The logic of native enhancer-promoter compatibility and cell-type-specific gene expression variation

Takeo Narita[1], Yoshiki Higashijima[1], Sinan Kilic[1], Elina Maskey[1], Katrin Neumann[2], Chunaram Choudhary[1,*]

[1]Department of Proteomics, The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark.

[2]Stem Cell Engineering, Center for Molecular and Cellular Bioengineering, Biotechnology Center, Technische Universität Dresden, 01307 Dresden, Germany.

*Correspondence to: chuna.choudhary@cpr.ku.dk

Abstract

Cis-regulatory enhancers are essential for differential expression of developmental and housekeeping genes. However, the specificity of native mammalian enhancers and how it shapes cell-type-specific gene expression landscapes remain largely unknown. We show that endogenous enhancers are broadly compatible with the promoters of developmental and housekeeping genes. Broad enhancer compatibility affords 'retrofitting' new regulatory capabilities to housekeeping genes that evolved before the advent of enhancers. This enables cell-type-specific tuning of ubiquitously expressed genes. Segregation between enhancer-dependent and –independent type regulation is blurred. Within the same cell type, a single promoter can be activated by enhancers and non-enhancer promoter-regulatory elements (PREs). It is the tunable and integrated strengths of enhancers and PREs that quantitatively shape gene expression landscapes, within and across cell types. Our findings have broad implications for understanding cell-type-specific quantitative gene expression variation, as well as the emergence and rewiring of gene regulatory networks in disease and organismal evolution.

Introduction

Gene transcription in metazoans is controlled by promoters and enhancers[1]. An autonomous promoter consists of a core promoter (CP) and proximally occurring regulatory sequences (hereafter non-enhancer promoter regulatory elements; PREs). CPs define the transcription start site (TSS) and recruit the basal transcription machinery. PREs recruit transcription factors (TFs) and cofactors (COFs) and provide an activation signal to the CP. Many cell-type-specific (CTS) gene promoters lack potent PREs, and instead, are activated by enhancers, which act orientation and distance independently and can be located hundreds of kilobases away from the target CP[2,3]. In mammals, enhancers outnumber genes, many enhancers skip their nearest genes and contact distally located promoters, multiple promoters interact with one enhancer, multiple enhancers interact with one promoter, and enhancers can coalesce to form phase-

separated condensates[4-10]. A key question is, in such environments, how enhancers specifically activate some genes without affecting others.

In different analyses, enhancer-promoter (E-P) compatibility shows considerable differences. An early study reported that enhancers lack specificity and similarly activate diverse promoters[11]. Some of the recent analyses confirm this idea on a larger scale[12,13]. However, in other analyses, enhancers show remarkable specificity and activate some but not other promoters[9,14-17]. In massively parallel reporter assays (MPRAs), a majority of *Drosophila* and mouse enhancers show promoter specificity[18,19], and the compatibility information appears to be sequence-encoded in the enhancers and promoters[18,20]. Supporting E-P specificity, different groups of human enhancers use different cofactors (COFs)[21], and different COFs activate different promoters in *Drosophila*[22]. These findings present two seemingly diverging models of enhancer specificity. In one model, enhancers are broadly compatible with diverse promoters[11-13], while in another model, E-P compatibility is restricted such that enhancers activate selected promoters with high specificity[14-19]. An attractive aspect of the sequence-encoded E-P compatibility model is that it can rationalize why enhancers frequently skip their nearest genes and connect with distal genes[4-10], how specificity can be achieved in phase-separated enhancer condensate hubs[23-25], and how enhancers differentially activate alternative promoters within the same genes[22].

One of the limitations is that the current understanding of E-P specificity is largely derived using plasmid-based ectopic reporters. Configurations of native cis-regulatory elements are far more complex, and arguably, endogenous enhancer specificity could be affected by many additional factors, such as 3D genome organization, chromatin modifications, competition among promoters, and binding of proteins that can facilitate or impede specific E-P interactions[2,26-30]. Supporting this notion, a recent study identified thousands of candidate silencer elements in mammalian cells[31], and 3D contact frequency suggests that the closest gene is often not the likely target of native enhancers[4,32,33]. Because of these complexities in the native environment, two fundamental questions remain unresolved. (1) Do enhancers display restricted or broad promoter specificity in the endogenous context? (2) What is the basis of cell-type-specific gene expression variability, and how does E-P compatibility impact quantitative gene expression variation?

Here, we use a reporter-free approach to infer native E-P specificity on a genome-wide scale. Importantly, we use native E-P compatibility principles to explain how E-P compatibility shapes endogenous transcription landscapes.

Results

Considerations for studying native enhancer specificity

A major unresolved challenge in the field is to find generalizable principles of enhancer-dependent gene regulation in mammalian cells, in the native context, on a global scale, across diverse cell systems. We reasoned that it should be possible to infer native E-P compatibility if we could: (1) define the cell-type-specificity of the expressed genes, (2) map enhancers and define the distances between enhancers and their nearest active promoters, (3) estimate the

87 strength of enhancers, and (4) comprehensively quantify the enhancer-dependence of gene
88 expression. We further posited that, by performing these analyses across diverse cell types, it
89 should be possible to quantitatively assess enhancer and PRE function in cell-type-specific gene
90 expression variation.
91
92 CBP/p300-dependent enhancers constitute the largest and best-characterized group of
93 mammalian enhancers. To our knowledge, it is the only known group of enhancers that can be
94 mapped using chromatin marks, whose strength can be estimated (see below), and whose
95 activity can be rapidly ablated on a global scale[34]. Even in plasmid-based reporter assay,
96 CBP/p300 is uniquely required for the activity of natively accessible enhancers (**Extended Data**
97 **Fig. 1; Supplemental note 1**). For these reasons, we used CBP/p300-dependent enhancers as a
98 model for understanding the rules of enhancer-dependent transcription.
99
100 **Determination of cell-type-specificity of genes**
101 To investigate enhancer specificity and cell-type-specific gene regulation, we generated three
102 isogenic cell types by differentiating mouse embryonic stem cells (ESCs) into epiblast stem cells
103 (EpiSCs) and neural progenitor cells (NPCs), and confirmed their identity by cell-type-specific
104 gene expression (**Extended Data Fig. 2a**). To confirm the generality of the rules, we used six
105 human cell lines: HCT116, SH-SY5Y, Jurkat, KG-1, HepG2, and K562. The latter two are also used
106 as reference cell lines by the ENCODE consortium[35]. In each cell line, we performed RNA-seq to
107 define the cell-type-specificity of the expressed genes. Similar to a previous study [36], the
108 expressed genes were categorized as cell-type-specific (CTS), cell-type-enriched (CTE),
109 housekeeping (HK), and not classified (NC) (**Fig. 1a**; see Methods). Within each category, genes
110 from all cell lines were summed and shown in a separate category (Combined). In all cell lines,
111 HK and CTE genes constituted a roughly equal portion (~40% each), and CTS genes constituted
112 7-14% (**Fig. 1a**). A slightly lower fraction of CTS genes in human cancer cell lines possibly reflects
113 a lack of specific cytokines in the culture media and downscaling of CTS genes in long-term
114 cultured cancer cells[37].
115
116 **Quantification of enhancer-regulated genes**
117 We used the CBP/p300 catalytic inhibitor A-485[38] to comprehensively map the enhancer
118 dependency of genes. A-485 specificity has been demonstrated using CBP/p300 knockout
119 cells[39]. We further validated it by comparing the nascent transcription changes caused by A-485
120 and the CBP/p300 bromodomain-based degrader dCBP1[40] (**Extended Data Fig. 2b**). A-485 and
121 dCBP1-induced gene expression changes were highly correlated, and A-485 showed no sign of
122 off-target specificity. A-485-induced catalytic inhibition occurs more rapidly (<5 min) than
123 dCBP1-induced CBP/p300 depletion (>60 min)[34,40]; therefore, we chose A-485 for this study. A-
124 485-induced nascent transcription changes were measured using metabolic RNA labeling with
125 5-ethynyl uridine and sequencing (EU-seq). In ESC, EpiSC, and NPC, transcription was measured
126 at three different time points (30, 60, and 120 min), with a minimum of three biological
127 replicates, generating 9 independent quantification data points. A-485 caused rapid and
128 selective downregulation of CTS genes (*Cobl*, *Enpp3*, *Npas3*), but not HK genes (*Mcmbp*, *Cct4*)
129 (**Extended Data Fig. 3a**). In human cells, nascent transcription changes were quantified after 60
130 min of A-485 treatment, in two biological replicates. A-485 robustly downregulated enhancer

Fig. 1

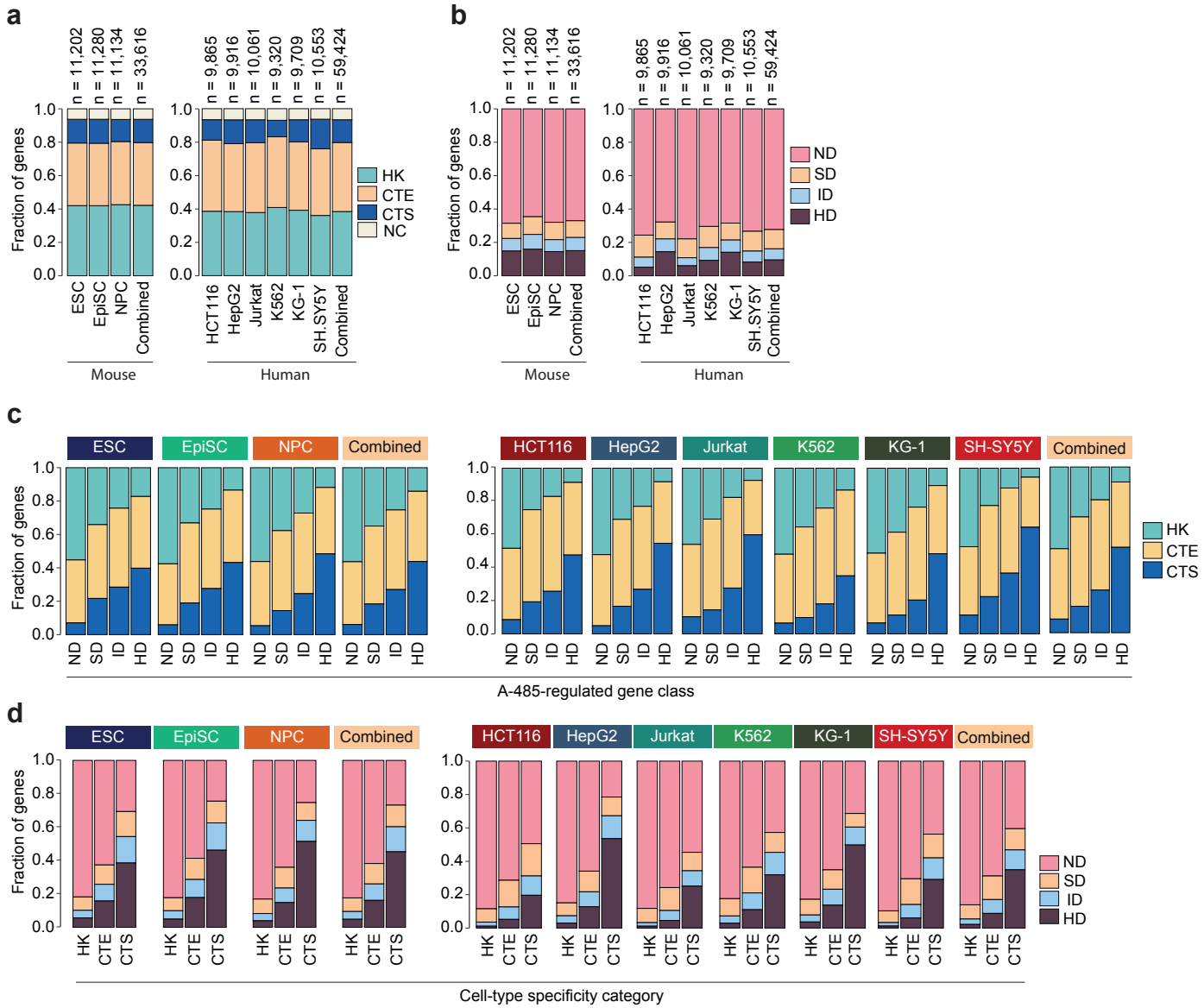


Fig. 1. Enhancers prominently regulate cell-type-specific genes. a, Cell-type specificity of genes expressed in the indicated mouse and human cell lines. HK, housekeeping; CTS, cell-type-specific; CTE, cell-type-enriched; NC, not classified into any of the three categories. In each cell line, the fraction of cell-type-specificity categories and the number (n) of expressed genes is indicated. Gene counts from all analyzed mouse cell lines and human cell lines are shown in a separate category (Combined). **b,** Fraction of A-485 downregulated genes in the indicated cell lines. HD, highly downregulated (≥2-fold decreased); ID, intermediate downregulated (>1.5-2.0-fold decreased); SD, slightly downregulated (>1.2-1.5-fold decreased); and ND, not downregulated (≤1.2-fold decreased) after A-485 treatment. Gene quantified from all analyzed mouse cell lines and human cell lines are shown in a separate category (Combined). (Biological replicates of EU-seq: n=5 for ESC, n=3 for EpiSC and NPC, n=2 for human cell lines). **c,** Fractional distribution of genes with the indicated cell-type-specificity categories, in the indicated cell lines and the depicted A-485-regulated gene categories. **d,** Fractional distribution of A-485-regulated gene class, in the indicated cell lines, and within the indicated cell-type-specificity categories.

target gene (*MYC*) without affecting non-target housekeeping gene (*CCT4*) (**Extended Data Fig. 3b**), confirming the suitability of our approach for separating enhancer-dependent and –independent genes.

Enhancers regulate genes of diverse cell-type-specificity

Based on A-485-induced nascent transcription changes, genes were classified as highly downregulated (HD; decreased by ≥2-fold), intermediate downregulated (ID; decreased by 1.5-2.0-fold), slightly downregulated (SD; decreased by 1.2-1.5-fold), and not changed (NC; decreased by ≤1.2-fold) (**Fig. 1b**). Within each category, gene numbers from all cell lines were summed and shown as a separate category (Combined).

Next, we analyzed the relationship between enhancer dependence and cell-type specificity of genes. With increasing A-485-induced downregulation, the fraction of CTS genes progressively increased, while the fraction of HK genes gradually decreased (**Fig. 1c**). In cell-type-specificity categories, a larger fraction of CTS genes was HD than HK genes, and most HK genes were ND (**Fig. 1d**). The relationship between cell-type-specificity and A-485-induced gene downregulation is highly similar in diverse mouse and human cell lines. These results show that enhancers more frequently regulate CTS genes, but not all CTS genes are enhancer regulated. Conversely, most HK genes are activated enhancer-independently, but as such HK genes are not insensitive to enhancers.

Validation of A-485-induced nascent transcription changes

CBP/p300 inhibition specifically impairs RNA Polymerase II (Pol2) recruitment and pause release at enhancer-regulated genes[34]. To independently validate EU-seq-based gene regulation, we quantified Pol2 binding in ESC, EpiSC, and NPC. A-485-induced changes in Pol2 binding were correlated (r=0.87-0.92) in replicate measurements (**Extended Data Fig. 4a**). Notably, the A-485-induced decrease in Pol2 binding was strongly correlated (r=0.80-0.86) with the decrease in nascent transcription (**Extended Data Fig. 4b**). Pol2 binding, both at TSS and gene body regions, was correspondingly reduced at EU-seq downregulated gene categories (**Extended Data Fig. 4c**). Robust, genome-scale correlation between the independent assays (nascent transcriptome analyses and Pol2 ChIP-seq), and across diverse types, strongly supports the quantitative accuracy of the measurements.

Enhancers mainly regulate proximal genes

Next, we investigated the relationship between enhancer strength, E-P distance, cell-type-specificity, and A-485-induced gene regulation. Previously, accurate estimation of enhancer strength had remained challenging. In the accompanying manuscript, we established histone H2B N-terminus multisite lysine acetylation (H2BNTac) as a genuine signature of active enhancers (Narita et al.). We used H2BK20ac as a representative H2BNTac mark to map enhancers. For comparison, we also included H3K27ac, a widely used hallmark of active enhancers. H2BK20ac and H3K27ac occupancy were analyzed by ChIP-seq in all 9 cell lines (**Extended Data Fig. 5a**). Regions commonly marked by H2BK20ac and H3K27ac were considered as candidate enhancers, and enhancer strength was separately estimated using (1) H3K27ac ChIP-seq signal, or (2) H2BK20ac ChIP-seq signal.

175
176 To infer the rules of native E-P specificity, we investigated the relationship between the
177 following four parameters: (1) the cell-type-specificity of genes, (2) the E-P distance, (3) the
178 enhancer strength, and (4) A-485-induced gene downregulation. First, in each cell type, genes
179 were grouped based on the cell-type specificity (CTS, CTE, HK). Then, within each cell-type-
180 specificity category, genes were further grouped based on the E-P distance (0-10kb, 10-50kb, or
181 >50kb). Next, within the above categories, genes were further sub-grouped based on the
182 strength of proximal enhancers. Finally, in each of these categories, the fraction of A-485-
183 regulated genes was determined.
184
185 H2BK20ac intensity-based enhancer ranking showed a strikingly clear relationship with all the
186 above parameters (**Fig. 2a**, **Extended Data Fig. 5b**). In ESC, EpiSC, and NSC, A-485-induced gene
187 downregulation directly reflected the strength of enhancers, their distance from nearest
188 promoters, and the cell-type-specificity of genes. The strongest enhancers (Top10%, Top25%)
189 had the greatest impact on the regulation of the nearest genes. Indeed, regardless of the cell
190 type specificity of genes, most genes are downregulated if they occurred in proximity to strong
191 enhancers (<10kb, Top10%). The linear distance also shows an unambiguous relationship with
192 gene regulation. Within the same enhancer rank category, the extent of A-485-induced gene
193 downregulation drops markedly if enhancers are located further away from the TSS (compare
194 0-10kb versus 10-50kb groups), indicating a clear relationship between enhancer distance and
195 gene regulation (**Fig. 2a**). Unlike H2B20ac, H3K27ac-based ranking showed no clear relationship
196 between enhancer strength, gene distance, and A-485-induced transcription downregulation in
197 ESC, EpiSC, and NPC (**Fig. 2b**). In particular, CTE, HK, and Combined gene categories displayed
198 no apparent relationship with H3K27ac-based enhancer ranking and gene regulation. These
199 results demonstrate that the relationship between enhancer position, strength, cell-type-
200 specificity and proximal gene regulation is far simpler and more predictable than suggested in
201 the current literature. However, these relationships are poorly exposed by the widely used
202 enhancer marker H3K27ac, and only become apparent when enhancer strength is defined by
203 H2BK20ac. This may explain why these simple relationships have remained hidden in plain
204 sight.
205
206 **Enhancers are broadly compatible with diverse promoters**
207 Of the all genes quantified in the mouse cell lines (n=33,616, **Fig. 1a**, Combined category), 1862
208 occurred in proximity to strong proximal enhancers (Top10%, <10kb), and 83% (HD 61%, ID
209 13%, SD 9%) of those were downregulated (**Fig. 2a**, Combined category). In contrast, 10152
210 genes lacked proximal enhancers (no enhancer within 50kb from TSS, NA category), and from
211 those, only 14% (HD 3%, ID 4%, SD 7%) were downregulated. This shows that genes with strong
212 proximal enhancers are more likely (odds ratio, 45, $p < 2.2e-16$, two-sided fisher test) to be HD
213 than those lacking proximal enhancers. In ESC, EpiSC, and NPC, 60-70% of HK genes with strong
214 proximal enhancers (Top10%, <10kb) are downregulated by A-485. In contrast, only 4-11% of
215 HK genes lacking proximal enhancers are downregulated, demonstrating that HK genes
216 occurring in proximity to strong enhancers are much more likely (odds ratio, ESC, 31, $p < 2.2e-$
217 16; EpiSC, 13, $p < 2.2e-16$; NSC, 60, $p < 2.2e-16$, two-sided fisher test) to be downregulated than
218 non-enhancer proximal genes. This difference would not be expected if HK CPs *per se* were

