Comparative cofactor screens show the influence of transactivation domains and core promoters on the mechanisms of transcription

Charles C. Bell[1,2,3]*, Jesse J. Balic[1,2#], Laure Talarmain[1,2#], Andrea Gillespie[1], Laura Scolamiero[1,2], Enid Y.N. Lam[1], Ching-Seng Ang[4], Geoffrey J. Faulkner[3,5], Omer Gilan[1,6], Mark A. Dawson[1,2,7,8]*

[1]Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

[2]Sir Peter MacCallum Department of Oncology, University of Melbourne, VIC, Australia

[3]Mater Research Institute, University of Queensland, TRI Building, Woolloongabba, QLD, Australia

[4]Bio21 Mass Spectrometry and Proteomics Facility, The University of Melbourne, Parkville, VIC, Australia

[5]Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia

[6]Australian Centre for Blood Diseases, Monash University, VIC, Australia

[7]Department of Haematology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

[8]Centre for Cancer Research, University of Melbourne, VIC, Australia

∗ Corresponding authors

# Equal contribution

Correspondence:
Professor Mark A. Dawson
Peter MacCallum Cancer Centre,
305 Grattan Street, Melbourne, VIC 3000, Australia
Email: mark.dawson@petermac.org

Dr Charles Bell
Peter MacCallum Cancer Centre,
305 Grattan Street, Melbourne, VIC 3000, Australia
Email: charles.bell@petermac.org

Abstract

Eukaryotic transcription factors (TFs) activate gene expression by recruiting cofactors to promoters. However, the relationships between TFs, promoters and their associated cofactors remain poorly understood. Here, we combine GAL4-transactivation assays with comparative CRISPR-Cas9 screens to identify the cofactors used by nine different TFs and core promoters in human cells. Using this dataset, we associate TFs with cofactors, classify cofactors as ubiquitous or specific, and discover transcriptional co-dependencies. Through a reductionistic, comparative approach, we demonstrate that TFs do not display discrete mechanisms of activation. Instead, each TF depends on a unique combination of cofactors, which influences distinct steps in transcription. In contrast, the influence of core promoters appears relatively discrete. Different promoter classes are constrained by either initiation or pause-release, which influences their dynamic range and compatibility with cofactors. Overall, our comparative cofactor screens characterize the interplay between TFs, cofactors, and core promoters, identifying general principles by which they influence transcription.

Introduction

Regulation of gene expression allows different cell states to arise from a single genome. This process is coordinated by transcription factors (TFs), which use DNA binding domains to recognise DNA sequences, and specialised activation domains (ADs) to recruit the transcriptional cofactors (cofactors) required for gene regulation[1,2]. The recruited cofactors can remodel chromatin, modify histones or act as multi-subunit protein complexes that link with the transcriptional machinery[3]. Importantly, cofactors do not demonstrate DNA sequence specificity and are generally recruited to specific loci by TFs.

Despite decades of research, our understanding of why certain cofactors are needed by different TFs remains incomplete[1,4,5]. Structural approaches, which have been critical in characterising the DNA binding domains of TFs, have been unable to provide insights into function of ADs as they are often unstructured[5,6]. Moreover, since transcription is such a complex process, traditional functional approaches are generally unable to deconvolute how each of the various regulatory inputs (such as the activating TF, collaborating TFs, core promoter or cellular context) influences which cofactors are used. Thus, exploring how TFs and core promoters influence the mechanisms of transcription requires a reductionist approach, where ADs and promoters can be isolated and studied independently of other variables.

In recent years, such systematic, synthetic approaches have provided important insights[7–11]. For example, a recent study assessed the requirement for individual cofactors across thousands of promoters and enhancers, identifying widespread variation in cofactor requirements[8]. However, the reciprocal approach, to define the entire range of cofactors needed by individual TFs or promoters, has yet to be performed[1,12]. Consequently, a number of key questions about cofactor specificity and cofactor-promoter compatibility remain unanswered[1,5,12–18].

Results

Establishing a transcription factor-based screening system

To address this challenge, we developed a screening system consisting of (i) a GAL4-DNA binding domain fused to a transactivation domain of interest and (ii) a reporter containing GAL4 binding motifs upstream of a mCMV promoter driving a fluorescent reporter (Extended Data Fig.1A-B). This reductionistic design controls the promoter and DNA-binding element to isolate how the activation domain alters the cofactors used for transcription. To enable investigation of all transcriptional regulators, including those required for cell survival, our

81   reporter construct contains an unstable GFP[19]. Using this reporter, we observed a dramatic

82   reduction of fluorescent signal within 24 hours (hrs) of transcription inhibition, prior to cell

83   death (Extended Data Fig.1C-D) and detected complete loss of expression upon knockout (KO)

84   of the catalytic subunit of RNA polymerase II (*POLR2A*) (Extended Data Fig.1E),

85   demonstrating our ability to identify common-essential proteins necessary for transcription.

86

87   We decided to investigate the cofactors needed by a representative set of nine transcription

88   factors with diverse and important functions in synthetic biology, development and disease.

89   These include VP64, c-MYB (MYB), EWSR1 (EWS), p65 (NF-κB), p53, IRF1, PU.1,

90   NOTCH and glucocorticoid receptor (GR). The AD regions used were largely unstructured,

91   enriched for acidic, proline and glutamine residues, and previously shown to have activation

92   potential (Extended Data Fig.1B and Supplementary Table 1,2).

93

94   To perform genetic screens using these ADs, we developed an isogenic, constant reporter line

95   by lentiviral integration of the reporter construct at high multiplicity of infection (MOI) into a

96   Cas9-K562 clone. Polyclonal, high MOI integration enabled detection of the unstable GFP,

97   while also reducing the potential for technical artefacts associated with a single integration site.

98   To illustrate the robustness of our approach, firstly, we introduced our reporter into the

99   endogenous *AAVS* safe-harbour locus and validated hits identified from our screens (Extended

100   Data Fig.2A). Secondly, we developed an independent reporter line through a separate viral

101   transduction, which is unlikely to share the same integration sites. We observed a very high

102   correlation (r=0.918, Pearson) between screens performed on lines derived from separate

103   transductions (Extended Data Fig. 2B). Lastly, we replicated the screen using an insulated

104   reporter introduced via piggyBac integration. In this setting, the AD is fused to an artificial

105   zinc-finger protein (ZFP), rather than GAL4, and recruited to an alternative DNA binding

106   sequence (Extended Data Fig.2C). Despite differences in the DNA binding domain and DNA

107   context, the results were highly correlated (r=0.830, Pearson) (Extended Data Fig.2C).

108

109   The nine GAL4-AD constructs were introduced into this constant reporter line, after which we

110   confirmed that GFP signal was completely GAL4 dependent (Extended Data Fig.2D). Some

111   TFs, such as VP64 and MYB, are known to be particularly dependent on certain cofactors. To

112   validate that we could capture this specificity, we confirmed that these ADs were

113   disproportionately affected by loss of MED25 and p300 respectively[20–23] (Extended Data
114   Fig.1E).

115

116   Having established the validity and reproducibility of our platform, we used the nine GAL4-
117   AD reporter lines to perform comparative CRISPR-Cas9 screens with a bespoke guide RNA
118   (gRNA) library targeting 1137 transcriptional regulators and chromatin-associated proteins
119   (Supplementary Table 3). To minimise technical variation, the screens were performed in
120   parallel, and cells with a reduction in transcription (as measured by GFP signal) were harvested
121   at three timepoints (Day 5, Day 6, Day 7 post guide library infection). Isolating at multiple
122   timepoints enabled robust quantitative comparisons of the effects of different cofactors on AD
123   activity (Extended Data Fig.3A, Supplementary Note 1). Altogether, we identified 239 genes
124   in the library as significantly enriched for at least one of the nine ADs (Supplementary Table
125   4). As expected, hits clustered together by STRING analysis, with enrichment for RNA
126   polymerase initiation and elongation, Mediator complex, SWI/SNF components and
127   SET/COMPASS family members (Extended Data Fig.3B-C).

128

129   To confirm that we could separate transcriptional effects from effects on viability, we
130   intersected the screen hits with dropout data from matched samples. Approximately 30% of
131   genes required for cell growth were not significantly enriched in any AD screen (Extended
132   Data Fig.3D). Similarly, integration with DEPMAP, identified that many common-essential
133   cofactors were not enriched for any AD (Extended Data Fig.3E). We also confirmed that there
134   was not substantial dropout of essential genes at the screen timepoints (Extended Data Fig.3F-
135   G), and essential genes did not display more variability relative to non-essential genes
136   (Extended Data Fig.3H). Together this demonstrates that our screens can identify essential
137   transcriptional regulators, divorcing their contribution to transactivation from their requirement
138   for cell viability.

139

140   ***Overview of the specificity of TF-cofactor interactions***
141   To represent this large dataset, we developed a spoke and wheel plot to provide an overview
142   of dependence on key transcriptional regulators across the 9 ADs (Fig. 1). As expected, the
143   entire RNA polymerase II (RNA Pol II) complex is necessary for all of the ADs (90/90 possible
144   enrichments). As are components of the preinitiation complex, such as TFIIA, TFIIB, TFIIE,
145   TFIIF, TFIIH, the FACT complex and DSIF components. Beyond these core transcriptional
146   proteins, other coactivator complexes, such as Mediator, SET/COMPASS, Integrator,

147 chromatin remodellers and transcriptional elongation components display more interesting

148 patterns.

149

150 Reassuringly, our approach identified several previously reported interactions between TFs and

151 cofactors, including MYB-p300[20,24], MYB-TAF12[25], VP64-MED25[21–23], p53-CDK8[26] and

152 NOTCH-WDR5[27] (Fig.1). In addition to these known interactions, several novel TF-cofactor

153 associations were observed. One that was particularly striking was a submodule of Integrator

154 containing INTS5, INTS2 and INTS8[28] , which is especially important for NF-κB activity

155 (Fig.1, Extended Data Fig.4A). To confirm that INTS5 is preferentially needed for the

156 endogenous activity of NF-κB, we treated K562 cells with TNF-α and assessed the effects in

157 control and INTS5 KO cells. By cell surface and ChIP-seq analyses, we also confirmed that

158 INTS5 KO has an impact on endogenous NF-κB activity (Extended Data Fig.4), illustrating

159 that our method was able to identify endogenous cofactor dependencies for specific TFs.

160

161 *Exploring cofactor specificity across activation domains*

162 Our comparative screens not only serve as an important resource, but also provide the

163 opportunity to obtain systematic insights into the transactivation process. To begin with, we

164 explored the relationship between cofactor specificity and quantitative contribution to

165 transcription. In general, we observed that cofactors that contribute broadly to activation by

166 most ADs, tend to display higher enrichment (Fig.2A). Interestingly, we observed very few

167 examples of cofactors with potent and highly selective requirement, with the notable exception

168 of NCOA1, which is a major dependency for Glucocorticoid Receptor-mediated

169 transactivation. This suggests that, in general, TFs do not to have highly specific, dedicated

170 cofactors that contribute strongly to transcription.

171

172 A lack of potent, selective cofactors does not necessarily imply generic mechanisms of

173 transactivation. Instead, our data suggests that rather than using dedicated cofactors, ADs may

174 achieve specificity by using unique combinations of cofactors (Fig.1). To explore this

175 possibility, we first identified heterogeneously enriched cofactors across the 9 screens,

176 anchoring the analysis on cofactors that are not expected to be variable (i.e. RNA Pol II). This

177 analysis identified ~100 cofactors that display heterogeneity in their requirement, some of

178 which were enriched for a large number of ADs, such as p300 or CHD1, and others that are

179 enriched for few ADs, such as SETD1B or NCOA1 (Fig.2B).

180   To identify patterns in our data, we performed Pearson-based hierarchical clustering of the top

181   50 heterogeneous cofactors. The data was bi-clustered to identify ADs with similar patterns of

182   cofactor dependency and cofactors with similar patterns of enrichment across ADs (Fig.2C).

183   Clustering of ADs demonstrates, that for the most part, ADs do not separate into a small number

184   of discrete groups. Interestingly, Glucocorticoid Receptor clustered away from the other ADs,

185   perhaps due to its structured AD (Fig.2C, Extended Data Fig.1B). Similarly, clustering of the

186   cofactors was relatively indiscrete, with no clear co-dependency relationships identified

187   between distinct cofactor complexes. The few co-dependent clusters identified were within

188   large multi-subunit complexes, such as MED16, MED24 and MED25 (Mediator), and INTS2,

189   INTS5 and INTS8 (Integrator) (Fig.2C). Together, this suggests that, while certain ADs utilise

190   submodules of multi-subunit complexes, each AD is dependent upon a different combination

191   of cofactors.

192

193   To provide further confidence in these results, we selected eleven heterogeneous cofactors, and

194   independently quantified their contribution to transactivation by each AD. Overall, with some

195   minor exceptions, we observed a high degree of concordance between the enrichment scores

196   reported by the screen and the reduction in GFP signal in each respective AD line,

197   demonstrating that the heterogeneity is genuine and reinforcing the quantitative nature of the

198   screens (Fig.2D, Extended Data Fig.5). Importantly, the heterogeneity was reproduced using

199   rapid protein degradation and specific inhibitors at a much earlier timepoint (Extended Data

200   Fig. 6). We also confirmed that transcriptional regulation of the GAL4-ADs (Extended Data

201   Fig.7A-C), GAL4-AD protein levels and stability (Extended Data Fig.7D), or GAL4-AD

202   chromatin occupancy (Extended Data Fig.7E), are unlikely to be major contributors to the

203   cofactor heterogeneity observed.

204

205   Taken together, our screen and validation data illustrate that cofactor dependence across

206   different ADs does not conform to simple patterns. Instead, our data suggests that each AD

207   uses a unique set of cofactors to activate transcription (Fig.1, Fig.2C-D).

208

209   ***Co-dependency between Mediator tail 2 and the kinase module***

210   While cofactors from distinct chromatin complexes did not show clear co-dependencies, our

211   Pearson-based clustering identified highly correlated, co-dependent cofactors within multi-

212   subunit complexes (Fig.1, Fig.2C). Clustering of all ~100 heterogeneous cofactors identified

213   strong correlations within tail 2 of the Mediator complex (MED16, MED23, MED24, MED25),

214   CDK12 and its associated cyclin CCNK, subunits of the Integrator complex (INTS2, INTS5,

215   INTS8) and PAF1/WDR61 (Extended Data Fig.8A-B).

216

217   Due to its key role in transcription, we were particularly interested in co-dependencies observed

218   within the Mediator complex. Mammalian Mediator is organised into 5 major modules – head,

219   middle, tail 1 (upper tail), tail 2 (lower tail) and sub-stoichiometric kinase module[29,30] (Fig.3A).

220   Whilst structural composition has been well-characterised, functional relationships within the

221   complex remain poorly understood. Our data suggests that the head, middle and tail 1 modules

222   are ubiquitously enriched (Fig.3A), with most of the functional heterogeneity observed in the

223   tail 2 module, kinase module and MED1, which is structurally positioned proximal to tail 2

224   (Fig.3A).

225

226   We noted that enrichment of tail 2 subunits is strongly correlated, and that they correlate with

227   the sub-stoichiometric kinase module (Fig.2C-D, Fig.3A, Extended Data Fig.8B). VP64, EWS,

228   NF-κB and p53 are highly dependent on tail 2 and the kinase module, while MYB and

229   Glucocorticoid Receptor are largely tail 2 independent and show minimal requirement for

230   kinase module subunits (Fig.2C-D, Fig.3A, Extended Data Fig.6A). To test whether this co-

231   dependency extends beyond our screening system, we performed RNA Pol II ChIP-seq upon

232   genetic deletion of tail 2 subunits (MED16, MED24, MED25), the kinase module (CDK8,

233   MED12, CCNC) and a core structural subunit (MED14) and assessed whether the effects were

234   correlated at endogenous genes. As expected[29,31], loss of the core subunit (MED14) led to a

235   marked, relatively uniform decrease in RNA Pol II levels (Fig.3B). In contrast, loss of tail 2

236   subunits (MED16, MED24, MED25) or the kinase module (CDK8, MED12, CCNC) resulted

237   in disproportionate effects on particular subsets of genes (Fig.3B)[32]. Importantly, the effects of

238   disrupting individual tail 2 and kinase subunits were highly correlated, with clear concordance

239   between the rank order of genes effected by MED16, MED24, MED25, MED12, CDK8 and

240   CCNC loss (Fig.3B-C). Consistent with our screens, correlation is strongest within tail 2, with

241   a minority of tail 2 dependent genes, not dependent on CDK8 and CCNC (Fig.3B-C).

