















Development of compact transcriptional effectors using high-throughput measurements in diverse contexts

Received: 8 May 2023

Accepted: 20 September 2024

Published online: 01 November 2024

 Check for updates

Josh Tycko ^{1,2,12}, Mike V. Van ^{3,12}, Aradhana¹, Nicole DelRosso⁴, Hanrong Ye⁵, David Yao ¹, Raeline Valbuena ¹, Alun Vaughan-Jackson^{1,6}, Xiaoshu Xu ⁷, Connor Ludwig⁷, Kaitlyn Spees¹, Katherine Liu ³, Mingxin Gu¹, Venya Khare⁵, Adi Xiyal Mukund⁴, Peter H. Suzuki⁷, Sophia Arana¹, Catherine Zhang⁸, Peter P. Du ⁸, Thea S. Ornstein⁵, Gaelen T. Hess ⁹, Roarke A. Kamber¹, Lei S. Qi ^{6,7,10}, Ahmad S. Khalil ^{5,11}, Lacramioara Bintu ⁷  & Michael C. Bassik ^{1,10} 

Transcriptional effectors are protein domains known to activate or repress gene expression; however, a systematic understanding of which effector domains regulate transcription across genomic, cell type and DNA-binding domain (DBD) contexts is lacking. Here we develop dCas9-mediated high-throughput recruitment (HT-recruit), a pooled screening method for quantifying effector function at endogenous target genes and test effector function for a library containing 5,092 nuclear protein Pfam domains across varied contexts. We also map context dependencies of effectors drawn from unannotated protein regions using a larger library tiling chromatin regulators and transcription factors. We find that many effectors depend on target and DBD contexts, such as HLH domains that can act as either activators or repressors. To enable efficient perturbations, we select context-robust domains, including ZNF705 KRAB, that improve CRISPRi tools to silence promoters and enhancers. We engineer a compact human activator called NFZ, by combining NCOA3, FOXO3 and ZNF473 domains, which enables efficient CRISPRa with better viral delivery and inducible control of chimeric antigen receptor T cells.

Over 10% of human genes encode proteins that localize to the nucleus and function to regulate gene expression at the level of mRNA transcription^{1–3}. These transcription factors (TFs) and chromatin regulator proteins represent a reservoir of potentially useful effector domains for constructing transcriptional and epigenomic perturbation tools for synthetic biology and dissecting biological processes^{4–8}. Transcriptional effector domains, also known as activators and repressors, are the regions of these proteins that can increase or silence transcription of a gene upon recruitment to its promoter^{1,9}. When fused to DNA-binding domains (DBD), they can be used as tools to manipulate

the expression of endogenous genes and long noncoding RNAs^{5,10,11} and control synthetic gene circuits¹² or endogenous regulatory elements^{13,14}, which has therapeutic potential^{15–17}.

However, transcriptional effectors can function differently depending on the biological context of their recruitment to a gene. For example, targeting of effector domains to the same genomic context but in different cell lines^{18–20} or in different mouse developmental stages *in vivo*²¹ can result in variable transcriptional and epigenetic effects. The target locus of effector recruitment can also have a role^{22–25}, wherein recruitment of effector domains to different promoters and enhancers

A full list of affiliations appears at the end of the paper.  e-mail: lbintu@stanford.edu; bassik@stanford.edu

in the same cell type results in distinct effects on transcription^{13,18,19,26–28}. Finally, the DBD can affect effector function^{19,29}.

We currently lack a system-level understanding of how frequently effector function depends on the target, cell type or DBD context, which is critical for identifying effectors that are robust across these conditions. Recently, pooled screening methods that measure effector strength for large libraries of domains have greatly expanded the list of sequences known to function as effectors^{30–34}, but these scalable methods use synthetic reporter genes instead of targeting endogenous loci and have not yet been used to compare activities across many contexts. Meanwhile, other salient features like deliverability, size, expression level and off-target toxicities remain uncharacterized for the majority of effectors.

To address these questions, we performed high-throughput recruitment (HT-recruit) to measure the effector function of 5,092 Pfam-annotated domains from human nuclear proteins (the Pfam library) across 17 biological contexts and integrated this data with two pre-existing datasets³². By adapting HT-recruit to use the programmable DBD dCas9, we were able to target endogenous promoters and enhancers. Using these large-scale datasets, we identify forms of context-specificity, which varied by effector family and were striking in the helix-loop-helix (HLH) family of domains. We then confirmed and expanded these results with screens of 114,288 tiling sequences from all human chromatin regulators and TFs (the CRTF tiling library) across two endogenous gene contexts, which we integrated with three pre-existing reporter gene datasets³⁴. This comprehensive testing of short human effector domains in human cells across multiple contexts allowed us to select robust activators and repressors to be used in synthetic biology tools. Finally, with additional measurements of effector expression level, deliverability and cellular toxicity, we developed improved effectors for CRISPR interference (CRISPRi), CRISPR activation (CRISPRa) and synthetic TF circuits. For repression, we find Krüppel-associated box (KRAB) domains are the most potent, and from the entire set of 336 human KRAB domains, we find several top KRAB domains that outperform the traditionally used ZNF10 KRAB. From the activation domains, we constructed a short potent combination comprising NCOA3, FOXO3 and ZNF473 domains (NFZ) whose improved utility for CRISPRa and synthetic TFs we demonstrate in macrophages, induced pluripotent stem cells (iPSCs) and T cells.

Results

HT-recruit quantifies effector function across contexts

We set out to measure transcriptional effectors quantitatively across biological contexts by performing HT-recruit screens using two DBDs (rTetR and dCas9) to recruit effectors to various gene targets in two cell types (Fig. 1a and Extended Data Fig. 1a). First, to study different

promoter contexts while keeping other parameters constant, we developed a set of reporters with varied promoter origins and strengths (Fig. 1b). To study how effector–promoter interactions differ across cell types, we installed these reporters at the AAVS1 safe harbor in both K562 and HEK293T cells (Fig. 1a). As expected, all minimal promoters (minimal cytomegalovirus (minCMV), non-TATA box-containing X chromosome and chromosome 21 (NTX and NT21) promoters) started OFF and could be activated with the VP64 activator domain (Fig. 1b). Two of the constitutive promoters, pEF1 α and UbC, started ON and could be repressed by the ZNF10 KRAB domain (Fig. 1b). The Rous sarcoma virus (RSV) promoter was rapidly silenced upon installation in both cell types (Extended Data Fig. 1b), so we did not use it for screens. In certain cases we observed promoter silencing to be cell-type specific—the phosphoglycerate kinase 1 (PGK) promoter was constitutively ON in K562 cells but was background silenced in HEK293T cells (Fig. 1b). Based on this observation, we decided to use PGK as a repressible promoter in K562 and as an activatable promoter in HEK293T cells (Fig. 1b). Overall, the expression levels for each promoter were similar across the two cell types ($R^2 = 0.86$) except for the PGK promoter (Extended Data Fig. 1c).

To extend HT-recruit to endogenous loci, we used dCas9 and targeted the promoters of endogenous genes encoding cell surface proteins. Targeting surface proteins allows us to use fluorescence antibodies to immunostain cells, thus providing a way to monitor single-cell gene expression variability during individual recruitment assays by flow cytometry and to magnetically separate a large number of ON and OFF cells during HT-recruit (Fig. 1a). For a repressible context, we targeted the highly expressed surface marker CD43 in K562 cells. First, we individually recruited either dCas9 alone or dCas9–KRAB with single guide RNAs (sgRNAs) targeting the CD43 transcriptional start site (TSS) and found two sgRNAs, sg10 and sg15, for which repression depended on the KRAB repressor. Similarly, we identified sgRNAs with which dCas9–VP64 or VPR could activate the lowly expressed *CD2*, *CD20* and *CD28* genes (Fig. 1c and Extended Data Fig. 1d,e).

Then, we generated lentiviral libraries with 5,092 Pfam domains from nuclear-localized human proteins and 499 random and 362 negative control sequences tiling the nonnuclear protein DMD (hereafter Pfam library) fused to rTetR or dCas9 and delivered them to cell lines with a low multiplicity of infection (MOI) such that most cells express one DBD-domain fusion. Using these components, we performed eight rTetR-targeted screens across six synthetic reporter contexts in two cell types and 11 dCas9-targeted screens with sgRNAs targeting endogenous genes plus two dCas9 screens at reporters using a sgRNA targeting the TetO motif. We also integrated data from two previous screens using rTetR and the Pfam library (Fig. 1d and Extended Data Fig. 1f,g). To measure the activities of sequences beyond annotated Pfam domains, we also used our recently designed 20 \times larger library

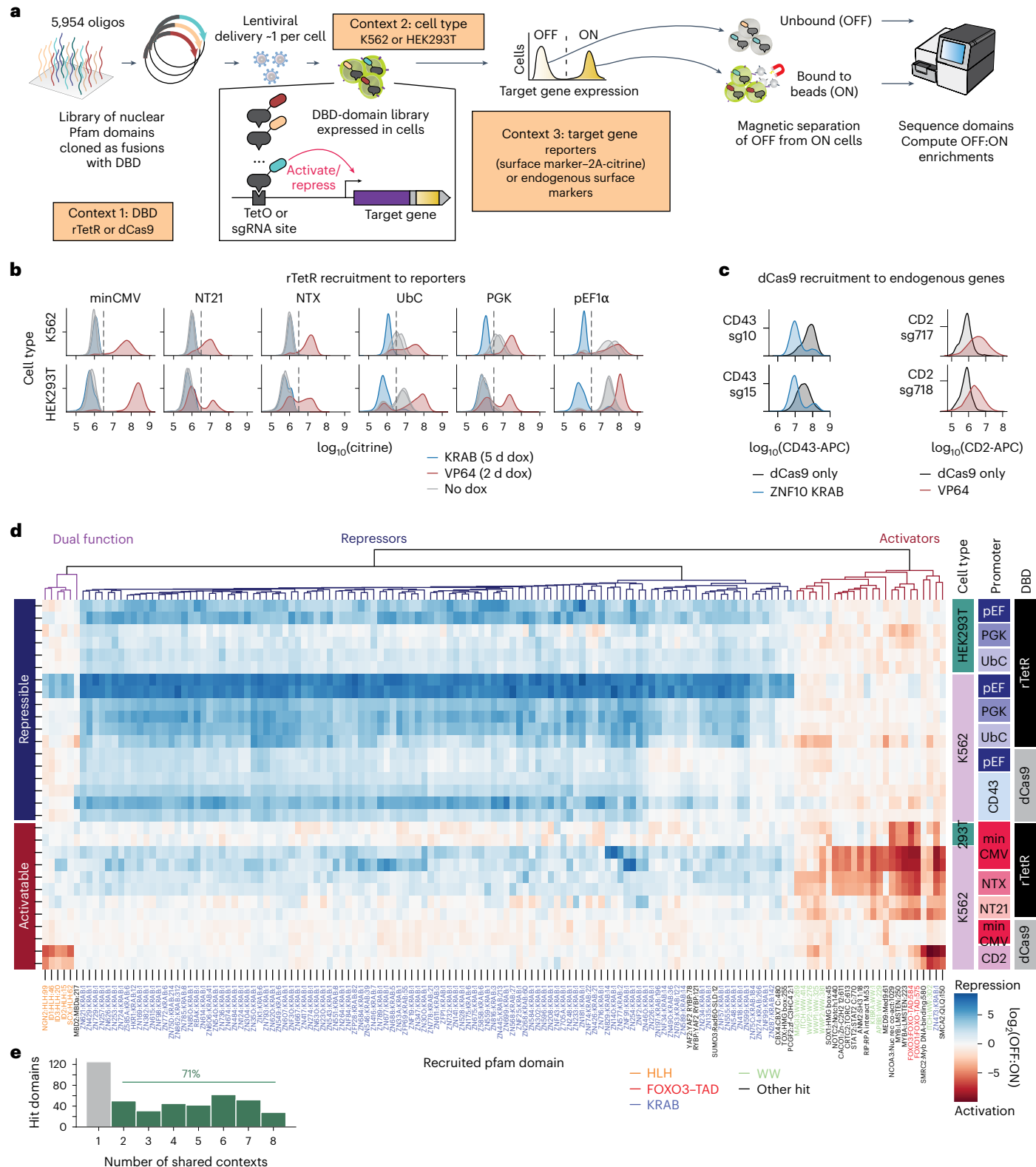
Fig. 1 | HT-recruit quantifies transcriptional effector functions across DBD, cell type and target gene contexts. a, Schematic representation of HT-recruit to quantify transcriptional effector function while varying the context of DBDs, cell type and target synthetic reporters or endogenous genes. A pooled library of Pfam domains from human nuclear proteins of ≤ 80 amino acids is synthesized as 300-mer DNA oligonucleotides, cloned downstream of the DBD (context 1) and delivered to cells (context 2) at a low MOI such that most cells express a single domain. The synthetic reporters encode a synthetic surface marker (lgx-hlg1-Fc-PDGFR β) and fluorescent marker (citrine), while the endogenous target genes encode surface markers (context 3). After the recruitment of Pfam domains, ON and OFF cells were magnetically separated using beads that bind these synthetic or endogenous surface markers, and the domains were sequenced in the bound and unbound populations to compute enrichments. **b**, Expression of synthetic reporters. Positive control effectors, such as ZNF10 KRAB repressor or VP64 activator, were stably delivered by lentivirus. Cells were treated with 1,000 ng ml⁻¹ doxycycline for 5 days for repression and 2 days for activation (or untreated as a negative control), and citrine expression was measured by flow cytometry after gating for rTetR delivery (mCherry⁺; $n = 2$ infection replicates

shown as histogram curves). **c**, Expression of endogenous surface marker genes in K562 cells measured by immunostaining and flow cytometry. dCas9 fusions and sgRNAs were delivered by lentivirus and selected for by blasticidin and puromycin, respectively. Data are gated for sgRNA delivery (mCherry⁺ in CD43 and GFP⁺ in CD2 samples) and for dCas9 (BFP⁺; $n = 1$ infection replicate). **d**, Transcriptional effector hits' activation and silencing activity across different target gene, DBD and cell-type recruitment contexts ($n = 2$ biological replicates per screen). A subset of effectors (columns) are shown that are a hit at a high threshold of 3 s.d. stronger than the median of the poorly expressed domains in ≥ 2 samples (rows; $n = 143$). Unbiased column clustering shows three major clusters of effectors (top). dCas9 targets pEF1 α and minCMV with sgTetO-1, CD2 with sg717 and CD43 with sg10 (upper two rows) and sg15. Column labels on the bottom show the protein, Pfam domain and domain start position within the protein; select Pfam domain families are colored. Rows are manually ordered, with the targets that are predominantly repressible above the activatable targets. **e**, Distribution of the number of screen contexts in which a Pfam domain was a hit effector in two replicates. Domains that are shared hits in multiple contexts are colored green.

that tiles all human CRTFs in 80 amino acid tiles at ten amino acid steps ($n = 128,565$ elements, hereafter CRTF tiling library)³⁴ and recruited it with dCas9 to CD2 and CD43 in K562 cells (Extended Data Fig. 2a,b). We additionally integrated data from three previously performed rTetR screens with the CRTF library at pEF, PGK and minCMV³⁴.

To minimize confounding effects due to protein stability, we identified domains that were well-expressed in both cell types and when fused to either DBD. Expression was measured by permeabilizing

the cells, staining with an anti-FLAG antibody, sorting into high and low protein expression level bins and then sequencing the domains in those two cell populations ($R^2 = 0.44-0.78$ across comparisons of replicates; Extended Data Fig. 3a). Overall, domain expression was similarly correlated between cell types ($R^2 = 0.6$; Extended Data Fig. 3b) and whether they were fused to rTetR or dCas9 ($R^2 = 0.56$; Extended Data Fig. 3c,d). We labeled a domain as well-expressed if its score was 1σ above the median of the random controls. Meanwhile, the poorly



expressed domains, which included most random sequence controls, served as a large set of negative controls for activation and repression screens to set our hit threshold.

After filtering out poorly expressed domains, we conservatively called activation and repression hit domains with $\log_2(\text{OFF:ON})$ scores 2 s.d. beyond the median of poorly expressed elements (higher scores for repressors and lower for activators). This pipeline called 435 Pfam domains as hits in both replicates of a context, with 71% of these hits being rediscovered in multiple contexts (Fig. 1e and Supplementary Table 1). We also called 1,261 tiled regions as hit domains from the larger CRTF tiling library (Methods; Supplementary Tables 2 and 3). The hits clustered into the following three categories: activators, repressors and dual-function effectors that are context-dependent activators or repressors (Fig. 1d). We then set out to investigate the diverse categories of context-dependent and context-robust effectors.

Activator function across targets and cell types

We first considered activator promoter-specificity when using rTetR to target the Pfam library at minimal promoter reporters in K562 cells. We found that these core promoters responded very similarly to activator recruitment, wherein 85% of well-expressed NTX activators also activate minCMV, and do so with well-correlated strengths ($R^2 = 0.8$; Fig. 2a,b and Extended Data Fig. 4a,b), which we individually validated with flow cytometry. In contrast, only 9% of the well-expressed minCMV activators, such as FOXO3-TAD, were able to re-activate the background-silenced PGK promoter in HEK293T cells (Extended Data Fig. 4c,d).

To extend these experiments to endogenous gene targets, we used dCas9 recruitment. First, we targeted the minCMV reporter using a TetO-targeting sgRNA. dCas9 recruitment resulted in fewer hits than rTetR at the same reporter, perhaps due to weaker and/or lower copy number recruitment; however, the strongest hit, the Med9 mediator component, was validated individually (Extended Data Fig. 4e–g). In contrast, when we used dCas9 to target the Pfam library to endogenous genes, we found 68 hits targeting CD2 with sg717, which systematically led to stronger activation than any other sgRNA targeting CD2, CD20 or CD28 (Fig. 2c and Extended Data Fig. 4h–l). Only 19% of rTetR-activator hits at minCMV recurred as dCas9 hits at CD2 (for example, ZNF473 KRAB and SMARCA2 QLQ); conversely, 89% of domains capable of activating CD2 were not hits in the rTetR- and reporter-based screens, including 24 HLH domains (for example, NGN2 and HAND2). The second strongest HLH activator domain was from HAND2, consistent with a recent report that full-length HAND2 protein can be an activator in a recruitment assay³³. Indeed, 77% of the HLH activator domain hits were from proteins that were also activators in that study ($P = 3.1 \times 10^{-3}$, Fisher's exact test; Extended Data Fig. 4m). Meanwhile, the two MYB LMSTEN domains were only activators when recruited with rTetR at reporters. Using the larger CRTF tiling library, dCas9 recruitment to CD2 identified 19 activator domains that were not hits with rTetR at minCMV. In total, 42% of the CD2-only activators again overlapped with HLH domains (for example, in MYOD1, PTF1A, MAX and ASCL5), while others were new activators that did not overlap any domain annotation (for example, PBRM1; Extended Data Fig. 5a–d). A notably strong shared activator hit with both rTetR at minCMV and dCas9 at CD2 was the DUX4 C-terminus, which interacts with histone acetyltransferase P300 (ref. 35). Therefore, rTetR recruitment to different minimal promoters identified similar activators, while dCas9 recruitment to an endogenous gene identified a largely distinct set of activators including HLH domains (Supplementary Tables 1–3).

We then investigated the cell-type dependence of different activators by comparing their effects in HEK293T versus K562 cells when recruited by rTetR at the minCMV reporter. In general, we found much greater differences between activators targeted to the same promoter in different cell types than when targeting distinct minimal promoters within the same cell type. Overall, only 19% of activators in K562

cells were shared across the cell types, including strong activators like FOXO-TADs, NCOAs and the ZNF473 KRAB domain (Fig. 2d,e and Supplementary Fig. 1a). Meanwhile, several WW, SRC Homology 3 (SH3), plant homeodomain (PHD) and the poorly characterized SMARCA2/SMARCA4 QLQ domains^{36,37} were much stronger activators in K562 cells, despite being similarly expressed in the two cell types as measured by FLAG staining (Fig. 2e, Extended Data Fig. 3c and Supplementary Fig. 1a–d). We hypothesized that one source of cell-type dependence of an effector could arise from competition for coactivators with its endogenous copy, and found that knocking down the endogenous SMARCA2/SMARCA4 resulted in increased activation by the recruited SMARCA2 QLQ (from 1.3% to 23.7% of HEK293T cells ON; Supplementary Fig. 1e–i).

Overall, while we cannot rule out all technical reasons that these activators are stronger in K562 cells (for example, related to the reporter chromatin state or rTetR), these results show that some activators function very differently across cell types (for example, KIBRA WW activates 91% of K562 versus 3% of HEK293T cells) while others are more consistent (for example, FOXO3-TAD activates 93% of K562 versus 87% of HEK293T cells; Supplementary Fig. 1a,b).

Repressor function across targets and cell types

We next tested repressor promoter-specificity using rTetR to target pEF1 α , PGK and UbC promoters in K562 cells. We found repression scores at the moderate strength PGK and UbC were highly correlated with each other, largely due to signal from KRAB repressors ($R^2 = 0.74$ for $n = 2,718$ well-expressed Pfam domains; Supplementary Fig. 2a), whereas the stronger pEF1 α was more silenced by weaker repressors such as HOX homeodomains (Supplementary Fig. 2b,c). We previously observed that HOX homeodomain repression strength correlated with the presence of an RKRR motif and positive charge in the homeodomain N-terminal region³², so we recruited motif- and charge-modifying HOX mutants to the pEF1 α reporter and confirmed they contribute to repression (Supplementary Fig. 2d,e). In sum, the similarly strong PGK and UbC reporters responded similarly to repressors, whereas the stronger pEF1 α reporter uniquely identified weak non-KRAB repressors and thus could be used to characterize their sequence dependencies.

