Article

Differential cofactor dependencies define distinct types of human enhancers

https://doi.org/10.1038/s41586-022-04779-x

Received: 3 October 2020

Accepted: 20 April 2022

Published online: 1 June 2022

Check for updates

Christoph Neumayr[1,2,7], Vanja Haberle[1,7], Leonid Serebreni[1,2], Katharina Karner[1], Oliver Hendy[1,2], Ann Boija[3], Jonathan E. Henninger[3], Charles H. Li[3,4], Karel Stejskal[1,5], Gen Lin[1], Katharina Bergauer[1], Michaela Pagani[1], Martina Rath[1], Karl Mechtler[1,5], Cosmas D. Arnold[1] & Alexander Stark[1,6]✉

All multicellular organisms rely on differential gene transcription regulated by genomic enhancers, which function through cofactors that are recruited by transcription factors[1,2]. Emerging evidence suggests that not all cofactors are required at all enhancers[3–5], yet whether these observations reflect more general principles or distinct types of enhancers remained unknown. Here we categorized human enhancers by their cofactor dependencies and show that these categories provide a framework to understand the sequence and chromatin diversity of enhancers and their roles in different gene-regulatory programmes. We quantified enhancer activities along the entire human genome using STARR-seq[6] in HCT116 cells, following the rapid degradation of eight cofactors. This analysis identified different types of enhancers with distinct cofactor requirements, sequences and chromatin properties. Some enhancers were insensitive to the depletion of the core Mediator subunit MED14 or the bromodomain protein BRD4 and regulated distinct transcriptional programmes. In particular, canonical Mediator[7] seemed dispensable for P53-responsive enhancers, and MED14-depleted cells induced endogenous P53 target genes. Similarly, BRD4 was not required for the transcription of genes that bear CCAAT boxes and a TATA box (including histone genes and LTR12 retrotransposons) or for the induction of heat-shock genes. This categorization of enhancers through cofactor dependencies reveals distinct enhancer types that can bypass broadly utilized cofactors, which illustrates how alternative ways to activate transcription separate gene expression programmes and provide a conceptual framework to understand enhancer function and regulatory specificity.

Multicellular organisms depend on differential gene transcription mediated by enhancers, which bind transcription factors (TFs) and recruit cofactors (COFs) to activate transcription[1]. Both COFs and the DNA-binding TFs are crucial for enhancer function[2] and transcription activation at the initiation, pause-release or elongation step[7,8]. Prominent COFs include the acetyltransferase P300 (also known as EP300) and the Mediator complex, which mediate histone modifications, RNA polymerase II (Pol II) recruitment and transcription initiation[7,9], and bromodomain-containing protein 4 (BRD4) and cyclin-dependent kinase 9 (CDK9), which mediate transcriptional pause-release and elongation[8,10].

Although COFs generally localize to active enhancers and promoters[11,12] and have long been thought to be universally required, emerging evidence suggests that different regulatory elements and genes might require different COFs[13,14]. For example, pharmacological inhibition of COFs leads to gene-specific rather than global effects (for example, for BRD4 (refs. [15,16]), CDK7 (ref. [5]) and CDK8 (ref. [17])). Meanwhile, cells can acquire resistance to BRD4 inhibition by deploying a BRD4-independent enhancer[3]. Similarly, several Mediator subunits are not necessary for the transcription of all genes[4,18]. These findings suggest that even essential COFs that localize to most or all active genes are not globally required for transcription and that individual enhancers can bypass some of the COFs. However, whether such examples reflect more general gene-regulatory principles, such as different enhancer types with distinct properties and regulatory roles, has remained unknown. Moreover, systematic analyses of COF requirements for enhancer-mediated transcription activation are lacking.

To systematically discern the dependency of enhancers on various COFs, we measured genome-wide enhancer activities in human HCT116 cells in the presence and absence of specific COFs. As many COFs are essential and their prolonged depletion affects cell viability[15,18], we used the auxin-inducible degron (AID) system[19] to rapidly induce the depletion of COF proteins. We coupled this to the quantitative assessment of enhancer activities for millions of fragments across the entire human genome using the plasmid-based massively parallel reporter assay STARR-seq[6] (Fig. 1a).

[1]Research Institute of Molecular Pathology, Vienna BioCenter, Vienna, Austria. [2]Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria. [3]Whitehead Institute for Biomedical Research, Cambridge, MA, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Institute of Molecular Biotechnology, Vienna BioCenter, Vienna, Austria. [6]Medical University of Vienna, Vienna BioCenter, Vienna, Austria. [7]These authors contributed equally: Christoph Neumayr, Vanja Haberle. ✉e-mail: stark@starklab.org

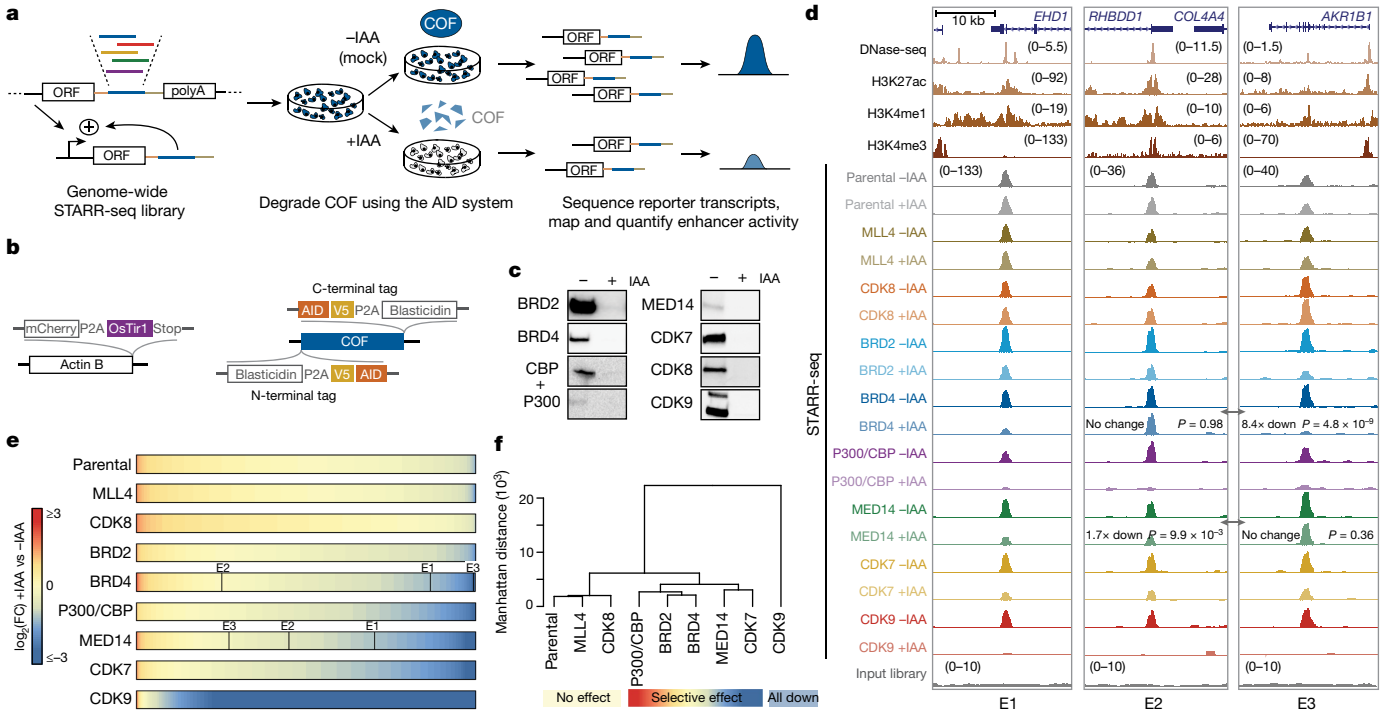


Fig. 1 | Rapid COF degradation coupled to STARR-seq reveals COF-specific effects on enhancer activity. a, Schematic of the experimental approach. HCT116 cells with a COF of interest tagged by an AID were transfected with a genome-wide STARR-seq library and treated with either IAA to degrade the COF or with a mock control. Enhancer activity across the entire human genome was quantified in the two conditions by sequencing and mapping reporter transcripts. **b**, Schematic of the COF tagging strategy. Left, the parental HCT116 cell line carries a heterozygous insertion of the OsTir1 ligase downstream of the gene encoding actin B. Right, an AID-tagged cell line was created for each COF through the homozygous insertion of a cassette containing an AID to either the N terminus or the C terminus of the respective COF gene in the parental cell line. **c**, Western blots of denoted COFs in the cell line in which the respective COF is tagged by AID, with and without IAA treatment for 1 h. Experiment was done once, validated by MS. Source gel data are provided in Extended Data Fig. 1a and Supplementary Fig. 1. **d**, Activity of three enhancers (E1–E3) measured by STARR-seq in different COF-AID cells with and without IAA treatment (normalized STARR-seq signal for merged replicates; adjusted *P* values of the edgeR negative binomial model). Endogenous chromatin accessibility and histone modifications in WT HCT116 cells are shown on the top. **e**, $\log_2$(fold change (FC)) values for a reference set of 6,249 enhancers, sorted individually for each COF-AID cell line from the least affected (or most upregulated) enhancers on the left to the most downregulated enhancers on the right. Three enhancers shown in **d** are marked for BRD4 and MED14 cell lines. **f**, Hierarchical clustering of parental and COF-AID cell lines based on the $\log_2$(FC) of enhancer activity in IAA-treated versus untreated cells shown in **e**.

COF-AID cells enable rapid COF depletion

To generate COF AID-tagged cell lines (COF-AID cells), we first created a parental cell line that uniformly expresses the *Oryza sativa* Tir1 (OsTir1) ligase (Fig. 1b, left), and subsequently knocked-in the AID tag homozygously at individual COF genes[19] (Fig. 1b, right). We created eight cell lines to deplete various COFs that regulate crucial steps of transcription: the bromodomain-containing BRD2 and BRD4; the structural core Mediator subunit 14 (MED14); the acetyltransferases P300 and CBP (both tagged in a single cell line; P300/CBP); the cyclin-dependent kinase CDK7 (a core TFIIH subunit); the Mediator kinase CDK8; the pTEFb kinase CDK9; and the methyltransferase MLL4 (as HCT116 cells lack the MLL4 paralogue MLL3 (ref. [20]), MLL4 depletion should deplete MLL3 and MLL4 functionality).

Treatment with auxin (specifically, 3-indoleacetic acid (IAA)) strongly depleted all tagged COFs after 1 h (Fig. 1c and Extended Data Fig. 1a). Shotgun mass spectrometry (MS) of IAA-treated MED14-AID cells revealed a greater than twofold depletion of all detectable Mediator subunits, which suggests that Mediator is disintegrated as expected (Extended Data Fig. 1b,c and refs. [4,21]). A targeted MS approach for all COFs after 3 h of IAA treatment revealed no (BRD4, CBP, CDK7, CDK8 and MLL4) or low (<15%; BRD2, P300, MED14 and CDK9) residual levels (Extended Data Fig. 1d). After 2 days, COF degradation strongly affected proliferation for all COFs except CDK8 and MLL4, for which proliferation was not affected even after 5 days (Extended Data Fig. 1e,f). This result is consistent with reports that CDK8 and MLL4 are not essential in HCT116 cells[22,23].

Enhancers have distinct COF dependencies

To assess enhancer activity changes following the loss of each COF, we performed STARR-seq in the parental and the eight COF-AID cell lines after mock or IAA treatment (Fig. 1a). In brief, we transfected the cells with a genome-wide STARR-seq library comprising more than 50 million genomic fragments of 1.2 kb (ref. [6]) (about 22 times genome coverage), treated half of the cells with water (mock) or IAA, and collected cellular RNA after 6 h (see Extended Data Fig. 2a for different time points of BRD4 depletion). We added spike-in RNA to total cellular RNA for normalization, and then isolated, amplified and quantified the poly-adenylated reporter transcripts by deep sequencing.

We performed three replicates per condition for the parental cell line, two for CDK9-AID cells and four for all other COF-AID cells (replicates had pairwise Pearson's correlation coefficients of ≥0.7; Extended Data Fig. 2a). We first defined a set of enhancers that were strongly active in at least one condition using all replicates and stringent thresholds (Methods), which detected between 141 and 1,979 enhancers per condition (fewer in COF-depleted conditions) and 6,249 enhancers in total.

Without IAA treatment, STARR-seq results from COF-AID cells were similar to the parental controls (Extended Data Fig. 2b), which suggests that COF and enhancer functions were maintained. The only exception was the double-tagged P300/CBP cells, which showed reduced enhancer activity in the absence of IAA (Extended Data Fig. 2c); this result might be due to significant pre-degradation of both COFs

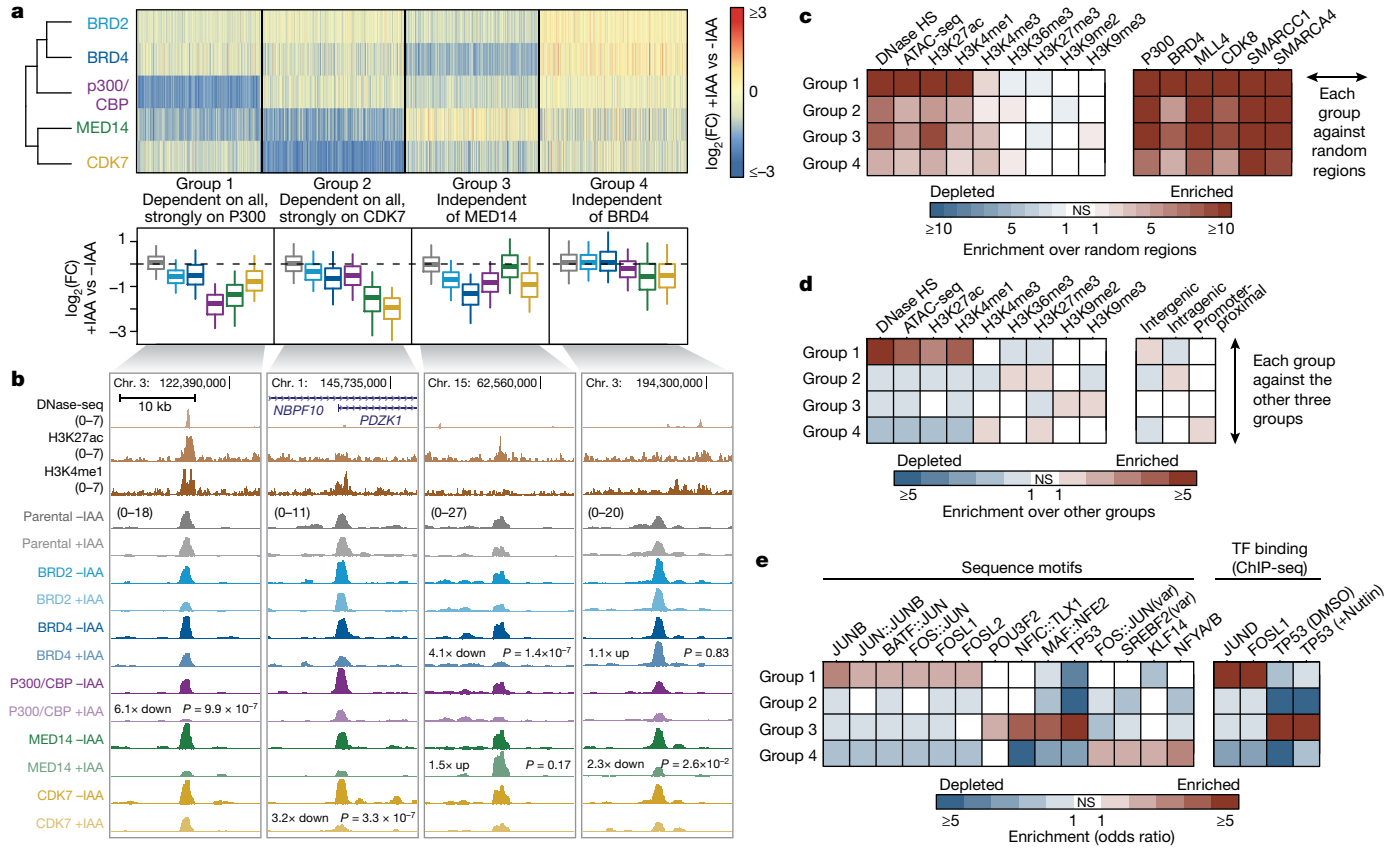


Fig. 2 | Differential COF requirements define distinct enhancer types with distinguishing sequence and chromatin features. a, $\log_2$(FC) of enhancer activity after individual COF degradation for four groups of enhancers defined by PAM clustering. Box plots summarize the values per COF for each group. n = 1,392; 1,660; 1,519 and 1,678 for groups 1, 2, 3 and 4, respectively. Boxes show the median and interquartile range, whiskers the 5th and 95th percentiles. **b**, Examples of enhancers from each of the four groups showing activity in different COF-AID cell lines with and without IAA treatment (normalized STARR-seq signal for merged replicates; adjusted P values of the edgeR negative binomial model). **c**, Enrichment of chromatin accessibility and histone modification ChIP-seq peaks (left) and various COF ChIP-seq peaks (right) from HCT116 cells for the four groups of enhancers against random control regions. **d**,**e**, Mutual enrichment of chromatin accessibility and histone modification ChIP-seq peaks (**d**, left), genomic localization (**d**, right), or TF motifs (**e**, left) and TF ChIP-seq peaks (**e**, right) for the four groups of enhancers. The enrichment for each group was calculated against the remaining three groups. Statistically significant (two-sided Fisher's exact test; P value ≤0.05) enrichments and depletions are in shades of red and blue, respectively. Non-significant (NS) fields are shown in white. Nutlin, Nutlin-3a.

(Extended Data Fig. 1d). However, the loss of enhancer activity was marginal compared with the effects after IAA-induced COF degradation (≤15% of enhancers, compare Extended Data Fig. 2c and d), and IAA treatment downregulated pre-affected and non-pre-affected enhancers to similar extents (Extended Data Fig. 2e), which suggests that P300/CBP-dependent enhancers can be studied.

Overall, COF depletion revealed different effects for different COFs. Degradation of CDK8 and MLL4 did not affect enhancer activity (Fig. 1d–f and Extended Data Fig. 2b,d,f), which is consistent with unaltered proliferation and reports that CDK8 and MLL4 are dispensable in HCT116 cells[22,23] (Extended Data Fig. 1e,f). By contrast, CDK9 depletion led to global inactivation of enhancers (Fig. 1d,e and Extended Data Fig. 2d,f), which is consistent with the role of CDK9 during pause-release and elongation[8,10].

Degradation of the remaining COFs had more selective effects, with some COFs, such as BRD2 and BRD4, having more similar effects than others (Fig. 1f), and some enhancers were downregulated, whereas others were unaffected or even upregulated (Fig. 1d,e and Extended Data Fig. 2d,f). For instance, BRD4 loss had no effect on an enhancer in the *RHBDD1* gene, but strongly impaired an enhancer in *AKR1B1*, whereas the opposite was true for MED14. Taken together, rapid COF degradation coupled to STARR-seq revealed differential COF dependencies for individual enhancers.

COF dependencies define four enhancer types

The result that not all enhancers depend similarly on all COFs suggests that there are enhancer groups with specific COF requirements. To reveal such groups, we clustered the 6,249 enhancers on the basis of enhancer activity change after degradation of each of the five COFs that showed selective effects (BRD2, BRD4, P300/CBP, MED14 and CDK7). Using partitioning around medoids (PAM, *k*-medoids), we defined four distinct groups of enhancers (Fig. 2a and Extended Data Fig. 3a) that accounted for ≥85% of the variance in the data (Extended Data Fig. 3b) and were reproducible with alternative clustering approaches (Extended Data Fig. 3c–e). The first two groups required all five COFs for full activity, whereby group 1 was more strongly dependent on P300/CBP and group 2 on CDK7 (Fig. 2a,b). Notably, the enhancers of groups 3 and 4 were not impaired by the degradation of MED14 or BRD4, respectively, thereby defining enhancer types that can function with limiting levels, or potentially entirely independently, of these two COFs (Fig. 2a, b).

Endogenous enhancer chromatin features in HCT116 cells were enriched in all four groups of enhancers compared with random control regions, including DNA accessibility, H3K27ac, H3K4me1 and COF binding (Fig. 2c; see Methods for published data sources). However, the groups differed in relative levels of chromatin marks and in genomic

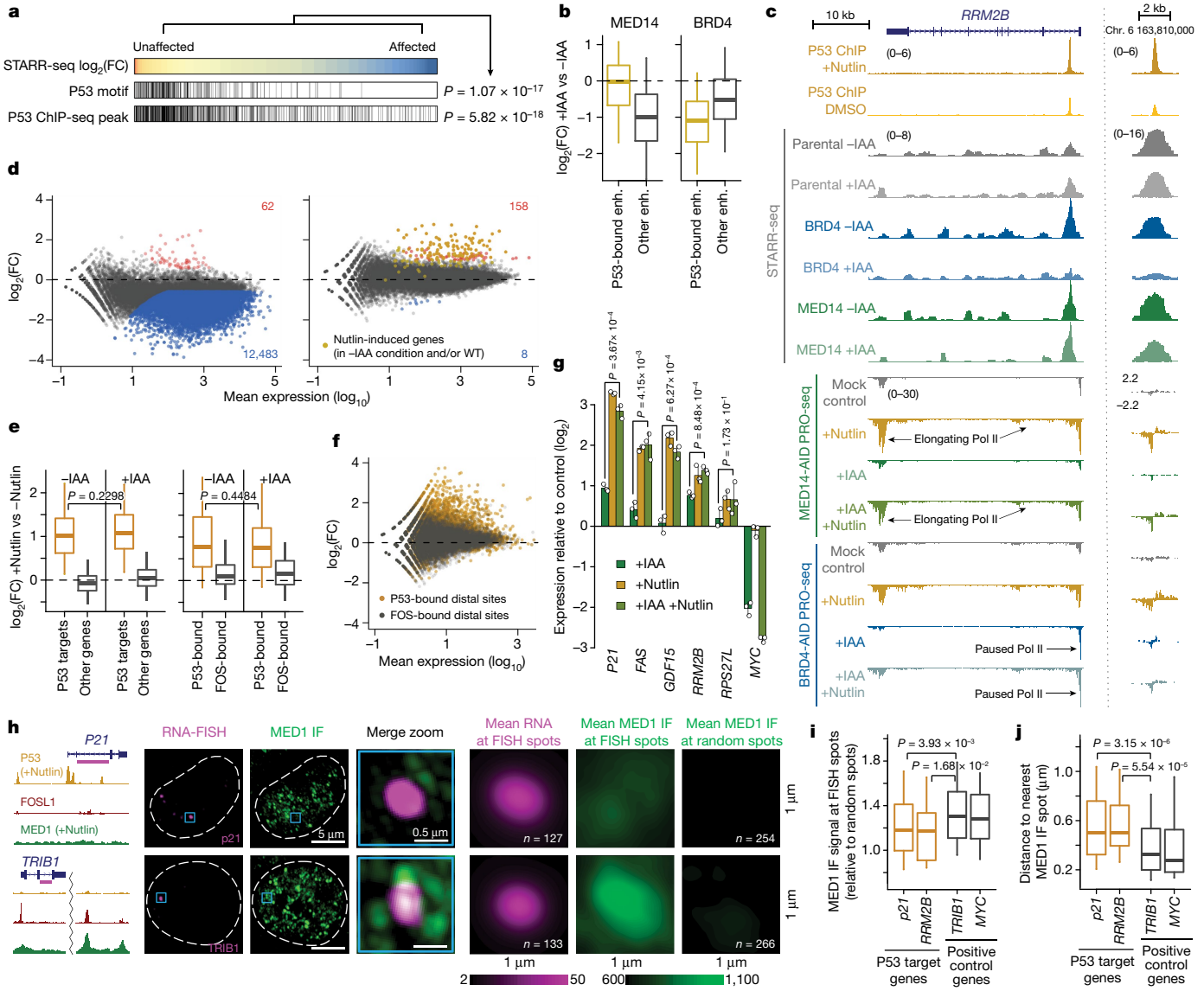


Fig. 3 | P53 target genes and enhancers are insensitive to MED14 depletion. **a**, P53 motifs and ChIP-seq peaks in STARR-seq enhancers sorted by level of downregulation after MED14 depletion. *P* values calculated using one-sided Fisher's exact test (top against bottom 20%). **b**, STARR-seq activity change for P53-bound ($n = 621$) versus other ($n = 5628$) enhancers (enh.) in MED14-AID and BRD4-AID cells. **c**, Enhancer activity (merged STARR-seq replicates) and nascent transcription (merged PRO-seq replicates) in the *RRM2B* locus after P53 induction with Nutlin-3a in MED14-AID cells and BRD4-AID cells with and without IAA treatment. **d**, Differential gene PRO-seq in MED14-AID cells (left, ±IAA; right, +IAA + Nutlin-3a versus +IAA only; FDR ≤ 0.05; FC ≥ 2; $n = 2$ independent replicates; yellow: 151 Nutlin-3a-induced genes from unperturbed cells (Extended Data Fig. 4b)). **e**, PRO-seq FC for P53 target genes (left; $n = 243$ (Extended Data Fig. 4c)) and distal TF P53-bound sites (right) in MED14-AID cells after Nutlin-3a ± IAA treatment. $n = 243$; 20,964; 233 and 346 for P53 targets, other genes, P53-bound enhancers and FOS-bound enhancers, respectively. **f**, Differential PRO-seq for distal P53-bound or FOS-bound enhancers after Nutlin-3a in IAA-treated MED14-AID cells. **g**, Expression (qPCR) of P53 targets in IAA-treated and/or Nutlin-3a-treated MED14-AID cells. $n = 3$ independent replicates; mean ± s.d.; *P* values calculated using two-sided Student's *t*-test. **h**, MED1 immunofluorescence (IF) with concurrent RNA FISH against P53-target *P21* (top) and control *TRIB1* gene (bottom) in Nutlin-3a-treated HCT116 cells. Left, gene loci with P53, FOSL1 and MED1 ChIP-seq signal and intronic FISH-target sequence (magenta). Dashed lines indicate nuclear periphery. Right, mean RNA FISH and MED1 IF signals centred on FISH spots or random spots (n = number of spots). **i**, MED1 IF signal at FISH spots, normalized to the mean MED1 IF signal at random spots. **j**, Distance between FISH spot and nearest MED1 IF spot. In **i** and **j**, $n = 127$, 50, 133 and 118 FISH spots for P21, RRM2B, TRIB1 and MYC, respectively. In **b**, **e**, **i** and **j**, boxes indicate the median and interquartile range, whiskers the 5th and 95th percentiles; *P* values calculated using two-sided Wilcoxon rank-sum test.