242   Importantly, the co-dependency is highly specific to tail 2 and the kinase module, as there was

243   minimal correlation with disruption of the core (MED14 KO) (Fig.3B-C).

244

245   To confirm these results are likely due to direct functional interplay between these submodules,

246   we developed dTAG lines[33] to rapidly degrade MED12, MED14 and MED25 (Fig.3D,

247 Extended Data Fig.8C). Using these degron lines and a CDK8 inhibitor, we monitored RNA

248 Pol II chromatin occupancy upon rapid perturbation of these subunits (Fig.3E and Extended

249 Data Fig.8D-F). Consistent with our CRISPR KO results, we confirmed a high correlation

250 between perturbations of tail 2 and kinase module subunits (MED12, MED25 and CDK8,

251 r=0.6–0.73, Pearson), with markedly less correlation with loss of the core (MED14, r=0.33–

252 0.45, Pearson) (Fig.3E). Taken together, our comparative screens uncover a previously

253 uncharacterised functional association within tail 2, and between tail 2 and the kinase module

254 of the Mediator complex.

255

256 ***TFs and cofactors influence different steps in transcription***

257 As transcription is a multistep process that involves promoter opening, initiation, pausing,

258 elongation and termination, we hypothesised that TFs may recruit diverse cofactors to influence

259 different steps in transcription. To explore this hypothesis, we used ChIP-nexus[34] to precisely

260 map RNA Pol II on the reporter construct in each of the GAL4-AD cell lines. Remarkably,

261 when the reporter is activated by different ADs, we observed differences in the ratio of RNA

262 Pol II within the gene body, relative to the amount adjacent to the TSS (Fig.4A), suggesting

263 that TFs have different capacities to facilitate RNA Pol II initiation and elongation.

264

265 Importantly, the degree of elongating RNA Pol II was associated with enrichment of cofactors

266 involved in RNA Pol II pause-release (Fig.4A). Notably, the two ADs with the highest

267 elongation ratios, VP64 and NF-κB, were most dependent on major pause-release regulators,

268 NELF and Integrator (Fig.4A). Interestingly, each of these ADs were most dependent on

269 different regulators of pausing (Fig.4A), supporting the prospect of multiple, independent

270 pause-release checkpoints[35]. We also observed that other regulators of elongation, such as

271 CDK8, CCNC, CDK9, CDK12, ELL1, PAF1 and CDC73, displayed a general correlation

272 between screen enrichment and the proportion of elongating RNA Pol II (Fig.4A).

273

274 While no cofactor alone is predictive of elongation potential, we were struck by the association

275 between the proportion of elongating RNA Pol II and dependence on Mediator kinase (CDK8)

276 and its associated cyclin (CCNC) (Extended Data Fig.9A-B). Based on the co-dependency

277 between CDK8 kinase module and tail 2 of Mediator (Fig.2, 3), we considered the prospect

278 that some TFs, such as VP64, interact with subunits in tail 2 to engage the kinase module to

279 potentiate elongation[21]. To test this idea, we assessed RNA Pol II in the VP64 cell line treated

280   with CDK8 inhibitor. Here, CDK8 inhibition reduced the ability of the VP64-AD to potentiate

281   elongation (Fig.4B) whereas elongation was unperturbed by CDK8 inhibition in cells

282   expressing MYB-AD (Fig.4B), which is not reliant on CDK8 or tail 2 for transactivation

283   (Fig.2C-D, Fig.3A). Together these findings suggest that these submodules of Mediator are

284   preferentially used by certain TFs to facilitate transcriptional elongation[36].

285

286   *Exploring the influence of core promoters on transcription*

287   Transactivation requires cofactors to converge onto a promoter where RNA Pol II is loaded.

288   Our screens suggest that TFs recruit diverse sets of cofactors to potentiate different steps in

289   transcription. However, it remains unclear how the core promoter influences which cofactors

290   are needed. To address this, we adapted our screening system to vary the core promoter while

291   maintaining a constant activation domain (Nf-KB) (Extended Data Fig.10A). We chose to

292   study promoters containing different well-characterised core promoter elements: (i) TATA

293   box, (ii) TATA-like element with reduced affinity for TBP[37–39], (iii) Initiator sequence (Inr)

294   and/or (iv) polypyrimidine Initiator (TCT) sequence (Extended Data Fig.10B). These core

295   motifs are associated with different classes of genes, suggesting that they influence how their

296   associated genes are regulated[40]. Many genes containing a TATA box have focussed promoters

297   that are tissue specific, have a large dynamic range and can be rapidly induced[41–45]. In contrast,

298   the TCT element is present in housekeeping genes, primarily ribosomal proteins[46], which are

299   often widely expressed across tissues and have a narrow dynamic range in gene expression[41].

300   Interestingly, previous reports indicate that these promoter classes are differentially responsive

301   to NF-kB[7], suggesting inherently different mechanisms of regulation; an observation we

302   confirmed using our transactivation system (Extended Data Fig.10C)

303

304   We hypothesised that our comparative cofactor screens could demonstrate how different core

305   promoters influence cofactor requirements, while also providing mechanistic insights into

306   cofactor-promoter compatibility[7]. To this end, we created nine independent promoter lines and

307   performed the screens using the same experimental design described earlier (Extended Data

308   Fig.3A and Methods). The results of these screens are summarised with another spoke and

309   wheel chart, which provides a global overview of how core promoters influence the

310   mechanisms of transcription (Fig.5A, Supplementary Table 5).

311

312   Using the same approach as the AD screens, we began by isolating the most heterogeneously

313   enriched cofactors across 9 core promoters and performed Pearson-based hierarchical

314   clustering to unbiasedly identify patterns in the data. Interestingly, this analysis separated the

315   highly responsive TATA/TATA-like promoters from less responsive TCT promoters,

316   suggesting that differential cofactor use underpins differences in compatibility (Fig.5B). In

317   stark contrast with the AD screens (Fig.2C), cofactor clustering across core promoters was

318   discrete (Fig.5B). Promoters containing similar core motifs displayed relatively similar

319   cofactor requirements, with cofactors clustering primarily based on their degree of dependence

320   at TATA/TATA-like or TCT promoters (Fig.5B).

321

322   Amongst the cofactors with differential enrichment between promoter classes were various

323   components of the TFIID complex. As expected, TBP was identified as a major requirement

324   for promoters with a TATA box (Fig.5A). Consistent with recent structural studies, in which

325   TAF11 and TAF13 form a bridge linking TBP to TFIID[47], we identified that these two subunits

326   are also needed together with TBP. Notably, components of this submodule of TFIID were

327   largely not necessary for activation at TCT-containing core promoters (Fig.5A-C). At TCT-

328   containing promoters, we instead identified an increased dependence on TAF1, TAF2, TAF7

329   and TAF8, which are structurally co-located and interact with promoter elements in a manner

330   distinct from TBP (Fig.5A-C)[47–50]. Our data suggests TFIID-mediated assembly of the pre-

331   initiation complex is required across all of these core promoters, however promoters lacking a

332   TATA box are more dependent on the TAF2/7/8 submodule for TFIID assembly.

333

334   We also observed that TATA/TATA-like and TCT promoters were differentially dependent on

335   submodules of Mediator (Fig.3), Integrator (Extended Data Fig.4) and several other cofactors

336   implicated in pause-release and elongation (Fig.1, Fig4A, Fig.5A-C). Interestingly, these

337   cofactors were generally much less enriched at TCT promoters, which are less responsive to

338   NF-κB-AD (Fig.5A-C). The inability for pause-release cofactors to contribute to transcription

339   at TCT promoters suggests that their reduced responsiveness may be due to differences in the

340   rate-limiting step for activation. To test this hypothesis, we performed RNA Pol II ChIP-nexus

341   on our reporter constructs containing the responsive (TATA) and unresponsive (TCT)

342   promoters activated by the NF-κB-AD. TATA promoters displayed clear evidence of RNA Pol

343   II accumulation at the pause site, suggesting that pausing is the rate-limiting step (Fig.6A). In

344   contrast, both unresponsive TCT promoters display a ~10-fold lower in accumulation of RNA

345   Pol II around the TSS and no discernible pausing of RNA Pol II (Fig.6A). This suggests that

346   output from these TCT promoters is constrained by the rate of RNA Pol II initiation rather than

347 pause-release. Based on these findings, we hypothesize that cofactor/promoter incompatibility

348 occurs when TFs recruit cofactors to promoters which do not activate the appropriate rate-

349 limiting step for transcription.

350

351 To provide further support for this hypothesis, we swapped promoter elements from the

352 responsive and unresponsive promoters in an attempt to alter the rate-limiting step and

353 therefore influence cofactor-promoter compatibility (Fig.6B). Replacing the TCT motif with

354 an Initiator motif from TATA or TATA-like promoters did not influence responsiveness,

355 demonstrating that Initiator motifs alone do not have a dominant influence on cofactor-

356 promoter compatibility (Fig.6B). However, adding a TATA or TATA-like element and an

357 Initiator motif into the unresponsive RPL30 promoter, completely restored responsiveness,

358 markedly increasing the dynamic range of gene expression (Fig.6B). Importantly, these

359 changes induced dependency on the pause-release cofactor, CDK8, suggesting that pause-

360 release and elongation became the rate-limiting step (Fig.6B).

361

362 TATA boxes increase the rate of transcriptional initiation by enabling efficient assembly of the

363 pre-initiation complex[51,52] and our screens, and prior work[53,54] suggest that the incompatible

364 TCT promoters are more dependent on a submodule of TFIID that is TBP-independent (Fig.5).

365 This suggests that by adding a TATA box to this unresponsive promoter, we increased the

366 initiation rate, changing the rate-limiting step from initiation to elongation, restoring cofactor-

367 promoter compatibility and increasing the dynamic range of gene expression. Overall, our data

368 supports a model of cofactor-promoter compatibility dictated by core promoter motifs that

369 result in different rate-limiting steps in transcription (Fig.6C).

Discussion

Using a synthetic, reductionistic screening system, we have provided key insights into how TFs and core promoters influence the cofactors required for transcription. Our comparative screens across 9 different ADs suggest that TFs rarely have highly specific, dedicated cofactors with a dominant contribution to transactivation. In contrast, each AD appears to use unique combinations of cofactors. Even when cofactors contribute to transactivation by many TFs (i.e., p300), they can still display variable contributions to transcription. Consequently, there are likely to be a large number of distinct mechanisms by which genes can be activated. Due to our reductionistic design, it is likely that cofactor preference is conferred by AD sequence, however precisely how these intrinsically disordered domains achieve this specificity remains unclear.

Exactly why each AD requires a unique combination of cofactors also remains unclear. One possibility is that this type of combinatorial specificity enables TFs to integrate a large amount of information about cellular state when activating their targets[55–57]. Another possibility is that cofactor specificity enables TFs to regulate distinct steps in transcription. While previous reports have suggested that different TFs can activate different steps[58–60], here we demonstrate that these differences are associated with differences in cofactor use, providing important mechanistic insights into this process. This ability for different TFs to activate different steps in transcription provides the capacity for kinetic synergy[61,62]. Kinetic synergy creates an AND logic, where only in the context of complementary TFs does transcription proceed at maximum efficiency. Therefore, TFs may use different sets of cofactors, to enable more complex logic downstream of DNA binding, or enable different kinetic behaviours, that would not be possible if each TF used the same set of cofactors[5,45,59,63].

We also extended our screens to address how core promoters influence cofactors use. In stark contrast to ADs, the influence of core promoters appears to be discrete. Core promoters with highly divergent sequences, from distinct origins (synthetic, viral or endogenous), but sharing similar promoter motifs, displayed similar cofactor requirements. This suggests that the logic of core promoters is relatively simple, an observation supported by recent deep learning and mutational approaches[64–66]. Despite similarities within promoter classes, distinct classes displayed dramatic differences for cofactors involved in transcriptional initiation and elongation. Importantly, these differences were associated with responsiveness to the NF-κB-AD, leading us to propose a rate-limiting step model of cofactor-promoter compatibility (Fig.

403   5C). TCT promoters, which often regulate genes with a narrow range of gene expression,

404   appear to be constrained by initiation rate, restricting their ability to respond to NF-κB-AD. In

405   contrast, TATA gene promoters, which often regulate dynamically expressed genes, appear to

406   be constrained by pause-release, enabling greater responsiveness and a higher dynamic range

407   in gene expression[67]. This raises the prospect that the pause-release checkpoint evolved to

408   enable a higher dynamic range for genes with promoter elements that efficiently assemble RNA

409   Pol II. Overall, our comparative cofactor screens support a model of transactivation in which

410   promoters establish different rate-limiting steps, that are activated by specific TFs, in order to

411   achieve effective regulation of ubiquitous or inducible gene expression.

412

413   Our comparative screens required a genetic approach and a reductionistic design. Whilst our

414   experiments using rapid protein degradation and transcriptional inhibitors suggest that our

415   major findings result from direct effects, future work assessing other candidates should also

416   employ methods with rapid kinetics. Further work will also be needed to consider how other

417   variables, such as different TF domains, chromatin context, regulation from distal enhancers,

418   and combinations of TFs, interact to regulate endogenous gene expression.

Author Contributions

C.C.B, O.G and M.A.D designed the research and interpreted data. M.A.D supervised the research, with assistance from C.C.B. C.C.B and M.A.D wrote the manuscript with helpful input from all the authors. C.C.B performed the experiments with assistance from J.J.B, L.S, C-S.A and O.G. G. J. F provided critical research support and input. L.T performed the bioinformatic analysis with assistance from A.G and E.Y.N.L and input from C.C.B and M.A.D.

Competing Interests

M.A.D. has been a member of advisory boards for GSK, CTX CRC, Storm Therapeutics, Celgene, and Cambridge Epigenetix and receives research funding from Pfizer. The remaining authors declare no competing interests.

Main figure legends

441 **Figure 1) Comparative CRISPR screens identify the cofactors needed by nine different**

442 **ADs**

443 Spoke and wheel plot demonstrating the enrichment of key cofactors in each of the 9 AD

444 screens. The colour in each wedge reflects the average fold enrichment for each AD. Cofactors

445 are organised into particular complexes based on their known complex associations or

446 molecular functions. Components of this figure were created with biorender.com.

447
448
449 **Figure 2) Transcription factors display a diverse range of activation mechanisms**

450 **(A)** Violin plot of maximum fold enrichment for each cofactor in the library across the 9 AD

451 screens. The number of genes in each category is displayed. **(B)** Violin plot of the coefficient

452 of variation for each gene across the 9 screens. Selected variable genes are listed. Genes that

453 were validated by independent KO experiments are highlighted in bold. **(C)** (Left) Heatmap

454 showing the enrichment of the top 50 most heterogeneously used cofactors across the 9 AD

455 screens. Heatmap is bi-clustered by Pearson correlation distance. Pearson correlation matrices

456 are displayed alongside the heatmap to enable visualisation of which cofactors (right) and ADs

457 (bottom) display correlated patterns of enrichment. (**D**) Heatmap comparing fold enrichment

458 in the screens (left) with fold change in GFP upon knockout of various candidate,

459 heterogeneous cofactors (right). Fold reduction in GFP calculated by dividing the average

460 fluorescence signal (M.F.I) in perturbed cells by the M.F.I in cells containing the safe guide

461 control at D5 after sgRNA infection. Fold reduction in GFP calculated based on at least two

462 sgRNAs per gene.