To identify repressors of endogenous genes, we used dCas9 recruitment. When targeting the pEF1 α reporter with dCas9, we found fewer hits than with rTetR (possibly due to lower recruitment copy number), and all the hits were KRAB repressors (Supplementary Fig. 2f–h). In contrast, targeting dCas9 to the endogenous CD43, we found that in addition to shared hits including KRAB domains and a zinc finger from the IRF-2BP1 corepressor protein³⁸, a distinct set of domains that were not hits in the rTetR screens showed repressor activity, including the NuRD-interacting methyl-binding protein domains such as the P66_CC domain from P66A and the MBDa domain from MBD2 (which bind one another³⁹) (Fig. 2f and Supplementary Fig. 2i–k). Relatedly, an MBD2B repressor was previously shown to silence a reporter in HEK293T cells with dCas9 recruitment⁴⁰. Another repressor hit unique to targeting CD43 with dCas9 was the SAM1 domain from the putative polycomb protein L3MBTL4 (whose analog is involved in silencing in *Drosophila* cells⁴¹). MBDa, P66_CC, IRF-2BP1 and SAM1 silencing of CD43 depended on which sgRNA was used, while KRAB repressors were efficient with either of two sgRNAs (Supplementary Fig. 2i,j). Meanwhile, chromo and chromoshadow domains were only hit with rTetR.

dCas9 recruitment of the CRTF tiling library to the CD43 promoter confirmed these results, with the strongest shared repressors being KRAB domains, and additionally revealing 274 repressor domains that were not hits at pEF1 α , including from additional methyl-binding domain proteins (Extended Data Fig. 6a–c). This larger library also uncovered repressors in the unannotated regions of proteins, including CDCA7L, which is known to be a repressive protein⁴². In total, 78% of CD43 repressor domains were also hits in the rTetR screens, and

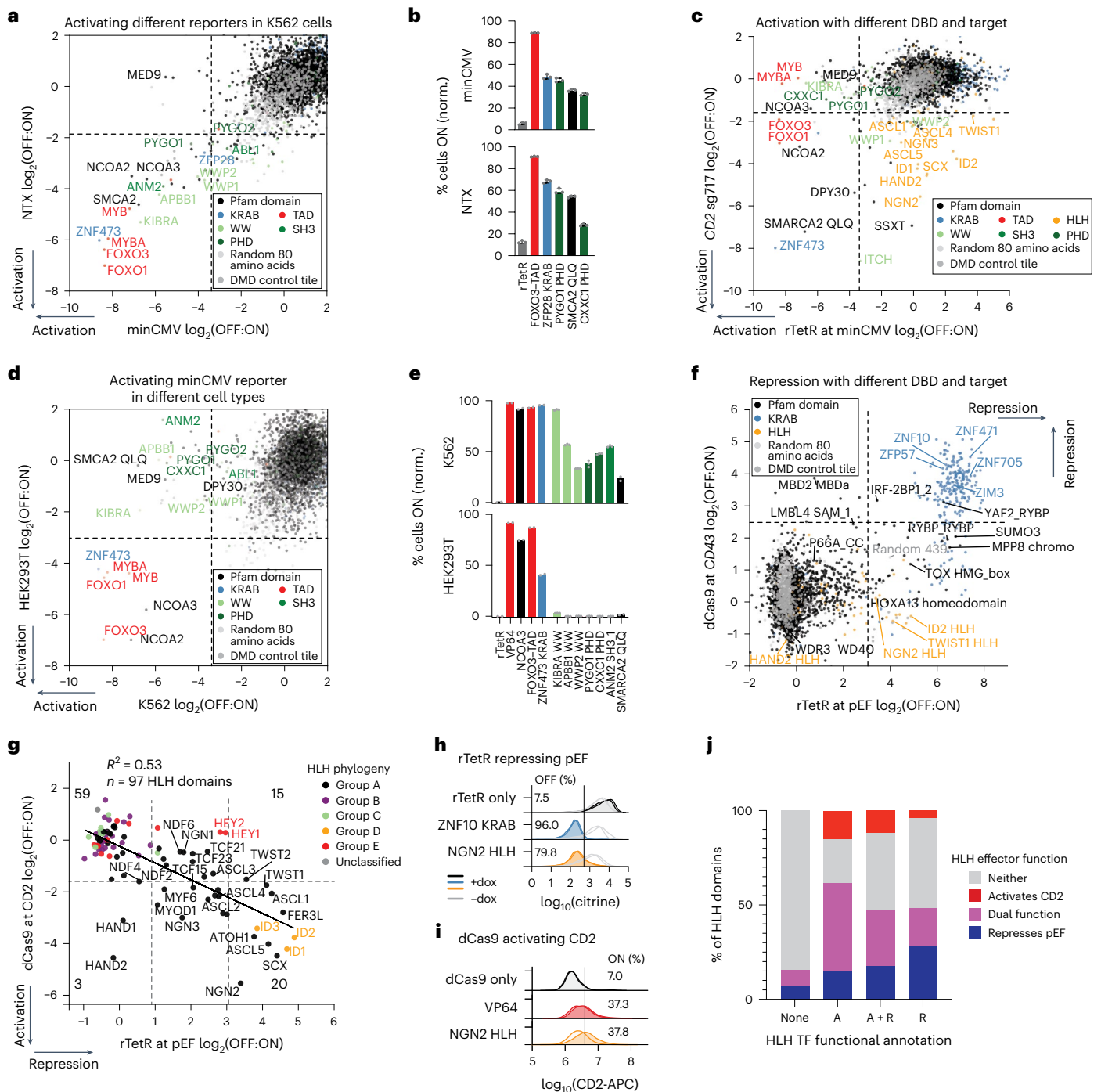


Fig. 2 | Context-dependent transcriptional effectors. **a**, HT-recruit with rTetR targeting two minimal promoters, minCMV and NTX, in K562 cells ($n = 2$ biological replicates). Dashed lines show hit thresholds at 2 s.d. above the median of the poorly expressed domains (same in **c**, **d** and **f**). **b**, Validation of activator domains across minimal promoter reporters in K562 cells. Individual rTetR-activator fusions or the rTetR-only negative control were delivered by lentivirus, and, after selection, cells were treated with 1,000 ng ml⁻¹ doxycycline for 2 days to induce reporter activation. The percentage of cells activated was measured by flow cytometry for the citrine reporter, after gating for delivery with mCherry. Bars show the average percentage of cells ON normalized to no-doxycycline control; error bars are s.d. ($n = 3$ infection replicates shown as dots). **c**, HT-recruit with dCas9 targeting endogenous gene *CD2* with sg717 compared with rTetR targeting the minCMV reporter in K562 cells ($n = 2$ biological replicates). **d**, HT-recruit with rTetR targeting the minCMV reporter in K562 and HEK293T cells ($n = 2$ biological replicates per cell type). **e**, Individual rTetR fusions were measured in minCMV reporter cells as in **b** ($n = 2-3$ infection

replicates shown as dots). **f**, HT-recruit with dCas9 targeting endogenous gene *CD43* with sg15 compared with rTetR targeting the pEF1 α reporter in K562 cells ($n = 2$ biological replicates). **g**, HT-recruit with dCas9 to activate *CD2* using sg717 compared with repression of pEF1 α promoter with rTetR in K562 cells, showing only the HLH domains within the Pfam library ($n = 2$ replicates per screen). Black line shows the best linear fit. The conservative hit threshold (black dashed line) was chosen to identify strong effectors. The gray dashed vertical line is equivalent to the strength of the weakest repressor that was individually validated (Methods). The HLH phylogenetic groups are shown in colored solid circles⁴⁴⁻⁴⁶. **h**, rTetR recruitment to the pEF1 α reporter in K562 cells with 6 days of treatment with 1,000 ng ml⁻¹ of doxycycline ($n = 2$ infection replicates shown as histograms). **i**, dCas9 recruitment to *CD2* in K562 cells ($n = 2$ sgRNAs, sg717 in darker and sg718 in lighter shade). **j**, Full-length HLH TFs were defined as not yet defined (none, $n = 45$), activators (A, $n = 13$), both (A + R, $n = 17$) or repressors (R, $n = 25$), in previous studies reviewed in ref. 46. Colors show the effector function of HLH domains from these TFs.

they overlapped similar annotations for SUMOylation, short linear interactions motifs and DBDs³⁴ (Extended Data Fig. 6d).

We then investigated the cell type-specificity of repressors using rTetR to target the Pfam library at pEF1 α reporters. The HEK293T screen identified strong repressors that were in agreement across cell types, such that 96% of repressor hits in HEK293T cells were also hits in K562 cells (including >200 KRAB domains; Supplementary Fig. 3a–d). When using the moderately strong UbC promoter, we found fewer hits than with pEF1 α , and again the majority were shared across cell types (for example, >130 KRAB domains; Supplementary Fig. 3e). Altogether, these results show that some repressors (for example, MDBa) strongly depend on recruitment context, while KRAB repressors are particularly robust across cell type, target and DBD contexts.

HLH effectors can activate or repress depending on the context

Effector domains are classified as activators or repressors; however, like the TFs that consist of these domains⁹, these effectors may have a dual activator–repressor function³⁴ that could be context-dependent. When comparing most activatable and repressible contexts, we saw no overlap in hits from the Pfam library; however, when considering the most sensitive repressible context (rTetR at pEF1 α in K562 cells), we found some activators could repress (Supplementary Fig. 4a–d). Certain minCMV strong activators either silence a percentage of the pEF1 α cells (that is, bifurcate the population) or super-activate it in a cell type-dependent manner (Supplementary Fig. 4e,f).

In addition, the large HLH family, which includes TFs shown to function either as an activator or repressor depending on context in yeast⁴³, stood out as a major source of dual-functioning effectors. Specifically, many of the same HLH effectors that activate CD2 when recruited with dCas9 also repressed pEF1 α when recruited with rTetR with correlated strength ($R^2 = 0.53$, $n = 97$ HLH domains; Fig. 2g–i and Extended Data Fig. 7a). Furthermore, we found that effector function is enriched within certain HLH phylogenetic groups^{44–46}; for example, HAND HLHs from group A are only activators, HEY HLHs from group E are only repressive, ID HLHs from group D are dual-functioning and other subfamilies (for example, HES from group E and all of group B) did neither in the contexts tested here (Fig. 2g and Extended Data Fig. 7b). These HLH effector functions partially corresponded with annotations from previous studies of the HLH TFs as activators, repressors or both (Fig. 2j)⁴⁶. These results were confirmed with the tiling library, where 74% of proteins with a dual-function tile that activates CD2 (but not minCMV) and represses pEF1 α were again HLH proteins. The hit tiles overlapped the HLH heterodimerization portion of their bHLH domains, and deletion experiments showed that this C-terminal portion was required for both activation and repression with a dual-functioning HLH (Extended Data Fig. 7c–g)³⁴.

Overall, while some effectors were highly context-dependent (for example, HLH and WW domains), we noticed many of the strongest effectors were consistent across contexts (for example, KRAB repressors), which suggests they would improve synthetic tools for manipulating transcription.

Efficient CRISPRi with ZNF705/ZNF471 KRAB repressors

We set out to identify repressors that are robust across contexts using the Pfam library because it was screened against the largest number of contexts. We performed an additional screen with this library at the *GATA1* locus to compare repressors at an enhancer (Supplementary Fig. 5a). *GATA1* is an essential gene, so we used the growth phenotype associated with targeting dCas9-repressors to its enhancer (eGATA1) as a proxy for repression strength^{14,47}. The KRAB domain found in ZNF705B/ZNF705D/ZNF705F (here called ZNF705) and ZNF471 were the strongest hit Pfam domains with the eGATA1 sgRNA, and they did not have growth effects with a control safe-targeting sgRNA (Supplementary Fig. 5b and Supplementary Table 1). Meanwhile, some domains did show growth effects with the safe-targeting sgRNA, suggesting

their expression is toxic. The most toxic were the cyclin-dependent kinase inhibitor domains from CDKN1A/CDKNB, which inhibit cell cycle progression by binding to CDK2 (ref. 48). The random sequences were substantially more toxic than the Pfam domains or DMD tile controls (Supplementary Fig. 5c,d), which could be due to random sequence misfolding⁴⁹. We confirmed these toxic domains were not confounding repressor hits when targeting the nonessential gene, *CD43* (Supplementary Fig. 5e).

To identify strong repressors that function well across many different contexts, we focused on the KRAB domains, which were commonly top hits. Furthermore, despite some effort, we could not identify more efficient non-KRAB repressors (Supplementary Note 1 and Supplementary Fig. 6a–l), consistent with KRAB repressors being most potent across contexts in the screens. Previously, using a deep mutational scan of the ZNF10 KRAB, we identified a mutant in the N-terminal KRAB domain region, WSR7EEE, that provides increased expression in cells and silencing strength when fused to rTetR³². We found the same is true when this mutant is fused to dCas9 (Fig. 3a,b and Extended Data Fig. 8a,b). This enhanced mutant KRAB specifically silences endogenous genes via heterochromatinization (as marked by H3K9me3), similar to reported data for the CRISPRi tools that use the WT KRAB from ZNF10 (refs. 5,50; Fig. 3c). We previously found that nearly all KRAB paralogs that repress in HT-recruit also interact with KAP1 as measured by mass spectrometry and colocalize with H3K9me3 by chromatin immunoprecipitation followed by sequencing (ChIP-seq), so we expect they also recruit H3K9me3 (ref. 32). Therefore, we used this enhanced ZNF10 KRAB (WSR7EEE) as a benchmark when determining how well the KRAB repressor paralogs work.

We ranked 323 KRAB domains by their scores across all of the repressor screens and found ZNF705 KRAB ranked 18 and ZNF471 KRAB ranked 1 (Fig. 3d). In agreement with a recent finding from a screen of 57 KRABs, our approach also identified ZIM3 KRAB (13) as a stronger repressor than ZNF10 KRAB (166)^{51,52}. Further supporting their context-robustness and reproducibility, we found ZNF705, ZNF471 and ZIM3 KRAB were hits in both replicates of 8, 7 and 6 of the original 8 repressive contexts, respectively (Fig. 1e and Supplementary Table 1). Directly comparing HT-recruit scores across contexts suggested that these top KRAB domains were stronger than ZNF10 KRAB but not meaningfully different from one another ($n = 12$ contexts; Extended Data Fig. 8c).

We set out to validate the top-ranked KRAB domains across different target contexts compared to the strong ZNF10 KRAB (WSR7EEE). Encouragingly, the KRAB domains from ZNF705 and ZNF471 both outperform the ZNF10 KRAB (WSR7EEE) at silencing CD43 (Fig. 3e). To test additional targets, we performed CRISPRi screens with a library including 405 sgRNAs targeting promoters of 37 essential genes (Extended Data Fig. 8d,e) and found that the ZNF705 KRAB consistently provides ~1.5 \times greater effect sizes across a range of ZNF10 baseline effects and sgRNA distances from the TSS (Fig. 3f and Extended Data Fig. 8d–g). We also found that ZNF705 KRAB consistently exhibits stronger effects at candidate *cis*-regulatory elements (Fig. 3g). These KRABs also completely silence a reporter in HEK293T cells (Extended Data Fig. 8h,i). Furthermore, ZNF705 KRAB more completely silenced an endogenous gene than ZNF10 KRAB across a population of THP-1 monocytes when stably delivered by lentivirus ($n = 2$ sgRNAs; Fig. 3h and Extended Data Fig. 8j). ZNF705 matched ZNF10 KRAB when transfected into iPSCs, which could be a saturated condition (Extended Data Fig. 8k,l). We note that the enhancing ZNF10 (WSR7EEE) mutation is in a unique N-terminal region that is spliced onto the ZNF10 KRAB A-box from a distinct exon not found in other KRAB genes. So, we could not transfer that same mutation to further enhance ZNF705/ZNF471/ZIM3.

To assess the KRAB repressors in a different DBD context, we used an engineered form of dCas12a (dEnAsCas12a; 1,308 amino acids)^{53,54}. dCas12a is weaker than dCas9 for recruiting repressors, so there is more need to improve its efficiency¹⁸. Fusing ZNF705 KRAB to dCas12a

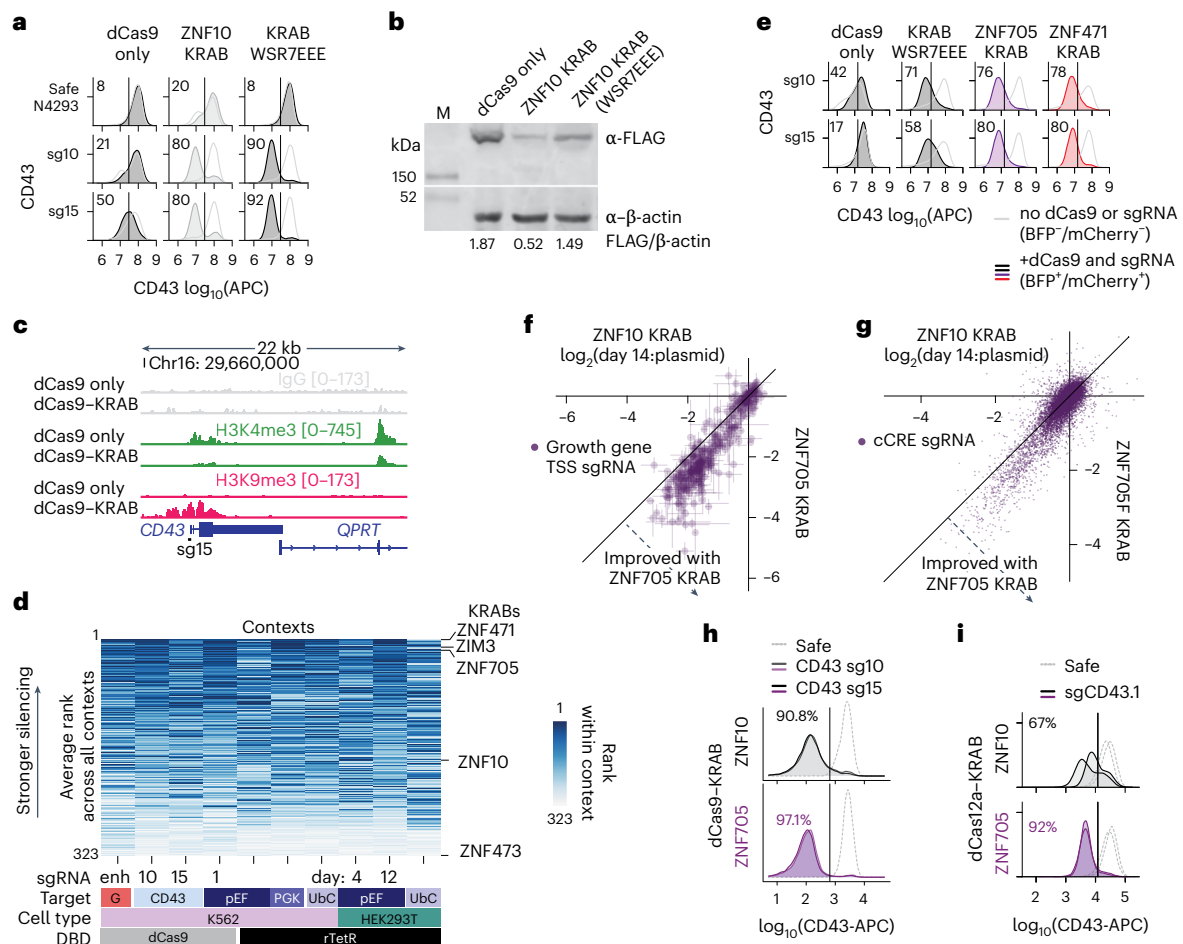


Fig. 3 | Improved CRISPRi using ZNF705 and ZNF471 KRABs. **a**, dCas9 recruitment in K562 cells. sgRNAs were delivered by lentivirus, followed by dCas9 constructs, and the cells were stained 7 days after dCas9 infection. The shaded histogram shows the cells after gating for both the dCas9 (BFP) and the sgRNA (mCherry), and their percentage of cells OFF is shown. The unshaded histogram shows the cells from the same sample that express neither and serve as an internal control ($n = 1$ infection replicate). **b**, Western blot of dCas9 only, ZNF10 KRAB and ZNF10 KRAB(WSR7EEE) fusions ($n = 1$ biological replicate). dCas9 and KRAB fusions were tagged with 3xFLAG. The band intensity ratio is below. **c**, Chromatin modifications mapped by CUT&RUN after dCas9 recruitment of ZNF10 KRAB(WSR7EEE) using sg15. Stable cell lines expressing the dCas9 fusion and the sgRNA were selected with antibiotics and fluorescent cell sorting before chromatin was analyzed. **d**, KRAB domains ranked by repression in HT-recruit screens across the target, cell type and DBD contexts (a higher average $\log_2(\text{OFF:ON})$ from two biological replicates is better ranked). The 323/336 KRAB domains that had no missing data across contexts are included. The red target labeled G shows the growth-based screen at the GATA1 enhancer. The HEK293T pEF screen has two time points as follows: a silencing measurement at 4 days after doxycycline addition and a memory measurement at day 12 (8 days after doxycycline removal). **e**, sgRNAs were delivered by lentivirus, followed by dCas9

constructs, and the cells were stained 9 days after dCas9 infection ($n = 1$). **f**, CRISPRi growth screen to compare KRAB repressors. A sgRNA library targeting 37 essential gene promoters was delivered into K562 cell lines that stably express either dCas9–KRAB. Cells were passaged for 14 days, and then the guides were sequenced to measure fitness effects (change from the original plasmid pool to the final day 14 genomic DNA). Greater depletion is a measure of stronger silencing of the essential genes. Each dot shows the average effect for a sgRNA, and the error bars show the s.d. from two screen replicates ($n = 405$ sgRNAs). The diagonal line represents the identity between KRAB domains. **g**, Results are shown similarly for the sgRNAs in the library that target cCREs of growth genes ($n = 10,155$ sgRNAs). **h**, sgRNAs were delivered to THP-1 monocytes by lentivirus, followed by dCas9 fusions. Eight days later, cells were stained. Gating and shading as in **a** ($n = 2$ targeting sgRNAs shown as histograms). **i**, dCas12a recruitment in K562 cells. dCas12a–KRAB fusions and then guide RNAs were delivered by lentivirus. Nine days after guide RNA infection, cells were stained. Data are gated for guide RNA expression (mCherry) and dCas12a expression (hemagglutinin (HA)-tag stain). Infection replicates are shown as separate histograms ($n = 2$ biological replicates shown as histograms). cCREs, candidate *cis*-regulatory elements.

resulted in a mean 1.5× more complete silencing of a population of cells than ZNF10 KRAB across multiple targets (Fig. 3i and Extended Data Fig. 8m). Encouragingly, a 99% identical ZNF705 family KRAB (from ZNF705E) was ranked 12 of >1,000 repressor hit tiles recruited in three contexts with the larger tiling library, further establishing ZNF705 KRAB as a particularly strong repressor (Extended Data Fig. 8n,o). Other highly ranked KRABs likely perform similarly well. Concordant with its low toxicity and previous RNA-seq results with ZNF10 KRAB^{55,56}, stable overexpression of dCas9–ZNF705 KRAB did not perturb the baseline transcriptome relative to a dCas9-only control (0.01% transcripts changed >4× by RNA-seq; Extended Data Fig. 8p).

Compact human activator combinations for lentiviral CRISPRa
We next ranked activators from HT-recruit measurements performed across the target, DBD and cell-type contexts and selected context-robust domains that were a hit in at least five activation screens (Fig. 4a). ZNF473 KRAB (Z) is an acidic domain that was a strong hit with both dCas9 recruitment to CD2 and most rTetR contexts. Meanwhile, NCOA3 (N) and FOXO3–TAD (F), which bind p300 (refs. 57–59), were rTetR hits and stronger than ZNF473 in HEK293T cells. Previous ChIP experiments showed that the full-length proteins from which we sourced these effectors are associated with active chromatin marks—NCOA3 is enriched near H3K27ac and H3K4me1 chromatin

modifications⁶⁰, FOXO3 colocalizes with RNA polymerase II and H3K27ac^{61,62} and ZNF473 is enriched at H3K27ac peaks³². We found that these domains were well-expressed across DBDs and cell types and there was no toxicity associated with their expression (Extended Data Fig. 3d and Supplementary Fig. 5e). We had previously found that the trimmed Pfam-annotated domains (Z = 44 amino acids, N = 48 amino acids and F = 41 amino acids) for these three activators performed equally to the 80 amino acid sequence centered on the domain with rTetR³², so we used the trimmed sequences to build activator tools (Fig. 4b). This trimmed FOXO3 sequence was within the 11th activator tile in the larger library ranked by the average of the minCMV and CD2 screens (Supplementary Table 2).