localization (Fig. 2d). Group 1 contained the highest proportion of endogenously accessible enhancers (open across many cell types; Extended Data Fig. 3f,g) and were most highly enriched for H3K27ac and H3Kme1 (Fig. 2c,d). By contrast, group 2 enhancers were subtly enriched for H3K36me3, a gene-body mark, and intragenic localization (Fig. 2c,d). Groups 3 and 4 contained enhancers accessible in HCT116 cells and enhancers accessible only in other cell types (Extended Data Fig. 3f), which is indicative of chromatin-mediated silencing in HCT116 cells[6]. Indeed, both groups displayed a relative enrichment of repressive H3K27me3 (group 4) and H3K9me2 and H3K9me3 marks (group 3) (Fig. 2d).

The four groups most notably differed in their sequences and contained specific TF motifs. Group 1 enhancers were highly enriched for the AP-1 family (FOS and JUN) motifs and their combinations (Fig. 2e and Extended Data Fig. 3h), whereas group 3 enhancers were most strongly enriched for P53 (also known as TP53) motifs, and group 4 enhancers

for NFY (CCAAT box) motifs. Published chromatin immunoprecipitation with sequencing (ChIP-seq) datasets confirmed the preferential binding of these TFs to endogenous enhancers of the different groups (Fig. 2e), which suggests that transactivation by different TFs requires different sets of COFs.

Mediator independence of P53 targets

The finding that enhancers characterized by P53 motifs and endogenous p53 binding are insensitive to MED14 depletion (Fig. 2a,b,e) suggests that P53-mediated activation might be Mediator-independent. This is consistent with reports that some active or stress-inducible promoters do not associate with Mediator in yeast[24]. However, it is also unexpected, as P53 directly interacts with Mediator[7,25,26], and most activators of stress-responsive genes recruit Mediator[24].

We first confirmed that P53 motifs and P53 binding[27] are the most strongly enriched in enhancers that show the least dependence on MED14 (Fig. 3a), whereas motifs for FOS and JUN, for example, were enriched in MED14-dependent enhancers (Extended Data Fig. 3i,j). Consistently, MED14 depletion did not affect P53-bound enhancers, whereas the activity of enhancers not bound by P53 decreased on average by about twofold (Fig. 3b and Extended Data Fig. 3k). This difference was specific to MED14 depletion, whereas, for example, BRD4 depletion reduced enhancer activity irrespective of P53 binding (Fig. 3b and Extended Data Fig. 3k), as exemplified by an enhancer in the first intron of the P53 target gene *RRM2B*, which was strongly affected by depletion of BRD4 but not MED14 (Fig. 3c).

We next assayed the transcriptional response of endogenous P53 target genes using PRO-seq after depleting MED14. IAA treatment for 3 h led to global transcriptional downregulation of almost all genes (Fig. 3d, left), which is consistent with the dependence on Mediator of most enhancers in HCT116 cells (Fig. 2a) and confirms effective depletion of Mediator. However, when we treated MED14-depleted cells with the small-molecule Nutlin-3a, which activates P53 signalling[27], the transcriptional response was essentially identical as in MED14-non-depleted cells and in wild-type (WT) HCT116 cells (Fig. 3d, right, and Extended Data Fig. 4a,b). Indeed, direct P53 target genes activated by Nutlin-3a treatment in WT HCT116 cells (Extended Data Fig. 4c) were upregulated to the same extent in both MED14-depleted and control cells, including the well-known P53 targets *FAS*, *RPS27L* and *RRM2B* (Fig. 3c, e, left, and Extended Data Fig. 4d,e). Consistent with the induction of P53 target genes, we also observed the specific upregulation of nascent bidirectional transcription from p53-bound enhancers in the vicinity of those genes (Fig. 3c, right, and Extended Data Fig. 4f) to the same extent in both MED14-depleted and control cells (Fig. 3e, right). This result confirms that the endogenous enhancers are activated despite MED14 depletion (Fig. 3f). In addition, we confirmed the induction at the mature mRNA level for several well-known P53 targets, including *P21* (also known as *CDKN1A*) through quantitative PCR (qPCR) (Fig. 3g). After MED14 depletion, Nutlin-3a treatment led to an induction of all assayed P53 targets to similar final levels as without depletion, whereas the transcription of Mediator-dependent control genes, including *MYC*, was impaired.

In contrast to MED14 depletion, BRD4 depletion significantly reduced the induction of both P53 target genes and p53-bound enhancers, as measured by PRO-seq and qPCR (Fig. 3c and Extended Data Fig. 4a,d–i), which demonstrates that unlike MED14, BRD4 is required for a strong P53 response. Furthermore, degradation of either TAF1 or CDK9 completely abolished the induction of p53 target genes (Extended Data Fig. 4j–l), which indicates that P53-mediated activation depends on functioning initiation and pause-release steps, both of which seem to occur in MED14-depleted cells.

Taken together, these results show that P53-mediated activation is insensitive to limiting levels of MED14. This is consistent with either P53 target enhancers being highly efficient in recruiting residual MED14 (Extended Data Fig. 1d) or functioning independently of MED14 through non-canonical Mediator subcomplexes, which presumably contain MED1 or MED17 that can directly interact with P53 (refs. [25,26,28]). To discern between these possibilities, we performed MED1 ChIP-seq in MED14-AID and in WT HCT116 cells after IAA and/or Nutlin-3a treatment. In unperturbed cells, MED1 bound to many endogenously active enhancers, including a previously described enhancer cluster at the *MYC* locus (Extended Data Fig. 5a–c). MED1 ChIP signals were increased at endogenous MED14-dependent enhancers compared with MED14-independent enhancers, and the majority, including those in the *MYC* locus and at MED14-dependent enhancers, were lost following MED14 depletion (Extended Data Fig. 5d,e). Thus, Mediator-dependent enhancers bind detectable levels of Mediator, which is effectively depleted by MED14 degradation. By contrast, we did not detect MED1 ChIP-seq signals at P53 target enhancers in any condition, which suggests that these enhancers do not recruit high levels of MED1, at least not like MED14-dependent enhancers (for example, MYC enhancers; Extended Data Fig. 5e).

To assess Mediator binding to P53 target genes (*P21* and *RRM2B*) and Mediator-dependent control genes (*TRIB1* and *MYC*) by an independent approach, we combined MED1 immunofluorescence (IF) with RNA fluorescence in situ hybridization (FISH) against nascent transcripts in WT HCT116 cells treated with Nutlin-3a for 3 h. In this condition, the gene loci of both groups of genes were strongly detected by FISH, which enabled the quantification of MED1 IF signals at 127 *P21* and 133 *TRIB1* gene loci (Fig. 3h; see Extended Data Fig. 5f for *RRM2B* and *MYC*). Consistent with the ChIP-seq data, the MED1 signal at individual gene loci was significantly lower for P53 target genes than controls (Fig. 3h,i). Moreover, MED1 spots were significantly farther from P53 target genes than from controls (Fig. 3j), which was not due to overall differences in the number of MED1 spots (Extended Data Fig. 5g). This result demonstrates that P53 target genes do not recruit substantial amounts of MED1 and suggests that P53-mediated activation does not require the full or canonical Mediator complex that contains MED14 and MED1 (ref. [7]).

To assess whether the P53 response is independent of additional Mediator subunits, we measured the induction of known p53 target genes by qPCR in cells depleted of different Mediator subunits from the head, tail and middle modules, including the two subunits previously reported to interact with p53, MED1 and MED17 (refs. [25,26,28]). Depletion of all targeted subunits by AID or small interfering RNAs (siRNAs) had no effect on P53 target gene induction, which was the same as in unperturbed cells (Extended Data Fig. 6a–d). To extend our findings to another cell type and organism and to cells that are permanently devoid of non-essential Mediator subunits, we chose knockout (KO) mouse lymphoma CH12 cells, lacking the MED1, MED19, MED20, MED26 or MED29 Mediator subunit, or the entire Mediator tail (MED15, MED16, MED23, MED24 and MED25)[18]. The known P53 target genes *P21*, *Fas* and *Rrm2b* were induced in all KO cells, including cells lacking the P53 interacting subunit MED1 (MED17 is essential and could not be tested; Extended Data Fig. 6e). Only the MED19-KO and tailless cells had undetectable levels of *P21* in all conditions, which was potentially a result of clonal selection, but both strongly induced *Fas* and *Rrm2b*.

Overall, the results regarding enhancer activities and nascent transcription after MED14 depletion, the lack of detectable MED1 binding and the dispensability of various Mediator subunits for P53 targets in human and mouse cells suggest that P53-mediated transcription activation is independent of full or canonical Mediator[7] (Discussion).

TATA boxes confer BRD4 independence

Group 4 enhancers remained active or even increased in activity in the absence of BRD4 (Fig. 2a), and were often associated with closed chromatin, repressive histone marks (Fig. 2d) and individual repeat elements (Fig. 4a). In particular, the long terminal repeat families LTR12C

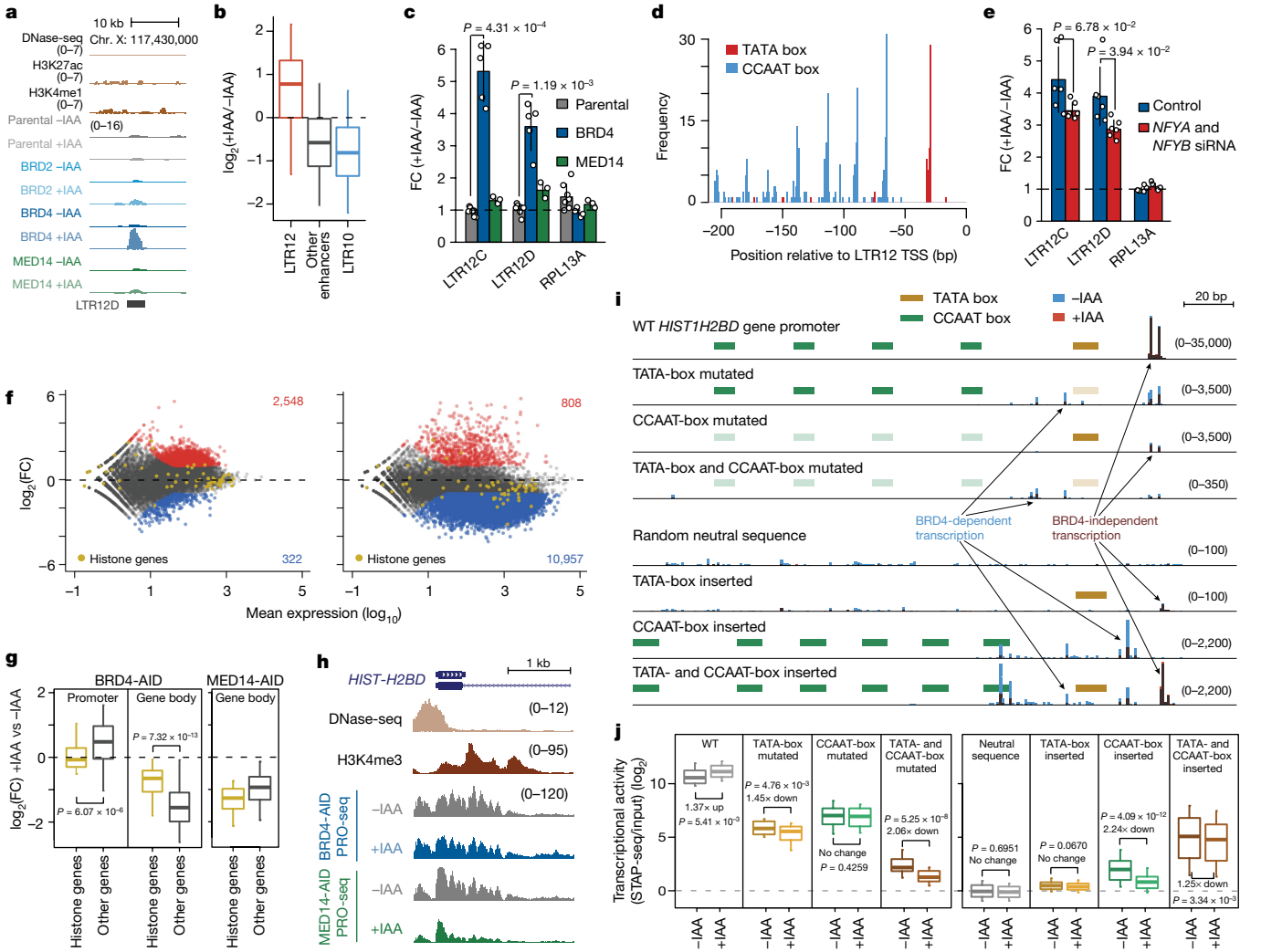


Fig. 4 | Combination of the TATA box and the CCAAT box renders the transcription of LTR12 retrotransposons and histone genes independent of BRD4. a, LTR12D element with increased enhancer activity after BRD4 degradation. **b**, Enhancer activity change after BRD4 depletion for LTR12-overlapping ($n = 117$), LTR10-overlapping ($n = 198$) and all other ($n = 5{,}935$) enhancers. **c**, Change in endogenous LTR12 expression (qPCR) after IAA treatment of BRD4-AID and MED14-AID cells. $n = 7$, 5 and 3 independent replicates for parental, BRD4-AID and MED14-AID cells, respectively. **d**, Occurrence of TATA and CCAAT boxes in LTR12 repeats with STARR-seq activity, relative to their endogenous TSSs. **e**, Change in endogenous LTR12 expression (qPCR) after BRD4 depletion before and after NFYA and NFYB knockdown. $n = 6$ independent replicates. **f**, Differential analysis (±IAA) of PRO-seq in the promoter-pause region (+1 to +150 bp; left) and the gene body (+150 bp to gene end; right) for BRD4-AID cells (FDR ≤ 0.05; FC ≥ 2; yellow indicates histone genes; $n = 2$ independent replicates). **g**, Change in PRO-seq signal in the promoter-pause region and the gene body in BRD4-AID cells (left) and the gene body in MED14-AID cells (right) for histone genes ($n = 50$) versus all other expressed genes ($n = 11{,}869$). **h**, PRO-seq signal at *HIST1H2BD* in BRD4-AID and MED14-AID cells ±IAA (normalized signal for merged replicates). **i**, Transcription (base-pair resolution; Extended Data Fig. 9b) from WT and mutated *HIST1H2BD* promoters (top) and from neutral sequences with inserted LTR12-derived TATA and/or CCAAT boxes (bottom). The mean normalized STAP-seq signal across barcodes and replicates ($n = 2$ independent replicates, 5 barcodes per sequence) in +IAA (red) versus –IAA (blue) BRD4-AID cells is overlaid. **j**, STAP-seq signal for WT and mutated versions of histone and LTR12 promoters (left; $n = 50$) and for random neutral sequences with inserted TATA and/or CCAAT boxes (right; $n = 90$, 120 and 900 for WT, single insertions and double insertions, respectively). In **b**, **g**, **j**, boxes indicate the median and interquartile range, whiskers the 5th and 95th percentiles; *P* values calculated using two-sided Wilcoxon rank-sum test. In **c** and **e**, mean ± s.d.; *P* values calculated using two-sided Student's *t*-test.

and LTR12D were enriched in upregulated enhancers (Extended Data Fig. 7a), and LTR12 elements detected in STARR-seq displayed strongly increased enhancer activity after BRD4 depletion, unlike the related LTR10 elements and most enhancers that generally lost activity (Fig. 4b and Extended Data Fig. 7b). Furthermore, endogenous LTR12C and LTR12D were strongly upregulated (per qPCR analysis) after prolonged BRD4 degradation, which is consistent with effects of inhibiting histone deacetylases[29,30], but not after MED14 depletion (Fig. 4c). This upregulation also occurred in K562 and A549 cells after BRD4 depletion (Extended Data Fig. 7c).

LTR12 elements contain a TATA box promoter and multiple CCAAT boxes[29,30] (Fig. 4d and Extended Data Fig. 7d), which were also the most highly enriched motifs in BRD4-independent enhancers (Fig. 2e) and in enhancers upregulated following BRD4 depletion (Extended Data Fig. 7e). As CCAAT boxes in LTR12 bind the NFY TFs[30], which maintain nucleosomal-depleted regions[31], we tested whether NFY is required for LTR12 expression by depleting the NFY subunits A and B through RNA interference (RNAi) in BRD4-depleted HCT116 cells (Extended Data Fig. 7f–h) and A549 cells (Extended Data Fig. 7i–k). NFYA and NFYB depletion significantly reduced the upregulation of LTR12C and LTR12D

after BRD4 depletion in both cell types (Fig. 4e and Extended Data Fig. 7h,j). Thus, NFY contributes to the upregulation of LTR12C and LTR12D following BRD4 loss and is potentially involved in the mechanism that confers BRD4 independence.

Gene ontology analysis for genes with a CCAAT box and a TATA box promoter structure revealed terms related to nucleosome assembly and DNA packaging (Extended Data Fig. 8a), and identified histone genes as top hits. Indeed, promoters of histone genes have a precisely positioned TATA box and proximal upstream CCAAT boxes (Extended Data Fig. 8b). To test whether histone genes are transcribed in the absence of BRD4, we performed PRO-seq after BRD4 depletion. Consistent with the function of BRD4 in pause-release and in line with previous reports[32,33], BRD4 depletion led to a global pause-release defect characterized by the loss of the Pol II signal in gene bodies and a gain in the promoter-proximal pause region (Fig. 4f). However, histone genes were much less affected compared with other genes after BRD4 depletion and with histone genes after MED14 depletion (Fig. 4g,h and Extended Data Fig. 8c). This result suggests that histone gene transcription is independent of BRD4 but dependent on MED14. Indeed, a re-analysis of published datasets using nascent transcription after BRD4 inhibition or degradation[32,33] confirmed that transcription of histone genes occurs independent of BRD4 (Extended Data Fig. 8d).

The above results suggest that LTR12 elements and histone gene promoters contain TATA-box-compatible enhancers that can activate the heterologous TATA box promoter in STARR-seq and their cognate TATA box promoters in vivo in a BRD4-independent manner. The elements are also orientation-independent in STARR-seq as expected for bona fide enhancers (Extended Data Fig. 8e,f). To examine whether there is a functional link between TATA and CCAAT boxes and BRD4-independent transcription, we made use of the fact that these elements function as autonomous promoters and assessed the transcriptional activity of hundreds of WT and mutated sequences in BRD4-AID cells with or without IAA treatment (Extended Data Fig. 9a,b). To this end, we used a massively parallel reporter assay with single base-pair resolution[34] with a synthetic oligonucleotide library comprising 240-bp-long fragments, each with five unique barcodes. To test the necessity of motifs, we selected ten BRD4-independent promoters, including LTR12 elements and histone gene promoters, and generated WT sequences and variants that were mutant for either TATA or CCAAT boxes or both (Extended Data Fig. 9a). To test for motif sufficiency, we inserted the TATA and/or CCAAT boxes into 18 different transcriptionally inactive random sequences, preserving the arrangement of these motifs in BRD4-independent promoters.

This resulted in highly reproducible transcriptional activities and initiation patterns (Extended Data Fig. 9b,c) that confirmed BRD4-independent transcription of histone gene promoters and LTR12 elements (Fig. 4i,j and Extended Data Fig. 9c). Mutations in TATA boxes impaired transcription from the cognate transcription start site (TSS) and BRD4-independence, as seen by a further reduction in transcription following IAA treatment. By contrast, mutations in CCAAT boxes resulted in a strong loss of transcription, but the remaining transcription was still BRD4-independent. Mutations in both motifs further reduced transcriptional activity, and any remaining transcription was strongly BRD4-dependent (Fig. 4i,j and Extended Data Fig. 9c).

Consistently, inserting a TATA box into inactive sequences resulted in low levels of BRD4-independent transcription from a single TSS (Fig. 4i,j), which is in line with observations that TATA boxes on their own support only low levels of transcription[34]. Inserting only CCAAT boxes increased transcription from dispersed ectopic initiation sites, and this transcription was highly dependent on BRD4. Inserting both motifs together resulted in strong transcription from a single TSS that was less dependent on BRD4 and to varying levels of BRD4-dependent transcription from ectopic sites (Fig. 4i,j and Extended Data Fig. 9d).

Taken together, these results demonstrate that a TATA box promoter is necessary and sufficient to confer BRD4 independence, whereas CCAAT boxes act as enhancers to boost BRD4-independent transcription but cannot themselves confer BRD4 independence. As STARR-seq uses a promoter with mixed features and multiple TSSs[6], we speculate that BRD4-independent enhancers activate TATA-box-associated TSSs, whereas BRD4-dependent enhancers are presumably not compatible with the TATA box and activate other TSSs within the same promoter.