Fig. 2

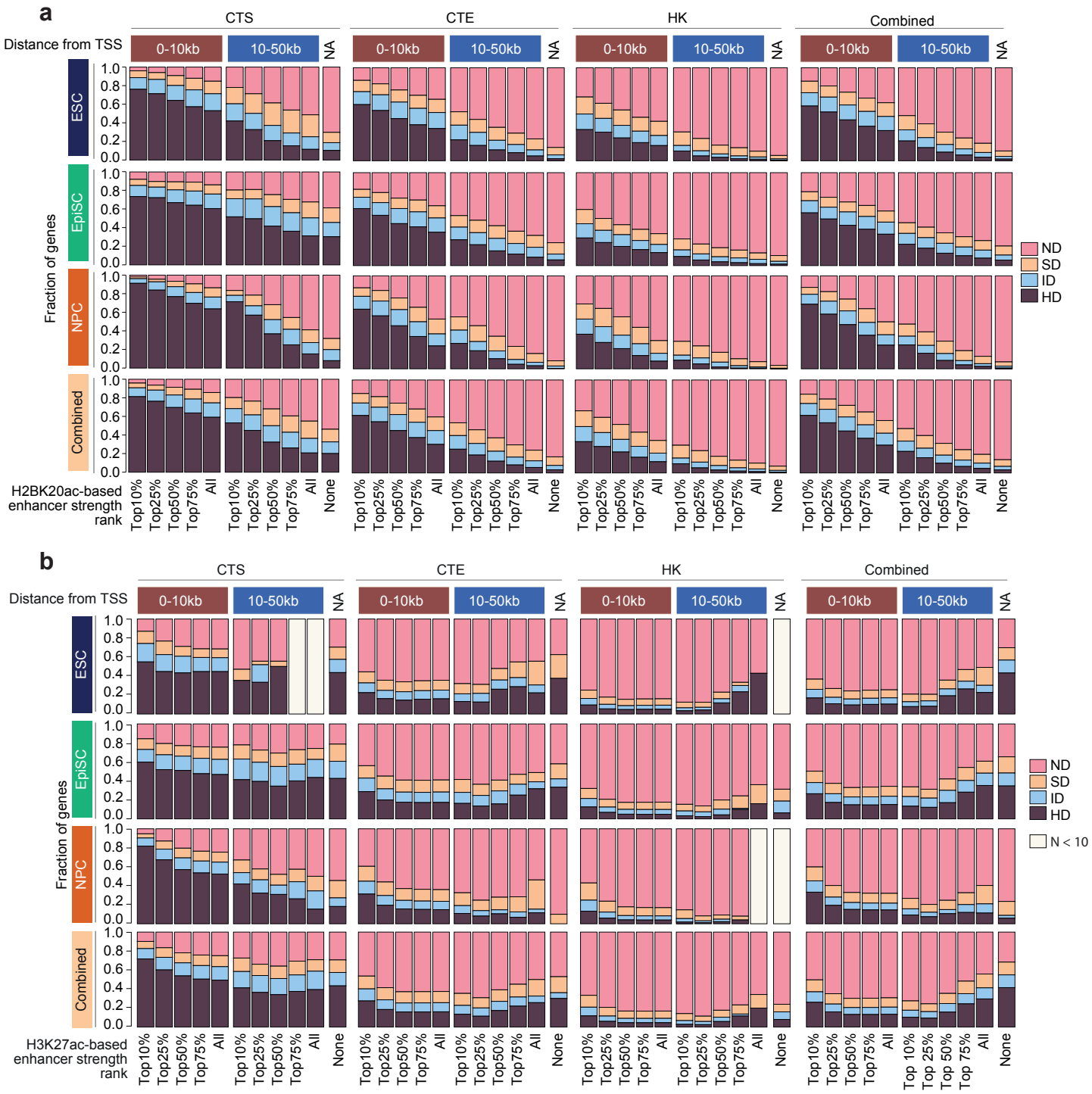


Fig. 2. Enhancer-dependent gene regulation is strongly associated with enhancer strength, distance, and cell-type specificity of genes. a, Relationship between H2BK20ac intensity-based enhancer ranking and A-485-induced gene regulation. For the indicated cell lines, genes were grouped based on cell type specificity (CTS, CTE, HK, and Combined). Enhancers (marked with both H3K27ac + H2BK20ac) were ranked based on H2BK20ac ChIP intensity. Enhancers were rank-ordered and grouped into the indicated percentile categories. Genes were further grouped based on the shortest distance between TSS and the enhancer in each percentile category. Within different groups, the fraction of A-485 regulated gene categories are shown. All, all enhancers; None, no enhancer within 50kb from TSS. b, Relationship between H3K27ac intensity-based enhancer ranking and A-485-induced gene regulation. Enhancers used for these analyses are H3K27ac peak regions and enhancer strength was defined by the H3K27ac ChIP signal. The relationship between H3K27ac defined enhancer strength and distance with cell-type-specificity and A-485-induced gene regulation was analyzed as described in panel a. The groups where the number of genes is small (N< 10) are not shown.

219 insensitive to enhancers. If we consider a gene being enhancer-sensitive, even if it is only
220 downregulated in one of the cell lines, about half of the expressed genes appear enhancer
221 regulated by at least >1.2 fold (**Extended Data Fig. 5c**). Extrapolating these observations to
222 hundreds of other cell types, it can be anticipated that a large majority of mammalian genes are
223 impacted by enhancers in at least one cell type, even though in any one cell type, enhancers
224 only regulate a small subset of genes.
225
226 To further understand enhancer compatibility, we manually inspected several genes in ESC that
227 occurred proximally to strong enhancers (Top10%, <10kb) but were not downregulated by A-
228 485. In most instances, the same enhancers occurred in proximity to other genes and regulated
229 those genes. For example, *Rpl3* and *Polr2e* are insensitive to proximal enhancers. The same
230 enhancers also localize in proximity to *Syngr1* and *Gpx4*, respectively, and regulate these genes,
231 which occur more proximal to the enhancers (**Extended Data Fig. 6a**). Regulation of one of the
232 proximal genes may explain a lack of regulation of the other proximal genes, which often
233 localize more distally and have stronger promoters. Overall, of the 463 strong enhancer-
234 proximal CTS and HK genes in ESC, ~20 could be manually confirmed as enhancer insensitive,
235 including *Eif4a2*, *Ube2s*, and *Zfp936* (**Extended Data Fig. 6b**). In these instances, no immediate
236 enhancer-proximal genes were downregulated. Promoters of these genes could be
237 incompatible with enhancers, but there can also be other reasons for their apparent
238 insensitivity, such as long mRNA half-life or binding of insulators. Consistent with this idea,
239 *MRPL23* is not regulated by the proximal enhancers because CTCF-Cohesin-mediated looping
240 connects the enhancer to distally located *H19* and *IGF2*[41,42]. *H19* and *IGF2* have short introns
241 and very long mRNA half-life[43,44]. Therefore, their nascent transcript levels are not reduced and
242 they also appear enhancer insensitive. But a zoom-in view of the *IGF2* intron regions confirms
243 that it is indeed downregulated in HepG2 (**Extended Data Fig. 6c**). *IL32* and *KCNQ* have proximal
244 enhancers in more than one cell line yet show A-485 sensitivity in only one of them (**Extended**
245 **Data Fig. 7**). This indicates that, as such, a lack of enhancer insensitivity in one cell line does not
246 necessarily establish the insensitivity of a CP to enhancers. We conclude that enhancers have
247 very broad, but not necessarily universal, compatibility with diverse promoters.
248
249 **Validation of enhancer-promoter compatibility**
250 To confirm broad E-P compatibility, we used the *Nanog* upstream (-5kb) enhancer and
251 positioned it in proximity to promoters of six different genes. The *Mcmbp*, *Sec23ip*, *Fbxo9*, *Ick*,
252 and *Mthfsd* are broadly expressed and not regulated by A-485 in any of the analyzed cell lines
253 in the native state (**Extended Data Fig. 8**). The *Foxc2* promoter is accessible in ESC, but the gene
254 is marked with the repressive mark H3K27me3 and not expressed in the analyzed cell lines. The
255 enhancer was stably inserted in the first intron of *Mcmbp*, *Sec23ip*, *Fbxo9*, and *Ick*. *Foxc2* is
256 intronless, and thus, the enhancer was positioned downstream of the gene. For *Mthfsd*, the
257 enhancer was inserted upstream of TSS. Reassuringly, in all cases, regardless of the cell-type-
258 specificity of genes and orientation of the enhancer, enhancer insertion significantly (p=
259 <0.0001) increased the expression of the proximal genes, supporting the idea of broad
260 enhancer compatibility (**Fig. 3a**).
261
262 **Strong enhancers frequently occur near their target genes**

Fig. 3

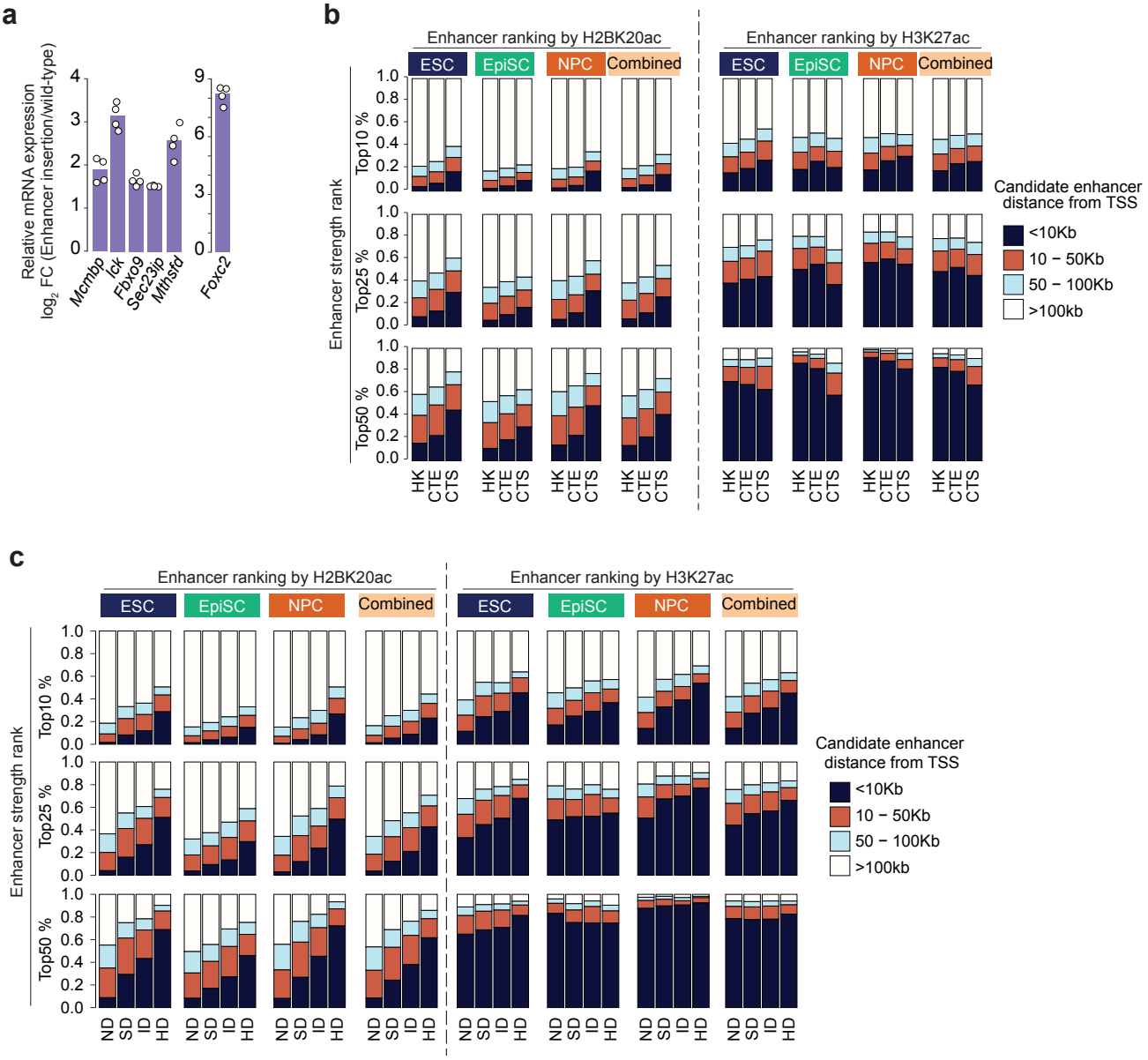


Fig. 3. Strong enhancers frequently occur proximal to CTS and A-485 downregulated genes. a, Enhancer insertion increases expression of proximal genes. For five of the genes (*Mcmbp, Fbxo9, Ick, Sec23ip, Mthfsd*), the Nanog upstream (-5kb) enhancer was inserted in the first intron of the genes. Foxc2 is intronless, and hence the enhancer was positioned downstream of the gene (see Methods). The relative mRNA expression was measured in wild-type and enhancer-inserted cells using quantitative RT-PCR. **b-c,** In the indicated mouse cell lines, the expressed genes were classified based on their cell type-specificity (b) or A-485-regulated gene class (c). Enhancer strength was determined using H2BK20ac (left panels) or H3K27ac (right panels) ChIP-signal, and enhancers were grouped into Top10%, 25%, and 50% categories. Only commonly marked H3K27ac and H2BK20ac regions were used in these analyses. Distance between candidate enhancer and TSS was determined by the distance between the commonly marked H3K27ac and H2BK20ac region and the nearest TSS. The bar charts show the relationship between cell-type-specificity (b) or A-485-dependent gene regulation (c), enhancer rank category, and enhancer-TSS distance.

263 Next, we investigated the reasons for the relatively greater impact of enhancers on CTS than HK
264 genes (**Fig. 1d**). H2BK20ac-based, but not H3K27ac-based, enhancer ranking shows that the
265 strongest proximal enhancers occur more frequently in proximity to CTS genes than to HK
266 genes (**Fig. 3b, Extended Data Fig. 9a**). For example, in ESC, EpiSC, and NPC, 17-32% of CTS
267 genes harbored Top25% enhancers within 10kb, whereas only 6-9% of HK genes had Top25%
268 enhancers within 10kb of their TSS. This shows that strong enhancers are positioned more
269 frequently (odds ratio, ESC, 3.9, p <2.2e-16; EpiSC 3.4, p <2.2e-16; NSC 6.7, p <2.2e-16, two-
270 sided fisher test) near CTS genes as compared to HK genes. Notably, this difference was
271 negligible or very small (odds ratio, ESC, 1.2, p =0.056; EpiSC 1.0, p =0.8411; NSC 0.80, p
272 =0.0063, two-sided fisher test) if enhancers occurred further away (>10kb) from TSS, where a
273 similar fraction of enhancers occurred near HK, CTS, and CTE genes. Analysis of the A-485
274 downregulated gene class confirmed this trend and showed that a greater fraction of HD genes
275 harbor strong proximal enhancers (Top10%, 25%, <10kb) than ND genes (**Fig. 3c, Extended Data**
276 **Fig. 9b**). The results show that enhancers are positioned non-randomly, occur less frequently
277 near ND genes, and those that occur in proximity to ND genes are weaker and located distally
278 from TSS. Thus, the infrequent occurrence of strong proximal enhancers is one of the reasons
279 for the modest regulation of HK genes.
280
281 **Enhancers boost transcription of ubiquitously expressed genes**
282 To confirm and further understand the reason for the quantitative differences, we manually
283 inspected tracks of A-485-regulated genes. We find that, within the same cell type, neighboring
284 genes are quantitatively differentially affected (**Fig. 4a**). As an example, A-485 almost
285 completely ablates *WNT16* and *MIR31HG* expression in HCT116, but the proximally-occurring
286 genes *FAM3C* and *MTAP* are only partially downregulated. *FAM3C* and *MTAP* expression is
287 reduced to a level that is similar to other cell lines in which the gene is not enhancer regulated.
288 We refer to this type of partial regulation (*FAM3C*, *MTAP*) as enhancer-dependent
289 transcriptional 'boosting' to discriminate it from the canonical strong enhancer-dependent
290 regulation (*WNT16*, *MIR31HG*), where virtually all transcription is dependent on enhancer
291 activity (**Fig. 4a**). Of note, H3K4me3 indicates two alternative TSS for *MIR31HG*, expression from
292 one of the promoters is fully ablated whereas the other promoter retains weak expression,
293 indicating differences in the autonomous activities of the two alternative promoters. Enhancer-
294 dependent boosting is observed in all cell lines and frequently impacts genes that are
295 ubiquitously expressed, some of them conserved down to single-celled eukaryotes, or generally
296 important for human cell proliferation, such as *RAD21*, USP7, *MPHOSPH10*, *PSMA6*, and
297 *RNF168* (**Extended Data Fig. 10**). Enhancer-dependent boosting of HK genes, which are
298 conserved in non-metazoan organisms, shows that regulatory enhancer capacity can be
299 'retrofitted' onto genes that evolved prior to the advent of enhancers and are generally
300 transcribed enhancer-independently.
301
302 **Enhancer-dependent boosting is functionally consequential**
303 To consider the biological implications of enhancer 'retrofitting', we analyzed HK genes with
304 known heightened cell-type-specific roles. *Wdr5*, *Rif1*, and *Ulk1* are evolutionarily conserved
305 from budding yeast to humans and their elevated expression is linked to their heightened
306 function in stem cells. In DepMap analyses[45], *WDR5* is essential (1054/1054) in human cell lines,