We fused these trimmed activators, or two commonly used synthetic activators (VP64 and VPR)^{63,64}, individually to dCas9 in a lentiviral vector with BFP as a delivery marker. Given previous observations of synergistic activation from concatenated activators (including VP64 compared to VP16; VPR compared to its components VP64, p65 and Rta; and the synergistic activation mediator (SAM), which combines VP64 with a >450 amino acids MCP-p65-HSF1 component recruited via MS2 stem-loops in the sgRNA)^{4,64–68}, we reasoned that fusing N, F and Z could improve activation. Indeed, when we combined any of these domains together as bipartite activators, we observed synergistic activation of the *CD2* gene (Fig. 4c). Further fusion of all three domains together to create the tripartite activator, NFZ, led to even stronger activation (Fig. 4c).

At 145 amino acids long, NFZ is more compact than the tripartite 523 amino acid VPR and does not use viral components. We found that NFZ was more efficiently delivered by lentivirus, more efficiently generated stable cell lines after blasticidin selection and provided a similar or higher level of activation across three target genes (Fig. 4c–f and Extended Data Fig. 9a–d). Furthermore, we could use lentivirus to deliver NFZ, but not VPR, to J774 mouse macrophages, which are more difficult to infect than K562 cells, and successfully select a stable line in which we could activate endogenous genes (Fig. 4g and Extended Data Fig. 9e,f). NFZ was also a potent activator in THP-1 monocytes and iPSCs (Fig. 4h,i). Additionally, by changing the N, F and Z domain orientations, we identified NZF as a moderately stronger activator with lentiviral delivery efficiencies between VPR and NFZ (Fig. 4f and Extended Data Fig. 9a,b,g).

NFZ may be more deliverable by lentivirus than VPR due to differences in transduction efficiency and/or cell toxicity related to the

expression of VPR. While VPR is the largest effector, the lentiviral insert size (between central polypurine tract (cPPT) and woodchuck hepatitis virus post-transcriptional regulatory element (WPRE)) is 8.5 kb, which does not exceed packaging limits⁶⁹. Meanwhile, transcriptional squelching is associated with VP16-derived activators^{70,71}, and VPR can be toxic in cell lines⁷², which could affect the packaging cells and/or transduced cells. When using transient plasmid delivery, we observed similar or better activation with VPR than NFZ (Extended Data Fig. 9h,i), despite observing lower expression of VPR transcripts (Extended Data Fig. 9j,k; consistent with others' observations of low dCas9-VPR expression^{73,74}), suggesting NFZ is specifically advantageous for viral delivery.

Notably, the tripartite fusions were strong activators even with dCas9 targeting *CD20*, *CD28*, *CD2* (with its sgRNAs beyond sg717) or the minCMV reporter, which were all contexts with poor activation from nearly any single domain in the Pfam library (Fig. 4c,f, Extended Data Fig. 9g–i,l–n and Supplementary Table 1). We found no further improvement by fusing a fourth domain, QLQ, or using homotypic combinations (Extended Data Fig. 9l–o and Supplementary Note 2).

One of the most potent currently used CRISPRa tools is the larger SAM system⁶⁶. SAM consists of the following three components: dCas9–VP64, a sgRNA engineered to contain the MS2 stem-loop and a 472 amino acid helper protein MPH (MS2-binding protein fused to P65 and HSF1 human activators; Supplementary Fig. 7a). We found that we could replace the viral VP64 in SAM with human NFZ/NZF and activate endogenous *CD2* with comparable efficiency (Supplementary Fig. 7b–g and Supplementary Note 3). We also found the two-component dCas9–NFZ easier to deliver than the three-component SAM (which required an additional helper protein virus, thus resulting in fewer cells co-infected with all components), but that SAM can be more potent if successfully delivered.

Fusing compact activator NFZ to smaller CRISPR DBDs

dCas9 is a relatively large DBD (1,368 amino acids), and it is difficult to multiplex with multiple sgRNAs. To make a CRISPRa system that can be multiplexed, we ported NFZ to dEnAsCas12a (1,308 amino acids)^{54,75} and found that NFZ activated *CD2* approximately tenfold more strongly than VP64 (Fig. 4e,j and Supplementary Fig. 8a,b). As with dCas9, we could not establish a stable dCas12a–VPR cell line using lentivirus (7.6 kb) and blasticidin selection, but we were successful with dCas12a–NFZ (Supplementary Fig. 8c,d).

Fig. 4 | Compact NFZ activator improves CRISPRa systems in different cell types.

a, HT-recruit scores for activators that were hit in ≥ 5 samples across the target, cell type and DBD contexts ($n = 2$ replicates per rTetR screen and $n = 1–2$ replicates per sgRNA for dCas9 screens shown as columns). The rows are clustered in an unbiased manner. Labels for NCOA3 (1,045–1,092; N), FOXO3 FOXO–TAD (604–644; F) and ZNF473 KRAB (5–48; Z) are colored. **b**, Size of N, Z and F human activator domains and combinations compared to viral activators. **c**, The effectors were recruited with dCas9 in K562 cells. sgRNAs were delivered by lentivirus, and cells were selected with puromycin. Then dCas9 fusions were delivered by $10\times$ concentrated lentivirus. Four days after dCas9 delivery, cells were stained and measured by flow cytometry ($n = 2$ infection replicates shown as histograms). Left, *CD2* expression after gating cells for sgRNA delivery (GFP). Middle, the average percentage of cells with dCas9 delivery (BFP) is written in blue. Right, *CD2* expression after gating for both sgRNA and dCas9 delivery. The darker histogram is *CD2* sg717, and the lighter shade is sg718. **d**, Comparison of activation and dCas9-activator delivery (BFP) after gating for sg717 delivery (GFP). Color shows the smoothed density of flow cytometry data. The percentage of BFP⁺ cells is labeled ($n = 2$ infection replicates). **e**, Schematics of CRISPR activator systems using the NFZ tripartite activator fused to dCas9, dEnAsCas12a or dCasMINI_V4. **f**, dCas9 activators were delivered by $5\times$ concentrated lentivirus into stable sgRNA-expressing cell lines; 5 days later, blasticidin selection was started, and then 4 days later, cells were stained. Top, the percentage of cells ON (APC) is shown after gating for the sgRNA (GFP⁺) and dCas9 (BFP⁺). All of the VPR samples had <250 BFP⁺ cells counted

(associated with poor infection and survival of blasticidin selection), while all other samples had >6k BFP⁺ cells (mean = 53k). Bottom, the percentage of BFP⁺ cells is shown ($n = 5–7$ sgRNAs per effector). The *CD28* sgRNAs infected with VPR poorly survived blasticidin selection, and <300 cells total could be counted, so they are not shown. **g**, dCas9–NFZ was stably delivered to J774 mouse macrophages by lentivirus and selected with blasticidin. Activation of endogenous *Gpr84* or *Actc1* measured by qPCR after lentiviral sgRNA delivery ($n = 3$ qPCR replicates). NT sgRNAs are negative controls. **h**, THP-1 monocytes stably expressing the sgRNAs were generated using lentiviral delivery, followed by infection with dCas9-activator lentiviruses. Cells were stained for *CD2* expression 8 days later. Gated for sgRNA delivery (GFP) and dCas9 (BFP; $n = 1$ infection per sgRNA). **i**, iPSCs were transfected with dCas9 activator and sgRNA plasmids and then stained 2 days later. Gated for sgRNA delivery (GFP) and dCas9 (BFP; $n = 3$ transfection replicates shown as histograms). The percentage of cells ON with the gene-targeting sgRNA, using the threshold shown by the black vertical line, is shown. **j**, dEnAsCas12a activators on a marker-less plasmid were co-electroporated into K562 cells with a gRNA-mCherry expression plasmid. After 3 days, cells were stained. Gated for high mCherry ($n = 2$ replicates). **k**, dCasMINI activators on a plasmid with an mCherry marker were cotransfected into HEK293T GFP reporter cells with a gRNA-BFP expression plasmid. Two days later, GFP reporter activation was measured by flow cytometry with gates for BFP⁺/mCherry⁺ cells ($n = 6$ transfection replicates shown as dots, and bar shows mean). Np NLS, nucleoplasmic nuclear localization signal; NT, nontargeting.

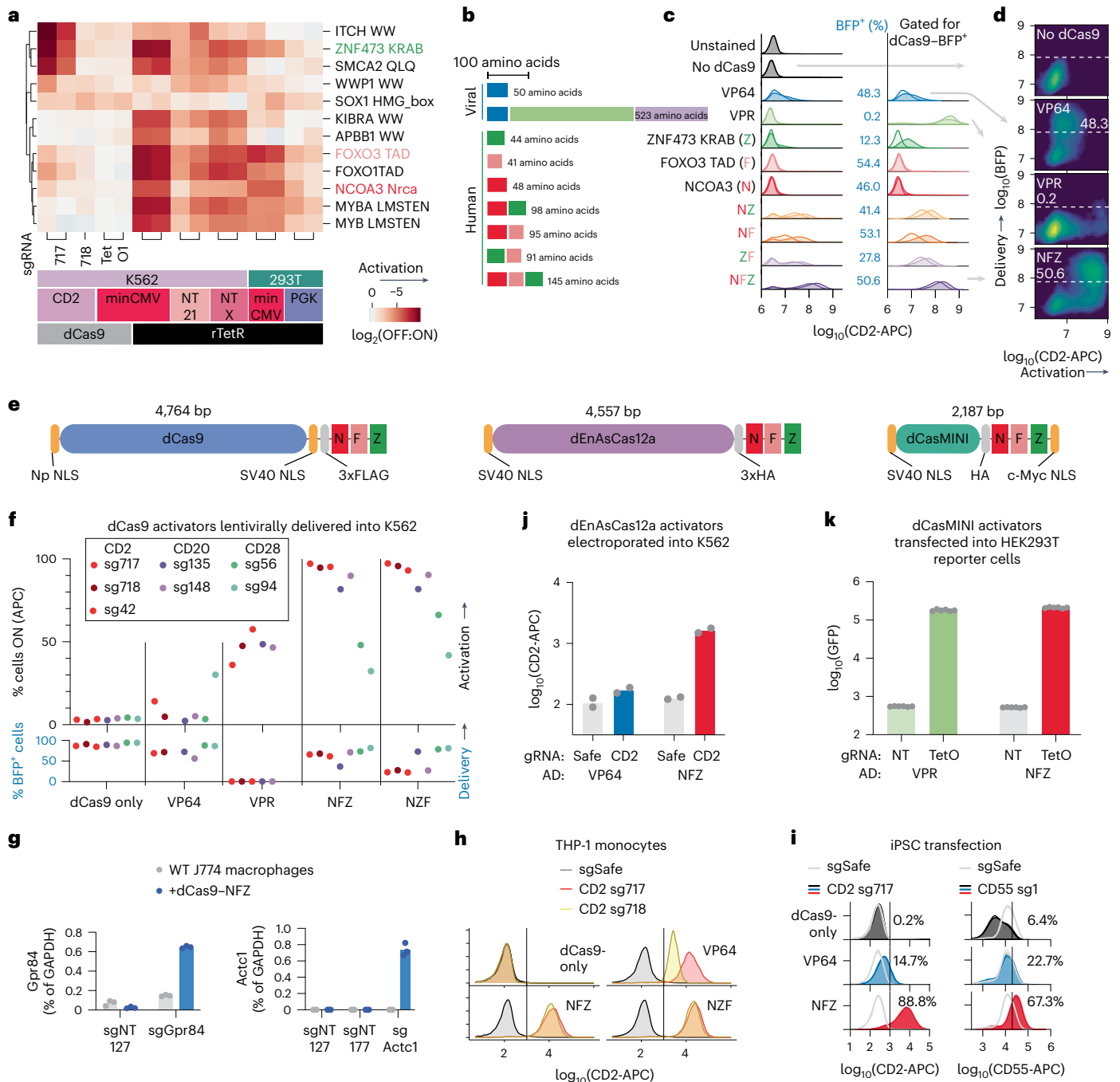
Recently, a very compact dCas12f DBD called dCasMINI (529 amino acids) was engineered as a CRISPRa system in human cells⁷⁶. We found that dCasMINI-NFZ (2.2 kb) strongly activated HEK293T reporter cells (Fig. 4e,k and Supplementary Fig. 9a–e).

Inducible expression with the compact activator NFZ

Another useful application of compact human activators is engineering circuits that tunably control transgene expression using a synthetic TF that conditionally activates a promoter in the presence of a small molecule⁷⁷. Previously, the zinc finger-homeodomain DBD (ZFHD1)⁷⁸ was fused to FK506 binding protein (FKBP)-dimerization domains that recruit HSF1 and p65 activators in the presence of rapamycin⁷⁹. We designed a >1 kb smaller version of this design using NFZ (Fig. 5a and Supplementary Fig. 10a). Upon transfection, we observed rapamycin-dependent reporter transgene inducibility and low levels

of leakiness with NFZ relative to P65 + HSF1 (Fig. 5b and Supplementary Fig. 10b–d). Next, we used ZFHD1–NFZ to inducibly express the human hepatocyte growth factor (HGF) gene (2.2 kb)^{80–84} and observed a rapamycin dose-dependent increase in secreted HGF protein, with minimal leakiness in untreated cells (Fig. 5c). This compact circuit could enable longer proteins (such as HGF) to be inducibly expressed from limited-size viral vectors like adeno-associated virus (AAV).

Inducible gene expression circuits can also expand the therapeutic potential of chimeric antigen receptor (CAR) T cells by providing control over CAR expression. Such a circuit, built using synthetic zinc finger transcription regulators (synZiFTR), was recently shown to be effective for dose- and time-dependent control of CARs in therapeutically relevant primary T cells both in vitro and in mouse models of cancer⁸⁵. To minimize immunogenicity, synZiFTRs use a human activator domain. Here by replacing the previously used p65 activator with NFZ,



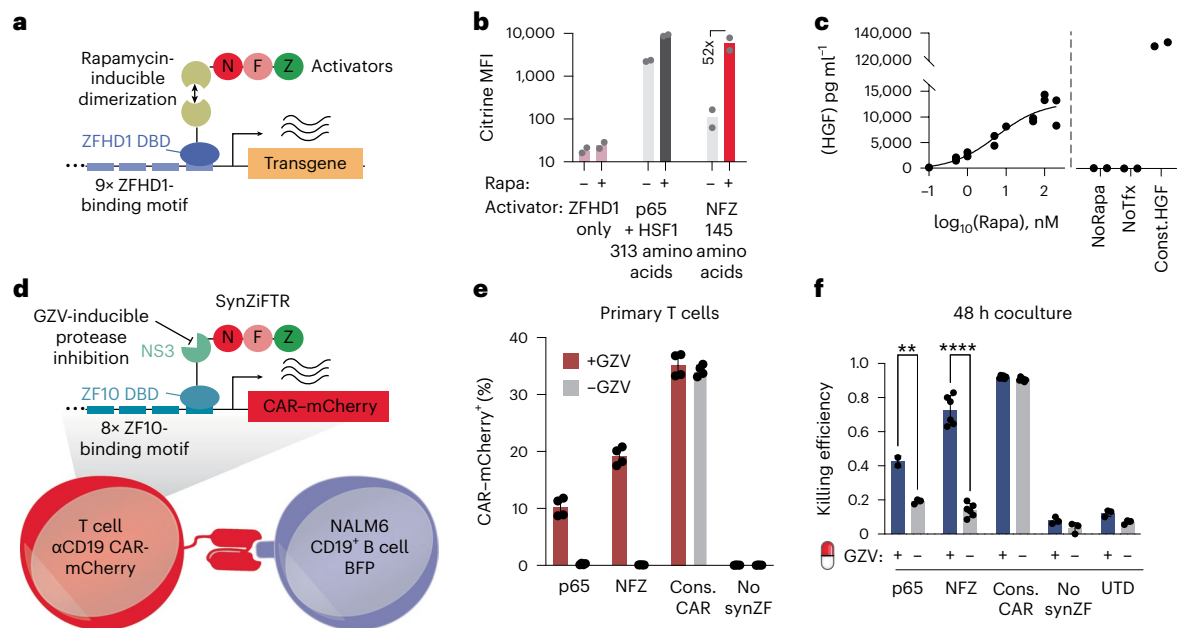


Fig. 5 | Compact NFZ activator improves therapeutically relevant inducible systems.

a, Schematic representation of the minimized rapamycin-inducible transcription cassette using the ZFHD1 DBD fused to NFZ. **b**, Rapamycin-inducible expression of citrine with ZFHD1 recruitment of NFZ or a control activator combination, p65 + HSF1. In total, 3 μg of plasmids were transfected into HEK293T cells; 1 day later, 10 nM rapamycin was added, and then 2 d later, citrine MFI was measured by flow cytometry ($n = 2$ transfection replicates shown as dots, and bar shows mean). **c**, Rapamycin-inducible expression of HGF with ZFHD1 recruitment of NFZ, after transfection in HEK293T cells. Secreted HGF protein concentrations were measured in the cell culture supernatant by ELISA after 2 days of rapamycin treatment with varied doses. Transfection of no plasmid and a constitutive pEF1α-HGF plasmid served as a negative and positive control, respectively ($n = 2$ transfection replicates; the curve shows nonlinear regression fit to dose-response data). **d**, Schematic representation of synZiFTR circuit for drug-regulated control of CAR T cells. SynZiFTR uses the FDA-approved drug Grazoprevir (GZV) to inducibly inhibit NS3 protease-mediated self-cleavage of

an engineered zinc finger activator. Upon induction, the activator is recruited to synthetic ZF10-binding motifs⁸⁵ to activate an anti-CD19 CAR fused to a mCherry reporter. Bottom illustration shows a synZiFTR CAR T cell targeting a CD19⁺ cancer cell. **e**, Efficiency of activating CAR-mCherry as measured using the vertical threshold shown in Extended Data Fig. 10c ($n = 1-2$ infection replicates measured with two technical replicates each shown as dots, and bar shows mean and error bars show s.d.). ‘Cons. CAR’ control is the constitutive spleen focus-forming virus (SFFV) promoter, and ‘no synZF’ control only contains the CAR-mCherry transgene. **f**, After 72 h of induction with GZV, SynZiFTR T cells were cocultured in the continuous presence of GZV for 48 h with a CD19⁺ NALM6 precursor leukemia B cell line engineered to express BFP. Killing efficiency was estimated by measuring the reduction in BFP⁺ cells compared to a NALM6-only negative control by flow cytometry ($n = 1-2$ infection replicates measured with three technical replicates each shown as dots, bars show mean and error bars show s.d.). ** $P < 0.005$ and **** $P < 0.00005$, by two-sided unpaired t test. UTD, untransduced; MFI, mean fluorescence intensity.

we found that we could increase induced CAR expression in primary T cells and correspondingly increase pro-inflammatory antitumor cytokine secretion and CAR-dependent cancer cell-killing efficiency (Fig. 5d–f and Extended Data Fig. 10a–h).

Discussion

A better understanding of transcriptional effector context-specificity is needed to improve strategies for manipulating transcription. Our systematic effector measurements across cell type, genomic target and DBD contexts revealed both context-dependent effectors (for example, WW and HLH domains) and context-independent effectors that can improve transcriptional perturbation tools (for example, the tripartite NFZ activator and the ZNF705 KRAB repressor).

For activation, NFZ may overcome delivery challenges associated with the size, number of components or reported toxicity of expressing VPR and SAM^{72,86–88} (Fig. 4d). NFZ is a compact, nonviral and low-toxicity activator that was better delivered by lentivirus than VPR (likely due to lower impacts on packaging cell or recipient cell health) and than SAM (because it does not require an additional helper protein component). In some applications, the domain orientation NZF provides stronger activation, but NFZ showed better lentiviral delivery and less leaky activation in an inducible circuit. Furthermore, NFZ shows promise as a compact and stably deliverable activator that could be easily multiplexable with dCas12a if further optimized, potentially in combination with other recent improvements for that DBD^{75,89–91}. Overall, our data suggest that NFZ could be preferable over VPR or SAM in contexts

where the larger systems cannot be equally well-delivered and where human-derived activators are needed, such as synZiFTR CAR T cells. However, we caution that the domain junctions in NFZ could produce immunoreactive peptides and that NFZ may not improve delivery or avoid toxicity in all contexts. CRISPRi/CRISPRa expression alone, with certain effectors, can perturb baseline gene expression of some genes^{56,68}, in a manner that correlates with the effectors’ toxicity⁵⁵, and NFZ could still have this issue despite being less toxic than VPR.

For repression, we find that KRABs are the strongest family across contexts and that ZNF705, ZNF471, ZIM3 and several other top-ranked KRAB domains are all similarly strong, demonstrating that repressors substantially stronger than ZNF10 KRAB are abundant upon screening all 336 KRABs. This wealth of options of strong KRAB domains could be useful to the field of synthetic biology, which now does not need to rely on only ZIM3 or ZNF10. Further studies could determine whether there are meaningful differences between these top KRABs or whether non-KRAB repressors from this resource outperform these top KRABs singly or in combinations in certain contexts (for example, target sites, cell types or recruiter fusion protein configurations).

Furthermore, our larger tiling library both confirmed context effects we originally observed with fewer elements in the Pfam library (for example, in the HLH, KRAB and MBD families) and provides a valuable resource of transcriptional effectors, including from unannotated protein regions, that function on dCas9 at endogenous genes. These transcriptional repressors and activators, in addition to others^{32,34,40,51,74,92–96}, can now be combined with varied DBDs⁸⁵ to

enable gene regulation tools that are more compact, humanized and deliverable.

Looking beyond the context-robust effectors, this and other recent work^{97,98} describe the extent of context dependence across effectors. Overall, for both activators and repressors, differences across cell types were more likely to be changes in magnitude, whereas differences across target and DBD were more likely to be complete changes from effective to ineffective (or even from activating to repressing). More specific trends can be found within domain families such as the target-dependent HLH domains or the cell-type-dependent WW and SH3 domains that interact with proline-rich motifs^{99,100} that are common in activators¹⁰¹. This resource of effector domains measured at endogenous sites provides useful starting points to explore mechanisms that explain which domains work at which loci. Integrating this effector function data with protein interaction maps could identify cofactors that help explain context dependence.

Scalable synthetic systems can control for variability across contexts and isolate a single biological variable to measure, but they have technical limitations. For example, across effectors that function as activators, repressors or both, we generally find stronger effects in K562 versus HEK293T cells, and it seems possible that a global effect, for example, on rTetR recruitment strength, contributes to this result. While these cell lines are useful for screening and we validated some effectors in other cells including J774 macrophages, THP-1 monocytes, primary T cells and WTC11 iPSCs, further assessment of effectors in a broader set of cell types, including primary and stem cells, will be important to uncover dependencies that are particularly relevant in disease and development. We also found that dCas9 HT-recruit screens varied in their dynamic range depending on the sgRNA and target. Measurements in additional endogenous target contexts^{102,103}, including sgRNAs with varied proximity to target TSS and enhancers, will be needed to dissect the complex relationships between target context and effector function. While 80 amino acid domain libraries are powerful due to their scalability, programmability, focus on short functional regions and the deliverability of the compact domains, using longer sequences could uncover additional forms of context dependence as longer sequences are more likely to encode multiple functions and layers of regulation. Isolating domains can reveal functions that are masked in the full-length context, but these functions can also be unrelated to the protein's normal function. Although termed transcriptional effectors, these assays do not directly measure mRNA synthesis, and it is still possible that some hits affect processes downstream of transcription, such as mRNA maturation. Toward understanding the mechanisms behind effector context dependence, it would next be useful to better measure and manipulate the endogenous factors that co-occupy the targeted promoters and enhancers.