463

464 **Figure 3) A direct co-dependent relationship between the tail 2 and kinase modules of the**

465 **Mediator complex**

466 **(A)** Spoke and wheel plot of the fold enrichment of Mediator complex subunits across the nine

467 ADs. **(B)** Waterfall plots showing the change in RNA pol II levels at Mediator dependent genes

468 after individual subunit KO. Mediator dependent genes defined as genes with at least 30%

469 reduction in RNA pol II signal upon MED14 KO. Each sample is compared to a safe guide

470 control to calculate a change in RNA pol II ChIP-seq signal. The genes are ordered based on

471 the degree of reduction in each KO sample enabling direct comparison of whether the same

472 genes are affected by removal of different subunits. Spearman rank order correlation is

473 displayed on each waterfall plot. The colour of each sample reflects which submodule it

474   belongs to. Orange = middle/core, blue = tail 2 and purple = kinase module. **(C)** Correlation
475   matrix of the spearman rank order correlation coefficient upon KO of various components of
476   the Mediator complex. **(D)** Western blot for HA-dTAG-tagged MED12, MED14 and MED25
477   after 4hrs of dTAGV-[1] or CDK8i treatment. Alpha-Tubulin displayed as loading control.
478   Representative blot of two biological replicates. Blot performed on matched samples from
479   ChIP-seq and SLAM-seq experiments. **(E)** Scatter plots demonstrating the correlation between
480   the change in RNA pol II ChIP-seq signal upon MED12, MED14, MED25 degradation (4hrs
481   dTAGV-[1]) and CDK8i treatment (4hrs). Fold change calculated by comparing each sample to
482   matched DMSO treated control. Red line and r value reflect Pearson correlation.

483

484   **Figure 4) Transcription factors use different cofactors to regulate different steps in**
485   **transcription**
486   **(A)** (Left) RNA pol II ChIP-nexus coverage across the reporter construct in each of the GAL4-
487   AD cell lines. Blue signal reflects reads from the sense strand, red signal reflects reads from
488   the anti-sense strand and shaded blue signal reflects cumulative signal between both strands.
489   Elongation index is the inverse of the pausing index i.e., total gene body signal divided by the
490   total promoter signal. (Right) Heatmap displaying fold enrichment of various key regulators of
491   pause-release and transcriptional elongation in each AD screen. **(B)** RNA pol II ChIP-nexus
492   coverage across the reporter construct in the GAL4-VP64 and GAL4-MYB cell lines treated
493   for 1hr with DMSO or CDK8i. Blue and red lines reflect cumulative ChIP-nexus signal in
494   DMSO and CDK8i treatment respectively. Quantification of the change in GFP upon CDK8i,
495   CDK8 KO and CCNC KO in these two cell lines is also shown. For GFP quantification, n = 3
496   biological replicates, error bars = S.E.M.

497

498   **Figure 5) Comparative screens shows discrete cofactor preferences dictated by core**
499   **promoter elements**
500   **(A)** Spoke and wheel plot demonstrating the enrichment of key cofactors in each of the 9 core
501   promoter screens. The colour in each wedge reflects the average fold enrichment for each AD.
502   Cofactors are organised into particular complexes based on their known complex associations
503   or molecular functions. For ease of comparison, the same cofactors are displayed as Fig. 1. **(B)**
504   (Left) Heatmap showing the enrichment of the top 50 most heterogeneously used cofactors
505   across the 9 core promoter screens. Heatmap is bi-clustered by Pearson correlation distance.
506   Pearson correlation matrices are displayed alongside the heatmap to enable visualisation of
507   which cofactors (right) and core promoters (bottom) display correlated patterns of enrichment.

508   Display is consistent with Fig. 2C to enable direct comparison of the clustering across different

509   ADs and core promoters. **(C)** Validation of screen data through quantification change in GFP

510   signal upon individual cofactor KO. Relative GFP signal is calculated by dividing the average

511   fluorescence signal (M.F.I) cells containing the safe guide control by the M.F.I in perturbed

512   cells at D5 after sgRNA infection. n = 2 sgRNAs per gene, error bars = S.E.M. Components of

513   this figure were created with biorender.com.

514

515   **Figure 6) Cofactor-promoter compatibility is influenced by the rate limiting step in**

516   **transcription**

517   **(A)**   RNA pol II ChIP-nexus coverage across the reporter construct in responsive and

518   unresponsive promoter lines activated by GAL4-NF-κB. Blue signal reflects reads from the

519   sense strand, red signal reflects reads from the anti-sense strand and shaded blue signal reflects

520   cumulative signal between both strands. The top number adjacent to the graph reflects the

521   cumulative normalised read counts from both strands. The bottom two numbers reflect the

522   normalised read counts from each strand respectively. **(B)** Luciferase assays performed with

523   different promoter constructs with or without GAL4-NF-κB demonstrating that adding of a

524   TATA box restores responsiveness and cofactor-promoter compatibility. CDK8i was dosed for

525   12hrs prior to luciferase assay. n=3 technical replicates. Error bars = S.E.M. Data is

526   representative of two independent biological replicates. **(C)** Rate limiting step model of

527   cofactor-promoter compatibility. Pause-release cofactors have limited ability to activate

528   promoters where initiation is the rate-limiting step. + symbols do not directly correspond to

529   any quantitively information, instead reflecting a conceptual model.

530

531   **Figure – Representative FACS gating**

532   Representative gating strategy used to identify GFP positive cells throughout the manuscript.

533   Live cells are identified by FSC-A and SSC-A, single cells identified by FSC-A and FSC-H

534   and GFP positive cells gated relative to a negative control population.

535

A

		Elongation index
Compatible	CMV	4.2
	SCP	3.8
Incompatible	MYB	8.7
	RPL30	8.1

B

SCP — TATA / Inr
- No TAD
- NfKB
- NfKB + CDK8i

HBB — weakTATA / Inr
- No TAD
- NfKB
- NfKB + CDK8i

RPL30 — TCT
- No TAD
- NfKB
- NfKB + CDK8i

RPL30 with SCP initiator — TCT / Inr
- No TAD
- NfKB
- NfKB + CDK8i

RPL30 with HBB initiator — TCT / Inr
- No TAD
- NfKB
- NfKB + CDK8i

RPL30 with SCP initiator and TATA — TATA / Inr
- No TAD
- NfKB
- NfKB + CDK8i

SCP with RPL30 TCT — TATA / TCT
- No TAD
- NfKB
- NfKB + CDK8i

HBB with RPL30 TCT — weakTATA / TCT
- No TAD
- NfKB
- NfKB + CDK8i

Luciferase relative to renilla

C

	Initiation rate	Elongation rate	Pausing index	Transcriptional output	Rate-limiting step
SCP	+	+	−	+	Balanced
SCP+NfKB	++++	+++	+	+++	Elongation
RPL30	+	+	−	+	Balanced
RPL30+NfKB	++	+++	−	++	Initiation

Representative FACS gating

References

1. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
2. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
3. Roeder, R. G. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* **579**, 909–915 (2005).
4. Ferrie, J. J., Karr, J. P., Tjian, R. & Darzacq, X. "Structure"-function relationships in eukaryotic transcription factors: The role of intrinsically disordered regions in gene regulation. *Mol Cell* 1–15 (2022) doi:10.1016/j.molcel.2022.09.021.
5. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev* **43**, 73–81 (2017).
6. Levine, M., Cattoglio, C. & Tjian, R. Looping Back to Leap Forward: Transcription Enters a New Era. *Cell* **157**, 13–25 (2014).
7. Haberle, V. *et al.* Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).
8. Neumayr, C. *et al.* Differential cofactor dependencies define distinct types of human enhancers. *Nature* **606**, 406–413 (2022).
9. Stampfel, G. *et al.* Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**, 147–151 (2015).
10. Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
11. Alerasool, N., Leng, H., Lin, Z.-Y., Gingras, A.-C. & Taipale, M. Identification and functional characterization of transcriptional activators in human cells. *Mol Cell* **82**, 677-695.e7 (2022).
12. Nemčko, F. & Stark, A. Proteome-scale identification of transcriptional activators in human cells. *Mol Cell* **82**, 497–499 (2022).
13. Donczew, R., Warfield, L., Pacheco, D., Erijman, A. & Hahn, S. Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. *Elife* **9**, 1–27 (2020).
14. Bergman, D. T. *et al.* Compatibility rules of human enhancer and promoter sequences. *Nature* **106**, 1–22 (2022).
15. Martinez-Ara, M., Comoglio, F., van Arensbergen, J. & van Steensel, B. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol Cell* (2022) doi:10.1016/j.molcel.2022.04.009.
16. Arensbergen, J. Van, Steensel, B. Van & Bussemaker, H. J. In search of the determinants of enhancer – promoter interaction specificity. *Trends Cell Biol* **24**, 695–702 (2014).
17. Galouzis, C. C. & Furlong, E. E. M. Regulating specificity in enhancer-promoter communication. *Curr Opin Cell Biol* **75**, 102065 (2022).
18. Kim, S. & Wysocka, J. Deciphering the multi-scale, quantitative cis-regulatory code. *Molecular Cell* vol. 83 373–392 Preprint at https://doi.org/10.1016/j.molcel.2022.12.032 (2023).
19. Schmid-Burgk, J. L., Höning, K., Ebert, T. S. & Hornung, V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat Commun* **7**, 12338 (2016).
20. Pattabiraman, D. R. *et al.* Interaction of c-Myb with p300 is required for the induction of acute myeloid leukemia (AML) by human AML oncogenes. *Blood* **123**, 2682–2690 (2014).
21. Vojnic, E. *et al.* Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* **18**, 404–409 (2011).

601   22.   Mittler, G. *et al.* A novel docking site on Mediator is critical for activation by VP16 in
602   mammalian cells. *EMBO Journal* **22**, 6494–6504 (2003).

603   23.   Yang, F., DeBeaumont, R., Zhou, S. & Näär, A. M. The activator-recruited
604   cofactor/Mediator coactivation subunit ARC92 is a functionally important target of the
605   VP16 transcriptional activator. *Proc Natl Acad Sci U S A* **101**, 2339–2344 (2004).

606   24.   Sandberg, M. L. *et al.* c-Myb and p300 regulate hematopoietic stem cell proliferation
607   and differentiation. *Dev Cell* **8**, 153–166 (2005).

608   25.   Xu, Y. *et al.* A TFIID-SAGA Perturbation that Targets MYB and Suppresses Acute
609   Myeloid Leukemia. *Cancer Cell* **33**, 13-28.e8 (2018).

610   26.   Donner, A. J., Szostek, S., Hoover, J. M. & Espinosa, J. M. CDK8 Is a Stimulus-
611   Specific Positive Coregulator of p53 Target Genes. *Mol Cell* **27**, 121–133 (2007).

612   27.   Chung, C. Y. *et al.* Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary
613   Tumorigenesis. *Cell Rep* **16**, 472–486 (2016).

614   28.   Zheng, H. *et al.* Identification of Integrator-PP2A complex (INTAC), an RNA
615   polymerase II phosphatase. *Science (1979)* **370**, (2020).

616   29.   El Khattabi, L. *et al.* A Pliable Mediator Acts as a Functional Rather Than an
617   Architectural Bridge between Promoters and Enhancers. *Cell* **178**, 1145–1158 (2019).

618   30.   Abdella, R. *et al.* Structure of the human Mediator-bound transcription preinitiation
619   complex. *Science (1979)* **372**, 52–56 (2021).

620   31.   Jaeger, M. G. *et al.* Selective Mediator dependence of cell-type-specifying
621   transcription. *Nat Genet* 825299 (2020) doi:10.1038/s41588-020-0635-0.

622   32.   Warfield, L., Donczew, R., Mahendrawada, L. & Hahn, S. Yeast Mediator facilitates
623   transcription initiation at most promoters via a Tail-independent mechanism. *Mol Cell*
624   **82**, 4033-4048.e7 (2022).

625   33.   Nabet, B. *et al.* The dTAG system for immediate and target-specific protein
626   degradation. *Nat Chem Biol* **14**, 431–441 (2018).

627   34.   Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional
628   initiation. *Nat Genet* **49**, 1045–1051 (2017).

629   35.   Aoi, Y. *et al.* NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II
630   Pause-Release. *Mol Cell* **78**, 261-274.e5 (2020).

631   36.   Steinparzer, I. *et al.* Transcriptional Responses to IFN-γ Require Mediator Kinase-
632   Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions.
633   *Mol Cell* **76**, 485-499.e8 (2019).

634   37.   Leach, K. M. *et al.* Characterization of the human β-globin downstream promoter
635   region. *Nucleic Acids Res* **31**, 1292–1301 (2003).

636   38.   Stewart, J. J., Fischbeck, J. A., Chen, X. & Stargell, L. A. Non-optimal TATA
637   elements exhibit diverse mechanistic consequences. *Journal of Biological Chemistry*
638   **281**, 22665–22673 (2006).

639   39.   Stewart, J. J. & Stargell, L. A. The Stability of the TFIIA-TBP-DNA Complex Is
640   Dependent on the Sequence of the TATAAA Element. *Journal of Biological*
641   *Chemistry* **276**, 30078–30084 (2001).

642   40.   Smale, S. T. & Kadonaga, J. T. The RNA Polymerase II Core Promoter. *Annu Rev*
643   *Biochem* **72**, 449–479 (2003).

644   41.   Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley*
645   *Interdiscip Rev Dev Biol* **1**, 40–51 (2012).

646   42.   Morachis, J. M., Murawsky, C. M. & Emerson, B. M. Regulation of the p53
647   transcriptional response by structurally diverse core promoters. *Genes Dev* **24**, 135–
648   147 (2010).

649 43. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise Maps of RNA Polymerase
650   Reveal How Promoters Direct Initiation and Pausing. *Science (1979)* **339**, 950–953
651   (2013).
652 44. Gilchrist, D. A. *et al.* Pausing of RNA polymerase II disrupts DNA-specified
653   nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–551 (2010).
654 45. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: A nexus
655   of gene regulation. *Genes Dev* **33**, 960–982 (2019).
656 46. Parry, T. J. *et al.* The TCT motif, a key component of an RNA polymerase II
657   transcription system for the translational machinery. *Genes Dev* **24**, 2013–2018 (2010).
658 47. Patel, A. B. *et al.* Structure of human TFIID and mechanism of TBP loading onto
659   promoter DNA. *Science (1979)* **362**, eaau8872 (2018).
660 48. Petrenko, N., Jin, Y., Dong, L., Wong, K. H. & Struhl, K. Requirements for RNA
661   polymerase II preinitiation complex formation in vivo. *Elife* **8**, 1–19 (2019).
662 49. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-
663   initiation complex assembly. *Nature* (2016) doi:10.1038/nature17394.
664 50. Petrenko, N. & Struhl, K. Comparison of transcriptional initiation by rna polymerase ii
665   across eukaryotic species. *Elife* **10**, 1–23 (2021).
666 51. Hoopes, B. C., LeBlanc, J. F. & Hawley, D. K. Contributions of the TATA box
667   sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J Mol
668   Biol* **277**, 1015–1031 (1998).
669 52. Yean, D. & Gralla, J. Transcription reinitiation rate: a special role for the TATA box.
670   *Mol Cell Biol* **17**, 3809–3816 (1997).
671 53. Serebreni, L. *et al.* Functionally distinct promoter classes initiate transcription via
672   different mechanisms reflected in focused versus dispersed initiation patterns. *EMBO J*
673   **42**, (2023).
674 54. Wang, Y. L. *et al.* TRF2, but not TBP, mediates the transcription of ribosomal protein
675   genes. *Genes Dev* **28**, 1550–1555 (2014).
676 55. Klumpe, H. E. *et al.* The context-dependent, combinatorial logic of BMP signaling.
677   *Cell Syst* **13**, 388-407.e10 (2022).
678 56. Su, C. J. *et al.* Ligand-receptor promiscuity enables cellular addressing. *Cell Syst* **13**,
679   408-425.e12 (2022).
680 57. Klumpe, H. E., Garcia-Ojalvo, J., Elowitz, M. B. & Antebi, Y. E. The computational
681   capabilities of many-to-many protein interaction networks. *Cell Systems* vol. 14 430–
682   446 Preprint at https://doi.org/10.1016/j.cels.2023.05.001 (2023).
683 58. Danko, C. G. *et al.* Signaling Pathways Differentially Affect RNA Polymerase II
684   Initiation, Pausing, and Elongation Rate in Cells. *Mol Cell* **50**, 212–222 (2013).
685 59. Blau, J. *et al.* Three functional classes of transcriptional activation domain. *Mol Cell
686   Biol* **16**, 2044–2055 (1996).
687 60. Harden, T. T., Vincent, B. J. & DePace, A. H. Transcriptional activators in the early
688   Drosophila embryo perform different kinetic roles. *Cell Syst* **14**, 258–272 (2023).
689 61. Herschlag, D. & Johnson, F. B. Synergism in transcriptional activation: a kinetic view.
690   *Genes Dev* **7**, 173–179 (1993).
691 62. Martinez-Corral, R. *et al.* Transcriptional kinetic synergy: A complex landscape
692   revealed by integrating modeling and synthetic biology. *Cell Syst* **14**, 324–339 (2023).
693 63. Scholes, C., DePace, A. H. & Sánchez, Á. Combinatorial Gene Regulation through
694   Kinetic Control of the Transcription Cycle. *Cell Syst* **4**, 97-108.e9 (2017).
695 64. Dudnyk, K., Shi, C. & Zhou, J. Sequence basis of transcription initiation in human
696   genome. *bioRxiv* (2023) doi:/10.1101/2023.06.27.546584.

