.020>.
16. Sahni, N., Yi, S., Taipale, M., Fuxman Bass, J.I., Coulombe-Huntington, J., Yang, F., Peng, J., Weile, J., Karras, G.I., Wang, Y., et al. (2015). Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* 161, 647–660. <https://doi.org/10.1016/j.cell.2015.04.013>.
17. Scarpato, M., Federico, A., Ciccodicola, A., and Costa, V. (2015). Novel transcription factor variants through RNA-sequencing: the importance of being “alternative.”. *Int. J. Mol. Sci.* 16, 1755–1771. <https://doi.org/10.3390/ijms16011755>.
18. Talavera, D., Orozco, M., and de la Cruz, X. (2009). Alternative splicing of transcription factors’ genes: beyond the increase of proteome diversity. *Comp. Funct. Genomics* 2009, 905894. <https://doi.org/10.1155/2009/905894>.
19. Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. <https://doi.org/10.1038/nature07509>.
20. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. *Cell* 172, 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>.
21. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 47, D766–D773. <https://doi.org/10.1093/nar/gky955>.
22. Deveson, I.W., Brunck, M.E., Blackburn, J., Tseng, E., Hon, T., Clark, T.A., Clark, M.B., Crawford, J., Dinger, M.E., Nielsen, L.K., et al. (2018). Universal Alternative Splicing of Noncoding Exons. *Cell Syst.* 6, 245–255.e5. <https://doi.org/10.1016/j.cels.2017.12.005>.
23. Sheynkman, G.M., Tuttle, K.S., Laval, F., Tseng, E., Underwood, J.G., Yu, L., Dong, D., Smith, M.L., Sebra, R., Willems, L., et al. (2020). ORF Capture-Seq as a versatile method for targeted identification of full-length isoforms. *Nat. Commun.* 11, 2326. <https://doi.org/10.1038/s41467-020-16174-z>.
24. Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* 136, 777–793. <https://doi.org/10.1016/j.cell.2009.02.011>.
25. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774. <https://doi.org/10.1101/gr.135350.111>.
26. Sinitcyn, P., Richards, A.L., Weatheritt, R.J., Brademan, D.R., Marx, H., Shishkova, E., Meyer, J.G., Hebert, A.S., Westphall, M.S., Blencowe, B.J., et al. (2023). Global detection of human variants and isoforms by deep proteome sequencing. *Nat. Biotechnol.* 41, 1776–1786. <https://doi.org/10.1038/s41587-023-01714-x>.
27. López, A.J. (1995). Developmental role of transcription factor isoforms generated by alternative splicing. *Dev. Biol.* 172, 396–411. <https://doi.org/10.1006/dbio.1995.8050>.
28. Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. *Gene* 514, 1–30. <https://doi.org/10.1016/j.gene.2012.07.083>.
29. Taneri, B., Snyder, B., Novoradovsky, A., and Gaasterland, T. (2004). Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. *Genome Biol.* 5, R75. <https://doi.org/10.1186/gb-2004-5-10-r75>.
30. Vuzman, D., and Levy, Y. (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. Biosyst.* 8, 47–57. <https://doi.org/10.1039/c1mb05273j>.
31. Li, J., Wang, Y., Rao, X., Wang, Y., Feng, W., Liang, H., and Liu, Y. (2017). Roles of alternative splicing in modulating transcriptional regulation. *BMC Syst. Biol.* 11, 89. <https://doi.org/10.1186/s12918-017-0465-6>.
32. Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O’Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* 147, 132–146. <https://doi.org/10.1016/j.cell.2011.08.023>.
33. Barbaux, S., Niaudet, P., Gubler, M.C., Grünfeld, J.P., Jaubert, F., Kuttann, F., Fékété, C.N., Souleyreau-Therville, N., Thibaud, E., Fellous, M., et al. (1997). Donor splice-site mutations in WT1 are responsible for Frasier syndrome. *Nat. Genet.* 17, 467–470. <https://doi.org/10.1038/ng1297-467>.
34. Bickmore, W.A., Oghene, K., Little, M.H., Seawright, A., van Heyningen, V., and Hastie, N.D. (1992). Modulation of DNA binding specificity by alternative splicing of the Wilms tumor wt1 gene transcript. *Science* 257, 235–237. <https://doi.org/10.1126/science.1321494>.
35. Belluti, S., Rigillo, G., and Imbriano, C. (2020). Transcription Factors in Cancer: When Alternative Splicing Determines Opposite Cell Fates. *Cells* 9, 760. <https://doi.org/10.3390/cells9030760>.
36. Schaefer, T.S., Sanders, L.K., Park, O.K., and Nathans, D. (1997). Functional differences between Stat3alpha and Stat3beta. *Mol. Cell Biol.* 17, 5307–5316. <https://doi.org/10.1128/MCB.17.9.5307>.
37. Zhang, H.-X., Yang, P.-L., Li, E.-M., and Xu, L.-Y. (2019). STAT3beta, a distinct isoform from STAT3. *Int. J. Biochem. Cell Biol.* 110, 130–139. <https://doi.org/10.1016/j.biocel.2019.02.006>.
38. Flouriot, G., Brand, H., Denger, S., Metivier, R., Kos, M., Reid, G., Sonntag-Buck, V., and Gannon, F. (2000). Identification of a new isoform of the human estrogen receptor-alpha (hER- α) that is encoded by distinct transcripts and that is able to repress hER- α activation function 1. *EMBO J.* 19, 4688–4700. <https://doi.org/10.1093/emboj/19.17.4688>.
39. Shi, L., Dong, B., Li, Z., Lu, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T., Lin, B., et al. (2009). Expression of ER- α 36, a novel variant of estrogen receptor α , and resistance to tamoxifen treatment in breast cancer. *J. Clin. Oncol.* 27, 3423–3429. <https://doi.org/10.1200/JCO.2008.17.2254>.
40. Gadea, G., Arsic, N., Fernandes, K., Diot, A., Jorruiz, S.M., Abdallah, S., Meuray, V., Vinot, S., Anguille, C., Remenyi, J., et al. (2016). TP53 drives invasion through expression of its Δ 133p53 β variant. *eLife* 5, e14734. <https://doi.org/10.7554/eLife.14734>.
41. Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y.A., Murray, R.R., et al. (2016). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* 164, 805–817. <https://doi.org/10.1016/j.cell.2016.01.029>.
42. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804. <https://doi.org/10.1126/science.1075090>.
43. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. <https://doi.org/10.1016/j.cell.2014.08.009>.
44. Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science* 324, 1720–1723. <https://doi.org/10.1126/science.1162327>.
45. Mukund, A.X., Tycko, J., Allen, S.J., Robinson, S.A., Andrews, C., Sinha, J., Ludwig, C.H., Spees, K., Bassik, M.C., and Bintu, L. (2023). High-throughput functional characterization of combinations of transcriptional activators and repressors. *Cell Syst.* 14, 746–763.e5. <https://doi.org/10.1016/j.cels.2023.07.001>.

46. Morales, J., Pujar, S., Loveland, J.E., Astashyn, A., Bennett, R., Berry, A., Cox, E., Davidson, C., Ermolaeva, O., Farrell, C.M., et al. (2022). A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* 604, 310–315. <https://doi.org/10.1038/s41586-022-04558-8>.
47. Kriventseva, E.V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand, M.S., and Sunyaev, S. (2003). Increase of functional diversity by alternative splicing. *Trends Genet.* 19, 124–128. [https://doi.org/10.1016/S0168-9525\(03\)00023-4](https://doi.org/10.1016/S0168-9525(03)00023-4).
48. Alekseyenko, A.V., Kim, N., and Lee, C.J. (2007). Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* 13, 661–670. <https://doi.org/10.1261/rna.325107>.
49. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. <https://doi.org/10.1126/science.1262110>.
50. Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., et al. (2019). Gene expression across mammalian organ development. *Nature* 571, 505–509. <https://doi.org/10.1038/s41586-019-1338-5>.
51. Raj, B., O'Hanlon, D., Vessey, J.P., Pan, Q., Ray, D., Buckley, N.J., Miller, F.D., and Blencowe, B.J. (2011). Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol. Cell* 43, 843–850. <https://doi.org/10.1016/j.molcel.2011.08.014>.
52. Wilanowski, T., Tuckfield, A., Cerruti, L., O'Connell, S., Saint, R., Parekh, V., Tao, J., Cunningham, J.M., and Jane, S.M. (2002). A highly conserved novel family of mammalian developmental transcription factors related to *Drosophila* grainyhead. *Mech. Dev.* 114, 37–50. [https://doi.org/10.1016/S0925-4773\(02\)00046-1](https://doi.org/10.1016/S0925-4773(02)00046-1).
53. Weber, D., Wiese, C., and Gessler, M. (2014). Hey bHLH transcription factors. *Curr. Top. Dev. Biol.* 110, 285–315. <https://doi.org/10.1016/B978-0-12-405943-6.00008-7>.
54. Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). Human genomics. The human transcriptome across tissues and individuals. *Science* 348, 660–665. <https://doi.org/10.1126/science.aaa0355>.
55. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
56. Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. *Biochemistry* 45, 6873–6888. <https://doi.org/10.1021/bi0602718>.
57. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842–1855.e16. <https://doi.org/10.1016/j.cell.2018.10.042>.
58. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, eaar3958. <https://doi.org/10.1126/science.aar3958>.
59. Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D.K., Carmi, M., and Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol. Cell* 79, 459–471.e4. <https://doi.org/10.1016/j.molcel.2020.05.032>.
60. 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.M. (2011). Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat. Methods* 8, 1059–1064. <https://doi.org/10.1038/nmeth.1748>.
61. 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. <https://doi.org/10.1038/s41586-020-2188-x>.
62. Chothani, S.P., Adami, E., Widjaja, A.A., Langley, S.R., Viswanathan, S., Pua, C.J., Zhihao, N.T., Harmston, N., D'Agostino, G., Whiffin, N., et al. (2022). A high-resolution map of human RNA translation. *Mol. Cell* 82, 2885–2899.e8. <https://doi.org/10.1016/j.molcel.2022.06.023>.
63. Albert, R. (2005). Scale-free networks in cell biology. *J. Cell Sci.* 118, 4947–4957. <https://doi.org/10.1242/jcs.02714>.
64. Berger, M.F., and Bulyk, M.L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* 4, 393–411. <https://doi.org/10.1038/nprot.2008.195>.
65. Ma, L., Sham, Y.Y., Walters, K.J., and Towle, H.C. (2007). A critical role for the loop region of the basic helix-loop-helix/leucine zipper protein Mix in DNA binding and glucose-regulated transcription. *Nucleic Acids Res.* 35, 35–44. <https://doi.org/10.1093/nar/gkl987>.
66. Englert, C., Vidal, M., Maheswaran, S., Ge, Y., Ezzell, R.M., Isselbacher, K.J., and Haber, D.A. (1995). Truncated WT1 mutants alter the subnuclear localization of the wild-type protein. *Proc. Natl. Acad. Sci. USA* 92, 11960–11964. <https://doi.org/10.1073/pnas.92.26.11960>.
67. Larsson, S.H., Charlieu, J.P., Miyagawa, K., Engelkamp, D., Rassoulzadegan, M., Ross, A., Cuzin, F., van Heyningen, V., and Hastie, N.D. (1995). Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell* 81, 391–401. [https://doi.org/10.1016/0092-8674\(95\)90392-5](https://doi.org/10.1016/0092-8674(95)90392-5).
68. Siggers, T., Reddy, J., Barron, B., and Bulyk, M.L. (2014). Diversification of Transcription Factor Paralogs via Noncanonical Modularity in C2H2 Zinc Finger DNA Binding. *Mol. Cell* 55, 640–648. <https://doi.org/10.1016/j.molcel.2014.06.019>.
69. Eddy, S.R. (2011). Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7, e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
70. Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., 3rd, and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* 24, 1429–1435. <https://doi.org/10.1038/nbt1246>.
71. Kock, K.H., Kimes, P.K., Gisselbrecht, S.S., Inukai, S., Phanor, S.K., Anderson, J.T., Ramakrishnan, G., Lipper, C.H., Song, D., Kurland, J.V., et al. (2023). DNA binding analysis of rare variants in homeodomains reveals novel homeodomain specificity-determining residues. Preprint at bioRxiv. <https://doi.org/10.1101/2023.06.16.545320>.
72. He, F., Borchers, W., Song, T., Wei, X., Das, M., Chen, L., Daughdrill, G.W., and Chen, J. (2019). Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. *Proc. Natl. Acad. Sci. USA* 116, 8859–8868. <https://doi.org/10.1073/pnas.1903077116>.
73. Krois, A.S., Dyson, H.J., and Wright, P.E. (2018). Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. USA* 115, E11302–E11310. <https://doi.org/10.1073/pnas.1814051115>.
74. Bjarnason, S., Mclvor, J.A.P., Prestel, A., Demény, K.S., Bullerjahn, J.T., Kragelund, B.B., Mercadante, D., and Heidarsson, P.O. (2024). DNA binding redistributes activation domain ensemble and accessibility in pioneer factor Sox2. *Nat. Commun.* 15, 1445. <https://doi.org/10.1038/s41467-024-45847-2>.
75. Ginell, G.M., Emenecker, R.J., Lotthammer, J.M., Usher, E.T., and Holehouse, A.S. (2024). Direct prediction of intermolecular interactions driven by disordered regions. Preprint at bioRxiv. <https://doi.org/10.1101/2024.06.03.597104>.
76. Steimle, J.D., and Moskowitz, I.P. (2017). TBX5: A Key Regulator of Heart Development. *Curr. Top. Dev. Biol.* 122, 195–221. <https://doi.org/10.1016/bs.ctdb.2016.08.008>.
77. Yamak, A., Georges, R.O., Sheikh-Hassani, M., Morin, M., Komati, H., and Nemer, M. (2015). Novel exons in the *tbx5* gene locus generate