Overall, our work highlights the utility of a high-throughput process for protein engineering in synthetic biology that is enabled by the ability to synthesize and test large pooled libraries of domains for function directly in human cells. In the future, HT-recruit could be adapted to systematically discover and characterize effectors for the other forms of gene and epigenetic editing, including base and prime editing¹⁰⁴.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-024-02442-6>.

References

1. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
2. Medvedeva, Y. A. et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database* **2015**, bav067 (2015).
3. Göös, H. et al. Human transcription factor protein interaction networks. *Nat. Commun.* **13**, 766 (2022).
4. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR–Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
5. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
6. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
7. Sanson, K. R. et al. Optimized libraries for CRISPR–Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
8. Gao, Y. et al. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat. Methods* **13**, 1043–1049 (2016).
9. Soto, L. F. et al. Compendium of human transcription factor effector domains. *Mol. Cell* **82**, 514–526 (2021).
10. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. 3rd Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA* **95**, 14628–14633 (1998).
11. Joung, J. et al. Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* **548**, 343–346 (2017).
12. Rivera, V. M. et al. A humanized system for pharmacologic control of gene expression. *Nat. Med.* **2**, 1028–1032 (1996).
13. Hilton, I. B. et al. Epigenome editing by a CRISPR–Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, 510–517 (2015).
14. Fulco, C. P. et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
15. Segal, D. J. et al. Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. *J. Biol. Chem.* **279**, 14509–14519 (2004).
16. Bailus, B. J. et al. Protein delivery of an artificial transcription factor restores widespread Ube3a expression in an Angelman syndrome mouse brain. *Mol. Ther.* **24**, 548–555 (2016).
17. Thakore, P. I. et al. RNA-guided transcriptional silencing in vivo with *S. aureus* CRISPR–Cas9 repressors. *Nat. Commun.* **9**, 1674 (2018).
18. O'Geen, H. et al. dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. *Nucleic Acids Res.* **45**, 9901–9916 (2017).
19. Amabile, A. et al. Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell* **167**, 219–232 (2016).
20. Hathaway, N. A. et al. Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460 (2012).
21. Ying, Y. et al. The Krüppel-associated box repressor domain induces reversible and irreversible regulation of endogenous mouse genes by mediating different chromatin states. *Nucleic Acids Res.* **43**, 1549–1561 (2015).
22. O'Geen, H., Tomkova, M., Combs, J. A., Tilley, E. K. & Segal, D. J. Determinants of heritable gene silencing for KRAB-dCas9+DNMT3 and Ezh2-dCas9+DNMT3 hit-and-run epigenome editing. *Nucleic Acids Res.* **50**, 3239–3253 (2022).
23. Cano-Rodríguez, D. et al. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat. Commun.* **7**, 12284 (2016).

24. Hong, C. K. Y. & Cohen, B. A. Genomic environments scale the activities of diverse core promoters. *Genome Res.* **32**, 85–96 (2022).
25. Sahu, B. et al. Sequence determinants of human gene regulatory elements. *Nat. Genet.* **54**, 283–294 (2022).
26. O'Geen, H. et al. Ezh2-dCas9 and KRAB-dCas9 enable engineering of epigenetic memory in a context-dependent manner. *Epigenetics Chromatin* **12**, 26 (2019).
27. Nuñez, J. K. et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503–2519 (2021).
28. Kearns, N. A. et al. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
29. Cano-Rodriguez, D. & Rots, M. G. Epigenetic editing: on the verge of reprogramming gene expression at will. *Curr. Genet. Med. Rep.* **4**, 170–179 (2016).
30. Sanborn, A. L. et al. Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *eLife* **10**, e68068 (2021).
31. Arnold, C. D. et al. A high-throughput method to identify trans-activation domains within transcription factor sequences. *EMBO J.* **37**, e98896 (2018).
32. Tycko, J. et al. High-throughput discovery and characterization of human transcriptional effectors. *Cell* **183**, 2020–2035 (2020).
33. Alerasool, N., Leng, H., Lin, Z.-Y., Gingras, A.-C. & Taipale, M. Identification and functional characterization of transcriptional activators in human cells. *Mol. Cell* **82**, 677–695 (2022).
34. DelRosso, N. et al. Large-scale mapping and mutagenesis of human transcriptional effector domains. *Nature* **616**, 365–372 (2023).
35. Choi, S. H. et al. DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
36. Alfert, A., Moreno, N. & Kerl, K. The BAF complex in development and disease. *Epigenetics Chromatin* **12**, 19 (2019).
37. Treich, I., Cairns, B. R., de los Santos, T., Brewster, E. & Carlson, M. SNF11, a new component of the yeast SNF-SWI complex that interacts with a conserved region of SNF2. *Mol. Cell. Biol.* **15**, 4240–4248 (1995).
38. Childs, K. S. & Goodbourn, S. Identification of novel co-repressor molecules for interferon regulatory factor-2. *Nucleic Acids Res.* **31**, 3016–3026 (2003).
39. Gnanapragasam, M. N. et al. p66 α -MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2–NuRD complex. *Proc. Natl Acad. Sci. USA* **108**, 7487–7492 (2011).
40. Yeo, N. C. et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018).
41. Robinson, A. K. et al. The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile α motif (SAM) domain. *J. Biol. Chem.* **287**, 8702–8713 (2012).
42. Chen, K., Ou, X.-M., Chen, G., Choi, S. H. & Shih, J. C. R1, a novel repressor of the human monoamine oxidase A. *J. Biol. Chem.* **280**, 11552–11559 (2005).
43. McIsaac, R. S., Petti, A. A., Bussemaker, H. J. & Botstein, D. Perturbation-based analysis and modeling of combinatorial regulation in the yeast sulfur assimilation pathway. *Mol. Biol. Cell* **23**, 2993–3007 (2012).
44. Ledent, V., Paquet, O. & Vervoort, M. Phylogenetic analysis of the human basic helix–loop–helix proteins. *Genome Biol.* **3**, RESEARCH0030 (2002).
45. Atchley, W. R. & Fitch, W. M. A natural classification of the basic helix–loop–helix class of transcription factors. *Proc. Natl Acad. Sci. USA* **94**, 5172–5176 (1997).
46. Torres-Machorro, A. L. Homodimeric and heterodimeric interactions among vertebrate basic helix–loop–helix transcription factors. *Int. J. Mol. Sci.* **22**, 12855 (2021).
47. Tycko, J. et al. Mitigation of off-target toxicity in CRISPR–Cas9 screens for essential non-coding elements. *Nat. Commun.* **10**, 4063 (2019).
48. Russo, A. A., Jeffrey, P. D., Patten, A. K., Massagué, J. & Pavletich, N. P. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A–Cdk2 complex. *Nature* **382**, 325–331 (1996).
49. Kubota, H. Quality control against misfolded proteins in the cytosol: a network for cell survival. *J. Biochem.* **146**, 609–616 (2009).
50. Thakore, P. I. et al. Highly specific epigenome editing by CRISPR–Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
51. Alerasool, N., Segal, D., Lee, H. & Taipale, M. An efficient KRAB domain for CRISPRi applications in human cells. *Nat. Methods* **17**, 1093–1096 (2020).
52. Replogle, J. M. et al. Mapping information-rich genotype–phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575 (2022).
53. Tak, Y. E. et al. Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. *Nat. Methods* **14**, 1163–1166 (2017).
54. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
55. Replogle, J. M. et al. Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *eLife* **11**, e81856 (2022).
56. Gemberling, M. P. et al. Transgenic mice for in vivo epigenome editing with CRISPR-based systems. *Nat. Methods* **18**, 965–974 (2021).
57. Karlsson, E. et al. Disordered regions flanking the binding interface modulate affinity between CBP and NCOA. *J. Mol. Biol.* **434**, 167643 (2022).
58. Chen, H. et al. Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* **90**, 569–580 (1997).
59. Wang, F. et al. Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. *Proc. Natl Acad. Sci. USA* **109**, 6078–6083 (2012).
60. Percharde, M. et al. Ncoa3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming. *Genes Dev.* **26**, 2286–2298 (2012).
61. Eijkelenboom, A. et al. Genome-wide analysis of FOXO3 mediated transcription regulation through RNA polymerase II profiling. *Mol. Syst. Biol.* **9**, 638 (2013).
62. Eijkelenboom, A., Mokry, M., Smits, L. M., Nieuwenhuis, E. E. & Burgering, B. M. T. FOXO3 selectively amplifies enhancer activity to establish target gene regulation. *Cell Rep.* **5**, 1664–1678 (2013).
63. Maeder, M. L. et al. CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
64. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
65. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
66. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).

67. Zhou, H. et al. In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR–dCas9-activator transgenic mice. *Nat. Neurosci.* **21**, 440–446 (2018).
68. Dominguez, A. A. et al. CRISPR-mediated synergistic epigenetic and transcriptional control. *CRISPR J.* **5**, 264–275 (2022).
69. Counsell, J. R. et al. Lentiviral vectors can be used for full-length dystrophin gene therapy. *Sci. Rep.* **7**, 44775 (2017).
70. Ptashne, M. & Gann, A. A. F. Activators and targets. *Nature* **346**, 329–331 (1990).
71. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563–564 (1988).
72. Jones, R. D. et al. An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat. Commun.* **11**, 5690 (2020).
73. Wang, K. et al. Systematic comparison of CRISPR-based transcriptional activators uncovers gene-regulatory features of enhancer–promoter interactions. *Nucleic Acids Res.* **50**, 7842–7855 (2022).
74. Mahata, B. et al. Compact engineered human mechanosensitive transactivation modules enable potent and versatile synthetic transcriptional control. *Nat. Methods* **20**, 1716–1728 (2023).
75. Guo, L. Y. et al. Multiplexed genome regulation in vivo with hyper-efficient Cas12a. *Nat. Cell Biol.* **24**, 590–600 (2022).
76. Xu, X. et al. Engineered miniature CRISPR–Cas system for mammalian genome regulation and editing. *Mol. Cell* **81**, 4333–4345 (2021).
77. Bhatt, B., García-Díaz, P. & Foight, G. W. Synthetic transcription factor engineering for cell and gene therapy. *Trends Biotechnol.* **42**, 449–463 (2023).
78. Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. Structure-based design of transcription factors. *Science* **267**, 93–96 (1995).
79. Rivera, V. M. et al. Long-term pharmacologically regulated expression of erythropoietin in primates following AAV-mediated gene transfer. *Blood* **105**, 1424–1430 (2005).
80. Suzumura, K. et al. Adeno-associated virus vector-mediated production of hepatocyte growth factor attenuates liver fibrosis in mice. *Hepatol. Int.* **2**, 80–88 (2008).
81. Schievenbusch, S. et al. Combined paracrine and endocrine AAV9 mediated expression of hepatocyte growth factor for the treatment of renal fibrosis. *Mol. Ther.* **18**, 1302–1309 (2010).
82. Lee, S. H. et al. Intrathecal delivery of recombinant AAV1 encoding hepatocyte growth factor improves motor functions and protects neuromuscular system in the nerve crush and SOD1-G93A transgenic mouse models. *Acta Neuropathol. Commun.* **7**, 96 (2019).
83. Matsuda, E., Obama, Y. & Kosai, K.-I. Safe and low-dose but therapeutically effective adenovirus-mediated hepatocyte growth factor gene therapy for type 1 diabetes in mice. *Life Sci.* **268**, 119014 (2021).
84. Morishita, R. et al. Combined analysis of clinical data on HGF gene therapy to treat critical limb ischemia in Japan. *Curr. Gene Ther.* **20**, 25–35 (2020).
85. Li, H.-S. et al. Multidimensional control of therapeutic human cell function with synthetic gene circuits. *Science* **378**, 1227–1234 (2022).
86. Jia, Y. et al. Next-generation CRISPR/Cas9 transcriptional activation in *Drosophila* using flySAM. *Proc. Natl Acad. Sci. USA* **115**, 4719–4724 (2018).
87. Ewen-Campen, B. et al. Optimized strategy for in vivo Cas9-activation in *Drosophila*. *Proc. Natl Acad. Sci. USA* **114**, 9409–9414 (2017).
88. Yamagata, T. et al. CRISPR/dCas9-based *Scn1a* gene activation in inhibitory neurons ameliorates epileptic and behavioral phenotypes of Dravet syndrome model mice. *Neurobiol. Dis.* **141**, 104954 (2020).
89. Magnusson, J. P., Rios, A. R., Wu, L. & Qi, L. S. Enhanced Cas12a multi-gene regulation using a CRISPR array separator. *eLife* **10**, e66406 (2021).
90. Griffith, A. L. et al. Optimization of Cas12a for multiplexed genome-scale transcriptional activation. *Cell Genom.* **3**, 100387 (2023).
91. Hsiung, C. C.-S. et al. Engineered CRISPR–Cas12a for higher-order combinatorial chromatin perturbations. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-024-02224-0> (2024).
92. Vora, S. et al. Rational design of a compact CRISPR–Cas9 activator for AAV-mediated delivery. Preprint at *bioRxiv* <https://doi.org/10.1101/298620> (2018).
93. Ma, D., Peng, S., Huang, W., Cai, Z. & Xie, Z. Rational design of mini-Cas9 for transcriptional activation. *ACS Synth. Biol.* **7**, 978–985 (2018).
94. Omachi, K. & Miner, J. H. Comparative analysis of dCas9–VP64 variants and multiplexed guide RNAs mediating CRISPR activation. *PLoS ONE* **17**, e0270008 (2022).
95. Mukund, A. X. et al. High-throughput functional characterization of combinations of transcriptional activators and repressors. *Cell Syst.* **14**, 746–763 (2023).
96. Ludwig, C. H. et al. High-throughput discovery and characterization of viral transcriptional effectors in human cells. *Cell Syst.* **14**, 482–500 (2023).
97. Jacobs, J., Pagani, M., Wenzl, C. & Stark, A. Widespread regulatory specificities between transcriptional corepressors and enhancers in *Drosophila*. *Science* **381**, 198–204 (2022).
98. Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V. & Hackett, J. A. Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. *Nat. Genet.* **56**, 1168–1180 (2022).
99. Bedford, M. T., Chan, D. C. & Leder, P. FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *EMBO J.* **16**, 2376–2383 (1997).
100. Macias, M. J., Wiesner, S. & Sudol, M. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* **513**, 30–37 (2002).
101. Gerber, H. P. et al. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**, 808–811 (1994).
102. Akhtar, W. et al. Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc.* **9**, 1255–1281 (2014).
103. Li, X. et al. Chromatin context-dependent regulation and epigenetic manipulation of prime editing. *Cell* **187**, 2411–2427 (2024).
104. Velimirovic, M. et al. Peptide fusion improves prime editing efficiency. *Nat. Commun.* **13**, 3512 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

¹Department of Genetics, Stanford University, Stanford, CA, USA. ²Department of Neurobiology, Harvard Medical School, Boston, MA, USA.

³Department of Biology, Stanford University, Stanford, CA, USA. ⁴Biophysics Program, Stanford University, Stanford, CA, USA. ⁵Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, MA, USA. ⁶Chan Zuckerberg Biohub—San Francisco, San Francisco, CA, USA.

⁷Department of Bioengineering, Stanford University, Stanford, CA, USA. ⁸Department of Cancer Biology, Stanford University, Stanford, CA, USA.

⁹Department of Biomolecular Chemistry and Center for Human Genomics and Precision Medicine, University of Wisconsin—Madison, Madison, WI, USA.

¹⁰Sarafan ChEM-H, Stanford University, Stanford, CA, USA. ¹¹Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA.

¹²These authors contributed equally: Josh Tycko, Mike V. Van. ✉e-mail: lbintu@stanford.edu; bassik@stanford.edu

Methods

Cell lines and cell culture

Experiments presented here were carried out in K562 (American Type Culture Collection (ATCC), CCL-243, female), HEK293T (ATCC, CRL-3216), LentiX-293T (Takara Bio, 632180), primary human CD8⁺ T cells, NALM6 (ATCC), MCF10a (ATCC), THP-1 (ATCC), WTC11 (Gladstone Institute) and J774 (ATCC) cells. Cells were cultured in a controlled humidifier at 37 °C and 5% CO₂. K562 cells were cultured in RPMI 1640 (Gibco, 11-875-119) media supplemented with penicillin (10,000 IU ml⁻¹), streptomycin (10,000 µg ml⁻¹), 2 mM L-glutamine and 10% Tet-approved FBS (Omega Scientific, 20014T). HEK293T or LentiX cells were maintained in Dulbecco's modified Eagle medium (DMEM; Gibco, 10569010) supplemented with 25 mM D-glucose (Gibco), 1 mM sodium pyruvate (Gibco), 1× GlutaMAX (Gibco), penicillin (10,000 IU ml⁻¹), streptomycin (10,000 µg ml⁻¹) and 10% Tet-approved FBS (Omega Scientific, 20014T). When HEK293T cells reached 80% confluence, they were gently washed with 1× Dulbecco's phosphate-buffered saline (DPBS) (Life Technologies) and passaged using 0.25% trypsin (Life Technologies). MCF10a cells were maintained in DMEM/F12 (Invitrogen, 11330-032) supplemented with insulin (10 µg ml⁻¹), cholera toxin (100 ng ml⁻¹), hydrocortisone (0.5 mg ml⁻¹), EGF (20 ng ml⁻¹), penicillin (10,000 IU ml⁻¹), streptomycin (10,000 µg ml⁻¹) and 5% horse serum (Invitrogen, 16050-122). J774 cells were cultured in 10 cm plates in DMEM media supplemented with 2 mM glutamine, 100 U ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin and 10% heat-inactivated FBS and were passaged two to three times weekly by exchanging media when cells reached ~90% confluency, incubating for 24 h and scraping. For long-term storage, cells were resuspended in freezing media (10% dimethyl sulfoxide (Sigma-Aldrich) and cell media) in a cryovial and frozen at -80 °C. THP-1 cells were cultured in RPMI 1640 (Gibco, 11-875-119) media supplemented with penicillin (10,000 IU ml⁻¹), streptomycin (10,000 µg ml⁻¹), 2 mM L-glutamine and 10% heat-inactivated FBS (Omega Scientific, 20014T). WTC11 (Gladstone Institute) human iPSCs were cultured in mTeSR Plus medium (STEMCELL Technologies) and cultured on Geltrex-coated plates (Gibco). Cells were routinely passaged at approximately 80% confluency by enzymatic lifting with Accutase (Innovative Cell Technologies), and media was supplemented with 10 µM Y27632 (MedChemExpress) for the first 24 h after passaging. Primary CD8⁺ T cells were isolated using the EasySep Human CD8⁺ T cell Isolation Kit (STEMCELL Technologies) from the whole peripheral blood provided by healthy donors at the Blood Donor Center at Boston Children's Hospital through a protocol approved by the Boston University Institutional Review Board. Cells were activated by Human T-activator CD3/CD28 for T cell expansion and activation Dynabeads (Gibco, 11131D) and cultured in X-Vivo 15 Medium (Lonza, 04-418Q) supplemented with 5% human AB serum (Valley Biomedical, HP1022), 10 mM N-acetyl L-cysteine (Sigma-Aldrich, A9165), 55 µM 2-mercaptoethanol (Gibco) and 100 U ml⁻¹ interleukin 2 (PeproTech, 200-02). NALM6 target cell lines were cultured in RPMI 1640 medium supplemented with 5% FBS, 1% L-glutamine (Gibco, A2916801) and 1% penicillin-streptomycin.

Generating reporter cell lines

The minCMV, human pEF1α from the *EEF1A1* gene, RSV, human UbC and human PGK promoters were selected to span a range of basal expression levels. To include more minimal promoters with low basal expression, two promoters (NTX and NT21, which both lack a TATA box, in contrast to minCMV) were selected from a published resource¹⁰⁵. These promoters were each cloned into a reporter plasmid with homology arms for integration into the AAVS1 safe harbor locus (Supplementary Table 4).

Reporter cell lines were generated by transcription activator-like effector nuclease (TALEN)-mediated homology-directed repair via integration of a donor construct into the AAVS1 locus of cells as follows: 1.2×10^6 K562 cells were electroporated in Amaxa solution

(Lonza Nucleofector 2b, setting T-16), or 8×10^4 HEK293T cells were transfected using Lipofectamine LTX (Invitrogen), according to manufacturer's instructions, with 1,000 ng of reporter donor plasmid and 500 ng of each TALEN-L (Addgene, 35431) and TALEN-R (Addgene, 35432) plasmid (targeting upstream and downstream the intended DNA cleavage site, respectively). Cells were selected with 500 ng ml⁻¹ puromycin (InvivoGen) starting 48 h after electroporation/transfection for 5 days or until all of the negative control cells died. The TALEN plasmids were gifts from F. Zhang¹⁰⁶. These cells were not authenticated.

Domain library design and cloning

The nuclear protein Pfam domain library was designed previously³². Briefly, we queried the UniProt database¹⁰⁷ for human genes that can localize to the nucleus. We then retrieved Pfam-annotated domains using the ProDy searchPfam function¹⁰⁸. We filtered for domains that were 80 amino acids or shorter and excluded the C2H2 zinc finger DBDs, which are highly abundant, repetitive and not expected to function as transcriptional effectors. We retrieved the sequence of the annotated domain and extended it equally on either side to reach 80 amino acids total. Duplicate sequences were removed, and then codon optimization was performed for human codon usage, removing BsmBI sites and constraining GC content to between 20% and 75% in every 50 nucleotide window using DNA Chisel¹⁰⁹. A total of 499 random controls of 80 amino acids lacking stop codons were computationally generated as controls. A total of 362 elements tiling the DMD protein in 80 amino acid tiles with a 10 amino acid sliding window were also included as controls because DMD was not thought to be a transcriptional regulator. In total, the library consists of 5,954 elements.

The CRTF tiling library was previously designed³⁴. Briefly, 735 chromatin regulators and 1294 TF proteins were tiled with 80 amino acid tiles with a 10 amino acid tile sliding window. In addition, 2,028 random sequence controls (lacking stop codons) were computationally generated. Duplicate sequences were removed, sequences were codon optimized for human codon usage, 7xC homopolymers were removed, BsmBI restriction sites were removed, rare codons (less than 10% frequency) were avoided and the GC content was constrained to be between 20% and 75% in every 50 nucleotide window (performed with DNA Chisel¹⁰⁹). The library includes a total of 128,565 elements (including small sublibraries that were not analyzed here), which were split across five subpools that were separately amplified and cloned into the dCas9 recruitment plasmid pJT216.

Oligonucleotides with lengths of 300 nucleotides to encode these protein sequences and flanking cloning and amplification adapters were synthesized as pooled libraries (Twist Biosciences; Supplementary Table 6). The methods used to amplify these oligonucleotides, clone them into a lentiviral dCas9 recruitment vector pJT216 (Addgene, 187320; Supplementary Table 5), produce lentiviral libraries and deliver them to cells are described in detail in Supplementary Note 4.