697 65. Li, X. C., Fuqua, T., van Breugel, M. E. & Crocker, J. Mutational scans reveal
698   differential evolvability of Drosophila promoters and enhancers. *Philosophical*
699   *Transactions of the Royal Society B: Biological Sciences* **378**, (2023).
700 66. Sahu, B. *et al.* Sequence determinants of human gene regulatory elements. *Nat Genet*
701   **54**, 283–294 (2022).
702 67. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. Prevalence of the
703   initiator over the TATA box in human and yeast genes and identification of DNA
704   motifs enriched in human TATA-less core promoters. *Gene* **389**, 52–65 (2007).
705
706

Methods

Cell culture

A clonal K562 Cas9 cell line was generated previously to ensure high efficiency CRISPR editing[68]. K562 cells were cultured in RPMI-1640 supplemented with 20% FCS, streptomycin (100ug/ml), penicillin (100 units/ml) and Glutamax, under standard culture conditions (5% CO_2, 37°C). HEK293ET cells were grown in DMEM supplemented with 10% FCS, streptomycin (100ug/ml), penicillin (100 units/ml) and Glutamax, under standard culture conditions (5% CO_2, 37°C). All cell lines were subjected to regular mycoplasma testing and underwent short tandem repeat (STR) profiling.

Lentivirus production and transduction

Lentivirus was prepared by transfecting HEK293ET cells with plasmid:pVSV-G:psPAX2 plasmids in a 3:2:1 ratio using PEI reagent. The viral supernatant was collected 48-72hrs following transfection, filtered through a 0.45 μm filter and added to cells.

Drug treatment

Senexin A (CDK8i) (Selleckchem) was dosed at 10uM. GAL4-GR cells were dosed with 1uM Dexamethasone (Sigma) to induce GR activity. Recombinant human TNF-α (Peprotech) treatment was performed for 6hrs at 25ng/ml. For acute dTAG degradation experiments, cells were dosed with dTAG-V1 (Tocris Biosciences) at 500nM for 4hrs. For GFP half-life analysis of the reporter system, cells were treated with triptolide (10uM)

Flow cytometry analyses

Flow cytometry analyses were performed on the LSRFortessa X-20 flow cytometer (BD Biosciences). Data were analysed with FlowJo v10 software (Tree Star). Cell sorting was performed on the FACSAria or Fusion 5 flow sorter (BD Biosciences).

Cloning of screening system

pKLV-U6gRNA(BbsI)-Puro2ABFP vector was used as the base vector for cloning the lentiviral GAL4-AD vector. The entire gRNA, Puromycin and BFP regions of the plasmid were removed and replaced by the GAL4-DBD together with an IgA linker through standard cut and paste cloning. The EF1a, IRES and mCherry were then introduced sequentially. The vector was designed for simple cut and paste replacement of the AD region downstream of the GAL4-

740 DBD with alternative ADs. ADs of interested were amplified by PCR from cDNA expression
741 vectors obtained from Addgene. Primers used for amplification and the final AD sequences are
742 listed in Supplementary Table 2 and 6.

743

744 The base for the reporter construct was obtained from Addgene (#79199). The original vector
745 is a lentiviral vector that includes the 5xUAS upstream of a minimal CMV promoter.
746 Downstream of this promoter, Turbo-GFP-PEST was subcloned from another Addgene vector
747 (#67180). To produce different promoter reporters, promoter regions were obtained from
748 previous publications or from the eukaryotic promoter database (Supplementary Table 2). For
749 endogenous promoters, a region was selected to capture a 100bp window around the centre of
750 the CAGE signal reported on the Eukaryotic Promoter Database. Oligos corresponding to these
751 regions were synthesized by IDT, annealed and cloned into the 5xUAS-mCMV-Turbo-GFP-
752 PEST construct, replacing the mCMV promoter (Supplementary Table 2).

753

754 Generation of GAL4-AD and promoter reporter cell lines
755 For AD screens, the lentiviral 5xUAS mCMV Turbo-GFP-PEST reporter was introduced into
756 clonal K562-Cas9-Blasticidine cell line using high titre virus to achieve a high MOI. High MOI
757 infection is necessary to minimise locus specific effects and to ensure robust detection of GFP
758 signal upon activation by lower potency ADs. Into this reporter line, each of the GAL4-AD
759 constructs were introduced at a high MOI (~90-100% of cells infected) by lentiviral integration.
760 For promoter screens, the lentiviral 5xUAS Turbo-GFP-PEST reporter with variable promoters
761 were integrated into a clonal K562-Cas9-Blasticidine cell line using high titre virus to achieve
762 a high MOI. For all cell lines, GFP positive cells were sorted until a pure and stable GFP
763 population was obtained.

764

765 sgRNA design and cloning
766 sgRNAs were designed using the IDT CRISPR design tool or were obtained from the
767 sequences of guides in the pooled guide library. sgRNAs were cloned into the pKLV-
768 U6gRNA(BbsI)-Puro$_{2A}$BFP vector using standard golden gate cloning.

769

770 sgRNA and primer sequences
771 The sequences for sgRNA sequences and relevant primer sequences are included in
772 Supplementary Table 6.

773

Validation and quantification of screen hits

All experiments validating and quantifying the effect of cofactor knockout were performed by quantifying the mean fluorescence intensity (M.F.I. or average fluorescence signal) of the GFP reporter at day 5 after infection with the relevant sgRNA. The relative GFP signal was calculated by dividing average GFP signal in the AD line infected with an sgRNA of interest by the same AD line infected with a control sgRNA targeting a safe genomic locus (safe guide). A number of examples of FACS plots used to calculate the change in GFP signal are provided in Extended Data Fig. 5. Each AD line was infected with the same batch of virus in parallel to minimise technical variation between the quantification. Cofactors of interest were validated with at least 2 sgRNAs and at least 2 biological replicates.

HDR mediated AAVS knock-in

The 5xUAS, mCMV promoter and Turbo-GFP-PEST were cloned by Gibson assembly into the pMK232 (CMV-OsTIR1-Puro), which contained homology arms for the AAVS locus. The sgRNA targeting the AAVS was introduced into the pX330-mCherry vector, which contains both Cas9 and the gRNA. The AAVS-reporter repair template and the pX330-Cas9-AAVS gRNA vector was electroporated into K562 cells using the Neon Transfection system (Thermo Fischer) with settings optimised for K562 cells. Single cell clones were sorted 5 days after transfection and grown out for 2 weeks to obtain sufficient cells for analysis. AAVS knock-in clones were identified by In-Out PCR and Sanger sequencing.

Guide library design, generation and cloning

To assess the requirement of transcriptional regulators, a bespoke library of gRNAs that targets over 1137 known chromatin and transcriptional regulators was designed (Supplementary Table 3). The library was designed through a combination of searches for genes containing domains enriched in transcriptional regulators and manual curation. Each gene was targeted with 6 independent gRNAs. As controls, the library also contains a large number of guides targeting safe regions and guides that do not target any genomic locus. The total library contains 7239 gRNAs. The oligo pool was synthesized by CustomArray (Genescript). The sgRNA pool was PCR amplified and pot cloned into the pKLV-U6gRNA(BbsI)-Puro$_{2A}$BFP vector using standard cut and paste cloning. The ligated product was electroporated into Electrocompetent cells (Lucigen) and grown in liquid culture overnight at 37 degrees before being extracted by Maxiprep. Low skewing of the plasmid was confirmed by sequencing of the cloned plasmid

807   pool. The library had a skew ratio of 4.35 (counts for top 10% of guides divided by counts for
808   bottom 10% of guides). A skew ratio below 10 is considered acceptable [69].

809

810   **Comparative CRISPR screens**

811   Prior to beginning the screens, reporter cell lines were sorted to obtain a pure GFP positive
812   population. Sufficient cells were used to maintain 1000-fold representation at all stages of the
813   screening process. The cells were transduced with an appropriate volume of viral supernatant
814   to ensure only a single guide was present in most cells (MOI < 0.3, mean of 0.21,
815   Supplementary Table 7). At day 5, 6 and 7 after guide library infection, at least 1 million guide
816   positive (BFP positive), GFP negative cells (< 25% of the mean fluorescence intensity of the
817   entire population) were sorted (refer to Supplementary Note 1 for FACS plots and further
818   details). This translates to approximately 1000-fold representation of the library. Guide positive
819   cells (at least 10 million cells) were also sorted as a library control at each time point to provide
820   a library control reference to calculate enrichment (>1000-fold representation). Cells were
821   pelleted after sorting and stored at -80°C until genomic DNA extraction was performed. Four
822   of the GAL4-AD cell lines from the comparative CRISPR screens were also maintained until
823   day 14 after guide infection to test which genes are required for cell growth. These four AD
824   lines were used as independent replicates for the dropout analysis.

825

826   Genomic DNA was extracted using Monarch® Genomic DNA purification kit (New England
827   Biolabs), according to the manufacturer's instructions. PCR was conducted to maintain guide
828   representation, using Q5® High Fidelity DNA Polymerase (New England Biolabs). PCR was
829   performed through a one-step PCR with 28 cycles. 500ng of template was added to each PCR.
830   One step PCR helps to avoid excessive amplification, by minimising sample processing. PCR
831   was optimised to ensure it stays within the exponential phase. Depending on the gDNA
832   concentration from the extraction, approximately 10-20 PCR reactions were performed per
833   screen sample and between 30-50 PCRs per library control. PCR products were pooled and
834   sequenced on the NextSeq500 using 75bp single-end chemistry. The samples were sequenced
835   with the following summary statistics: Min = 3.86×10^6 reads, Q1 = 6.24×10^6 reads, Q2=
836   7.37×10^6 reads, Q3 = 8.51×10^6 reads, Max = 12×10^6 reads.

837

838   **Analysis of comparative CRISPR screens**

839   The sequence reads were trimmed to remove the constant portion of the sgRNA sequences with

840   cutadapt v4.3[70], then mapped to the reference sgRNA library with Bowtie2[71]. After filtering to

841   remove multi-mapping reads, the read counts were computed for each sgRNA. After obtaining

842   guide counts for all samples, a series of processing steps were performed to calculate mean

843   fold enrichment values for each gene in each screen (see Supplementary Note 1 for further

844   details). Firstly, guides that were very lowly represented (below 2.5th percentile) were filtered

845   from the analysis, since their low representation caused extreme fold change values[72,73] (see

846   Supplementary Note 1 for further details). The counts were then normalised to sequencing

847   depth before calculating the fold enrichment for each guide by dividing the counts for each

848   gene in the screen samples by the counts in the library control. This resulted in a total of 18

849   individual fold enrichments for each gene (6 guides, 3 timepoints). To remove outliers, we

850   filtered any guides that had a fold enrichment below 0.1 or greater than 10 (see Supplementary

851   Note 1 for further details). To improve quantification, further outliers were removed by filtering

852   guides that were more than 4-fold away from the mean fold change value. The filtering steps

853   applied to the data are benchmarked using two highly correlated replicates of the GAL4-NF-

854   κB screen (r=0.92, Pearson) (Supplementary Note 1).

855

856   Using this filtered guide list, we calculated a final fold enrichment score for each gene in each

857   screen. In order to calculate what fold enrichment score should be considered statistically

858   significant, a permutation test was performed for each screen (Supplementary Note 1). The

859   permutation test shows what fold change distribution would be expected if you randomly

860   sampled fold enrichment scores from guides in the data. Specifically, 6 guides were randomly

861   sampled from each timepoint providing a vector of 18 values. The mean fold enrichment was

862   calculated across these 18 values. Random sampling was performed 10000 times to produce a

863   random sampling distribution. Genes that had an average fold enrichment above the 95th

864   percentile of this random distribution, as well as at least 1/3rd of the guides (6/18) above the

865   95th percentile of this distribution, were considered significant.

866

867   To identify heterogeneous regulators, the coefficient of variation was calculated for each gene.

868   To produce the heatmaps displayed in Fig. 2 and Fig. 5, the top 50 most heterogeneous

869   cofactors were then bi-clustered based on their Pearson correlation distance. SPC24, SPC25

870   and NUF2 were filtered from the variability analysis as the function in regulating genome

871   ploidy has been previously demonstrated to result in spurious variability across screens[74].

872   Two relevant regulators of transcription, *CCNC* and *SETD1A* were not represented in the

873   library. To calculate fold enrichment scores for these genes, we performed independent KO

874   experiments using at least 3 sgRNAs and calculated the reduction in GFP signal as described

875   above. The fold reduction in GFP signal for these genes is displayed as the fold enrichment

876   score in Figure 1, 3 and 5.

877

878   To define which genes affected cell growth, MAGeCK v0.5.6 analysis was performed using

879   the D14 timepoint from 4 of the AD screens, comparing them to the plasmid as the D0

880   reference[75]. Any genes with an adjusted p-value below 0.05 were considered significant.

881

882   **ZFP-VP64 piggyBac reporter screen**

883   A replicate screen was performed using an alternative integration method and alternative DNA

884   binding domain to demonstrate the broad relevance and reproducibility of our findings. In order

885   to perform this screen, a construct in which VP64-AD is recruited through an artificial ZFP

886   protein[76] was developed (37ZFP, Addgene #176627). To generate the DNA binding-AD

887   construct, the GAL4-VP64 vector was cut and the GAL4-DBD was replaced by the ZFP DNA

888   binding domain. A fully insulated PiggyBac reporter construct was also developed. This

889   reporter construct contains the A1 insulator[77] upstream and downstream of a GFP-PEST

890   reporter. GFP-PEST is activated by a minimal CMV promoter with 6 upstream binding sites

891   for 37ZFP-VP64. The reporter construct was cloned by a combination of Gibson assembly and

892   standard cut and paste cloning. The binding sites for 37ZFP were obtained by PCR from

893   Addgene vector #176627.

894

895   Upon successful cloning of both the reporter and ZFP-AD constructs, a reporter line was

896   generated by piggyBac-mediated integration of the insulated ZFP-reporter construct. 1 million

897   Cas9-K562 cells were transfected with 1ug of the reporter and 250ng of HyBase transposase[78]

898   using the Lonza 4D nucleofection system. Transfected cells were then infected with the ZFP-

899   VP64 construct by lentiviral integration at high MOI (~90-100% of cells infected). GFP

900   positive cells were then isolated by FACS sorting. The screen was performed and analysed as

901   described above, in order to enable direct comparison of the results. Guides targeting *FKBP1A*

902   and *FKBP1C* were excluded from the analysis as they directly target the ZFP sequence.

903

904   **Cofactor KO ChIP-seq and RNA-seq experimental design**

905   sgRNAs targeting *MED12*, *MED14*, *MED16*, *MED24*, *MED25*, *CDK8* and *CCNC* were
906   lentivirally introduced into a K562-Cas9 clone. Cells were grown for 4 days after infection at
907   which timepoint they were harvested for ChIP-seq. For INTS5 RNA-seq and ChIP-seq,
908   sgRNAs targeting INTS5 were introduced into a K562-Cas9 clone. Cells were grown for 4
909   days post guide infection and treated with PBS or TNF-α for 6hrs before being harvested for
910   ChIP-seq or RNA-seq.