To further investigate the role of TATA boxes in conferring BRD4 independence, we analysed heat-shock genes, which are well-studied models of TATA box promoters and proximally bound activators[35]. In brief, we induced heat shock for 1 h at 43 °C in BRD4-AID cells pretreated with water (mock) or IAA and analysed the expression of four heat-shock genes by qPCR. In three different cell lines, all tested genes were strongly induced after heat shock irrespective of BRD4 depletion[36] (Extended Data Fig. 9e), whereas CDK9 depletion abolished gene induction as expected (Extended Data Fig. 9f). This dependence on CDK9 but not on BRD4 suggests that the CDK9-containing complex pTEFb is recruited by other means, presumably by the super elongation complex (SEC) that functions at stress-related genes[37]. Indeed, the simultaneous depletion of the two SEC subunits AFF1 and AFF4 led to a mild but significant reduction in heat-shock gene induction (Extended Data Fig. 9g). This provides evidence that SEC might aid in the recruitment of CDK9 to support full inducibility of heat-shock genes independently of BRD4 (ref. [36]).

Taken together, our data show that transcription from TATA box promoters is insensitive to BRD4 depletion and allows BRD4-independent transcription of different types of genes through different TATA-box-compatible enhancers. Thus, specific classes of genes and their associated enhancers have distinct COF requirements and can function independently of broadly deployed COFs, possibly through alternative mechanisms, to regulate specific steps in transcription.

Discussion

Here we reported on distinct enhancer types with different COF dependencies that further differ in TF binding, chromatin modifications, genomic localization and the transcriptional response of nearby genes to COF depletion (Extended Data Fig. 9h, i). We anticipate that enhancer classifications will be refined when additional COFs are considered. However, when we AID-tagged and depleted three additional COFs (BRD7, BRD9 and MLL1; Extended Data Fig. 10a), STARR-seq with a focused library covering about 0.4% of the human genome (11.7 Mb) did not reveal any changes in enhancer activity (Extended Data Fig. 10b,c). In steady-state HCT116 cells, these factors might act redundantly with others or could only be required following stimuli[38] or during cellular transitions[39].

The results for MED14 suggest that P53-mediated transcription might be independent of the Mediator complex, a finding that is difficult or impossible to formally prove given the essentiality of Mediator. That is, residual MED14 or partial Mediator complexes may enable activation of P53 target genes in MED14-depleted cells. Although selective rescue of P53 targets by residual MED14 seems less probable given that Mediator does not preferentially localize to these genes in any condition (Fig. 3h and Extended Data Fig. 5d–f), diverse Mediator subcomplexes exist in yeast[40] and in humans[21,41] and could be recruited, for example, through MED17 and P53 interactions[26]. Although the depletion of individual Mediator subunits by AID (four subunits), RNAi (MED17) or genetic depletion in stable KO cells[18] (five subunits) and the combined depletion of five Mediator tail subunits in stable KO cells did not impair P53 target gene transcription (Extended Data Fig. 6a–e), it is possible that these subunits function partially redundantly or in subcomplexes of variable composition. Redundancy between Mediator subunits has been observed in yeast[42–44], and stable partial human Mediator complexes could be reconstituted[21,41], including a Mediator head and middle module that included MED17 but not MED14 (ref. [21]). Alternatively, P53 targets might require levels of Mediator below the detection limits of this study, or other factors

and conditions such as high local Pol II concentrations[45], Mediator bypass via BRD4 and/or CDK9 (which are both required), or compensation by mobilized CDK9 (ref. [32]) might partially substitute for Mediator function at these genes. Finally, Pol II may initiate at these promoters through different mechanisms with distinct rate-limiting steps, potentially involving preinitiation complexes with different protein composition[46].

The finding that TATA boxes can confer BRD4-independence to LTR12 repeats, histone genes and heat-shock genes, a classical model of TATA box promoter genes regulated primarily at the pause-release step, suggests that there are alternative mechanisms to recruit CDK9, for example through the SEC complex[47,48] or TFs[49]. Notably, many enhancers required either MED14 or BRD4 (Figs. 3b and 2a, compare groups 3 and 4). As MED14 and BRD4 function mainly in initiation or pause-release, respectively, groups 2 and 3 enhancers might regulate distinct steps of transcription. The fact that both Mediator-independent and BRD4-independent enhancers relate to genes activated following stress suggests that rapidly inducible genes might have exploited this concept by circumventing certain regulatory steps (regulatory shortcutting) or by overcoming particular steps before actual induction (regulatory priming). Priming and regulation at the pause-release step is, for instance, well known for heat-shock-inducible genes[50].

Together with the recent finding that promoters show distinct compatibilities towards different enhancers and specific COFs[51], our results that enhancers differ consistently in their COF dependencies and that gene regulatory programmes differentially utilize these enhancer types is an important step towards understanding gene-regulatory specificities and determining innovative targets for the precise modulation of gene expression.

Online content

Any methods, additional references, Nature Research 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/s41586-022-04779-x.

1. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43**, 73–81 (2017).
2. Nakagawa, T., Yoneda, M., Higashi, M., Ohkuma, Y. & Ito, T. Enhancer function regulated by combinations of transcription factors and cofactors. *Genes Cells* **23**, 808–821 (2018).
3. Rathert, P. et al. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543–547 (2015).
4. Jaeger, M. G. et al. Selective mediator dependence of cell-type-specifying transcription. *Nat. Genet.* **52**, 719–727 (2020).
5. Chipumuro, E. et al. CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126–1139 (2014).
6. Muerdter, F. et al. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods* **15**, 141–149 (2018).
7. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
8. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
9. Vo, N. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. *J. Biol. Chem.* **276**, 13505–13508 (2001).
10. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, R106 (2017).
11. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
12. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
13. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
14. Krebs, A. R., Karmodiya, K., Lindahl-Allen, M., Struhl, K. & Tora, L. SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. *Mol. Cell* **44**, 410–423 (2011).
15. Zuber, J. et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524–528 (2011).
16. Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
17. Pelish, H. E. et al. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **526**, 273–276 (2015).
18. El Khattabi, L. et al. A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers. *Cell* **178**, 1145–1158.e20 (2019).
19. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).
20. Watanabe, Y. et al. Frequent alteration of MLL3 frameshift mutations in microsatellite deficient colorectal cancer. *PLoS ONE* **6**, e23320 (2011).
21. Cevher, M. A. et al. Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat. Struct. Mol. Biol.* **21**, 1028–1034 (2014).
22. Liang, J. et al. CDK8 selectively promotes the growth of colon cancer metastases in the liver by regulating gene expression of TIMP3 and matrix metalloproteinases. *Cancer Res.* **78**, 6594–6606 (2018).
23. Hu, D. et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.* **33**, 4745–4754 (2013).
24. Fan, X., Chou, D. M. & Struhl, K. Activator-specific recruitment of Mediator in vivo. *Nat. Struct. Mol. Biol.* **13**, 117–120 (2006).
25. Meyer, K. D., Lin, S.-C., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nat. Struct. Mol. Biol.* **17**, 753–760 (2010).
26. Ito, M. et al. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol. Cell* **3**, 361–370 (1999).
27. Andrysik, Z. et al. Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. *Genome Res.* **27**, 1645–1657 (2017).
28. Drané, P., Barel, M., Balbo, M. & Frade, R. Identification of RB18A, a 205 kDa new p53 regulatory protein which shares antigenic and functional properties with p53. *Oncogene* **15**, 3013–3024 (1997).
29. Brocks, D. et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* **49**, 1052–1060 (2017).
30. Krönung, S. K. et al. LTR12 promoter activation in a broad range of human tumor cells by HDAC inhibition. *Oncotarget* **7**, 33484–33497 (2016).
31. Oldfield, A. J. et al. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. *Nat. Commun.* **10**, 3072 (2019).
32. Winter, G. E. et al. BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol Cell* **67**, 5–18.e19 (2017).
33. Muhar, M. et al. SLAM-seq defines direct gene-regulatory functions of the BRD4–MYC axis. *Science* **360**, 800–805 (2018).
34. Arnold, C. D. et al. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat. Biotechnol.* **35**, 136–144 (2017).
35. Lis, J. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 347–356 (1998).
36. Zheng, B. et al. Acute perturbation strategies in interrogating RNA polymerase II elongation factor function in gene expression. *Genes Dev.* **35**, 273–285 (2021).
37. Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018).
38. Galbraith, M. D. et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* **153**, 1327–1339 (2013).
39. Kubo, N., Hu, R., Ye, Z. & Ren, B. MLL3/MLL4 histone methyltranferase activity dependent chromatin organization at enhancers during embryonic stem cell differentiation. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.17.435905 (2021).
40. Kang, J. S. et al. The structural and functional organization of the yeast Mediator complex. *J. Biol. Chem.* **276**, 42003–42010 (2001).
41. Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C. & Cramer, P. Structure of the human Mediator–RNA polymerase II pre-initiation complex. *Nature* **594**, 129–133 (2021).
42. Lee, D., Kim, S. & Lis, J. T. Different upstream transcriptional activators have distinct coactivator requirements. *Gene Dev.* **13**, 2934–2939 (1999).
43. Petrenko, N., Jin, Y., Wong, K. H. & Struhl, K. Evidence that Mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. *eLife* **6**, 155 (2017).
44. Anandhakumar, J., Moustafa, Y. W., Chowdhary, S., Kainth, A. S. & Gross, D. S. Evidence for multiple Mediator complexes in yeast independently recruited by activated heat shock factor. *Mol. Cell. Biol.* **36**, 1943–1960 (2016).
45. Cho, W.-K. et al. RNA polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, 1123 (2016).
46. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C. & Tjian, R. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420**, 439–445 (2002).
47. Lin, C. et al. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol. Cell* **37**, 429–437 (2010).
48. Lin, C. et al. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev.* **25**, 1486–1498 (2011).
49. Bugai, A. et al. P-TEFb activation by RBM7 shapes a pro-survival transcriptional response to genotoxic stress. *Mol. Cell* **74**, 254–267.e10 (2019).
50. Lis, J. T., Mason, P., Peng, J., Price, D. H. & Werner, J. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**, 792–803 (2000).
51. Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).

Methods

Cell culture

HCT116 cells were purchased from the American Type Culture Collection (ATCC, CCL-247) and cultured in DMEM with 10% heat-inactivated FCS (SigmaAldrich, F7524) and 1% L-glutamine (LifeTech Austria/Invitrogen, 25030024). HCT116 cells are near-diploid, chromosomally stable (P53 WT) and do not elicit interferon responses after reporter plasmid transfection[6]. For proliferation assays, cells were seeded into 6-well plates with 2×10^5 cells per well as a starting seeding density with or without the addition of indole-3-acetic acid sodium salt (IAA/auxin, SigmaAldrich, I5148-2G) 500 µM final concentration. For up to five consecutive days, cells were counted (Countess II Thermo Fisher, AMQAX1000) in 24 h intervals. K562 BRD4-AID cells were obtained from ref. [33]. and cultured in RPMI-1640 with 10% FCS (SigmaAldrich, F7524). CH12 mouse lymphoma cell lines (WT and KO for different Mediator subunits) were obtained from ref. [18]. and were cultured in RPMI-1640 with 10% FCS (SigmaAldrich, F7524), 1% penicillin–streptomycin and 50 µM of β-mercaptoethanol (Thermo Fisher Scientific). All cell lines tested negative for mycoplasma.

Cloning and characterization of genome-editing events

The SpCas9 knock-in homology-dependent recombination strategy and cloning of vectors were performed based on ref. [33]. The parental cell line was generated through the insertion of the knock-in cassette '500 bp 5′HA-mCherry-P2A-OsTir1-3×Myc-500 bp 3′HA' downstream of the *Actinb* gene. A stretch of 500 bp homology arms (HAs) flanking the regions upstream and downstream of the *Actinb* stop codon were obtained by PCR on human genomic DNA (Promega, G304A). A total of 20 µg of the knock-in cassette (cloned into a MCS of a pbluescript vector) and the lentiCRISPR v2 vector comprising SpCas9 and gRNA (Addgene plasmid, 52961) against the *Actinb* stop codon were electroporated (at equimolar concentrations) into 5×10^6 HCT116 cells using a Maxcyte STX electroporation device (GOC1). After 25 min of a recovery phase, medium was added and cells were grown for 3 days. Afterwards, cells were single-cell sorted on the basis of the mCherry signal (approximately 0.5–1% of total population). After 14 days, outgrowing clones were lysed (Biozym, 101094) and genotyped, and potential knock-in candidates were further validated by western blotting against 3×Myc tag (Merck, 05-724). Within an established *O. sativa* Tir1 (OsTir1) heterozygote-tagged parental clone (*Ostir*$^{+/-}$), tagging of individual COFs with the AID system was performed. IAA-inducible destabilization domain constructs were cloned into lentiviral vector (Addgene plasmid, 14748)[33] for either amino-terminal COF tagging '5′HA-blasticidin-P2A-V5-AID-spacer-3′HA' or carboxy-terminal COF tagging '5′HA-spacer-AID-V5-P2A-blasticidin-3′HA'. N-terminal or C-terminal tagging constructs were electroporated with the lentiCRISPR v2 containing gRNA against individual COFS with Maxcyte STX. After 25 min of recovery at 37 °C, medium (DMEM with 10% FCS and 1% L-glutamine) was added, and cells were grown for 3 days. Cells were trypsinized, transferred (1×10^6) into 6-well plates and selected for 10 days on blasticidin (10 µg ml^{-1}; eubio, ant-bl-10p). Outgrowing colonies were collected and single-cell sorted for mCherry and against GFP. As described in ref. [33], the Addgene plasmid no. 4748 construct expresses a constitutively active GFP, which enabled negative FACS selection against potential vector backbone integrations. After 14 days, grown out colonies were individually collected, lysed with DNA extraction solution (Biozym, 101094) and genotyped by Sanger sequencing. Potential candidates were investigated by western blotting against the integrated V5-tag (Thermo Fisher, R960-25) or antibodies against endogenous proteins (Supplementary Table 1).

PITCh knock-in HCT116 cells

Cloning of PITCh vectors was based on ref. [52]. pX330S-2-PITCh (Addgene, plasmid no. 63670) containing PITCh gRNA was cloned using Golden Gate assembly into the pX330A-1×2 vector (Addgene, plasmid no. 58766), which expresses Cas9 and the gRNA against a target locus. Knock-in cassettes flanked by 40 bp microhomology arms were cloned into the pCRIS-PITChv2-FBL vector (Addgene, plasmid no. 63672). A total of 20 µg (13 µg pX330A-1×2 and 7 µg pCRIS-PITChv2-FBL) was electroporated into 5×10^6 cells using Maxcyte STX. Follow-up steps were similarly performed as described in the previous section 'Cloning and characterization of genome-editing events'.

Western blotting

Cells (1×10^6) were collected, centrifuged at 300*g* for 5 min, washed with 1× PBS and lysed in 75 µl RIPA buffer containing protease inhibitor (Roche, 11836170001). For complete lysis, cells were incubated on ice for 30 min, sonicated four times for 30 s each with a sonicator (Diagenode Bioruptor) and treated with 1 µl benzonase endonuclease (Sigma Aldrich, E1014-5KU) for 30 min to solubilize the chromatin-bound proteins. Afterwards, samples were centrifugated for 10 min at 12,000 r.c.f. and 4 °C, after which 40 µl 2× Laemmli buffer (Bio-Rad, 1610737) was added. Samples were vortexed, boiled for 5 min at 95 °C and centrifuged for 2 min at 12,000 r.c.f. Next, samples and marker (Invitrogen, LC5602) were loaded on the protein gel (Bio-Rad, 4561083) using 1× SDS running buffer with 120 V for 1 h and 20 min. Separated proteins were transferred by wet-transfer (Bio-Rad, 1703930) onto a methanol-activated membrane (Millipore, PVDF, 0.45 µm, IPFL00010) with a transfer time of 1 h at 100 V. After transfer, the membrane was incubated for 10 min with TBST and blocked for 30 min in TBST + 5% milk (Bio-Rad, 1706404) on a rotating platform at room temperature. Next, the membrane was incubated in TBST + 5% milk comprising the primary antibody (Supplementary Table 1) overnight at 4 °C. After overnight incubation, the membrane was washed three times with TBST for 15 min and incubated with secondary antibody (Supplementary Table 1) for 2 h on a rotating platform at room temperature. Last, the membrane was washed three times for 15 min in TBST before protein visualization by ECL detection (ChemiDOC Imager, Bio-Rad, 170-5060).

MS analysis of COF-depleted cell nuclei

Cells (1×10^6) were treated with water (mock) or 500 µM IAA for 1 or 3 h. Afterwards, cells were collected with 1× trypsin, washed with 1× PBS and centrifuged for 3 min at room temperature at 500*g*. The supernatant was removed and the cell pellet was resuspended in about 100 µl of cytoplasmic extraction buffer (1× solution: 10 mM HEPES, 60 mM KCl, 1 mM EDTA, 0.075% (v/v) NP40, 1 mM DTT and 1 mM PMSF). Cells were incubated on ice for 3 min and centrifuged for 5 min at 4 °C at maximum speed. The cytoplasmic extract was removed from the nuclear pellet and washed three times with 100 µl cytoplasmic extraction buffer without detergent NP40. Next, pellets were frozen in liquid nitrogen and stored at −80 °C for subsequent processing steps.

Sample preparation for MS

Samples for MS analysis were prepared using an iST kit (PreOmics, P.O.00027) according to the manufacturer's instructions. Frozen pellets from the nuclear extraction step were incubated for 10 min with 50 µl lysis buffer at 95 °C. To shear long DNA fragments, cold lysate was added and the sample was sonicated with an ultrasonication probe for 20 s (amplitude 50%, cycle 0.5 s; UP100H, Hielscher). The total protein concentration was determined by measuring tryptophan fluorescence. The protein lysate was transferred into the cartridge, mixed with 50 µl lysate buffer and digested overnight at 37 °C. Digestion was quenched with 100 µl of Stop solution. Peptides were bound to sorbent in the cartridge by centrifugation at room temperature at 3,800*g* for 3 min. Then a wash with 200 µl of Wash1 and then with of Wash2 solution was performed. The flow through was discarded, and cleaned peptides were eluted from the cartridge in two steps by adding 100 µl of Elute buffer and centrifugation at room temperature at 3,800*g* for 3 min. The peptide solution was placed into a SpeedVac machine until

completely dry. The sample was then resuspended in 50 µl of 0.1% trifluoroacetic acid (TFA) and sonicated in an ultrasonication bath for 5 min to facilitate peptide solubilization. The peptide solution was stored at −80 °C before further use.

Peptide separation

The nano HPLC system used was an UltiMate 3000 RSLC nano system coupled to a Q Exactive HF-X mass spectrometer, equipped with an EASY-spray ion source (Thermo Fisher Scientific) and a JailBreak 1.0 adaptor insert as the spray emitter (Phoenix S&T). Peptides were loaded onto a trap column (Thermo Fisher Scientific, PepMap C18, 5 mm × 300 µm i.d., 5 µm particles, 100 Å pore size) at a flow rate of 25 µl min^{-1} using 0.1% TFA as the mobile phase. After 10 min, the trap column was switched in line with the analytical column (Thermo Fisher Scientific, PepMap C18, 500 mm × 75 µm i.d., 2 µm, 100 Å). For shotgun MS analysis, peptides were eluted using a flow rate of 230 nl min^{-1} and a binary 3 h gradient of 220 min. The gradient started with mobile phases of 98% A (water:formic acid, 99.9:0.1 v/v) and 2% B (water:acetonitrile:formic acid, 19.92:80:0.08 v/v/v), increasing to 35% B over the next 180 min, followed by a gradient over 5 min to 90% B, held for 5 min and decreasing over 2 min back to gradient 98% A and 2% B for equilibration at 30 °C. For parallel reaction monitoring, peptides were eluted using a flow rate of 230 nl min^{-1} and a binary 1 h gradient of 105 min. The gradient started with mobile phases of 98% A (water:formic acid, 99.9:0.1 v/v) and 2% B (water:acetonitrile:formic acid, 19.92:80:0.08 v/v/v) and held for 10 min, increasing to 35% B over the next 60 min, followed by a gradient over 5 min to 95% B, held for 5 min and decreasing over 2 min back to gradient 98% A and 2% B for equilibration at 30 °C.

Shotgun MS analysis

The Q Exactive HF-X mass spectrometer was operated in the data-dependent mode using a full scan (*m/z* range of 380–1,500, nominal resolution of 60,000, target value of 1×10^6 ions) followed by MS/MS scans of the 10 most abundant ions. MS/MS spectra were acquired using a normalized collision energy of 28, an isolation width of 1.0 *m/z*, a resolution of 30,000 and a target value of 1×10^5 ions. Precursor ions selected for fragmentation (exclude charge state 1, 7, 8 and >8) were placed on a dynamic exclusion list for 60 s. Additionally, the minimum AGC target was set to 5×10^3, and the intensity threshold was calculated to be 4.8×10^4. The peptide match feature was set to 'preferred', and the exclude isotopes feature was enabled. For peptide identification, the RAW files were loaded into Proteome Discoverer (v.2.3.0.522, Thermo Scientific). All the created MS/MS spectra were searched using MSAmanda v.2.0.0.9849 (ref. [53]). For the first step search, the RAW files were searched against the SwissProt human database (2019-02-23; 20,333 sequences and 11,357,489 residues) using the following search parameters: the peptide mass tolerance was set to ±5 ppm and the fragment mass tolerance to 15 ppm; the maximal number of missed cleavages was set to 2; and the result was filtered to 1% false discovery rate (FDR) on the protein level using the Percolator algorithm integrated in Thermo Proteome Discoverer. A subdatabase was generated for further processing. For the second step, the RAW files were searched against the created subdatabase called Neumayr_20190223_QExHFX4_med14_human_step1.fasta. For the search parameters, β-methylthiolation on cysteine was set as a fixed modification, and the following were set as variable modifications: oxidation on methionine; deamidation on asparagine and glutamine; acetylation on lysine; phosphorylation on serine, threonine and tyrosine; methylation on lysine and arginine; dimethylation on lysine and arginine; trimethylation on lysine; ubiquitinylation residue on lysine; and biotinylation on lysine. Monoisotopic masses were searched within unrestricted protein masses for tryptic enzymatic specificity. The peptide mass tolerance was set to ±5 ppm and the fragment mass tolerance to ±15 ppm. The maximal number of missed cleavages was set to two. The result was filtered to 1% FDR on the peptide level using the Percolator algorithm integrated in Thermo Proteome Discoverer. Peptide areas were quantified using IMP-apQuant[54]. Statistical significance of differentially abundant peptides and proteins between different conditions was determined using a paired LIMMA test[55].

Parallel reaction monitoring

The Q Exactive HF-X mass spectrometer was operated using a mixed MS method, which consisted of one full scan (*m/z* range of 380–1,500, resolution of 15,000 and target value of 1×10^6 ions) followed by the parallel reaction monitoring of targeted peptides from an inclusion list (isolation window of 0.7 *m/z*, normalized collision energy (NCE) of 30, resolution of 30,000 and AGC target of 2×10^5 ions). The maximum injection time variably changed based on the number of targets in the inclusion list to use up the total cycle time of 3 s. The scheduling window was set to 4 min for each precursor. A list of peptides, including basic MS information used for parallel reaction monitoring analysis, and proteins of interest and seven normalization proteins are displayed in Supplementary Table 1. Data processing and manual evaluation of results were performed in Skyline-daily[56] (64-bit, v.19.0.9.190). For data processing, peptides that had at least three specific peptide fragments were used. Proteins of interest were quantified on the basis of integrated ion intensities over retention time of peptides from the inclusion list. To account for different amounts between samples, these values were normalized on the basis of a set of seven abundant/housekeeping proteins (Supplementary Table 1).