Fig. 4

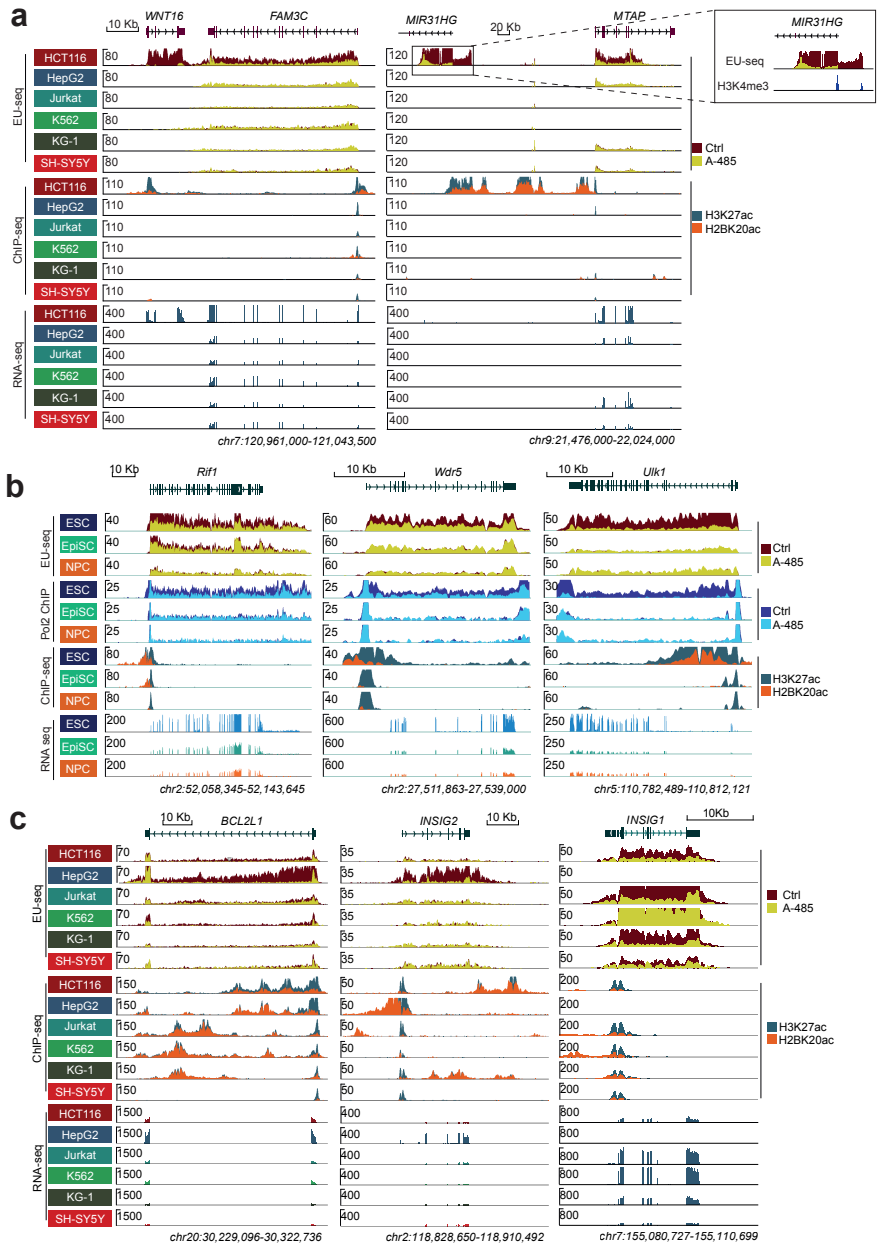


Fig. 4. Enhancers cell type specifically increase the expression of ubiquitously expressed genes. a, Gene tracks showing quantitative differences in A-485-induced gene downregulation. *WNT16* and *MIR31HG* are cell type specifically expressed in HCT116, and the A-485 treatment completely abrogates their transcription. In contrast, the proximal genes, *FAM3C* and *MTAP*, respectively, are expressed in other cell lines, but their expression is elevated in HCT116, and A-485 treatment specifically reduces FAM3C expression in HCT116. Unlike *WNT16 and MIR31HG, FAM3C* and *MTAP* expression in HCT116 is not abrogated by A-485 but quantitatively reduced. After A-485 treatment, the expression of *FAM3C* and *MTAP* in HCT116 becomes similar to that of other cell lines in which the gene is not affected by A-485. **b,** Genome browser tracks of *Rif1, Wdr5,* and *Ulk1* showing their cell-type-specific differential expression and A-485-induced regulation in ESC, EpiSC, and NPC. Overlaid tracks show nascent transcription and Pol2 binding in untreated and A-485 treated (1h) cells. *Rif1* is strongly downregulated in ESC, weakly downregulated in EpiSC, and remains unaffected in NPC. *Wdr5* and *Ulk1* are downregulated in ESC only. Changes in nascent transcription are mirrored in a cell-type-specific decrease in Pol2 binding. H3K27ac and H2BK20ac mark proximal candidate cis-regulatory elements, and RNA-seq shows cell-type-specific quantitative variation in the expression of the indicated genes. **c,** Genome browser view of *BCL2L1, INSIG1,* and *INSIG2* showing a loss of *INSIG1* expression, and a greatly elevated expression of *BCL2L1* and *INSIG2* in HepG2. *BCL2L1* is variably expressed, with HepG2 showing the strongest expression and the highest A-485-induced downregulation. *INSIG2* expression is greatly elevated in HepG2, and its expression is specifically downregulated in HepG2. H3K27ac and H2BK20ac show candidate cis-regulatory regions in each cell line.

307 and its increased expression regulates pluripotency[46]. Autophagy, and the autophagy-regulating
308 kinase ULK1, regulate pluripotency and early stages of embryonic development[47,48]. RIF1 is a
309 conserved regulator of replication timing, ESCs express a high level of RIF1, and its deletion
310 causes the most dramatic changes in the replication timing in stem cells[49,50]. *Rif1* knockdown
311 impairs mESC proliferation and pluripotency[51]. Consistent with their heightened functional
312 importance in stem cells, ESC expresses the highest levels of *Rif1*, *Wdr5*, and *Ulk1* (**Fig. 4b**).
313 *Wdr5* and *Ulk1* expression is selectively enhancer-boosted in ESC. *Rif1* expression is strongly
314 enhancer-boosted in ESC, modestly boosted in EpiSC, while the gene is expressed enhancer
315 independently in NPC (**Fig. 4b**). Notably, in all these instances, expression of the genes is not
316 ablated, but reduced, and equilibrates at a similar steady-state level in all 3 cell types after the
317 A-485 treatment. This demonstrates that these genes are broadly expressed independently of
318 enhancers and mounting of enhancers onto them increases the transcript dosage cell-type
319 specifically.
320
321 To verify the functional relevance of enhancer-dependent *Rif1* boosting in ESC, we conditionally
322 deleted the *Rif1* upstream enhancer (**Extended Data Fig. 11a**). *Rif1* expression was reduced
323 ~80% after the enhancer deletion (**Extended Data Fig. 11b**). For comparison, we also fused RIF1
324 with GFP-FKBP12^{F36V} to acutely deplete RIF1 protein using the dTAG approach [52] (**Extended**
325 **Data Fig. 11c-d**). This allowed us to compare the phenotypic effect of enhancer-dependent *Rif1*
326 boosting with RIF1 depletion. Within 3-5 days after inducing enhancer deletion or RIF1
327 depletion, ESC growth was reduced (**Extended Data Fig. 11e-f**), indicating the importance of
328 enhancer-dependent *Rif1* boosting for ESC proliferation.
329
330 To investigate the potential relevance of enhancer-dependent boosting in the disease context,
331 we analyzed genes that show preferential essentiality in HepG2, HCT116, Jurkat, K562, and SH-
332 SY5Y in DepMap[45] (gene essentiality data were not available for KG-1). Except for HCT116,
333 other cell lines included between 2-5 enhancer-regulated genes among the Top10 cell line
334 preferential essential genes, including cell-type-specific TFs (*MYB, RUNX1, HNF1A, ISL1*) and
335 signaling proteins (*ALK, IRS1, IRS2*) (**Extended Data Fig. 12a**). Interestingly, ubiquitously
336 expressed *INSIG2* and *BCL2L1* score among the Top10 preferentially essential genes in HepG2.
337 *BCL2L1* expression is strongly boosted in HepG2 by proximal enhancers (**Fig. 4c**). Most notably,
338 *INSIG2* scores as an essential gene in just 2 (out of 1054) cell lines, with HepG2 showing the
339 highest dependency. INSIG2 and INSIG1 are crucial regulators of cholesterol homeostasis[53].
340 *INSIG1* and *INSIG2* are expressed in virtually all tissues (**Extended Data Fig. 12b**). Conspicuously,
341 HepG2 lacks *INSIG1* expression and shows dramatic enhancer-dependent boosting of *INSIG2*,
342 whereas in all other cell lines *INSIG2* is expressed enhancer-independently (**Fig. 4c**). This implies
343 that enhancer-mediated *INSIG2* boosting provides a compensatory mechanism for the loss of
344 *INSIG1* in HepG2, offering a plausible explanation for the unique essentiality of *INSIG2* in
345 HepG2.
346
347 Further supporting the functional relevance of enhancer-dependent boosting, a recent study
348 showed that an oncogenic enhancer increases the expression of *SEPHS2* in acute myeloid
349 leukemia (AML) and creates unique sensitivity of AML cells to dietary selenium restriction[54].
350 Enhancer deletion only partially (<50%) reduces *SEPHS2* expression, but significantly impairs

351 AML cell proliferation[54]. Our work independently confirms enhancer-dependent boosting
352 (~50%) of *SEPHS2* in AML-derived KG-1, but not in HCT116, SH-SY5Y, or Jurkat (**Extended Data**
353 **Fig. 12c**). We extend the prior findings and show that *SEPHS2* expression is also enhancer-
354 boosted in the non-AML cell line HepG2, but by a different enhancer than the one used in AML
355 cells (**Extended Data Fig. 12c**). Previously, it remained unclear why deletion of the enhancer
356 only partially reduced *SEPHS2* expression in AML cells, and why deletion of the same enhancer
357 did not decrease *SEPHS2* expression in non-AML cells[54]. We find that *SEPHS2* can be expressed
358 enhancer-independently, and its expression can be boosted cell-type-specifically, by distinctly
359 positioned enhancers, offering plausible answers to the unresolved questions[54]
360 (**Supplementary note 2**).
361
362 **Enhancer and PRE strengths are tuned independently and cell-type-specifically**
363 Next, we sought to quantitatively assess the relative impact of enhancers and PREs in gene
364 activation. Nascent transcript abundance provides a reasonably direct measure of the promoter
365 output, which in turn reflects the sum of enhancer and PRE contribution. A-485-induced
366 decrease in nascent transcription should reflect enhancer contribution, and the remaining gene
367 expression after A-485 treatment should reflect PRE-dependent autonomous promoter
368 strength. In unperturbed cells, nascent transcript abundance of HD, ID, SD, and ND genes is
369 quite similar and varies by >3 orders of magnitude, but after A-485 treatment, expression of SD
370 and ID genes is sustained at a much higher level than HD genes (**Extended Data Fig. 13a**).
371 Interestingly, enhancers appear to similarly contribute to the expression of genes with varying
372 promoter strength, but the relative impact of enhancers appear stronger in genes that have
373 weaker promoters (**Extended Data Fig. 13b**). This shows that HD promoters are mainly
374 activated by enhancers, ND genes are activated by PREs, and SD and ID genes are partially
375 activated by PREs, and their expression is boosted by enhancers.
376
377 Enhancer strength is tuned greatly across cell types. For example, the TFs *Sox2*, *Spry2*, *Spry4*,
378 *ID1*, and *SOX4* are regulated by enhancers in all cell lines expressing the genes, yet their nascent
379 transcript abundance varies by up to ~30 fold in different cell types (**Extended Data Fig. 14a-b**).
380 Enhancer strength is also tuned for non-TF proteins, such as *LDLR*, *SLC43A3*, *HMGCS1*, and *CDK6*
381 (**Fig. 5a, Extended Data Fig. 14c**). Thus, for the same genes, enhancer strength can be tuned by
382 orders of magnitude. Similar to enhancers, PRE strength is also adjustable cell-type-specifically
383 (**Fig. 5b, Extended Data Fig. 15**). Alternative promoters of the same genes, such as *Cbx7*, *Rbpj*,
384 *ARHGEF18*, and *EIF2AK4*, are differentially activated by enhancers or PREs (**Extended Data Fig.**
385 **16**). Yet more remarkably, expression of cell-type-specific genes is adjusted by tuning either
386 enhancer or PRE strength (**Fig. 5c, Extended Data Fig. 17**). The combined function of enhancers
387 and PREs also impacts broadly expressed genes, and can even affect some of the most highly
388 expressed genes, such as *MALAT1*, although in such instances the measurable enhancer impact
389 appears relatively modest in cell lines that have strong PREs (**Extended Data Fig. 18**).
390
391 Strikingly, for the same genes, regulation can be switched cell-type-specifically. Histone
392 *MACROH2A* is expressed enhancer-independently in 5 cell lines, including Jurkat, which has
393 nearby enhancers (**Fig. 5d**). Yet, *MACROH2A* transcription is completely switched to being
394 enhancer-dependent in KG-1. Although this type of switching is infrequent, we found several

Fig. 5

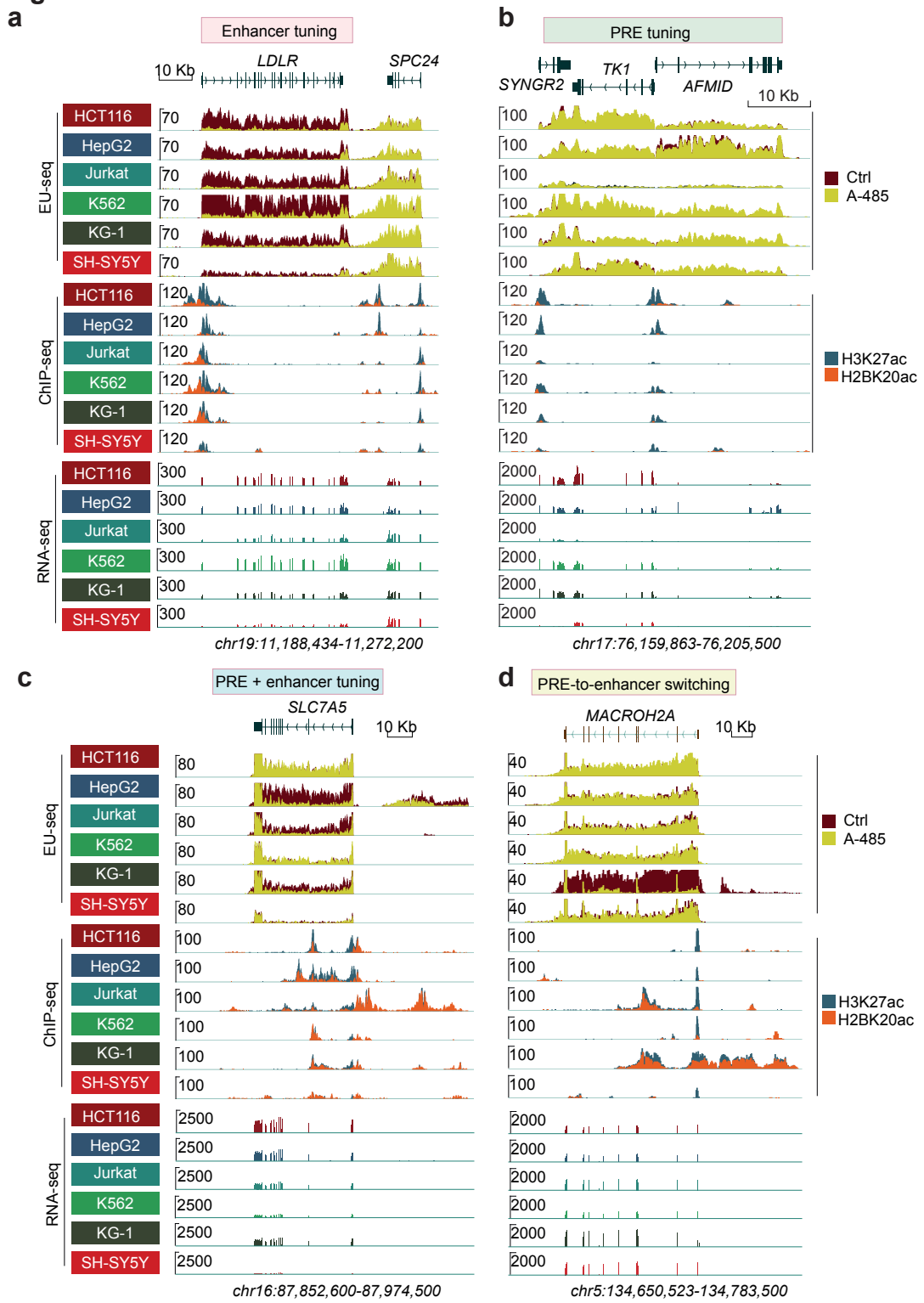


Fig. 5. Tunable and integrated strengths of PREs and enhancers control cell-type-specific gene expression. a, Gene tracks showing cell-type-specific variability in enhancer-dependent LDLR expression. LDLR is downregulated by A-485 in all cell lines but its expression is highly variable across cell lines, showing gene expression tuning by cell-type-specific tuning of enhancer strength. Proximally located SPC24 is activated by PREs and remains unaffected by A-485. **b,** Genome browser tracks showing enhancer-independent, cell-type-specific quantitative variation in gene expression, likely through changing PRE strength**. c,** Genome browser tracks showing cell-type-specific quantitative variation in *SLC7A5* expression, either through changing PRE strength (SH-SY5Y, K562, HCT116) or through the combination of changing PRE and enhancer strength (KG-1, Jurkat, HepG2). **d,** Genome browser tracks showing promoter-enhancer switching of histone *MACROH2A* regulation. The expression of *MACROH2A* in KG-1 is abrogated by A-485, but its expression remains unaffected in all the other cell lines.

395 examples (*ISOC1*, *FECH*, *MED14*, *CSK, GSTO1*) showing this type of dramatic change in
396 regulation (**Extended Data Fig. 19**). Cells showing PRE-to-enhancer switching harbor H2BK20ac
397 marked strong proximal candidate enhancers. The concordance between the acquisition of cell-
398 type-specific candidate enhancers and change in gene regulation suggests that PRE-to-enhancer
399 switching occurs as a consequence of cell-type-specific enhancer mounting. Of note, the overall
400 expression of some of the genes, such as *MED14* and *FECH*, is similar in all cell lines, yet the
401 type of regulation differs, suggesting that PRE strength is reduced in cells that have acquired
402 strong proximal enhancers. Collectively, these results show (1) that enhancer and PRE strength
403 are tuned independently of each other, (2) enhancer and PRE strengths are adjusted gene-
404 specifically and cell-type-specifically, and (3) the same genes can be activated jointly or
405 interchangeably by PREs and enhancers. From this, we conclude that, in the same cell type, a
406 single CP can accommodate input signals from enhancers and PREs.
407
408 **DNA sequences discriminate between autonomous and enhancer-dependent promoters**
409 Ultimately, gene regulatory information must be encoded in the DNA sequences. We sought to
410 identify sequence features that discriminate between enhancer-dependent and –independent
411 promoters (i.e. ND versus HD promoters). First, we compared the frequency of two main
412 mammalian core promoter motifs: TATA box and Initiator (Inr). TATA-box occurred more
413 frequently among HD than ND genes (**Extended Data Fig. 20a-b**), but, even among HD genes,
414 84-91% lacked a TATA-box. CpG island (CGI) density is higher in promoters of HK genes than CTS
415 genes[55,56]. We found that CGI frequency gradually decreases (ND>SD>ID>HD1>HD2) with A-485-
416 induced gene downregulation (**Fig. 6a; Extended Data Fig. 20c**). Normalized CpG density
417 (observed to expected, O/E, ratio) is inversely associated with enhancer-dependent gene
418 regulation. The O/E ratio was highest in ND promoters and gradually decreased with the extent
419 of A-485-induced gene downregulation (**Fig. 6b; Extended Data Fig. 20c**).
420
421 To test if promoter sequences can discriminate ND and HD promoters, we trained gapped k-
422 mer support vector machine (SVM)-based kernels (gkm-SVM) (see Methods). The classifier was
423 able to distinguish ND promoters from null sequences (i.e. randomly selected sequences of
424 matching GC and repeat content) with very high accuracy (mouse: AUROC 0.95, AUPRC 0.93;
425 human: AUROC 0.96, AUPRC 0.95) (**Fig. 6c; Extended Data Fig. 21a**). The classifier also
426 separated HD promoters from null sequences, but with a lower accuracy (mouse: AUROC 0.86,
427 AUPRC 0.83; human: AUROC 0.84, AUPRC 0.81), indicating that ND promoters are more
428 information-rich than HD promoters. Notably, ND and HD promoters could be discriminated
429 with good accuracy (mouse: AUROC 0.88, AUPRC 0.88; human: AUROC 0.85, AUPRC 0.83).
430
431 Next, we tested the performance of the sequence-based model in separating A-485 regulated
432 gene classes. We used A-485-induced gene regulation information from the isogenic mouse cell
433 lines and trained classifiers using the following three gene sets. (1) positive set: all HD genes
434 combined from the three cell lines; negative set: ND genes in all the three cell lines. (2) positive
435 set: HD genes from one of the cell lines; negative set: ND genes in all three cell lines. (3) positive
436 set: HD genes exclusively downregulated in one of the cell types; negative set: ND genes in all
437 three cell lines.
438