911

912   **Cloning & Generation of dTAG knock-in cell lines**
913   The dTAG degron (FKBP12^{F36V}) was selectively knocked-in into the N-termini of *MED12*,
914   *MED14* and *MED25*. Cloning of knock-in plasmids and the subsequent generation of knock-
915   in cells lines were conducted as previously described[33,79]. Briefly, the hU6-PITCh-gRNA
916   cassette from pX330S-2-PITCh (Addgene, #63670) was subcloned into pX330A-1x2
917   (Addgene, # 58766) via BsaI digestion to create an all-in-one CRISPR-Cas9 vector labelled
918   pX330-A+S. sgRNAs against target loci were ligated into pX330-A+S via BpiI digestion and
919   golden-gate assembly. Donor vectors were constructed by using PCR to add 20bp
920   microhomology sequences against target loci to pCRIS-PITChv2-Puro-dTAG (Addgene,
921   #91793). 750ng each of the paired gRNA (px330-A+S) and donor vectors (pCRIS-PITChv2)
922   were electroporated into $3x10^5$ K562 cells in Buffer R using the Neon™ Transfection System
923   10uL kit (Thermo Fisher). Cells were allowed to recover for 48 hours, followed by 5-7 days of
924   puromycin selection (2ug/mL). Single cell clones were isolated via FACS sorting into 96-wells
925   and allowed to expand. Genomic DNA was isolated using DirectPCR Lysis Reagent® (Viagen
926   Biotech) according to manufacturer's instructions and directly used as input for genotyping
927   PCRs. Clones demonstrating homozygous knock-in were further validated by sanger
928   sequencing of the homozygous knock-in gel extracted product, as well as further validated by
929   immunoblot analysis. Supplementary Table 6 lists dTAG cloning and PCR oligo sequences.

930

931   **Chromatin immunoprecipitation (ChIP)**
932   For each ChIP, at least 20 million cells were crosslinked for 15 mins with 1% formaldehyde.
933   Crosslinked material was sonicated to approximately 200-1000bp using the Covaris
934   Ultrasonicator S2. Sonicated material was incubated overnight with each antibody in IP buffer
935   (10mM Tris-HCl pH8, 1% Triton X-100, 0.1% sodium deoxycholate, 90mM NaCl), then
936   incubated for 3hrs with 50ul of either Protein A or Protein G Dynabeads (Thermo Fisher).
937   Antibody bound beads were washed twice with low salt wash buffer (20mM Tris-HCl pH8,

938 2mM EDTA, 1% Triton X-100, 0.1% SDS, 150mM NaCl) and once with high salt wash buffer

939 (20mM Tris-HCl pH8, 2mM EDTA, 1% Triton X-100, 0.1% SDS, 500mM NaCl) and once

940 with TE, before the ChIP material was eluted and de-crosslinked overnight at 65 degrees in

941 elution buffer (1% SDS, 100mM NaHCO3). DNA was purified using Qiagen Minelute

942 columns. All ChIP antibodies were used at ~10ug per IP. Sequencing libraries were prepared

943 using the ThruPLEX® DNA-seq kit (Takara Bio). Libraries were size selected between 200-

944 500bps and sequenced on the NextSeq500 using the 75bp single-end chemistry. The following

945 antibodies were used for ChIP: Mouse anti-RNA polymerase II antibody clone CTD4H8 (6ul)

946 (Merck Millipore, 05-623), Rabbit anti-NF-kB p65 antibody clone D14E12 (10ul) (Cell

947 Signalling, 8242), mouse anti-GAL4-DBD antibody clone RK5C1 (10uL) (sc-510). For

948 quantification of occupancy on the GAL4-reporter construct, ChIP-qPCR was performed using

949 primers specific to the promoter of the GAL4-reporter construct and compared to a gene desert

950 negative control region.

951

952 **ChIP-seq analysis**

953 Reads were aligned to the human genome (GRCh38) with Bowtie2[71]. Duplicate reads and reads

954 mapping to blacklist regions or mitochondria were removed. ChIP–seq coverage across

955 selected genomic regions was calculated with BEDtools v2.31.0[80]. To define which genes are

956 Mediator dependent, the 10000 genes with the most RNA pol II signal across the gene were

957 isolated. From this list, genes with at least 30% reduction in total RNA pol II signal in the

958 MED14 KO were defined as Mediator dependent (1020 genes). For each KO sample, we then

959 calculated the change in RNA Pol II signal by comparison to the SAFE guide control. For the

960 mediator dTAG degron ChIP-seq analysis, correlation plots were performed on the top 10000

961 genes by RNA Pol II signal.

962

963 **ChIP-nexus**

964 ChIP-nexus was performed as described previously[34,81,82]. Briefly, the immunoprecipitation

965 and washes were performed using the same conditions as the ChIP protocol. Upon completion

966 of these steps, the DNA was end-repaired, A-tailed, adaptors ligated, exonuclease treated,

967 circularized on Dynabeads as described previously[34,81,82]. DNA was then eluted from the beads,

968 and PCR was performed to produce sequencing ready libraries. For ChIP-nexus performed on

969 the different GAL4-AD lines, DNA was sequenced on the NextSeq500 using the 75bp single-

970 end chemistry. For ChIP-nexus on the different promoter lines, DNA was sequenced on the

971   NextSeq500 using 75bp single-end chemistry, which was configured to produce paired-end

972   37bp reads.

973

974   **ChIP-nexus data analysis**

975   ChIP-nexus analysis was performed as previously described[81]. Specifically, reads that passed

976   the Illumina quality filter were filtered for the presence of a fixed barcode. The fixed barcode

977   was used to demultiplex the samples, before being filtered from the reads by Cutadapt v4.3[70].

978   The random barcode on each read was retained to enable UMI based quantification. Any

979   remaining adapter sequences were trimmed, and reads were removed if they were less than

980   22bp in length. Reads were then aligned using Bowtie2[71] to a human genome (GRCh38) that

981   was modified to contain an additional chromosome with the relevant reporter construct

982   sequence. Duplicate reads were removed, and each sample was converted into two bigwig files,

983   one with strand specific information and one that aggregates the reads from both strands. The

984   sequence coverage across the reporter construct was then visualised using IGV. To enable

985   better visualisation of the disproportionate effect of CDK8i on elongation (Figure 4B), CDK8i

986   samples are scaled to the height of the DMSO samples. The scale reflects the number of reads

987   in the DMSO sample.

988

989   **RNA-seq**

990   RNA was extracted using the Qiagen RNeasy kit. RNA concentration was quantified with a

991   NanoDrop spectrophotometer (Thermo Scientific). Libraries were prepared using a plate based

992   in-house library prep method based on DRUG-seq[83]. The method results in 3' RNA-seq

993   libraries containing a UMI and a well ID on read 1, and the unique transcript information on

994   read 2. Libraries were sequenced on the NextSeq500 using the 75bp single end chemistry, run

995   with paired-end settings. 25bps was allocated for the Read1 and 50bps for Read2.

996

997   **RNA-seq analysis**

998   Fastq files were demultiplexed and mapped to the human genome (hg19) using STARsolo

999   v2.7.9a[84]. Downstream processing of the output counts matrix was then performed in R. Using

1000   the Seurat package[85], the raw counts matrix was transformed and then subject to differential

1001   gene expression analysis using DESeq2 v3.18[86]. Genes were classified as differentially

1002   expressed if they had an absolute fold change above 1.5 and a p-value <0.05. TNF target genes

1003   were defined as genes with a fold increase greater than 1.5 with a p-value < 0.05 and displayed

1004   a p65 binding defined by MACS2 v2.2.7 within 10kb of the gene.

SLAM-seq

SLAM-seq was performed as previously described with minor modifications[87]. Briefly, a total of 10^7 K562 cells were treated with dTAGV-^{1} (Tocris) as described above for a total of 4 hours. In the final hour of treatment, cells were labelled with 200uM of 4-thioruidine (4sU; Cayman Chemical) to capture nascent mRNA transcripts. RNA extraction was performed using TRIzol (Ambion) following the manufacturer's instructions, with the addition of 1mM DTT to the isopropanol precipitation and ethanol wash steps. Total RNA was eluted in nuclease-free H_2O containing 0.1mM DTT. 10ug of total RNA was treated with fresh 10mM iodoacetamide (Pierce) in reaction buffer containing 50mM Tris, pH 8.0 and 50% DMSO at 50°C for 15 minutes, light-protected and shaking at 1000rpm. The alkylation reaction was quenched with 20mM of fresh DTT, followed by adding 1ug of unlabelled Drosophila S2 RNA. Total RNA was cleaned up using the RNeasy MinElute Cleanup Kit (Qiagen) and DNAse treated. Libraries were prepared using 500ng of material using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD (Lexogen; V1 kit with single 6nt i7 indexes) according to the standard manufacturer's protocol. Sequencing of cDNA libraries was performed on the Illumina NextSeq2000 with 100bp single-end configuration. SLAM-seq was performed in biological triplicate.

SLAM-seq analysis

Quality assessment was performed on sequenced reads using FastQC v0.11.6 and adapters were trimmed using TrimGalore v0.6.6 and Cutadapt v4.3[70]. Read alignment to HG38 reference genome, read filtering, SNP calling and masking, and feature calling were performed with SlamDunk v0.2.4[88]. Reads containing at least 2 T>C conversions were retained as nascent transcripts. edgeR v3.38.1[89] was used to perform TMM normalisation, before performing differential expression analysis with limma voom v3.52.1[90]. Genes defined as down-regulated were those with a negative fold change value and an adjusted p-value < 0.05.

GAL4-AD protein expression levels and stability

As commercially available GAL4 antibodies performed unreliably in western blot, it was necessary to introduce an epitope tag to the GAL4-AD constructs. The DBD region of each GAL4-AD construct was replaced by a 3xFLAG tagged GAL4-DBD synthesized by IDT. FLAG-GAL4-ADs constructs were introduced at a high MOI into K562-Cas9 cells containing the 5xUAS mCMV GFP-PEST reporter to confirm that their transactivation capacity was not impacted by the addition of the FLAG tag. Cells were sorted based on the same relative levels of GAL4-AD expression as the original GAL4-AD lines used for the screens. This was

1039   achieved by direct comparison of the expression level of the IRES mCherry in the original AD

1040   cell lines and the newly derived AD cell lines. To assess protein expression and stability, one

1041   million of each of the FLAG-GAL4-AD cell lines were treated with either DMSO or

1042   cycloheximide (100uM) for 24hrs before being harvested for western blot.

1043

1044   **Immunoblot**

1045   Cells were lysed in RIPA lysis buffer (50mM Tris (pH 7.4), 150mM NaCl, 1mM EDTA, 1%

1046   Triton X-100, 0.5% sodium deoxycholate, 0.5% SDS supplemented with protease and

1047   phosphatase inhibitors (Roche)) and protein concentration was determined using a BCA

1048   protein assay kit (Pierce). Normalised concentrations of lysate were reduced and denatured for

1049   5 min at 95°C in Laemmli buffer containing 10% beta-mercaptoethanol and subsequently

1050   electrophoresed on 4–15% precast polyacrylamide gels (Mini-PROTEAN® TGX; Bio-Rad)

1051   under denaturing conditions. Proteins were wet transferred onto PVDF using the Mini Trans-

1052   Blot Electrophoretic Transfer Cell System (Bio-Rad) at 100V (400mA) for 60 min at 4°C.

1053   Membranes were blocked for 1 hour at room temperature in Intercept® Blocking Buffer (LI-

1054   COR) and subsequently probed with the following primary antibodies diluted at 1:1000 in

1055   blocking buffer supplemented with 0.1% Tween-20 overnight at 4°C on a roller: anti-HA-TAG

1056   (Cell Signaling, #2367), anti-MED12 (Bethyl Laboratories, #A300-774A), anti-MED14

1057   (Abcam; #ab72141), anti-MED25 (Abcam, #ab221741), anti-MYC (Abcam, #ab32072), anti-

1058   alpha-tubulin (Cell Signaling Technology; #3873), mouse anti-FLAG M2 (Sigma, #F3165),

1059   rabbit anti-LAMIN-B1 clone D4Q4Z (Cell Signaling, #12586). After washing, membranes

1060   were probed with the appropriate IRDye-conjugated secondary antibodies (LI-COR, 926-

1061   68071, 926-32210) diluted at 1:10,000 for 1 hour at room temperature. Membranes were

1062   scanned using an Odyssey ® Infrared Imaging System (LI-COR).

1063

1064   **Luciferase assays**

1065   To generate the reporter constructs for the luciferase assays, promoters of interest were cloned

1066   into the pGL4.35 (luc2p/9xgal4uas/hygro) luciferase construct (Promega, E1370) downstream

1067   of the 9xUAS site. Luciferase constructs were introduced into HEK293T cells by transient

1068   transfection using PEI. The luciferase construct of interest was co-transfected with or without

1069   the relevant GAL4-AD and the pRL Renilla control (Promega, E2261). Cells were harvested

1070   48hrs after transfection. Luciferase signal and Renilla signal were analysed using the Dual-

1071   Luciferase reporter system (Promega, E1910) using the Cytation 3 plate-reader (BioTek).

1072

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. Experiments were not randomized. The investigators were not blinded to the allocation during experiments and outcome assessment.

Data availability

All high-throughput sequencing data relevant to this study have been deposited in the NCBI Gene Expression Omnibus under primary accession code: GSE198944. All of the relevant source data has been provided with the manuscript.

Code availability

The manuscript does not include any custom code beyond implementation of pre-existing publicly available software packages. All computational analysis can be reproduced from the descriptions provided in the methods using the listed publicly available software.

Methods-only References

68. Burr, M. L. *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* 1–17 (2019) doi:10.1016/j.ccell.2019.08.008.

69. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828–863 (2017).

70. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10 (2011).

71. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

72. Michlits, G. *et al.* CRISPR-UMI: Single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat Methods* **14**, 1191–1197 (2017).

73. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675–686 (2015).

74. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559-2575.e28 (2022).

75. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).

76. Zhu, R., del Rio-Salgado, J. M., Garcia-Ojalvo, J. & Elowitz, M. B. Synthetic multistability in mammalian cells. *Science (1979)* **375**, (2022).

77. Liu, M. *et al.* Genomic discovery of potent chromatin insulators for human gene therapy. *Nat Biotechnol* **33**, 198–203 (2015).

78. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* **108**, 1531–1536 (2011).

79. 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).

80. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

81. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat Biotechnol* **33**, 395–401 (2015).

82. Shao, W., Alcantara, S. G. M. & Zeitlinger, J. Reporter-Chip-nexus reveals strong contribution of the drosophila initiator sequence to RNA polymerase pausing. *Elife* **8**, 1–25 (2019).

83. Ye, C. *et al.* DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat Commun* **9**, 1–9 (2018).

84. Kaminow, B., Yunusov, D., Dobin, A. & Spring, C. STARsolo : accurate , fast and versatile mapping / quantification of single-cell and single-nucleus RNA-seq data. 1–35 (2021) doi:10.1101/2021.05.05.442755.

85. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

86. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

87. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**, 1198–1204 (2017).

88. Neumann, T. *et al.* Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics* **20**, (2019).

89. 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).

1137   90.   Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. *Voom: Precision Weights Unlock*
1138   *Linear Model Analysis Tools for RNA-Seq Read Counts. Genome Biology* vol. 15
1139   http://genomebiology.com/2014/15/2/R29 (2014).

A

EF1a | Gal4-DBD (1-147) | AD | IRES | mCherry

Gal4-DBD — AD — Coactivators X,Y,Z

5xUAS — minimal promoter (mCMV) | Fast-degrading GFP

B

VP64 (aa 1-64)

c-MYB (aa 275-334)

EWSR (aa 1-279)

NF-κB (P65) (aa 315-551)

P53 (aa 1-81)

IRF1 (aa 84-241)

PU.1(aa 1-118)

NOTCH (aa 2090-2556)

GR (aa 524-777)

C

Gal4-NfKB
5xUAS-mCMV-Turbo-PEST

Dead cells/debris

Live cells

- DMSO
- 24hr Triptolide

- DMSO
- 2hr Triptolide
- 4hr Triptolide
- 6hr Triptolide
- 8hr Triptolide
- 16hr Triptolide
- 24hr Triptolide

D

Triptolide 10uM

E

| | Gal4 KO | POLR2A KO | MED25 KO | P300 KO |

Gal4-VP64

Gal4-MYB

- Safe guide control
- Guide 1
- Guide 2

Dead cells/debris

Live cells

- Safe guide control
- Pol2 sgRNA - D5 post infection

Dead cells/debris

Live cells

- Safe guide control
- Pol2 sgRNA - D5 post infection

A

AAVS knockin Clone 1

AAVS knockin Clone 2

B

r=0.918

C

r=0.830

D

VP64   MYB   EWS   NF-κB   IRF1

P53   PU.1   NOTCH   GR

Non-targetting
Gal4 sgRNA1
Gal4 sgRNA2

A Day 0 → Day 5 → Day 6 → Day 7 → Day 14

ADs screened
VP64, MYB, EWS
NF-κB, P53, IRF1
PU.1, NOTCH, GR

gRNAs for 1.3k transcription regulators

D14 library control

Dropout

D5 library control — D5 screen — D6 library control — D6 screen — D7 library control — D7 screen

Dropout analysis
Guide abundance
D0   D14

Screen enrichment
AD1 AD2 AD3 ...
Gene X
Gene Y
Gene Z
...