- protein isoforms with distinct expression domains and function. *J. Biol. Chem.* 290, 6844–6856. <https://doi.org/10.1074/jbc.M114.634451>.
78. Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J., and Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* 19, e46255. <https://doi.org/10.15252/embr.201846255>.
 79. 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. <https://doi.org/10.1016/j.sbi.2021.06.011>.
 80. Garcia, D.A., Johnson, T.A., Presman, D.M., Fettweis, G., Wagh, K., Rinaldi, L., Stavreva, D.A., Paakinaho, V., Jensen, R.A.M., Mandrup, S., et al. (2021). An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. *Mol. Cell* 81, 1484–1498.e6. <https://doi.org/10.1016/j.molcel.2021.01.013>.
 81. Kumar, D.K., Jonas, F., Jana, T., Brodsky, S., Carmi, M., and Barkai, N. (2023). Complementary strategies for directing in vivo transcription factor binding through DNA binding domains and intrinsically disordered regions. *Mol. Cell* 83, 1462–1473.e5. <https://doi.org/10.1016/j.molcel.2023.04.002>.
 82. Mosca, R., Céol, A., Stein, A., Olivella, R., and Aloy, P. (2014). 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 42, D374–D379. <https://doi.org/10.1093/nar/gkt887>.
 83. Shen, W.-K., Chen, S.-Y., Gan, Z.-Q., Zhang, Y.-Z., Yue, T., Chen, M.-M., Xue, Y., Hu, H., and Guo, A.-Y. (2023). AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res.* 51, D39–D45. <https://doi.org/10.1093/nar/gkac907>.
 84. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 43, D512–D520. <https://doi.org/10.1093/nar/gku1267>.
 85. Zhang, C.-H., Gao, Y., Jadhav, U., Hung, H.-H., Holton, K.M., Grodzinsky, A.J., Shivdasani, R.A., and Lassar, A.B. (2021). Creb5 establishes the competence for Prg4 expression in articular cartilage. *Commun. Biol.* 4, 332. <https://doi.org/10.1038/s42003-021-01857-0>.
 86. Naqvi, S., Martin, K.J., and Arthur, J.S.C. (2014). CREB phosphorylation at Ser133 regulates transcription via distinct mechanisms downstream of cAMP and MAPK signalling. *Biochem. J.* 458, 469–479. <https://doi.org/10.1042/BJ20131115>.
 87. Kopelman, N.M., Lancet, D., and Yanai, I. (2005). Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat. Genet.* 37, 588–589. <https://doi.org/10.1038/ng1575>.
 88. Grishkevich, V., and Yanai, I. (2014). Gene length and expression level shape genomic novelties. *Genome Res.* 24, 1497–1503. <https://doi.org/10.1101/gr.169722.113>.
 89. Lister, J.A., Close, J., and Raible, D.W. (2001). Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev. Biol.* 237, 333–344. <https://doi.org/10.1006/dbio.2001.0379>.
 90. Hultman, K.A., Bahary, N., Zon, L.I., and Johnson, S.L. (2007). Gene Duplication of the zebrafish kit ligand and partitioning of melanocyte development functions to kit ligand a. *PLOS Genet.* 3, e17. <https://doi.org/10.1371/journal.pgen.0030017>.
 91. Marshall, A.N., Montealegre, M.C., Jiménez-López, C., Lorenz, M.C., and van Hoof, A. (2013). Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. *PLoS Genet.* 9, e1003376. <https://doi.org/10.1371/journal.pgen.1003376>.
 92. Laudet, V. (2011). The origins and evolution of vertebrate metamorphosis. *Curr. Biol.* 21, R726–R737. <https://doi.org/10.1016/j.cub.2011.07.030>.
 93. Ortiga-Carvalho, T.M., Sidhaye, A.R., and Wondisford, F.E. (2014). Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nat. Rev. Endocrinol.* 10, 582–591. <https://doi.org/10.1038/nrendo.2014.143>.
 94. Hsu, T., Gogos, J.A., Kirsh, S.A., and Kafatos, F.C. (1992). Multiple zinc finger forms resulting from developmentally regulated alternative splicing of a transcription factor gene. *Science* 257, 1946–1950. <https://doi.org/10.1126/science.1411512>.
 95. Talavera, D., Vogel, C., Orozco, M., Teichmann, S.A., and de la Cruz, X. (2007). The (in)dependence of alternative splicing and gene duplication. *PLoS Comput. Biol.* 3, e33. <https://doi.org/10.1371/journal.pcbi.0030033>.
 96. Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* 573, 45–54. <https://doi.org/10.1038/s41586-019-1517-4>.
 97. Yang, J., Chung, C.-I., Koach, J., Liu, H., Navalkar, A., He, H., Ma, Z., Zhao, Q., Yang, X., He, L., et al. (2024). MYC phase separation selectively modulates the transcriptome. *Nat. Struct. Mol. Biol.* 31, 1567–1579. <https://doi.org/10.1038/s41594-024-01322-6>.
 98. Ji, D., Shao, C., Yu, J., Hou, Y., Gao, X., Wu, Y., Wang, L., and Chen, P. (2024). FOXA1 forms biomolecular condensates that unpack condensed chromatin to function as a pioneer factor. *Mol. Cell* 84, 244–260.e7. <https://doi.org/10.1016/j.molcel.2023.11.020>.
 99. Tripathi, S., Shirnekhi, H.K., Gorman, S.D., Chandra, B., Baggett, D.W., Park, C.-G., Somjee, R., Lang, B., Hosseini, S.M.H., Pioso, B.J., et al. (2023). Defining the condensate landscape of fusion oncoproteins. *Nat. Commun.* 14, 6008. <https://doi.org/10.1038/s41467-023-41655-2>.
 100. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. *Science* 356, eaal3321. <https://doi.org/10.1126/science.aal3321>.
 101. Linares, A.J., Lin, C.-H., Damianov, A., Adams, K.L., Novitch, B.G., and Black, D.L. (2015). The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. *eLife* 4, e09268. <https://doi.org/10.7554/eLife.09268>.
 102. Remesal, L., Roger-Baynat, I., Chirivella, L., Maicas, M., Brocal-Ruiz, R., Pérez-Villalba, A., Cucarella, C., Casado, M., and Flames, N. (2020). PBX1 acts as terminal selector for olfactory bulb dopaminergic neurons. *Development* 147, dev186841. <https://doi.org/10.1242/dev.186841>.
 103. Xu, Y., Zhao, W., Olson, S.D., Prabhakara, K.S., and Zhou, X. (2018). Alternative splicing links histone modifications to stem cell fate decision. *Genome Biol.* 19, 133. <https://doi.org/10.1186/s13059-018-1512-3>.
 104. Veiga, R.N., de Oliveira, J.C., and Gradia, D.F. (2021). PBX1: a key character of the hallmarks of cancer. *J. Mol. Med. (Berl)* 99, 1667–1680. <https://doi.org/10.1007/s00109-021-02139-2>.
 105. Mary, L., Leclerc, D., Gilot, D., Belaud-Rotureau, M.-A., and Jaillard, S. (2022). The TALE never ends: A comprehensive overview of the role of PBX1, a TALE transcription factor, in human developmental defects. *Hum. Mutat.* 43, 1125–1148. <https://doi.org/10.1002/humu.24388>.
 106. Riback, J.A., Zhu, L., Ferrolino, M.C., Tolbert, M., Mitrea, D.M., Sanders, D.W., Wei, M.-T., Kriwacki, R.W., and Brangwynne, C.P. (2020). Composition-dependent thermodynamics of intracellular phase separation. *Nature* 581, 209–214. <https://doi.org/10.1038/s41586-020-2256-2>.
 107. Qian, D., Ausserwoger, H., Arter, W.E., Scrutton, R.M., Welsh, T.J., Kartanas, T., Ermann, N., Qamar, S., Fischer, C., Sneideris, T., et al. (2023). Linking modulation of bio-molecular phase behaviour with collective interactions. Preprint at bioRxiv. <https://doi.org/10.1101/2023.11.02.565376>.
 108. Hsiao, P.W., and Chang, C. (1999). Isolation and characterization of ARA160 as the first androgen receptor N-terminal-associated coactivator in human prostate cells. *J. Biol. Chem.* 274, 22373–22379. <https://doi.org/10.1074/jbc.274.32.22373>.
 109. Volpe, M., Shpungin, S., Barbi, C., Abrahm, G., Malovani, H., Wides, R., and Nir, U. (2006). trnp: A conserved mammalian gene encoding a

- nuclear protein that accelerates cell-cycle progression. *DNA Cell Biol.* 25, 331–339. <https://doi.org/10.1089/dna.2006.25.331>.
110. Esgleas, M., Falk, S., Forné, I., Thiry, M., Najas, S., Zhang, S., Mas-Sanchez, A., Geerlof, A., Niessing, D., Wang, Z., et al. (2020). Trnp1 organizes diverse nuclear membrane-less compartments in neural stem cells. *EMBO J.* 39, e103373. <https://doi.org/10.15252/embj.2019103373>.
 111. Mizutani, A., Matsuzaki, A., Momoi, M.Y., Fujita, E., Tanabe, Y., and Momoi, T. (2007). Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochem. Biophys. Res. Commun.* 353, 869–874. <https://doi.org/10.1016/j.bbrc.2006.12.130>.
 112. Zammarchi, F., de Stanchina, E., Bournazou, E., Supakorndej, T., Martires, K., Riedel, E., Corben, A.D., Bromberg, J.F., and Cartegni, L. (2011). Antitumorigenic potential of STAT3 alternative splicing modulation. *Proc. Natl. Acad. Sci. USA* 108, 17779–17784. <https://doi.org/10.1073/pnas.1108482108>.
 113. Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* 18, 205–214. <https://doi.org/10.1093/bib/bbw008>.
 114. Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., et al. (2023). A transcription factor atlas of directed differentiation. *Cell* 186, e26.
 115. Sapiro, L., Salzillo, A., Ragone, A., Illiano, M., Spina, A., and Naviglio, S. (2020). Targeting CREB in Cancer Therapy: A Key Candidate or One of Many? *Cancers* 12, 3166. <https://doi.org/10.3390/cancers12113166>.
 116. Aragón, E., Wang, Q., Zou, Y., Morgani, S.M., Ruiz, L., Kaczmarek, Z., Su, J., Torner, C., Tian, L., Hu, J., et al. (2019). Structural Basis for Distinct Roles of SMAD2 and SMAD3 in FOXP1 Pioneer-Directed TGF- β Signaling. *Genes & Development* 33, 1506–1524. <https://doi.org/10.1101/gad.330837.119>.
 117. Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature* 329, 219–222. <https://doi.org/10.1038/329219a0>.
 118. Koenig, R.J., Lazar, M.A., Hodin, R.A., Brent, G.A., Larsen, P.R., Chin, W.W., and Moore, D.D. (1989). Inhibition of thyroid hormone action by a non-hormone binding c-erbA protein generated by alternative mRNA splicing. *Nature* 337, 659–661. <https://doi.org/10.1038/337659a0>.
 119. Zhang, Y., Ho, T.D., Buchler, N.E., and Gordân, R. (2021). Competition for DNA binding between paralogous transcription factors determines their genomic occupancy and regulatory functions. *Genome Res.* 31, 1216–1229. <https://doi.org/10.1101/gr.275145.120>.
 120. Feng, S., Rastogi, C., Loker, R., Glassford, W.J., Tomas Rube, H., Bussemaker, H.J., and Mann, R.S. (2022). Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors. *Nat. Commun.* 13, 3808. <https://doi.org/10.1038/s41467-022-31501-2>.
 121. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. *Cell* 170, 564–576.e16. <https://doi.org/10.1016/j.cell.2017.06.010>.
 122. IGVF Consortium (2024). Deciphering the impact of genomic variation on function. *Nature* 633, 47–57. <https://doi.org/10.1038/s41586-024-07510-0>.
 123. Lee, D.S.W., Strom, A.R., and Brangwynne, C.P. (2022). The mechanobiology of nuclear phase separation. *APL Bioeng.* 6, 021503. <https://doi.org/10.1063/5.0083286>.
 124. Spies, N., Burge, C.B., and Bartel, D.P. (2013). 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res.* 23, 2078–2090. <https://doi.org/10.1101/gr.156919.113>.
 125. Churko, J.M., Garg, P., Treutlein, B., Venkatasubramanian, M., Wu, H., Lee, J., Wessells, Q.N., Chen, S.-Y., Chen, W.-Y., Chetal, K., et al. (2018). Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. *Nat. Commun.* 9, 4906. <https://doi.org/10.1038/s41467-018-07333-4>.
 126. Ang, Y.-S., Rivas, R.N., Ribeiro, A.J.S., Srivas, R., Rivera, J., Stone, N.R., Pratt, K., Mohamed, T.M.A., Fu, J.-D., Spencer, C.I., et al. (2016). Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* 167, 1734–1749.e22. <https://doi.org/10.1016/j.cell.2016.11.033>.
 127. Gonzalez-Teran, B., Pittman, M., Felix, F., Thomas, R., Richmond-Buccola, D., Hüttenhain, R., Choudhary, K., Moroni, E., Costa, M.W., Huang, Y., et al. (2022). Transcription factor protein interactomes reveal genetic determinants in heart disease. *Cell* 185, 794–814.e30. <https://doi.org/10.1016/j.cell.2022.01.021>.
 128. Rual, J.-F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.-O., et al. (2004). Human ORFome version 1.1: a platform for reverse proteomics. *Genome Res.* 14, 2128–2135. <https://doi.org/10.1101/gr.2973604>.
 129. Santoso, C.S., Li, Z., Lal, S., Yuan, S., Gan, K.A., Agosto, L.M., Liu, X., Pro, S.C., Sewell, J.A., Henderson, A., et al. (2020). Comprehensive mapping of the human cytokine gene regulatory network. *Nucleic Acids Res.* 48, 12055–12073. <https://doi.org/10.1093/nar/gkaa1055>.
 130. Lu, Y., Berenson, A., Lane, R., Guelin, I., Li, Z., Chen, Y., Shah, S., Yin, M., Soto-Ugaldi, L.F., Fiszbein, A., et al. (2024). A large-scale cancer-specific protein-DNA interaction network. *Life Sci. Alliance* 7, e202402641. <https://doi.org/10.26508/lsa.202402641>.
 131. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Erratum: near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 888. <https://doi.org/10.1038/nbt0816-888d>.
 132. Mariani, L., Weinand, K., Vedenko, A., Barrera, L.A., and Bulyk, M.L. (2017). Identification of Human Lineage-Specific Transcriptional Coregulators Enabled by a Glossary of Binding Modules and Tunable Genomic Backgrounds. *Cell Syst.* 5, 654. <https://doi.org/10.1016/j.cels.2017.12.011>.
 133. Fang, Z., Liu, X., and Peltz, G. (2023). GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39, btac757. <https://doi.org/10.1093/bioinformatics/btac757>.
 134. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
 135. Emig, D., Salomonis, N., Baumbach, J., Lengauer, T., Conklin, B.R., and Albrecht, M. (2010). AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res.* 38, W755–W762. <https://doi.org/10.1093/nar/gkq405>.
 136. Li, G., Mahajan, S., Ma, S., Jeffery, E.D., Zhang, X., Bhattacharjee, A., Venkatasubramanian, M., Weirauch, M.T., Miraldi, E.R., Grimes, H.L., et al. (2024). Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy. *Sci. Transl. Med.* 16, eade2886. <https://doi.org/10.1126/scitranslmed.ade2886>.
 137. 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. <https://doi.org/10.1038/nmeth.1764>.
 138. Wingender, E., Schoeps, T., Haubrock, M., Krull, M., and Dönitz, J. (2018). TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 46, D343–D347. <https://doi.org/10.1093/nar/gkx987>.
 139. Salehi-Ashtiani, K., Yang, X., Derti, A., Tian, W., Hao, T., Lin, C., Makowski, K., Shen, L., Murray, R.R., Szeto, D., et al. (2008). Isoform discovery by targeted cloning, 'deep-well' pooling and parallel sequencing. *Nat. Methods* 5, 597–600. <https://doi.org/10.1038/nmeth.1224>.

140. Frankish, A., Carbonell-Sala, S., Diekhans, M., Jungreis, I., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Arnan, C., Barnes, I., et al. (2023). GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res.* *51*, D942–D949. <https://doi.org/10.1093/nar/gkac1071>.
141. Nellore, A., Jaffe, A.E., Fortin, J.-P., Alquicira-Hernández, J., Collado-Torres, L., Wang, S., Phillips, R.A., III, Karbhari, N., Hansen, K.D., Langmead, B., et al. (2016). Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. *Genome Biol.* *17*, 266. <https://doi.org/10.1186/s13059-016-1118-6>.
142. Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberte, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., et al.; FANTOM Consortium; the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. *Nature* *507*, 462–470.
143. Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* *27*, i275–i282. <https://doi.org/10.1093/bioinformatics/btr209>.
144. 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. <https://doi.org/10.1038/s41467-023-42445-6>.
145. Berger, M.F., and Bulyk, M.L. (2006). Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. *Methods Mol. Biol.* *338*, 245–260. <https://doi.org/10.1385/1-59745-097-9:245>.
146. Dreze, M., Monachello, D., Lurin, C., Cusick, M.E., Hill, D.E., Vidal, M., and Braun, P. (2010). High-quality binary interactome mapping. *Methods Enzymol.* *470*, 281–315. [https://doi.org/10.1016/S0076-6879\(10\)70012-4](https://doi.org/10.1016/S0076-6879(10)70012-4).
147. Schmeier, S., Alam, T., Essack, M., and Bajic, V.B. (2017). TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. *Nucleic Acids Res.* *45*, D145–D150. <https://doi.org/10.1093/nar/gkw1007>.
148. Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J.X., Kreisberg, R., Kauffman, S.A., Huang, S., and Shmulevich, I. (2013). Gene-pair expression signatures reveal lineage control. *Nat. Methods* *10*, 577–583. <https://doi.org/10.1038/nmeth.2445>.
149. Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.-I., et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* *6*, 83–90. <https://doi.org/10.1038/nmeth.1280>.
150. Choi, S.G., Olivet, J., Cassonnet, P., Vidalain, P.-O., Luck, K., Lambourne, L., Spirohn, K., Lemmens, I., Dos Santos, M., Demeret, C., et al. (2019). Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* *10*, 3907. <https://doi.org/10.1038/s41467-019-11809-2>.
151. Rodriguez, J.M., Pozo, F., Cerdán-Vélez, D., Di Domenico, T., Vázquez, J., and Tress, M.L. (2022). APPRIS: selecting functionally important isoforms. *Nucleic Acids Res.* *50*, D54–D59. <https://doi.org/10.1093/nar/gkab1058>.
152. Akdel, M., Pires, D.E.V., Pardo, E.P., Jänes, J., Zalevsky, A.O., Mészáros, B., Bryant, P., Good, L.L., Laskowski, R.A., Pozzati, G., et al. (2022). A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* *29*, 1056–1067. <https://doi.org/10.1038/s41594-022-00849-w>.
153. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* *22*, 2577–2637. <https://doi.org/10.1002/bip.360221211>.
154. Tien, M.Z., Meyer, A.G., Sydykova, D.K., Spielman, S.J., and Wilke, C.O. (2013). Maximum Allowed Solvent Accessibilities of Residues in Proteins. *PLoS One* *8*, e80635. <https://doi.org/10.1371/journal.pone.0080635>.
155. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* *356*, eaaj2239. <https://doi.org/10.1126/science.aaj2239>.
156. McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.-P., Lane, T.J., and Pande, V.S. (2015). MDTraj: a modern, open library for the analysis of molecular dynamics trajectories. *Biophys. J.* *109*, 1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015>.
157. Lotthammer, J.M., Ginell, G.M., Griffith, D., Emenecker, R.J., and Holehouse, A.S. (2024). Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat. Methods* *21*, 465–476. <https://doi.org/10.1038/s41592-023-02159-5>.
158. Joseph, J.A., Reinhardt, A., Aguirre, A., Chew, P.Y., Russell, K.O., Espinosa, J.R., Garaizar, A., and Collepardo-Guevara, R. (2021). Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nat. Comput. Sci.* *1*, 732–743. <https://doi.org/10.1038/s43588-021-00155-3>.
159. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* *596*, 590–596. <https://doi.org/10.1038/s41586-021-03828-1>.
160. Emenecker, R.J., Griffith, D., and Holehouse, A.S. (2021). Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* *120*, 4312–4319. <https://doi.org/10.1016/j.bpj.2021.08.039>.
161. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
162. Pohl, A., and Beato, M. (2014). bwtool: a tool for bigWig files. *Bioinformatics* *30*, 1618–1619. <https://doi.org/10.1093/bioinformatics/btu056>.
163. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissán, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* *2017*, PO.17.00011. <https://doi.org/10.1200/PO.17.00011>.
164. Suehnholz, S.P., Nissan, M.H., Zhang, H., Kundra, R., Nandakumar, S., Lu, C., Carrero, S., Dhaneshwar, A., Fernandez, N., Xu, B.W., et al. (2024). Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. *Cancer Discov.* *14*, 49–65. <https://doi.org/10.1158/2159-8290.CD-23-0467>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-GST Monoclonal Antibody [3G10/1B3]	Abcam	Cat# AB92; RRID: AB_307067
Polyclonal Anti-GST Tag Antibody Produced in Rabbit	Sigma	Cat# G7781; RRID: AB_259965
Stabilized Peroxidase Conjugated Goat Anti-Mouse (H + L)	Invitrogen	Cat# 32430; RRID: AB_1185566
Anti-GST Alexa Fluor 488 Conjugate	Invitrogen	Cat# A11131; RRID: AB_143157
Anti-SP2 Polyclonal Antibody	Proteintech	Cat# 25000-1-AP; RRID: AB_2879839
Anti-GAPDH Monoclonal Antibody (14C10)	Cell Signaling Technology	Cat# 2118T; RRID: AB_561053
Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	Invitrogen	Cat# 31460; RRID: AB_228341
Bacterial and virus strains		
DH5 α Chemically Competent Cells	Made in house	N/A
TOP10 Chemically Competent Cells	Thermo Fisher	C4040-03
One Shot™ <i>ccdB</i> Survival™ 2 T1 ^R Competent Cells	Thermo Fisher	A10460
Biological samples		
Total RNA - Human Fetal Normal Tissue: Liver	BioChain	R1244149-50 / A601607
Total RNA - Human Adult Normal Tissue: Liver	BioChain	R1234149-50 / B705065
Total RNA - Human Fetal Normal Tissue: Brain	BioChain	R1244035-50 / B210035
Total RNA - Human Adult Normal Tissue: Brain	BioChain	R1234035-50 / B805061
Total RNA - Human Fetal Normal Tissue: Heart	BioChain	R1244122-50 / B512118
Total RNA - Human Adult Normal Tissue: Heart	BioChain	R1234122-50 / B604038
Chemicals, peptides, and recombinant proteins		
NanoLuc substrate	Promega	N1110
Critical commercial assays		
PURExpress <i>In Vitro</i> Protein Synthesis Kit	New England BioLabs	E6800L
1-Step Human Coupled IVT Kit	Thermo Fisher Scientific	88882
Dual-Glo Luciferase Assay System	Promega	E2920
SMARTer® PCR cDNA Synthesis Kit	Takara	634926
Deposited data		
GTEX	GTEX consortium ⁴⁹	gtexportal.org
Developmental RNA-seq	Cardoso-Moreira ⁵⁰	E-MTAB-6814
Ribo-seq data	Chothani et al. ⁶²	GSE182372
Human Protein Atlas	Thul et al. ¹⁰⁰	proteatlas.org
TF mORF Atlas	Joung et al. ¹¹⁴	GSE216481
The Cancer Genome Atlas Breast Cancer Data	NCI Genomic Data Portal	portal.gdc.cancer.gov
TBX5 and CREB1 PBMs	This paper	GEO: GSE253638
TBX5 ChIP	Churko et al. ¹²⁵ ; Ang et al. ¹²⁶ ; Gonzalez-Teran et al. ¹²⁷	GSE81585; GSE85631; GSE159411
Original code	This paper	Zenodo: 14969075
Experimental models: Cell lines		
HEK293T	ATCC	CRL-3216
HepG2	ATCC	HB-8065
MCF-7	ATCC	HTB-22
U2OS	ATCC	HTB-96