We also used previously cloned plasmid libraries wherein these domains were cloned onto the rTetR or rTetR(SE-G72P) mutant version of rTetR with reduced leakiness^{32,34,110}. In addition, we integrated previous HT-recruit datasets that also used these previously cloned libraries, with rTetR used for the Pfam library at the pEF reporter in K562 cells and rTetR(SE-G72P) used for all other screens.

HT-recruit to measure repressor activity at reporters

For the nuclear Pfam domain repressor screens, 4×10^7 K562 PGK or UbC reporter cells per replicate were infected with 72 ml of the lentiviral library by spinfection. Cells were started in T150 flasks. Two days later, cells were ~10% mCherry⁺ by flow cytometry (ZES Cell Analyzer), and selection began with 10 µg ml⁻¹ blasticidin (InvivoGen) in T225 flasks. Cells were maintained in log growth conditions each day by diluting cell concentrations back to 5×10^5 cells per ml, with at least 1×10^8 cells total remaining per replicate, such that the lowest maintenance coverage was >10,000× cells per library element. On day 7 postinfection,

recruitment was induced by treating the cells with $1,000 \text{ ng ml}^{-1}$ doxycycline (Tocris Bioscience) for 5 days, and then cells ($>5,000\times$ coverage) were collected for magnetic separation.

HEK293T screens were performed in T225 flasks and split daily via dissociation with $1\times$ TrypLE (Gibco). For HEK293T cells with the pEF reporter, 1.25×10^7 cells per replicate were infected by spinfection with 25 ml of virus in six-well plates with 5 ml per well, resulting in 85% mCherry⁺ 2 days later. Two days after infection, $1 \mu\text{g ml}^{-1}$ doxycycline and $10 \mu\text{g ml}^{-1}$ blasticidin were added. Four days later, 75% of cells ($\sim 6,500\times$ coverage) were collected for magnetic separation, and the remaining cells were washed twice and resuspended in fresh media. Eight days later, those cells ($10,000\times$ coverage) were collected for magnetic separation.

For HEK293T cells with the UbC reporter, 1.65×10^7 cells per replicate were infected by overnight incubation with varied doses of virus in T225 flasks. The screen continued with the high dose (4.13 ml), resulting in $\sim 24\%$ mCherry⁺ cells 2 days later. Two days after infection, $10 \mu\text{g ml}^{-1}$ blasticidin was added. Five days later, the cells were frozen. Later, cells were thawed, and then blasticidin was added and refreshed daily for 6 days. Then, $1 \mu\text{g ml}^{-1}$ doxycycline was added, and 4 days later, cells ($>6,500\times$ coverage) were collected for magnetic separation followed by sequencing (described in detail in Supplementary Note 4).

HT-recruit to measure activation activity at reporters

For the nuclear Pfam domain activator screens, 4×10^7 K562 NTX or NT21 reporter cells per replicate were infected with 72 ml of the lentiviral library by spinfection. Cells were started in T150 flasks. Two days later, cells were $\sim 10\%$ mCherry⁺ by flow cytometry (ZE5 Cell Analyzer), and selection began with $10 \mu\text{g ml}^{-1}$ blasticidin (InvivoGen) in T225 flasks. Cells were maintained in log growth conditions each day by diluting cell concentrations back to 5×10^5 cells per ml, with at least 1×10^8 cells total remaining per replicate, such that the lowest maintenance coverage was $>10,000\times$ cells per library element. On day 7 postinfection, recruitment was induced by treating the cells with $1,000 \text{ ng ml}^{-1}$ doxycycline (Tocris Bioscience) for 2 days, and then cells ($>6500\times$ coverage) were collected for magnetic separation.

The protocol was similar for HEK293T cells, but cells were maintained daily in a T225 flask at $>1.4 \times 10^7$ via dissociation with $1\times$ TrypLE (Gibco). For HEK293T cells with the minCMV and silenced PGK reporters, 1.65×10^7 cells per replicate were infected by overnight incubation with varied doses of virus in T225 flasks. The screen continued with the high dose (4.13 ml), resulting in $\sim 24\%$ mCherry⁺ cells 2 days later. Two days after infection, $10 \mu\text{g ml}^{-1}$ blasticidin was added. The minCMV screen continued—5 days later, $1 \mu\text{g ml}^{-1}$ doxycycline was added, and 2 days later, the cells ($>2,000\times$ coverage) were collected for magnetic separation. Meanwhile, 5 days after blasticidin addition, PGK cells were frozen. Later, PGK cells were thawed, and then blasticidin was added and refreshed daily for 6 days. Then, $1 \mu\text{g ml}^{-1}$ doxycycline was added, and 2 days later, cells ($>6,500\times$ coverage) were collected for magnetic separation followed by sequencing (described in detail in Supplementary Note 4). Additionally, cells from the HEK293T minCMV reporter screen were collected for FLAG staining to measure fusion protein expression levels (Supplementary Note 4).

CRISPR HT-recruit to measure transcriptional effectors at endogenous genes and reporters

CRISPR HT-recruit screens were performed with dCas9 as the DBD and a sgRNA targeting either an endogenous surface marker (CD2, CD20, CD28 or CD43) or the TetO site upstream a reporter with the synthetic surface marker (TetO-minCMV or TetO-pEF1 α ; Supplementary Table 9). First, the sgRNA was stably delivered to K562 cells by lentivirus and selected with puromycin for 3–4 days. The cells were confirmed to be $>95\%$ mCherry⁺ by flow cytometry (BD Accuri C6).

For the first replicate of the CD2 (sg717 and sg718) and CD43 screens, 1.35×10^7 of these cell lines per replicate were infected with

20 ml of the $4\times$ -concentrated (LentiX) lentiviral dCas9-Pfam library by spinfection (split and spun in $2 \times 10 \text{ ml}$ volumes in 50 ml Falcon tubes). The cells were $\sim 10\%$ BFP⁺ by flow cytometry (ZE5) after 5 days (as measured from a set-aside plate that did not receive antibiotic selection). For the CD20, CD28, other CD2 sgRNAs and the second replicate of the CD2 sg717 screens, 1.35×10^7 of these cell lines per replicate were infected with 20 ml of the $4\times$ -concentrated (LentiX) lentiviral dCas9-Pfam library by spinfection in 4 ml per well in six-well plates. The cells were $\sim 68\%$ BFP⁺ by flow cytometry (ZE5) after 5 days. For the TetO sgRNA and the second replicate of the CD43 screens, 1.5×10^7 cells were spinfected with 9 ml of $4\times$ -concentrated virus split in three wells of a six-well plate, resulting in $\sim 50\%$ BFP⁺ cells.

For all of these samples, cells were maintained in T175 flasks. Two to three days after infection, selection began with $10 \mu\text{g ml}^{-1}$ blasticidin (InvivoGen). Cells were maintained in log growth conditions each day by diluting cell concentrations back to 5×10^5 cells per ml (and replenishing blasticidin), with at least 4.2×10^7 cells total remaining per replicate, such that the lowest maintenance coverage was $>5,000\times$ cells per library element. On day 10 postinfection, cells ($>20,000\times$ coverage) were collected for magnetic separation.

For the dCas9-CRTF screens, lentivirus for the library was generated using $16 \times 15 \text{ cm}$ dishes of HEK293T cells and then concentrated $4\times$ using LentiX. Then 1.15×10^8 K562-sgRNA cells (sg717 or sg15) per replicate were infected with 72 ml of the lentiviral library by spinfection for 2 h with two separate biological replicates of the infection, resulting in $18\text{--}23\%$ BFP⁺ cells in unselected cells after 4 days. Two days after infection, the cells were selected with $10 \mu\text{g ml}^{-1}$ blasticidin (InvivoGen). Cells were $>95\%$ BFP⁺ by the final time point. On day 11 postinfection, 5×10^8 cells ($>3,000\times$ coverage) were collected for magnetic separation.

An additional growth-based CRISPR HT-recruit screen was performed with the Pfam domain library and a sgRNA targeting the eGATA1 enhancer (which has a fitness effect when repressed) or a safe-targeting negative control with no fitness effect. For the first replicate, 1.35×10^7 cells of these sgRNA cell lines per replicate were infected with 20 ml of the $4\times$ -concentrated (LentiX) lentiviral dCas9-Pfam library by spinfection (split and spun in $2 \times 10 \text{ ml}$ volumes in 50 ml Falcon tubes). Cells were maintained in T175 flasks. The cells were $\sim 10\%$ BFP⁺ by flow cytometry (ZE5) after 5 days (as measured from a set-aside plate that did not receive antibiotic selection). Three days after infection, cells were selected with $10 \mu\text{g ml}^{-1}$ blasticidin (InvivoGen) for 7 days. The second replicate was performed later, with 1.5×10^7 cells spinfected with 9 ml of virus split in three wells of a six-well plate, resulting in $\sim 60\%$ BFP⁺ cells. Cells were maintained in log growth conditions each day by diluting cell concentrations back to 5×10^5 cells per ml (and replenishing blasticidin), with at least 4.2×10^7 cells total remaining per replicate, such that the lowest maintenance coverage was $>5,000\times$ cells per library element. On day 13 postinfection, cells ($>20,000\times$ coverage) were collected for genomic DNA extraction and sequencing (described in Supplementary Note 4). Additionally, screen cells with safe N4293 and CD43 sg15 were collected for FLAG staining to measure fusion protein expression levels (Supplementary Note 4).

HT-recruit analysis

The HT-recruit-Analyze pipeline¹¹¹ was used to run Bowtie to align Illumina sequencing reads to an index of domain-encoding sequences and compute effector enrichment scores. Sequencing reads were trimmed to remove the 19–24 bp stagger plus constant primer handle region from the beginning of the read. For most samples, reads were further trimmed to retain only the first 131 bp of the domain-encoding sequence. Alignment parameters were optimized for a high alignment rate ($>80\%$) and a low ambiguous alignment rate ($<1\%$). With these metrics, for the Pfam library, mismatch tolerance was set to Bowtie -m 3 (as each domain is codon optimized so they can differ at more DNA base pairs than amino acid residues), while for the tiling library,

no mismatches were allowed. This mismatch tolerance approach is tailored toward ignoring sequencing errors while removing reads with synthesis or PCR errors because it allows mismatches if the sum of their Q scores is <30 and the number of mismatches in the initial 28 bp is less than $-m$.

Then, the $\log_2(\text{OFF:ON})$ score was computed using sample depth-normalized counts in the OFF (unbound) and ON (bound) samples. To avoid inaccurate measurements from low coverage, if an element had fewer than five counts in one sample (bound or unbound fraction), its count was set to five. If an element had fallen below that threshold in both samples, it was removed from the data tables. For downstream analyses, the Pfam and tiling libraries were further filtered for elements with >5 or >20 counts, respectively, in both samples.

Analysis was performed similarly for the FLAG-based expression level measurements, using the FLAG high and low samples. An element was defined as well-expressed if its $\log_2(\text{FLAG high:low})$ score was greater than 1 s.d. above the median of the random controls.

Then, an element from the Pfam library was defined as an activator or repressor hit if its $\log_2(\text{OFF:ON})$ score was 2 s.d. below (for activators) or above (for repressors) the median of the poorly expressed domains (as defined by FLAG measurements with that DBD and cell type). This approach provided a conservative hit threshold that was chosen to identify robust effectors, but subthreshold domains may also have some effector activity. For some analyses of the pEF1 α rTetR screen in K562 cells, we also report a lower threshold (0.9), equivalent to the strength of the weakest repressor (DUF1087) that was individually validated³². The tally of domains that are hits in both replicates of a screen excludes growth-based screens, screen memory time points after doxycycline washout and screens with one replicate (sgCD2_718, sgCD20_135, sgCD20_148, sgCD20_275, sgCD28_07, sgCD28_56, sgCD28_94, sgCD2_39, sgCD2_42 and sgCD2_89). Additional analysis of the tiling library is described in Supplementary Note 4.

CRISPRi effector benchmarking growth screens

Detailed descriptions of the methods for three CRISPRi benchmarking screens using growth readouts are in Supplementary Note 4.

Individual recruitment assays

The methods used to perform individual recruitment assays are described in detail in Supplementary Note 4. In total, these were performed (1) by transient transfection or lentiviral delivery; (2) with rTetR, dCas9, dCas12a, dCasMINI, MS2, rapamycin-inducible ZFHD1 or synZiFTR recruitment; (3) in K562, HEK293T, THP-1, J774, iPSC or T cells.

Western blot for effector fusion expression

K562 cells were transduced with a lentiviral vector containing a dCas9-3XFLAG-effector-T2A-BFP-BSD and 3 days later selected with blasticidin ($10 \mu\text{g ml}^{-1}$) for 15 days until complete. Greater than 80% of the cells appeared BFP⁺ by microscopy. In total, 5–10 million cells were lysed in lysis buffer (1% Triton X-100, 150 mM NaCl, 50 mM Tris (pH 7.5) and protease inhibitor cocktail). Protein amounts were quantified using the Pierce BCA Protein Assay Kit (Bio-Rad). Equal amounts were loaded onto a gel and transferred to a PVDF membrane. The membrane was probed using FLAG M2 monoclonal antibody (mouse; Sigma-Aldrich, F1804; 1:1,000) and β -actin antibody (rabbit; Abcam, ab8227; 1:1,000) as primary antibodies. Antirabbit 680 (IRDye 680LT donkey antirabbit IgG secondary antibody; Li-Cor, 926-68023) and antimouse 800 (IRDye 800CW goat anti-mouse IgG secondary antibody; Li-Cor, 926-32210) at 1:10,000 dilution were used as secondary antibodies. Blots were imaged on a Li-Cor Odyssey CLx. Band intensities were quantified using ImageJ¹¹².

Chromatin modification mapping and RNA-seq

Stable K562 cell lines expressing dCas9 fusions were analyzed by CUT&RUN and RNA-seq (details are given in Supplementary Note 4).

Analysis software for flow cytometry data

The flow cytometry data were analyzed with the MATLAB program EasyFlow (<https://antebilab.github.io/easyflow/>) or the Python program Cytoflow (<https://cytoflow.github.io/>).

Statistical analyses

Nonlinear regression, the Kruskal–Wallis test and two-way analysis of variance were performed in Prism 9.3.1. Other statistical analyses were performed in Python using SciPy¹¹³. No methods were used to determine whether the data met the assumptions of the statistical approach.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All Illumina sequencing data generated in this study are deposited in the National Center for Biotechnology Information Sequence Read Archive database (PRJNA1160796)¹¹⁴. Processed data are given in Supplementary Tables 1–3 (<https://doi.org/10.5281/zenodo.13756269>)¹¹⁵. Plasmids generated in this study are available from Addgene (<https://www.addgene.org/browse/article/28228872/>). An uncropped scan of the western blot is provided as source data. Source data are provided with this paper.

Code availability

Code to analyze HT-recruit domain screens (<https://github.com/bintulab/HT-recruit-Analyze>) and sgRNA screens (<https://github.com/elifesciences-publications/dmorgens-castle>) is available online.

References

- Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).
- Sanjana, N. E. et al. A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* **7**, 171–192 (2012).
- UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
- Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
- Zulkower, V. & Rosser, S. DNA Chisel, a versatile sequence optimizer. *Bioinformatics* **36**, 4508–4509 (2020).
- Roney, I. J., Rudner, A. D., Couture, J.-F. & Kærn, M. Improvement of the reverse tetracycline transactivator by single amino acid substitutions that reduce leaky target gene expression to undetectable levels. *Sci Rep.* **6**, 27697 (2016).
- Tycko, J. HT-recruit-analyze. *GitHub* github.com/bintulab/HT-recruit-Analyze (2020).
- Rueden, C. T. et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinform.* **18**, 529 (2017).
- Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
- Tycko J., et al. Development of compact transcriptional effectors using high-throughput measurements in diverse contexts. Datasets. *NCBI Sequence Read Archive* <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1160796> (2024).
- Tycko J., et al. Development of compact transcriptional effectors using high-throughput measurements in diverse contexts. Datasets. *Zenodo* <https://doi.org/10.5281/zenodo.13756269> (2024).
- Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).
- Leers, J., Treuter, E. & Gustafsson, J. A. Mechanistic principles in NR box-dependent interaction between nuclear hormone receptors and the coactivator TIF2. *Mol. Cell. Biol.* **18**, 6001–6013 (1998).

118. UniProt Consortium UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
 119. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
 120. Varadi, M. et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
 121. Horlbeck, M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).
 122. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
- J.T., M.V.V. and A. performed HT-recruit, other screens and individual validations. G.H. designed and tested sgRNAs. A. and K.L. designed and performed dCas12a experiments. J.T., D.Y., N.D., A.X.M. and P.H.S. designed oligonucleotide libraries. P.H.S. characterized homeodomain mutants. J.T., A. and D.Y. performed CRISPRi benchmarking screens. M.G. and R.A.K. performed experiments with J774 cells. K.L. performed experiments with MCF10a cells. R.V. and C.Z. performed experiments with THP-1 cells. A.V.-J. and S.A. performed experiments with iPSC. A. and R.V. performed RNA-seq. X.X. designed and performed dCasMINI experiments with support from L.S.Q. J.T., V.K. and P.P.D. designed and performed MS2-recruitment experiments. H.Y. designed, performed and analyzed synZiFTR T cell experiments with support from A.S.K. and assistance from T.S.O. J.T. and M.V.V. analyzed the data and prepared the figures, with additional analysis of the tiling library by N.D. J.T., M.V.V., L.B. and M.C.B. wrote the manuscript with input from all authors.

Acknowledgements

We thank members of the Bassik Lab and Bintu Lab for helpful conversations and assistance. We acknowledge the Program in Cellular and Molecular Medicine Flow and Imaging Cytometry Resource at Boston Children's Hospital and help with their cytometers from V. Haridas and A. Agarwal. We thank M.E. Greenberg (Department of Neurobiology, Harvard Medical School) for providing laboratory space for some experiments. J.T. is supported by the F99/K00 Fellowship of the National Institutes of Health (NIH-1F99DK126120-01 and NIH-4K00DK126120-03). M.C.B. is supported by a grant from Stanford ChEM-H and an NIH Director's New Innovator Award (1DP2HD08406901). This work was supported by grants from NIH/Encyclopedia of DNA Elements (ENCODE) 5UM1HG009436-02 (to M.C.B.), NIH/National Institute of General Medical Sciences (NIGMS) R35M128947 (to L.B.), NIH/National Human Genome Research Institute (NHGRI) 5R01HG011866 (to M.C.B. and L.B.), NIH/National Institute of Biomedical Imaging and Bioengineering (NIBIB) R01EB029483 (to A.S.K.), National Science Foundation Graduate Research Fellowship Program (NSF GRFP) 2234657 (to T.S.O.) and NSF EF-1921677 (to A.S.K.). M.V.V. was supported by the NIH T32 training grant (T32GM007276). N.D. was supported by grants from NSF GRFP (DGE-1656518) and ARCS Foundation. X.X. was supported by the Chan Zuckerberg Initiative Neurodegeneration Collaborative Pairs Phase 2. L.S.Q. is a Chan Zuckerberg Biohub—San Francisco Investigator. A.S.K. is supported by the Department of Defense Vannevar Bush Faculty Fellowship (N00014-20-1-2825).

Author contributions

J.T., M.C.B. and L.B. designed the overall study. J.T., M.V.V., A. and C.L. generated reporter cell lines. J.T. and K.S. cloned pooled libraries.

Competing interests

J.T., L.B. and M.C.B. are inventors on provisional patents related to this work and acknowledge outside interest in Stylus Medicine. L.S.Q. is a founder and scientific advisory board member of Epicrispr Biotechnologies and Refuge Biotechnologies. X.X. and L.S.Q. are inventors on provisional patents related to dCasMINI. A.S.K. is an inventor on patents related to synZiFTR, is a scientific advisor for and holds equity in Senti Biosciences and Chroma Medicine and is a cofounder of K2 Biotechnologies. All other authors declare no competing interests.

Additional information

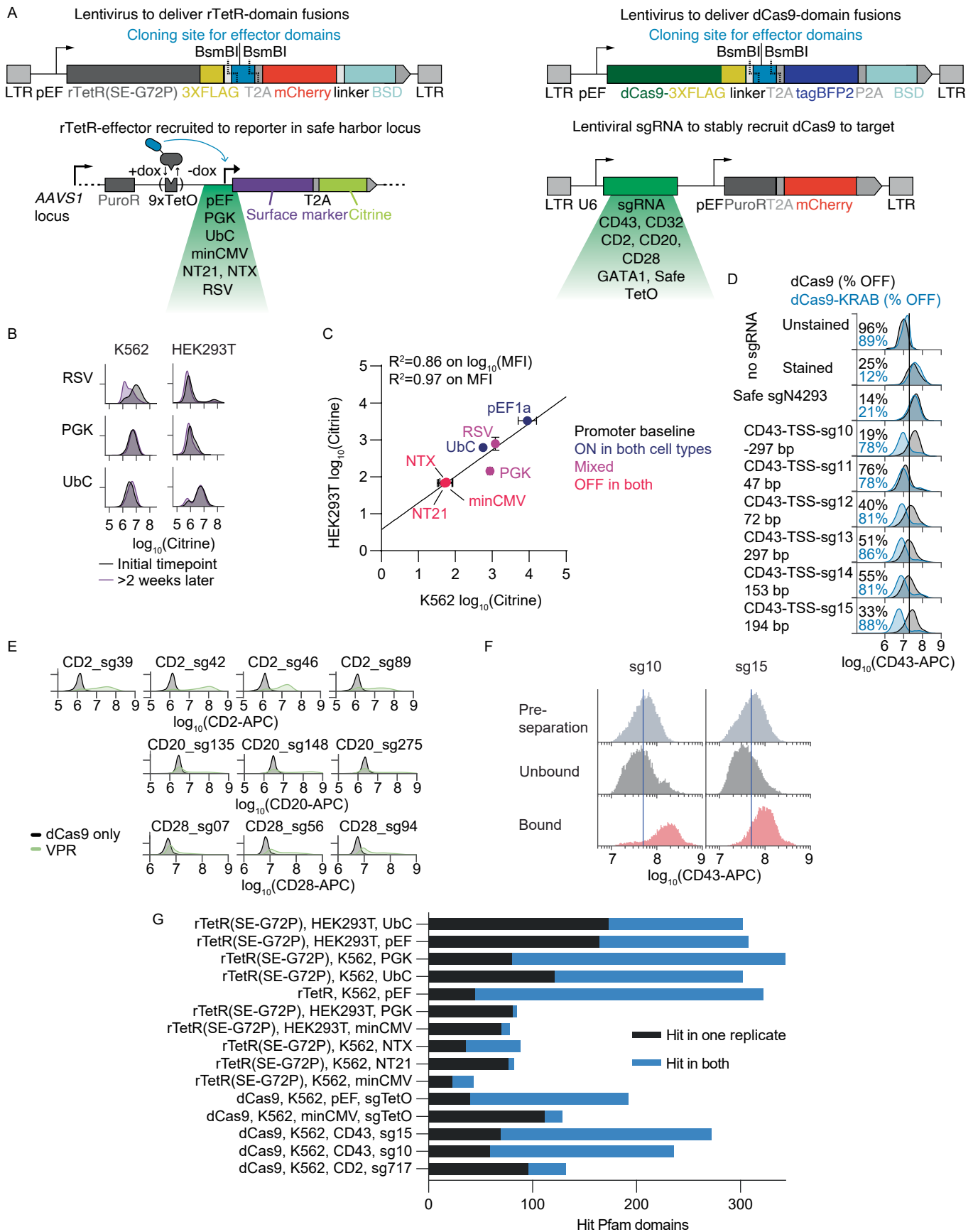
Extended data is available for this paper at <https://doi.org/10.1038/s41587-024-02442-6>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-024-02442-6>.