STARR-seq

Cells were grown in square plates (Thermo Scientific, 166508) with a seeding density of about 20 million cells per square plate 2 days before transfection. For genome-wide screens, 4×10^8 cells were used, whereas for BAC screens, 4×10^7 cells were used. A genome-wide (Addgene, 99296) or a BAC STARR-seq library utilizing the ORI as a core promoter[6] was electroporated using Maxcyte STX into 85% confluent *OsTir1*$^{+/-}$ COF-AID$^{+/+}$ cells. After 30 min of recovery, cells were split in two conditions: those that received medium containing water and those that received medium containing IAA (500 µM final concentration; 2×10^8 cells). After 6 h, cells were collected, and total RNA was isolated using a RNeasy Maxi kit (Qiagen, 75162) containing β-mercaptoethanol supplemented RLT buffer. Spike-in control was added in a 1:1,000 ratio to the isolated total RNA. Subsequent steps were carried out as described in refs. [6,57]. In brief, mRNA was isolated using Oligo-dT25 beads (Invitrogen, 61005) followed by 1 h of TurboDNaseI treatment (Invitrogen, AM2238) at 37 °C. Subsequently, mRNA was cleaned using AMPure XP beads (Beckman Coulter, A63882) at 1:1.8 ratio (RNA:beads) followed by reverse transcription by SuperScript III (Invitrogen, 18080093) using a gene-specific primer (GSP) using the following conditions: 50 °C for 1 h, 70 °C for 15 min, and 4 °C for 10 min. Afterwards, cDNA was treated with RNaseA (Thermo Fisher, EN0531) for 1 h at 37 °C followed by clean-up using AMPure XP beads at 1:1.8 ratio. Next, 'junction PCR', which allows enrichment of reporter transcripts, was performed using KAPA 2× HiFi (KapaBiosystems, KK2601) utilizing the following thermocycler program: 98 °C for 45 s, 98 °C for 15 s and 65 °C for 30 s for 16 cycles; 72 °C for 70 s; and 72 °C for 120 s. Samples were then purified with AMPure XP at 1:0.8 ratio (DNA:beads). Afterwards, 'sequencing ready PCR', which amplifies STARR-seq transcripts, was performed on the junction PCR products using Illumina primers with the following thermocycler program: 98 °C for 45 s, 98 °C for 15 s and 65 °C for 30 s for 5 cycles; 72 °C for 45 s; and 72 °C for 120 s. Illumina adapter-containing STARR-seq library fragments were cleaned using SPRIselect beads (Beckman Coulter, B23318) with a stringent ratio of 1:0.5 (DNA:beads) and deep sequenced, paired-end, on an Illumina HiSeq2500 or NextSeq550 platform following the manufacturer's protocol, recovering 15–20 million (genome-wide) or 1.5–2 million (BAC) reads per sample. Deep sequencing base-calling was performed with CASAVA 1.9.1.

STARR-seq spike-in controls

To accurately quantify changes in enhancer activity after COF degradation and facilitate detection of potential global loss, we used spike-in controls for normalization of STARR-seq signals. In total, 13 neutral/enhancer sequences (Supplementary Table 2) from either the human or mouse genome were cloned into the STARR-seq vector[6] (Addgene, 99296) downstream of the ORI into the 3′ UTR. Five human spike-in sequences were flanked by a 2-bp unique *Drosophila melanogaster* sequence to distinguish spike-in reads from genome-wide STARR-seq reads and cloned in one orientation. Four promoter-proximal mouse enhancers were cloned in both orientations. All individually cloned vectors were pooled equimolar and electroporated into HCT116 cells. Total RNA was collected after 6 h and stored at −80 °C. Spike-in was added to each genome-wide STARR-seq screen in a ratio of 1:1,000 at the total RNA isolation step.

PRO-seq

The PRO-seq protocol was adapted from ref. [58]. as follows. A total of 1×10^7 COF-AID cells or WT HCT116 cells per replicate were collected and nuclei were isolated after the following treatments: (1) 3 h DMSO (mock); (2) 3 h 500 μM IAA (MED14-AID and BRD4-AID); (3) 3 h 10 μM Nutlin-3a (Sigma, SML0580); or (4) 3 h 500 μM IAA and subsequent 3 h 10 μM Nutlin-3a (MED14-AID and BRD4-AID). Spike-in control (S2 cells; 1% of total human cells) were added at the level of nuclei permeabilization step. Subsequent nuclear-run-on was performed for 3 min at 37 °C with biotin-labelled CTPs (Perkin Elmer, NEL542001EA) followed by RNA extraction and base hydrolysis. Biotin nuclear-run-on RNA was enriched using M280 streptavidin beads (Invitrogen, 112.06D) and precipitated by phenol–chloroform treatment. Next, 3′ RNA adapters were ligated, and second biotin RNA enrichment followed by RNA 5′-cap modification by TAP (Biozym, 187005) treatment was performed. Furthermore, 5′-hydroxyl repair by PNK (NEB, M0201S) and subsequent 5′ adapter ligation was carried out. Afterwards, cDNA was generated from enriched RNA by reverse transcription (Super Script III Reverse Transcriptase, Invitrogen, 18080-044). A total of 10 μl of the cDNA library was amplified by KAPA Amplification reaction (Roche, 7959028001) on a qPCR machine (Bio-Rad CFX Connect RealTime System). The KAPA reaction comprised 10 μl cDNA, 1 μl forward primer 35 μM (RP1-RP20), 1 μl of reverse primer 35 μM (RP1: 5′-AATGATACGGCGACCACCG AGATCTACAGTTCAGAGTTCTACAGTCCGA-3′), 25 μl 2× KAPA SYBR master mix and 13 μl water. The following PCR program was used: 98 °C for 45 s, 98 °C for 15 s, 60 °C for 30 s, 72 °C for 30 s and 72 °C for 10 s. Samples were removed from the qPCR machine after 12–15 cycles and cleaned with Ampure beads (Beckman, A63881) in a 1:1.4 ratio (sample:beads). DNA bound to the beads was eluted in 11 μl water and deep sequenced single-end on an Illumina HiSeq2500 platform following the manufacturer's protocol. Deep sequencing base-calling was performed with CASAVA 1.9.1.

P53 induction for qPCR

HCT116 COF-AID cells (5×10^5 per replicate) were treated for 3 h (MED14-AID, BRD4-AID, CDK9-AID and TAF1-AID cells) or 12 h (MED15-AID, MED19-AID and MED1-AID cells) with 500 μM IAA (SigmaAldrich, I5148-2G) or water (mock) at 37 °C. This was followed by 6 h of treatment with 10 μM Nutlin-3a (Sigma, SML0580) or DMSO (mock). Mouse CH12 KO cells were treated for 6 h with 30 μM Nutlin-3a (Sigma, SML0580) or DMSO (mock).

Oxidative stress induction

HCT116 MED14-AID cells (5×10^5 cells per replicate) were treated for 3 h with 500 μM IAA (SigmaAldrich, I5148-2G) or water (mock) at 37 °C. This was followed by 4 h treatment with 100 μM H_2O_2 or water (mock).

Heat-shock induction

HCT116 (parental, BRD4-AID, CDK9-AID and MED14-AID), K562 (BRD4-AID) and A549 (BRD4-AID) cells (5×10^5 cells per replicate) were treated for 3 h with 500 μM IAA (SigmaAldrich, I5148-2G) or water (mock) at 37 °C. This was followed by heat shock for 1 h at 43 °C.

Induction of LTR12 transcription

BRD4-AID cells (HCT116, K562 and A549) were treated for 18 h with 500 μM IAA (SigmaAldrich, I5148-2G) or water (mock) at 37 °C to observe induction of LTR12 transcription after BRD4 depletion.

siRNA-mediated knockdown

For gene knockdown by siRNA, 3×10^5 cells were plated into single 6-well plates 5 h before transfection. Lipofectamine 2000 (5 μl; Thermo Fisher, 11668027) was added to 250 μl OptiMEM (Invitrogen, 31985062) and incubated for 5 min. Meanwhile, siRNAs against target genes (10 nM final concentration, IDT) were mixed with 250 μl OptiMEM, the mixes were combined, incubated for 20 min and dropwise added to the cells. For NFYA and NFYB knockdown, BRD4-AID cells (HCT116 or A549) were used. Six hours after addition of NFYA and NFYB siRNAs, IAA (500 μM final concentration) or water (mock) was added for 18 h for a total of 24 h knockdown. For AFF1 and AFF4 knockdown, parental HCT116 cells (containing OsTir1) were used. After 24 h, knockdown cells were heat shocked for 1 h at 43 °C. For MED17 knockdown, parental HCT116 cells were used. At 18 h after the addition of MED17 siRNA, Nutlin-3a (10 μM final concentration) or DMSO (mock) was added for 6 h for a total of 24 h knockdown.

qPCR

Following the different treatments, cells were washed with 1× PBS, trypsinized for 3 min at 37 °C with 500 μl trypsin and collected after the addition of 500 μl medium. Cells were centrifuged at 500*g* and washed with 1× PBS. PBS was removed and cells were lysed using Qiashredder columns (Qiagen, 79654) followed by total RNA extraction using a RNeasy mini prep kit (Qiagen, 74104), with β-mercaptoethanol-supplemented RLT buffer. Isolated RNA (2 μg) was treated with 2 μl TurboDNase and 2 μl TurboDNase buffer (Invitrogen, AM2238) for 30 min at 37 °C in a thermocycler. Afterwards, 2 μl DNase inactivation reagent (Ambion, AM1906) was added, samples were vortexed for 2 min with 20 s breaks in between and centrifuged for 5 min at 10,000*g*. A volume of 10 μl of RNA was used for reverse transcription and comprised 1 μl $d(T)_{18}$ primer (NEB, S1316S) for mRNA or random hexamers (Bioline, 38028) for LTRs, 1 μl dNTPs (NEB, 4475), 1 μl RNase inhibitor (Thermo Fisher, EN0531), 1 μl SuperScript III (Invitrogen, 18080093), 1 μl DTT (Invitrogen, 18080093; within the SSIII kit), 4 μl forward strand buffer (Invitrogen, 18080093, within the SSIII kit) and 1 μl water. The reaction was mixed and heated to 25 °C for 5 min, 50 °C for 50 min, 70 °C for 15 min and 4 °C for 10 min in a thermocycler. Afterwards, samples were diluted to total of 100 μl, and 2 μl was used for qPCR. Reaction setup/sample comprised 10 μl SYBR-Green (Promega, A6002), 1 μl forward primer (10 μM final concentration), 1 μl reverse primer (10 μM final concentration), 7 μl water and 2 μl DNA. The qPCR setup/whole plate program consisted of 95 °C for 2 s, 95 °C for 3 min, 60 °C for 30 s, read plate, go back to step 2 for 39 times (40 cycles in total).

MED1 ChIP-seq

MED14-AID HCT116 cells were cultured as described above. Medium was removed and 1% formaldehyde in PBS for 15 min was used to fix cells. Glycine (0.5 ml, 2.5 M) was added to each plate and left to sit for 5 min. Medium was removed and plates were washed with PBS. PBS (10 ml) was added to the plate and cells were scraped off. Cell pellet was spun down and flash frozen in liquid nitrogen and stored at −80 °C, with about 140 million cells in each tube. All buffers contained freshly prepared complete protease inhibitors (Roche, 11873580001). Frozen crosslinked cells were thawed on ice and then resuspended in lysis buffer I (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and protease inhibitors) and rotated for 10 min at 4 °C, then spun at 1,350 r.c.f. for 5 min at 4 °C. The pellet was resuspended in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA,

0.5 mM EGTA and protease inhibitors) and rotated for 10 min at 4 °C and spun at 1,350 r.c.f. for 5 min at 4 °C. The pellet was resuspended in sonication buffer (20 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA 1 mM EGTA, 1% Triton X-100, 0.1% sodium-deoxycholate, 0.1% SDS and protease inhibitors) and then sonicated on a Misonix 3000 sonicator for 10 cycles at 30 s each on ice (18–21 W) with 60 s on ice between cycles. Sonicated lysates were cleared once by centrifugation at 16,000 r.c.f. for 10 min at 4 °C. Input material was reserved and the remainder was incubated overnight at 4 °C with magnetic beads bound with anti-MED1 antibody (Bethyl, A300-793A) to enrich for DNA fragments bound by MED1. Beads were washed with each of the following buffers: washed twice with sonication buffer (20 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA 1 mM EGTA, 1% Triton X-100, 0.1% sodium-deoxycholate and 0.1% SDS); once with sonication buffer with high salt (20 mM HEPES pH 7.5, 500 mM NaCl, 1 mM EDTA 1 mM EGTA, 1% Triton X-100, 0.1% sodium-deoxycholate and 0.1% SDS); once with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40 and 0.5% sodium-deoxycholate); and once with TE buffer. DNA was eluted off the beads by incubation with agitation at 65 °C for 15 min in elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA and 1% SDS). Crosslinks were reversed for 12 h at 65 °C. To purify eluted DNA, 200 ml TE was added and then RNA was degraded by the addition of 2.5 ml of 33 mg ml^{-1} RNase A (Sigma, R4642) and incubation at 37 °C for 2 h. Protein was degraded by the addition of 4 µl of 20 mg ml^{-1} proteinase K (Invitrogen, 25530049) and incubated at 55 °C for 30 min. DNA was purified using a Qiagen PCR purification kit, eluted in buffer EB, and deep sequenced single-end on an Illumina HiSeq2500 platform following the manufacturer's protocol.

MED1 immunofluorescence with RNA FISH

Immunofluorescence (IF) with concurrent RNA FISH was performed as previously described[59,60]. In brief, coverslips were coated at 37 °C with 5 µg ml^{-1} poly-L-ornithine (Sigma-Aldrich, P4957) for 30 min and 5 µg ml^{-1} of laminin (Corning, 354232) for 2 h. HCT116 cells were plated on the pre-coated cover slips and grown for 24 h. For the last 3 h, the cells were treated with 10 µM Nutlin-3a (Sigma, SML0580) or DMSO (mock) followed by fixation using 4% paraformaldehyde (VWR, BT140770) in PBS for 10 min. After washing cells three times in PBS, the coverslips were put into a humidifying chamber or stored at 4 °C in PBS. Permeabilization of cells was performed using 0.5% Triton X-100 (Sigma Aldrich, X100) in PBS for 10 min, followed by three PBS washes. Cells were blocked with 4% IgG-free BSA (VWR, 102643-516) for 30 min, and anti-MED1 antibody (Bethyl, A300-793A) was added at a concentration of 1:500 in PBS for 4–16 h. Cells were washed with PBS three times, followed by incubation with secondary antibody at a concentration of 1:5,000 in PBS for 1 h. After washing twice with PBS, cells were fixed using 4% paraformaldehyde (VWR, BT140770) in PBS for 10 min. After two washes of PBS, wash buffer A (20% Stellaris RNA FISH wash buffer A (Biosearch Technologies, SMF-WA1-60) and 10% deionized formamide (EMD Millipore, S4117)) in RNase-free water (Life Technologies, AM9932) was added to cells and incubated for 5 min. A total of 12.5 µM RNA probe (Biosearch Technologies, Stellaris RNA FISH Probe) in hybridization buffer (90% Stellaris RNA FISH hybridization buffer (Biosearch Technologies, SMF-HB1-10) and 10% deionized formamide) was added to cells and incubated overnight at 37 °C. After washing with wash buffer A for 30 min at 37 °C, the nuclei were stained with 20 mg ml^{-1} Hoechst 33258 (Life Technologies, H3569) for 5 min, followed by a 5-min wash in wash buffer B (Biosearch Technologies, SMF-WB1-20). Cells were washed once in water, followed by mounting the coverslip onto glass slides with Vectashield (VWR, 101098-042) and finally by sealing the cover slip with nail polish (Electron Microscopy Science, 72180). Images were acquired on a RPI Spinning Disk confocal microscope with a ×100 objective using MetaMorph acquisition software and a Hammamatsu ORCA-ER CCD camera (W. M. Keck Microscopy Facility, MIT). Images were post-processed using Fiji Is Just ImageJ (FIJI). RNA FISH probes were custom-designed and generated by Biosearch Technologies (Stellaris RNA FISH) to target *P21*, *RRM2B*, *TRIB1* and *MYC* intronic regions to visualize nascent RNA (Supplementary Table 1).

Oligonucleotide library of TATA box and CCAAT box motif mutations

Eight instances of LTR12 elements overlapping a STARR-seq peak and promoters of two histone genes insensitive to BRD4 depletion were used as representative BRD4-independent promoters. For each candidate, the extended promoter sequence consisting of 205 nt upstream and 35 nt downstream of the CAGE-defined TSS was selected and scored against the TATA box (TBP-binding motif) and the CCAAT box (NFYA- and NFYB-binding motif) position-weight matrix from the JASPAR database[61] with the R package seqPattern v.1.14.0. All motif instances with a match above 90% were replaced by a fixed, low scoring sequence with similar nucleotide content as follows: CCAATCAS→AACTGACC for CCAAT box motifs and STATAWAWRS→TGCAAGTCTT for the TATA box motif, creating mutants for the TATA box, the CCAAT box or both motifs together. For the gain-of-function approach, 18 transcriptionally inert 240-bp-long genomic regions were randomly selected. TATA box and/or CCAAT box motif instances from the ten BRD4-independent promoters were inserted into these neutral backgrounds by preserving the original number and arrangement of the motifs. Double motif insertions were designed for all 18 random sequences and motifs from all 10 BRD4-independent promoters, and single motif insertions for 6 random sequences and motifs from 4 promoters. Each 240-nt-long candidate sequence was present in the library 5 times and barcoded with a unique 10 nt random barcode at the 3′ end. Barcode sequences were designed to match the GC content of the human 5′ UTRs[62] and to differ from each other by at least 3 nucleotides. Designed 250-nt-long candidate sequences are provided in Supplementary Table 7. Sequences were flanked by the Illumina i5 (25 bp; 5′-TCCCTACACGACGCTCTTCC GATCT) and i7 (25 bp; 5′-GTTCAGACGTGTGCTCTTCCGATCT) adaptor sequences upstream and downstream, respectively, serving as constant linkers for amplification and cloning. The pool of 2,000 synthesized 300-mer oligonucleotides was obtained from Twist Biosciences.

STAP-seq

The STAP-seq input library was generated by cloning the amplified synthetic oligonucleotide pool into a human STAP-seq screening vector (Addgene, ID 125150) as previously described[34,51]. About 80 µg of input library was transfected into 4 × 10^7 BRD4-AID HCT116 cells using MaxCyte STX. Two independent transfections (biological replicates) were performed. After 30 min of recovery phase, cells were split in two conditions: those that received medium containing water or those that received medium with IAA (500 µM final concentration). Total RNA was isolated 6 h after electroporation followed by polyA+ RNA purification and turbo DNase treatment (Ambion, AM2238). Spike-in control was added in a 1:100 ratio to the isolated total RNA. STAP-seq RNA processing and cDNA amplification was performed as previously described[51]. Samples were sequenced paired-end on an Illumina NextSeq 550 platform following the manufacturer's protocol and base-calling was performed with CASAVA 1.9.1.

STAP-seq spike in controls

To accurately quantify changes in transcriptional activity after BRD4 degradation, we used spike-in controls for normalization of STAP-seq signals. Previously described spike-in mix consisting of nine mouse extended promoters cloned into a human STAP-seq spike-in vector (Addgene, ID 125152) was used[51]. WT HCT116 cells were electroporated with the spike-in plasmid mix and total RNA was isolated after 6 h as described above and stored at −80 °C. Spike-in RNA was added to each STAP-seq screen in a ratio of 1:100 at the total RNA isolation step.

STARR-seq data processing

Paired-end 50-bp-long STARR-seq reads were mapped using Bowtie[63] (v.1.2.2), first to the reference hg19 genome allowing up to three

mismatches and then to the reference consisting of five human (flanked by *D. melanogaster*) and four mouse spike-in sequences allowing one mismatch. Only read pairs that were uniquely mapping were kept. Mapped reads were sorted and indexed with samtools v.0.1.19 and combined into paired-end fragments with the R/Bioconductor[64] package GenomicAlignments v.1.18.1. Summary of reads mapping to the reference genome and spike-in sequences for each sample is provided in Supplementary Table 2.

STARR-seq normalization by spike-in

For each spike-in sequence, the number of paired-end fragments mapping exactly to sequence ends and spanning the entire cloned spike-in sequence in the correct orientation was counted. For mouse spike-in sequences that were cloned in both orientations, mappings in the two orientations were considered separately. For each individual STARR-seq sample, relative abundance (proportion) of each of the 13 cloned spike-in sequences was calculated and scaled by dividing with the mean across the 13 sequences. These relative abundances were used to normalize the STARR-seq signal between IAA treated and control condition for each AID-tagged COF as follows. For each individual sample (replicate) the median of scaled relative abundances across 13 spike-in sequences was taken and used to calculate the ratio between paired treated and control samples (these samples stem from the same STARR-seq library transfection and differ only in the treatment). The control sample was then set to 1, and the scaling factor for the treated sample was expressed relative to the control using the calculated ratio. Finally, for each AID-tagged COF, the mean scaling factor across the replicates was taken to make the normalization more robust and less sensitive to variability between replicates. For P300/CBP-AID, we did not use spike-in for normalization because it is not reliable in this case. p300/CBP regulates the transcription of rRNAs by Pol I[65]; therefore depletion of P300/CBP leads to drastic changes in total cellular RNA abundance. Our normalization approach relies on adding spike-in RNA in a fixed ratio to total RNA and assumes that the bulk of total cellular RNA is not changing, so it cannot be used in the case of P300/CBP depletion. All spike-in counts, relative abundances and calculations of scaling factors are provided in Supplementary Table 2. The final scaling factor for each AID-tagged COF was used to normalize the STARR-seq coverage in IAA treatment relative to control and was supplied as a custom scaling factor in differential analysis.

Detection and quantification of enhancer activity

For each AID-tagged COF and condition, unique STARR-seq fragments (after removing duplicates) from all replicates were combined and used for peak calling with MACS2 v.2.1.2.1. Genome-wide STARR-seq library input was previously sequenced[6] and used here as background for peak calling. Only peaks at 1% FDR with enrichment over input ≥3 on both strands and at least 3 tags per million (corresponding to about 25 fragments) were kept and combined into a reference set of 6,249 STARR-seq enhancers. The number of unique fragments for peak calling and peaks called per COF and condition is provided in Supplementary Table 2. Note that due to COF depletion, the number of peaks called per condition varies, yet all enhancer activity changes are re-evaluated independently of these initial peak calling for each of the 6,249 enhancers in the reference set. To quantify enhancer activity, the number of STARR-seq fragments overlapping each enhancer in the reference set was counted in each individual STARR-seq sample (replicate). A raw count table is provided in Supplementary Table 3, and was used for subsequent differential analysis.

Differential analysis of COF-AID STARR-seq

Differential analysis between IAA-treated and control conditions was performed per COF-AID cell line with the R/Bioconductor package edgeR[66] (v.3.24.3), always using the same reference set of 6,249 STARR-seq enhancers. The scaling factor calculated from spike-in was supplied as a custom scaling factor for normalization to allow accurate assessment of changes in enhancer activity and possible detection of global effects. Significant changes in enhancer activity were called at 5% FDR (Extended Data Fig. 2d) . Corrected $\log_2$(fold change) values and multiple-testing adjusted *P* values from edgeR for all enhancers in the reference set were used for downstream analyses and are provided in Supplementary Table 3. To assess the effect of COF tagging on enhancer activity (in the absence of IAA), we also performed differential analysis between control condition of each COF and the parental cell line with edgeR, calling significant changes at 5% FDR (Extended Data Fig. 2c).

Clustering of COF-AID STARR-seq screens

To group the different COF-AID cell lines on the basis of enhancer activity, we used normalized COF STARR-seq signals from merged replicates per COF and condition (IAA treatment and control). Hierarchical clustering was performed using Manhattan distance between normalized STARR-seq signals (Extended Data Fig. 2b). To group the COF-AID cell lines based on changes in enhancer activity after IAA treatment, we performed hierarchical clustering using Manhattan distance between $\log_2$(fold change) values (Fig. 1f).