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Yeast Yc1867	Reece-Hoyes et al. ⁶⁰	N/A
Yeast Y1Has2	Reece-Hoyes et al. ⁶⁰	N/A
Yeast Y8800	Luck et al. ⁶¹	N/A
Yeast Y8930	Luck et al. ⁶¹	N/A
Oligonucleotides		
Primers used for cloning	Data S1	N/A
Primer used for PBMs	Integrated DNA Technologies	Agilent Primer1
Recombinant DNA		
pDONR223	Rual et al. ¹²⁸	N/A
pDEST-P4P1R	ThermoFisher	N/A
pMW#2	Addgene	13349
pMW#3	Addgene	13350
pGADT7-GW	Addgene	61702
pDEST-AD2 μ	Reece Hoyes et al. ⁶⁰	N/A
pGL4.23[luc2/minP]-GW	Fuxman Bass et al. ⁴	N/A
pEZY3-VP160	Santoso et al. ¹²⁹	N/A
pGL4.74[hRluc/TK]	Promega	E6921
pEZY3-DB	Lu et al. ¹³⁰	N/A
Gateway pDEST15 Vector	Thermo Fisher Scientific	11802014
pt7CFE1-NHis-GST-CHA [Gateway compatible]	Thermo Fisher Scientific	88871
pDEST-N2H-N1	Addgene	125547
pDEST-N2H-N2	Addgene	125548
pDEST-N2H-F1	Addgene	125551
pDEST-N2H-F2	Addgene	125552
pDEST-DB	Luck et al. ⁶¹	N/A
pDEST-AD-CYH2	Luck et al. ⁶¹	N/A
pcDNA3.1-ccdB-EGFP	Gift from Taipale Lab	N/A
N-term GST pDEST 27	Invitrogen	11812013
Software and algorithms		
kallisto	Bray et al. ¹³¹	pachterlab.github.io/kallisto
HMMER v3.3	Eddy ⁶⁹	hmmer.org
AlphaFold v2.3.1	Jumper et al. ⁵⁵	https://github.com/google-deepmind/alphafold
upbm v0.99.0	Kock et al. ⁷¹	github.com/pkimes/upbm
GENRE	Mariani et al. ¹³²	github.com/BulykLab/MEDEA
GSEApY v1.1.3	Fang et al. ¹³³	gseapy.readthedocs.io
Mathematica-based image analysis interface	Riback et al. ¹⁰⁶	N/A
STAR v2.4.0	Dobin et al. ¹³⁴	github.com/alexdobin/STAR
AltAnalyze v2.1.4	Emig et al. ¹³⁵ and Li et al. ¹³⁶	http://altanalyze.org/
Other		
Custom tRNA Mix	New England BioLabs	N6842Z
XT Reducing Agent 20X 1610792	BioRad	1610792
XT Sample Buffer 4X	BioRad	1610791
BlueStain 2 Protein Ladder, 5-245kDa	GoldBio	P008
GST Glutathione S-Transferase	Thermo Fisher Scientific	20237
Tween 20	Sigma-Aldrich	P1379
XT MES Running Buffer 20X	BioRad	1610789
RPI Dry Milk Powder	Research Products International	M17200

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Phosphate Buffered Saline 10X PBS for Western Blots and IP	Sigma-Aldrich	P7059
1-Step Transfer Buffer	Thermo Scientific	84731
Thermo Sequenase Cycle Sequencing Kit [Discontinued]	Thermo Fisher Scientific	785001KT
Cytiva Thermo Sequenase Reaction Buffer 10X	Cytiva Life Sciences	E790000Y
Cytiva Thermo Sequenase DNA Polymerase (with Thermoplasma acidophilum Inorganic Pyrophosphatase (TAP))	Cytiva Life Sciences	E790000Y
Deoxynucleotide (dNTP) Solution Set, 100mM	New England BioLabs	N0446S
Amersham CyDye Fluorescent Nucleotides Cy3-dUTP	Cytiva Life Sciences	PA53022
Triton X-100	Sigma-Aldrich	T9284
Nonfat Dried Milk Bovine	Sigma-Aldrich	M7409
Deoxyribonucleic acid, single stranded from salmon testes	Sigma-Aldrich	D7656
Molecular Grade Bovine Serum Albumin	New England BioLabs	B9200
KOD HotStart Polymerase	Novagen	71086
BP Clonase	Life Technologies	11791043
Spectinomycin	Sigma Aldrich	S4014-5G
3-AMINO-1,2,4-TRIAZOLE >=95% (TLC)	Sigma Aldrich	A8056-100g
CYCLOHEXIMIDE FROM MICROBIAL SOURCE	Sigma Aldrich	C7698-5G
96-well, flat bottom, cell culture microplate	Greiner	655083
Axygen™ Oasis™ Robotic Tips	Fisher Scientific	14-222-223
Coming™ Untreated 245mm Square BioAssay Dishes	Fisher Scientific	07-200-600
96WL R-BTM PLT TCT W/LP 100/CS	Fisher Scientific	07 200 720A
PEI MAX 40000	Polysciences Inc	24765
SWIM primer	Life Technologies	Custom
Furimazine	Pasteur Institute	Custom
FuGENE HD Transfection Reagent	Promega	E2311
Lipofectamine™ 3000 Transfection Reagent	Invitrogen	L3000001
CHAPS Detergent	Thermo Fisher	28299
Unstained protein ladder	Thermo Fisher	84786
Bolt™ Sample Reducing Agent	Thermo Fisher	B0004
Bolt™ MOPS SDS Running Buffer	Thermo Fisher	B0001

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Yeast strains

For the eY1H assay, DNA-bait strains had been generated in the Y1HaS2 yeast background and were previously published^{4,60,129} and TF-prey strains were generated in the Y α 1867 yeast background. For the Y2H assay, AD-ORF strains were generated in the Y8800 background and DB-ORF strains were generated with the Y8930 background.

Human cell lines and cell culture

HEK293T, HEPG2, and MCF-7 cells were maintained in DMEM and U2OS cells were maintained in RPMI, each supplemented with 10% FBS and 1% antibiotic-antimycotic. Cells were passaged every 2–3 days at a ratio of 1:4, were kept in a sterile incubator at 37°C and 5% CO₂, and regularly tested for mycoplasma contamination. Only low passage number cells were used in mammalian one-hybrid, NanoLuc two-hybrid, and Western blot experiments.

METHOD DETAILS

TFiso1.0 clone collection generation and validation

A PCR-based method was used to amplify and clone the coding regions of TF isoforms, similarly to that in Yang et al.⁴¹ Gene-specific anchoring primers (Data S1) were used to PCR amplify ORF sequences from reverse transcribed RNA obtained from fetal and adult brain, heart, and liver tissues obtained from BioChain.

Catalog number	Description	Lot number
R1244149-50	Total RNA - Human Fetal Normal Tissue: Liver	A601607
R1234149-50	Total RNA - Human Adult Normal Tissue: Liver	B705065
R1244035-50	Total RNA - Human Fetal Normal Tissue: Brain	B210035
R1234035-50	Total RNA - Human Adult Normal Tissue: Brain	B805061
R1244122-50	Total RNA - Human Fetal Normal Tissue: Heart	B512118
R1234122-50	Total RNA - Human Adult Normal Tissue: Heart	B604038

Primers were targeted against TF protein-coding transcripts based on sequences from GENCODE v21, with TFs defined as a union of two datasets: Reece-Hoyes et al.¹³⁷ and TFClass.¹³⁸ In subsequent data analysis, we updated the list of TFs used to Lambert et al.²⁰ and the GENCODE version to v30.

Cloning was performed in two stages: an initial pilot stage followed by the main stage, with clones resulting from both stages combined in the final collection. In the pilot stage, TFs targeted for cloning were selected based on long-read RNA-seq data of brain, heart, and liver, obtained from PacBio (private communication). In the full stage, the majority (> 90%) of the TFs targeted for cloning were selected based on having either protein-protein interactions (PPIs) or protein-DNA interactions (PDIs) in other ongoing single-isoform-per-gene PPI and PDI mapping projects from our labs. An additional 45 TF genes with the potential to have measurable differences in DNA binding were selected, the criteria for this was having an annotated alternative isoform with differences, relative to its reference isoform, in the DBD or in the 15 a.a. either side of the DBD, with both isoforms less than 80 kDa and less than 80% predicted disorder. Finally, an additional 25 TF genes with variants associated with neurodegenerative diseases (private communication) were added.

Cloning of isoforms of selected target genes was carried out as described previously.^{41,139} Reverse transcription (RT) was carried out using a SMARTer® PCR cDNA Synthesis Kit (TaKaRa) with oligo (dT)16 primers according to the manufacturer's instructions. The resultant cDNAs were used as templates for PCR amplification using KOD HotStart Polymerase (Novagen) and ORF-specific primers (Data S1). Up to 4 primer pairs targeting alternative N- and C-termini were used for each TF gene. The resulting amplicons, which may contain more than one alternatively spliced isoform were transferred into pDONR223 by Gateway™ BP reaction (Life Technologies) followed by transformation into *E. coli* DH5 α . Transformed *E. coli* cells were plated on LB agar containing 50 mg/L spectinomycin for overnight growth at 37°C, after which up to 24 colonies were isolated for each primer pair using a Qpix 2 XT colony picking robot (Molecular Devices - Genetix).

These cloned TF isoforms were then combined with existing cloned TFs from our ORFeome collections.^{41,61} PCR artifacts, duplicates, cases where there were multiple isoforms per well, and incomplete or otherwise erroneous cloned isoforms, were removed following Illumina short-read sequencing of the corresponding clones. The final set of ORFs was then chosen based on subsequent multiplexed, full-length, long-read sequencing.

The short-read sequencing step was performed mostly as described in Yang et al.,⁴¹ pooling individual *E. coli* strains carrying plasmids encoding different TF genetic loci. Plasmid DNA minipreps from the pooled *E. coli* were prepared on a Qiagen BioRobot® Universal System according to the manufacturer's instructions and processed to make an Illumina sequencing library, during which Illumina adapter sequences, i7 and i5, were incorporated as plate indexes. The library was then paired-end-sequenced using an Illumina platform (MiSeq or NextSeq 500).

The long-read sequencing templates were generated by PCR using a method where the forward primers contained well specific barcodes and the reverse primers contained plate specific barcodes, enabling the clones to be pooled into a small number of aliquots that were processed for long-read sequencing on a PacBio RS II system (Pacific Biosciences of California, Inc.).

All isoforms were assessed by expert human annotators using methods and standards developed by GENCODE^{25,140} to identify well supported isoforms suitable for further study. To assess orthogonal support for transcript structures, each isoform was aligned to the human reference genome (GRCh38) and compared to the contemporaneous GENCODE gene annotation, all available long transcriptomic data, RNAseq data and RNAseq supported introns from the Intropolis dataset¹⁴¹ and CAGE transcription start site data.¹⁴² Having determined support for transcript structure, putative novel CDSs were similarly investigated to assess support for translation initiation sites and novel coding exons using aligned protein sequence data, Ribo-seq data, and PhyloCSF¹⁴³ constraint data. For the 246 genes in TFiso1.0, our clone collection is missing a total of 504 GENCODE-annotated isoforms. We did not observe

any obvious systematic difference between cloned and uncloned GENCODE-annotated isoforms, with the exception of some missing alternative N- or C-terminal isoforms that likely dropped out of our collection due to PCR difficulties (Figure 2B).

There were 28 cases where two or more clones encoded identical amino acid sequences. These were tested in the Y1H, M1H, and Y2H assays, and data from the duplicate clones were filtered out, keeping the clone that had the least drop-outs in the assays.

The inclusion of novel isoforms in our clone collection necessitated a numbering system that expands upon GENCODE annotation: we refer to TF isoforms by their gene name and clone ID and supplement these with the matching GENCODE transcript name for annotated isoforms.

TF isoform annotations

Transcription factors, as defined by Lambert et al., were downloaded from <http://humantfs.ccb.utoronto.ca/> v1.01.²⁰

Isoforms were matched to the CDS sequence of transcripts in the GENCODE basic set of GENCODE v30. Two cloned isoforms matched identical sequences to isoforms of two genes, HSFY1 and HSFY2, where the reference and alternative isoforms have identical CDS. We arbitrarily annotated these clones as HSFY1. For analysis, GENCODE transcripts with identical amino acid sequences but differences in the UTRs were merged into one protein isoform.

The reference isoform of a gene was defined in the vast majority of cases by the MANE select transcript. In the cases where a MANE select transcript was not available, the APPRIS principle isoform was used. If that was also not available, the longest isoform was chosen. The cloned reference isoform for a gene in TFiso1.0 was defined as the reference isoform, if it was cloned. If the reference isoform was not cloned, the cloned reference isoform was defined by the APPRIS annotated isoforms preferring principle over alternative. If no MANE or APPRIS annotated isoforms were cloned or available, then the cloned reference isoform was defined as the longest isoform matched to a GENCODE transcript. In the final case, if no GENCODE-matched isoforms were cloned, then the cloned reference isoform was the longest cloned isoform.

Alignment of TFiso1.0 isoforms with prior curated isoforms from the literature were manually determined based on the original evidence provided, including reported exon position and sequence length.

Western blots of novel TF isoforms

The controls used to validate the reported isoform bands were produced by transient transfection of 3.5 million HEK293T cells with 5.8 µg plasmid DNA using Lipofectamine™ 3000 Transfection Reagent (Invitrogen L3000001) with the manufacturer's recommended volumes for a 10-cm plate. The plasmid constructs were generated using standard Gateway cloning methods to insert the ORFs in N-terminal GST-tagged pDEST27 for mammalian expression.