Correspondence and requests for materials should be addressed to Lacramioara Bintu or Michael C. Bassik.

Peer review information *Nature Biotechnology* thanks Prashant Mali and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

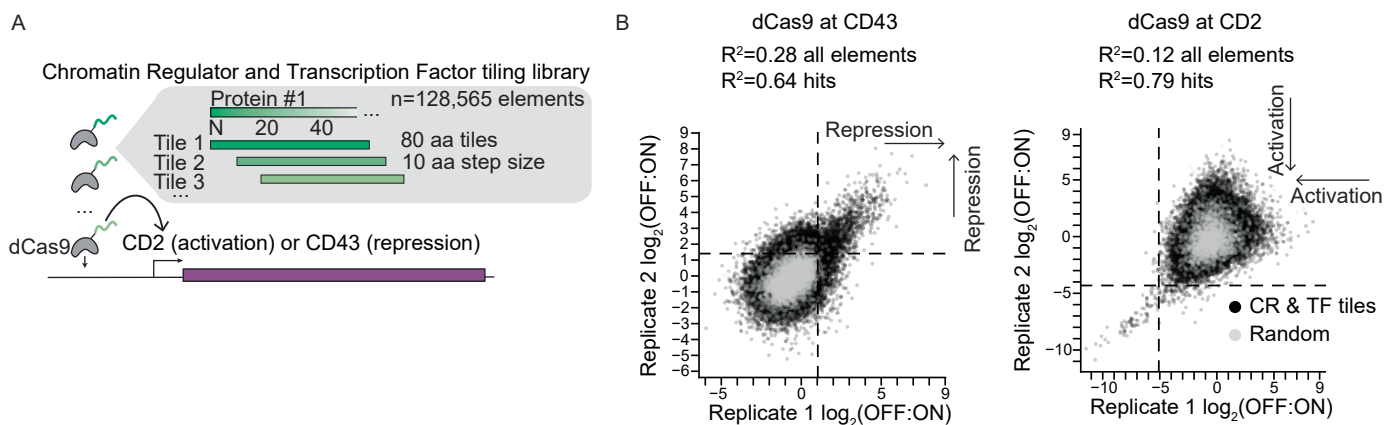
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | HT-recruit to varied reporters and endogenous gene targets in K562 and HEK293T cells. **a**, Schematics of recruitment constructs. These can be used to recruit effectors to either reporter constructs that are integrated into the AAVS1 safe harbor, or to endogenous genes. Safe sgRNAs target the genome at a safe location¹¹⁶, and the TetO sgRNA targets the synthetic reporter at an overlapping location as rTetR (the TetO motif upstream of the promoter). **b**, Observation of background silencing at reporters (n = 1 replicate per promoter type). Reporters were stably integrated at the AAVS1 safe harbor locus in both cell types by TALEN-mediated homology-directed repair. **c**, Mean fluorescent intensity (MFI) of the citrine reporter under different promoters. Each dot represents a mean from 3 replicates, and error bars show SD. Promoters with red text represent reporters that are OFF in both cell types. **d**, Silencing was measured by CD43 surface marker immunostaining and flow cytometry 7 days after lentiviral sgRNA infection in stable cell lines. Data are gated for sgRNA (mCherry⁺, only in samples with a sgRNA) and dCas9 (BFP⁺) delivery

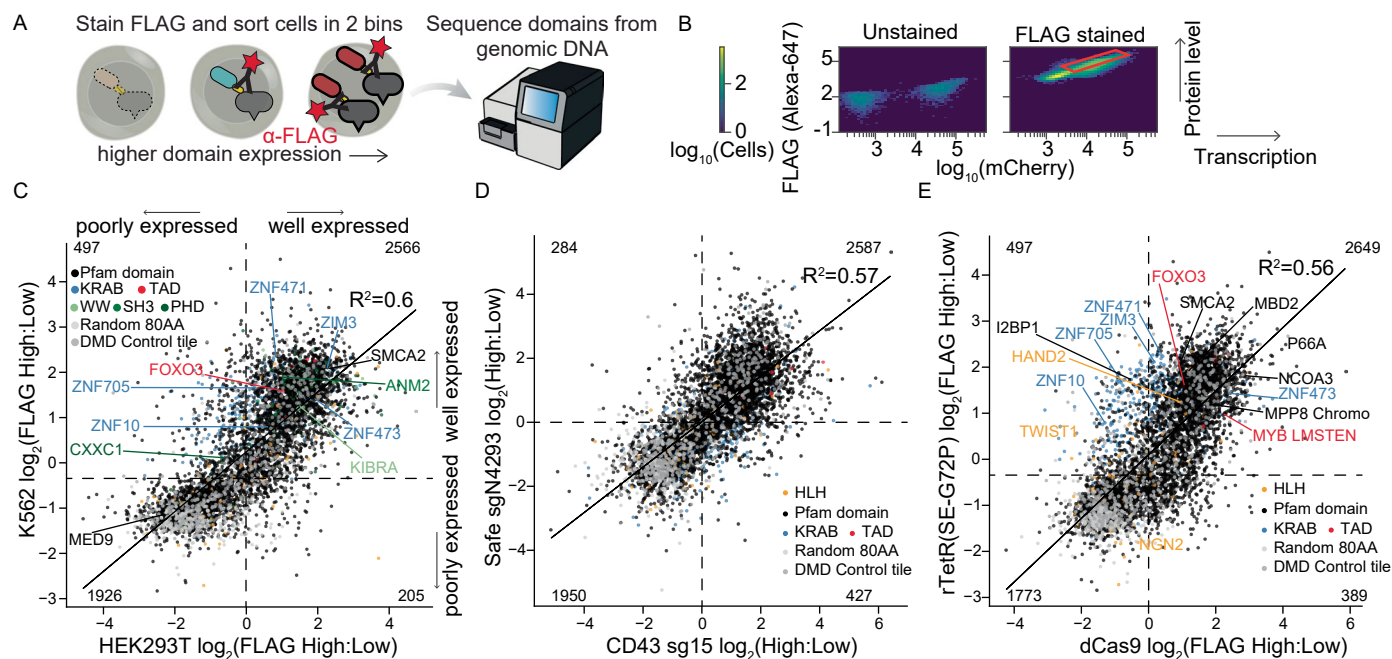
(n = 1 infection replicate). **e**, dCas9-activators targeting surface marker genes in K562 cells. sgRNAs were stably installed by lentiviral delivery and puromycin selection. Then 500 ng of dCas9 plasmids were electroporated into 1e6 cells. Two days later, cells were stained for surface CD2 (APC), CD20 (APC) or CD28 (PE) expression and analyzed by flow cytometry after gating for dCas9 (BFP) and the stably expressed sgRNA (GFP; n = 1 electroporation replicate). **f**, Magnetic separation using anti-CD43 antibody and protein G Dynabeads performed 9 days after lentiviral delivery of dCas9-Pfam library in K562 cells expressing sgRNAs that target CD43. Separation is measured by flow cytometry with gates for dCas9 (BFP⁺) and sgRNA (mCherry⁺). **g**, Overlap in hit Pfam domains in both biological replicates for HT-recruit screens, defined as elements that had ≥ 5 sequencing reads in bound and unbound and $\log_2(\text{OFF:ON})$ scores 2 standard deviations beyond (that is, higher for repressors, lower for activators) the median of the poorly expressed controls.



Extended Data Fig. 2 | dCas9 fusions to tiles of all human chromatin regulators and transcription factors uncovers unannotated effectors.

a, Schematic of a library tiling all human chromatin regulator and transcription factor (CR and TF) proteins in 80 amino acid tiles with a 10 amino acid step size ($n = 128,565$ elements)³⁴. This library was fused to dCas9 and used to target CD43 with sg15 and CD2 with sg717. **b**, Replicates of CR and TF library fused to dCas9

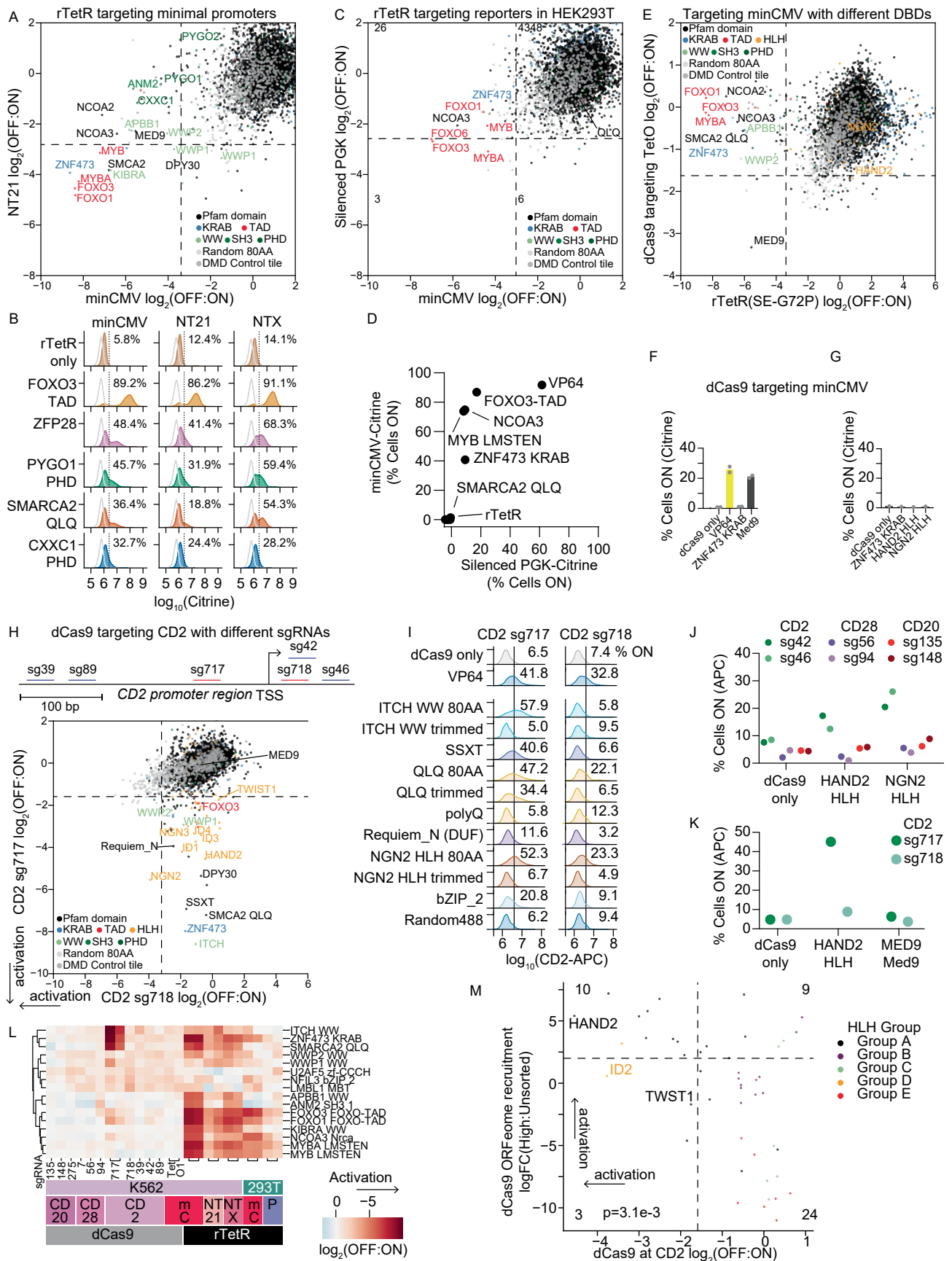
and recruited to CD43 or CD2 in K562 cells. Hit threshold shown at 2 standard deviations above (for CD43 screen) or below (CD2) the median of the random controls. Elements with >20 sequencing counts in both the bound (ON) and unbound (OFF) samples are included. The linear regression goodness of fits (R^2) is shown for all elements and for the subset that are hits in both replicates.



Extended Data Fig. 3 | Pooled measurement of domain expression level across cell types and DBDs.

a, Schematic of pooled approach to measuring domain expression. Cells are permeabilized and then stained with a fluorescent anti-FLAG antibody. Then, cells are sorted into high and low FLAG level bins using gates that account for transcription variability by using the fluorescent delivery marker (mCherry for rTetR constructs, and BFP for dCas9 constructs), which is found on the same transcript after the T2A cleavage signal. Genomic DNA is extracted, and then the domains are sequenced in those two cell populations. **b**, Gating strategy for sorting based on FLAG stain intensity in HEK293T cells. Example gate shown in red accounts for variation across cells in transcription level of the rTetR-3XFLAG-effector-T2A-mCherry transcript by using mCherry fluorescence. **c**, Comparison of Pfam domain library expression levels between

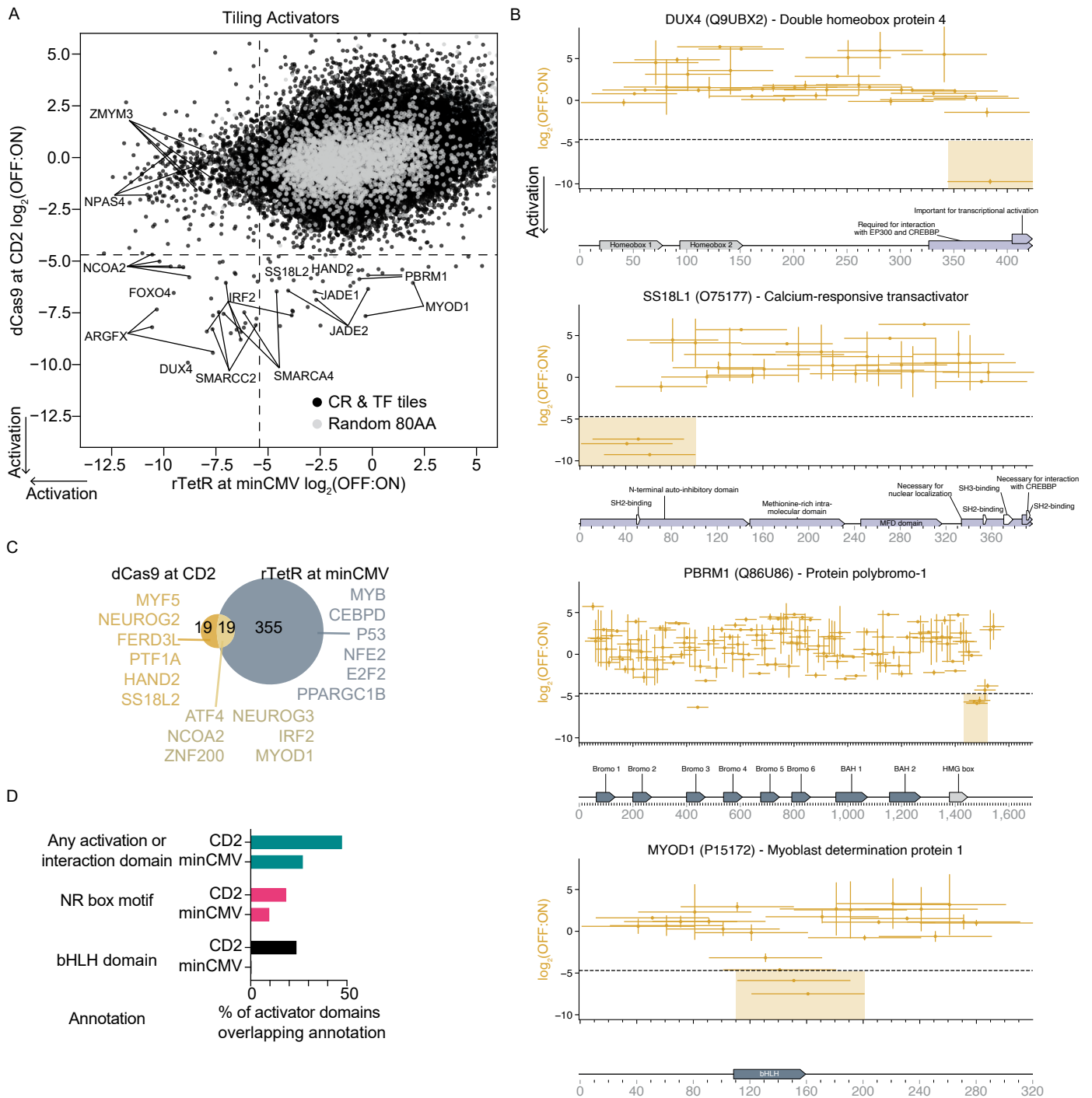
HEK293T and K562 cells (Spearman $\rho = 0.84$). In both cell types, the DBD was rTetR(SE-G72P) and the cell line was the minCMV reporter ($n = 2$ replicates). Data were filtered for elements with >5 reads in both FLAG high and low samples. Well-expressed domains were identified based on a hit threshold (dashed lines) set 1 S.D. above the median of the random controls. The number of library elements in each quadrant is labeled in the corners. **d**, Comparison of Pfam domain library expression levels when fused to dCas9 and delivered to cell lines expressing either a safe control sgRNA or a sgRNA targeting CD43 ($n = 1$ replicate). **e**, Comparison of Pfam domain library expression levels when fused to dCas9 or rTetR(SE-G72P) and delivered to K562 cells ($n = 2$ replicates). For dCas9, one replicate each from the Safe sgN4293 and CD43 sg15 cells are averaged.



Extended Data Fig. 4 | See next page for caption.

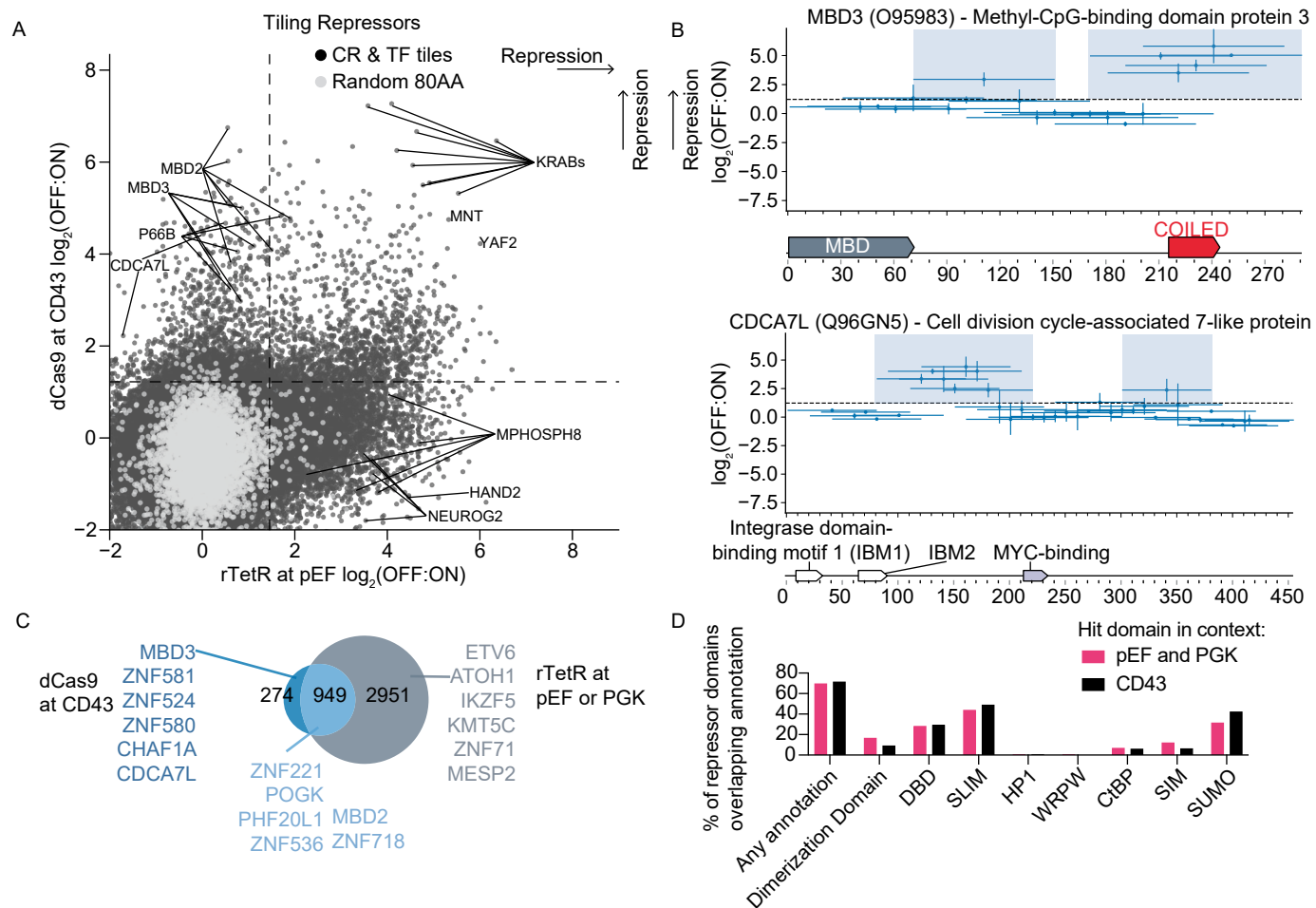
Extended Data Fig. 4 | Activators function more similarly at minimal promoters than with a different DBD or at a silenced promoter. **a**, HT-recruit with rTetR in K562 cells ($n = 2$ biological replicates). Dashed lines show hit thresholds at 2 standard deviations above the median of the poorly expressed domains. **b**, Validation of activator domains across core promoter reporters in K562 cells. rTetR-activator fusions or the rTetR-only negative control were delivered by lentivirus to reporter cells. After selection, cells were treated with 1000 ng/ml doxycycline for 2 days to induce reporter activation. The percent of cells activated was measured by flow cytometry for the citrine reporter, after gating for delivery with mCherry. Percentages normalized to no-doxycycline control and shown as an average from 3 biological replicates. **c**, HT-recruit with rTetR in HEK293T cells ($n = 2$ biological replicates per promoter). rTetR-domain fusions were recruited to the reporter with 1000 ng/ml doxycycline for 2 days. The number of Pfam domains in each quadrant is labeled. **d**, Individual validations of activators in HEK293T as measured by average percentage of cells ON normalized to no-doxycycline control. rTetR-activators were stably delivered by lentivirus. Cells were treated with 1000 ng/ml doxycycline for 2 days to induce reporter activation ($n = 2$ independently transduced replicates for each promoter type). **e**, HT-recruit with dCas9 recruitment of activators with a sgRNA that binds the TetO site upstream of the minCMV reporter in K562 cells ($n = 2$ biological replicates). **f**, dCas9 activators recruited with a sgRNA that binds the TetO site ($n = 2$ infection replicates shown as dots). dCas9 fusions were delivered by lentivirus, selected with blasticidin starting on day 5 and cells were analyzed on day 9. Flow cytometry measurements were gated for dCas9 and TetO_sg1 using BFP and mCherry, respectively. **g**, Other dCas9 activators were recruited in K562 cells with the TetO sgRNA ($n = 1$ replicate). dCas9 fusions were delivered by lentivirus, selected with blasticidin starting on day 3 and cells were analyzed on day 9 as in **f**. **h**, HT-recruit with the dCas9-Pfam domain library targeted to the CD2 gene TSS using two different guides ($n = 2$ replicates for

sg717, and $n = 1$ for sg718) in K562 cells. Screen measurement was taken 10 days after library delivery. Schematic shows locations of the CD2-targeting guides. **i**, Recruitment of dCas9-activator hits at the CD2 gene using two different guides in K562 cells. sgRNA were stably delivered by lentivirus and selected for with puromycin, then dCas9 fusion plasmids were delivered by electroporation, then cells were analyzed 3 days later by flow cytometry for surface stained CD2 after gating for dCas9 (BFP) and sgRNA (GFP). The percentage of cells ON is shown ($n = 1$ electroporation replicate). The 80 amino acid sequences match the library elements, while the trimmed sequences match the annotated Pfam domain. The polyQ is a homopolymer of 15 repeated glutamines, which is also found at the C-terminus of the 80 amino acids QLQ and is not present in the trimmed QLQ. bZIP_2 domain from CEBPE was filtered due to low counts in the screen. **j**, dCas9-HLHs were delivered to K562 cells by lentivirus and selected for with blasticidin, and then sgRNAs were delivered by lentivirus and selected with puromycin. Then 8 days after sgRNA delivery, the cells were stained for the targeted surface markers and measured by flow cytometry ($n = 1$ infection replicate). Data were gated for dCas9 with BFP and sgRNA with GFP. **k**, Nine days after lentiviral delivery of dCas9 fusions, K562 cells were immunostained with CD2 antibody to measure gene activation by flow cytometry ($n = 1$ replicate). **l**, HT-recruit scores for activators that were hit in ≥ 5 samples across target, cell type and DBD contexts ($n = 2$ replicates per rTetR and sg717 screens, and $n = 1$ for other dCas9 screens shown as columns). The rows are clustered in an unbiased manner. mC = minCMV and P = PGK. **m**, Comparison of HLH domain HT-recruit scores from the dCas9 screen at CD2 with sg717 with published data, wherein the ORFeome was fused to dCas9 and recruited to activate a reporter gene³³. Dashed lines show hit thresholds. The number of elements in each quadrant is labeled (not all corresponding HLH full-length genes were included in the ORFeome study). P-value from 2-sided Fisher's exact test is shown. The HLH phylogenetic groups are shown as colors⁴⁴⁻⁴⁶.