439 The model trained on the combined gene set performed best and separated HD and ND genes
440 in all cell lines (**Fig. 6d**). Notably, the classifier also discriminated held-out SD and ID genes,
441 which were not included in the training dataset. A model trained using ND and HD genes from
442 individual cell lines also performed well, and separated genes in the trained cell type as well as
443 in the held-out cell types (**Fig. 6e**). The model trained on genes that were exclusively HD in one
444 of the cell types showed modest performance, both in the cell type in which it was trained and
445 in the held-out cell types (**Extended Data Fig. 22**). A model trained in human cell lines also
446 performed well (**Extended Data Fig. 23a-b**). The model trained in mouse cell lines could also
447 separate A-485 regulated gene class in human cell lines, albeit with decreased performance
448 (**Extended Data Fig. 23c**). These results indicate that promoter-proximal sequences include
449 information that can reasonably predict their autonomous activation. This indicates that, to
450 some extent, sequence features that segregate ND and HD promoters are evolutionarily
451 conserved.
452
453 **Validating greater intrinsic autonomous activity of ND than HD promoters**
454 To identify sequence motifs that segregate ND and HD promoters, we derived *de novo* motifs
455 from importance scores using the trained gkm-SVM kernel. Sequences enriched in ND
456 promoters are CG-rich and match motifs of TFs, such as NRF1, GABPA, YY1, ETS1, ZNF143, and
457 THAP11, which are broadly expressed and known to bind promoters of HK genes [57-61] (**Extended**
458 **Data Fig. 21b**). To confirm this prediction, we analyzed the binding of NRF1, GABPA, ZNF143,
459 and YY1, as well as the chromatin remodeler INO80 that binds to YY1, in ESC. As a reference, we
460 also included CTCF, which is directly implicated in chromatin loop formation and enhancer-
461 promoter communication[28,62]. Except for CTCF, a larger fraction of ND promoters were
462 occupied by these factors than HD promoters, and YY1 and INO80 co-occupied a majority of ND
463 promoters (**Fig. 7a**). Notably, at the occupied promoters, there was no difference in CTCF
464 binding in ND and HD promoters, but GABPA, YY1, INO80, and NRF1 binding was significantly
465 different and progressively decreased (ND>ID>SD>HD) with increasing A-485-induced gene
466 downregulation (**Fig. 7b**). These results validate the model´s prediction and suggest that
467 sequence differences in promoter-proximal regions possibly dictate differential recruitment of
468 housekeeping TFs to ND and HD promoters.
469
470 The differences in TF binding are sportive, but not demonstrative, of the differences in ND and
471 HD promoter activity. To directly assess the autonomous activity of ND and HD promoters, we
472 used the SuRE (survey of regulatory elements) dataset, which possibly represents the most
473 comprehensive measurement of autonomous promoter activity in mammalian cells[63]. SuRE
474 covered 75% of promoters that are active in K562 and promoters of actively transcribed genes
475 had higher SuRE scores than promoters of non-expressed genes (**Fig. 7c**). The fraction of $SuRE^+$
476 promoters decreased with increasing A-485-induced gene downregulation; 83% of ND
477 promoters were $SuRE^+$ whereas only 59% of HD promoters were $SuRE^+$ (**Fig. 7d**). Notably, in
478 $SuRE^+$ promoters, SuRE scores progressively decreased with the extent of A-485-induced gene
479 downregulation. ND promoters had the highest SuRE scores, and HD promoters had the lowest
480 SuRE scores. Mirroring significant differences in SuRE scores, the level of PRE-driven
481 endogenous transcription (the remaining nascent transcription after A-485 treatment) at A-
482 485-regulated gene class is significantly different (**Fig. 7e**).

Fig. 6

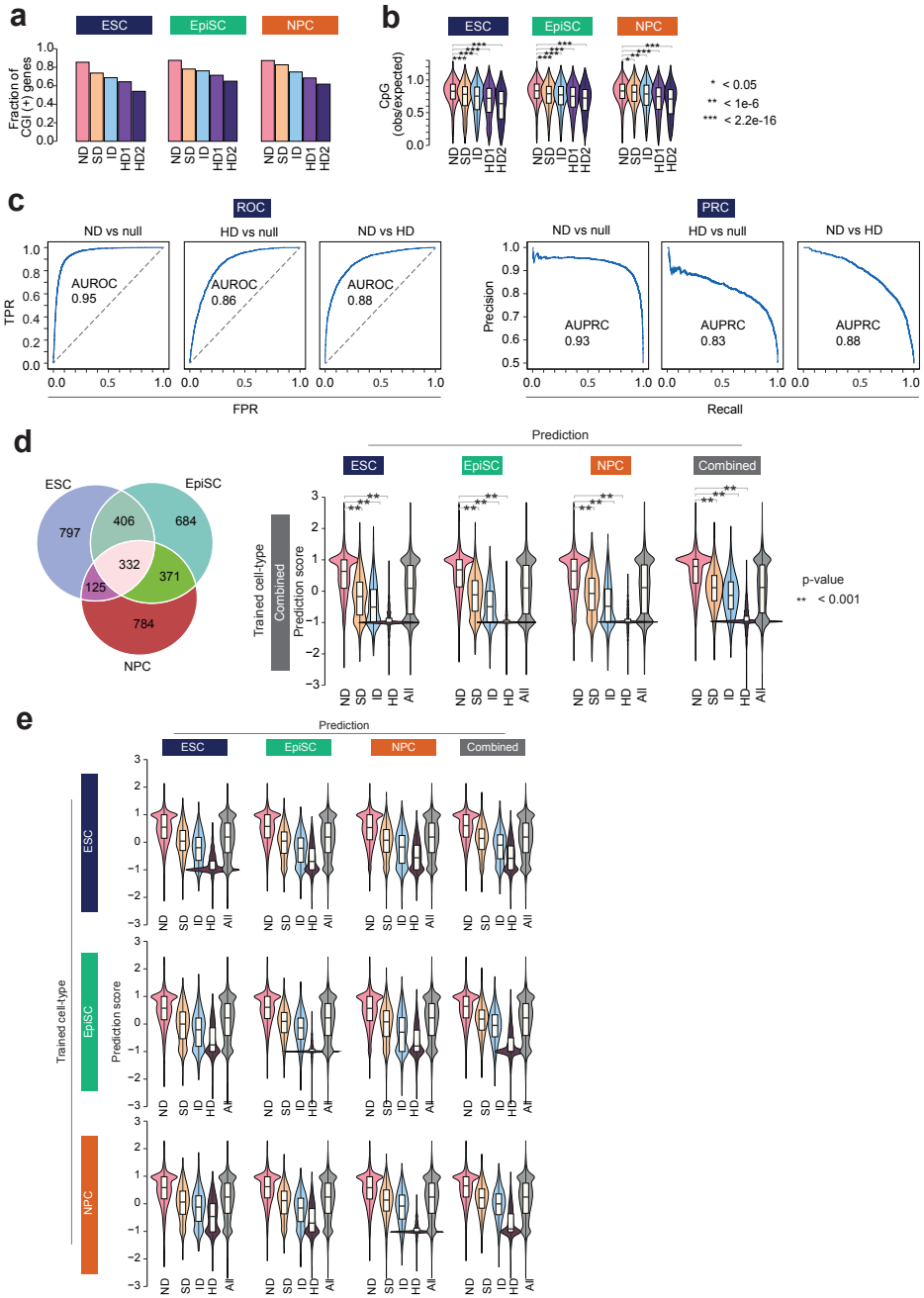


Fig. 6. Sequence differences separate HK and CTS promoters. a-b, Promoter CpG island (CGI) density inversely relates to enhancer-dependent gene regulation. The fraction of CGI positive (CGI +) genes (a) and normalized (observed to expected, O/E) CpG density (b) among the indicated A-485 regulated gene categories. A-485-regulated genes were classified as specified in Fig. 1b, and HD genes were further subdivided into two groups: HD1 (downregulated ≥2-4-fold) and HD2 (downregulated ≥4-fold). ** $p < 0.001$, Mann–Whitney U test. c, Promoter sequences can discriminate ND and HD genes. The receiver operating characteristic (ROC, left panel) and precision-recall cure (PRC, right panel) show that promoters of ND and HD genes can be separated from each other, and random genomic sequences (null sequence). The classifiers were trained using DNA sequences, 500bp upstream from TSS of ND and HD genes from ESC, EpiSC, and NPC, or randomly selected genomic sequences of similar length and GC content (see Methods for details). The area under the ROC curve (AUROC) and area under the PRC (AUPRC) are shown. **d-e,** Promoter sequence differences predict enhancer-dependence of genes. The classifier was trained using ND and HD genes as positive and negative sets respectively. The Venn diagram shows number of HD genes in the three cell lines. A gkm SVM model was trained using HD genes combined from the three cell lines and performance was tested in each of the cell lines. **e,** A gkm SVM model was trained using ND (negative set) and HD (positive set) genes from one of the indicated cell lines. Performance of the model was tested on the trained cell line as well as in the other two held-out cell lines.

Fig. 7

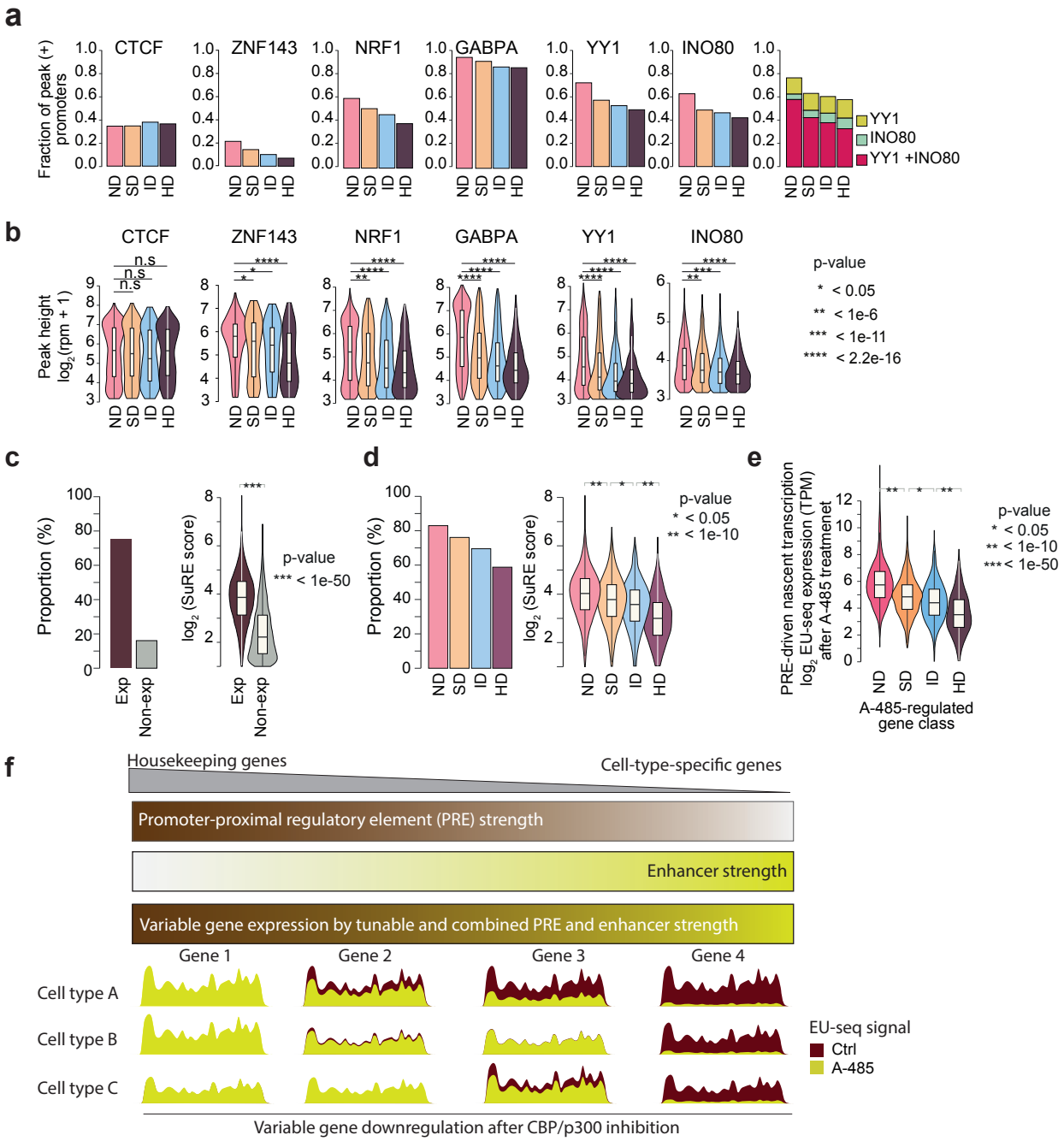


Fig. 7. Distinct TF binding and autonomous activities of enhancer-dependent and -independent promoters. a, Shown is the promoter binding of the indicated transcription factors (TFs) and INO80 at the indicated A-485 regulated gene categories in mouse ESC. Bar charts show the fraction of promoters bound by the indicated proteins. Violin plots, below each bar chart, show the ChIP intensity of the indicated proteins at promoters of the respective gene class. ChIP-seq data were re-processed from references listed in the Methods section. **b,** Portion of K562 promoters whose activity was quantified in SuRE analyses[63] (left panel) and SuRE scores for the indicated groups of promoters (right panel). Exp: promoters of genes expressed in K562; Non-exp: promoters of genes not actively transcribed in K562. **c,** Portion of the indicated A-485-regulated gene class promoters scored in the SuRE dataset (left panel), SuRE scores in the indicated groups of promoters (right panel). **d,** Shown are the nascent transcript expression levels (TPM) in the indicated A-485-regulated gene class. Nascent transcription is quantified using EU-seq, in K562 cells treated with or without A-485 (1h). This analysis only included promoters that scored in SuRE analyses. **e, f,** Proposed model for quantitative gene expression variation in mammalian cells through tunable and integrated strengths of PREs and enhancers.

483
484 The concordance between SuRE scores and A-485-induced gene downregulation is remarkable
485 considering that SuRE assesses promoter activity in an ectopic reporter, whereas in EU-seq, we
486 indirectly infer the autonomous activity of native promoters by eliminating enhancer function.
487 Together, these findings support the model that enhancer-dependent and -independent gene
488 expression is segregated through differences in PREs, which are predominantly positioned in
489 proximity to core promoters. The autonomous promoter activity progressively decreases from
490 ND>ID>SD>HD.
491
492 **Discussion**
493 This work provides a first genome-wide view of the endogenous E-P specificity in mammalian
494 cells, demonstrating that enhancers are broadly compatible and predominantly activate the
495 nearest accessible promoters. Our findings revise the idea that the nearest promoter is not the
496 true target of human enhancers[4,32,33].
497
498 Reporter-based systems are powerful in revealing E-P specificity in isolation but do not directly
499 inform how context-dependent native transcription landscapes are adjusted cell-type-
500 specifically. In our approach, enhancer specificity is inferred indirectly, but a salient advantage
501 is that it directly informs the functional impact of enhancers on endogenous genes. Thus, our
502 approach complements the existing methods and offers much-needed in vivo validation for
503 reporter-based observations. Importantly, our work goes beyond revealing native E-P
504 compatibility; we use the compatibility rules to illuminate quantitative gene expression
505 variation within and across cell types. The work showcases the usefulness of improved
506 enhancer markers and comparative analyses for uncovering and explaining the complexities of
507 native gene expression. For example, had we not used H2BK20ac as an enhancer marker, the
508 striking relationships between enhancer distance, strength, and gene regulation could not have
509 been uncovered. Similarly, without studying multiple cell types, it would have been impossible
510 to uncover the regulatory intricacies and explain how E-P compatibility shapes gene expression
511 variability across cell types.
512
513 Below, we discuss the major findings of our work and their conceptual implications in
514 understanding gene regulation and the evolution of new regulatory networks.
515
516 **Enhancers prominently act locally**
517 High-resolution Micro-C-based chromatin contact maps suggest a median enhancer-promoter
518 distance of ~100kb [64,65], and lower resolution enhancer-promoter interaction analyses imply
519 that as many as half of enhancers skip their proximal genes to activate distally located genes[4-10].
520 If enhancers were to frequently skip their nearest active promoter, we would expect many
521 enhancer-proximal genes to remain unaffected by CBP/p300 inhibition, and many non-
522 enhancer proximal genes to be downregulated. In all cell lines investigated, most of the genes
523 occurring in proximity to strong enhancers are downregulated after CBP/p300 inhibition,
524 whereas genes that are located distally are seldom affected. If a large fraction of enhancers
525 were incompatible with nearest promoters and frequently skipped their nearest active

526 promoters, we would not expect such a clear relationship between enhancer-promoter
527 distance and gene regulation.
528
529 Our finding, derived from tens of thousands of endogenous enhancers and promoters, suggests
530 that enhancers act prominently on proximal open promoters. The local enhancer action may
531 help explain why global ablation of TADs and large architectural loops have a limited impact on
532 gene regulation [66,67]. We find that the activation potential of enhancers decreases sharply with
533 the distance from the promoter, and weakly impacts genes beyond >50kb distance, providing
534 genome-scale validation for reporter-based analyses in a single TAD[68]. We do note robust
535 evidence showing that enhancers can and do act at long genomic distances [3]. We acknowledge
536 an important role of chromatin looping, but this mechanism may be more broadly used for
537 reducing large enhancer-promoter distances rather than for skipping nearest genes and
538 activating distal genes. Gene skipping may occur, but much less frequently than implied
539 currently.
540
541 **Enhancers and PREs jointly shape transcription landscapes**
542 Our work reveals that mammalian gene regulation is remarkable plastic and that gene
543 regulation is not strictly segregated into 'enhancer-dependent' and 'enhancer-independent'
544 types. We propose that PRE and enhancer function in gene regulation are not mutually
545 exclusive; instead, cell-type-specific transcription landscapes are quantitatively shaped through
546 tunable and integrated strengths of PREs and enhancers, providing crucial in vivo validation of
547 reporter-based observations[12,13]. Notably, enhancer and PRE strength are tuned independently
548 of each other, across diverse cell types, providing a highly modular mechanism for dynamic
549 gene regulation. We find hundreds of genes (SD and ID gene class) that are not strictly
550 enhancer-dependent, but whose expression is boosted by enhancers, showing that the
551 combined gene activation by enhancers and PREs is not a rarity, but a widespread feature of
552 mammalian genes. This finding demonstrates that, in the same cell type, in its native context, a
553 single CP can accommodate input signals from PREs and enhancers. We suggest that enhancers
554 and PREs can act as independent regulatory units and broad compatibility is likely a general
555 property of mammalian CPs.
556
557 **Broad enhancer compatibility expands the regulatory complexity**
558 Qualitatively, many human genes are ubiquitously expressed, but their expression is
559 quantitatively altered across tissues[69]. If enhancer compatibility was constrained, as implied in
560 the enhancer specificity model, it would render a large portion of human genes beyond the
561 scope of enhancers. Broad compatibility expands enhancer function beyond CTS genes and
562 affords a widespread role to enhancers in quantitative boosting of HK genes across tissues in
563 development and disease. In the context of evolution, we suggest that broad E-P compatibility
564 provided an elegant mechanism for 'retrofitting' new regulatory complexity onto HK genes,
565 whose origin pre-dates the advent of enhancers. At the same time, keeping the CP architecture
566 preserved allowed the cell to use the same conserved basal machinery for transcribing all
567 genes. We suggest that broad enhancer compatibility is not an accidental invention but a key
568 design principle with important functional relevance.
569