Clustering of STARR-seq enhancers

We clustered enhancers on the basis of change in their activity after depletion of five individual COFs (BRD2, BRD4, P300/CBP, MED14 and CDK7) with *k*-medoids (Fig. 2a). PAM (*k*-medoids) was performed on $\log_2$(fold change) values using the PAM algorithm implemented in the R package cluster v.2.0.7-1. To determine the optimal number of clusters, PAM was initially run with varying number of clusters from 1 to 10, and for each run, the proportion of variance explained by clustering was calculated as ratio of within-cluster variance and between-cluster variance. Clustering into 4 clusters explained more than 85% of the variance and further increasing the number of clusters led to less than 5% gain (Extended Data Fig. 3b), so we selected 4 as the optimal number of clusters. To make the clustering robust, we ran PAM with $k = 4$ clusters independently 1,000 times, each time using different randomly chosen data points as initial centroids. For each enhancer, we then calculated the number of times it was assigned to each of the four clusters and assigned it to the most frequent cluster. The clustering was robust, with the majority of enhancers (>86%) assigned to the same cluster >50% of the time. To further confirm the robustness of the defined enhancer groups (size of groups and enhancer group membership), we used two alternative clustering approaches. We performed hierarchical clustering using Euclidean distance metric, and defined five clusters by cutting the dendrogram. For each hierarchical cluster we calculated the percentage of enhancers that are assigned to each of the four originally defined PAM enhancer groups. This revealed an almost 1:1 correspondence between hierarchical clusters and originally defined PAM clusters, with more than 80% of enhancers in each hierarchical cluster belonging to a single originally defined enhancer group (Extended Data Fig. 3c,d). We also used uniform manifold approximation and projection (UMAP) algorithm to reduce the dimensionality and visualize the data. This revealed a clear separation of originally defined enhancer groups in two-dimensional UMAP representation (Extended Data Fig. 3e).

Annotation of enhancers with TF motifs and transposable elements

All TF motifs from the JASPAR 2020 vertebrate core collection[61] of 579 non-redundant motifs were considered, and the occurrence of these motifs at different score thresholds in the hg19 genome assembly was downloaded directly from the JASPAR database (http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_pfms_jaspar.zip). Only the most highly scoring motif occurrences, with a score in the top 1 percentile of the scores for the respective motif, were kept. These motif occurrences were overlapped with STARR-seq

enhancers, and a binary matrix denoting which motifs are present in each enhancer was constructed. For annotation of enhancers with transposable elements, the annotation of repeats from RepatMasker for hg19 genome assembly was downloaded from the UCSC Table Browser[67].

Annotation of enhancers with TF and COF binding and histone modifications

Various published datasets for the HCT116 cell line were downloaded from the Gene Expression Omnibus (GEO) repository and ENCODE database, including chromatin accessibility[68,69], ChIP-seq for different histone modifications[68,70], TFs[27,68] and COFs[23,38,70,71]. All accession numbers of used published datasets are listed in Supplementary Table 4. Raw sequencing data were downloaded from GEO or the Sequence Read Archive SRA), and reads were mapped with Bowtie v.1.2.2 to hg19 genome assembly allowing only unique mapping. Peaks were called with MACS2 v.2.1.2.1. against matching input (if available) using only unique reads and default MACS2 parameters, keeping peaks at 5% FDR. For datasets from ENCODE, the peaks files were downloaded and used directly in downstream analyses. ChIP-seq peaks from individual datasets were overlapped with STARR-seq enhancers, and a binary matrix denoting which TF, COF or histone modification peaks are present in each enhancer was constructed.

Motif, TF and COF binding and histone modification enrichment analysis

For enrichment analysis, a binary matrix denoting which enhancers overlap which motifs, repeat elements, TF and COF binding sites or histone modifications was used. To create a random background for assessing enrichment, STARR-seq peaks were shifted by 10 kb and the resulting shifted regions were annotated with motifs, TF and COF binding sites and histone modifications as described above. Two-sided Fisher's exact test was used to assess the enrichment or depletion of a particular feature in a specific group of enhancers, either against random regions or against enhancers in other groups. Enrichment and depletion values (odds ratios) of different features across different groups of enhancers were visualized in the form of a heatmap, showing only significant enrichments ($P \leq 0.05$; Fig. 2c–e and Extended Data Fig. 3h).

Multiple alignment of LTR12 elements

Sequences of LTR12 family retrotransposons overlapping STARR-seq enhancers were multiple aligned using ClustalW algorithm implemented in the R package msa v.1.14.0. Multiple alignment was visualized with ggmsa v.0.0.2 package (Extended Data Fig. 7d).

Gene and TSS annotation

To obtain a non-redundant set of genes and their precise associated TSSs for accurate quantification of PRO-seq signals in different gene regions, we pre-processed and refined gene annotation as follows. We took all coding and long noncoding transcripts from Ensembl v.82 for hg19 genome assembly and removed transcripts shorter than 300 bp. For each group of transcripts that have the same annotated TSS, we kept only the longest one. We annotated these non-redundant transcripts with CAGE TSS clusters from FANTOM5 (ref. [72]) as follows. For each transcript (unique annotated TSS), we identified the strongest CAGE TSS within a window encompassing 500 bp upstream and 500 bp downstream of the annotated TSS, excluding the coding sequence. Then, for each selected CAGE TSS (that was possibly associated with multiple annotated transcripts), we kept the closest transcript and corrected its annotated TSS to the CAGE TSS. The resulting non-redundant transcript/gene annotation with precise CAGE-corrected TSSs was used in all downstream analyses.

Gene ontology analysis

We assessed whether genes with promoters containing CCAAT and TATA boxes are enriched for a particular gene ontology (GO) term by calculating hypergeometric *P* values for every GO term with the R/Bioconductor package GOstats[73] (v.2.48.0), using genes containing CCAAT and TATA boxes as a foreground and all other annotated genes as a background. Only terms with $P \leq 1 \times 10^{-4}$ were considered significant and sorted by the enrichment. The top 5 enriched terms for each of the 3 GO categories (biological process, molecular function and cellular compartment) are shown (Extended Data Fig. 8a).

PRO-seq data processing

Single-end 50-bp-long PRO-seq reads contained a 8-bp long unique molecular identifier (UMI) at the 5′ end, which was removed before mapping and kept track of. From the remaining 42 bp, the Illumina adapter was trimmed with cutadapt v.1.18. Reads longer than 15 bp after adapter trimming were mapped using Bowtie[63] (v.1.2.2) to a reference consisting of hg19 and dm3 (spike-in) genome allowing up to 2 mismatches. Multimapping was allowed to up to 1,000 positions, and all multimapping reads were randomly assigned to one mapping position. For reads that mapped to the same genomic position, we collapsed those that had identical UMIs as well as those for which the UMIs differed by a single nucleotide to ensure the counting of unique nascent RNA molecules. To generate the coverage of PRO-seq signal, that is, exact positions of Pol II molecules associated with 3′ end of nascent transcripts, only the first nucleotide of each read was considered, and the strand was swapped to match the direction of transcription. Summary of reads mapping to the reference genome and spike-in genome, and counts of reads with unique UMIs for all PRO-seq samples is provided in Supplementary Table 5.

Differential analysis of PRO-seq

Differential analysis was performed using a non-redundant set of genes with CAGE-corrected TSSs. For each gene, the region from the TSS up to 150 bp downstream (+1 to +150) was defined as the 'promoter + pause region', and the rest of the annotated gene was defined as 'gene body'. For BRD4 depletion in BRD4-AID cells (Fig. 4f,g), the number of unique (UMI collapsed) PRO-seq read 5′ ends falling into these two regions was counted for each gene. Differential analysis was performed with DESeq2 (v.1.22.2)[74] for 'promoter + pause' and 'gene body' region separately to capture the pause-release defect. For MED14 depletion in MED14-AID cells and induction of P53 target genes by Nutlin-3a in WT, MED14-AID and BRD4-AID cells (Fig. 3d,e and Extended Data Fig. 4a,d), the number of unique (UMI collapsed) PRO-seq read 5′ ends falling into the whole gene region was counted and differential analysis was performed on the entire gene. Raw PRO-seq counts used for differential analysis are provided in Supplementary Table 6. To allow accurate assessment of changes in enhancer activity after different treatments and possible detection of global effects, we used spike-in based normalization. A scaling factor for normalization between conditions was calculated from relative abundance of reads mapping to spike-in genome (dm3) in combined replicates for each condition. Spike-in normalization factors were supplied as custom scaling factors to DESeq2, with all replicates of the same condition receiving the same scaling factor. These scaling factors were also used to normalize PRO-seq coverage of combined replicates per condition for visualization in the genome browser. Spike-in read counts, relative abundances and calculations of scaling factors are provided in Supplementary Table 5.

qPCR data analysis

All treatments for qPCR were done in at least three independent biological replicates and each sample was measured at least two times (technical replicates). Raw CT values of technical replicates were averaged and then normalized to a reference gene: *GAPDH* for all human WT and AID-tagged cell lines and *Actb* for mouse CH12 WT and KO cell lines. When calculating a ratio to a control (no treatment) condition, the normalized value for each individual replicate of the treated condition was divided by the normalized value for the corresponding replicate

of the control condition. Obtained ratios therefore accounted for the variance in both treated and control conditions and were used to calculate the standard deviation shown in all qPCR barplots and to perform two-sided Student's *t*-test (Figs. 3g and 4c,e and Extended Data Figs. 4i,j,k, 6a–d, 7c,f, h–j and 9e–g).

MED1 ChIP-seq data processing and analysis

Single-end 50-bp-long reads were mapped using Bowtie v.1.2.2 to the reference hg19 genome, allowing up to 3 mismatches and only uniquely mapping reads were retained. A summary of reads mapping to the reference genome for each sample is provided in Supplementary Table 5. To generate genome-wide coverage, mapped reads were extended to 500 bp with GenomicRanges v.1.34.0. and the coverage was normalized to reads per million. Unique reads were used to call peaks with MACS2 v.2.1.2.1 for each condition and treatment against the respective input, using default MACS2 settings (adjusted $P \leq 0.05$). For WT HCT116 cells, unique reads from two independent biological replicates were combined before peak calling to obtain a common set of peaks per condition. Peaks from different conditions were sequentially combined to obtain a non-redundant set of reference peaks and ChIP-seq signal (ChIP-seq coverage over input) from different datasets centred at the reference peak summits was visualized (Extended Data Fig. 5c). The MED1 ChIP-seq signal was quantified and compared at two different types of STARR-seq enhancers: (1) P53-bound (overlapping a P53 ChIP-seq peak in HCT116 cells after Nutlin-3a treatment[27]) enhancers insensitive to MED14 depletion according to differential analysis of MED14 STARR-seq; and (2) accessible (according to DHS-seq) and H3K27ac-marked enhancers significantly downregulated after MED14 depletion according to differential analysis of MED14 STARR-seq (Extended Data Fig. 5e).

Analysis of MED1 IF with RNA FISH

Three-dimensional image data gathered in RNA FISH and IF channels for about 120 cells per FISH probe (gene) were processed with custom Python and Matlab scripts as previously described[59,60]. In brief, FISH foci were manually identified in individual *z*-stacks through intensity thresholds, centred along a box of size l = 1 µm, and stitched together in three dimensions across *z*-stacks. Only cells with one or two FISH foci were considered for downstream analyses. For every RNA FISH focus identified, the signal from the corresponding location in the IF channel was gathered in the l × l square centred at the RNA FISH focus at every corresponding *z*-slice. The IF signal centred at FISH foci for each FISH and IF pair were then combined, and an average intensity projection calculated, thereby providing averaged data for the IF signal intensity within a l × l square centred at FISH foci. The same process was carried out for the FISH signal intensity centred on its own coordinates, thereby providing averaged data for the FISH signal intensity within a l × l square centred at FISH foci. As a control, this same process was carried out for an IF signal centred at randomly selected nuclear positions within the nuclear volume determined from DAPI staining through the *z*-stack image as described in detail in ref. [59]. Average MED1 IF intensity projections centred at FISH foci were visualized using the same intensity colour range for all genes, ranging from minimal to maximal observed IF intensity within the 1 × 1 µm area (Fig. 3h and Extended Data Fig. 5f). For quantitative comparison of MED1 IF signal between different genes (Fig. 3i), the MED1 IF signal at each FISH focus was normalized to the average signal at random spots from the same dataset to account for the difference in overall MED1 IF intensity between different datasets.

STAP-seq reads processing

STAP-seq sequencing reads were processed as previously described[51]. In brief, paired-end STAP-seq reads were mapped to a reference containing 250-bp-long sequences of 2,000 barcoded WT and mutant promoter oligonucleotides and to the 9 mouse spike-in promoter sequences using Bowtie[63] (v.1.2.2) allowing only 1 mismatch. Before mapping, the 10-nt-long UMI was removed from the 5′ end of the forward read and kept track of for later counting. Only uniquely mapping read pairs for which the reverse read mapped exactly to the oligonucleotide end were kept, ensuring they correspond to reporter transcripts transcribed from that particular cloned barcoded promoter candidate. For read pairs that mapped to the same positions, we collapsed those that had identical UMIs as well as those for which the UMIs differed by a single nucleotide to ensure the counting of unique reporter transcripts. Tag counts at each position represent the sum of the 5′-most position of UMI collapsed fragments. Total read counts mapping to promoter oligonucleotide library and spike-in promoters are summarized in Supplementary Table 5.

STAP-seq data analysis

Tag counts at each position in each screened promoter candidate were quantified in different conditions and datasets as described above and represent the number of unique RNA molecules initiated at that position (Supplementary Table 7). Raw counts were normalized by the spike-in as previously described[51]. In brief, the number of unique RNA molecules originating from each of the nine spike-in mouse promoters was quantified as described above, and the counts were used to calculate the scaling factor from each individual spike-in promoter. Final normalization factor was calculated as median of factors derived from individual spike-in promoters and is provided in Supplementary Table 5. For comparison of transcriptional output between WT promoters or neutral sequences and their mutated variants, the sum of normalized counts in the 5-bp window centred at the cognate/expected TSS (position 206 in the 250-bp-long promoter candidate) was considered and was corrected for the abundance of each promoter sequence in the input STAP-seq library (Fig. 4j). For visualization of transcriptional output per position in a specific promoter variant (WT or mutant), the signal from five instances of that promoter variant present in the library (each barcoded with a different unique barcode) was combined (Fig. 4i and Extended Data Fig. 9c,d).

Statistics and data visualization

All statistical calculations and graphical displays were performed in R statistical computing environment[75] (v.3.5.1). In all box plots, the central line denotes the median, the box encompasses the 25th to 75th percentiles (interquartile range) and the whiskers extend from the 5th to 95th percentiles of the data. In all bar plots, the bar height denotes the mean and error bars denote the standard deviation. Heatmaps were created with R package gplots v.3.0.1. Coverage data tracks were visualized in the UCSC Genome Browser[67] and used to create displays of representative genomic loci.

Reporting summary

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

Data availability

All raw deep-sequencing data (STARR-seq, PRO-seq, ChIP-seq and STAP-seq) and associated processed data generated in this study have been deposited in the NCBI GEO database under accession number GSE156741. Previously published datasets re-analysed in this study are available in the GEO repository under the following accession numbers: GSE100432 (genome-wide STARR-seq input library), GSE97889 (ATAC-seq), GSE71510 (H3K4me1, H3K4me3, H3K27ac, SMARCC1 and SMARCA4 ChIP-seq), GSE51176 (P300 and MLL4 ChIP-seq), GSE57628 (BRD4 ChIP-seq), GSE38258 (CDK8 ChIP-seq) and GSE86164 (P53 ChIP-seq). Peak files for the following ChIP-seq datasets are available from ENCODE (https://www.encodeproject.org/): DNase-seq (ENCFF001SQU, ENCFF001WIJ, ENCFF001WIK, ENCFF175RBN, ENCFF228YKV, ENCFF851NWR, ENCFF927AHJ, ENCFF945KJN

and ENCFF360XGA), H3K36me3 (ENCFF467KXG, ENCFF742ZBG and ENCFF922EIA), H3K27me3 (ENCFF237TTT, ENCFF991HKN and ENCFF029ZPV), H3K9me2 (ENCFF586SOS, ENCFF808XMV and ENCFF346SOF), H3K9me3 (ENCFF751VFZ, ENCFF577FKU and ENCFF909UTX), JUND (ENCFF001UDY, ENCFF001UDZ, ENCFF950JTT and ENCFF088WYS) and FOSL1 (ENCFF001UDW and ENCFF001UDX). The vertebrate transcription factor motif collection is available from the JASPAR database (http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_ pfms_jaspar.zip). The SwissProt Human database is available athttps://www.uniprot.org/proteomes/UP000005640. No restrictions on data availability apply.

Code availability

All custom codes used for data processing and computational analyses are available from the authors upon request.

52. Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I. T. & Yamamoto, T. MMEJ-assisted gene knock-in using TALENs and CRISPR–Cas9 with the PITCh systems. *Nat. Protoc.* **11**, 118–133 (2016).
53. Dorfer, V. et al. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* **13**, 3679–3684 (2014).
54. Doblmann, J. et al. apQuant: accurate label-free quantification by quality filtering. *J. Proteome Res.* **18**, 535–541 (2019).
55. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
56. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
57. Neumayr, C., Pagani, M., Stark, A. & Arnold, C. D. STARR-seq and UMI-STARR-seq: assessing enhancer activities for genome-wide-, high-, and low-complexity candidate libraries. *Curr. Protoc. Mol. Biol.* **128**, e105 (2019).
58. Mahat, D. B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
59. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
60. Guo, Y. E. et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **13**, 720–726 (2019).
61. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
62. Zhang, L., Kasif, S., Cantor, C. R. & Broude, N. E. GC/AT-content spikes as genomic punctuation marks. *Proc. Natl Acad. Sci. USA* **101**, 16855–16860 (2004).
63. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
64. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
65. Pelletier, G. et al. Competitive recruitment of CBP and Rb-HDAC regulates UBF acetylation and ribosomal transcription. *Mol. Cell* **6**, 1059–1066 (2000).
66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
67. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
68. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
69. Ponnaluri, V. K. C. et al. NicE-seq: high resolution open chromatin profiling. *Genome Biol.* **18**, 122–15 (2017).
70. Mathur, R. et al. ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat. Genet.* **49**, 296–302 (2017).
71. Baranello, L. et al. RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* **165**, 357–371 (2016).
72. The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
73. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
75. The R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
76. Allen, M. A. et al. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *eLife* **3**, R106 (2014).

Acknowledgements We thank F. Muerdter, M. Muhar, J. Zuber and U. Schoeberl (all Research Institute of Molecular Pathology (IMP), Vienna BioCenter) for advice and help establishing the AID system and PRO-seq, respectively; R. Imre (IMP and Institute of Molecular Biotechnology, Vienna BioCenter) for help with analysing MS data; D. Hnisz (MPI Berlin) for help with Mediator ChIP-seq; R. C. Casellas, J. Kalchschmidt and S. K. Jung (NIH NIAMS) for sharing the mouse Mediator KO cell lines and data; D. Taatjes (University of Colorado), C. Bernecky (IST) and R. Young (MIT) for discussions; B. Sabari (UTSW) for feedback and help; and C. Buecker and H. Thomas (MPL), C. Plaschka (IMP) and A. Andersen (Life Science Editors) for comments on the manuscript; the staff at the IMP/IMBA in-house FACS facility, in particular G. Schmauss, and staff at the Molecular Biology Service. Deep sequencing was performed at the Vienna BioCenter Core Facilities. V.H. is supported by the Human Frontier Science Program (grant number LT000324/2016-L) and A.B. by the Swedish Research Council Postdoctoral Fellowship (VR 2017-00372). Research in the Stark group is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 647320) and by the Austrian Science Fund (FWF, F4303-B09). Basic research at the IMP is supported by Boehringer Ingelheim and the Austrian Research Promotion Agency (FFG).

Author contributions C.N. and A.S. conceived the project. C.N., K.B., O.H. and L.S. generated the CRISPR–Cas9-edited cell lines. M.P., M.R. and K.B. cultured cells and performed transfections. C.N., K.K. and C.D.A. performed STARR-seq. C.N., C.D.A., L.S., O.H. and M.P. performed PRO-seq, and C.D.A. performed STAP-seq. A.B. and C.H.L. performed MED1 ChIP-seq, and A.B. and J.E.H. performed and analysed MED1 IF with RNA FISH. K.S. performed the MS experiments (under the supervision of K.M.). C.N., L.S., C.D.A., M.P. and O.H. performed qPCR experiments. V.H. and G.L. performed the computational analyses. C.N., V.H. and A.S. interpreted the data and wrote the manuscript. A.S. supervised the project.

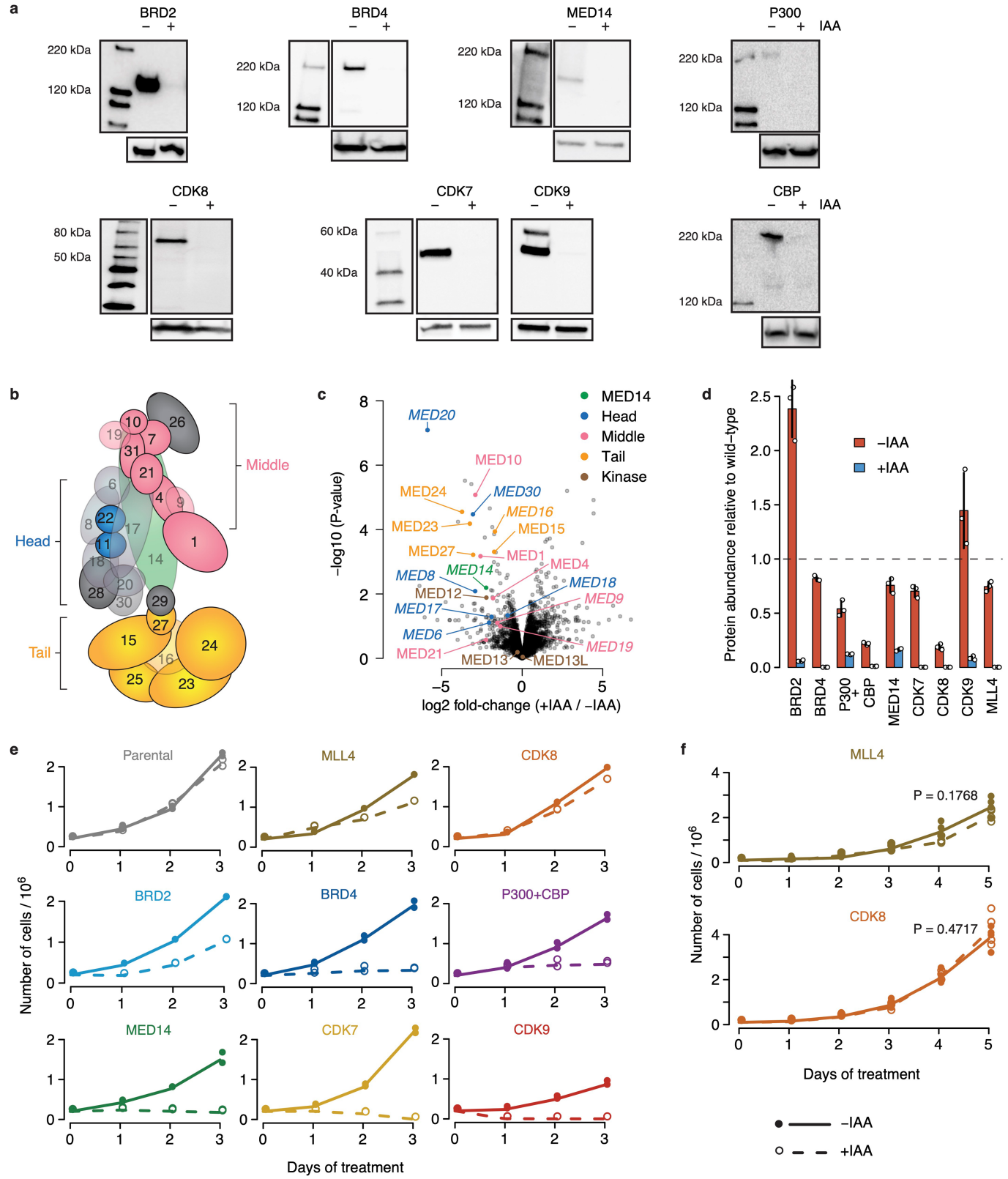
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04779-x.