B

RNA polymerase initiation and elongation
Mediator complex
SWI/SNF
SET/COMPASS complex
Ribosome biogenesis

C All hits
GR NOTCH P53 EWS MYB IRF1 PU.1 NFKB VP64

Fold enrichment
8 6 4 2 1

D
EWS GR IRF1 MYB NF-κB NOTCH PU.1 VP64 P53

Significantly enriched in every screen

Genes required for cell growth

Fold enrichment
8 6 4 2 1

Not significantly enriched in any screen

E
Mean toxicity score (DEPMAP)
Mean fold enrichment (TAD screens)

r = -0.54

#TFs require
0 1 2 3 4 5 6 7 8 9

POLR2H
POLR2G
POLR2K POLR2F
POLR2C
POLR2D
POLR2E
POLR2B
POLR2L

F Essential genes defined by D14 dropout
Average guide counts per gene relative to Day 0
Non-essential genes
Essential genes
Day 5 Day 6 Day 7 Day 14

G Essential genes defined by DEPMAP
Average guide counts per gene relative to Day 0
Non-essential genes
Essential genes
Day 5 Day 6 Day 7 Day 14

H Essential genes defined by D14 dropout
Coefficient of variation
n.s
Non-essential   Essential

Essential genes defined by DEPMAP
Coefficient of variation
*
>
Non-essential   Essential

A Integrator complex

AD legend

B

Safe sgRNA
INTS5 sgRNA
Safe sgRNA +TNF
INTS5 sgRNA +TNF

ICAM1 APC

Safe sgRNA
INTS5 sgRNA
Safe sgRNA +TNF
INTS5 sgRNA +TNF

CD69 PerCP-Cy5-5

C RNA-seq

No TNF   TNF

Log 2 Fold change
1.5
1
0.5
0
−0.5
−1
−1.5

TNF target genes

NF-κB ChIP-seq

No TNF   TNF

Read counts
80
70
60
50
40
30
20
10
0

NfKB peaks around TNF targets (all peaks around each gene)

1kb 0 1kb   0 1kb

D

p=0.01

Relative RNA pol2 levels (TNF vs. No TNF)

TNF target genes
Other genes

Safe Control   INTS5 KO

E

Relative RNApol2 levels (INTS5 KO + TNF vs. Safe + TNF)

TNF target genes

Ranking of genes by effect size

F

Safe
Safe + TNF

NF-κB

Safe
INTS5 KO
Safe + TNF
INTS5 KO + TNF

RNA pol2

ICAM1

G

Safe
Safe + TNF

NF-κB

Safe
Safe + TNF
INTS5 KO
INTS5 KO + TNF

RNA pol2

MYC

A

B All hits required for EF1a expression

Many EF1a dependent genes are not required for all ADs and no clear simple pattern of certain ADs requiring the same factors as EF1a

C 50 most heterogeneous cofactors ranked by requirement for EF1a

A number of the top heterogenuously required cofactors are **not** required by EF1a.

Those that are, do **not** display a simple pattern that would indicate their heterogenity is caused by indirect effects

D

E

A Including p53

CDK8i   r²=0.18, n.s
CDK8 KO r²=0.54, *
CCNC KO r²=0.59, *

p53

Fold reduction in GFP signal

Elongation index

B Excluding p53

CDK8i   r²=0.91, ***
CDK8 KO r²=0.80, **
CCNC KO r²=0.87, ***

Fold reduction in GFP signal

Elongation index

A

EF1a | Gal4-DBD (1-147) | NFκB | IRES | mCherry

Coactivators
NFκB
Promoter bound factors
Gal4-DBD

5xUAS | 9x Core Promoters | Fast-degrading GFP
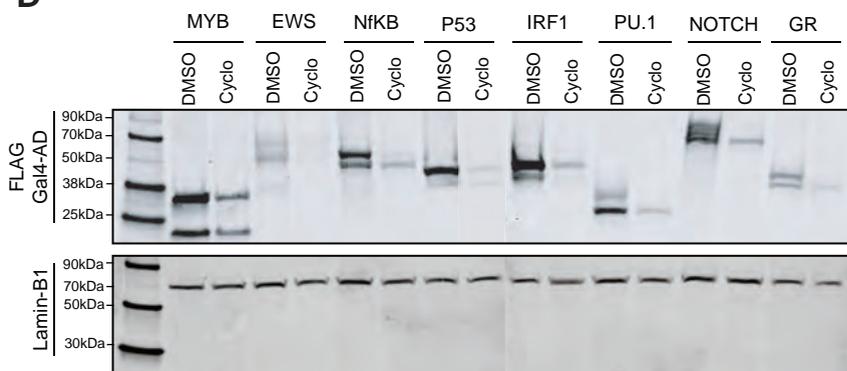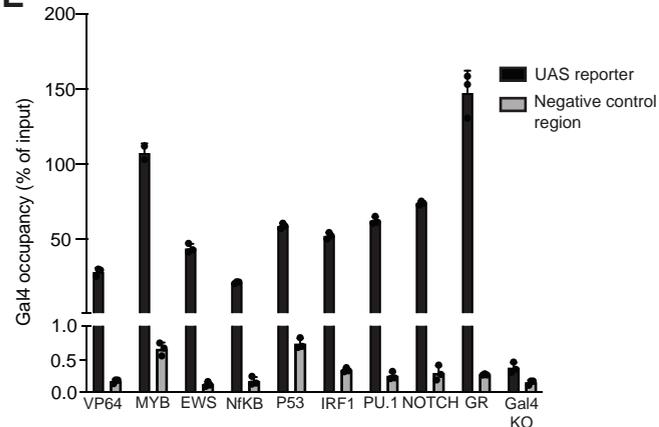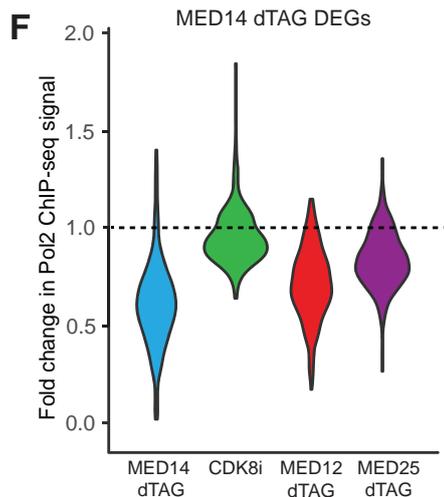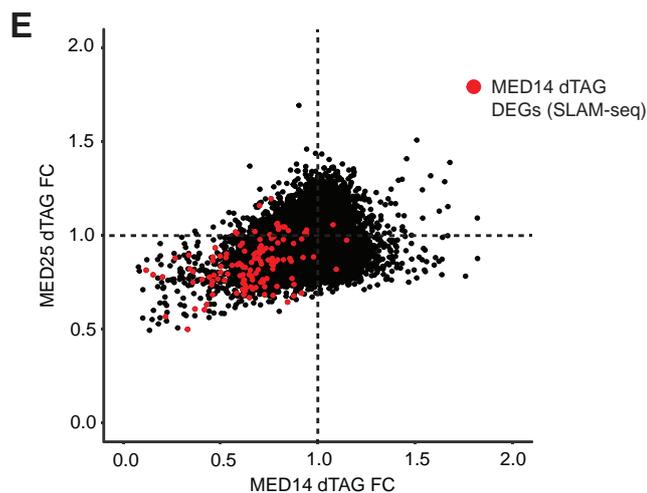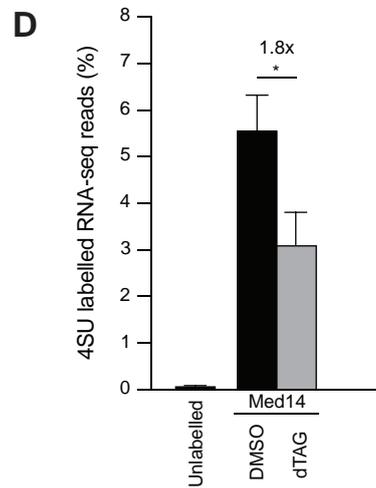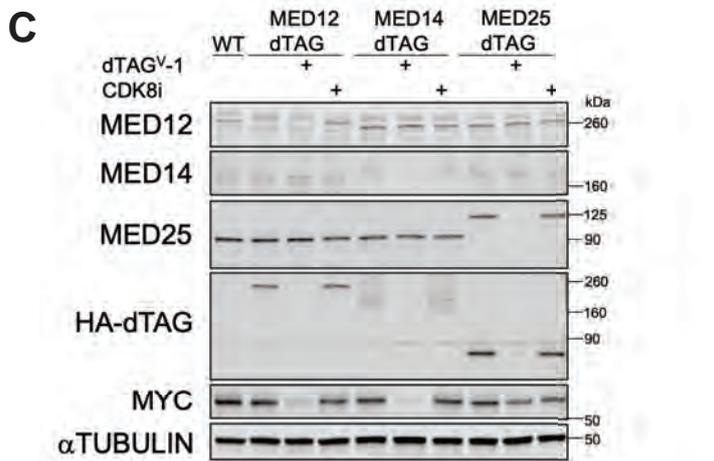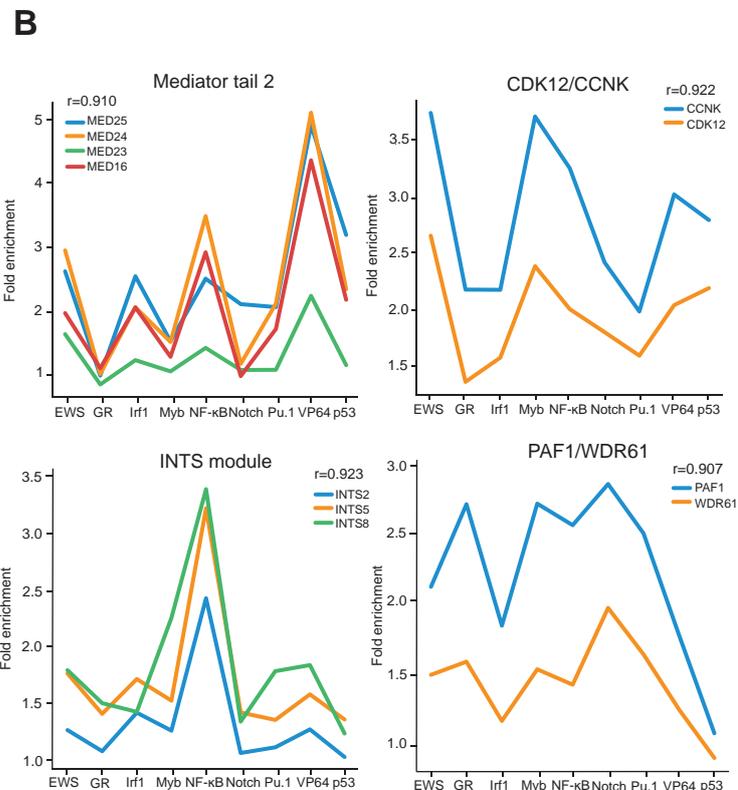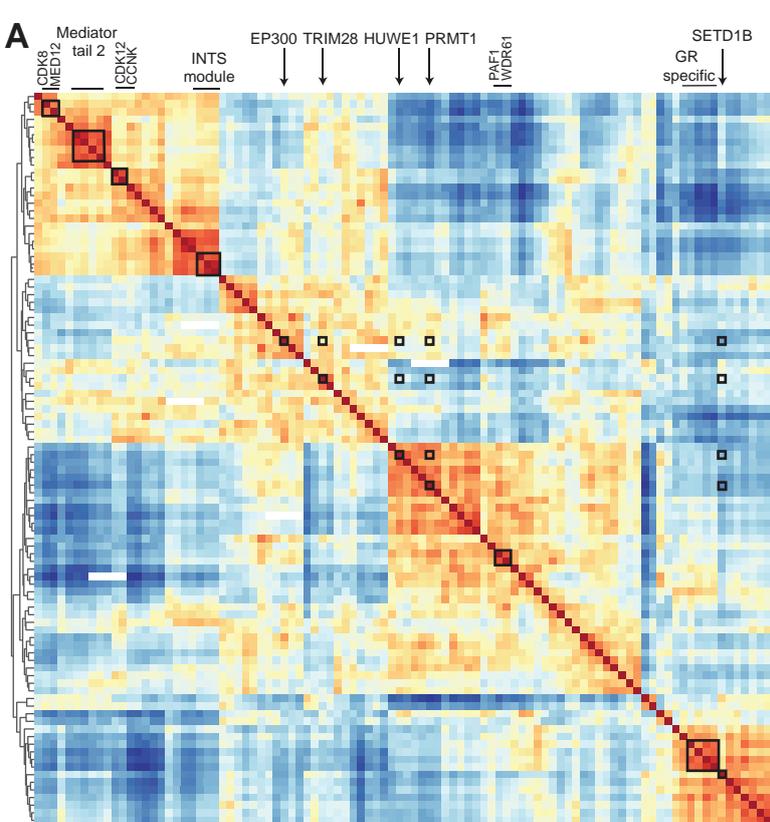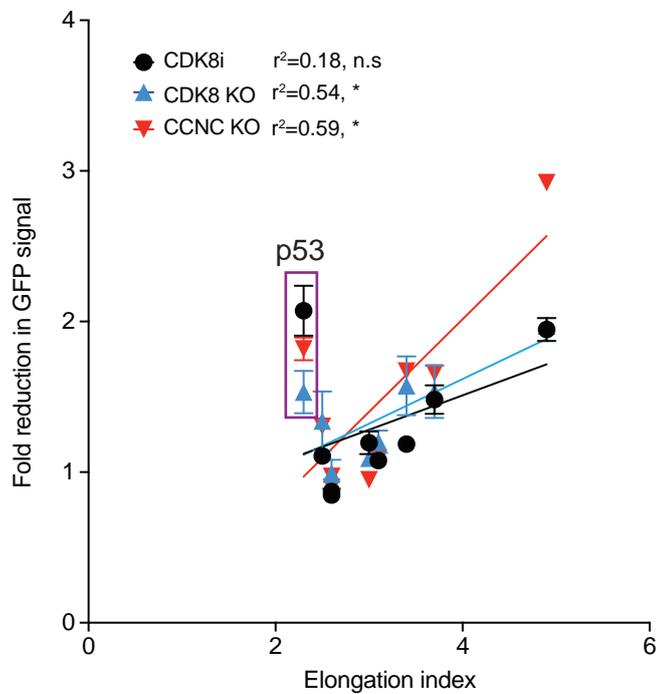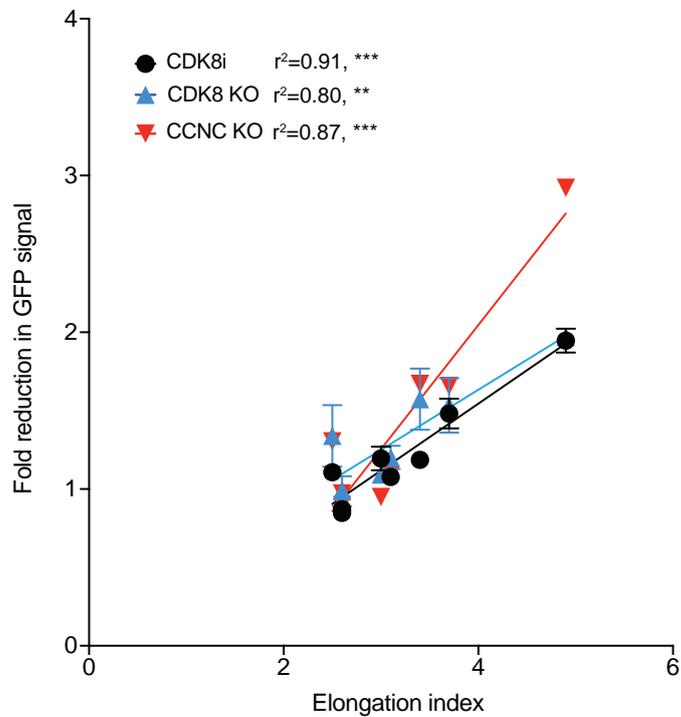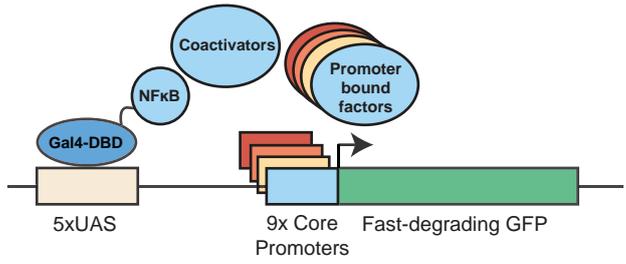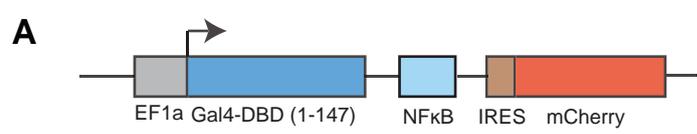
B