570 It is important to emphasize that broad E-P compatibility implies that enhancers are capable of
571 activating most promoters, but it does not mean that all promoters are quantitatively activated
572 to the same extent. We find that CTS genes are most prominently regulated by enhancers, CTE
573 genes are affected to an intermediate level, and HK genes are least affected. At least two
574 different factors contribute to this. (1) Strong enhancers are more often positioned near CTS
575 genes than CTE and HK genes. HK genes have weaker enhancers that are often located more
576 distally. (2) Native genes are activated by the integrated activity of enhancer and PRE. The
577 relative impact of an enhancer on transcription depends on the autonomous strength of the
578 promoter, and the relative impact of the enhancer is related to promoter strength. Because the
579 relative strength of PREs can change in different cell types, relative enhancer responsiveness
580 can be variable across cell types.
581
582 **PREs can sustain partial gene activation without enhancers**
583 The combined function of PREs and enhancers has implications for understanding transcript
584 dosage regulation. It has been noted that deletion of well-defined enhancers often causes only
585 modest or no downregulation of target genes[70]. Currently, this phenomenon is often
586 rationalized by transcriptional buffering by redundantly functioning 'shadow enhancers'[71,72].
587 Our findings show that hundreds of enhancer target genes are only partially downregulated
588 even when enhancer function is ablated globally. This suggests that along with shadow
589 enhancers, PREs have a previously unappreciated, widespread role in sustaining partial gene
590 activation in the absence of enhancer activity. Thus, just like shadow enhancers, dual regulation
591 by enhancers and PREs may provide robustness in gene regulation by buffering against
592 deleterious mutations in one of the regulatory elements.
593
594 **Broad compatibility lowers the threshold for evolving new regulatory networks**
595 Non-coding genetic variants can create de-novo enhancers that can act as new regulatory
596 modules[73-75]. Indeed, the evolution of organismal complexity primarily involves mutations in
597 the cis-regulatory sequences that quantitatively alter the expression of functionally conserved
598 proteins[76]. Restricted enhancer compatibility imposes greater constraints on the emergence of
599 new gene regulatory networks because it requires a simultaneous change in incompatible CP
600 sequences. Broad E-P compatibility lowers the threshold for implementing new regulatory
601 capacities and facilitates the emergence and re-wiring of regulatory networks in disease and
602 organismal evolution.
603
604 **Sequence differences in PREs segregate autonomous and non-autonomous promoters**
605 We propose that enhancer-dependent and -independent genes are mainly segregated by the
606 sequence-encoded differences in their PREs. Enhancer-independent genes, which prominently
607 include HK genes, harbor strong PREs that potently recruit ubiquitously expressed TFs and act
608 as 'built-in' activation modules. Enhancer-dependent genes, which mostly include CTS genes,
609 contain weaker PREs. Consequently, CTS promoters have weaker autonomous activation
610 potential. This brings two advantages (1) weaker autonomous promoter activity avoids
611 ubiquitous expression of CTS genes, and (2) weaker autonomous promoter activity generates a
612 dependence on enhancers for their activation. CGI density appears to be one of the features
613 that afford greater autonomy to PREs. This can rationalize higher CGI density in HK promoters

614 than in CTS promoters[57,61,77,78], stronger binding of ubiquitously expressed TFs to CGI-rich
615 promoters[58-60,63], the stronger activity of CGI promoters in reporter assays[63], and increased
616 promoter activity with increasing CG dinucleotide content[79]. Of note, for simplification, we
617 refer to promoters as autonomous and non-autonomous, but this binary classification is
618 inaccurate. Promoters have a graded extent of autonomous activation potential. Importantly,
619 the autonomous activity of promoters is not fixed; for the same promoters, activation strength
620 varies across cell lines. Thus, the autonomous activity of promoters neither means that
621 promoter strength is untunable nor does it imply insensitivity of such promoters to distal
622 enhancers.
623
624 Together, the presented findings provide major conceptual insights into endogenous enhancer
625 specificity and cell-type-specific gene regulation and will facilitate efforts toward accurate
626 modeling of enhancer-dependent transcription and functional interpretation of genetic
627 variants. A remaining goal for future studies is to clarify whether there are other classes of
628 distal enhancers that act independently of CBP/p300 and if their regulatory principles are any
629 different from CBP/p300-dependent enhancers investigated in this work.
630
631 **Supplementary note 1**
632 Recently it was shown that different groups of human enhancers require different COFs [21]. To
633 evaluate the requirement of CBP/p300 enhancer activity in STARR-seq, we re-analyzed the
634 published data. Specifically, we analyzed the requirement of CBP/p300 for enhancers that are
635 accessible in the native context. The reason we focused on chromatin accessibility is that
636 chromatin accessibility is a hallmark of endogenously active enhancers, and it is one of the most
637 fundamental differences between native enhancers and enhancers cloned in plasmids.
638 Surprisingly, among all COFs analyzed, only CBP/p300 showed a clear relationship with
639 endogenous chromatin accessibility and enhancer reporter activity (**Extended Data Fig. 1**).
640 Enhancers showing the greatest CBP/p300 dependence had the highest chromatin accessibility,
641 enhancers with intermediate CBP/p300 dependence had an intermediate level of accessibility,
642 and enhancers that were unaffected by CBP/p300 depletion lacked chromatin accessibility. This
643 shows that, even in the ectopic plasmid-based reporters, CBP/p300 is required for the activity
644 of enhancers that are open in the native chromatin context, supporting our assumption that
645 CBP/p300 is required for the activation of natively accessible enhancers. The lack of association
646 between STARR-seq activity and chromatin accessibility is consistent with the findings that
647 STARR-seq can identify open and closed enhancers[13,80]. Pioneering factors can bind to closed
648 chromatin, but we are not aware of concrete evidence showing a major contribution of closed
649 enhancers to native gene regulation, without acquiring accessibility under the conditions in
650 which they are active.
651
652 **Supplementary note 2**
653 Deletion of the oncogenic *SEPHS2* enhancer did not impair expression in normal hematopoietic
654 cells or cells that originated from other tissues[54]. Our work independently confirms this finding
655 and shows that the enhancer boosts (~50%) *SEPHS2* expression in the AML-derived KG-1, but
656 not in HCT116, SH-SY5Y, or Jurkat (**Extended Data Fig. 12c**). Extending the prior findings, we
657 find that *SEPHS2* expression is also enhancer-boosted in the non-AML cell line HepG2, but the

658 position of enhancers is completely different in KG-1 and HepG2. Our results offer a plausible
659 explanation for some of the questions that remained unresolved in previous work. For example,
660 why is *SEPHS2* expression only partially reduced in enhancer deleted cells, and is the remaining
661 *SEPHS2* expression sustained by shadow enhancers or other mechanisms? In our analyses,
662 *SEPHS2* expression in KG-1 is only reduced by ~50%, even when CBP/p300 activity is ablated
663 globally, suggesting that remaining *SEPHS2* expression is likely sustained by PREs, not by yet
664 undefined distal shadow enhancers. Why did deletion of the enhancer not impact *SEPHS2* in
665 non-AML cells? Our findings suggest two different reasons for this. (1) Enhancers positions in
666 KG-1 and HepG2 are completely different, showing that *SEPHS2* expression can be regulated by
667 different enhancers in different cell types, which can explain why deletion of an enhancer
668 sequence may affect expression in one cell type but not in another, even if the gene is
669 regulated by enhancers in both the cell types. (2) *SEPHS2* is expressed enhancer-independently
670 in HCT116, SH-SY5Y, and Jurkat. This demonstrates that the same gene can be activated by an
671 enhancer in one cell type but not in others, and thus, the effect of enhancer deletion will be
672 cell-type-specific. This exemplifies how studying gene regulation in multiple cell types can
673 provide an improved mechanistic understanding of quantitative gene regulation. Had we used
674 only one cell type, or simply analyzed enhancer activity with MPRA, it would be almost
675 impossible to obtain these insights. Together, the above results support the idea that enhancer-
676 dependent boosting of HK genes is an important feature of mammalian gene regulation, and it
677 provides a mechanism for quantitatively adjusting transcript dosage in physiology and disease.
678

679 **Acknowledgments**

680 We thank the members of the Choudhary lab for their helpful discussions. The Novo Nordisk
681 Foundation Center for Protein Research is financially supported by the Novo Nordisk
682 Foundation (NNF14CC0001). C.C. was supported by the Hallas Møller Investigator Award
683 (NNF14OC0008541) from Novo Nordisk Foundation. This work was funded by the European
684 Commission FP7 grant (SyBoSS FP7-242129). Y.H. was supported by a Grant-in-Aid for JSPS
685 Overseas Postdoctoral Fellows. S.K. was supported by a Lundbeck Foundation Fellowship (R347-
686 2020-2170). We thank the SyBoSS partners Konstantinos Anastassiadis, A. Francis Stewart,
687 Austin Smith, and Bill Skarnes for sharing E14TG2a Oct4-IRES-Puro mouse ESCs. We thank Jonas
688 Walter for generating Cre-ERT2 expressing mESCs. We thank Dr. Christopher J. Ott for providing
689 dCBP1. We thank the CPR Imaging Platform, the CPR Big Data Management Platform, and the
690 CPR and DanStem Genomics Platform for their assistance. We thank the ENCODE Consortium
691 and the ENCODE production laboratory for generating and sharing the data.
692

693 **Author contributions**

694 T.N. and C.C. conceived the project. K.N. differentiated ESCs and generated the isogenic EpiSC
695 and NPC cell lines. T.N., Y.H., S.K., E.M. performed experiments and analyzed data. T.N.
696 performed bioinformatic analyses. T.N. and C.C. wrote the manuscript with input from all co-
697 authors.
698

699 **Conflict of interest**

700 The authors declare no conflict of interest.
701

Material and Methods

Cell lines and culture conditions

E14TG2a Oct4-IRES-Puro mouse ESC (mESC)[81] were used for all genomics experiments. Where indicated, E14TG2a mESC (Sigma-Aldrich, Cat# 08021401) were used for manipulation (insertion or deletion) of enhancers. mESC were cultured in a custom-made (C.C.Pro GmbH, Oberdorla) N2B27 medium consisting of a 1:1 mix of DMEM/F12 and Neurobasal medium, but lacking Arginine and Lysine. Before use, the media was supplemented with 640 µM L-Arginine, 550 µM L-Lysine, 100 µM 2-mercaptoethanol, 150 µM sodium pyruvate, 0.5x B27 supplement (Thermo Fisher Scientific), 0.5x N2 supplement (Thermo Fisher Scientific), 1% penicillin/streptomycin, and cell type-specific growth factors or inhibitors. For mESCs the medium was supplemented with 1000 units/ml leukemia inhibitory factor (LIF) (Merck Millipore), 1 µM PD0325901, and 3 µM CT-99021 (custom-made by ABCR GmbH, Karlsruhe).

EpiSCs were derived from E14Tg2A Oct4-IRES-Puro ESCs by culturing them on dishes coated with 10ng/ml fibronectin (Merck Millipore), in N2B27 supplemented with 20ng/ml Activin A and 12ng/ml Fibroblast growth factor 2 (FGF2) (MPI-CBG protein facility, or Peprotech) for 14-15 passages. NPCs were generated from EpiSCs by seeding 0.5-1x10^5 cells per cm^2 to fibronectin-coated dishes in N2B27 without growth factors[82,83]. After 5 days, the developed neural rosettes were dissociated with Accutase (PAA) and seeded to 0.1% gelatinized tissue culture plates in N2B27 supplemented with 10ng/ml FGF2 and 10 ng/ml Epithelial growth factor (EGF) (PeproTech). All cell lines were passaged regularly (every 2 to 4 days) with Accutase (PAA) or TrypLE (Thermo Fisher Scientific), and ESCs and EpiSCs were occasionally selected for Oct4 expression with 1µg/ml puromycin (Sigma). HepG2 and HCT116 were cultured in DMEM media supplemented with 10% FBS and 1% penicillin/streptomycin. Jurkat, KG-1, K562, and SH-SY5Y cells were cultured in RPMI-1640 media supplemented with 10% FBS, and 1% penicillin/streptomycin. All cells were cultured at 5% CO_2, 37 °C. Where indicated, the cells were treated with the following inhibitors: A-485 (10µM), or dCBP-1 (100nM).

RNA-seq

Following cell lysis with RLT lysis buffer (Qiagen), total RNA was extracted using RNeasy Mini Kit (Qiagen). 1ug of total RNA was used for enriching polyadenylated mRNA, the samples were processed using TruSeq Stranded mRNA Library Prep Kit (Illumina) as per manufacturer's instructions and sequenced using a Next-Seq 500 Sequencer (Illumina).

Nascent transcription analyses by EU-seq

Nascent transcription analyses were performed as described previously[34]. ESC EU-seq data was previously published (GSE146328) and re-processed for the downstream analysis. EpiSC, and NPC were treated with DMSO or A-485 (10µM) for 30, 60, and 120 min. dCBP1-induced nascent transcription changes were quantified after treating ESC with DMSO or dCBP1 (100nM) for 120 min. In all human cell lines, nascent transcription changes were quantified in cells treated with DMSO or A-485 (10µM) for 60 min. Cells were pulse-labeled with 0.5 mM 5-EU for 20 min before sample collection. Following 5-EU labeling, cells were quickly washed (1x) with PBS, and immediately lysed in RLT buffer, and RNA was prepared using RNeasy Mini Kit (Qiagen). The

746 labeled RNA was subsequently enriched using the Click-iT® Nascent RNA Capture Kit (Thermo
747 Fisher Scientific) according to the manufacturer's instructions. RNA-seq libraries were prepared
748 using the NEBNext Ultra kit (NEB) as per the manufacturer's instructions and subsequently
749 sequenced using a Next-Seq 500 Sequencer (Illumina).
750
751 **Reference genome annotation**
752 Mouse and Human genome annotation and reference genome were downloaded from the
753 GENCODE (mouse; GRCm38 release 25, human; Grch37 release 29)[37]. As a reference, the
754 transcript type of "protein coding" and "lincRNA" was chosen as representative genes. For
755 building cell-type-specific gene models, the expressed transcripts were selected using the
756 criteria of Transcripts per million transcripts (TPM) > 2 in RNA-seq and the presence of
757 H3K4me3 ChIP seq peak within 1Kb from TSS. The longest isoform was selected if a single
758 H3K4me3 peak is proximal to multiple transcripts.
759
760 **RNA-seq**
761 Adaptors and low-quality sequences (-q 20) were trimmed using Cutadapt
762 (doi:https://doi.org/10.14806/ej.17.1.200). Reads were mapped to the reference genome using
763 STAR (version 2.6.1a)[84] with removing the non-canonical junctions. Reads mapped to tRNA and
764 rRNA regions were removed using Bedtools (version 2.23)[85]. The number of reads mapped to
765 exon was counted on a gene-basis using HTseq (version 0.11.1)[86] TPM was calculated using R.
766
767 **EU-seq**
768 Adaptor and low-quality sequence trimming was performed as described in the RNA-seq
769 section. Read sequences were aligned to mm10 (mouse) using bwa aln with default parameters
770 (BWA version 0.7.10[87]. Multi-mapped reads and reads with more than three mismatches were
771 removed using samtools (version 1.4)[88]. Reads mapped to rRNA and tRNA region obtained from
772 UCSC genome browser were removed using Bedtools. For 30, 60, and 120 min timepoint data,
773 based on the Pol2 elongation rates, maximum 30, 90, and 240 Kb gene body regions from TSS
774 were used for differential gene expression (DGE) analysis. The number of reads mapped to
775 defined regions was counted using HTseq. Log2 fold-change (FC) and p values were calculated
776 in individual time points using DEseq2[89] with the default scaling method (median of relative
777 abundance). The fold changes from distinct timepoints were summarized by calculating the
778 averages. Enhancer dependencies were determined by enhancer regulation based on the EU-
779 seq DGE changes after the A-485 treatment. Expressed genes in each cell line were determined
780 as described in the reference genome annotation section and further filtered the low expressed
781 genes if the mean EU-seq read count of control and A-485 treatment condition is less than 20.
782 ND as following criteria: < 1.2-fold changes of down-regulation after A-485 treatment. SD: < 1.5-
783 and >= 1.2-fold changes of down-regulation. ID: < 2- and >= 1.5-fold changes of down-
784 regulation. HD: >= 2-fold changes of down-regulation. For gene track visualization, we used
785 deeptools and bigWigMerge (UCSC tool)[90] to generate bigwig files using scaling factor
786 calculated from DESeq2.
787
788 **ChIP-Seq**

789 Reads were mapped to the reference genome by using bwa aln with default parameters (BWA
790 version 0.7.10). Multi-mapped reads, duplicated reads, or reads with more than three
791 mismatches were removed by samtools[88]. Reads mapped to the DAC Blacklisted Regions
792 (https://www.encodeproject.org/annotations/ENCSR636HFF/) were omitted from the
793 downstream analysis. Peak regions were called using LanceOtron with the default model (wide-
794 and-deep_jan-2021)[91]. The peaks proximal within 2kb are merged using Bedtools[92], and poorly
795 enriched peaks of maximum peak height < 8 reads mapped per million (rpm) were omitted.
796 Peak height was calculated using bamCompare[93] with the following parameters (centerReads,
797 minMappingQuality 10, 20bp bin, smooth length 400bp, extend reads 200bp, rpm
798 normalization, and input rpm value is subtracted). Where indicated, peak regions were
799 classified into three classes "Promoter" (TSS ±1kb), "Gene body" (Exon, intron, 5' UTR, and
800 3'UTR, excluding "Promoter" regions), and "Distal" by using ChIPseeker[94]. IGV[95] was used for
801 the gene track visualization. For Pol2 ChIP-seq, we applied scaling factors based on the EU-seq
802 results. We had assumed that Pol2 binding remains unaffected at promoters of genes that are
803 not downregulated (ND) by A-485 in EU-seq analyses, and the scaling factors are calculated by
804 DE seq default normalization methods using the Pol2 ChIP read counts at ND gene promoters
805 (±2kb from TSS). For the quantification of Pol2 binding changes in the gene body regions, a
806 maximum of 90kb gene body regions from TSS were used as described in the EU-seq section.
807
808 **Defining enhancer strength**
809 Enhancer regions were defined as overlapping regions of H2BK20ac and H3K27ac peaks. For
810 H2BK20ac-based analyses, enhancer strength was calculated by subtracting input rpm from the
811 H2BK20ac rpm value. For H3K27ac-based analyses, enhancer strength was calculated by
812 subtracting input rpm from the H3K27ac rpm value.
813
814 **Prediction model for ND and HD promoters**
815 Support vector machines with gapped k-mer kernels (gkm-SVMs)[96] were used to learn
816 predictive models of ND and HD promoter sequences. Genes categorized as ND in all the cell
817 lines (mouse: ESC, EpiSC and NPC, human: HCT116, HepG2, Jurkat, K562, KG-1, and SH-SY5Y)
818 are defined as combined ND genes (mouse: N= 5603, human: N= 3258), and genes categorized
819 as HD genes in any cell lines are defined as combined HD genes (mouse: N=3499, human:
820 N=3154). Combined SD and ID genes are determined by the highest down-regulation of EU-seq
821 DGE after A-485 treatment in all the cell lines; combined ID: < 2- and >= 1.5-fold changes of
822 down-regulation in any cell lines and not classified as combined HD genes, combined SD: < 1.5-
823 and >= 1.2-fold changes of down-regulation regulation in any cell lines and not classified as
824 combined HD nor combined ID genes. DNA sequences of 100, 250, 500, and 1Kb upstream or
825 downstream from TSS were used for generating prediction models. Null sequences were
826 obtained by randomly selecting repeat- and GC- matching sequences from the corresponding
827 genome using the genNullSeqs function of R package gkmSVM[96]. The approximately equal
828 number of positive and negative sequence sets were used as input (<10% batch size difference).
829 We tested gkm kernels and gkm radial basis function (gkm RBF) kernel using the gkmtrain
830 function (lsgkm 1.10)[97]. The model performance was evaluated by the area under the precision-
831 recall curve (AUPRC; calculated using ROCR R package)[98] by 10-fold cross-validation. Gkm kernel
832 trained by sequences of upstream 500bp from TSS with word length 10, maximum mismatch 3