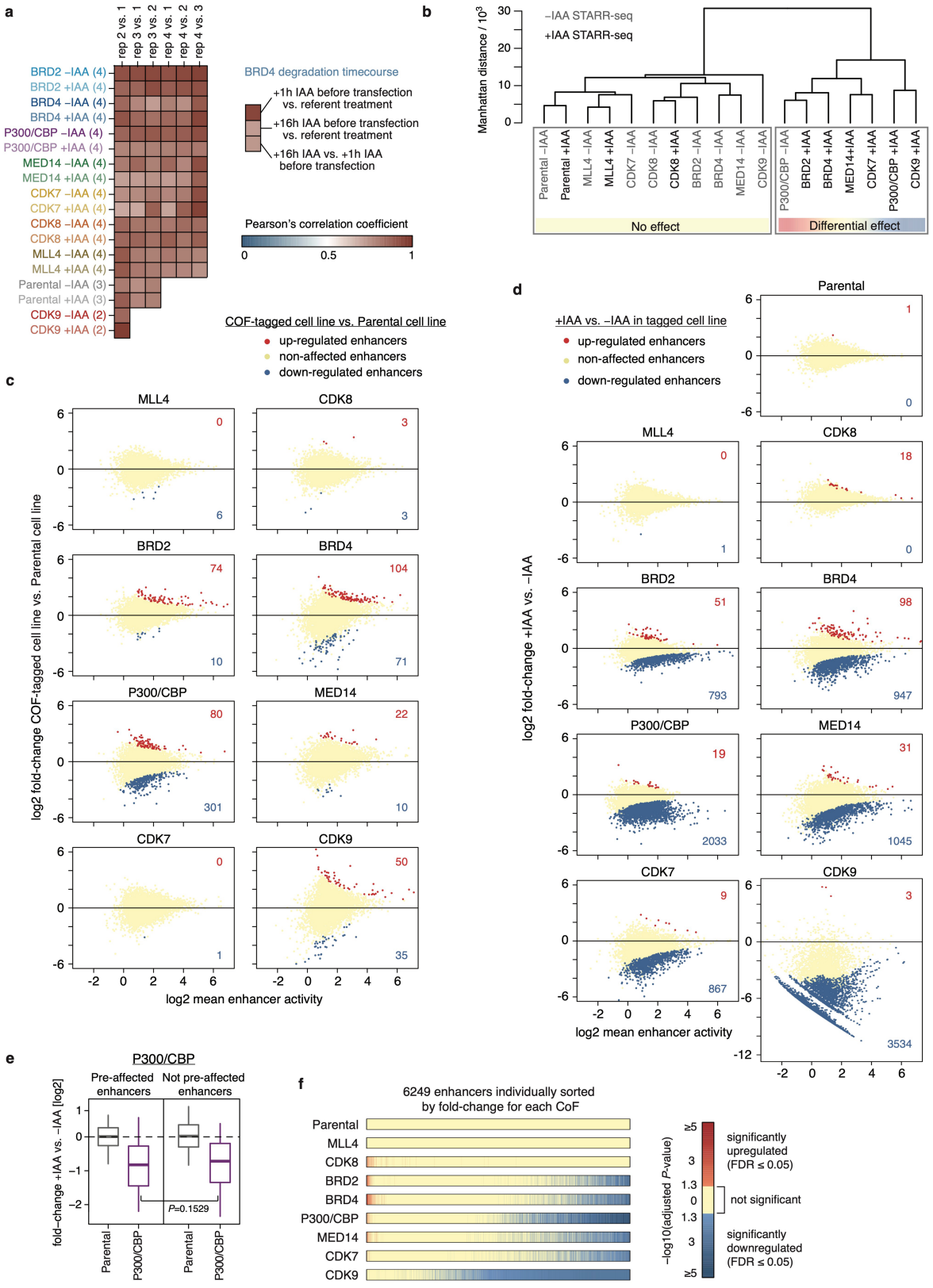
Correspondence and requests for materials should be addressed to Alexander Stark.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.



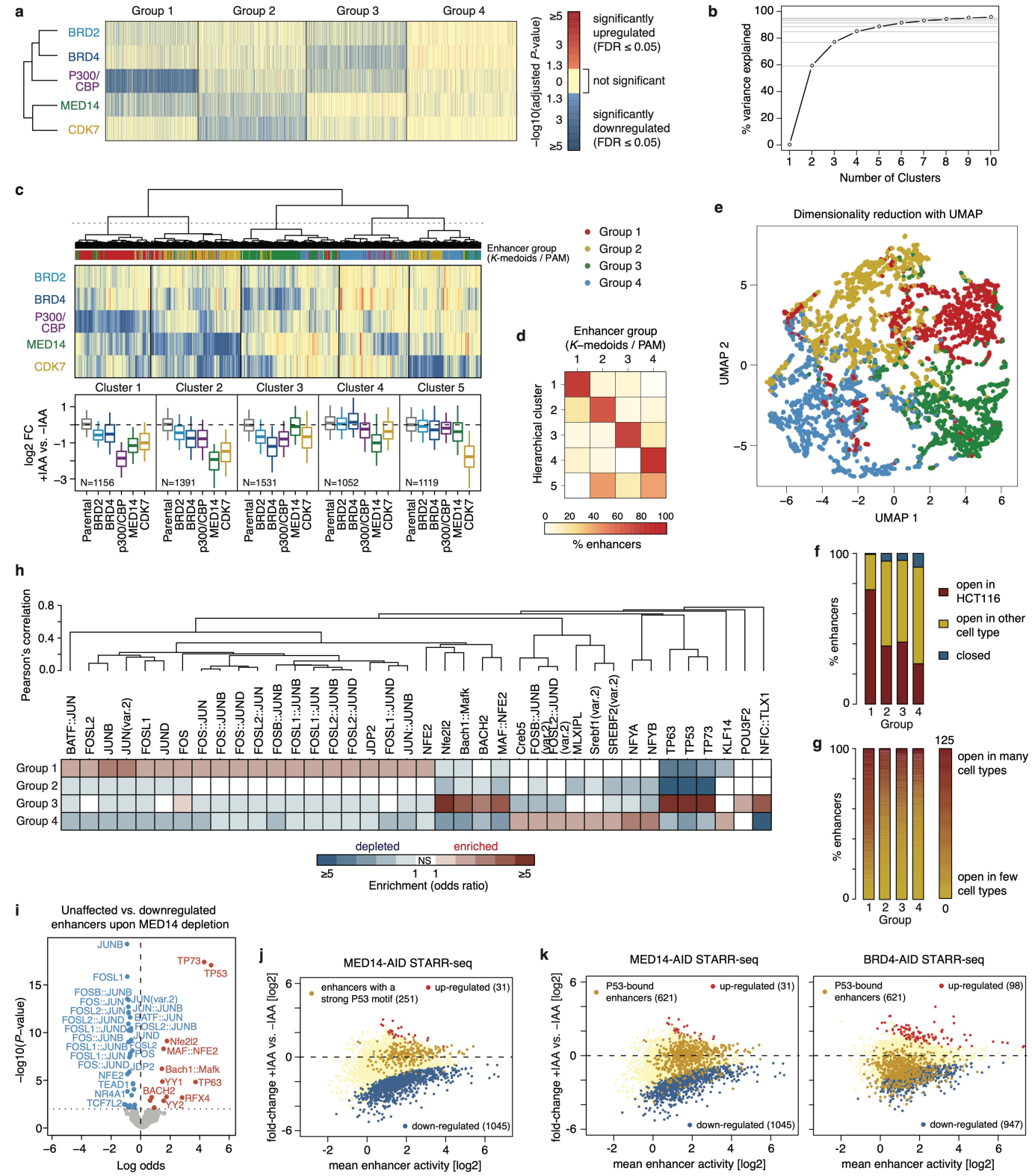
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Validation of cofactor degradation and effect on cell growth. a, Western blots of denoted cofactors (COF) in the cell line where the respective COF is tagged by AID, without and with auxin (IAA) treatment for 1h. done once; validated by mass spectrometry; gel source data: Supplementary Figure 1. **b**, Schematic of the Mediator complex structure with head, middle and tail domains shown in different colors. Core structural subunit MED14 targeted in this study is shown in green. Subunits that cannot be detected anymore in mass-spectrometry upon MED14 depletion are semi-transparent. **c**, Protein abundance change as measured by shot-gun mass-spectrometry upon MED14 depletion by IAA treatment. All detected Mediator subunits are marked and colored according to different Mediator modules/domains shown in b. Subunits marked in italic were not detected anymore (i.e. were below detection limit) in all replicates of IAA treatment. N = 3 independent replicates. **d**, Protein abundance of denoted COFs as measured by targeted mass-spectrometry approach in the cell line where the respective COF is tagged by AID, without and with IAA treatment for 3h. N = 3 independent replicates; mean ± s.d. shown. **e**, Growth curves over a course of 3 days comparing untreated (solid line) and IAA-treated (dashed line) cells, for each COF-AID cell line. N = 2 independent replicates. **f**, Growth curves over a course of 5 days comparing untreated (solid line) and IAA-treated (dashed line) cells for MLL4- and CDK8-AID cell line. N = 3 independent replicates. *P* value of two-sided Student's *t*-test at final day 5 timepoint is shown.



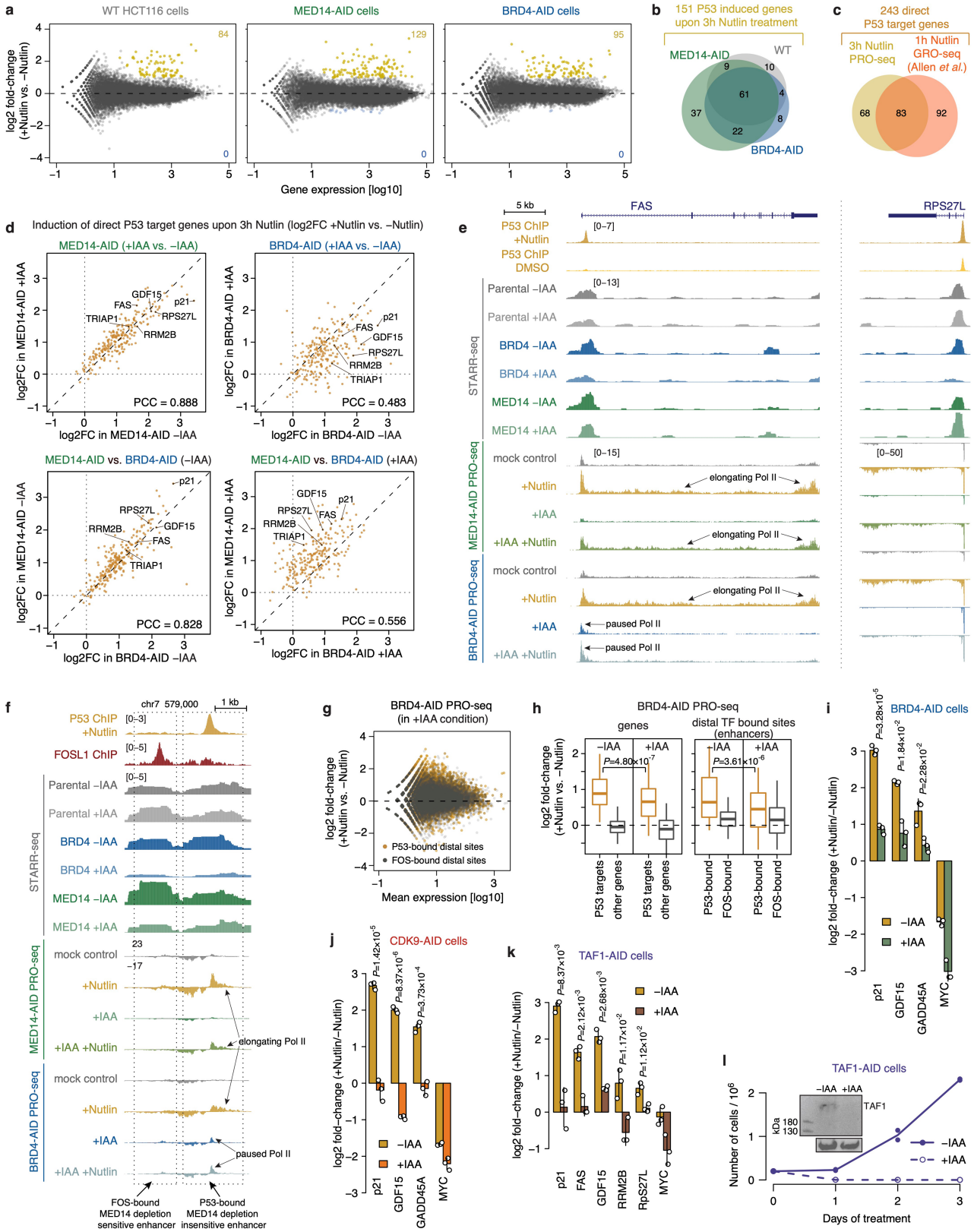
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Effect of cofactor tagging and targeted cofactor degradation on enhancer activity. a, Pearson's correlations for pair-wise comparisons of replicates for each cofactor (COF) with and without IAA treatment calculated across a reference set of 6249 enhancers. For majority of COFs there are 4 independent replicates in each condition, except for our positive and negative controls, CDK9 and the Parental cell line, that have 2 and 3 replicates per condition, respectively. Inset on the right shows correlations between BRD4 samples pre-treated with IAA before STARR-seq library transfection, i.e. with an extended period of protein degradation. **b**, Hierarchical clustering of untreated and IAA-treated Parental and different COF-AID cell lines based on enhancer activity for a reference set of 6249 enhancers. All untreated cell lines (except p300/CBP which shows high level of COF pre-degradation in absence of IAA) cluster together with the Parental cell line, as well as IAA-treated MLL4- and CDK8-AID cell lines. **c**, Differential analysis of STARR-seq enhancer activity between each individual COF-AID cell line and Parental cell line without any treatment to assess the effect of COF-tagging on enhancer activity. Number of significantly up- or down-regulated enhancers is denoted (FDR ≤ 0.05). **d**, Differential analysis of STARR-seq enhancer activity for each COF-AID cell line with and without IAA treatment to assess the effect of COF degradation on enhancer activity. Number of significantly up- or down-regulated enhancers is denoted (FDR ≤ 0.05). **e**, Log2 fold-change in enhancer activity for enhancers pre-affected by P300 and CBP tagging (left; N = 301) and the rest of non-affected enhancers (right; N = 5948) in Parental and P300/CBP-AID cells upon IAA treatment. Boxes: median and interquartile range; whiskers: 5th and 95th percentiles. *P* values: two-sided Wilcoxon rank-sum test. **f**, Significance of change in enhancer activity (*P* values from differential analysis corrected for multiple testing/FDR) for a reference set of 6249 enhancers sorted individually by fold-change in each COF-AID cell line, from unaffected (or upregulated) enhancers on the left to most downregulated enhancers on the right.



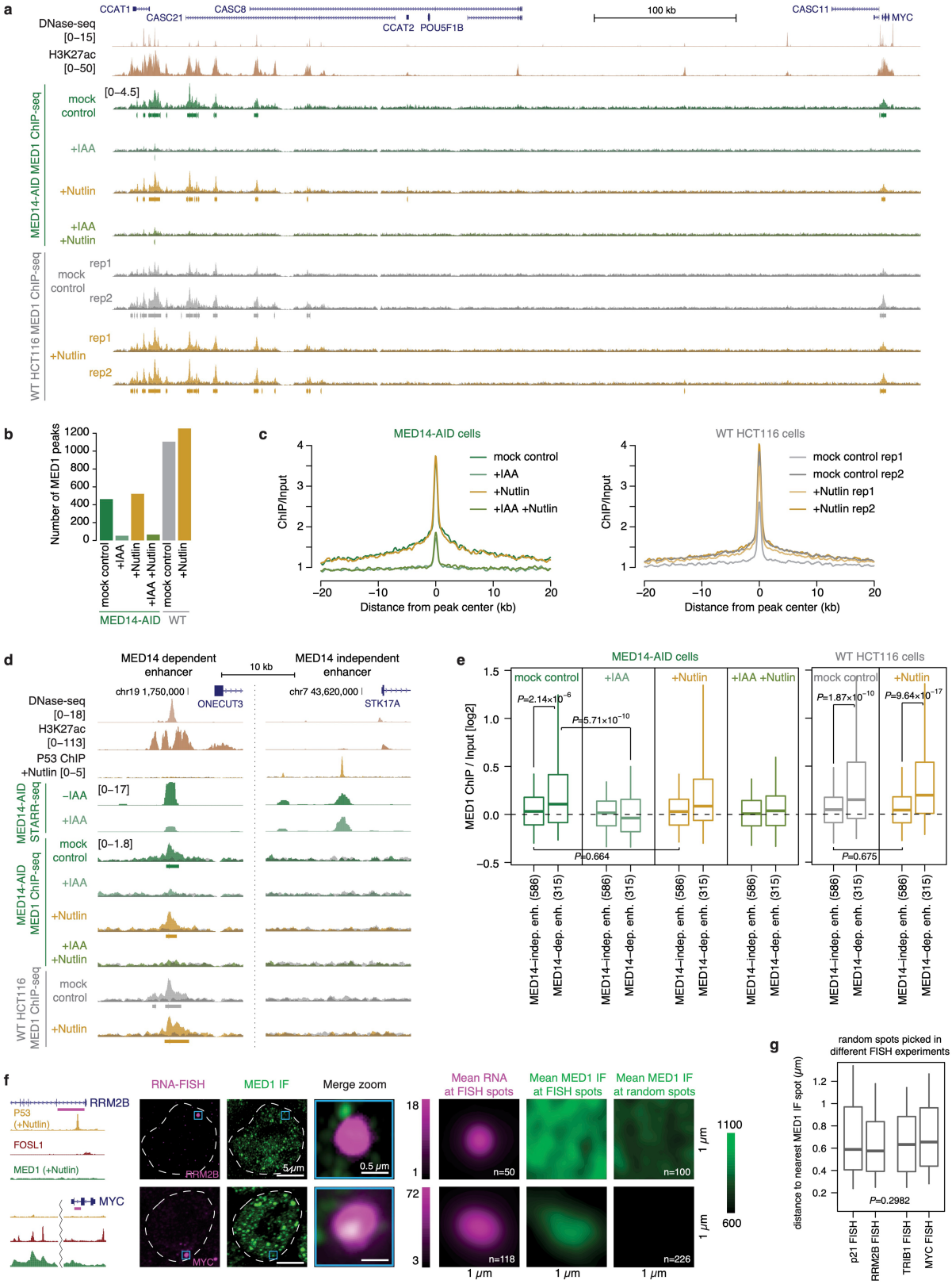
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Features of the four different groups of enhancers. a, Significance of change in enhancer activity (*P* values from differential analysis corrected for multiple testing/FDR) upon individual cofactor degradation for four groups of enhancers defined by PAM (partitioning-around-medoids) clustering. Significant *P* values (FDR≤0.05) for down- and up-regulated enhancers are shown in shades of blue and red, respectively. Non-significant *P* values are shown in yellow. N = 1392, 1660, 1519, 1678 for Groups 1-4, respectively. b, Percent of variance explained by clustering of 6249 enhancers with partitioning around medoids (PAM) algorithm into different number of clusters. Four clusters explain ~85% of the variance. c, Hierarchical clustering of enhancers based on change in enhancer activity upon individual cofactor degradation. Boxplots summarize the log2 fold-change values per COF for each of the 5 clusters defined by cutting the dendrogram as denoted with a dashed line. Enhancer group assignment (from PAM clustering shown in Fig. 2a) is denoted by the coloured stripe below the dendrogram. N = 1156, 1391, 1531, 1052, 1119 for Groups 1-5, respectively. Boxes: median and interquartile range; whiskers: 5th and 95th percentiles. d, Agreement between clusters defined by hierarchical clustering and enhancer groups defined by PAM. For each hierarchical cluster (row) percent of enhancers falling into each PAM enhancer group is shown. e, Two-dimensional visualization of the data after dimensionality reduction with UMAP algorithm. Points represent individual enhancers coloured by their group membership (from PAM clustering). f, Percent of enhancers accessible/open according to DNase-seq in HCT116 cells or in other cell types in the four groups of enhancers defined in Fig. 2a. g, Percent of enhancers accessible/open according to DNase-seq in different number of cell lines ranging from enhancers closed in all cell lines (0 - yellow) to enhancers open in many/all (125 - red) cell lines assayed by DNase-seq in ENCODE. h, Mutual enrichment of transcription factor motifs for the four groups of enhancers. For each motif from the JASPAR vertebrate core collection of 579 non-redundant TF motifs (http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_pfms_jaspar.zip) the enrichment/depletion in each group is assessed against the remaining three groups using two-sided Fisher's exact test and only motifs with *P* value ≤ 0.001 and odds-ratio ≥ 2 are shown. The motifs are hierarchically clustered based on pair-wise Pearson's correlation between motif position-weight matrices (PWMs) to group together similar motifs. A selection of representative motifs from these groups of similar motifs is shown in Fig. 2e. i, Enrichment analysis of 579 non-redundant TF motifs from the JASPAR vertebrate core collection (http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_pfms_jaspar.zip) between unaffected and down-regulated enhancers upon MED14 depletion. Significantly enriched and depleted motifs (two-sided Fisher's exact test; P value ≤0.01) are shown in red and blue, respectively. **j**, Differential analysis of enhancer activity upon MED14 depletion with enhancers containing a P53 motif marked in yellow. **k**, Differential analysis of enhancer activity upon MED14 (left) or BRD4 (right) depletion with enhancers overlapping a P53 ChIP-seq peak marked in yellow.