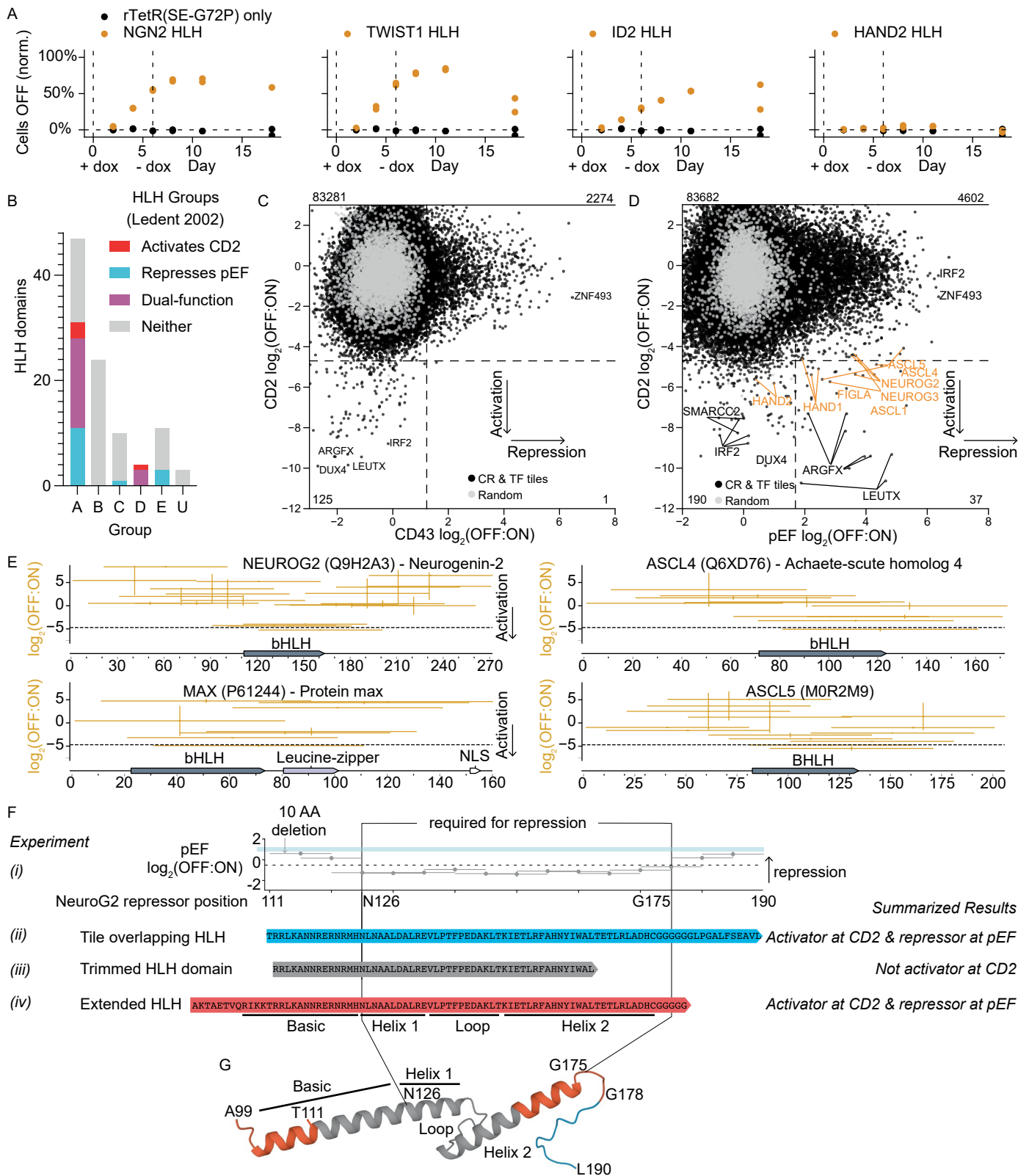
Extended Data Fig. 5 | dCas9 fusions to tiles of all human chromatin regulators and transcription factors uncovers unannotated and HLH activators. **a**, dCas9 recruitment of CR and TF tiles to CD2 compared with rTetR recruitment to minCMV. Dashed lines show average of hit thresholds ($n = 2$ replicates per screen). **b**, Proteins with activator hit tiles. Each horizontal line is a tile, and vertical bars show the range ($n = 2$ screen replicates). Dashed horizontal line is the hit-calling threshold based on random controls. Hit domains, defined as the sequence from the start of the first hit tile to the end of the last hit tile for a stretch of 1 or more consecutive hit tiles (that are below the hit threshold in

both replicates), are shaded. UniProt annotations and Pfam domains are shown below. **c**, Overlap of hit activator domains from different contexts. Shared hits are defined as hit domains with any overlapping sequence. Proteins containing the top 6 strongest hit domains are listed, and for the shared hit category, the proteins with the top 6 strongest CD2 activators are listed. **d**, Percentage of hit domains overlapping annotations. NR box motif is associated with recruitment of nuclear receptor coactivators¹⁷. Some domains are hits in both contexts or overlap multiple annotations.



Extended Data Fig. 6 | dCas9 recruitment of tiling library to CD43 uncovers unannotated repressors. **a**, Comparison of dCas9 recruitment of the CRTF tiling library to CD43 with rTetR recruitment to pEF1 α . Dashed lines show hit threshold ($n = 2$ replicates per screen) based on random 80 amino acids controls. **b**, Tiling methyl-binding domain protein MBD3 and CDCA7L. Each horizontal line is a tile, and vertical bars show the range ($n = 2$ screen replicates). Dashed horizontal line is the hit-calling threshold based on random controls. **c**, Overlap of hit repressor domains from different contexts. Proteins containing the top 6 strongest hit

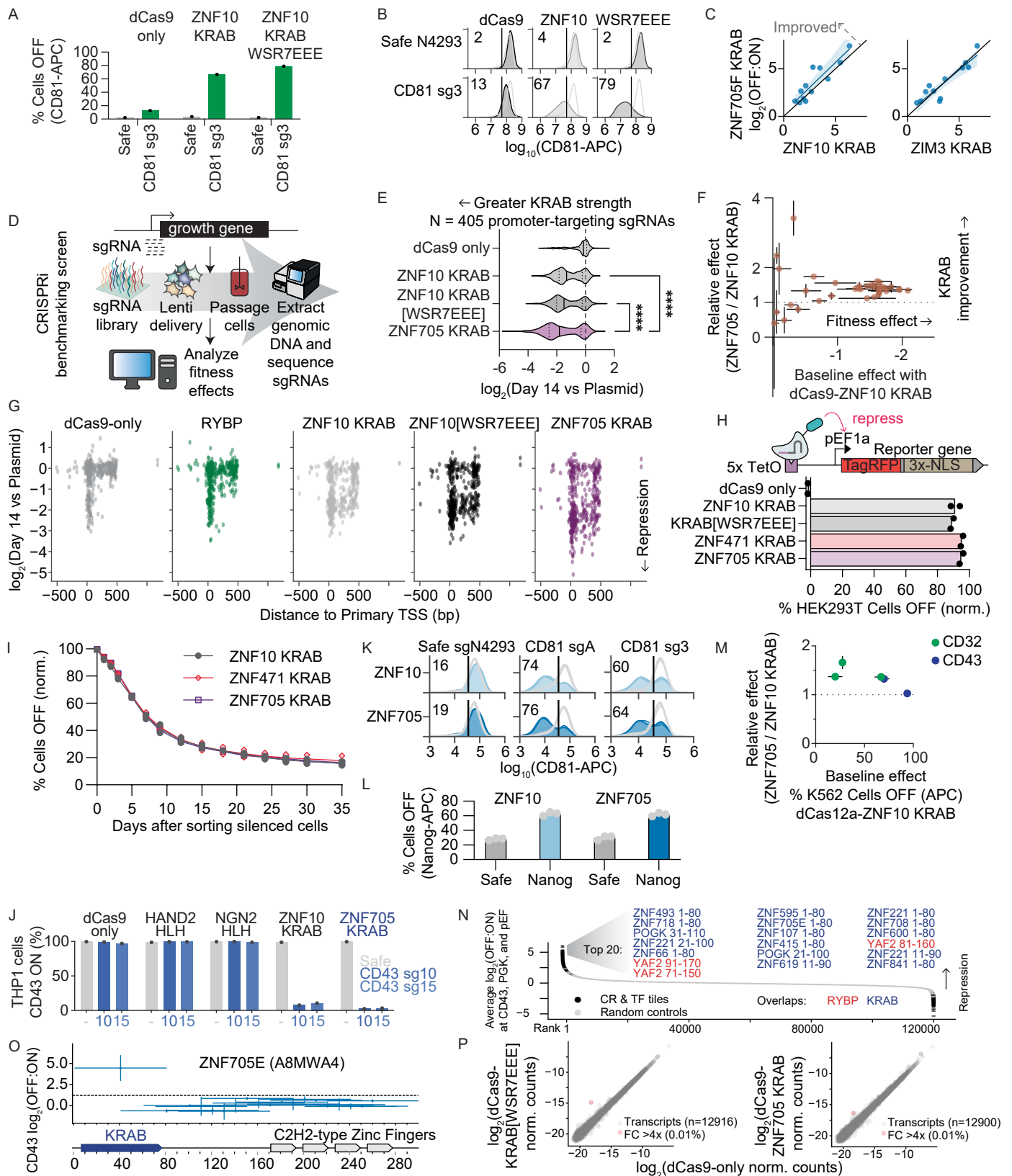
domains are listed, and for the shared hit category, the proteins with the top 6 strongest CD43 repressors are listed. **d**, Percentage of hit domains overlapping a curated set of annotations of interest. SLIM are short linear interaction motifs. HP1, WRPW and CtBP motifs are associated with co-repressor recruitment. SIM are SUMO-interacting motifs, and SUMO are SUMOylation sites. 'Any annotation' refers to any of these. To conservatively select hit domains from the rTetR screens, here we used domains that are hit with both pEF and PGK. Some domains are hits in both rTetR and dCas9 contexts or overlap multiple annotations.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Dual-functioning HLH domains. **a**, rTetR fusions were delivered by lentivirus to K562 cells with the pEF1 α reporter. Cells were selected with blasticidin, and then cells were treated with doxycycline for 6 days, then doxycycline was washed out to measure silencing memory. The percentage of cells silenced was measured by flow cytometry for the citrine reporter after gating for delivery with mCherry and normalizing to the matched no-doxycycline control (n = 2 infection replicates shown as dots). **b**, Classification of HLH effector hits from the Pfam library screens at pEF with rTetR and CD2 with dCas9 in K562 cells. Hits are shown as colors and compared with phylogenetic grouping defined in refs. 44–46 and reported as used here in ref. 46. For this analysis, an HLH domain is counted as repressing pEF if the average $\log_2(\text{OFF:ON})$ across biological replicates is above the lower threshold (>0.9, equivalent to the weakest individually validated repressor; Methods). U, unclassified. **c**, dCas9 recruitment of the CRTF tiling library to CD2 and CD43. Dashed lines show hit thresholds at 2 standard deviations below or above the median of the random controls (n = 2 replicates per screen). **d**, Comparison of dCas9 recruitment of the CRTF tiling library to CD2 with rTetR recruitment to pEF (n = 2 replicates per screen). Labels are orange for HLH proteins. **e**, Tiling HLH proteins and targeting CD2. Each horizontal line is a tile, and vertical bars show the range (n = 2 screen replicates). Dashed horizontal line is the hit-calling threshold. UniProt annotations and Pfam domains are shown below. The bHLH domains contain a short N-terminal basic

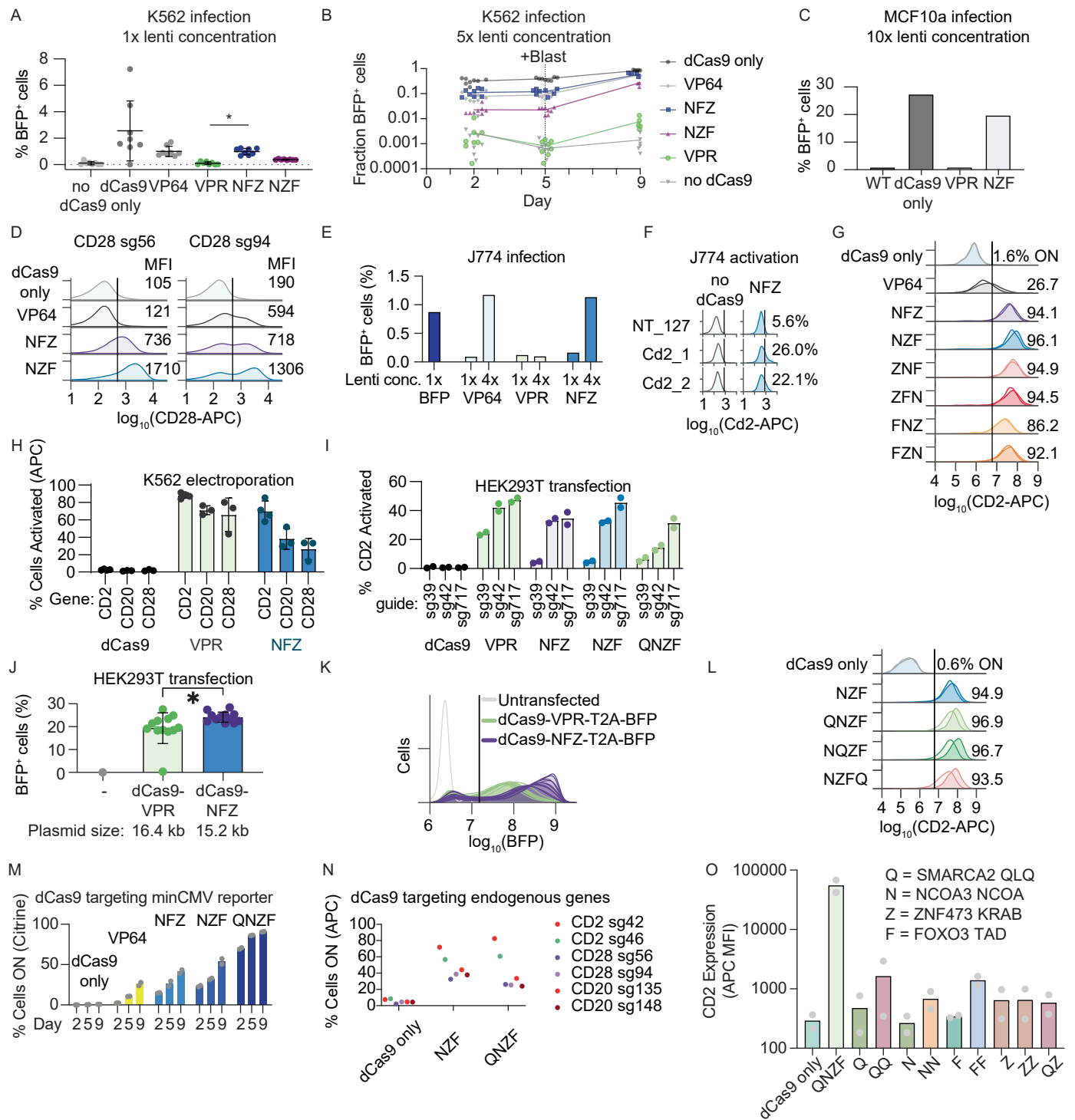
DNA-binding region followed by an HLH heterodimerization region. **f**, Integrated analysis of NGN2 experiments. From the top, (i) shows deletion scan data for an 80 amino acids repressor tile in NGN2 that contains the HLH domain³⁴. Each short horizontal bar depicts a 10 amino acid deletion from the 80 amino acids tile, which was recruited to the pEF reporter. The light blue shaded region shows the reference score for the full 80 amino acids tile. Bars below the dashed line are deletions that ablate repressor function. Dots show the middle of the deletion, and vertical error bars show the range of two biological replicates. Below, individual recruitment experiments (ii–iv) with NGN2 are depicted, with colored sequences showing what was recruited and the summarized results written on the right. (ii) The blue sequence is the same tile that was deletion scanned in (i). (iii) The trimmed Pfam-annotated HLH domain is shown in gray, and (iv) the extended 80 amino acids HLH domain from the Pfam library is shown in red. These recruitment assay data are shown in **d**, Fig. 2g and Extended Data Fig. 4i. **g**, AlphaFold structural prediction of the NGN2 region, accessed via UniProt ID Q9H2A3 (refs. 118–120). Residues that delimit the tested domains are labeled. The HLH region shown to be necessary by deletion scanning (N126 to G175) is bracketed. The boundary between the basic region and helix 1 is not structurally distinguishable and is determined here based on the presence of basic residues. Colors correspond with **f**.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | KRAB mutant and paralogs with improved CRISPRi efficiency. **a**, dCas9 constructs were delivered by lentivirus to K562 cells, and then sgRNAs were delivered by lentivirus. Three days later, the cells were selected with puromycin and blasticidin. Then, 9 days after sgRNA infection, the cells were stained for CD81 expression, fixed and analyzed by flow cytometry with gating for both dCas9 (BFP) and sgRNA (mCherry; $n = 1$ infection replicate). **b**, Shaded histogram shows the cells expressing both the dCas9 vector (BFP) and the sgRNA (mCherry); their percentage of cells OFF is shown. The unshaded histogram shows the cells from the same sample that express neither ($n = 1$ replicate). **c**, HT-recruit scores for selected KRAB domains compared across cell type, DBD and target gene contexts that measured repression directly or by using a growth phenotype associated with repressing GATA1. Each dot shows the average of two screen biological replicates ($n = 12$ contexts). Black line shows identity, and blue line shows linear regression with shaded 95% confidence interval estimated using a bootstrap. **d**, Schematic of a CRISPRi benchmarking screen for comparing KRAB repressors (Methods). **e**, CRISPRi benchmarking screen. Greater depletion over 14 days of growth relative to the original plasmid pool representation is associated with stronger effector-mediated silencing of the essential gene promoters. Violin shows distribution of average sgRNA-level depletion from two screen replicates; solid line shows median; dotted lines are quartiles ($N = 405$ sgRNAs, **** $P < 0.0001$ by Kruskal–Wallis test). The dashed line at 0 represents the median of the safe-targeting negative controls. **f**, Comparison KRABs from CRISPRi benchmarking screen targeting the promoters of 37 essential genes. Effect sizes are the \log_2 (fold-change) of sgRNA representation after 14 days of growth relative to the original plasmid pool representation. The median effect across 8–10 sgRNAs per gene was computed, and each dot shows its average for two infection replicate screens. Horizontal and vertical error bars show the ranges. Dashed line shows parity between KRAB domains. **g**, sgRNA distance from the primary TSS of 37 essential genes as defined by FANTOM^{121,122} ($n = 405$ sgRNAs shown as dots, average of 2 screen replicates). **h**, Top, schematic of dCas9-repressor recruitment at the pEF1 α -TagRFP-T reporter in HEK293T using TetO-targeting sgRNA. Bottom, transient transfection of dCas9–KRAB paralogs with a sgRNA targeting the TetO sites upstream of the reporter gene (bar shows mean from $n = 2$ biological replicates shown as dots). Percentage of cells OFF was normalized to safe-targeting sgRNA. **i**, After targeting the dCas9–KRAB paralog fusions for 5 days by transient transfection at the TagRFP reporter in HEK293T cells, silenced cells were sorted, and memory dynamics was measured