833 and informative column 7 was chosen as the best model based on the AUPRC. The trained
834 model was used for scoring all the expressed gene promoters by gkmpredict function (lsgkm)[97].
835
836 Nucleotide contribution scores were calculated for all the nucleotides in all the positive and
837 dinucleotide shuffled negative sequence sets (gkm explain function from lsgkm). We used TF–
838 Modisco[99] to obtain de novo motifs from the contribution scores of positive and negative
839 sequence sets. We used TOMTOM (version 5.4.1)[100] for comparing the computed motif position
840 weight matrix (PWM) with the motif references. The motif references included
841 JASPAR2022_CORE_vertebrates_non-redundant, uniprobe_mouse, jolma2013,
842 HOCOMOCOv11_core_HUMAN_mono (for human motif),
843 HOCOMOCOv11_core_MOUSE_mono (for mouse motif), and JASPAR2018_POLII.
844
845 **Annotation of CpG island, TATA box, and Inr promoter elements**
846 250bp upstream and downstream sequences from TSS were used to calculate GC content and
847 CpG dinucleotide frequencies. CpG island was defined as GC content > 0.5 and observed
848 CpG/expected CpG > 0.6 across the 500 bp region. TATA box and Inr promoter elements are
849 referenced from Eukaryotic Promoter Database (EPD)[101].
850
851 **Overlap of promoter regions with SuRE peaks**
852 K562 SuRE peaks were downloaded from the supplemental section of van Arensbergen et al. [63].
853 $SuRE^+$ active and inactive promoter, all the promoter regions in the reference genome
854 annotation were classified into $SuRE^+$ and $SuRE^-$ based on the criteria where it is present within
855 the range of -250 - +50bp from TSS. The proportions of $SuRE^+$ active and inactive promoters are
856 calculated by gene basis. SuRE peak overlap with A-485 downregulated gene class promoter
857 was calculated by promoter-basis.
858
859 **Evaluating chromatin accessibility of COF-AID STARR-seq regions**
860 COF-AID STARR-seq regions were downloaded from the supplemental information of Neumayr
861 et.al. [21]. HCT116 ATAC-seq data were downloaded from the GEO portal (GSE97889).
862 Accessibility was calculated by maximum peak height and normalized reads density within
863 STARR-seq regions as described in the ChIP-seq section.
864
865 **Gene classification by cell type-specificity**
866 We used the FANTOM5 CAGE dataset as a reference to classify cell-type specificity of genes
867 [69,102]. In the mouse reference, we excluded data from "whole," "lactating," and "pregnant"
868 stages. The resultant 151 mouse tissue expression profile (44 tissues from 11 developing
869 stages) and 76 human tissue expression profiles (68 tissues from adult and fetal stages) were
870 converted into a binary expression matrix using a gene expression threshold of TPM ≥ 2. Genes
871 were classified in mouse and human cell lines using the following criteria. [mouse] HK: genes
872 expressed in all the three used cell lines and expressed in >= 80% of tissues in the CAGE dataset.
873 CTE: genes expressed in all the three used cell lines and expressed in < 80% of tissues. CTS:
874 genes expressed in =< 2 used cell lines. [human] HK: genes expressed in all the six cell lines and
875 expressed in >= 80% of tissues. CTE: genes expressed in 4 or 5 used cell lines and expressed in

876 >= 80% of tissues or genes expressed in 6 used cell lines and expressed in < 80% of tissues. CTS:
877 genes expressed in =< 3 (out of 6) used cell lines.
878
879 **Statistical analysis**
880 Otherwise mentioned, p-values are calculated using Mann–Whitney U test and corrected for
881 multiple comparisons using Benjamini & Hochberg method (R package stats version 3.6.2).
882
883 **Use of publically available data**
884 The following publicly available data were used: Mouse and Human gene annotations were
885 from Genecode (https://www.gencodegenes.org); FANTOM5 CAGE dataset was from
886 ArrayExpress (https://www.ebi.ac.uk/arrayexpress/); Promoter Elements were referenced from
887 EPD (https://epd.epfl.ch//index.php); Motif references were downloaded from MEME Suite
888 (https://meme-suite.org/meme/index.html) and JASPAR (https://jaspar.genereg.net/). The
889 following publicly available sequencing datasets were reanalyzed: the RNA-seq in KG-1
890 (GSE171552), HepG2 (GSE134745), Jurkat (GSE134745), K562 (GSE134745), SH-SY5Y
891 (GSE162644) and HCT116 (GSE144165); EU-seq in ESC treated without and with A-485
892 (GSE146328); the ChIP-seq on ZFP143 (GSE39263), NRF1 (GSE678674), GABPA (GSE116704),
893 INO80 (GSE158533), YY1 (GSE178982), CTCF (GSE178982) in ESC; H3K4me3 in K562
894 (GSE163049), HepG2 (GSE111000), HCT116 (GSE127960), Jurkat (GSE164184), and SH-SY5Y
895 (GSE90045), ATAC-seq of HCT116 (GSE97889).
896
897 **Cloning of plasmids for genome editing**
898 Oligonucleotide pairs encoding sgRNAs (primers 1-14) were phosphorylated, annealed, and
899 cloned into plasmids pX330 or pX458 by one-pot BbsI digestion/T4 DNA ligation as described[103]
900 followed by exonuclease digestion. Oligonucleotide pairs (primers 15-26) encoding intronic
901 genomic target sites for the *Nanog* enhancer insertion (in proximity to *Mcmbp*, *Ick*, *Sec23ip*,
902 *Foxo9*, *Mthfsd*, and *Foxc2*) were phosphorylated, annealed, and inserted by T4 DNA ligation
903 into the pRRP recombination reporter[104] digested with SacI and AatI. Templates for HDR-based
904 editing were cloned by amplifying genomic regions with KOD Xtreme Hot Start DNA (primers
905 27-42) followed by amplification for homology arms with added multiple cloning site regions
906 using PrimeStar MAX DNA polymerase (primers 43-88) and the pieces assembled by Gibson
907 assembly into a linearized GeneArt pMA vector. A casette encoding an N-terminal Puromycin-
908 P2A was inserted by restriction/ligation cloning to the homology arms containing loxP sites
909 around the *Rif1* enhancer. For the degron editing of *Rif1*, a cassette encoding N-terminal
910 Puromycin-P2A-GFP-FKBP12^{F36V} was similarly inserted to a Rif1 N-terminal homology region. A
911 segment of the *Nanog* enhancer was inserted by digestion/ligation cloning into homology
912 regions proximal to *Mcmbp*, *Sec23ip*, *Ick*, *Fbxo9*, *Foxc2*, and *Mthfsd*. A list of oligonucleotides
913 used for genome editing, genomic confirmation of clones, and RT-PCR is provided in the
914 Supplemental Table 1.
915
916 **Genome editing of mESC**
917 For generating Cre-ERT2 expressing mESCs (E14TG2a, Sigma-Aldrich, cat# 08021401) were
918 stably transfected with a PiggyBac CreERT2-expressing plasmid[105]. For generating *Rif1* enhancer
919 fl/fl cells, the Cre-ERT2 expressing cells were seeded to gelatin-coated wells in a 6-well plate.

920 The day after, 1250ng pX330 and 1250ng HDR templates were mixed with 7.5µL Lipofectamine
921 2000 in a total volume of 500µL OptiMEM, allowed to stand 20min and then added to the cells.
922 After 24h transfection, cells were trypsinized and split 1:20 to a gelatin-coated 10cm dish. After
923 attaching 24h, the cells were selected with 0.5µg/mL puromycin for the next 6-8 days. Correctly
924 modified clones were identified by genotyping PCR with KOD Xtreme Hot Start DNA
925 polymerase.
926
927 For inserting *Nanog* enhancer in proximity to *Mcmbp*, *Fbxo9*, *Ick*, *Sec23ip*, *Mthfsd*, and *Foxc2*,
928 mESC (E14TG2a, Sigma-Aldrich, cat# 08021401) were seeded to gelatin-coated wells in a 6-well
929 plate for each transfection. The day after 200ng pRRP reporter, 1000ng pX330/458, and 1500ng
930 HDR template with the *Nanog* enhancer were transfected with lipofectamine and OptiMEM as
931 outlined above. 24h after transfection cells were trypsinized and split to 1:5 and 1:20 in gelatin-
932 coated 10cm dishes. 2-3h after splitting, cells were transiently selected (24h) with puromycin
933 (2µg/mL). Subsequently, the cells were washed once with PBS, and fresh media, without
934 puromycin, was added. Colonies were picked 6-8 days after transfection and genotyped as
935 described above (primers as for the cloning, or 91-94).
936
937 **mRNA isolation and RT-qPCR for mRNA quantification**
938 Total RNA was isolated using an RNeasy mini kit (Qiagen) as described for EU-seq sample
939 collection. Isolated RNA (500 ng) was reverse-transcribed to cDNA by using PrimeScript RT
940 master mix (Takara). PCR was performed using Mx3005P (Agilent) with SYBR® Premix EX Taq™
941 II (Takara) and primers listed in the table (99-112). The relative expression levels were
942 calculated using *β-actin* mRNA as a reference.
943
944 **Confirmation of *Rif1* enhancer deletion and RIF1 depletion, and cell proliferation analysis**
945 Following trypsinization, *Rif1* enhancer fl/fl or or GFP-FKBP12^{F36V}-RIF1 mESC were re-seeded in
946 gelatin-coated wells in 6-well plates with media supplemented with DMSO (1:1000) (both cell
947 lines), 4-OHT (100nM) (*Rif1* enhancer fl/fl cells), or dTAG-13 (100nM) (GFP-FKBP12^{F36V}-RIF1
948 cells). The cells were imaged 3 and 5 days after the treatment to monitor colony growth. For
949 confirming dTAG13-induced depletion of GFP-FKBP12^{F36V}-RIF1 cells, 1.2e4 cells were seeded in
950 a Geltrex coated 96-well CellCarrier Ultra plate and allowed to attach overnight. The next day,
951 the cells were stained with SiR-DNA (0.1µM) for 1h. Then dTAG-13 (100nM) was added to the
952 cells. The cells were imaged immediately after adding d-TAG13 and again after 1h of dTAG-13
953 addition, on a Perkin Elmer Phenix Harmony confocal spinning disk microscope using 488nm
954 and 647nm excitation.
955
956 To quantify cell viability, the cells were seeded in triplicates in 6-well plates as described above.
957 After 3 days of treatment, the medium from the supernatant was collected, the cells washed
958 once with PBS, wash collected, followed by TrypLE dissociation for 2-3min. Dissociated cells
959 were collected in the same tube as the supernatant and the PBS wash. The wells were washed
960 once more with PBS and collected in the same tubes followed by centrifugation 2min at 400g.
961 The cells were dissociated with a 20µL pipette tip and resuspended in 1mL medium followed by
962 dilution 1:2 with 0.4% trypan blue and loading of 10µL to a Countess cell culture chamber slide
963 and quantification on a Countess 3 cell counter (Thermo Scientific).

964
965
966 ## References
967

968 1 Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of
969 regulatory elements. *Nat Rev Genet* **21**, 71-87, doi:10.1038/s41576-019-0173-8 (2020).
970 2 Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology.
971 *Science* **361**, 1341-1345, doi:10.1126/science.aau0320 (2018).
972 3 Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression
973 control. *Nat Rev Genet* **20**, 437-455, doi:10.1038/s41576-019-0128-0 (2019).
974 4 Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of
975 gene promoters. *Nature* **489**, 109-113, doi:10.1038/nature11279 (2012).
976 5 Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological
977 basis for transcription regulation. *Cell* **148**, 84-98, doi:10.1016/j.cell.2011.12.014 (2012).
978 6 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in
979 human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).
980 7 Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target
981 genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612,
982 doi:10.1038/ng.3963 (2017).
983 8 Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for
984 functional genetic variants. *Nat Genet* **53**, 110-119, doi:10.1038/s41588-020-00745-3
985 (2021).
986 9 van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants
987 of enhancer-promoter interaction specificity. *Trends Cell Biol* **24**, 695-702,
988 doi:10.1016/j.tcb.2014.07.004 (2014).
989 10 Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture
990 mapping. *Nature* **543**, 519-524, doi:10.1038/nature21411 (2017).
991 11 Kermekchiev, M., Pettersson, M., Matthias, P. & Schaffner, W. Every enhancer works
992 with every promoter for all the combinations tested: could new regulatory pathways
993 evolve by enhancer shuffling? *Gene Expr* **1**, 71-81 (1991).
994 12 Bergman, D. T. *et al.* Compatibility rules of human enhancer and promoter sequences.
995 *Nature*, doi:10.1038/s41586-022-04877-w (2022).
996 13 Sahu, B. *et al.* Sequence determinants of human gene regulatory elements. *Nat Genet*
997 **54**, 283-294, doi:10.1038/s41588-021-01009-4 (2022).
998 14 Butler, J. E. & Kadonaga, J. T. Enhancer-promoter specificity mediated by DPE or TATA
999 core promoter motifs. *Genes Dev* **15**, 2515-2519, doi:10.1101/gad.924301 (2001).
1000 15 Gehrig, J. *et al.* Automated high-throughput mapping of promoter-enhancer interactions
1001 in zebrafish embryos. *Nat Methods* **6**, 911-916, doi:10.1038/nmeth.1396 (2009).
1002 16 Wefald, F. C., Devlin, B. H. & Williams, R. S. Functional heterogeneity of mammalian
1003 TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* **344**,
1004 260-262, doi:10.1038/344260a0 (1990).
1005 17 Li, X. & Noll, M. Compatibility between enhancers and promoters determines the
1006 transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila
1007 embryo. *EMBO J* **13**, 400-406 (1994).

1008 18 Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and
1009 housekeeping gene regulation. *Nature* **518**, 556-559, doi:10.1038/nature13994 (2015).
1010 19 Martinez-Ara, M., Comoglio, F., van Arensbergen, J. & van Steensel, B. Systematic
1011 analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol Cell*,
1012 doi:10.1016/j.molcel.2022.04.009 (2022).
1013 20 Pachano, T. *et al.* Orphan CpG islands amplify poised enhancer regulatory activity and
1014 determine target gene responsiveness. *Nat Genet* **53**, 1036-1049, doi:10.1038/s41588-
1015 021-00888-x (2021).
1016 21 Neumayr, C. *et al.* Differential cofactor dependencies define distinct types of human
1017 enhancers. *Nature* **606**, 406-413, doi:10.1038/s41586-022-04779-x (2022).
1018 22 Haberle, V. *et al.* Transcriptional cofactors display specificity for distinct types of core
1019 promoters. *Nature* **570**, 122-126, doi:10.1038/s41586-019-1210-7 (2019).
1020 23 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation
1021 and gene control. *Science* **361**, doi:10.1126/science.aar3958 (2018).
1022 24 Lim, B. & Levine, M. S. Enhancer-promoter communication: hubs or loops? *Curr Opin*
1023 *Genet Dev* **67**, 5-9, doi:10.1016/j.gde.2020.10.001 (2021).
1024 25 Grosveld, F., van Staalduinen, J. & Stadhouders, R. Transcriptional Regulation by
1025 (Super)Enhancers: From Discovery to Mechanisms. *Annu Rev Genomics Hum Genet* **22**,
1026 127-146, doi:10.1146/annurev-genom-122220-093818 (2021).
1027 26 Bailey, S. D. *et al.* ZNF143 provides sequence specificity to secure chromatin interactions
1028 at gene promoters. *Nat Commun* **2**, 6186, doi:10.1038/ncomms7186 (2015).
1029 27 Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by
1030 alternative chromatin loop formation. *Proc Natl Acad Sci U S A* **105**, 20398-20403,
1031 doi:10.1073/pnas.0808506106 (2008).
1032 28 Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent
1033 gene activation. *Nat Struct Mol Biol* **28**, 152-161, doi:10.1038/s41594-020-00539-5
1034 (2021).
1035 29 Klein, J. C. *et al.* A systematic evaluation of the design and context dependencies of
1036 massively parallel reporter assays. *Nat Methods* **17**, 1083-1091, doi:10.1038/s41592-
1037 020-0965-y (2020).
1038 30 Kragesteen, B. K. *et al.* Dynamic 3D chromatin architecture contributes to enhancer
1039 specificity and limb morphogenesis. *Nat Genet* **50**, 1463-1473, doi:10.1038/s41588-018-
1040 0221-x (2018).
1041 31 Doni Jayavelu, N., Jajodia, A., Mishra, A. & Hawkins, R. D. Candidate silencer elements
1042 for the human and mouse genomes. *Nat Commun* **11**, 1061, doi:10.1038/s41467-020-
1043 14853-5 (2020).
1044 32 Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-
1045 range associations. *Nature* **504**, 306-310, doi:10.1038/nature12716 (2013).
1046 33 Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by
1047 complex genomic signatures on looping chromatin. *Nat Genet* **48**, 488-496,
1048 doi:10.1038/ng.3539 (2016).
1049 34 Narita, T. *et al.* Enhancers are activated by p300/CBP activity-dependent PIC assembly,
1050 RNAPII recruitment, and pause release. *Mol Cell* **81**, 2166-2182 e2166,
1051 doi:10.1016/j.molcel.2021.03.008 (2021).