TATA box promoters

TSS

mCMV   TAGGCGTGTACGGTGGGGAGGCC**TATA**TAAGCAGAGCTCGTTTA<u>GTGAAC</u>GTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGC

SCP   TAGGTCTA**TATA**AGCAGAGCTCGTTTAGTGAACC<u>GTCAGAT</u>CGCCTGGAGACGTCGAGCCGAGTGGTTGTGCCTCCATAGAA

YB-TATA TCTAGAGGG**TATA**TAATGGGGGCCA

MLP   TGGGGC**TATA**AAAGGGGGTGGGGGCGCGTTCGT<u>CCTCACT</u>CTCTTCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGA

HSP1A1 TGAAAAGGCGGGTCTCCGTGACGACT**TATA**AAAGCCCAGGGGCAAGCGGTCC<u>GGATAAC</u>GGCTAGCCTGAGGAGCTGCTGC

TATA-like / weak TATA promoters

HSV-TK CGCATATTAAGGTGACGCGTGTGGCCTCG<u>AACACCG</u>AGCGACCCTGCAGCGACCCGCTTAA

HBB   TCCCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT<u>ACATTT</u>GCTTCT

TCT promoters

MYB   ATCAACCTGTTTCCTCCTCCTCCTTCTCCTCCTCCTCCGTGACCTCCTCCTCCTCCTCTTTCTCCTGAGAAACTTCGCCCCAGCGGTGCGGAGCGCCGCTGCGCAGCCGGGGAG

RPL30   CGGTCCCGCAGTTCCGGCTCTGCCGTGAAGAGCTTTGCATTGTGGGAAGTCTTTCCTTTCTCGTTCCCCGGCCATCTTAG

C
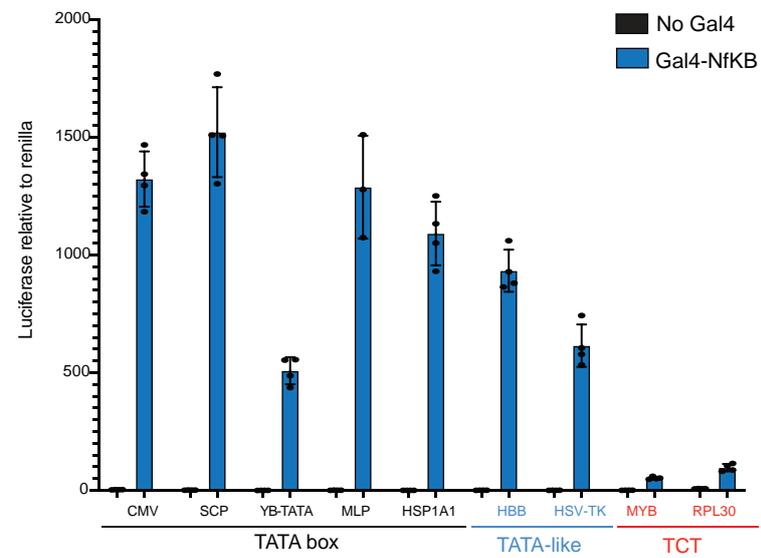
Supplementary Note 1 – designing and analysing comparative CRISPR screens

Our manuscript pioneers the use of comparative CRISPR screens. Most prior studies that have used CRISPR screens have been designed with the intention of identifying individual hits that regulate a specific phenotype of interest[1]. These hits are then validated and studied in further detail. For that purpose, it is not necessary to obtain a highly quantitative measure of the degree of enrichment of individual genes, as the hits do not need to be compared across different screening conditions. Here, our goal was to *compare across* screens to obtain more general insights into the specificity of transactivation domains and core promoters. We hoped to use the screens not only to identify important cofactors, but to provide insights into their degree of requirement across different conditions. This requires a relatively reproducible and precise measure of guide enrichment.

Since most CRISPR screens are not designed to obtain quantitative information, the analysis tools currently available are not designed to handle multiple timepoints, or to compare enrichment values across screens. As a result, it was necessary for us to develop bespoke screen design and analysis methods that would provide the most accurate quantitative information. In this Supplementary Note, we discuss the rationale and provide justification for the screen design and analyses steps we performed. We also demonstrate how the inclusion of certain filters in the analyses improve the reproducibility of the data.

Throughout, we benchmark our analysis approach by using two independent biological replicates performed on GAL4-NfKB activating the minimal CMV promoter. Upon processing the data according to our pipeline, the two replicates have a very high degree of correlation (r=0.918, Pearson), demonstrating that our design and analysis pipeline produces highly reproducible data.
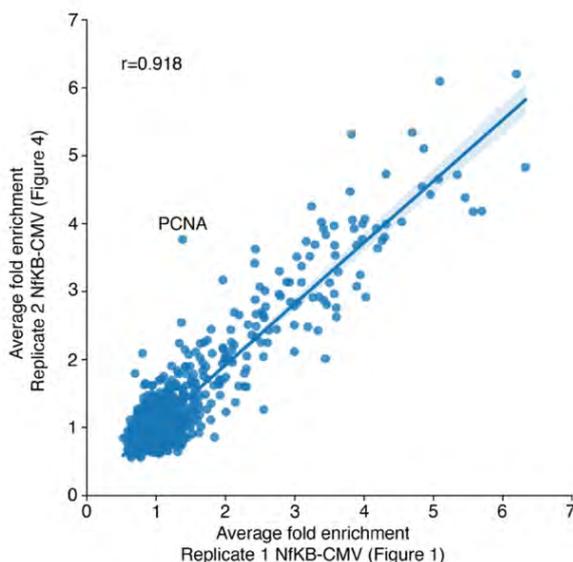
Supplementary Note Fig. 1) Correlation between the fold enrichment scores in the two replicates of the GAL4-NfKB mCMV screens. r = Pearson correlation. PCNA is outlined as the only major outlier. Error bands reflect 95% confidence interval of Pearson correlation.

Demonstrating the importance of filtering outliers for accurate quantification of the screen data
In the analysis of CRISPR screen data, fold enrichment values are calculated for each guide in the library by dividing the number of counts in the screen sample by the number of counts in

the matched library control. Our library contains 6 different guides targeting each gene and we isolated samples at 3 timepoints (D5, D6 and D7 after guide library infection). This resulted in a total of 18 guide enrichment values for each gene in each screen. These 18 enrichment values were averaged, resulting in a single fold enrichment value per gene per screen.

Upon initial inspection of the raw data, we noticed single sgRNAs targeting a gene with fold enrichment values of up to 500-fold, when other independent sgRNAs against the same gene were not enriched. These are examples of outliers potentially caused by random sampling variation, or due to random overamplification in the PCR step, which is an issue that has been previously reported in CRISPR screens[2]. Inclusion of these guides in the analysis would obscure the calculation of an accurate fold enrichment value. Consequently, we realised that some filtering of the guides would be necessary to remove outliers.

The following filtering steps were applied during the processing of the data:
1) Removal of guides with low guide counts in the plasmid pool (lowest 2.5% of guides)
2) Removal of guides that have a fold enrichment value below 0.1 or above 10-fold
3) Removal of guides that have a fold enrichment value that > 4-fold from the mean fold enrichment value (mean of 18 values – i.e., all 6 sgRNA's for a particular gene assessed at D5, D6, D7)

We demonstrate how each of these steps helped to improve the quality of the data below.

Removal of lowly represented guides:
To reduce the prevalence of outlier guides, we initially removed guides with low counts, which is a common processing step in CRISPR screens[3]. Low representation can cause sampling error in the dataset resulting in erratic fold change values. Prior studies remove the bottom 5% of guides[3], however due to our low skew ratio (4.38) and our desire to retain as many guides as possible, we decided to only remove the bottom 2.5% of guides.

Filtering guides with fold enrichment below 0.1 or above 10:
Even upon removal of these lowly represented guides, we still observed guides with unrealistic, high fold enrichment values that were not correlated between biological replicates (Supplementary Note Fig. 2, top right). These guides interfere with an accurate calculation of fold enrichment values. Therefore, we reasoned that further filtering was necessary. In order to set appropriate thresholds for filtering outlier guides from the screen data, we leveraged our two biological replicates of the GAL4-NfKB mCMV screen. We noted that none of the sgRNAs with fold enrichment values above 10 were correlated between the two replicates suggesting that a score above 10 is very unlikely to reflect a real enrichment score (Supplementary Note Fig. 2, top right). Therefore, we filtered any guides with a fold enrichment value above 10. We also reasoned that a depletion of 10-fold was just as likely to be due to sampling error as an enrichment above 10, so as not to bias the data in a specific direction, we also removed guides with a fold enrichment below 0.1. This filter dramatically improved the reproducibility between the replicates (Supplementary Note Fig. 2, bottom left).

Filtering guides with fold enrichment value > 4-fold from the mean:
The final filter we applied, removed guides that were four-fold away from the mean fold change calculated from the 18 individual enrichment values (i.e., all 6 sgRNAs for a particular gene assessed at D5, D6, D7). This filter was applied because we noted that some target genes were being classified as significant due to a single outlier guide (enrichment value < 10 and > 0.1) that caused the fold enrichment for that gene to increase beyond the significance cut-off.

Introduction of this filter clearly reduces the dispersion of the data (observable by the reduced spread of the values around a fold change of 1). This filter also increased the Pearson correlation between replicates from r=0.890 to r=0.918 (Supplementary Note Fig. 2, bottom right).
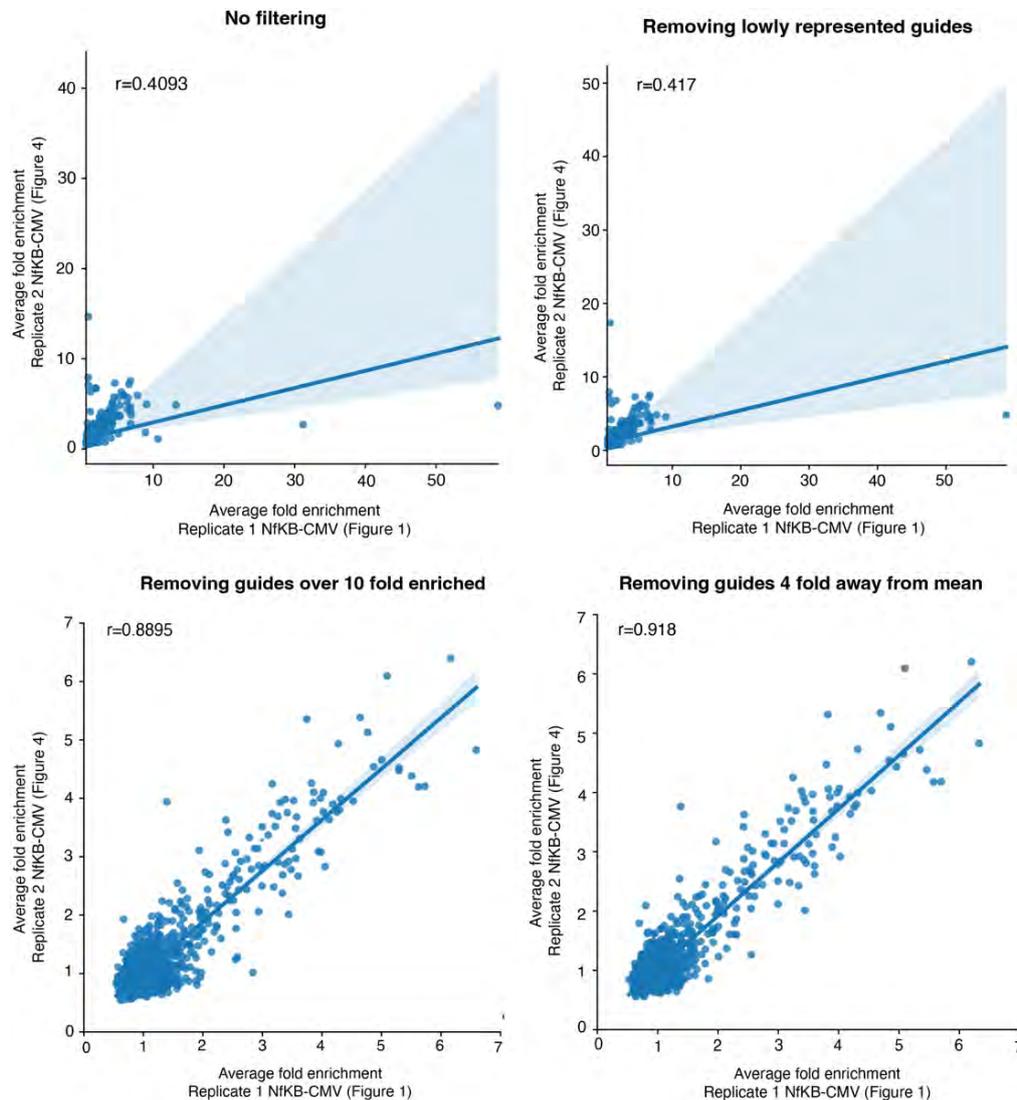
Supplementary Note Fig. 2) Correlation plot of fold enrichment scores of all genes in the library across two replicates of GAL4-NfKB mCMV screen with different guide filtering applied. (Top Left) No filter applied. (Top Right) Removing bottom 2.5% represented guides in the plasmid pool. (Bottom left) removing guides with a fold enrichment over 10-fold. (Bottom right) removing guides with a fold enrichment more than 4-fold away from the mean value of all 18 guides. r=Pearson correlation. Error bands reflect 95% confidence interval of Pearson correlation.

The three filters we applied to the raw guide values were the only methods of post-processing performed on the raw data prior to the calculation of a fold enrichment value. By using the biological replicates of the same screen, we can see that each of these filtering steps increased the reproducibility and were necessary to obtain an accurate fold enrichment value.

Demonstrating the importance of using multiple timepoints
Based on our prior experience with CRISPR screens[4–9], we reasoned that it would be necessary to include multiple timepoints as part of our screen design to reduce the potential effects of sampling error, by increasing the number of sampling events per guide. Since our screens were

performed at a relatively early timepoint, we also reasoned that using multiple timepoints would help to minimise overlooking guides that work with different kinetics. Once again, we can compare the results from the two independent biological replicates, to assess how aggregating across the three timepoints (D5, D6 and D7 post guide infection) helped to improve the reproducibility of the results.