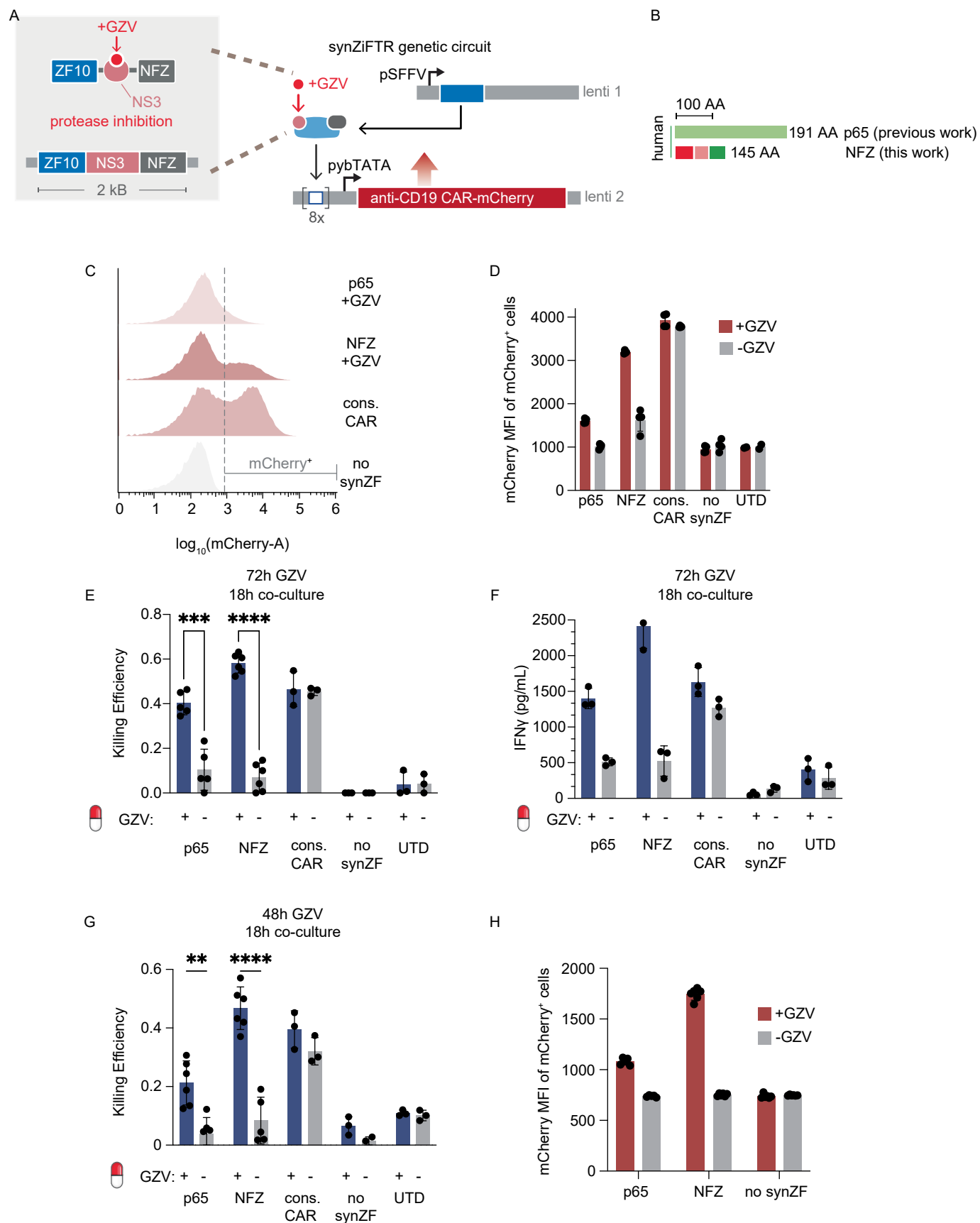
by flow cytometry throughout 35 days. Each dot is a biological replicate ($n = 2$). The percentage of cells OFF was normalized to safe-targeting sgRNA. **j**, sgRNAs and then dCas9 fusions were delivered by lentivirus to THP-1 monocytes. Eight days later, cells were stained for surface CD43 expression and analyzed by flow cytometry, with gates applied for sgRNA expression (mCherry) and dCas9 expression (BFP; $n = 1$ infection replicate). KRAB data are shared with Fig. 3e. **k**, iPSCs were transfected with dCas9–KRAB and sgRNA plasmids and then stained and measured by flow cytometry 2 days later. Data were gated for sgRNA delivery (GFP for safe and sgA, or mCherry for sg3) and dCas9 (BFP; $n = 3$ transfection replicates shown as shaded histograms). The gray unshaded histogram shows the cells from the same sample that express neither and serve as an internal control. The percentage of cells OFF with the gene-targeting sgRNA, using the threshold shown by the black vertical line, is shown. **l**, iPSC were similarly transfected, then permeabilized and stained for intracellular Nanog protein expression and measured by flow cytometry 2 days later. Data were gated for sgRNA delivery (GFP) and dCas9 (BFP; $n = 3$ transfection replicates shown as dots, and bar shows mean). **m**, Comparison of baseline silencing with ZNF10 KRAB and relative improvement with ZNF705 KRAB when dCas12a fusions were recruited to silence CD43 or CD32, with the same method as in Fig. 3f. Dots show the average for two infection replicates of a guide RNA. Horizontal and vertical error bars show the ranges. Dashed line shows parity between KRAB domains. **n**, Chromatin regulator and transcription factor tiles and random controls are ranked by their mean repression scores from the pEF, PGK and CD43 screens with the larger library ($n = 2$ replicates per screen). The ZNF705E tile is 99% identical to the ZNF705B/ZNF705D/ZNF705F KRAB Pfam domain, which was not itself included in the tiling library because most of the >300 KRAB-containing ZNF proteins were excluded from the library design due to space constraints. The top 20 ranked tiles are listed, with their protein name, tile start and end position, colored by whether they overlap a KRAB or RYBP domain. **o**, Tiling ZNF705E. Each horizontal line is a tile, and vertical bars show the range ($n = 2$ screen replicates). Dashed horizontal line is the hit-calling threshold based on random controls. UniProt annotations are shown below, associated with the UniProt accession ID written above. **p**, RNA-seq of K562 cells after stable lentiviral delivery of dCas9–KRABs and blasticidin selection compared to cells expressing dCas9-only ($n = 3$ technical replicates). No sgRNA was included. Red dots show transcripts with fold-change (FC) >4 in either direction. The total number of transcripts (after filtering for at least 10 counts) is labeled.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Characterization of transcriptional control tools using compact human activators. **a**, dCas9-activator delivery to K562 cells by lentiviral infection. Flow cytometry measuring BFP delivery marker was performed 3 days after infection (n = 8 infection replicates). *P < 0.0001 by unpaired 2-sided t test for comparison of NFZ and VPR. **b**, 5× LentiX-concentrated lentivirus was used and cells were analyzed by flow cytometry at 3 time points. Blasticidin selection was initiated on day 5. Each dot is an independent infection (n = 8 replicates per dCas9 construct). **c**, Lentiviral delivery to MCF10a breast cancer cells with 10× LentiX concentration, measured by flow cytometry after 4 days (n = 1 replicate). **d**, dCas9 activators recruited to CD28 in K562 cells (n = 1 infection replicate). The mean fluorescence intensity (MFI) is labeled. Vertical black line shows the threshold chosen to quantify the percentage of cells ON in Fig. 4f. **e**, Lentiviral delivery to J774 mouse macrophages with or without 4× LentiX concentration. 1e5 cells were infected with 1 mL of virus for 24 hours. Flow cytometry was performed 6 days after infection (n = 1 replicate). BFP is a positive control lentivirus (pEF-PuroR-P2A-BFP). **f**, Activation of endogenous Cd2 in J774 macrophages after lentiviral sgRNA delivery (n = 1 infection replicate). Data are gated for sgRNA delivery (GFP⁺) and for fair comparison, not gated for dCas9 delivery. Non-targeting (NT) sgRNAs are negative controls. **g**, Optimization of the tripartite activator by changing the N, F and Z domains' orientation. The various configurations were fused onto dCas9 and targeted to the CD2 gene in K562 cells. sgRNAs and then dCas9 fusions were delivered by unconcentrated lentivirus and selected with puromycin and blasticidin, respectively. Activation was measured 8 days after dCas9 infections by immunostaining CD2 with an APC-conjugated antibody followed by flow cytometry. The average percentage of cells ON is shown. The darker-shaded histogram is CD2 sg717, and the lighter shade is sg718 (n = 1 infection per sgRNA). dCas9-only and VP64 control data are shared with Fig. 1c. **h**, dCas9 activators targeting CD2, CD20 and CD28 surface marker genes in K562 cells. sgRNAs (CD2: sg39, sg42, sg46, sg89; CD20: sg135, sg148, sg275; CD28: sg7, sg56, sg94) were stably installed by lentiviral delivery and puromycin selection. Then 500 ng of dCas9 plasmids were electroporated into 1e6 cells. Two days later, cells were stained for surface CD2 (APC), CD20 (APC) or CD28 (PE) expression and analyzed by flow cytometry after gating for dCas9 (BFP) and the stably expressed sgRNA (GFP). Each dot represents a different sgRNA targeting the gene (n = 3 for CD20 and CD28; n = 4 for CD2), and error bars show SD. Control data are shared with Extended Data Fig. 1e. **i**, Percentage of CD2 endogenous gene activated 3 days after transient transfection of dCas9 activators and a sgRNA in HEK293T cells. Cells were immunostained for CD2 (APC) expression and analyzed by flow cytometry after gating for transfection (GFP on the sgRNA

plasmid). Each dot represents an independently transfected biological replicate (n = 2). **j**, dCas9-activator delivery 2 days after plasmid transfection in HEK293T cells measured by flow cytometry for the BFP marker on the dCas9-effector-T2A-BFP-P2A-BlastR transcript, with no gating on co-transfected delivery markers. dCas9-NFZ is significantly better delivered than dCas9-VPR with a greater fraction of BFP⁺ cells using a linear gate at 1.5×10^7 (P = 0.0281, two-tailed t-test). Each dot is an individual co-transfection of 500 ng of dCas9 plasmid and 300 ng of a sgRNA plasmid in a 24-well plate (n = 11 transfection replicates for NFZ and VPR) or an untransfected negative control. Bar shows mean, and error bars show standard deviation. **k**, BFP expression level in the same samples after accounting for overall delivery efficiencies by gating for transfectable cells based on the presence of GFP (from the co-transfected sgRNA plasmid). Each line is an individual co-transfection or the untransfected control, and the black line shows the linear gate for BFP⁺ cells (n = 10 for VPR and 11 for NFZ). **l**, Fusion of SMARCA2 QLQ (Q) to dCas9-NZF. Nine days after lentiviral delivery of dCas9 activators, K562 cells were immunostained with CD2 antibody to measure gene activation by flow cytometry, and the average percentage of cells ON is shown. The darker-shaded histogram is CD2 sg717, and the lighter shade is sg718 (n = 1 replicate per sgRNA). **m**, dCas9 recruitment of activators with a sgRNA that binds the TetO site upstream the minCMV reporter in K562 cells (n = 2 infection replicates). dCas9 fusions were delivered by lentivirus and cells were analyzed 2, 5 and 9 days later with blasticidin selection starting at day 5. Flow cytometry measurements were gated for dCas9 and TetO_sg1 using BFP and mCherry, respectively (VPR measurements not shown due to having <100 BFP⁺ cells). Control data are shared with Extended Data Fig. 4g. Bars show mean, and error bars show standard deviation. **n**, dCas9 activators were delivered to K562 cells by lentivirus and selected for with blasticidin, and then sgRNAs were delivered by lentivirus and selected for with puromycin. Then 8 days after sgRNA delivery, the cells were stained for the targeted surface marker genes and measured by flow cytometry. Surface marker expression is shown after gating for dCas9 with BFP and sgRNA with GFP, and dCas9-only data are shared with Extended Data Fig. 4j, which was performed in parallel. **o**, Homotypic combination of Q, N, Z and F activators fused onto dCas9 and delivered stably by lentivirus to target CD2 in K562 cell lines stably expressing the sgRNA. The mean fluorescence intensity (MFI) of CD2 staining (Alexa 647) of the cell population after gating for delivery of the sgRNA (GFP) and dCas9 fusion (BFP) is shown. Staining was performed 9 days after dCas9 fusion infection. Each point is a sgRNA (sg717 or sg718), and bars show the mean of two different sgRNAs (n = 2 infection replicated).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Inducible chimeric antigen receptor (CAR) expression with compact human activator domains in primary T cells. **a**, Schematic of the synZiFTR genetic circuit for inducible CAR expression. Lenti 1 constitutively expresses a synthetic transcription factor with a zinc finger 10 (ZF10) DBD fused to an NS3 protease domain and an activator domain. Upon protease inhibition by GZV treatment, the activator is inducibly stabilized and recruited to the second lentivirus, which contains 8× ZF10-binding motifs upstream of a minimal pybTATA to drive expression of an anti-CD19 chimeric antigen receptor (CAR) fused to mCherry. **b**, Schematic of activator domains recruited with synZiFTR. **c**, Distribution of CAR-mCherry expression levels in primary T cells. Cells were co-infected with synZiFTR-expressing and CAR-mCherry lentiviruses (or reporter-alone controls), and then expression was induced with 1 μM GZV for 72 hours. Cells were analyzed by flow cytometry with no gating applied for delivery of synZiFTR (n = 1 representative replicate shown). **d**, Lentiviruses expressing synZiFTRs with p65 or NFZ activator domains were normalized for titer and then delivered to primary human T cells. Simultaneously, cells were co-infected with a ZF10-targeting CAR-mCherry lentivirus, or with a constitutive CAR as a control. After 3 days of infection and 3 days of puromycin selection, cells were treated with GZV or DMSO for 72 hours. Then mCherry induction was measured by flow cytometry (n = 2 infection replicates with 2 technical replicates each shown as dots, bars show mean and error bars show standard deviation).

'No synZF' control virus only contains the CAR-mCherry transgene, and UTD is untransduced. Data are gated for mCherry⁺ cells. **e**, After 72 hours of induction with GZV, SynZiFTR T cells were co-cultured overnight for 18 hours with the NALM6 BFP-expressing cancer cell line while being continuously treated with GZV. Killing efficiency was estimated by measuring the reduction in BFP⁺ cells compared to a NALM6-only negative control by flow cytometry (n = 2 infection replicates measured with 3 technical replicates each shown as dots, bars show mean and error bars show standard deviation). ***P < 0.005, ****P < 0.00005, by unpaired 2-sided t-test. **f**, At the same time point, supernatant was collected from the cells and interferon-γ secretion was measured by ELISA (n = 3 technical replicates shown as dots, bars show mean and error bars show standard deviations). **g**, After 48 hours of induction with GZV, SynZiFTR T cells were co-cultured overnight for 18 hours with the NALM6 cells and then measured by flow cytometry (n = 2 infection replicates measured with 3 technical replicates each shown as dots, bars show mean and error bars show standard deviation). **P < 0.0005, ****P < 0.00005, by unpaired 2-sided t-test. UTD is untransduced. **h**, An independent batch of primary T cells was transduced with synZiFTR lentiviruses. Cells were induced with GZV for 48 hours and then measured by flow cytometry (n = 2 infection replicates measured with 3 technical replicates each shown as dots, and bar shows mean).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data collected on the Cytex Aurora cytometer used SpectroFlo software (version 3.3).

Data analysis

Non-linear regression, Kruskal-Wallis test and two-way ANOVA were performed in Prism 9.3.1. Other statistical analyses were performed in Python using SciPy version 1.13 (Virtanen et al., 2020). The flow cytometry data were analyzed with the MATLAB program EasyFlow (<https://antebilab.github.io/easyflow/>) or the Python program Cytoflow (<https://cytoflow.github.io/>). This code was used to analyze HT-recruit domain screens (<https://github.com/bintulab/HT-recruit-Analyze>) and sgRNA screens (<https://github.com/elifesciences-publications/dmorgens-castle>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All Illumina sequencing data generated in this study are deposited in the NCBI SRA database (XX ID). The UniProt database (www.uniprot.org) was queried to identify nuclear-localized proteins and Pfam domains in proteins.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample size calculation was performed. Sample size values (n) are provided in figure legends and vary depending on the experiment.

Data exclusions

The CRTF tiling library includes a small subpool of elements that are not tiles exactly matching human proteins (e.g. they encode a mutation relative to wild-type protein) and these were not analyzed here. The CRISPRi benchmarking screen data includes sgRNAs that do not target promoters or enhancers (e.g. they target CTCF binding sites or other regions) and these were not analyzed here.

Replication

All screens with hits that were further developed were repeated with a second independently transduced biological replicate, and all hits were selected on basis of passing the hit threshold in both replicates. Select hits were individually validated in at least one follow-up experiment each.

Randomization

Randomization was not applicable to these kinds of experiments. Pooled screening mitigates element-wise bias by having each element measured within the same sample, handled the same way.

Blinding

Blinding was not performed because it was not practical. Pooled screening mitigates element-wise bias by having each element measured within the same sample, handled the same way.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

The following primary antibodies were used: allophycocyanin (APC)-labeled anti-CD2 antibody (1:50 dilution, 130-116-253, Miltenyi-Biotec), APC-labeled anti-CD43 antibody (1:500, clone 4-29-5-10-21, eBioscience, Catalog # 17-0438-42), APC-labeled anti-CD20 antibody (1:100, clone L26, eBioscience, Catalog # 14-0202-82), or APC-labeled anti-CD28 antibody (1:20, clone CD28.2, eBioscience, Catalog # 17-0289-42). 5 μ L / 1 x 10⁶ cells of α -FLAG-Alexa647 (RNDsystems, IC8529R) was used on permeabilized cells. APC-labeled anti-CD81 antibody (clone 1D6, eBioscience, Catalog # 17-0819-42) and APC-labeled anti-CD32 Monoclonal Antibody (clone 6C4, eBioscience, Catalog # 17-0329-42) were also used on cells for flow cytometry. For Western Blot, membrane was probed using FLAG M2 monoclonal antibody (1:1000, mouse, Sigma-Aldrich, F1804) and beta-actin antibody (1:1000, Rabbit, Abcam, ab8227) as primary antibodies. Anti-rabbit 680 (IRDye[®] 680LT Donkey anti-Rabbit IgG Secondary Antibody, 926-68023, Li-Cor) and anti-mouse 800 (IRDye[®] 800CW Goat anti-Mouse IgG Secondary Antibody, 926-32210, Li-Cor) at 1:10000 dilution were used as secondary antibodies. H3K4me3 (0.5 μ g/ml, EpiCypher), H3K9me3 (0.9 μ g/ml, Diagenode C15410193), H3K27me3 (1:50, Cell Signaling Technology 97335), H2AK119Ub (1:100, Cell Signaling Technology 82405), and IgG (0.5 μ g/ml, EpiCypher) antibodies were used for CUT&RUN. APC-labeled anti-CD55 (clone JS11, Biolegend, Catalog # 311312), APC-labeled anti-CD2 (Miltenyi, Catalog # 130-098-579), APC-labeled anti-CD81 (clone 1D6, Invitrogen, Catalog 17-0819-42), anti-NANOG primary antibody (clone 1E6C4, Santa Cruz, Catalog sc-293121), and APC labeled Goat-anti-Mouse secondary antibody (Invitrogen, Catalog # A865) were used on iPSC flow experiments.

Validation

Antibodies were used following manufacturer's recommendations and validated to function as expected with positive controls. For flow cytometry at repressible target surface marker genes (CD43, CD81, CD32), we confirmed high staining in cell lines expected to express the surface marker based on public RNA-seq, and loss of staining upon repression with dCas9 KRAB at the surface marker TSS. Inversely for activatable target surface marker genes (CD2, CD55, CD20, CD28), we validated low/no staining in the negative control and higher staining upon targeting with CRISPRa at the surface marker TSS. For CUT&RUN we observed enrichment of repressive histone modifications at a set of genes in K562 that were previously shown to have enrichment for those modifications. For Western blot, we confirmed bands at the expected size in control lanes.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Experiments presented here were carried out in K562 (ATCC CCL-243, female), HEK293T (ATCC CRL-3216), LentiX-293T (Takara Bio #632180), Primary human CD8+ T-cells, NALM6 (ATCC), MCF10a (ATCC), THP-1 (ATCC), WTC11 (Gladstone Institute), and J774 (ATCC) cells.

Authentication

Not authenticated.

Mycoplasma contamination

Not tested.

Commonly misidentified lines (See [ICLAC](#) register)

Not commonly misidentified.

ChIP-seq

Data deposition

Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session

(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	<p>cells were first washed with 1% BSA (Sigma) in 1x DPBS (Life Technologies) and spun down and supernatant was aspirated without disturbing the pellet. If adherent, cells were dissociated with 1x TrypLE (Gibco) before washing. 100 ul of cells were then incubated on ice for 1h with fluorophore conjugated primary antibody. The following primary antibodies were used: 10 ul of allophycocyanin (APC)-labeled anti-CD2 antibody (clone LT2; Invitrogen, Catalog # MA1-10132), 5 uL of APC-labeled anti-CD81 antibody (clone 1D6, eBioscience, Catalog # 17-0819-42), 5 ul of APC-labeled anti-CD43 antibody (clone 4-29-5-10-21, eBioscience, Catalog # 17-0438-42), 5 uL of APC-labeled anti-CD32 Monoclonal Antibody (clone 6C4, eBioscience, Catalog # 17-0329-42), 5 uL of APC-labeled anti-CD20 antibody (clone L26, eBioscience, Catalog # 14-0202-82), or 5 uL of APC-labeled anti-CD28 antibody (clone CD28.2, eBioscience, Catalog # 17-0289-42). Afterwards, cells were washed 1-3 times with 1x PBS or 3 times with 1% BSA/DPBS and then analyzed by flow cytometry. For J774 cells, APC anti-mouse Cd2 Antibody (BioLegend 100112, 1:100 dilution) was used.</p> <p>For iPSC, cells were lifted by incubation with Accutase and washed once with DPBS + 1% FBS + 10 µg/mL human IgG (Sigma, Catalog # I8640-50MG). Cells were then incubated with antibodies diluted 1:100 in DPBS/FBS/hlgG for 1 h on ice. For NANOG staining, iPSC were first fixed in 4% PFA at room temperature for 10 minutes, before permeabilization by incubation in DPBS/FBS/hlgG + 0.1% Triton for 30 minutes at room temperature. Cells were then stained for 1 h with NANOG primary antibody, washed, and stained again with APC-anti-mouse secondary. Antibodies used were: APC-labeled anti-CD55 (clone JS11, Biolegend, Catalog # 311312), APC-labeled anti-CD2 (Miltenyi, Catalog # 130-098-579), APC-labeled anti-CD81 (clone 1D6, Invitrogen, Catalog 17-0819-42), anti-NANOG primary antibody (clone 1E6C4, Santa Cruz, Catalog sc-293121), and APC labeled Goat-anti-Mouse secondary antibody (Invitrogen, Catalog # A865).</p> <p>Transduced CD8+ primary T-cells were induced with 1 µM GZV or DMSO for 48-72h after removal of activation beads, and added in co-culture with BFP+ NALM6 target cells at a 1:1 effector:target ratio overnight (18h) or for 48h at 37°C (100,000 T-cells:100,000 NALM6 cells per well in a 96-well plate). After the co-culture period, cells were analyzed via flow cytometry to quantify BFP+ cell count. BFP+ population was gated on control wells containing NALM6 cells only without T-cells. Killing efficiency was calculated as the reduction in BFP+ cell count compared to that in the control well.</p>
Instrument	ZE5 Cell Analyzer (Bio-Rad Laboratories), BD Accuri C6, or Attune NxT Flow Cytometer (ThermoFisher).
Software	Everest Software (Bio-Rad Laboratories) and CytoFlow (https://cytoflow.readthedocs.io/en/stable/)
Cell population abundance	At least 5k viable cells were analyzed when measuring ON and OFF statistics. The exception is viral delivery of a transgene associated with toxicity or poor survival of antibiotic selection, where in some cases fewer than 5k viable cells could be collected.
Gating strategy	FSC/SSC gates were set for viable cells. The gate for "OFF" was set at the left shoulder of a high-expressing negative control population (No-dox, or dCas9-only) such cells undergoing repression are called as OFF. The gate for "ON" was set at the right shoulder of low-expressing negative control cells (No-dox or dCas9-only) such that any activated cells are called as ON.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.