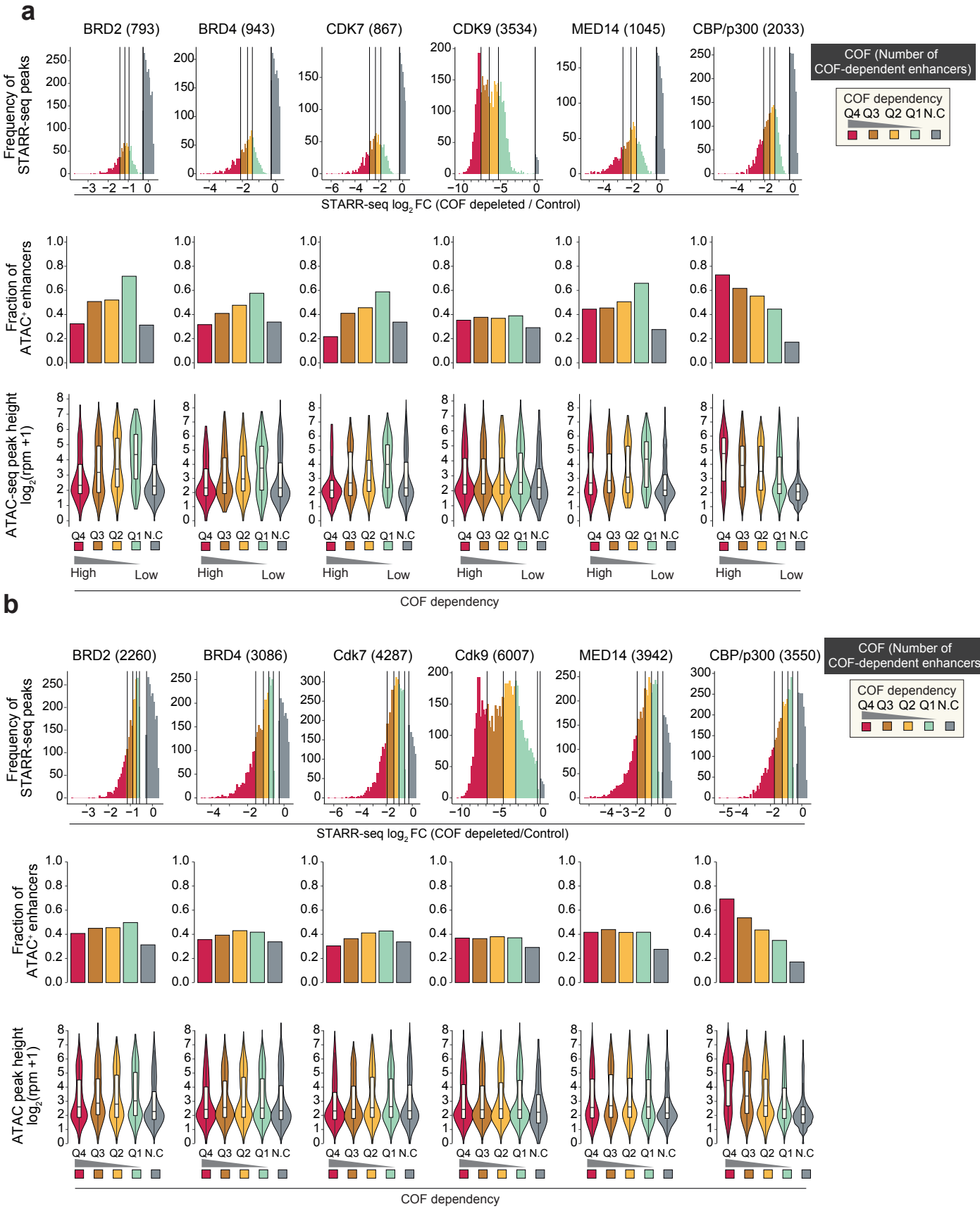
1052 35 Consortium, E. P. *et al.* Expanded encyclopaedias of DNA elements in the human and
1053 mouse genomes. *Nature* **583**, 699-710, doi:10.1038/s41586-020-2493-4 (2020).
1054 36 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**,
1055 1260419, doi:10.1126/science.1260419 (2015).
1056 37 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes.
1057 *Nucleic Acids Res* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).
1058 38 Lasko, L. M. *et al.* Discovery of a selective catalytic p300/CBP inhibitor that targets
1059 lineage-specific tumours. *Nature* **550**, 128-132, doi:10.1038/nature24028 (2017).
1060 39 Weinert, B. T. *et al.* Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of
1061 the CBP/p300 Acetylome. *Cell* **174**, 231-244 e212, doi:10.1016/j.cell.2018.04.033 (2018).
1062 40 Vannam, R. *et al.* Targeted degradation of the enhancer lysine acetyltransferases CBP
1063 and p300. *Cell Chem Biol* **28**, 503-514 e512, doi:10.1016/j.chembiol.2020.12.004 (2021).
1064 41 Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the
1065 H19/Igf2 locus. *Nature* **405**, 486-489, doi:10.1038/35013106 (2000).
1066 42 Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor.
1067 *Nature* **451**, 796-801, doi:10.1038/nature06634 (2008).
1068 43 Milligan, L. *et al.* H19 gene expression is up-regulated exclusively by stabilization of the
1069 RNA during muscle cell differentiation. *Oncogene* **19**, 5810-5816,
1070 doi:10.1038/sj.onc.1203965 (2000).
1071 44 Yeh, L. C., Adamo, M. L., Duan, C. & Lee, J. C. Osteogenic protein-1 regulates insulin-like
1072 growth factor-I (IGF-I), IGF-II, and IGF-binding protein-5 (IGFBP-5) gene expression in
1073 fetal rat calvaria cells by different mechanisms. *J Cell Physiol* **175**, 78-88,
1074 doi:10.1002/(SICI)1097-4652(199804)175:1<78::AID-JCP9>3.0.CO;2-9 (1998).
1075 45 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516,
1076 doi:10.1016/j.cell.2017.06.010 (2017).
1077 46 Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem
1078 cell core transcriptional network. *Cell* **145**, 183-197, doi:10.1016/j.cell.2011.03.003
1079 (2011).
1080 47 Tsukamoto, S. *et al.* Autophagy is essential for preimplantation development of mouse
1081 embryos. *Science* **321**, 117-120, doi:10.1126/science.1154822 (2008).
1082 48 Suvorova, II, Knyazeva, A. R., Petukhov, A. V., Aksenov, N. D. & Pospelov, V. A.
1083 Resveratrol enhances pluripotency of mouse embryonic stem cells by activating
1084 AMPK/Ulk1 pathway. *Cell Death Discov* **5**, 61, doi:10.1038/s41420-019-0137-y (2019).
1085 49 Yamazaki, S. *et al.* Rif1 regulates the replication timing domains on the human genome.
1086 *EMBO J* **31**, 3667-3677, doi:10.1038/emboj.2012.180 (2012).
1087 50 Cornacchia, D. *et al.* Mouse Rif1 is a key regulator of the replication-timing programme
1088 in mammalian cells. *EMBO J* **31**, 3678-3690, doi:10.1038/emboj.2012.214 (2012).
1089 51 Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in
1090 mouse embryonic stem cells. *Nat Genet* **38**, 431-440, doi:10.1038/ng1760 (2006).
1091 52 Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation.
1092 *Nat Chem Biol* **14**, 431-441, doi:10.1038/s41589-018-0021-8 (2018).
1093 53 Luo, J., Yang, H. & Song, B. L. Mechanisms and regulation of cholesterol homeostasis.
1094 *Nat Rev Mol Cell Biol* **21**, 225-245, doi:10.1038/s41580-019-0190-7 (2020).

1095 54 Eagle, K. *et al.* An oncogenic enhancer encodes selective selenium dependency in AML.
1096 *Cell Stem Cell* **29**, 386-399 e387, doi:10.1016/j.stem.2022.01.003 (2022).
1097 55 Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the
1098 human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A*
1099 **103**, 1412-1417, doi:10.1073/pnas.0510310103 (2006).
1100 56 Zhu, J., He, F., Hu, S. & Yu, J. On the nature of human housekeeping genes. *Trends Genet*
1101 **24**, 481-484, doi:10.1016/j.tig.2008.08.004 (2008).
1102 57 Rozenberg, J. M. *et al.* All and only CpG containing sequences are enriched in promoters
1103 abundantly bound by RNA polymerase II in multiple tissues. *BMC Genomics* **9**, 67,
1104 doi:10.1186/1471-2164-9-67 (2008).
1105 58 Curina, A. *et al.* High constitutive activity of a broad panel of housekeeping and tissue-
1106 specific cis-regulatory elements depends on a subset of ETS proteins. *Genes Dev* **31**,
1107 399-412, doi:10.1101/gad.293134.116 (2017).
1108 59 Domcke, S. *et al.* Competition between DNA methylation and transcription factors
1109 determines binding of NRF1. *Nature* **528**, 575-579, doi:10.1038/nature16462 (2015).
1110 60 Michaud, J. *et al.* HCFC1 is a common component of active human CpG-island promoters
1111 and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy.
1112 *Genome Res* **23**, 907-916, doi:10.1101/gr.150078.112 (2013).
1113 61 Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**,
1114 1010-1022, doi:10.1101/gad.2037511 (2011).
1115 62 Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in
1116 mammalian chromosomes. *Cell* **159**, 374-387, doi:10.1016/j.cell.2014.09.030 (2014).
1117 63 van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in
1118 human cells. *Nat Biotechnol* **35**, 145-153, doi:10.1038/nbt.3754 (2017).
1119 64 Hsieh, T. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian
1120 Chromatin Folding. *Mol Cell* **78**, 539-553 e538, doi:10.1016/j.molcel.2020.03.002 (2020).
1121 65 Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture.
1122 *Mol Cell* **78**, 554-565 e557, doi:10.1016/j.molcel.2020.03.003 (2020).
1123 66 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324,
1124 doi:10.1016/j.cell.2017.09.026 (2017).
1125 67 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of
1126 Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922,
1127 doi:10.1016/j.cell.2017.05.004 (2017).
1128 68 Zuin, J. *et al.* Nonlinear control of transcription through enhancer-promoter
1129 interactions. *Nature* **604**, 571-577, doi:10.1038/s41586-022-04570-y (2022).
1130 69 Yu, N. Y. *et al.* Complementing tissue characterization by integrating transcriptome
1131 profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic*
1132 *Acids Res* **43**, 6787-6798, doi:10.1093/nar/gkv608 (2015).
1133 70 Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in
1134 mammalian development. *Nature* **554**, 239-243, doi:10.1038/nature25461 (2018).
1135 71 Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in
1136 development and disease. *Nat Rev Genet* **22**, 324-336, doi:10.1038/s41576-020-00311-x
1137 (2021).

1138 72 Hong, J. W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary
1139 novelty. *Science* **321**, 1314, doi:10.1126/science.1160631 (2008).
1140 73 Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**,
1141 206-216, doi:10.1038/nrg2063 (2007).
1142 74 Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in
1143 regulatory DNA. *Science* **337**, 1190-1195, doi:10.1126/science.1222794 (2012).
1144 75 Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature*
1145 **593**, 238-243, doi:10.1038/s41586-021-03446-x (2021).
1146 76 Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of
1147 morphological evolution. *Cell* **134**, 25-36, doi:10.1016/j.cell.2008.06.030 (2008).
1148 77 Bajic, V. B. *et al.* Mice and men: their promoter properties. *PLoS Genet* **2**, e54,
1149 doi:10.1371/journal.pgen.0020054 (2006).
1150 78 Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues.
1151 *Nature* **507**, 455-461, doi:10.1038/nature12787 (2014).
1152 79 Hartl, D. *et al.* CG dinucleotides enhance promoter activity independent of DNA
1153 methylation. *Genome Res* **29**, 554-563, doi:10.1101/gr.241653.118 (2019).
1154 80 Peng, T. *et al.* STARR-seq identifies active, chromatin-masked, and dormant enhancers in
1155 pluripotent mouse embryonic stem cells. *Genome Biol* **21**, 243, doi:10.1186/s13059-020-
1156 02156-3 (2020).
1157 81 Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous
1158 fusion. *Nature* **416**, 545-548, doi:10.1038/nature729 (2002).
1159 82 Pollard, S. M., Benchoua, A. & Lowell, S. Neural stem cells, neurons, and glia. *Methods*
1160 *Enzymol* **418**, 151-169, doi:10.1016/S0076-6879(06)18010-6 (2006).
1161 83 Ying, Q. L. & Smith, A. G. Defined conditions for neural commitment and differentiation.
1162 *Methods Enzymol* **365**, 327-341, doi:10.1016/s0076-6879(03)65023-8 (2003).
1163 84 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
1164 doi:10.1093/bioinformatics/bts635 (2013).
1165 85 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc*
1166 *Bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).
1167 86 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
1168 throughput sequencing data. *Bioinformatics* **31**, 166-169,
1169 doi:10.1093/bioinformatics/btu638 (2015).
1170 87 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
1171 transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
1172 88 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
1173 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
1174 89 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
1175 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8
1176 (2014).
1177 90 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006,
1178 doi:10.1101/gr.229102 (2002).
1179 91 Hentges, L. D., Sergeant, M. J., Downes, D. J., Hughes, J. R. & Taylor, S. LanceOtron: a
1180 deep learning peak caller for ATAC-seq, ChIP-seq, and DNase-seq. *bioRxiv*,
1181 2021.2001.2025.428108, doi:10.1101/2021.01.25.428108 (2021).

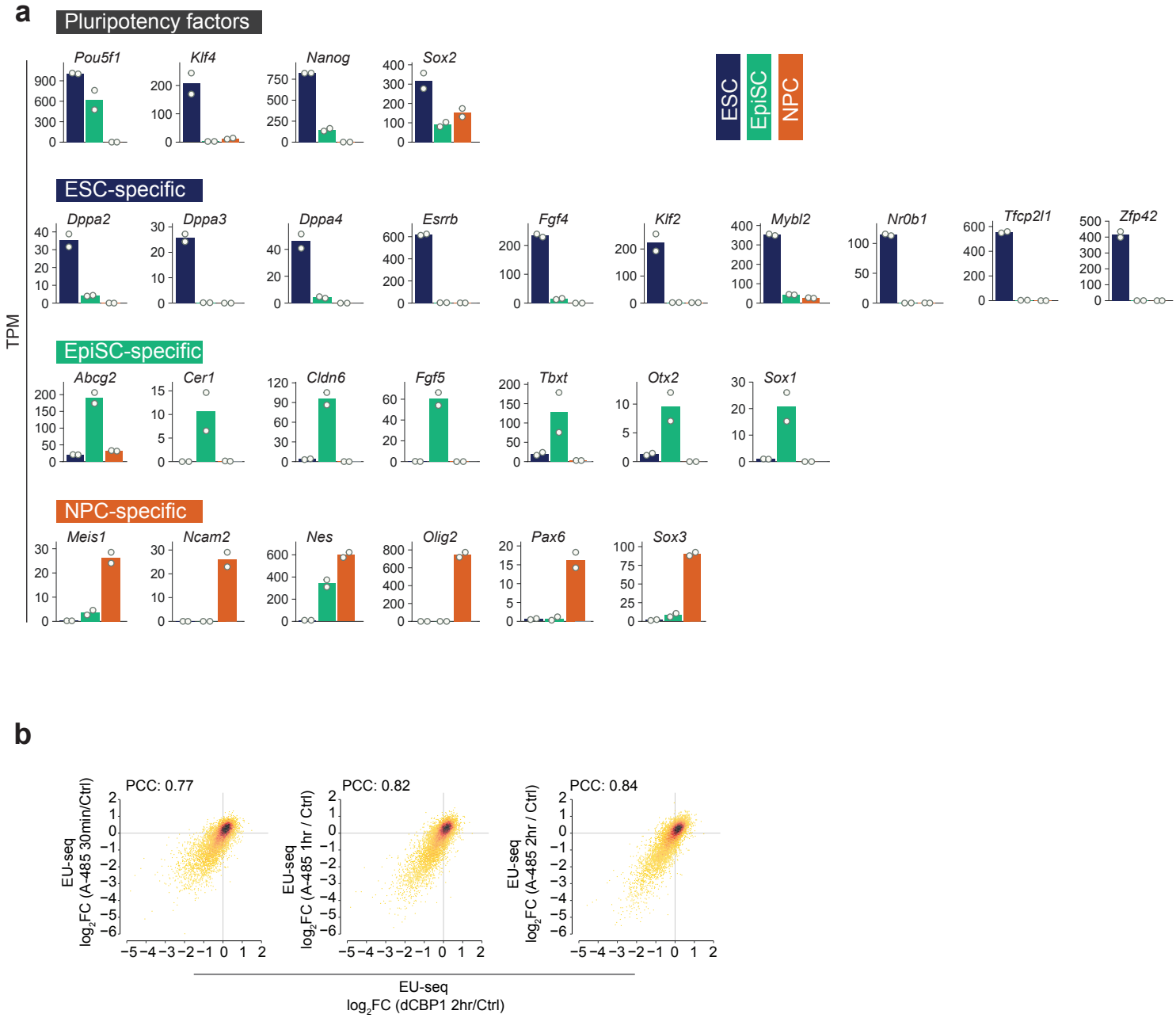
1182 92 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
1183 features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
1184 93 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
1185 analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).
1186 94 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak
1187 annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383,
1188 doi:10.1093/bioinformatics/btv145 (2015).
1189 95 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV):
1190 high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-
1191 192, doi:10.1093/bib/bbs017 (2013).
1192 96 Ghandi, M. *et al.* gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* **32**, 2205-
1193 2207, doi:10.1093/bioinformatics/btw203 (2016).
1194 97 Lee, D. LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinformatics* **32**, 2196-2198,
1195 doi:10.1093/bioinformatics/btw142 (2016).
1196 98 Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier
1197 performance in R. *Bioinformatics* **21**, 3940-3941, doi:10.1093/bioinformatics/bti623
1198 (2005).
1199 99 Shrikumar, A., Prakash, E. & Kundaje, A. GkmExplain: fast and accurate interpretation of
1200 nonlinear gapped k-mer SVMs. *Bioinformatics* **35**, i173-i182,
1201 doi:10.1093/bioinformatics/btz322 (2019).
1202 100 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*
1203 **37**, W202-208, doi:10.1093/nar/gkp335 (2009).
1204 101 Dreos, R., Ambrosini, G., Perier, R. C. & Bucher, P. The Eukaryotic Promoter Database:
1205 expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res* **43**, D92-96,
1206 doi:10.1093/nar/gku1111 (2015).
1207 102 Hon, C. C. *et al.* An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*
1208 **543**, 199-204, doi:10.1038/nature21374 (2017).
1209 103 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-
1210 2308, doi:10.1038/nprot.2013.143 (2013).
1211 104 Flemr, M. & Buhler, M. Single-Step Generation of Conditional Knockout Mouse
1212 Embryonic Stem Cells. *Cell Rep* **12**, 709-716, doi:10.1016/j.celrep.2015.06.051 (2015).
1213 105 Andersson-Rolf, A. *et al.* One-step generation of conditional and reversible gene
1214 knockouts. *Nat Methods* **14**, 287-289, doi:10.1038/nmeth.4156 (2017).
1215