All cell lysates were collected 24 hours after transfection in a solution containing 6.2 mg/mL CHAPS Detergent (Thermo Fisher 28299), 0.35 M NaCl, and 0.2 M Tris (pH 7.5). Following 30 minutes of centrifugation at 14,000 × g at 4°C, the supernatant was mixed with sterile glycerol to a final concentration of 25% glycerol before flash freezing with liquid nitrogen in single-use aliquots prior to storage at -80°C.

Western blots were performed by loading equal protein concentrations of HEK293T, HepG2, and MCF-7 lysates into 4-20% Tris-Glycine Mini Protein Gels. An unstained protein ladder (Thermo Fisher 84786) was used for molecular weight estimations. All samples except the ladder were incubated at 92°C for 5 minutes with XT Sample Buffer and Bolt™ Sample Reducing Agent (Thermo Fisher B0004), both at the manufacturer's recommended concentrations, before loading onto the gel. The gels were run for 40 minutes in Bolt™ MOPS SDS Running Buffer (Thermo Fisher B0001). The resulting gel was transferred to a Power Blotter Pre-cut nitrocellulose mini membrane using constant 2.5 A for 7 minutes on an Invitrogen Power Blotter–Semi-dry Transfer System. All blocking and antibody binding incubations were performed in a solution of PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 2 mM KH₂PO₄), 0.1% Tween 20™, and 0.03 g/mL of non-fat milk powder. All primary antibodies were used at the following dilutions: SP2 (Proteintech 25000-1-AP) 1:500, GAPDH (Cell Signaling 2118T) 1:5000. The secondary anti-rabbit antibody (Invitrogen 31460) was used at 1:2000 for SP2 and 1:10,000 for GAPDH. All blocking and antibody binding incubations were performed by placing the nitrocellulose membrane on a shaking platform at 70 rpm at room temperature for 60 minutes.

Visualization of the Western blots was accomplished by submerging the membrane with the two part SuperSignal West Ferato solution (Thermo Fisher) and waiting 6 minutes. Imaging was performed using standard chemiluminescence filter settings with an exposure time of 2 minutes.

Detection of protein-DNA interactions using enhanced yeast one-hybrid assays

Enhanced yeast one-hybrid (eY1H) assays were performed as described previously.^{4,16,60} TF isoform ORF clones from TFiso1.0 were transferred by Gateway LR cloning (ThermoFisher #11791100) to the destination vector pDEST-AD2µ (Walhout Lab)^{4,16,60} to generate fusion clones of each TF isoform with the yeast Gal4 activation domain (AD).

To generate TF-prey yeast strains, cloned TF isoform ORFs were transformed into haploid MATα type yeast strain Yα1867, as previously described^{4,16,60} and as follows. Yeast were inoculated in 1 L liquid YAPD media to a concentration of OD₆₀₀ = 0.15 and were then incubated at 30°C shaking at 200 rpm until they reached OD₆₀₀ = 0.5, washed with sterile water, and washed again with 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 before adding ~250 ng of the TF isoform clone. Six volumes of TE/LiAc + 40% polyethylene glycol were then added and samples were mixed gently ten times. Yeast were incubated at 30°C without shaking for 30 min followed by

42°C for 20 min, then resuspended in sterile water. Transformed yeast were plated on selective media lacking tryptophan to select for transformants.

DNA-bait yeast strains for 330 human enhancer and promoter sequences were previously generated using the Y1Has2 yeast strain.^{4,129} These DNA baits correspond to known human developmental enhancers selected from the Vista Enhancer Browser (enhancer.lbl.gov)⁴ and were tested against all TF ORFs. These baits were selected given their activity in different tissues during early embryonic development in order to maximize the number of TFs for which DNA binding could be detected. In addition, TF isoforms corresponding to MAX, STAT1, STAT3, PPARG, RARG, and RXRG were tested against a collection of 119 cytokine promoter DNA-baits using paired yeast one-hybrid assays as previously described.¹⁴⁴ For a complete list of baits with positive interactions, see [Data S2](#). Each DNA-bait strain carries two integrated copies of the enhancer or promoter cloned upstream of two reporter genes: *HIS3*, which allows yeast to grow in the absence of histidine and overcome inhibition by 3-Amino-1,2,4-triazole (3AT), and *LacZ*, which causes yeast colonies to turn blue in the presence of 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-gal).

eY1H assays were performed in 1,536-colony format using a high-density array ROTOR robot (Singer Instruments), which facilitated the comparison between TF isoforms by allowing simultaneous testing on the same array plate. TF-prey yeast strains were mated in a pairwise manner with 211 DNA-bait strains on permissive YAPD agar plates and incubated at 30°C for one day. Yeast were then transferred to selective media agar plates lacking uracil and tryptophan and incubated at 30°C for two days to select for successfully mated diploid yeast. The resulting diploid yeast colonies were finally transferred to selective media agar plates lacking uracil, tryptophan, and histidine, with 320 mg/L X-gal and 5mM 3AT. Readout plates were imaged 2, 3, 4, and 7 days after plating. Binding of the TF-AD fusion to the DNA-bait region results in expression of the *HIS3* and *LacZ* reporter genes, allowing colonies to visibly grow and turn blue on readout plates. eY1H and paired yeast one-hybrid assay images were manually analyzed by three independent researchers to identify interactions. Array coordinate “holes” - where yeast mating or transfer was unsuccessful - were identified and removed from analysis.

Validation of yeast one-hybrid protein-DNA interaction data using luciferase assays

For a random subset of TF isoform series (i.e., all isoforms of a TF in TFiso1.0) a random subset of DNA baits were selected that had at least one interaction identified by eY1H assays with at least one of the isoforms. These all-by-all combinations for each selected TF isoform series were validated in an orthogonal system by luciferase assays in HEK293T cells (see [Data S2](#)). Briefly, DNA-bait sequences were cloned upstream of the firefly luciferase reporter in a Gateway compatible pGL4.23[luc2/minP] vector.⁴ TF isoform ORFs were cloned into the Gateway compatible pEZY3-VP160 vector¹²⁹ such that TF isoforms are fused to 10 copies of the VP16 activation domain. HEK293T cells were plated in 96-well white opaque plates at a seeding density of ~10,000 cells/well and incubated for one day at 37°C with 5% CO₂. Cells were transfected using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol, with 80 ng of TF isoform (pEZY3-VP160) plasmid, 20 ng of DNA bait (pGL4.23) plasmid, and 10 ng of the renilla luciferase plasmid as a transfection normalization control. An empty pEZY3-VP160 plasmid co-transfected with the corresponding recombinant firefly luciferase plasmid were used as negative controls. Transfected cells were incubated for two days at 37°C with 5% CO₂. Firefly and renilla luciferase activities were measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol. Non-transfected cells were used to subtract background firefly and renilla luciferase activities, and then firefly luciferase activity was normalized to renilla luciferase activity in each well. Each TF isoform-DNA bait pair was tested in three biological replicates. In the event of TF isoform-DNA binding, the VP16 activation domains promote the expression of firefly luciferase, increasing the normalized luminescence over background levels.

Protein-DNA interaction assay using protein-binding microarrays

Full-length TF isoforms (3 isoforms of TBX5 and 2 isoforms of CREB1) were cloned from Gateway compatible Entry vectors (pDONR223) into N-terminal GST protein fusion expression Destination vectors and sequence-verified by long-read DNA sequencing via Plasmidsaurus. Specifically, TBX5 isoforms were cloned into a modified pT7CFE1-NHis-GST vector (Thermo 88871), which is compatible with mammalian *in vitro* transcription and translation (IVT) kits, and CREB1 isoforms were cloned into pDEST15-NGST (Thermo 11802014), which is compatible with the PURExpress IVT kit (New England Biolabs). Proteins were then expressed using either the 1-Step Human Coupled IVT Kit (for TBX5 isoforms) or PURExpress In Vitro Protein Synthesis Kit (for CREB1) (NEB E6800L), using the manufacturers' recommended protocols, with the exception of an addition of 1.5 μL custom tRNA mix (NEB N6842Z) to the 25 μL CREB1 PURExpress reactions. Protein expression was verified and quantified by Western blot using a dilution series of recombinant GST protein (Sigma G5663) as a standard. Primary rabbit anti-GST polyclonal antibody (Sigma G7781) (1:160,000) and secondary goat horseradish peroxidase-conjugated IgG monoclonal antibody (Pierce 31460) (1:200,000) were used for Western blotting.

Universal PBMs representing all 10-mers in 8 x 60K, GSE format (Agilent Technologies: AMADID #030236) were used. For TBX5 arrays, double-stranding of the PBM oligonucleotide arrays was performed as previously described.^{70,145} The polymerase we used for the TBX5 PBM arrays (Thermo Fisher Thermo Sequenase Cycle Sequencing Kit 785001KT) was discontinued by the manufacturer, and so, for doublestranding the arrays used in the CREB1 PBM double experiments, we instead used Cytiva Thermo Sequenase DNA polymerase (Cytiva E790000Y) using three times as much polymerase as in our standard PBM protocol⁶⁴ but otherwise following the manufacturer's protocol. PBMs were then performed as described,^{70,145} using 50 μg/mL Alexa-488-conjugated rabbit polyclonal anti-GST antibody (Invitrogen A11131) in PBS / 2% (wt/vol) nonfat dried milk. CREB1 isoforms were assayed at 400 nM

final concentration and TBX5 isoforms were assayed at 750 nM final concentration. PBMs were scanned in a GenePix 4400A microarray scanner. Each isoform was assayed on ≥ 2 independent arrays, and alternative isoforms were always assayed alongside their cognate reference isoforms in a separate ‘chamber’ on the same array. Because of the higher level of noise when assaying full-length TFs compared to extended DBDs, each reference isoform was assayed on an additional ≥ 2 independent arrays to ensure robust quantification and differential comparisons.

Protein-protein interaction assay with yeast two-hybrid (Y2H) Y2H Screens

The Y2H screens were performed mostly as described in Luck et al.⁶¹ with some modifications.

Yeast strains and transformation. Competent yeast strain Y8800, mating type MAT α (*leu2-3,112 trp1-901 his3 Δ 200 ura3-52 gal4 Δ gal80 Δ GAL2::ADE2 GAL1::HIS3@LYS2 GAL7::lacZ@MET2cyh2^R*) were transformed with individual AD-ORF constructs and plated onto yeast synthetic complete media¹⁴⁶ lacking tryptophan (SC-Trp) to select for AD-ORF plasmids.

Competent yeast strain Y8930, mating type MAT α , (*leu2-3,112 trp1-901 his3 Δ 200 ura3-52 gal4 Δ gal80 Δ GAL2::ADE2 GAL1::HIS3@LYS2 GAL7::lacZ@MET2cyh2^R*) were transformed with individual DB-ORF constructs and plated onto SC-Leu to select for DB-ORF plasmids. Haploid DB-ORF yeast strains were tested for auto-activation of the *GAL1::HIS3* reporter gene. Individual DB-ORF yeast strains were spotted on SC-Leu-His+1mM 3AT media and any strains showing growth were considered auto-activators (AAs) and removed from the collection of strains to be screened.

Primary Y2H Screens. Two first-pass Y2H screens were performed, in which all TF isoforms were tested against (1) the hORFeome v9.1 collection of $\sim 17,500$ ORF clones^{61,146} and (2) a subset of the hORFeome collection that were annotated as TFs or co-factors. The list of co-factors was taken from the union of the TcoF database¹⁴⁷ using the January 2017 update, and from Heinäniemi et al.¹⁴⁸

In both Y2H screens, TF isoforms were tested as fusions to the Gal4 activation domain (AD) in the Gateway compatible pDEST-AD-CYH2 vector, and screened against the hORFeome v9.1 fused to the Gal4 DNA-binding domain in the Gateway compatible pDEST-DB vector. To perform the screen pools of Y8930:DB-ORF yeast strains (baits) were mated against pools of Y8800:AD-ORF strains (preys). The TF isoforms AD-ORF yeast strains were combined into pools of 100 individual strains. In the large screen against the hORFeome the DB-ORF strains were combined into pools of 8 DB-ORF strains, and in the second focused screen DB-ORF strains were screened individually. These first-pass screens represent a systematic interrogation of ~ 13 million possible PPIs. To perform the mating, fresh overnight cultures of DB-ORF strains (either pools or individual) were mixed with AD-ORF strain pools and grown overnight at 30°C in liquid rich media (YEED). After overnight growth, the mated yeast cells were transferred into liquid SC-Leu-Trp media to select for diploids and again grown overnight at 30°C. Finally the yeast cells were spotted onto SC-Leu-Trp-His+1mM 3AT solid media to select for activation of the *GAL1::HIS3* reporter gene. In parallel, diploid yeast cells were transferred onto SC-Leu-His+1mM 3AT solid media supplemented with 1 mg/l cycloheximide (CHX) to test for spontaneous DB-ORF auto-activators. All AD-ORF plasmids carry the counter-selectable marker *CYH2*, which allows selection on CHX-containing media of yeast cells that do not contain any AD-ORF plasmid. After 72h incubation at 30°C, yeast that grew on SC-Leu-Trp-His+1mM 3AT media but not on SC-Leu-His+1mM 3AT+ 1 mg/l CHX media were picked into SC-Leu-Trp grown overnight and then processed to determine the identity of the respective bait and prey proteins. To identify the interacting bait and prey we used SWIM-Seq as described in Luck et al.⁶¹

The hits from the screen were then combined with PPIs from the subset of HuRi⁶¹ that was detected using Y2H v1 with the TF as the AD fusion, and Lit-BM-17,⁶¹ a dataset of literature-curated PPIs with multiple evidence including at least one experimental method that detects binary PPIs. These pairs were then tested in a series of initial pairwise Y2H experiments, testing each isoform of a TF gene against all interaction partners. These experiments were used to filter out pairs that were not positive with any of the isoforms of a TF gene, TF isoforms that were not positive with any interaction partner, and profiles of TF genes that were not positive (for at least one isoform) with at least two different partners and had at least two different isoforms with at least one positive interaction. These experiments were described as below, with the exception that the plate position of pairs was randomized, rather than keeping all isoforms of the same gene with the same partner on the same plate.

In the final pairwise test, we included additional pairs to test, so that we could compare the PPI profiles of paralogs to those of isoforms without the confounding effect of the sampling sensitivity of the screening. For a subset of paralogous TF genes and a control set of random paired non-paralogous genes (see the section on Paralogs definition), we additionally tested all isoforms of each paired gene with any additional interaction partners tested for the other gene.

Y2H Pairwise Test

Following first-pass screening, each protein isoform was pairwise tested for interaction with the candidate partners identified not only for itself but also for all first-pass partners of all other protein isoforms encoded by the same gene, thus minimizing biases due to incomplete sampling sensitivity.¹⁴⁹ To generate a final dataset of verified Y2H pairs, pairs were accepted if they showed (1) a valid growth score and (2) their ORF identities were confirmed by sequencing of the PCR products amplified from the tested colonies.

Briefly, interactors were inoculated in 200 μ L corresponding selection media and mated overnight at 30°C in 150 μ L liquid rich media (YEED). The following day, mated yeast cells were transferred into 150 μ L liquid SC-Leu-Trp media to select for diploids. After overnight incubation at 30°C, 5 μ L diploid yeast cells were spotted onto SC-Leu-Trp-His+1mM 3AT solid media to select for activation of the *GAL1::HIS3* reporter gene as well as on SC-Leu-Trp to control for successful mating. AA tests were included by mating each Y8930:DB-ORF against a Y8800:AD-null (containing no ORF), which was included on each test plate for each individual

Y8930:DB-ORF. Spots were scored for growth⁶¹ with scores of 0, 1, 2, 3, 4, and NA for cases where the spotting had failed. Growth scores of 0 and 1 were considered not growing. If a spot corresponding to either the pair or the corresponding AA-test did not grow on SC-Leu-Trp, the pair was scored NA. If the AA-test had a growth score of 4, the corresponding pair was scored NA. Interactions were scored positive if they had higher growth scores on SC-Leu-Trp-His+1mM 3AT solid media compared to the auto-activator test and had a growth score of at least 2; otherwise they were scored negative. All positive scored colonies were picked, lysed and the identity of the two interacting proteins was confirmed performing SWIM-seq.⁶¹ In addition, all SC-Leu-Trp plates were also sequenced. On each test plate, internal controls were included and in each batch of tested plates a positive reference set (PRS) and random reference set (RRS) was tested alongside the actual experiment.

Protein-protein interaction validation using mammalian NanoLuc two-hybrid (mN2H)

A random sample of PPIs identified by Y2H assays were validated in an orthogonal system by luciferase complementation assays in HEK293T cells,¹⁵⁰ along with positive and negative controls (Data S4). The tested positive Y2H pairs were a random sample of 300 pairs. The sample of negative pairs to test were selected in two steps: (1) negative pairs involving the same PPI partner with different isoforms of the same TF genes as the sampled positive pairs, where all negative pairs for each TF gene and partner combination were randomly selected to be included using a probability with a value of the fraction of positive pairs sampled for that gene; (2) an additional random sample of 150 negative pairs. The rationale behind this approach is that: by pairing the positive to negative pairs in step (1) we should reduce the variance when comparing positive to negative. The random selection in step (1) ensures a uniform random sample. Without it, the sample would be biased towards negative pairs with a larger number of positive PPIs involving other isoforms of the same TF gene. The additional pairs from step (2) were needed to increase the size of the negative sample to match the size of the positive sample.