To assess how aggregating the timepoints impacts the data, we can compare the correlation between individual timepoints from our two biological replicates, with the correlation when all three timepoints are aggregated. We found that the day 5 timepoint has greater variability between the two biological replicates, than the day 6 and day 7 timepoint (Supplementary Note Fig. 3), which suggests that while day 5 provides a window into the earliest changes that occur after gene knockout, these early changes are more prone to variability likely due to kinetic differences in sgRNA integration, expression and editing efficiency. At day 6 and day 7, the enrichments are far more consistent across the two biological replicates (Supplementary Note Fig. 3). However, aggregating the data across all three timepoints produces the highest correlation between the two biological replicates (Supplementary Note Fig. 3).
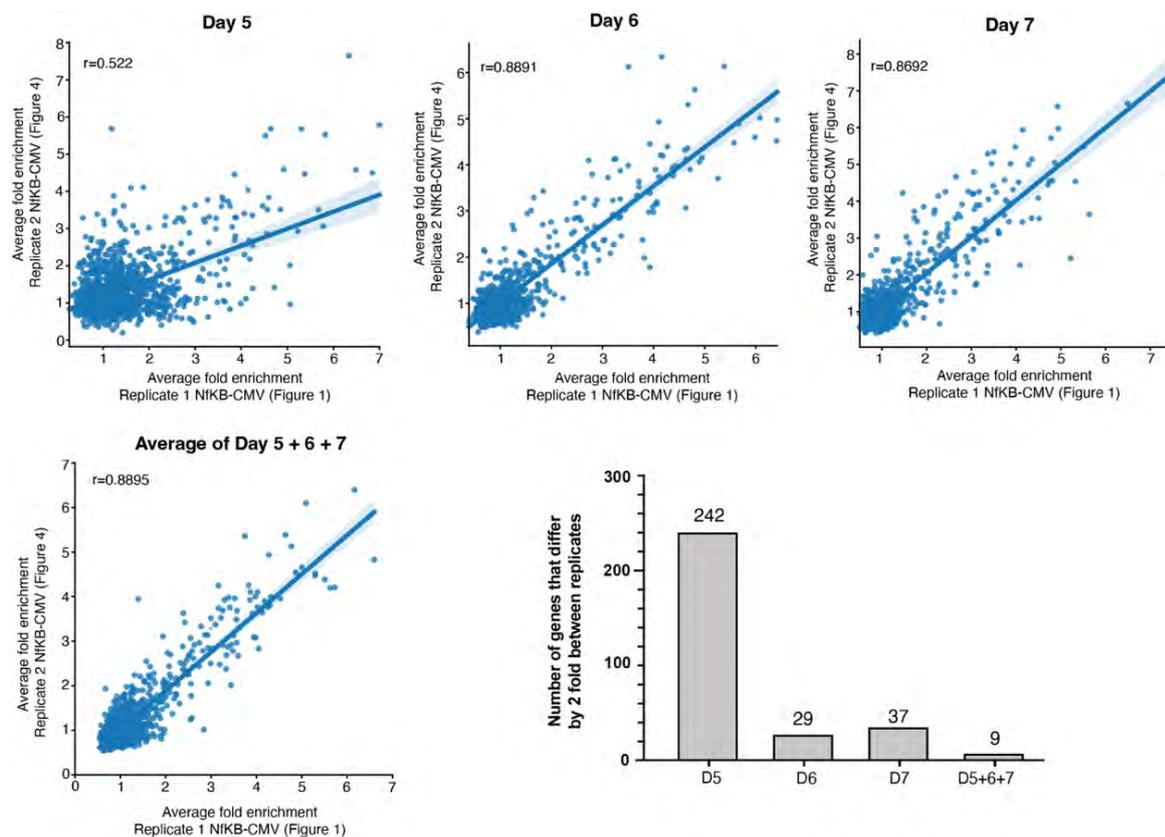
Supplementary Note Fig. 3) Correlation between two NfKB CMV biological replicates at each of the screen timepoints, as well as the data aggregated across the 3 timepoints. Each dot represents the fold enrichment of the 6 guides that target each gene, other than the aggregated plot, which represents the average of all 18 datapoints. R = Pearson correlation. (Bottom left) Bar plot shows the number of genes that differ in fold enrichment values by at least 2-fold between the two biological replicates at each timepoint and in the aggregated data. Error bands reflect 95% confidence interval of Pearson correlation.

To further reinforce the value of aggregating the three timepoints, we calculated how many genes display at least a 2-fold difference in enrichment values between the two biological

replicates. This analysis demonstrates that aggregating the data dramatically reduces the number of genes with a greater than 2-fold difference between replicates (Supplementary Note Fig. 3, Bottom Right). Only 9 genes display a difference of at least 2-fold between the two biological replicates when the data is aggregated. This reflects a greater than 3-fold improvement over the most reproducible individual timepoint (day 6). Overall, by aggregating across three timepoints, we can both capture the earliest consequences of gene knockout, while improving the reproducibility of the screens.

Permutation test approach to calculate an empirical p-value for each gene
The majority of the data presented throughout the manuscript uses the fold enrichment score to represent the requirement for a particular cofactor. For the most part, we avoided categorizing the data unnecessarily, as cofactor requirement appears to exist on a spectrum and is therefore not a categorical variable. However, for some of the analyses, it was helpful to define which genes were statistically significantly enriched in each screen. It is important to point out that classification as a hit based on statistical significance does not have any impact on the calculation of the fold enrichment score.

In order to calculate which genes should be considered significantly enriched in each screen, we used permutation testing to calculate empirical p-values. Permutation testing works by randomly sampling from the data to obtain an empirical distribution of the values. We obtained this distribution by randomly sampling 18-fold enrichment values (6 from each timepoint) from the 7240 guides in the library. Sampling was performed 10000 times on the data that had been pre-filtered to remove outliers (using the above methods). An independent permutation test was performed for each screen to ensure that we were accounting for screen specific differences in the distribution of values. The cut-off to categorize a gene as statistically significant was a fold enrichment score above the 95th percentile of this random sampling distribution, which equates to an empirical p value < 0.05. The 95th percentile of the random distribution was similar across all of the AD screens suggesting that the distribution of guide values was very similar across all of the screens.

The values that correspond to the 95th percentile of this random sampling distribution are provided in a table below:

AD screen	Enrichment score for p<0.05
GAL4-VP64	1.48
GAL4-MYB	1.46
GAL4-EWS	1.44
GAL4-NfKB	1.60
GAL4-P53	1.44
GAL4-IRF1	1.48
GAL4-PU1	1.54
GAL4-Notch	1.57
GAL4-GR	1.61
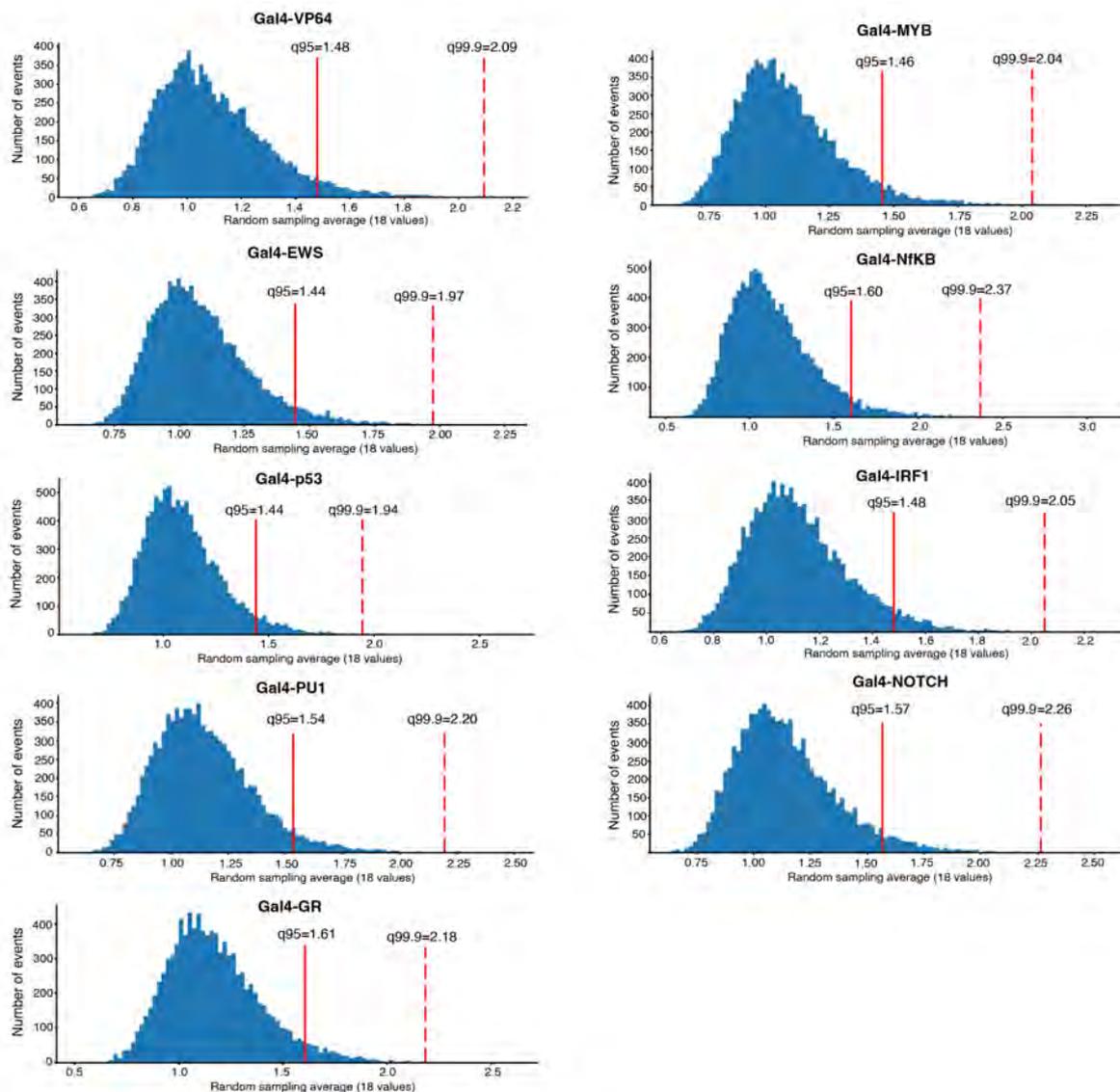
Supplementary Note Fig. 4) Random sampling distributions obtained from permutation test performed on each screen. Solid red line reflects the 95th percentile of the distribution which corresponds to an empirical p-value < 0.05. The dashed red line reflects the 99.9[th] percentile, which corresponds to an empirical p-value <0.001.

To provide extra stringency to what is classified as a hit, we also reasoned that genuine hits should have multiple guides enriched. Therefore, we applied an additional cut-off for significance that at least $1/3^{rd}$ of the guide values (6 out of 18 values) should display a fold enrichment score above the 95th percentile. This cut-off did not have a dramatic effect on the number of genes classified as hits, but did further increase the stringency of the data, by reducing the impact of certain individual guides.

Obtaining quantitative information from CRISPR screen data

As stated above our ambition for the comparative CRISPR screens was to obtain relatively quantitative data about the degree of requirement for different cofactors across a range of ADs and core promoters. Generally, CRISPR screens are considered to provide a binary output i.e., a gene is either identified as required or not. This is largely due to the design of the screens and the gates used to isolate cells displaying a particular phenotype. For our screens, as demonstrated above and throughout the manuscript, we were able to achieve relatively accurate

and sensitive quantification of the transcriptional effects of different cofactors. This was achieved by setting the gates for isolating cells at an intermediate distance (0.25 of the Mean Fluorescence Intensity (M.F.I) from the initial normal distribution of GFP expression (Supplementary Note Fig. 5). By setting the gate at an intermediate distance, we were able to isolate a different number of cells in the GFP-negative gate depending on the degree of effect on transcription (Supplementary Note Fig. 5, left). Setting the gate too close to the MFI of the unperturbed cells would result in genes with large effects on transcription having similar enrichment to genes with low effect sizes. Setting the gate too far from the main population of would result in genes with smaller effect sizes not being identified, reducing the sensitivity of the screens.
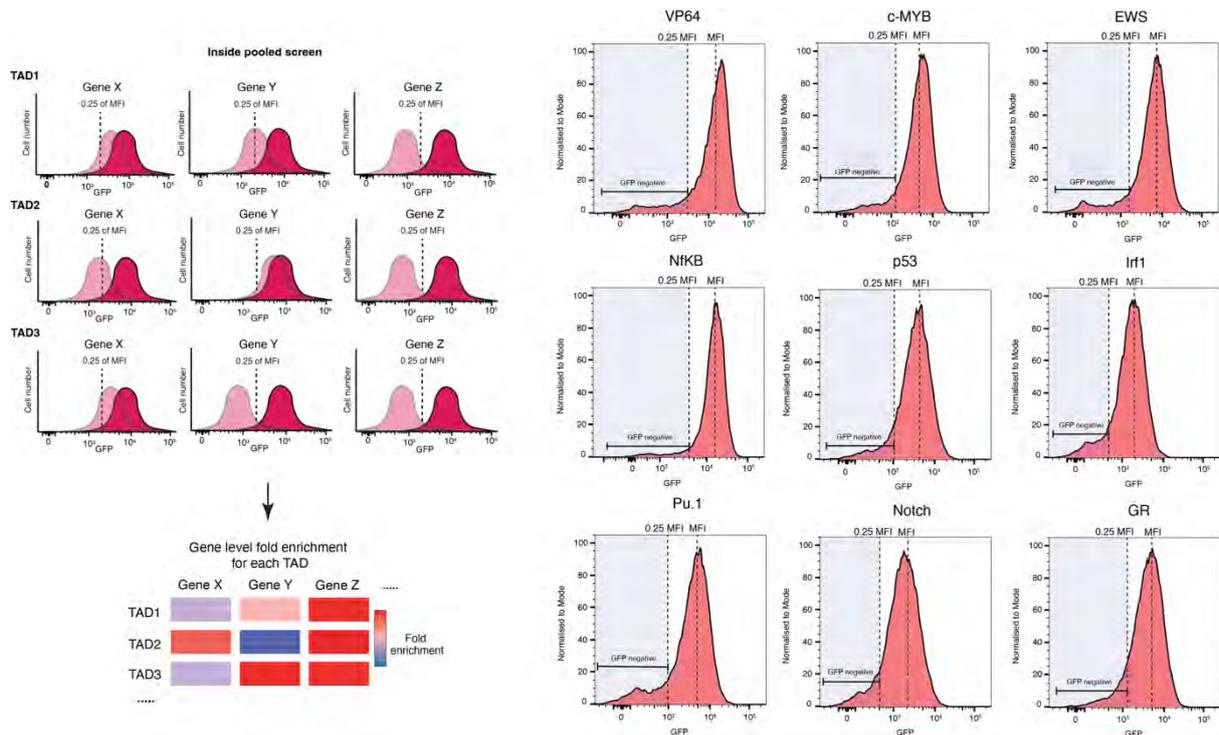
Supplementary Note Fig. 5) (Left) Schematic demonstrating the logic of the gating strategy that enables quantitative data to be obtained from a single gated-CRISPR screen. (Right) FACS gating strategy applied to each of the samples in the AD-screens. M.F.I = mean fluorescence intensity (i.e., average GFP signal).

Various pieces of evidence suggest that our screens resulted in quantitative data. The high degree of correlation in fold enrichment values (Extended Data Fig. 2B), correlation between the degree of requirement and specificity of the cofactors (Fig. 2A) support the quantitative nature of the screens. Likewise, our co-dependency analysis identified highly concordant patterns of requirement across different subunits in the same complex, which would not have been observable if the screen data was not highly quantitative (Extended Data Fig. 8B). Our validation experiments also demonstrate that there is a high degree of concordance between the fold enrichments identified in the screen and the fold change in transcription in our validation experiments (Fig. 2C-D). Altogether, these data demonstrate that our gating approach was capable of obtaining quantitative data from CRISPR screens. The accuracy and reproducibility of the quantification is likely the result of the many design and analyses choices detailed above.

Supplementary Note References

1. Michlits, G. *et al.* CRISPR-UMI: Single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat Methods* **14**, 1191–1197 (2017).
2. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675–686 (2015).
3. Burr, M. L. *et al.* CMTM6 maintains the expression of PD-L1 and regulates anti-Tumour immunity. *Nature* **549**, 101–105 (2017).
4. Sparbier, C. E. *et al.* Targeting Menin disrupts the KMT2A/B and polycomb balance to paradoxically activate bivalent genes. *Nat Cell Biol* (2023) doi:10.1038/s41556-022-01056-x.
5. Burr, M. L. *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* 1–17 (2019) doi:10.1016/j.ccell.2019.08.008.
6. Bell, C. C. *et al.* Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. *Nat Commun* **10**, 2723 (2019).
7. MacPherson, L. *et al.* HBO1 is required for the maintenance of leukaemia stem cells. *Nature* **577**, 266–270 (2020).
8. Gilan, O. *et al.* CRISPR–ChIP reveals selective regulation of H3K79me2 by Menin in MLL leukemia. *Nat Struct Mol Biol* (2023) doi:10.1038/s41594-023-01087-4.