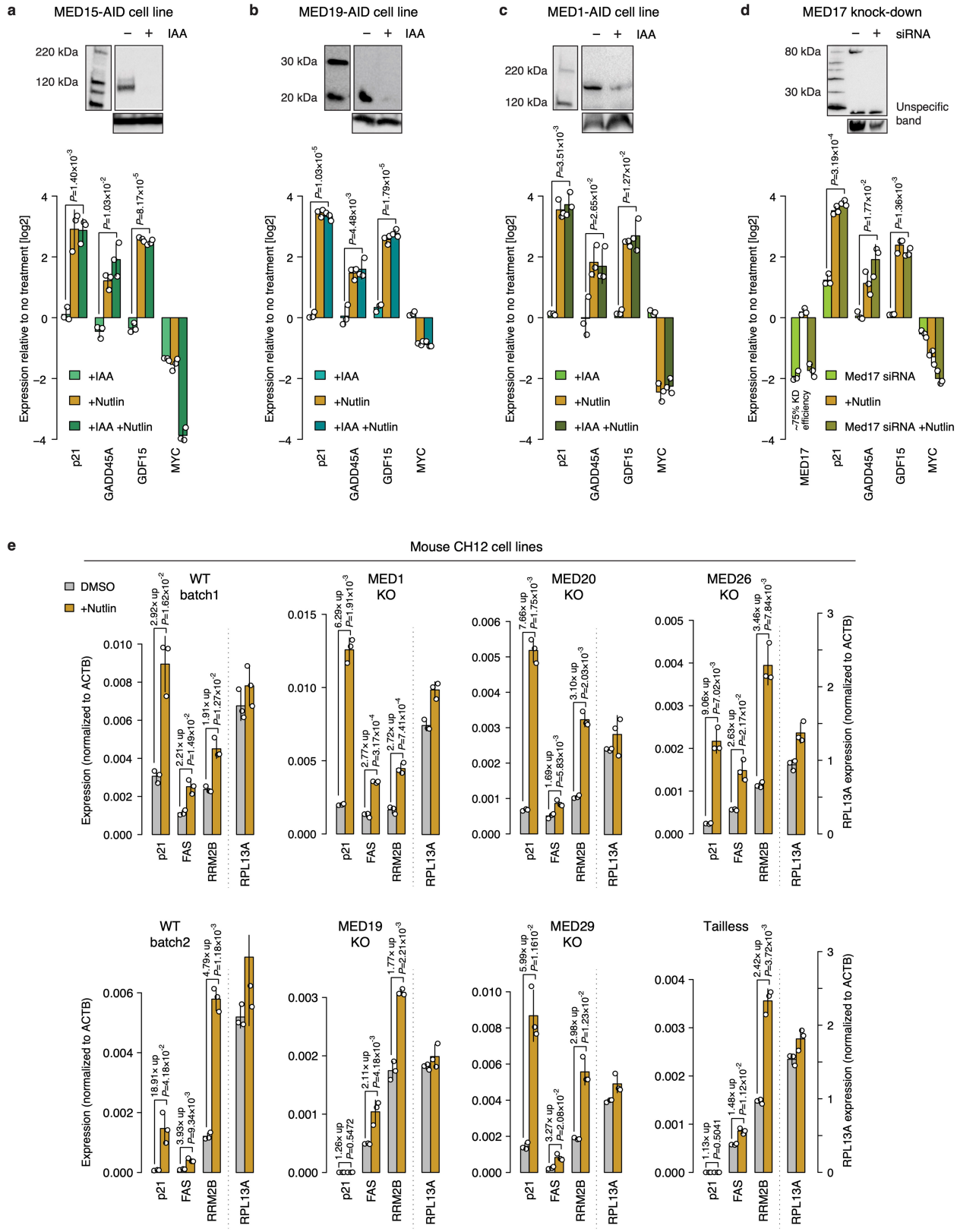
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Induction of P53 target genes and enhancers is insensitive to MED14 depletion, but sensitive to BRD4 depletion. a, Differential analysis of PRO-seq signal at genes between Nutlin-3a-treated and untreated WT HCT116 (left), MED14- (middle) or BRD4-AID (right) cells. Number of significantly upregulated genes in each cell line is denoted in yellow (FDR≤0.05 and fold-change≥2). N = 2 independent replicates for each condition. **b**, Venn diagram showing overlap of significantly upregulated genes in the 3 cell lines shown in panel a, defining in total 151 P53 target genes induced after 3h of Nutlin-3a treatment. **c**, Venn diagram showing overlap of 151 P53 target genes induced after 3h of Nutlin-3a (this study) and 175 P53 target genes defined previously after 1h of Nutlin-3a treatment[76], defining a set of 243 direct P53 target genes used in panels d and h, and in Fig. 3e. **d**, Comparison of induction of direct P53 target genes (defined in panel c) in different cell lines and conditions. Top row compares induction in MED14- (left) or BRD4-AID (right) cells when the respective factor is present (-IAA) or depleted (+IAA). P53 targets are induced to the same extent upon MED14 depletion, but their induction is impeded upon BRD4 depletion. Bottom row compares induction between the two cell lines in the condition without IAA (left) or with IAA (right). Without IAA both MED14- and BRD4-AID cells induce P53 target genes to the same extent, however with IAA the induction in the BRD4-AID cells is impeded compared to the MED14-AID cells. **e**, Loci of the P53 target genes FAS (left) and RPS27L (right) with intronic P53-bound enhancers. Enhancer activity in different COF-AID cell lines with and without IAA treatment is shown (normalized STARR-seq signal for merged replicates), together with nascent transcription (normalized PRO-seq signal for merged replicates) upon induction of P53 signalling with Nutlin-3a in MED14- and BRD4-AID cells with and without IAA treatment. Transcription of both genes is induced upon Nutlin-3a treatment in both conditions with MED14 present (-IAA) or degraded (+IAA), but is strongly reduced with BRD4 degraded due to a pause-release defect that persists upon Nutlin-3a treatment. Activity of their associated P53-bound enhancers is unchanged upon MED14 depletion but is abolished upon BRD4 depletion. **f**, Locus with a FOS-bound MED14-depletion sensitive (left) and a P53-bound MED14-depletion insensitive (right) enhancer. Activity in different COF-AID cell lines with and without IAA treatment is shown (normalized STARR-seq signal for merged replicates), together with nascent transcription (normalized PRO-seq signal for merged replicates) upon induction of P53 signalling with Nutlin-3a in MED14- and BRD4-AID cells with and without IAA treatment. Activity of the FOS-bound enhancer is strongly reduced by both MED14 and BRD4 depletion, whereas the activity of the P53-bound enhancer is unchanged upon MED14 depletion but is abolished upon BRD4 depletion. Endogenous bidirectional transcription of the P53-bound enhancer is induced upon Nutlin-3a treatment in both conditions with MED14 present (-IAA) or degraded (+IAA), but is reduced with BRD4 degraded due to a pause-release defect that persists upon Nutlin-3a treatment. **g**, Differential analysis of PRO-seq signal at distal P53 or FOS bound sites (enhancers) upon Nutlin-3a treatment in IAA-treated BRD4-AID cell line. **h**, Log2 fold-change of PRO-seq signal for direct P53 target genes (left; genes defined in panel c) and distal P53 bound sites around direct P53 target genes (enhancers; right) in BRD4-AID cell line upon Nutlin-3a induction in background with BRD4 present (-IAA) or depleted (+IAA). N = 151, 20964, 244, 359 for P53 targets, other genes, P53- and FOS-bound enhancers, respectively. Boxes: median and interquartile range; whiskers: 5th and 95th percentiles; *P* values: two-sided Wilcoxon rank-sum test. **i**-**k**, Endogenous induction of known P53 target genes with Nutlin-3a as measured by qPCR in BRD4- (**i**), CDK9- (**j**) or TAF1-AID (**k**) cells without or with IAA treatment, i.e. with the respective factor present or degraded. N = 3 independent replicates; fold-change for each replicate calculated independently by dividing the treatment value with the corresponding control value; mean ± s.d. shown; *P* values: two-sided Student's *t*-test. **l**, Growth curves over a course of 3 days comparing untreated (solid line) and IAA-treated (dashed line) TAF1-AID cells. N = 2 independent replicates. Inset shows Western blot for TAF1 in cells without and with IAA treatment for 1h.



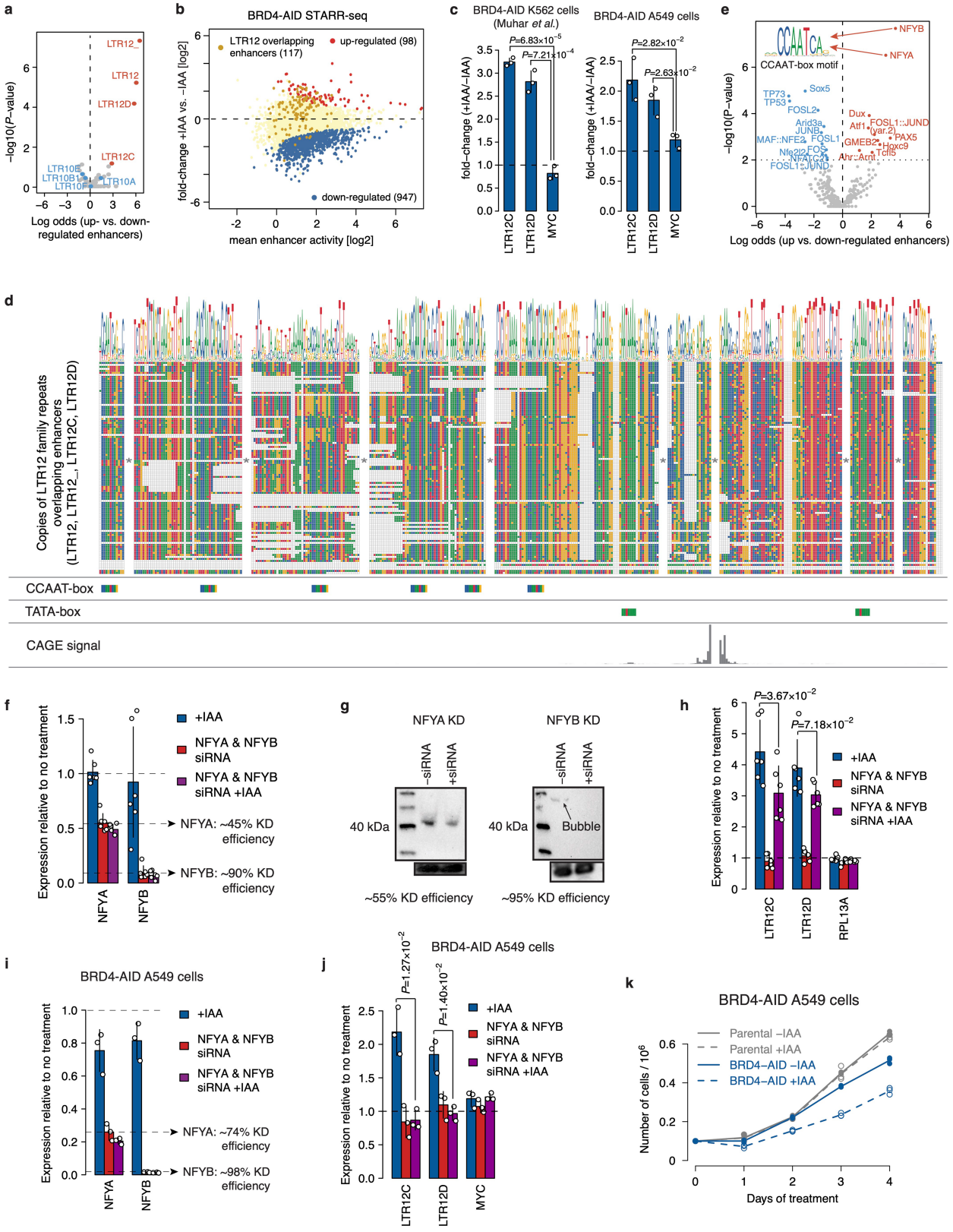
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | P53 target genes and enhancers are not bound by MED1. a, Locus of the MYC gene with an upstream cluster of endogenously active MED1-bound enhancers. ChIP-seq signal and called MED1 peaks in MED14-AID cells treated with IAA or/and Nutlin-3a and in WT HCT116 cells treated with Nutlin-3a are shown. **b**, Number of MED1 peaks called in each condition in MED14-AID and WT cells (MACS2, FDR ≤ 0.05). **c**, Average plot of MED1 ChIP-seq enrichment over input for a common set of MED1 peaks called in MED14-AID (638 peaks; left) and in WT HCT116 cells (1545 peaks; right). **d**, Example of an endogenously active MED14-dependent enhancer bound by MED1 (left) and a P53-bound MED14-independent enhancer not bound by MED1 (right). MED14-dependent enhancer is bound by MED1 in both WT and MED14-AID cells and this binding is abolished upon IAA treatment, i.e. upon MED14 depletion. P53-bound enhancer shows no MED1 binding in any condition, not even upon P53 induction with Nutlin-3a in either WT or MED14-AID cells. **e**, MED1 ChIP-seq enrichment over input for 2 groups of STARR-seq enhancers: 1) MED14-independent, P53-bound enhancers (N = 586) and 2) endogenously open and H3K27ac-marked MED14-dependent enhancers (N = 315), upon Nutlin-3a treatment in control and MED14-depleted MED14-AID cells (left) or in WT cells (right). While MED14-dependent enhancers show some MED1 binding in both WT and MED14-AID cells, which is abolished upon MED14 depletion (i.e. IAA treatment), P53-bound enhancers show no binding in any condition, including after Nutlin-3a treatment when these enhancers are activated. Boxes: median and interquartile range; whiskers: 5th and 95th percentiles. *P* values: two-sided Wilcoxon rank-sum test. **f**, MED1 IF with concurrent RNA FISH against P53 target gene RRM2B (top row) and Mediator-regulated positive control gene MYC (bottom row) in Nutlin-3a-treated WT HCT116 cells. Examples of individual cells with merged view of the FISH and MED1 IF signal at the FISH spot are shown on the left. Hoechst staining was used to determine the nuclear periphery, highlighted with a dashed white line. Mean RNA FISH and mean MED1 IF signal in 1x1μm window centred at FISH spots, or at random spots is shown on the right. Number of spots analysed is indicated in the lower right corner (n). **g**, Distribution of distance between each random spot and the nearest MED1 IF spot for random spots picked in different FISH experiments. Boxes: median and interquartile range; whiskers: 5th and 95th percentiles. *P* value: Kruskal-Wallis rank sum test.



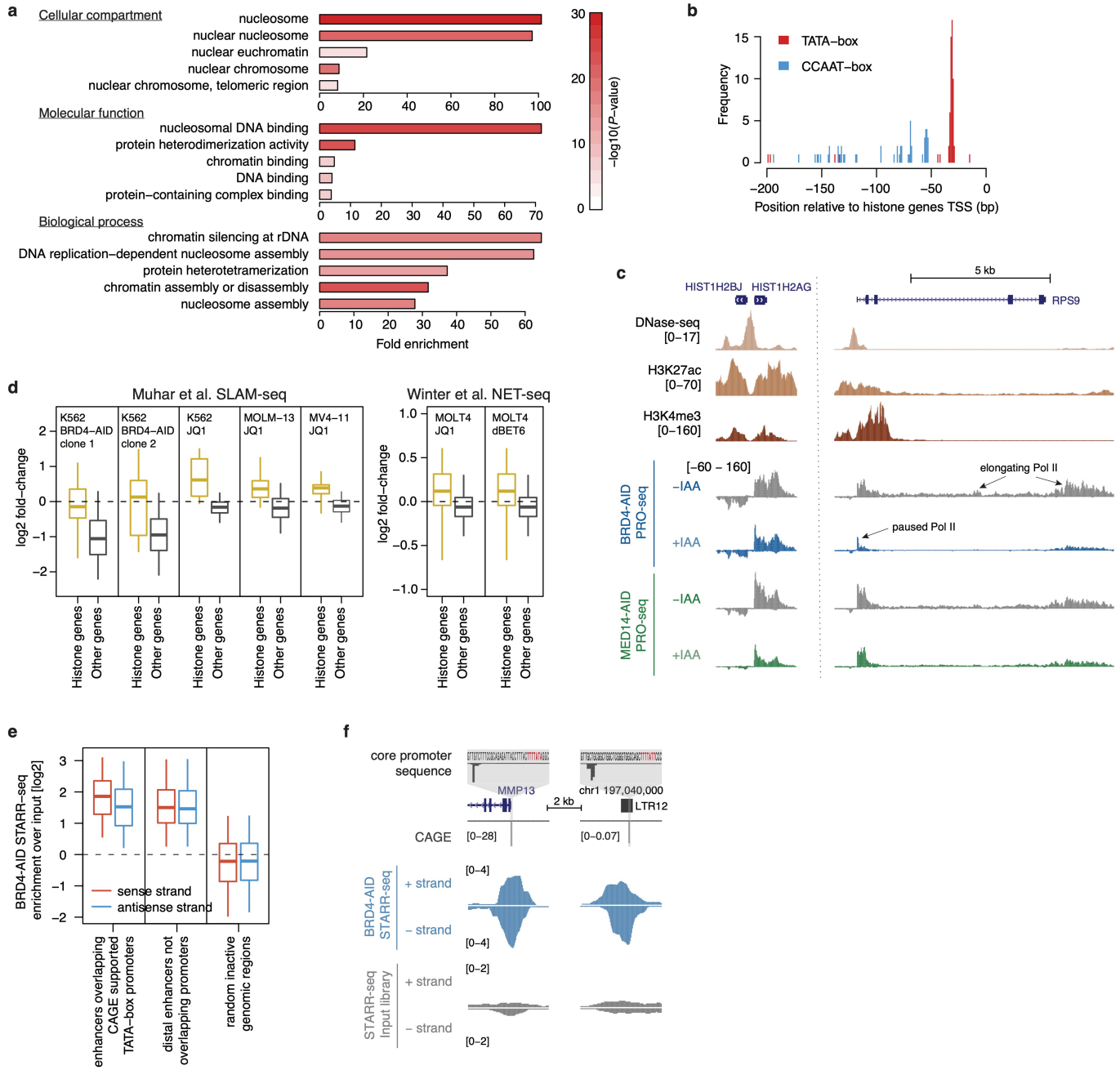
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | P53 target gene induction is independent of multiple Mediator subunits in human and mouse cells. a-c, Endogenous expression of known P53 target genes as measured by qPCR in IAA or/and Nutlin-3a-treated MED15- (**a**, tail module), MED19- (**b**, middle module) or MED1-AID (**c**, middle module) cells. Western blot of the denoted Mediator subunit in the respective COF-AID cell line, without and with IAA treatment for 3h is shown on top. gel s ource data: Supplementary Figure 1. **d**, Endogenous expression of known P53 target genes as measured by qPCR upon Nutlin-3a treatment before and after MED17 (head module) knock-down via RNAi in WT HCT116 cells. **e**, Endogenous expression of P53 target genes as measured by qPCR in DMSO or Nutlin-3a-treated mouse CH12 cells, either wild-type (WT) or knock-out (KO) cell lines for different Mediator subunits (cell lines from ref. [18]). Experiment was performed in two batches (shown in two rows), each time using a re-thawed WT cell line as a control. Tailless = quintuple knock-out for MED15, MED16, MED23, MED24 and MED25 subunits. In **a**-**g**, N = 3 independent replicates; fold-change for each replicate calculated independently by dividing the treatment value with the corresponding control value; mean ± s.d. shown; *P* values: two-sided Student's *t*-test.

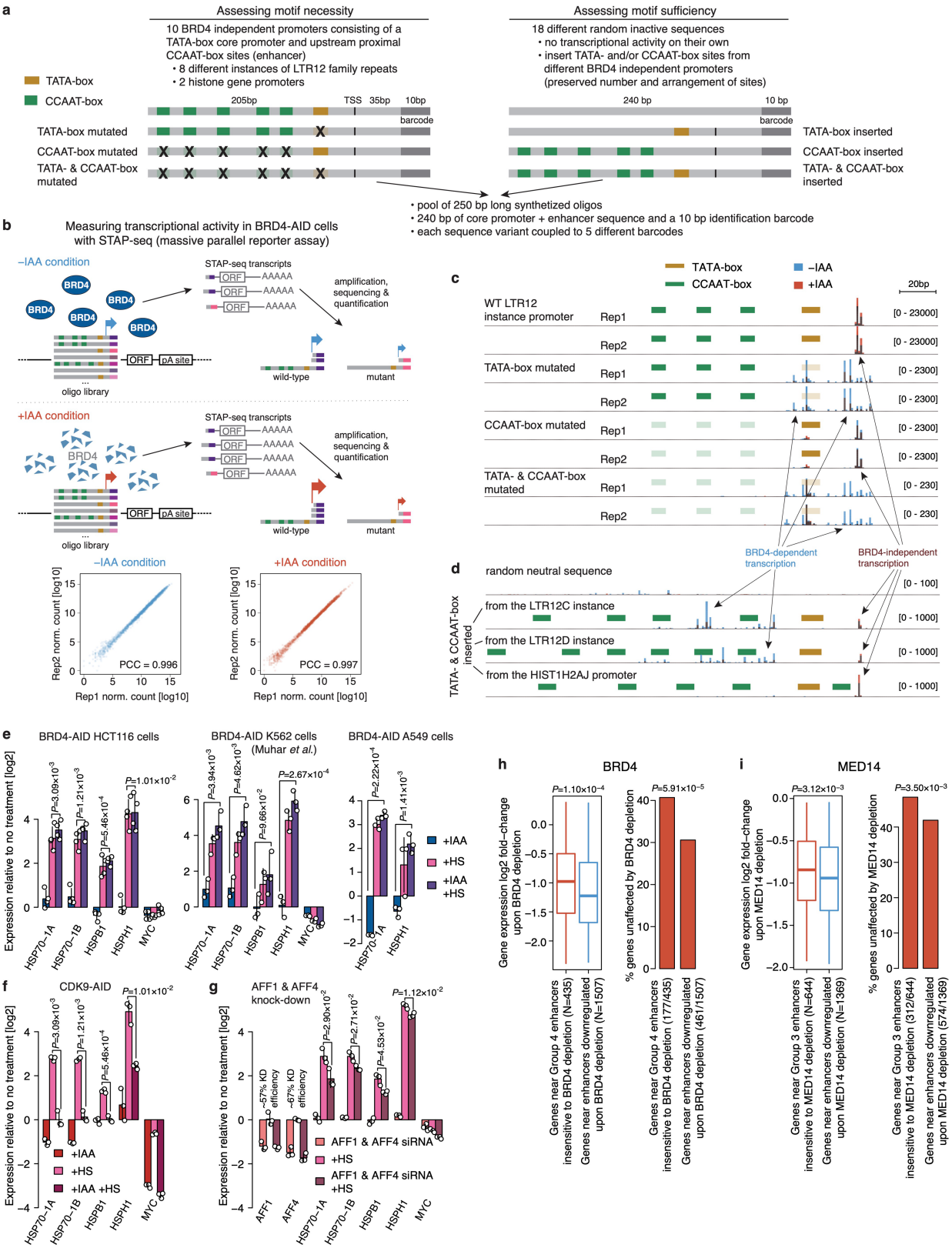


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | LTR12 family repeats act as BRD4 independent enhancers/promoters that contain a combination of TATA-box and multiple CCAAT-box motifs. a, Enrichment of retrotransposons in enhancers up- vs. down-regulated upon BRD4 depletion. b, Differential analysis of enhancer activity upon BRD4 depletion with LTR12 family repeat-overlapping enhancers marked in yellow. **c**, Fold-change of endogenous LTR12 expression as measured by qPCR in IAA-treated vs. untreated BRD4-AID K562 (left) and A549 (right) cells. In both cell lines BRD4 depletion leads to upregulation of LTR12C and D. **d**, Multiple alignment of LTR12 family repeats with detected enhancer activity in STARR-seq. Occurrences of CCAAT-box and TATA-box motifs, and the endogenous transcription initiation previously mapped by CAGE are marked below the alignment. e, Enrichment analysis of 579 non-redundant TF motifs from the JASPAR vertebrate core collection (http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_pfms_jaspar.zip) between upregulated and down-regulated enhancers upon BRD4 depletion in HCT116 cells. Significantly enriched and depleted motifs (two-sided Fisher's exact test; P value ≤0.05) are shown in red and blue, respectively. Logo of the most highly enriched CCAAT-box motif bound by NFYA/B is shown on the right. **f**, Endogenous expression of NFYA and NFYB as measured by qPCR without or with NFYA & NFYB siRNA treatment in BRD4-AID HCT116 cells. **g**, Western blots of NFYA (left) and NFYB (right) with and without treatment with the respective siRNA. gel source data: Supplementary Figure 1. **h**, Endogenous expression of LTR12C and D as measured by qPCR in IAA or/and NFYA & NFYB siRNA treated BRD4-AID HCT116 cells. **i**, Endogenous expression of NFYA and NFYB as measured by qPCR without or with NFYA & NFYB siRNA treatment in BRD4-AID A549 cells. **j**, Endogenous expression of LTR12C and D as measured by qPCR in IAA or/and NFYA & NFYB siRNA treated BRD4-AID A549 cells. **k**, Growth curves over a course of 4 days comparing untreated (solid line) and IAA-treated (dashed line) BRD4-AID and Parental A549 cells. N = 3 independent replicates. In **c**, **f**, **h**, **i** and **j**, mean ± s.d. shown; *P* values: two-sided Student's *t*-test. N = 3 (**c**, **i** and **j**) or N = 6 (**f** and **h**) independent replicates; fold-change for each replicate calculated independently by dividing the treatment value with the corresponding control value.