In this assay, interacting partners were cloned into the corresponding N2H gateway plasmids (pDEST-N1 or pDEST-N2). Briefly, 30,000 HEK293T cells were seeded in a 96-well, flat bottom, cell culture microplate (Greiner Bio-One, #655083), and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum at 37°C and 5% CO₂. Next day, cells were transfected with 100 ng of each N2H plasmid using linear polyethylenimine (PEI) to co-express the protein pairs fused with complementary NanoLuc fragments, F1 and F2. The following day, the media was removed and 50 μL of 100× diluted NanoLuc substrate (Promega, #N1110) or 100× diluted furimazine substrate (Yves Janin) was added to each well of a 96-well microplate containing the transfected cells. Plates were incubated for 3 min at room temperature. Luciferase enzymatic activity was measured using a TriStar luminometer (Berthold; 2 s integration time).

Transcriptional activity using mammalian one-hybrid assays

To measure TF isoform transcriptional activity, modified mammalian one-hybrid assays (M1H) were performed in HEK293T cells. Briefly, TF isoform ORFs were cloned into a Gateway compatible DB-pEZY3 vector such that TF isoforms would be N-terminally fused to the Gal4 DNA-binding domain (DB). Four copies of the yeast UAS site corresponding to the Gal4 DB were then cloned upstream of the firefly luciferase reporter gene in a Gateway compatible pGL4.23[luc2/minP] vector. The DB-pEZY3 and 4xUAS-pGL4.23 backbone vectors were generated for this study (Figure S2F). HEK293T cells were plated in 96-well white opaque plates at a seeding density of ~10,000 cells/well and incubated for one day at 37°C with 5% CO₂. Cells were then transfected using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol, with 80 ng of TF isoform (DB-pEZY3) plasmid, 20 ng of 4xUAS-pGL4.23 plasmid, and 10 ng of the renilla luciferase plasmid as a transfection normalization control. An empty DB-pEZY3 plasmid co-transfected with the 4xUAS-pGL4.23 were used as negative controls. Cells were incubated for 2 days after transfection at 37°C with 5% CO₂, and then firefly and renilla luciferase activities were measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol. Non-transfected cells were used to subtract background firefly and renilla luciferase activities, and then firefly luciferase activity was normalized to renilla luciferase activity in each well. In this assay, if the TF isoform is recruited to the UAS by the fused Gal4 DB this would lead to the expression or repression of the downstream firefly luciferase reporter genes depending on the endogenous activating or repressing activity of the TF isoform.

Condensate formation assay

Selection of isoforms for condensate assay

We selected 192 of our cloned isoforms (two 96 well-plates) to profile for condensate formation and localization in the condensates assay. We prioritized alternative isoforms based on showing differences from the reference isoform in either the PDI, PPI, or transcriptional activation assays. We restricted to genes where the MANE select isoform was cloned and the alternative isoform was cataloged in GENCODE. The assay criteria were: a difference in PDI profile with at least three DNA baits positive in at least one isoform of the gene; a difference in PPI profile, with at least three successfully tested PPIs, and at least one positive PPI for both the reference and alternative isoform; an 8-fold or greater difference in activation. These criteria were selected to try and get a roughly even split between differences in the three assays. The cloned reference and all alternative isoforms of a TF gene were selected, such that many other alternative TF isoforms that didn't show differences in the three assays were also included in those to be tested. This selection resulted in 50 TF genes. We then added an additional 11 TF genes which did not pass the selection above, but on manual inspection, showed differences between reference and alternative isoforms in our assay readouts that we judged to be interesting. We removed four cloned alternative isoforms that were not successfully tested in any of the three assays.

High-throughput confocal microscopy

We transferred the reference and alternative isoform clones by Gateway LR reactions into a mammalian expression vector pcDNA3.1-ccdB-EGFP containing a C-terminal EGFP tag. All these clones were subjected to high-content imaging for condensate formation in two cell lines, HEK293T and U2OS. HEK293T and U2OS cells were transfected using standard protocols with FuGENE HD Transfection Reagent (Promega, Cat. No. E2311) in a 96-well plate format in DMEM and RPMI media, respectively, supplemented with 10% FBS and appropriate amounts of penicillin and streptomycin. 48 hours after transfection, cells were stained with DAPI, and imaging was performed using a ZEISS LSM 880 confocal microscope using a 63x objective. For comparative purposes, all available reference and alternative isoforms of the same gene were included in the same 96-well plate, for high-content imaging. We filtered out proteins that were not expressed from our imaging screen analysis. Alternative isoform-mediated condensate calls (Gain-of-condensate or GOC, Loss-of-condensate or LOC, and unchanged) were obtained, by comparing to their reference isoform profile (i.e., condensate or non-condensate). All phase separation experiments were performed in duplicate.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene Annotation

TF families were defined by Lambert et al.²⁰ MANE select transcripts were obtained from the file MANE.GRCh38.v0.95.summary.txt.⁴⁶ APPRIS transcript annotations were obtained from the file APPRIS-annotations_human_GRCh38.p13_ensembl104.tsv.¹⁵¹

Protein domain annotation

Pfam domains were mapped to protein isoforms from GENCODE v30 and our TFiso1.0 clone collection using HMMER version 3.3 and Pfam version 32.0. Domain matches were filtered for E-value < 0.01 and c-Value < 0.01. Overlapping Pfam domains were removed, keeping the domain with the lowest E-value. Zinc Finger domains, defined by membership of Pfam clan CL0361, that were separated by 10 amino acids or less, were merged into a single ZF array domain. We manually curated a list of Pfam domains that corresponded to DNA binding domains (Data S1). For TFiso1.0, we manually inspected each cloned reference isoform that did not have an annotated DBD, finding that for two TFs, HIF1A and ZNF207, their DBDs were above the E-value cutoff, and so we implemented a manual override of the filter in those two cases. Effector (activation and repression) domains were obtained from the literature-curated database TRegDB⁸ and from two published systematic tiling screens.^{6,7} Nuclear localization and export sequence motifs (NLS/NES) were downloaded from UniProt on 2023-10-02. For the protein interaction partners, Pfam domains were filtered for E-value $\leq 10^{-5}$.

Proportion of alternative isoforms with domain affected

P-values and error bars for the fraction of domains in reference isoforms which are affected in alternative isoforms, were calculated using a null model where the domain is randomly positioned along the reference isoform. This is calculated by first, for each domain/reference-isoform/alternative-isoform combination, calculating a probability as the fraction of cases in which the alternative isoform affects a dummy domain, a contiguous set of amino acids the same length as the real domain in the reference isoform, of all possible positions of that dummy domain along the reference isoform. In the case of multiple domains of the same type on a single reference isoform, the probabilities of at least one of the domains being affected was calculated, assuming independence. This array of probabilities for a specific type of domain, with one value for each reference/alternative isoform pair, was used in a Poisson Binomial distribution to calculate p-values and confidence intervals. Because the independence assumption, in the case of multiple domains, is violated by the fact that domains do not overlap, we compared our approach with a more computationally intensive and less numerically precise approach of repeatedly randomly shuffling the positions of the domains along the reference isoform, not allowing overlap in the case of multiple domains of the same type, and we found that the two approaches gave consistent results (data not shown).

AlphaFold structural prediction

Predicted 3D structures of cloned TF isoforms were obtained using AlphaFold version 2.3.1. With options: `--model_preset=monomer_ptm --db_preset=full_dbs --max_template_date=2023-05-05`. To produce figures showing the approximate position of DNA relative to the isoform structure, the AlphaFold structures were aligned to experimental structures of the TF, or a homologous TF, bound to DNA. The experimental structures were manually selected after searching for the amino acid sequence of the DNA binding domain of the reference isoform on the PDB website. HEY1 was aligned to human CLOCK in CLOCK BMAL1 heterodimer, PDB: 4H10; CREB1 was aligned to mouse CREB1 homodimer, PDB: 1DH3; TBX5 was aligned to mouse TBX5, interacting with NKX2-5, PDB: 5FLV.

Predicted disorder values

Binary per-residue predictions of being in a disordered region, for each cloned isoform, were derived from the AlphaFold predicted structures by¹⁵²:

- (i) calculating accessible surface area (ASA) and secondary structure using DSSP¹⁵³
- (ii) normalizing to relative solvent accessibility (RSA), using maximum ASA values from¹⁵⁴

- (iii) calculating a sliding-window average RSA value for each residue, using a window of 20 aa both sides of the residue in question
- (iv) residues with this average RSA ≥ 0.5 were categorized as ‘disordered’, with RSA < 0.5 as ‘structured’
- (v) performed a correction for long alpha helices, which would generally be structured in binding, but have high solvent accessibility when looking at the monomer structure, for example in bZIP TFs. Residues within contiguous regions classed as alpha helix of 20 amino acids or longer were set to ‘structured’.

PPI partner classification

Protein interaction partners of the TF isoforms were categorized into one of four categories: *TF*, *cofactor*, *signaling*, or *other*. *TF* was based on the Lambert et al.²⁰ list, *cofactor* were proteins that were not classed as *TF* but appeared in the list of human cofactors from Animal TF DB v4.⁸³ *Signaling* were those partners not already classed as *TF* or *cofactor*, that were annotated with the gene ontology (GO) term ‘signaling’ (GO:0023052) or one of its related lower terms. GO annotations were generated by UniProt on 2023-07-28 and the ontology file was released 2023-07-27. All remaining partner proteins were classed as *other*.

We obtained a list of TF families that typically bind DNA as obligate heterodimers from Jolma et al. Nat Meth. 2013.³ The 22 families are: AP-2, ARID/BRIGHT, BED ZF, bHLH, bZIP, CENPB, E2F, EBF1, GCM, Grainyhead, HSF, IRF, MADF, MADS box, Myb/SANT, Nuclear receptor, p53, RFX, Rel, SAND, SMAD, STAT.

Domain-domain PPI annotation

A list of interacting domain pairs was obtained from 3did⁸² 2022-05 release. All possible domain pairs matching to the TF isoform and partner protein were initially mapped. These 61 domain pairs and their corresponding evidence were manually inspected, filtered for quality and duplicates were removed, resulting in a filtered list of 42 domain pairs. Further de-duplication was performed by collapsing the multiple different Pfam domains corresponding to bZIPs (bZIP_1/bZIP_2/bZIP_Maf), homeobox (Homeobox/Homeobox_KN), and PAS (PAS/PAS_3/PAS_9/PAS11) to a single domain each.

RNA-seq analyses

RNA-seq analyses were performed by pseudo aligning reads to transcriptome indices made using the following reference fasta files: (1) to estimate the relative abundance of annotated transcription factor isoforms (GENCODE version 30) alone (i.e., analyses in [Figure 1](#)), we used the GENCODE version 30 protein-coding transcripts fasta file (which includes full transcript sequences, including UTRs) as the reference, and (2) to estimate the abundance of both annotated TF isoforms and unannotated cloned isoforms in the TFiso1.0 collection (i.e., analyses in [Figures 2, 3, 4, 5, 6, and 7](#)), we produced a consensus fasta reference that includes the aforementioned GENCODE version 30 protein-coding transcripts as well as the CDS sequences of any unannotated, novel clones in our collection. In both cases, we generated index files for the software Kallisto (version 0.46.0) using the “kallisto index” command with default parameters. Pseudoalignment was then performed on fastq files using the “kallisto quant” command with default parameters to estimate transcript per million (TPM) estimates per isoform. To estimate relative isoform abundance, gene-level TPM values were computed as the sum of all isoform TPM values for a given gene, and individual isoform ratios were determined relative to the total gene TPM for any gene with a TPM > 1 . GTEx data were downloaded from the Sequence Read Archive prior to migration of the data to ANVIL following dbGAP approval (phs000424.v8.p2). Developmental RNA-seq data from Cardoso-Moreira et al.⁵⁰ were downloaded from ArrayExpress (accession number E-MTAB-6814). TCGA breast cancer data were downloaded as paired-end bam files from the NCI Genomics Data Commons portal following dbGAP approval (phs000178.v10.p8) and converted to paired-end fastq files using samtools (version 1.15). For these analyses, we included a subset of representative GTEx samples (n=1,201 samples), spanning the same 30 patients (where possible) for all 51 tissue regions (excluding cell lines).

Re-sampling GTEx data

Since the number of samples per condition and the number of conditions in the GTEx and Developmental RNA-seq datasets was very different, in order to compare isoform expression between adult and developing tissues, we created a randomly sampled subset of the GTEx dataset. To clarify: a condition in GTEx is one adult tissue type (e.g. “Liver”) and a condition in Developmental RNA-seq is a tissue/time-point (e.g. “Liver 10 weeks post conception”). There were 1-5 samples per condition, with a median of 2 samples and a total of 127 conditions in Developmental RNA-seq, and 5-379 samples per condition, with a median of 24 and a total of 51 conditions in GTEx. To generate the resampled GTEx dataset we cycled through the GTEx tissues creating dummy conditions by randomly sampling the total number and number of samples per condition of the Developmental RNA-seq dataset.

Ribo-seq analyses

Bulk Ribo-seq fastq files were downloaded from GEO (GSE182372). These single-end 36-nt reads were aligned to the human genome (hg38) using 2-pass alignment with STAR and a custom transcriptome (ENCODE version 30 + novel TFiso1.0 isoforms). To quantify exon-exon junctions, we used AltAnalyze^{135,136} version 2.1.4.

Pairwise sequence identity analyses

Amino acid sequences were aligned using the pairwise2 module of biopython, with the blosum62 substitution matrix, an open gap penalty of -10, an extend gap penalty of -0.5, and penalize_end_gaps=False.

Paralogs definition

Paralogs were downloaded from Ensembl Compara on 2023-10-26. The non-paralog control set of pairs was generated from the list of paralog pairs, by randomly re-pairing the genes, removing any pairs that were in the original paralogs list of any cases where a gene was paired with itself. Note that there are some pairs in the non-paralog control of TF genes from the same family. We repeated the analysis with a different non-paralog control that specifically excludes within-family pairs and it produced very similar results (data not shown).

Functional assay quantification

Jaccard distances of PPI and PDI data for a pair of TF isoforms were calculated as $1 - \text{number of common interaction partners} / \text{total number of interaction partners}$. Only partners that were successfully tested in both isoforms were included. For Y2H PPI data, we did not use values where one of the isoforms had no interactions, in order to try and avoid artifacts where the clone was not functional in the assay.

Violin plots

Violin plots were drawn with the Gaussian KDE in the python package seaborn but modified to fit bounded data by reflecting the probability density back from the bounds, with the bounds being 0-1 in the case of PDI/PPI profile Jaccard distance, a lower bound of 0 for absolute \log_2 fold change of activation, and 0%-100% for sequence similarity. The kernel bandwidth was set to 0.1/10% for Jaccard distance/sequence similarity and 0.5 for absolute \log_2 fold change of activation.

PBM analysis

For all PBM replicates, a scan corresponding to a photomultiplier tube (PMT) gain of 500 was selected, as this consistently resulted in the lowest proportion of both over-saturated and under-saturated probes. PBM pre-processing was then performed using the upbm data analysis pipeline as described previously.⁷¹ Briefly, probe intensities were background-subtracted, Cy3-normalized (to account for any biases resulting from double-stranding the array) and spatially de-biased. Reference isoforms then served as “anchors” for cross-array normalization for each gene. PBM inference to determine differentially bound DNA 8-mers was also performed using the upbm pipeline, which tests for a difference in the 8-mer affinity scores against a null hypothesis of zero. GPR files are available via GEO at accession GSE253638.

IDR:DBD inhibition analysis

DNA-contacting residues in the DLX4 and PKNOX1 homeodomains were determined based on the contacts of analogous residues observed in crystal structures of related homologs—DLX5 (PDB: 4RDU) for DLX4 and Meis1 (PDB: 5EGO¹⁵⁵) for PKNOX1—using a 5 Å cutoff distance between protein:DNA heavy atom pairs. Structural analysis was performed using MDTraj.¹⁵⁶ All DNA-contacting residues identified in the crystal structures share perfect sequence identity with the corresponding homolog.

Next, we used FINCHES to calculate mean-field interaction strengths between IDRs and surface residues of the DLX4 and PKNOX1 homeodomains, using residue interaction parameters derived from the Mpipi-GG forcefield.^{75,157,158} Briefly, FINCHES uses forcefield parameters developed for molecular simulations and repurposes them to create a sequence-specific intermolecular energy function to predict transient “chemically-specific” interactions between an IDR and a partner. In this case, the solvent-accessible surface residues from the DLX4 and PKNOX1 homeodomains were investigated.

Homeodomain structures were obtained from predicted models of the reference isoforms obtained from AlphaFold Protein Structure Database.^{55,159} IDRs were predicted using metapredict (v2.63).¹⁶⁰ Due to minor discrepancies in the boundaries between the homeodomain-flanking IDRs and high-confidence regions in the AlphaFold structures, we assigned the DLX4 homeodomain as residues 124 to 180 and the PKNOX1 homeodomain as residues 266 to 325. Structural topology diagrams were generated using DODO (<https://github.com/idptools/dodo>), with IDR dimensions predicted using ALBATROSS.¹⁵⁷

The interaction strength of each IDR and defined sets of surface residues was determined by first computing an interaction parameter (ϵ) between all overlapping 31-residue tiles of the IDR and the surface patches, representing the average favorability of IDR interaction over the given surface residues. These IDR:surface ϵ values were then summed along the length of the IDR to provide a total interaction strength for the surface with the IDR. This approach is analogous to prior work calculating apparent IDR:folded domain interaction.⁷⁵ We also note this simplifying approach does not take steric considerations of the IDR into account. We evaluated the interaction strengths for the set of DNA-contacting residues of each homeodomain and compared these with the non-contacting residues, summing the contributions for all IDRs present in each isoform. This allowed us to calculate an IDR:DNA binding residue versus an IDR:non-DNA binding residue comparison.

Code for reproducing this analysis is available at https://github.com/holehouse-lab/supportingdata/tree/master/2024/tf_isoforms_2024.

TBX5 ChIP-seq analysis

Uniformly processed human TBX5 ChIP-seq peaks were downloaded (bed and bigwig files) from ChIP-Atlas.⁷⁸ Only studies profiling wild-type TBX5 were considered (see [key resources table](#) for a list of Accession IDs), and only peaks with MACS2 q-values $< 1 \times 10^{-5}$ were considered. Any overlapping peaks were de-duplicated using the `bedtools161 merge` command, such that only one peak from one study was considered (randomly sampled) in any overlapping regions, resulting in a final list of 2,074 non-overlapping TBX5 ChIP peaks. Peak regions were then centered to the nucleotide with the highest ChIP signal and trimmed to 150 nucleotides using `bwtool162` and hg38 genome sequences were extracted for each centered peak using `bedtools161`. GENRE¹³² was then used to generate a list of matched genomic background sequences (of the same length) for k-mer enrichment analyses.

Csat analysis

Microscopy images were acquired on a Zeiss LSM 880 confocal laser scanning microscope equipped with a Plan-Apochromat 63x/1.4 oil DIC M27 objective with a pinhole size of 46 μm (1 Airy unit). mEGFP (expressed fused on the c-terminal of constructs) was excited with a 488 laser and imaged with emission filters 498–552 nm. As Csat curve analysis requires (1) construct expression levels that differ >10 fold and (2) ensuring a linear range of intensities in both condensates and the dilute phase (i.e., nucleoplasm or cytoplasm) which typically differ by >10 fold, laser power is optimized for each field. This optimization is done by the user in real-time at the scope with the guiding principles that the max pixel should be roughly 10% of the max of the detector, maximizing signal to noise while maintaining the linearity of the image digital units with a concentration of construct. Quantitative comparison between the different laser settings is achieved by converting each image to digital units (after background subtraction) referenced at 1% laser power (reported as AU). This conversion was done using an empirically measured conversion factor formula determined with conversion factors calculated from the slope of the pixel values when imaging the same field of view at 1% laser power and various other laser settings, only including those pixels within linear range in both images. Note that these images were not used for Csat analysis as pixels were frequently out of linear range, and each field was only imaged once to avoid the complication of photobleaching. All other microscope and camera settings were kept constant. For analysis, cells or nuclei at different expression levels were found and the rough region containing them was hand-segmented with polygons to remove extra-cellular debris and other nearby cells. The designation of containing condensates was user decided at this point prior to Csat quantification. Measurement of the dilute phase concentration was done manually by choosing a non-foci-containing location and getting the value at that pixel with a maximal Gaussian blur to lower that noise but without including foci. To approximate the total concentration (x-axis) the average pixel value was calculated in a 2D image in a binary mask for the cell. This binary mask for the cell was determined by taking the image in the polygon segmented region, blurring it with a pixel radius of 5, doing a morphological binarization using the dilute phase value, filling all holes in the object, and deleting small components. The exact command in Mathematica is `DeleteSmallComponents[FillingTransform[MorphologicalBinarize[Blur[image, 5], dil]]]` where image and dil is the polygon segmented image region and digital value for the dilute phase chosen, respectively. This was done using a custom-made Mathematica GUI based on that used in Riback et al. Nature 2020.¹⁰⁶

Human Protein Atlas localization validation

Subcellular localization data from the Human Protein Atlas¹⁰⁰ were downloaded on October 26, 2023. The following Human Protein Atlas subcellular localization annotations were considered “cytoplasmic”: Actin filaments, Cleavage furrow, Focal adhesion sites, Intermediate filaments, Centriolar satellite, Centrosome, Cytokinetic bridge, Microtubule ends, Microtubules, Midbody, Midbody ring, Mitotic spindle, Aggresome, Cytoplasmic bodies, Cytosol, Rods & rings, Mitochondria, Endoplasmic reticulum, Vesicles, Endosomes, Lipid droplets, Lysosomes, Peroxisomes, Golgi apparatus, Cell junctions, Plasma membrane. If one of the above localizations was observed in any localization column (Approved, Enhanced, Supported, Uncertain), we considered the protein to be “cytoplasmic” in [Figure S6A](#).

Classification of negative regulators and rewirers

We defined negative regulator alternative isoforms as those exhibiting one or more of the following: (1) those that show 0 PDIs while their cognate reference isoform shows ≥ 1 PDI or those that lose $\geq 10\%$ of the DBD; (2) those that show loss of activation compared to their reference isoform (reference has M1H signal ≥ 1 , alternative has M1H signal between 1 and -1, and alternative $\log_2\text{FC} \leq -1$ compared to reference) or loss of repression compared to their reference isoform (reference has M1H signal ≤ -1 , alternative has M1H signal between 1 and -1, and alternative $\log_2\text{FC} \geq 1$ compared to reference); (3) those that show 0 PPIs while their cognate reference isoform shows ≥ 1 PPI or those that lose all of 1 key type of PPI (within-family TFs of obligate dimers, signaling proteins, or transcriptional cofactors). To classify an alternative isoform as a negative regulator, we required that it show evidence of functionality in 1 additional assay as follows: (1) ≥ 1 PDI, (2) ≥ 1 PPI, or (3) M1H signal ≥ 1 (activation) or ≤ -1 (repression). Thus, only alternative isoforms with data from ≥ 2 assays (eY1H, Y2H, or M1H) were classified based on assay data alone; unless we considered the isoform a negative regulator due to loss of DBD (in those cases, evidence of functionality could come from Y2H or M1H assays alone). Finally, we layered subcellular localization on top of these categorizations (we did not consider it with the same initial weight as the Y1H, Y2H, and M1H assays as localization in and of itself is not evidence of TF functionality): any alternative isoforms whose localization changed from nuclear or both nuclear/cytoplasmic in the reference to solely cytoplasmic in either HEK293T or U2OS imaging assays were considered negative regulators. Any alternative isoforms with data in ≥ 2 functional assays (eY1H, Y2H, or M1H) that

had ≥ 1 difference in PPIs or PDIs, or $\log_2FC \geq 1$ or ≤ -1 (with at least 1 isoform having M1H signal above baseline), or a difference in subcellular localization in either cell line, were considered to be rewirers; any with 0 difference in PPIs or PDIs, and \log_2FC between -1 and 1 , and no differences in localization in either cell line were considered to be similar to the reference isoform. Only one isoform loses function across all tested axes (PPARG-3, see [Figure S7C](#)) and was filtered out of downstream analyses as likely non-functional.

TF Atlas mORF analyses

We downloaded data from the TF over-expression atlas (Joung et al.¹¹⁴): TF ORF library sequences (Table S1A from Joung et al.), TF over-expression scores (processed, Table S2B from Joung et al.), and single cell mRNA counts data (h5ad file, GEO: GSE216481). To intersect our clone collection with their library, we first attempted to match based on amino acid sequence; if this failed, we attempted to match based on annotated Ensembl transcript IDs. To examine the effect of TF over-expression on differentiation ([Figure S7G](#)), we used their processed “Diffusion difference” as the effect size and “Diffusion P-value” as the p -value. To examine how TF isoforms affect gene expression ([Figure 7L](#)), we re-processed scRNA-seq data from the author-provided h5ad matrix with associated cell barcode metadata. To control for variance in the number of cells expressed for each isoform and high dropout rate, we applied a metacell aggregation approach. In brief, 50 metacells were generated for each TF isoform, with each metacell produced from the average expression of 5 corresponding randomly selected representative cells with possible re-sampling. Differential expression between the TFiso1.0 alternative and reference isoform were generated in the software AltAnalyze,^{135,136} using the metaDataAnalysis function (fold > 0 , empirical Bayes t -test $p < 0.05$, FDR corrected). To perform gene set enrichment analyses ([Data S10](#)), we used the gseapy package prerank function to look for enrichment of MSigDB Hallmark 50 pathways and used the log fold-changes of gene expression between alternative and reference isoforms as the ranking metric. Thus, positive normalized enrichment scores indicate relative up-regulation in the alternative isoform compared to the reference isoform, while negative normalized enrichment scores indicate the reverse. We used an FDR cutoff of < 0.25 , which the authors of GSEA consider to be a reasonable threshold for hypothesis generation (GSEA Software Wiki, Broad Institute). For [Figure 7L](#), we removed redundant gene sets that overlap by more than 5% for clarity.

Paired tumor/normal TCGA analysis

We used paired (i.e., from the same patient) tumor/normal samples from breast cancer (BRCA), lung adenocarcinoma (LUAD), and head and neck squamous cell cancer (HNSCC) ([Data S9](#)). We only considered primary tumors and ignored metastases. In some cases, there were repeat samples from the same patient – we randomly sampled 1 paired sample in each case to make the final list of 112 BRCA patients, 105 LUAD patients, and 43 HNSCC patients. To find isoforms that show significant differences between tumor and normal controls, we performed paired Wilcoxon tests between the fractional isoform expression in normal samples compared to the fractional isoform expression in tumor samples, requiring a minimum gene-level expression of 1 TPM in both normal and tumor samples in a minimum of 20 such paired samples. We then corrected these p -values for multiple hypothesis testing using the Benjamini-Hochberg method and an FDR of 0.05 ([Data S9](#)). Oncogene and tumor suppressor annotations are from the OncoKB database.^{163,164}

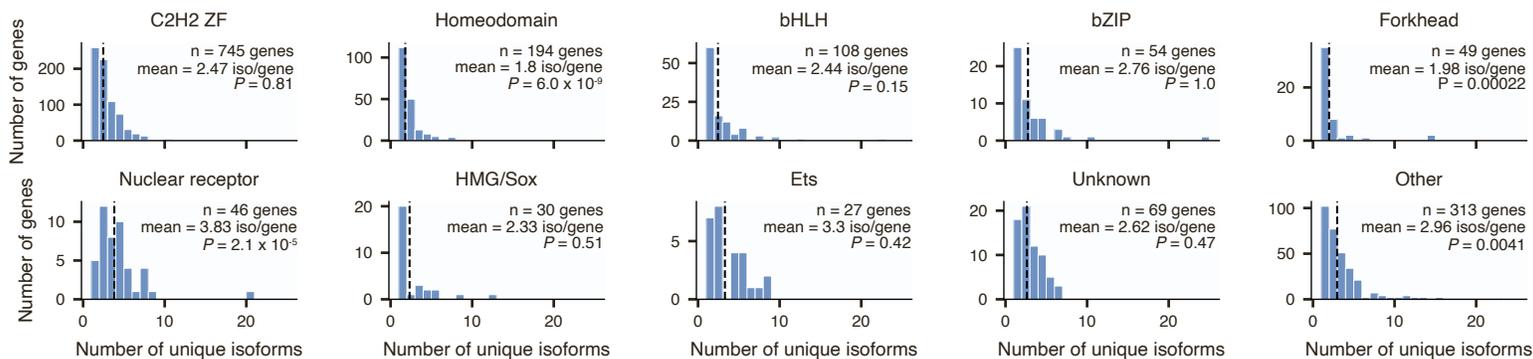
Supplemental information

**Widespread variation in molecular interactions
and regulatory properties
among transcription factor isoforms**

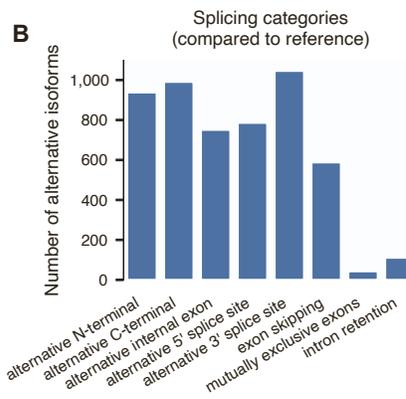
Luke Lambourne, Kaia Mattioli, Clarissa Santoso, Gloria Sheynkman, Sachi Inukai, Babita Kaundal, Anna Berenson, Kerstin Spirohn-Fitzgerald, Anukana Bhattacharjee, Elisabeth Rothman, Shaleen Shrestha, Florent Laval, Brent S. Carroll, Stephen P. Plassmeyer, Ryan J. Emenecker, Zhipeng Yang, Deepa Bisht, Jared A. Sewell, Guangyuan Li, Anisa Prasad, Sabrina Phanor, Ryan Lane, Devlin C. Moyer, Toby Hunt, Dawit Balcha, Marinella Gebbia, Jean-Claude Twizere, Tong Hao, Alex S. Holehouse, Adam Frankish, Josh A. Riback, Nathan Salomonis, Michael A. Calderwood, David E. Hill, Nidhi Sahni, Marc Vidal, Martha L. Bulyk, and Juan I. Fuxman Bass

Number of unique annotated isoforms per TF family

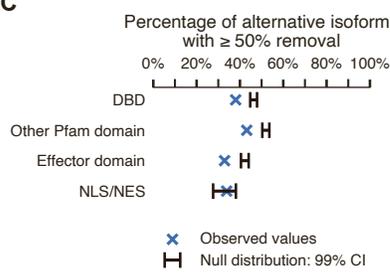
A



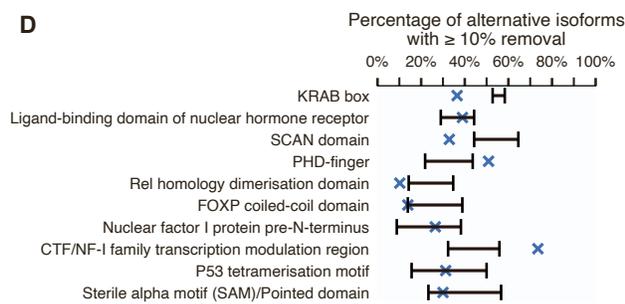
B



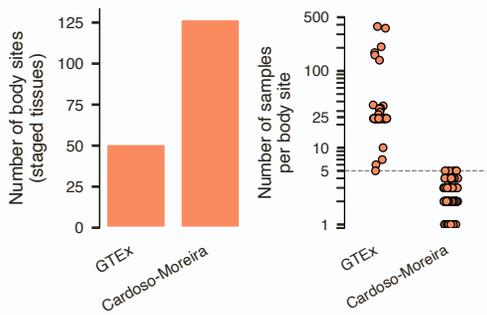
C



D



E



F

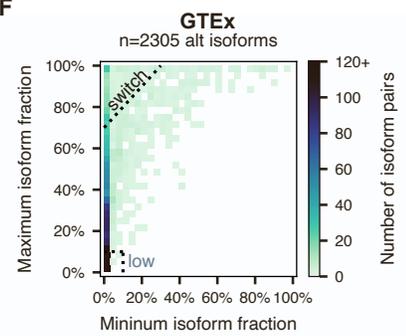


Figure S1

Figure S1: Sequence and expression diversity of annotated TF isoforms, related to Figure 1

- A.** Number of unique annotated protein isoforms per TF family. Mean number of isoforms per gene is shown as a dotted vertical line. Only TF families with ≥ 20 genes are shown; the remaining TF families are collapsed into the “other” category. P-values from two-sided Mann-Whitney tests.
- B.** Number of alternative isoforms that exhibit various sequence differences compared to their cognate reference isoforms. Categories are not mutually exclusive (so an alternative isoform could exhibit both an alternative N-terminal and exon skipping, for example).
- C.** Observed fraction of alternative isoforms with $\geq 50\%$ removal of various protein domains (blue X) compared to the expected fraction (black error bars, 99% CI) as defined by a null model assuming the domain is randomly positioned along the protein. DBD = DNA-binding domain; NLS/NES = nuclear localization/export signal.
- D.** Analogous to **C**, but showing specific domains that are collapsed in the “Other Pfam domains” category in **C**. Only domains with ≥ 30 annotation instances are shown.
- E.** Number of unique body sites (i.e., staged tissues) (left) and number of samples per body site (right) for both GTEx and Developmental RNA-seq. Developmental RNA-seq has more unique body sites, but fewer individual samples per body site, compared to GTEx.
- F.** Maximum isoform fraction compared to the minimum isoform fraction of alternative TF isoforms in re-sampled GTEx, where isoform fraction is defined as the expression level of an isoform normalized to the total expression level of its host gene. Dashed lines show the definitions used for isoforms that exhibit “switching” events and isoforms that remain lowly expressed. Only isoforms whose host genes are expressed at ≥ 1 TPM in ≥ 1 sample are shown.