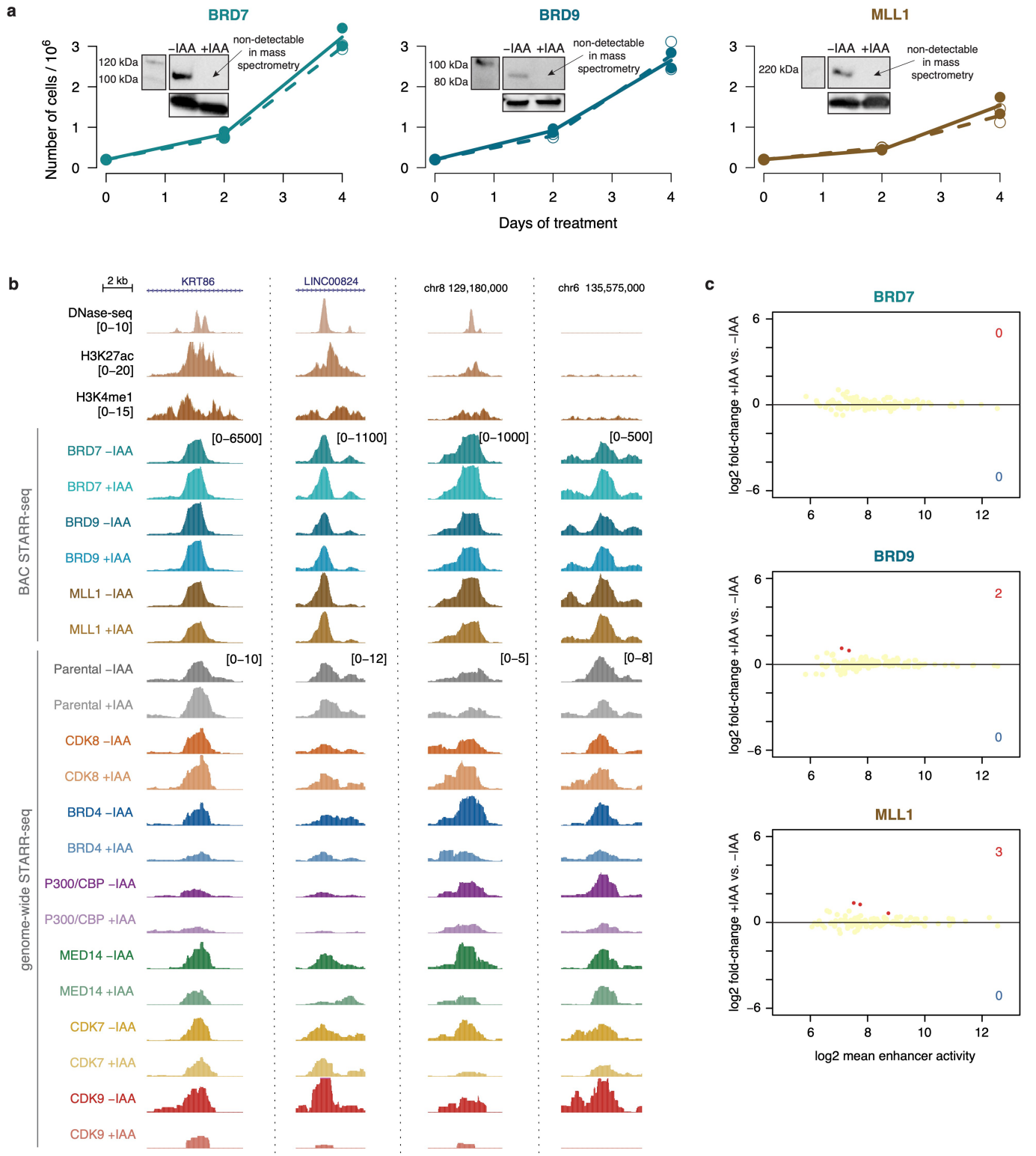


Extended Data Fig. 8 | Histone genes have a promoter with TATA-box and CCAAT-box motifs and do not require BRD4 for productive transcription. **a**, Gene ontology term enrichment for genes with promoters containing both TATA-box and CCAAT-box motifs. Top 5 terms for cellular compartment (top), molecular function (middle) and biological process (bottom) categories are shown. Bars show fold-enrichment and are colored according to the *P* value of the one-sided hypergeometric test. **b**, Occurrence of TATA- and CCAAT-boxes in histone genes promoters relative to TSSs. **c**, Loci of the histone genes HIST1H2BJ and HIST1H2AG (left) and ribosomal protein gene RPS9 (right) with enhancer activity (normalized STARR-seq signal for merged replicates) and nascent transcription (normalized PRO-seq signal for merged replicates) in BRD4- and MED14-AID cells with and without IAA treatment. While RPS9 shows typical pause release defect with loss of RNA polymerase II signal throughout the gene body and increase at the promoter, the two histone genes do not lose signal in the gene body and still have high levels of actively elongating RNA polymerase II. **d**, Log2 fold-change of endogenous nascent transcription for histone genes from previously published datasets. Left: SLAM-seq in different cell lines upon rapid BRD4 degradation via AID system or BRD4 inhibition by JQ1 (from ref. [33]); Right: NET-seq in MOLT4 cell line upon BRD4 inhibition by JQ1 or dBET6 (from ref. [32]). **e**, STARR-seq signal enrichment over input in BRD4-AID cell line separated by strand for enhancers overlapping TATA-box promoters (N = 190), distal enhancers not overlapping promoters (N = 4917) and random inactive regions (negative control; N = 5151). Sense strand corresponds to orientation of the gene for enhancers overlapping promoters and is randomly assigned for distal enhancers and random regions. In **d** and **e**, boxes: median and interquartile range; whiskers: 5th and 95th percentiles. **f**, Examples of STARR-seq enhancers overlapping TATA-box promoters with evidence of endogenous initiation (CAGE): promoter of the MMP13 gene (left) and an instance of LTR12 repeat element (right). STARR-seq signal in BRD4-AID cell line and input library coverage is shown for + and − strands separately. Fragments from both strands are enriched over input, *i.e.* these promoter-overlapping fragments work as enhancers in both orientations.



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

Extended Data Fig. 9 | Combination of a TATA-box core promoter and CCAAT-box-containing proximal enhancer is required and sufficient to drive high levels of BRD4 independent transcription. a, Design of a sequence library to assess the requirement and sufficiency of the TATA-box and CCAAT-box motifs in the core and proximal promoter region, respectively, for the BRD4-independent transcriptional activity with massive parallel reporter assay. For the loss of function approach (left) 10 different BRD4-independent promoters (from LTR12 repeats and histone genes) were selected and variants with either TATA- and/or CCAAT-box motifs mutated were designed. For the gain of function approach (right) the TATA- and/or CCAAT-box motifs from the 10 selected promoters were inserted into 18 randomly picked neutral sequences. Each sequence variant is present in the library 5 times, coupled to a different 10bp long barcode at the 3' end. **b**, Schematic of the massive parallel reporter assay (STAP-seq) to measure transcriptional activity at a single base-pair resolution in BRD4-AID cells without or with IAA treatment. 5' ends of transcripts arising from each sequence present in the library are captured, amplified and sequenced, and the sequenced tags are uniquely mapped to the sequence variant of origin via the 10bp identification barcode. Correlation between transcriptional activity across all sequences in the library measured in two independent replicates for IAA-treated (right) and untreated (left) cells is shown at the bottom. **c**, Transcriptional activity at single base-pair resolution measured by STAP-seq for wild-type (WT) and different mutant versions of the LTR12 promoter instance. Transcription from each sequence variant was assessed 5 times in the library (coupled to 5 different barcodes) and the mean normalized STAP-seq signal across different barcodes is shown for the 2 independent replicates. STAP-seq signal in IAA-treated (red) vs. untreated (blue) BRD4-AID cells is shown as semi-transparent overlay. **d**, Transcriptional activity at single base-pair resolution measured by STAP-seq for a random neutral sequence upon insertion of TATA- and CCAAT-box motifs from an LTR12C, an LTR12D instance or from the HIST1H2AJ promoter. **e**-**f**, Endogenous expression of known heat-shock responsive genes as measured by qPCR in IAA or/and heat-shock treated BRD4-AID HCT116 (left), K549 (middle) and A549 (right) cells (**e**), and CDK9-AID HCT116 cells (**f**). In all three BRD4-AID cell lines heat-shock genes are equally strongly induced with BRD4 present or depleted but fail to get induced with CDK9 depleted. **g**, Endogenous expression of AFF1, AFF4 and known heat-shock responsive genes as measured by qPCR without or with AFF1 & AFF4 siRNA treatment in HCT116 cells. The induction of heat-shock genes is decreased after AFF1 & AFF4 knock-down. In **e**-**g**, N = 3 independent replicates; fold-change for each replicate calculated independently by dividing the treatment value with the corresponding control value; mean ± s.d. shown; *P* values: two-sided Student's *t*-test. **h, i**, Changes in gene expression (log2 fold-change in PRO-seq signal) upon BRD4 (**h**) or MED14 (**i**) depletion for two groups of genes: (1) genes that have an enhancer insensitive to respective COF depletion (Group 4 enhancer for BRD4 or Group 3 enhancer for MED14) and (2) genes that have an enhancer downregulated upon respective COF depletion within 50 kb of their TSS. Number of genes in each group (N) is denoted in parentheses. Boxes: median and interquartile range; whiskers: 5th and 95th percentiles; *P* values: one-sided Wilcoxon rank-sum test. Barplots show percentage of genes in each group that are unaffected (not significantly downregulated) by COF depletion in PRO-seq. *P* values: one-sided Fisher's exact test.



Extended Data Fig. 10 | STARR-seq for additional AID-tagged cofactors shows no effect on enhancer activity. a, Growth curves over a course of four days comparing untreated (solid line) and IAA-treated (dashed line) cells, for BRD7- (left), BRD9- (middle) and MLL1-AID (right) cell line. N = 2 independent replicates. Insets show Western blot for the respective cofactor in cells without and with IAA treatment for 3h. Upon IAA treatment none of the cofactors were detectable either in Western blot or in mass spectrometry. **b**, Examples of four enhancers detected by STARR-seq in the BAC library. For each enhancer the activity in BRD7-, BRD9- and MLL1-AID cell lines in the BAC-STARR-seq screen with and without IAA treatment is shown (normalized STARR-seq signal for merged replicates), alongside with endogenous chromatin accessibility and histone modifications in wild-type HCT116 cells. For comparison, enhancer activity in different COF-AID cell lines from the genome-wide STARR-seq screen is shown. None of the enhancers are affected by the loss of neither BRD7, BRD9 nor MLL1, while they are sensitive to depletion of other COFs (e.g. BRD4, MED14 or CDK9). **c**, Differential analysis of STARR-seq enhancer activity for 114 enhancers detected in the BAC library in each COF-AID cell line with and without IAA treatment to assess the effect of COF degradation on enhancer activity. Number of significantly up- or down-regulated enhancers is denoted (FDR ≤ 0.05). Depletion of none of the three COFs has an effect on enhancer activity, suggesting that they are not required for enhancer activity in the unperturbed HCT116 cells.

nature research

Corresponding author(s): Dr. Alexander Stark

Last updated by author(s): Mar 17, 2022

Reporting Summary

Nature Research 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 Research 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	Deep sequencing base-calling was performed with CASAVA 1.9.1 Immunofluorescence images were acquired using MetaMorph v7.8 acquisition software
Data analysis	Read mapping and post-processing: Bowtie version 1.2.2, samtools version 0.1.19, R/Bioconductor package GenomicAlignments v.1.18.1 Coverage calculation, region intersection: R/Bioconductor package GenomicRanges version 1.34.0 Peak calling: MACS2 2.1.2.1 Differential analysis: R/Bioconductor packages DESeq2 version 1.22.2, edgeR 3.24.3 Coverage and clustering heatmaps: R package gplots version 3.0.1 Motif analysis: R/Bioconductor package seqPattern version 1.14.0 Gene ontology enrichment analysis: R/Bioconductor package GOstats version 2.48.0 Statistics, clustering and data visualisation: R statistical computing environment version 3.5.1 Gene annotation source and reference genome assembly: Ensembl version 83 for GRCh37/hg19 assembly Genomic coverage tracks visualisation: UCSC Genome Browser (http://genome.ucsc.edu) Mass spectrometry data processing and analysis: Proteome Discoverer v2.3.0.522, MSAmanda v2.0.0.9849, Skyline-daily56 v19.0.9.190 All other tasks were performed using custom Unix shell and R scripts, which are available upon request.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All raw deep sequencing data (STARR-seq, PRO-seq, ChIP-seq and STAP-seq) and associated processed data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession number GSE156741.

Previously published datasets for HCT116 cells re-analyzed in this study are available:

1) in the GEO repository under the following accession numbers:
GSE100432 (genome-wide STARR-seq input library)
GSE97889 (ATAC-seq)
GSE71510 (H3K4me1, H3K4me3, H3K27ac, SMARCC1 and SMARCA4 ChIP-seq)
GSE51176 (P300 and MLL4 ChIP-seq)
GSE57628 (BRD4 ChIP-seq)
GSE38258 (CDK8 ChIP-seq)
GSE86164 (P53 ChIP-seq)

2) from ENCODE (https://www.encodeproject.org/):
DNase-seq peaks (ENCFF001SQU, ENCFF001WIJ, ENCFF001WIK, ENCFF175RBN, ENCFF228YKV, ENCFF851NWR, ENCFF927AHJ, ENCFF945KJN, ENCFF360XGA)
H3K36me3 ChIP-seq peaks (ENCFF467KXG,ENCFF742ZBG,ENCFF922EIA)
H3K27me3 ChIP-seq peaks (ENCFF237TTT,ENCFF991HKN,ENCFF029ZPV)
H3K9me2 ChIP-seq peaks (ENCFF586SOS,ENCFF808XMV,ENCFF346SOF)
H3K9me3 ChIP-seq peaks (ENCFF751VFZ,ENCFF577FKU,ENCFF909UTX)
JUND ChIP-seq peaks (ENCFF001UDY,ENCFF001UDZ,ENCFF950JTT,ENCFF088WYS)
FOSL1 ChIP-seq peaks (ENCFF001UDW,ENCFF001UDX)

Vertebrate transcription factor motifs collection is available from the JASPAR database:
http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_pfms_jaspar.zip

SwissProt-human database is available at:
https://www.uniprot.org/proteomes/UP000005640

No restrictions on data availability apply.

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	For next-generation sequencing experiments we performed 2 to 4 biological replicates / independent transfections for STARR-seq, 2 biological replicates for PRO-seq, 1 or 2 biological replicates for ChIP-seq, and 2 biologcial replicates for STAP-seq, which is sufficient for statistical differential analysis with edgeR (Robinson et al. Bioinformatics 2009) or DEseq2 (Love et al. Genome Biology 2014). For single gene qPCR assays 3 to 6 biological replicates were performed, to have a minimum of 3 replicates for performing statistical analysis and obtaining P-values with Student's t-test.
Data exclusions	No data was excluded.
Replication	The reproducibility of experimental findings was verified by performing independent biological replicates. STARR-seq: 2 to 4 biological replicates; PRO-seq: 2 biological replicates; ChIP-seq: 2 biological replicates; STAP-seq: 2 biologcial replicates; qPCR experiments: performed at least 2 times independently. All replication attempts were successful.
Randomization	Not relevant because the samples were not grouped.
Blinding	Not relevant because the samples were not grouped.

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
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

V5-tag Mouse 1:1000 Thermo Fisher # R960-25; BRD4 Rabbit 1:1000 Cell Signaling #13440; P300 Rabbit 1:5000 Abcam #10485; CBP Rabbit 1:2000 Thermo Fisher PA5-27369; NFYA Mouse 1:200 Santa Cruz sc-17753; NFYB Rabbit 1:500 Thermo Fisher PA5-31913; Anti-MYC Mouse 1:2000 Millipore #05-724; Anti-mouse IgG HRP-linked 1:10000 Cell Signaling #7076S; Anti-rabbit IgG HRP-linked 1:10000 Cell Signaling #7074; MED1/TRAP220 Bethyl #A300-793A; Anti-Rabbit IgG Alexa Fluor 488 Life Technologies A11008

Validation

V5-tag Mouse 1:1000 Thermo Fisher # R960-25: This antibody is functionally tested against 20 ng of an E. coli expressed fusion protein containing a V5 epitope using a chemiluminescent substrate at a 1 minute exposure. This antibody has also been tested in Western blot against 25 ng of recombinant Positope™ protein. Using chemiluminescence as the detection method, no cross-reactivity has been observed in bacterial lysates. In mammalian lysates, a few cross-reactive proteins have been observed upon overexposure of blots.

BRD4 Rabbit 1:1000 Cell Signaling #13440: This antibody has been validated using SimpleChIP® Enzymatic Chromatin IP Kits. Masahiro Wakita, et. al. Nature Communication 2020; Yang et al. Am J Cancer Res. 2020

P300 Rabbit 1:5000 Abcam #10485: This antibody has been optimized for IP, WB and IHC-P. Several publications utilized this antibody: Mahmud Z et al. Biochem Pharmacol 163:391-403 (2019);Liu X et al. J Immunol 203:1548-1559 (2019)

CBP Rabbit 1:2000 Thermo Fisher PA5-27369: This antibody has been optimized for IP, WB and ChIP. Several publications validated and utilized this antibody: Martella, C., et al. 2020 J Virol 94(8):e02171-19; Castellaro, A.M., et al 2019 Cancers 11(2):189

NFYA Mouse 1:200 Santa Cruz sc-17753: Bellutti, F., et al. 2018. Cancer Discov. 8: 884-897; Deng, Y., et al. 2019. Mol. Pharmacol. 95: 507-518.; Dolfini, D., et al. 2019. Sci. Rep. 9: 12955; Junjappa, R.P., et al. 2019. Cancers 11 pii: E974.

NFYB Rabbit 1:500 Thermo Fisher PA5-31913: This antibody has been tested in IP, WB, ChIP, ICC/IF, IHC by Thermo Fisher (https://www.thermofisher.com/antibody/product/NFYB-Antibody-Polyclonal/PA5-31913).

This Anti-Myc Tag Antibody, clone 4A6 is validated for use in ChIP, IC, IF, IP, WB for the detection of Myc Tag Kong, X et al. Molecular medicine reports 11 2443-8 2015; Zhang, S; Nature neuroscience 18 386-92 2015

Anti-mouse IgG HRP-linked 1:10000 Cell Signaling #7076S: This product is thoroughly validated with CST primary antibodies and will work optimally with the CST western immunoblotting protocol, ensuring accurate and reproducible results. Fei Guo et al. 2020 Int J Oncol; Wang et al. 2020 Mol Med Rep; Ma et al. 2020 Mol Med Rep

Anti-rabbit IgG HRP-linked 1:10000 Cell Signaling #7074: This product is thoroughly validated with CST primary antibodies and will work optimally with the CST western immunoblotting protocol, ensuring accurate and reproducible results. Darvishi et al. 2020 Neoplasia; Zheng et al. 2020 Meol Med Rep; Ma et al. 2020 Mol Med Rep

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

ATCC HCT116 (ATCC® CCL-247™)
ATCC A549 (ATCC® CCL-185™)
ATCC K546 (ATCC® CCL-243™)
CH12 Mediator subunits KO cell lines were obtained directly from the authors: El Khattabi, L. et al. 2019 Cell 178:1145-1158.e20

Authentication

As HCT116, A549 and K546 cell lines were purchased directly from ATCC, visual inspection was used to confirm the morphology (compared to pictures provided by the vendor).

Mycoplasma contamination	All cell lines are negative for mycoplasma
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

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.*	NCBI GEO public database - accession GSE156741 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156741
Files in database submission	BigWig files with ChIP-seq or Input coverage: GSM4743283_Med14_msAuxin_Input.bw GSM4743284_Med14_msAuxin_MED1_ChIP.bw GSM4743285_Med14_msAuxin_Nutlin_Input.bw GSM4743286_Med14_msAuxin_Nutlin_MED1_ChIP.bw GSM4743287_Med14_psAuxin_Input.bw GSM4743288_Med14_psAuxin_MED1_ChIP.bw GSM4743289_Med14_psAuxin_Nutlin_Input.bw GSM4743290_Med14_psAuxin_Nutlin_MED1_ChIP.bw GSM5580156_WT_HCT116_DMSO_Input_rep1.bw GSM5580157_WT_HCT116_DMSO_MED1_ChIP_rep1.bw GSM5580158_WT_HCT116_DMSO_Input_rep2.bw GSM5580159_WT_HCT116_DMSO_MED1_ChIP_rep2.bw GSM5580160_WT_HCT116_Nutlin_Input_rep1.bw GSM5580161_WT_HCT116_Nutlin_MED1_ChIP_rep1.bw 3 GSM5580162_WT_HCT116_Nutlin_Input_rep2.bw GSM5580163_WT_HCT116_Nutlin_MED1_ChIP_rep2.bw ChIP-seq peaks files: GSM4743284_Med14_msAuxin_MED1_ChIP.narrowPeak.gz GSM4743286_Med14_msAuxin_Nutlin_MED1_ChIP.narrowPeak.gz GSM4743288_Med14_psAuxin_MED1_ChIP.narrowPeak.gz GSM4743290_Med14_psAuxin_Nutlin_MED1_ChIP.narrowPeak.gz GSE156736_WT_HCT116_DMSO_peaks.narrowPeak.gz GSE156736_WT_HCT116_Nutlin_peaks.narrowPeak.gz
Genome browser session (e.g. UCSC)	Link to a UCSC Genome Browser session with STARR-seq, ChIP-seq and PRO-seq data: https://genome.ucsc.edu/s/vhaberle/Neumayr_2020_COF_AID_STARRseq

Methodology

Replicates	1 (MED14-AID cell line) or 2 (WT HCT116 cells) biological replicates were perfomed for MED1 ChIP-seq.
Sequencing depth	All ChIP-seq samples were sequenced single-end for 50bp

Sample	Sequenced_reads	hg19_mapped_reads
Med14_msAuxin_MED1_ChIP	33861632	27834699
Med14_psAuxin_MED1_ChIP	41992502	34491083
Med14_msAuxin_Nutlin_MED1_ChIP	41605085	34043031
Med14_psAuxin_Nutlin_MED1_ChIP	46637301	38613324
Med14_msAuxin_Nutlin_Input	42437526	35584339
Med14_psAuxin_Input	40140290	33712525
Med14_msAuxin_Nutlin_Input	60048260	50386519
Med14_psAuxin_Nutlin_Input	35526938	29733956
WT_HCT116_DMSO_MED1_ChIP_rep1	41271565	33908509
WT_HCT116_DMSO_MED1_ChIP_rep2	43469813	35924100
WT_HCT116_Nutlin_MED1_ChIP_rep1	42264882	35012496
WT_HCT116_Nutlin_MED1_ChIP_rep2	44557702	36929187
WT_HCT116_DMSO_Input_rep1	45435562	37917322
WT_HCT116_DMSO_Input_rep2	47300612	39790373
WT_HCT116_Nutlin_Input_rep1	44244989	37034855
WT_HCT116_Nutlin_Input_rep2	37901513	31886143

Antibodies	MED1/TRAP220 Bethyl #A300-793A
Peak calling parameters	Read mapping with Bowtie: bowtie -p 8 -f -v 3 -m 1 --best --strata

Peak calling with MACS2:
macs2 callpeak -t Med14_msAuxin_MED1_ChIP.bed -c Med14_msAuxin_Input.bed -f BED -g hs --outdir peaks/ Med14_msAuxin_MED1_ChIP -n Med14_msAuxin_MED1_ChIP

Data quality

Sequencing data quality was accessed with FastQC version 0.11.8.

Software

Base-calling: CASAVA 1.9.1
Read mapping: Bowtie version 1.2.2.
Peak calling: MACS2 2.1.2.1
Coverage calculation: R/Bioconductor package GenomicRanges version 1.34.0.
Statistics and data visualisation: R statistical computing environment version 3.5.1
All other tasks were performed using custom Unix shell and R scripts, which are available upon request.