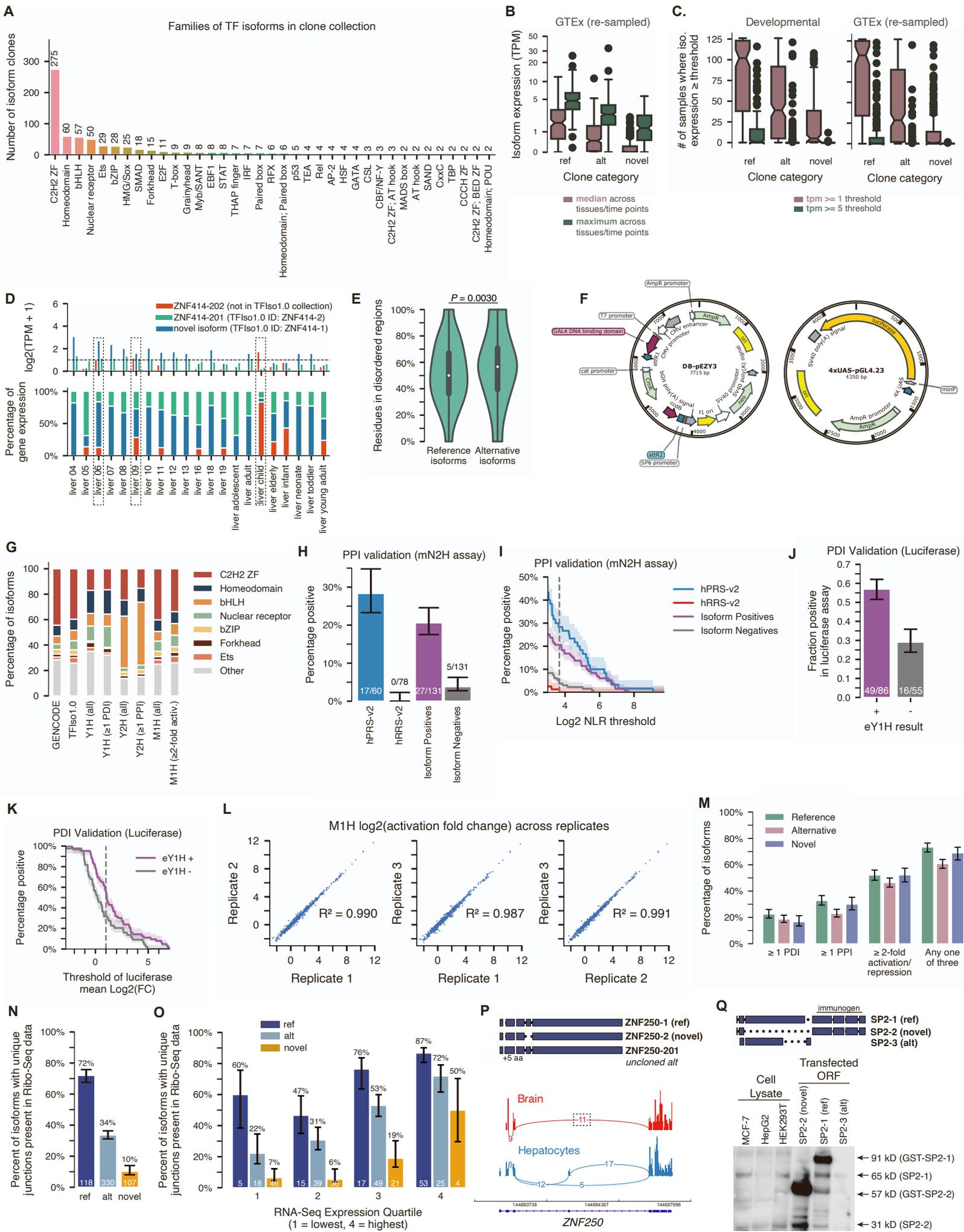


Figure S2

Figure S2: Overview of TFIso1.0 clone collection and TF molecular function assays, related to Figure 2

- A.** Number of clones in TFIso1.0 across all observed TF families.
- B.** Median and maximum expression levels (in TPM) in re-sampled GTEx RNA-seq data of reference, annotated alternative, and novel alternative isoforms in TFIso1.0.
- C.** Number of samples where reference, alternative, or novel TF isoforms are expressed ≥ 1 TPM or ≥ 5 TPM in Developmental RNA-seq and re-sampled GTEx. Box plots (**B** and **C**) show median, interquartile range (IQR), $1.5 \times$ IQR, and outliers.
- D.** Expression profile of a novel isoform in TFIso1.0, ZNF414-1. Log₂ TPM values (top) and isoform fraction (bottom) for each ZNF414 isoform. All liver samples from Developmental RNA-seq data are shown. Numbers in sample names correspond to weeks post-conception. Samples where ZNF414-1 is expressed ≥ 1 TPM are outlined.
- E.** Fraction of residues predicted to be in disordered regions comparing reference and alternative isoforms. White dot indicates the median, dark-gray box indicates IQR. P-value calculated using a two-sided permutation test.
- F.** Plasmids used in the M1H assay.
- G.** Percent of TF isoforms belonging to different TF families in GENCODE, the entire TFIso1.0 collection, those that have been successfully tested in each assay (“all” categories), and those that show evidence of function (≥ 1 PDI, ≥ 1 PPI, ≥ 2 -fold M1H activity) in each assay.
- H, I.** Results of testing our Y2H PPI data in the mN2H assay, along with positive and negative controls, displayed as a bar chart (**H**) and a titration across the readout value (**I**), with the cutoff displayed as a vertical dashed line. Error bars/bands are 68.3% Bayesian CI. hPRS-v2 = human positive reference set version 2; hRRS-v2 = human random reference set version 2.
- J, K.** Results of testing our Y1H PDI data in the luciferase assay, displayed as a bar chart (**J**) and a titration across the readout value (**K**). Error bars/bands are 68.3% Bayesian CI.
- L.** M1H activity fold-change correlation across 3 independent transfection replicates.
- M.** Proportion of isoforms exhibiting ≥ 1 PPI, ≥ 1 PDI, ≥ 2 -fold activation/repression in M1H, or any one of the three across reference, annotated alternative, and novel alternative isoforms, normalized to the total number of those isoforms in TFIso1.0. Error bars are 68.3% Bayesian CI.
- N, O.** Percentage of isoforms with unique exon-exon junctions that can be validated with Ribo-seq data (using a minimum threshold of at least 5 mapped reads in at least 1 sample), broken up by isoform category (**N**) and additionally by median RNA expression in the Developmental RNA-seq (**O**). The percentage values are annotated above the bars and the total numbers of isoforms in each category are annotated at the bottom of the bars.
- P.** Sashimi plot showing an example novel isoform, ZNF250-2, that was validated in the Ribo-seq data. The novel isoform was validated in both hepatocytes and brain, where it is the only isoform engaged by ribosomes (boxed number).
- Q.** Western blot showing endogenous expression of novel isoform SP2-2. Left lanes show unmodified cells, right lanes show the effect of over-expressing SP2 isoforms in HEK293T cells. The expression plasmid used expresses both unmodified and N-terminal GST-tagged (+26 kD) ORFs. SP2-3 was included as a negative control for the SP2 antibody, which used the C-terminal region of SP2-1 as an immunogen, and thus should not detect SP2-3. Less lysate was loaded into the transfection wells to minimize over-saturation of the primary GST-ORF bands compared to the endogenous bands.

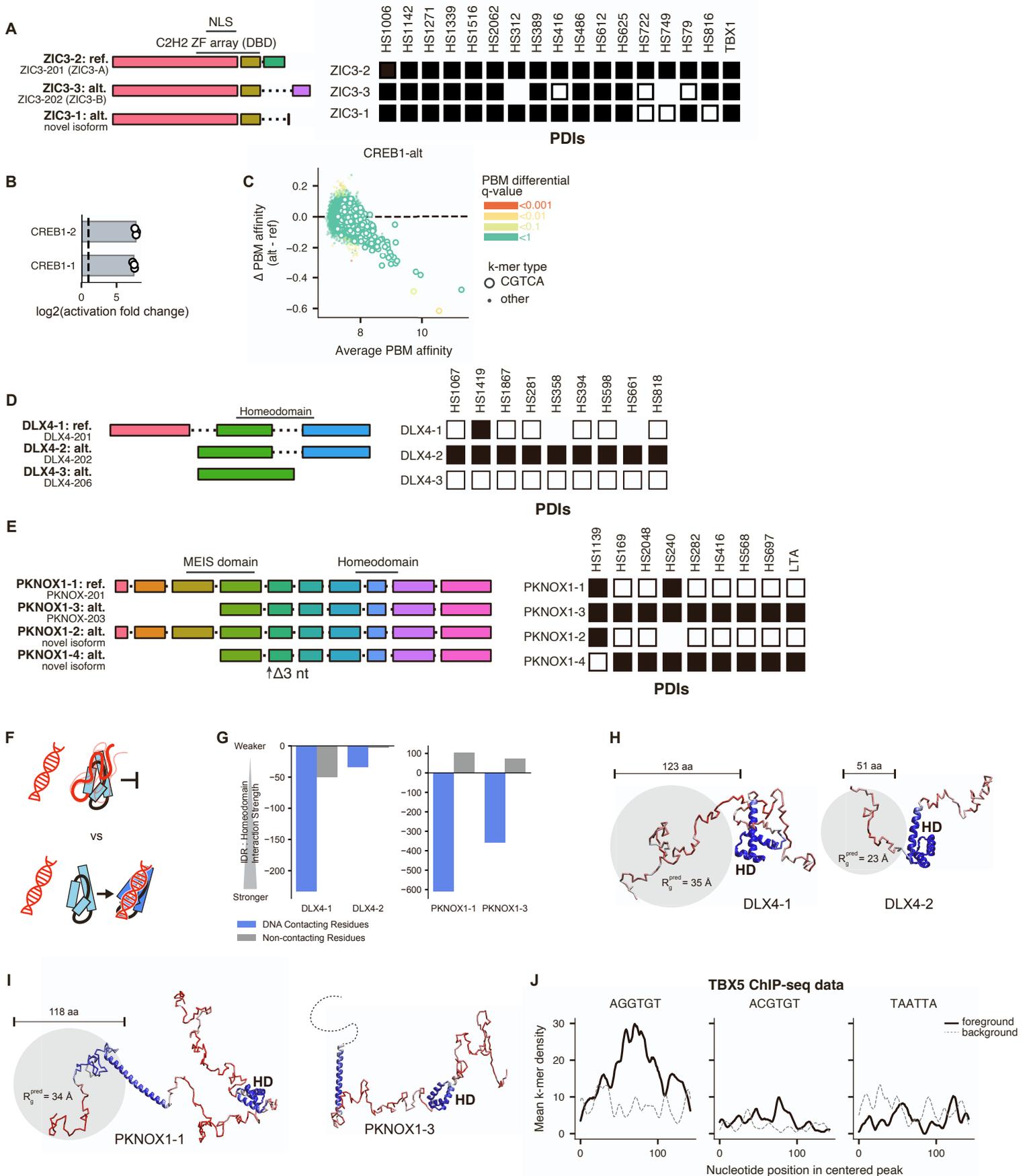


Figure S3

Figure S3: DNA binding preferences of TF isoforms, related to Figure 3

A. Left: exon diagrams of the 3 ZIC3 isoforms included in TFIso1.0. NLS = nuclear localization sequence. Right: PDI results from the Y1H assay. Missing boxes correspond to baits that were not successfully tested against one of the isoforms.

B. M1H activity results for CREB1 isoforms.

C. MA plot showing the PBM results comparing the alternative and reference isoforms of CREB1 for every 8-mer. Points are colored by the differential affinity q-value. Open circles correspond to 8-mers containing the canonical CREB1 5-mer CGTCA (or its reverse complement). Points below the dashed horizontal line correspond to 8-mers for which the alternative isoform shows reduced affinity compared to the reference isoform.

D. Left: exon diagrams of the 3 DLX4 isoforms. Right: PDI results from the Y1H assay.

E. Left: exon diagrams of the 4 PKNOX1 isoforms. Right: PDI results from the Y1H assay.

F. Schematic of the autoinhibitory model. IDRs with favorable interactions with those residues involved in DNA binding can compete with DNA, effectively acting as a locally tethered competitive inhibitor for DNA binding.

G. Predicted mean-field interaction between all IDRs present in each isoform and DNA contacting residues (blue bars, left) or all other residues (grey bars, right) on the homeodomains. More negative values reflect more attractive interactions.

H. Structural representation for DLX4 isoforms created with IDRs reconstructed at their predicted dimensions in isolation. All IDRs can interact directly with the homeodomain.

I. Structural representations for PKNOX1 isoforms created with IDRs reconstructed at their predicted dimensions in isolation. All IDRs can interact directly with the homeodomain.

J. Enrichment of the canonical TBX5 6-mer AGGTGT, the altered 6-mer ACGTGT, or a negative control 6-mer TAATTA (or each of their reverse complements) across TBX5 ChIP-seq peaks. Solid black lines show enrichment in ChIP peaks (foreground); dotted grey lines show enrichment in matched genomic negative control regions (background). Lines show the moving average of k-mer density, using a window of 8 nucleotides.

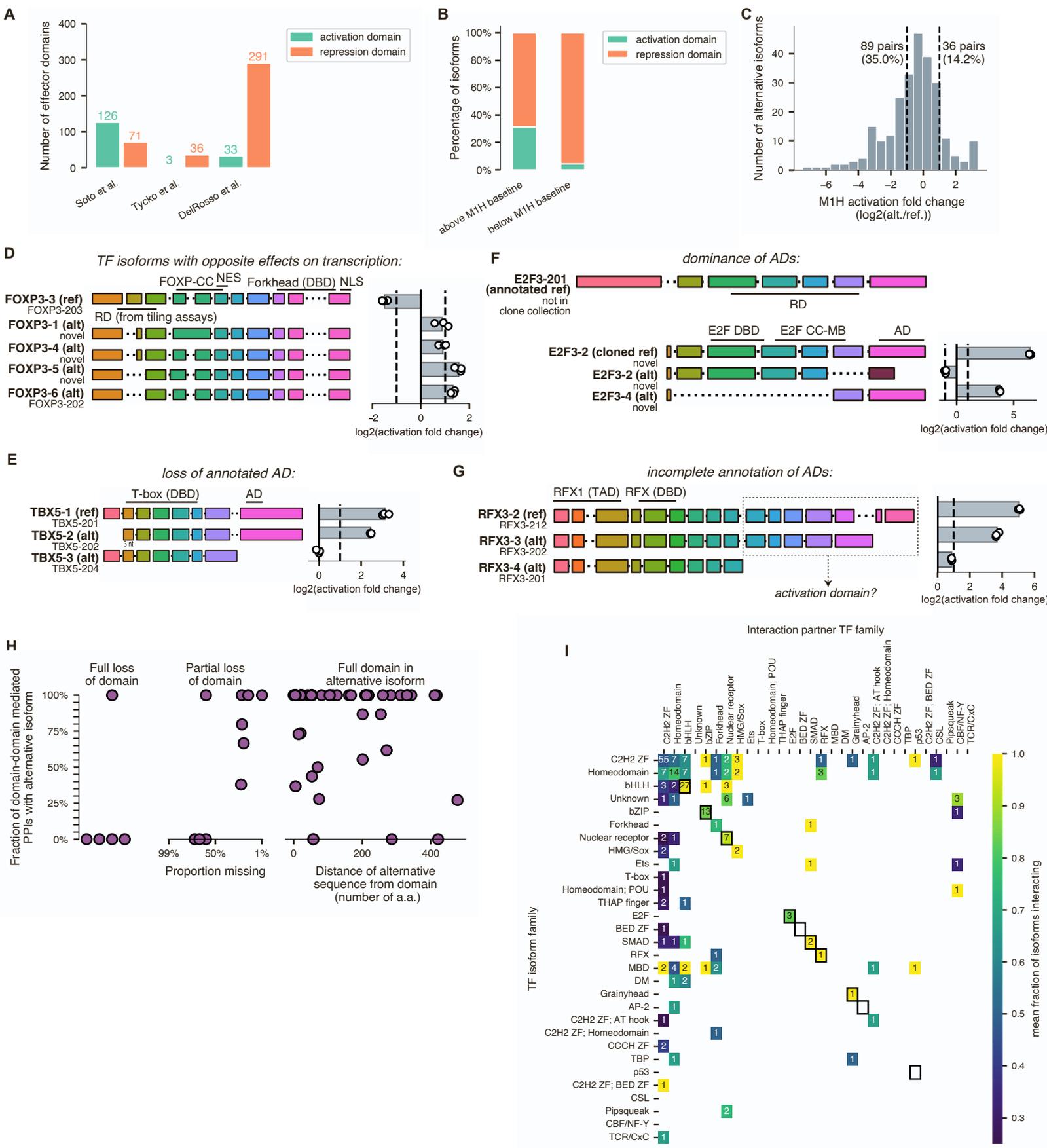


Figure S4

Figure S4: Transcriptional activity and protein binding preferences of TF isoforms, related to Figure 4

A. Number of activation and repression domains annotated in each of the 3 studies used in this work. Note that Soto et al. is based primarily on literature curation, whereas Tycko et al. and DelRosso et al. are each large-scale tiling screens.

B. Percent of TF isoforms containing an either annotated activation or repression domain that are either above (≥ 1) or below (≤ -1) the M1H baseline activity levels.

C. M1H activity changes ($\log_2(\text{alternate isoform activity/reference isoform activity})$) across all pairs assayed.

D-G. Examples of TF genes with isoforms that have: opposite effects on transcription (**D**); lose an annotated activation domain (**E**); show dominance of annotated activation domains over repression domains (**F**); and show potentially incomplete effector domain annotation (**G**). Left: exon diagrams. Right: M1H results.

H. Fraction of the subset of PPIs mapped to domain-domain interactions that are retained in each alternative isoform, relative to the reference isoform, in cases where the alternative isoform fully or partially loses the interacting domain, or contains the full domain.

I. Full heatmap showing the fraction of isoforms interacting for combinations of families of TF isoforms (y-axis) and families of TF PPI partners (x-axis). Within-family dimerizations are therefore denoted on the diagonal of the heatmap. TF families that bind DNA as obligate dimers are marked with outlined black boxes on the diagonal. The number within each cell indicates the number of PPIs that fall into that specific category, and the color denotes the mean fraction of isoforms interacting.

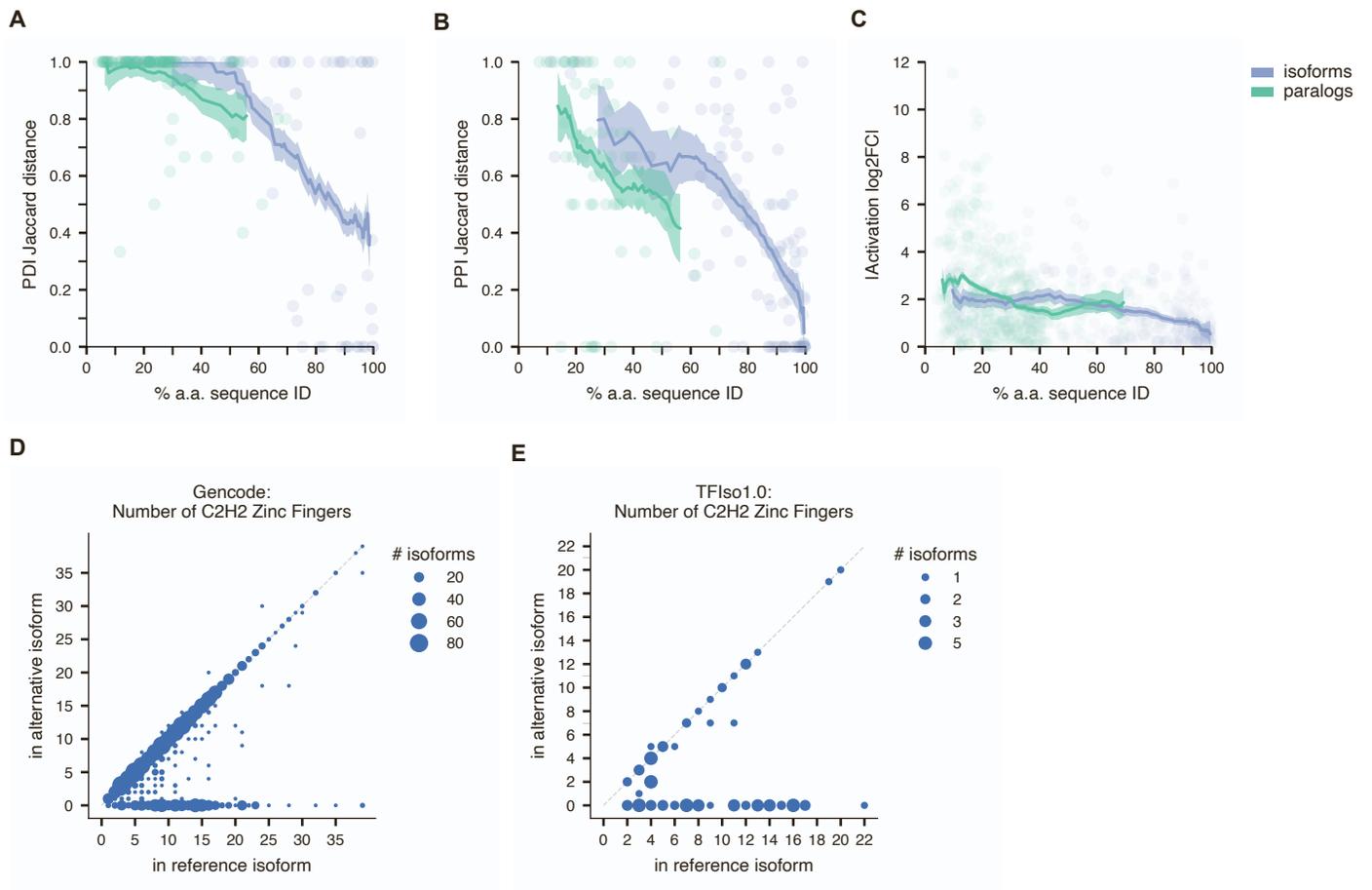


Figure S5

Figure S5: Functional differences between TF isoforms and TF paralogs, related to Figure 6

A-C. Jaccard distance in PDIs (**A**) or PPIs (**B**) or the absolute log₂ fold-change in M1H activity (**C**) between pairs of isoforms (blue) or paralogs (green) as compared to their pairwise amino acid sequence similarity. Lines show mean values across a sliding window of 40%; error bands are 68.3% Bayesian CI; P-values from two-sided permutation test.

D-E. Number of zinc fingers in annotated zinc finger array TFs in either the reference isoform or alternative isoform; size of the circles corresponds to the number of isoform pairs in each bin. **D**: considering all isoforms annotated in GENCODE; **E**: considering only isoforms in TFIso1.0.

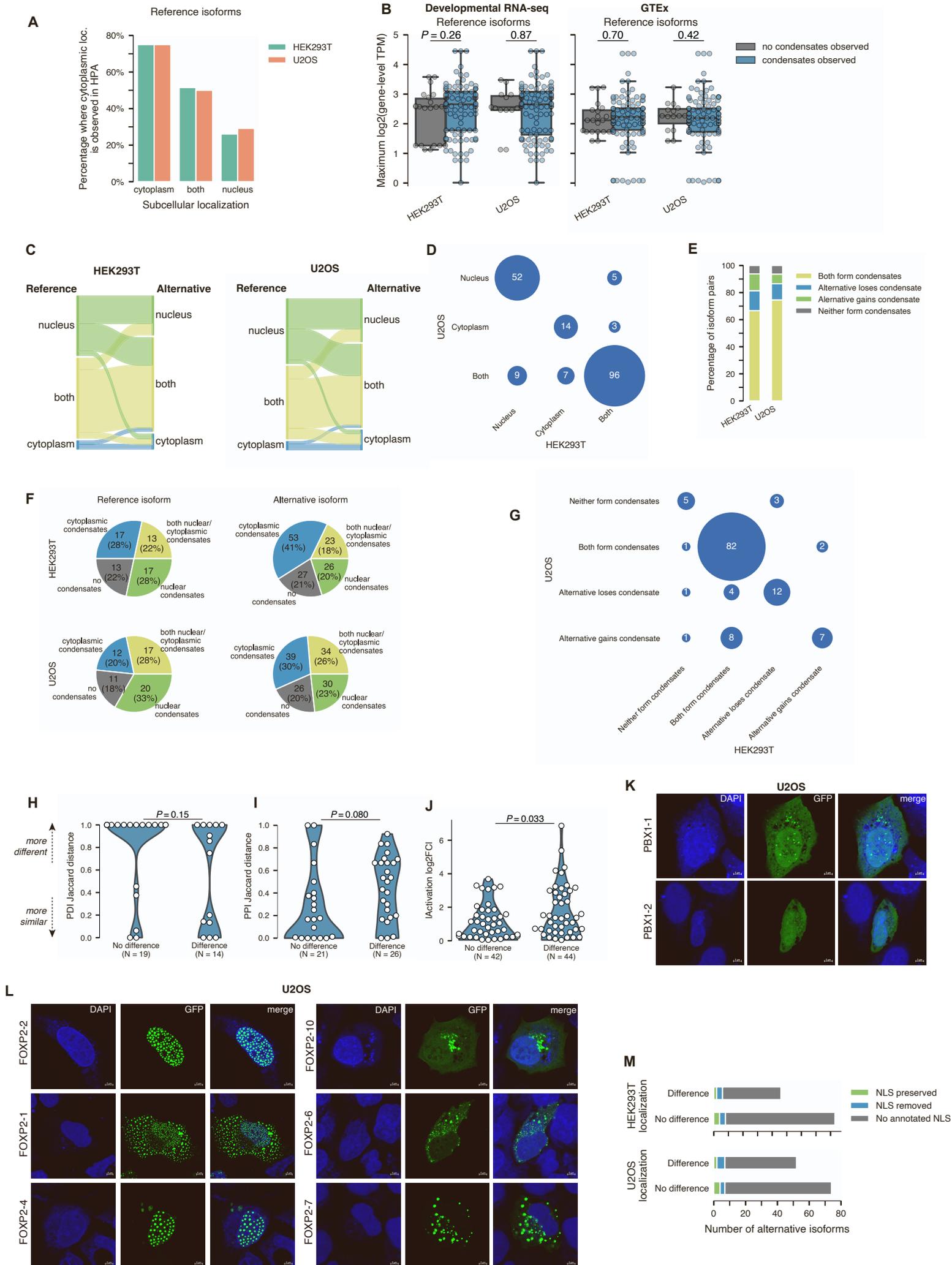


Figure S6

Figure S6: Condensate formation and subcellular localization differences between TF isoforms, related to Figure 6

A. Percentage of reference isoforms that show cytoplasmic localization in the Human Protein Atlas compared to their localization in our high-throughput imaging assay in either HEK293T or U2OS cells.

B. Maximum expression of reference isoforms (in TPM) in either Developmental RNA-seq or GTEx broken up by whether or not the reference isoform forms condensates in our high-throughput imaging assay in either HEK293T or U2OS cells. P-values shown are from a two-sided Mann-Whitney test. Circles are data points, box plots show median, interquartile range (IQR), and 1.5× IQR.

C. Localization of alternative isoforms as compared to their cognate reference isoforms in HEK293T and U2OS cells.

D Agreement in localization calls among all TF isoforms in HEK293T and U2OS cells. Size of the circle is proportional to the number of TF isoforms in that bin (shown in white).

E. Percent of reference-alternative isoform pairs where both show condensates, the alternative gains or loses condensates compared to the reference, or neither isoform shows condensates in either HEK293T or U2OS cells.

F. Condensate localization among reference and alternative isoforms in HEK293T and U2OS cells.

G. Agreement in condensate call differences among reference-alternative TF isoform pairs in HEK293T and U2OS cells. Size of the circle is proportional to the number of TF isoforms in that bin (shown in white).

H.-J. Differences in TF molecular functions (PDIs, **H**; PPIs, **I**; transcriptional activity, **J**) between alternative-reference TF isoform pairs that either show no difference in condensate formation or localization or those that do. For these analyses, only TF isoform pairs with consistent results across the two imaging cell lines were considered. P-values calculated using a two-sided permutation test.

K. Representative images of the expression of GFP fusions with PBX1 isoforms U2OS cells (63x magnification).

L. Representative images of the expression of GFP fusions with FOXP2 isoforms in U2OS cells (63x magnification).

M. Number of alternative isoforms with NLS preserved or lost, relative to the reference isoform, split by whether there was an observed difference in localization between the reference and alternative isoform in HEK293T or U2OS cells.

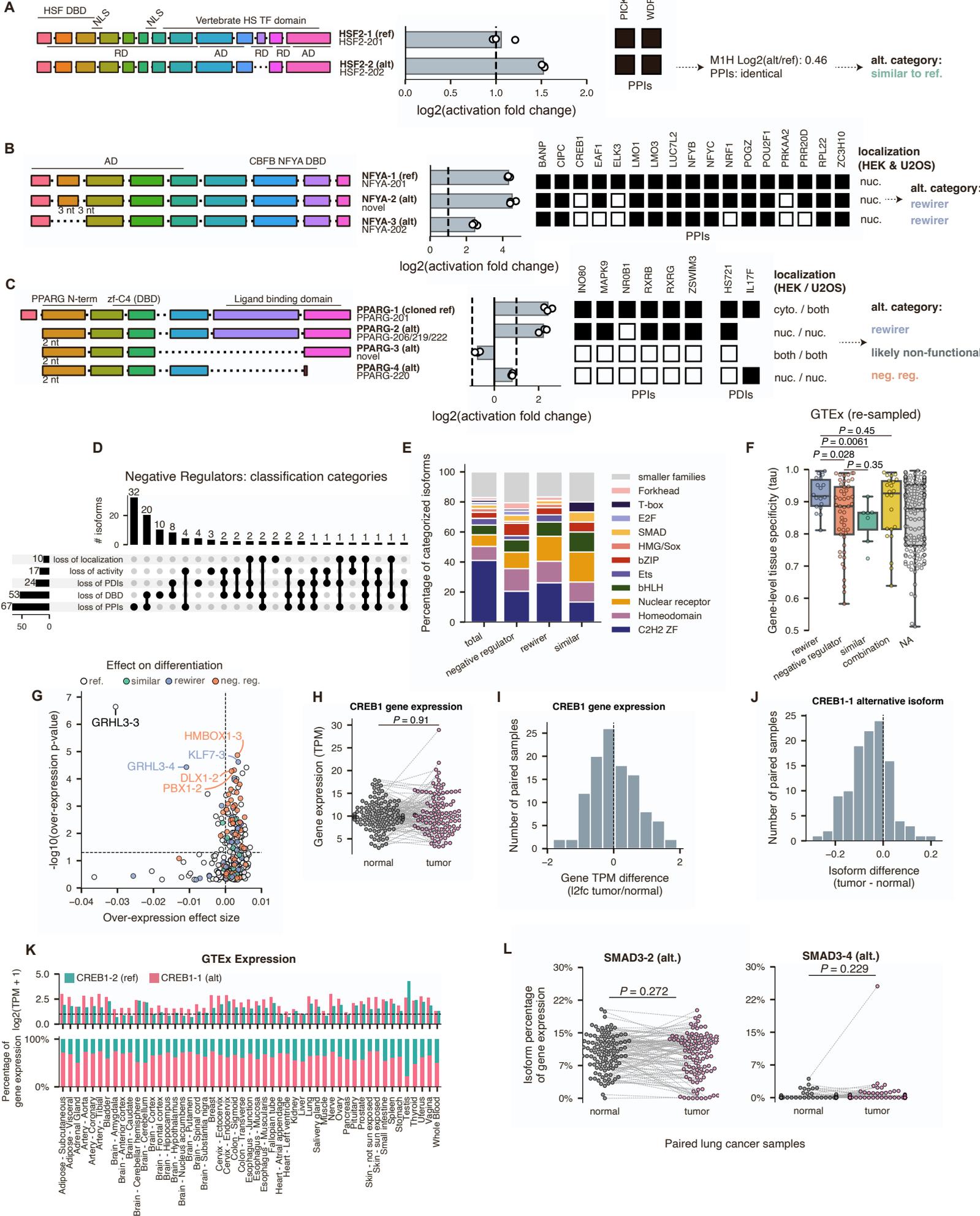


Figure S7

Figure S7: Alternative TF isoforms can function as negative regulators, related to Figure 7.

- A-C.** Examples of TF genes with isoforms that are similar to the reference (**A**), rewirers (**B, C**), negative regulators (**C**), and likely non-functional (**C**). For each gene, all assays (Y1H, Y2H, M1H, localization) with data are shown.
- D.** Details of negative regulator TF isoform classification, broken-down by the different assays.
- E.** Percentage of TF families among each category of alternative isoform.
- F.** Gene-level tissue specificities (tau metric), calculated from the re-sampled GTEx RNA-seq data, among TF genes with either only rewirer alternative isoforms, only negative regulator alternative isoforms, only alternative isoforms that are similar to reference, some combination of the above, or only alternative isoforms that were unable to be classified (NA). P-values are from a two-sided Mann-Whitney test. Circles are data points, box plots show median, interquartile range (IQR), and 1.5× IQR.
- G.** Effect of TF over-expression on differentiation (Diffusion difference and associated P-value) as calculated in the TF mORF Atlas, for isoforms in TFIso1.0.
- H.** Total CREB1 gene expression levels in matched breast cancer tumor and normal samples. P-value is from a two-sided Mann-Whitney test.
- I.-J.** Paired difference in total CREB1 gene expression (**I**) or CREB1-1 fraction (**J**) across matched breast cancer tumor and normal samples.
- K.** Expression levels of CREB1 isoforms in GTEx. Log2 TPM values (top) and isoform fraction (bottom) for each CREB1 isoform.
- L.** Isoform fraction of the additional two alternative SMAD3 isoforms in matched lung cancer tumor and normal samples (paired from the same patient, denoted using dotted lines). P-value is from a two-sided paired Wilcoxon test, adjusted for multiple hypothesis correction.