Extended Data Fig. 1



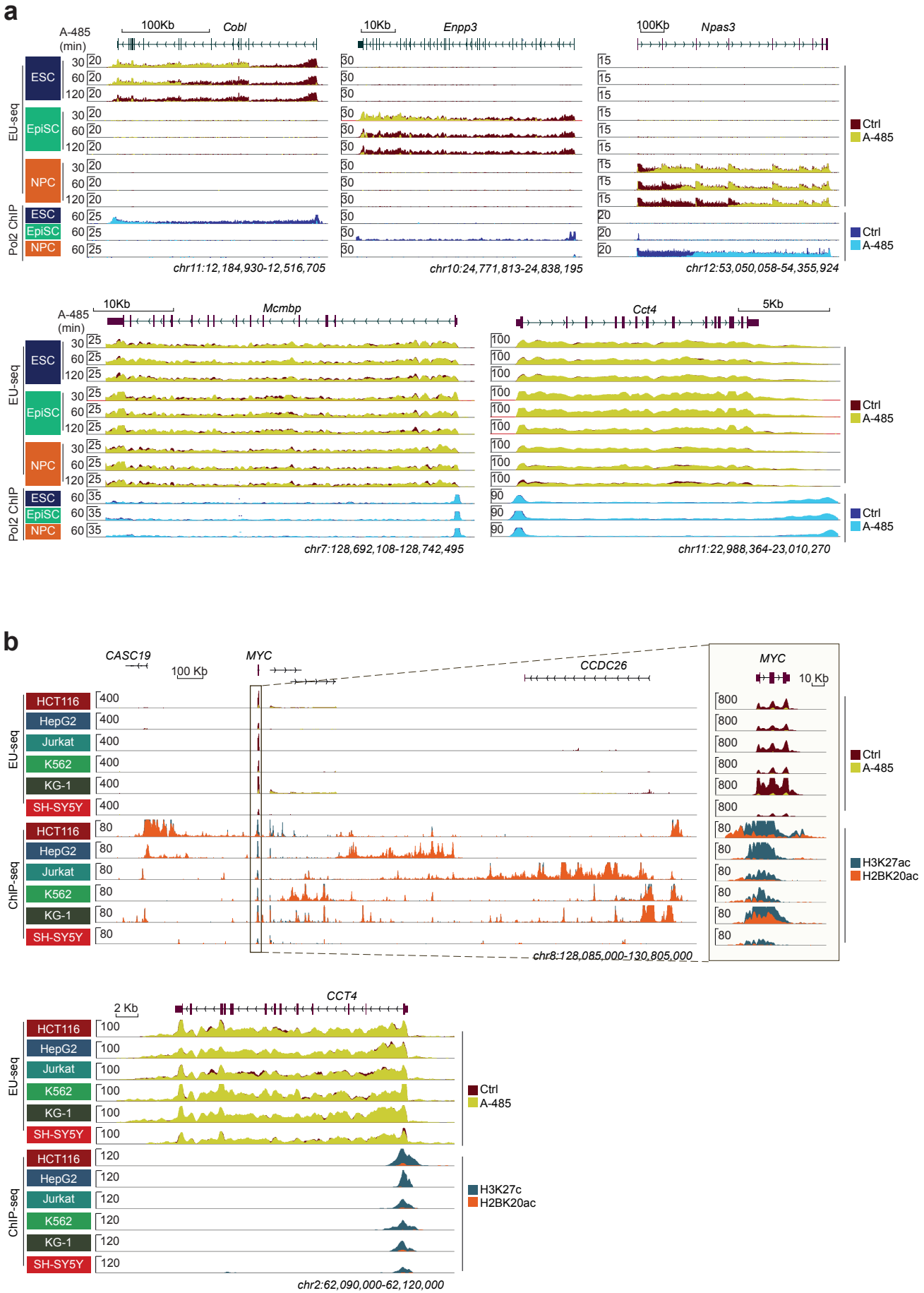
Extended Data Fig. 1. CBP/p300 is required for enhancer-dependent transcription activation in STARR-seq. a, The requirement of the six indicated cofactors (COFs; BRD2, BRD4, CDK7, CDK9, MED14, CBP/p300) for enhancer activity was assessed by Neumayr et al. using STARR seq21. The activity of 6249 candidate enhancers was quantified in wild-type cells or cells depleted of the indicated COFs. Based on the STARR-seq signal, the relative change in enhancer activity in wild-type and COF depleted cells was determined. For each COF, the number of enhancers that showed decreased activity in COF-depleted cells is indicated. The number of enhancers showing COF dependency was determined using statistical significance (p=<0.05) (shown in panel a) or fold downregulation (>1.5 fold decrease in enhancer activity in COF depleted cells) (shown in panel b). Based on enhancer activity fold-change in COF-depleted cells, enhancers were grouped into quartiles. Enhancers showing no change (N.C.) in activity after the depletion of the indicated COFs are shown as a separate group. In each of the above enhancer groups, native chromatin accessibility was analyzed in the cell line in which the enhancer activity was assayed (HCT116). Chromatin accessibility was determined by ATAC-seq. In the top panel, the frequency plot shows the fold-change distribution of enhancer activity in COF depleted and wild-type control. In the second raws, bar charts show the fraction of enhancers that are ATAC-seq^{+} in HCT116. In the last raw, violin plots show DNA accessibility by ATAC-seq in the corresponding group of enhancers. The fraction of enhancers showing a decrease in STARR-seq activity after the depletion of the indicated COFs is indicated. Analyses in panel a include enhancers whose activity in MPRA is statistically significantly decreased in the indicated COF depleted cells. Analyses in panel **b** include enhancers whose activity in MPRA is decreased by >1.5 fold (regardless of statistical significance) in the indicated COF depleted cells. Note that, among all analyzed COFs, CBP/p300 is the only COF that shows a strong association with the accessibility of enhancers in the native context. Enhancers whose activity is most strongly decreased after CBP/p300 depletion have the highest accessibility, enhancers that show intermediate dependency on CBP/p300 have intermediate accessibility, and enhancers that are insensitive to CBP/p300 depletion either lack or have very low DNA accessibility.

Extended Data Fig. 2



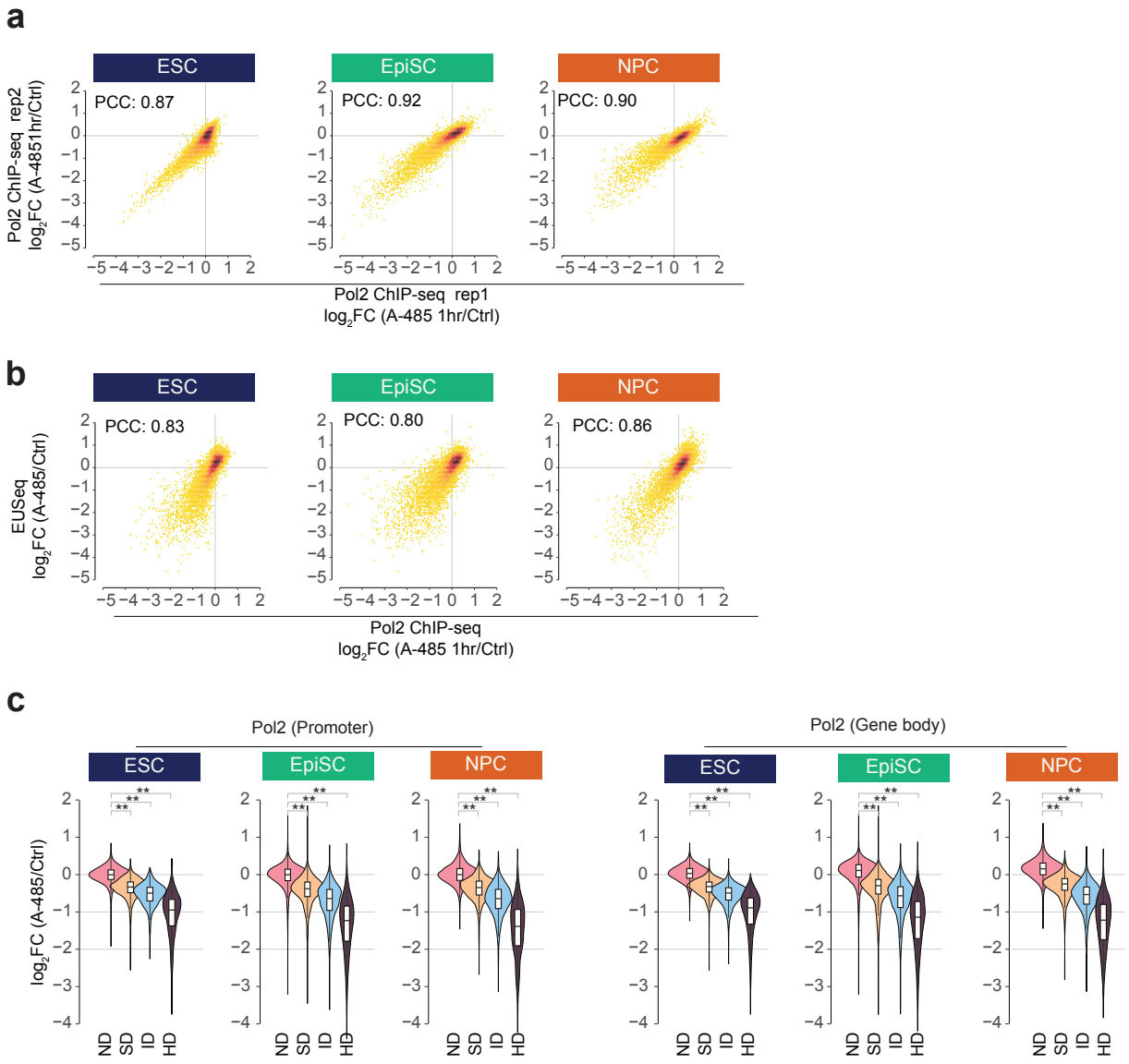
Extended Data Fig. 2. Confirmation of ESC, EpiSC, and NPC cell identity, and validation of A-485 target specificity. a, Expression profiles of cell type-specific transcription factors and signaling proteins in ESC, EpiSC, and NPC. Shown are the mRNA expression levels of the indicated genes, among the specified cell lines. Gene expression levels were quantified from two independent RNA-seq experiments. (TPM; transcripts per million). Two biological replicates were performed for RNA-seq in ESC, EpiSC, and NPC. **b,** CBP/p300 catalytic inhibition and acute CBP/p300 depletion cause similar gene transcription changes. ESCs were treated with the catalytic inhibitor A-485 for 30 min, 1h, and 2h, and with the CBP/p300 PROTAC dCBP1 for 2h. Nascent transcription changes were quantified using EU-seq. Shown are correlations between A-485 and dCBP1-induced transcription changes at the indicated time points. Note that A-485 and dCBP1-induced transcription changes are highly correlated, and A-485 shows no overt signs of off-target specificity. EU-seq data for A-485 treated ESC were re-processed from[34]. dCBP1-induced nascent transcription change was analyzed from a single experiment. PCC: Pearson´s correlation coefficient.

Extended Data Fig. 3



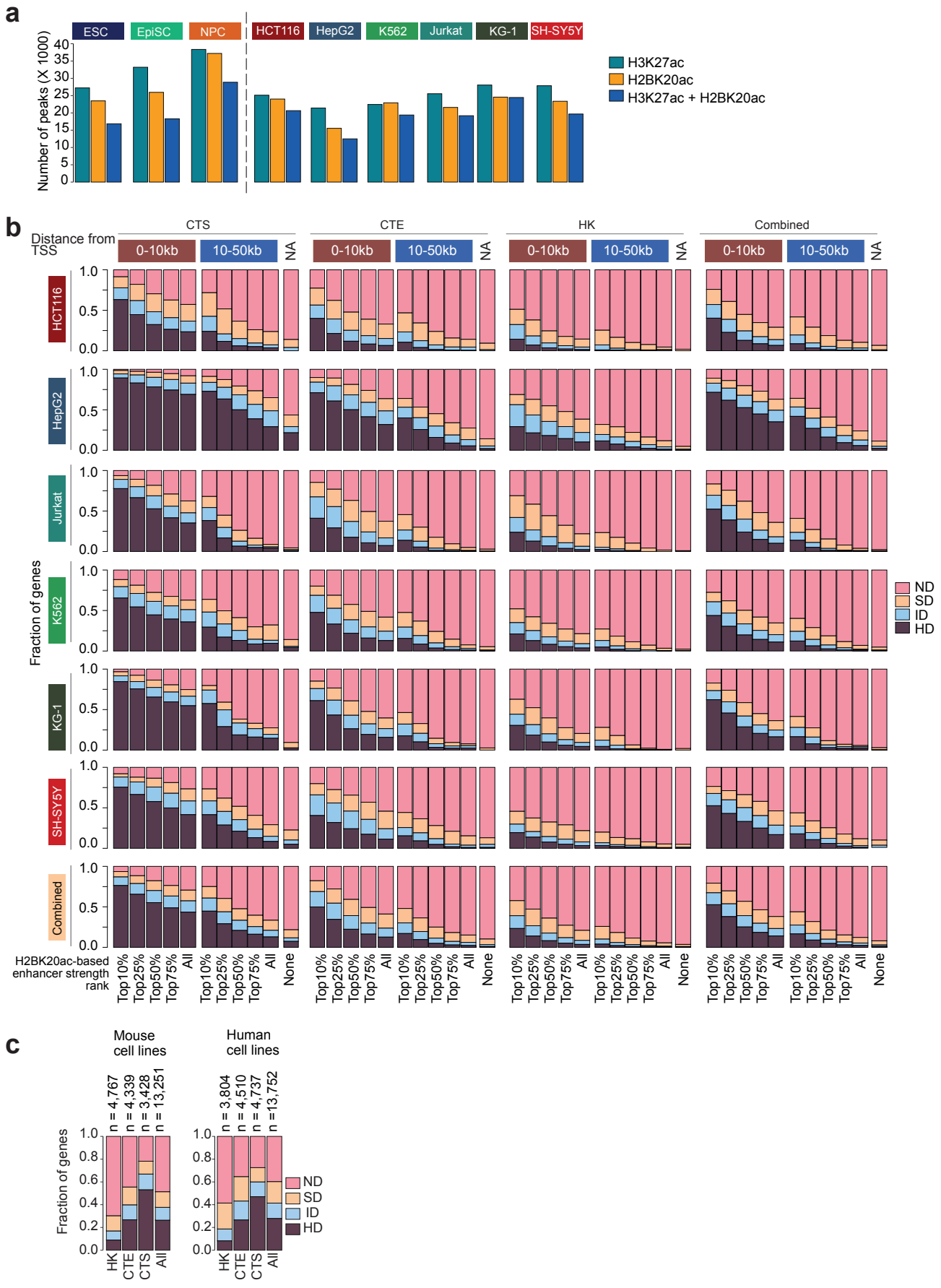
Extended Data Fig. 3. Confirmation of A-485-induced selective downregulation of enhancer-regulated genes in mouse and human cell lines. a, Representative nascent transcription tracks (top panels) in mouse cell lines showing A-485-induced downregulation of the indicated cell-type-specific, enhancer-dependent genes in (*Cobl, Enpp3, Npas3*), or housekeeping genes (*Mcmbp, Cct4*). The overlaid tracks show EU-seq signal in untreated and A-485-treated (30, 60, 120 min) cells, as indicated. Relatively long cell-type-specific genes were chosen as representative examples to demonstrate a time-dependent increase in the length of nascent transcript inhibition. The length of A-485-induced transcript inhibition roughly corresponds to the speed of Pol2 elongation. The overlaid Pol2 data (bottom panels) show Pol2 binding in the indicated treatment conditions and cell lines. Pol2 binding was analyzed by ChIP-seq, without or with A-485 treatment (60 min). Reduced Pol2 binding in ChIP-seq analyses independently confirms the above EU-seq results. *Mcmbp* and *Cct4* remain unaffected by A-485 in EU-seq and Pol2 ChIP-seq analyses in all cell lines. For ESC, EU-seq data were re-processed from[34]. EU-seq analyses in EpiSC and NPC were performed in minimum 3 independent biological replicates, and Pol2 ChIP-seq in ESC, EpiSC, and NPC were performed in two biological replicates. b, Confirmation of A-485-induced selective gene downregulation in human cell lines. Genome browser view of nascent transcription expression in the indicated human cell lines, and treatment conditions. For each cell line, overlaid tracks show EU-seq signals in untreated and A-485 treated conditions. Note that A-485 downregulates enhancer target (*MYC*), but not housekeeping (*CCT4*), genes. Nascent transcription was analyzed in the indicated 6 cell lines using EU-seq, without or with A-485 (60min). Two biological replicates were performed in all the indicated cell lines. Overlaid H3K27ac and H2BK20ac ChIP-seq tracks indicate the position of proximal candidate cis-regulatory elements.

Extended Data Fig. 4



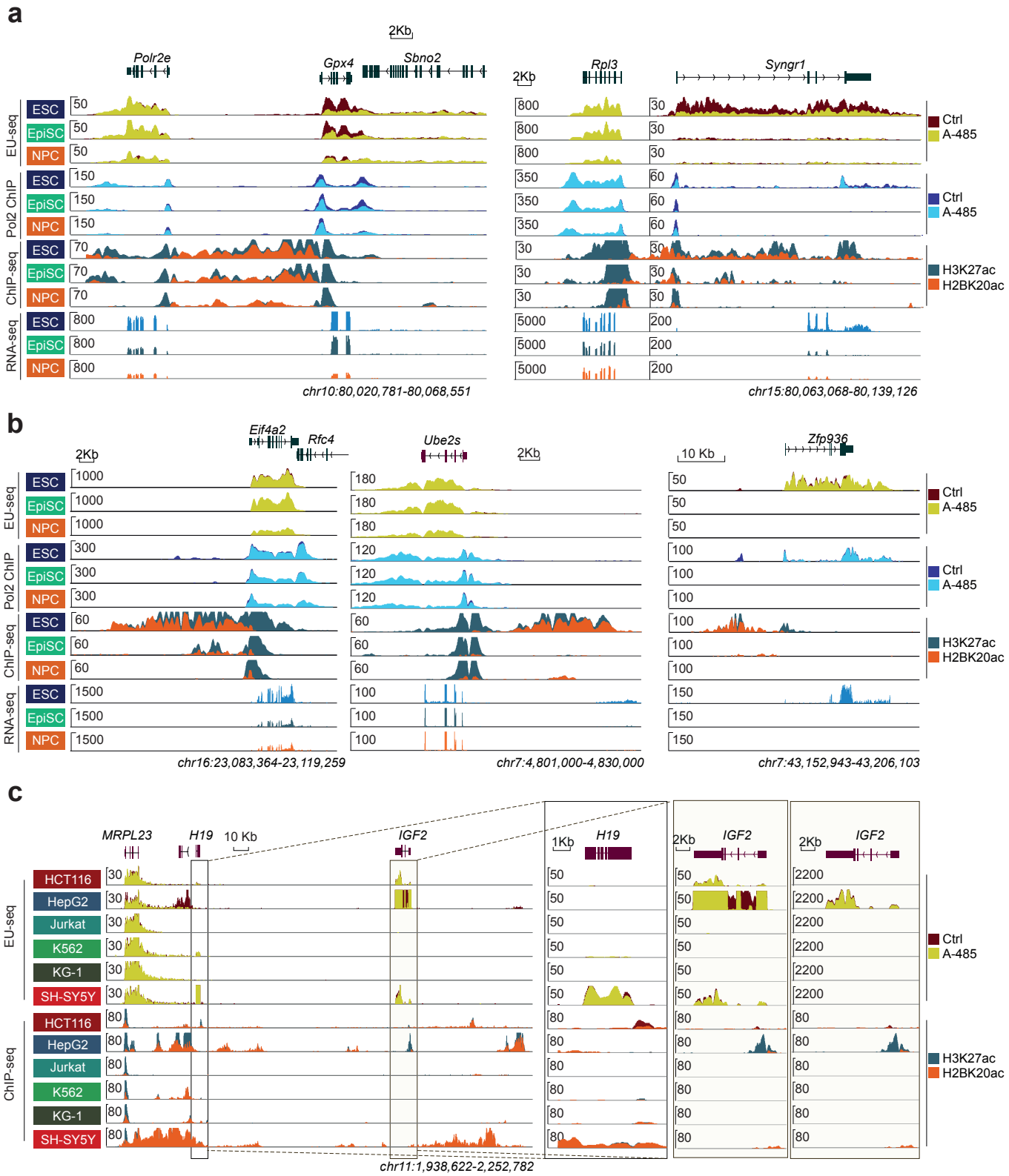
Extended Data Fig. 4. A-485-induced nascent transcription changes correlate with the changes in Pol2 binding. a, Shown are correlations between two independent Pol2 ChIP-seq replicate measurements in the indicated cell lines. Pol2 binding was analyzed without or with A-485 (1h) treatment. PCC: Pearson´s correlation coefficient. **b,** Correlation between A-485-induced nascent transcription changes (EU-seq) and Pol2 binding (Pol2 ChIP-seq) in the indicated cell lines. Note that, in all cell lines, A-485-induced change in Pol2 binding is strongly correlated with the change in EU-seq signal. **c,** A-485 treatment proportionally reduces Pol2 binding at A-485 downregulated gene class. Shown is the normalized Pol2 intensity in A-485 treated and untreated cells. Assuming that Pol2 binding remains unaffected at promoters of genes that are not downregulated (ND) by A-485 in EU-seq analyses, Pol2 binding untreated and A-485 treated cells were normalized based on Pol2 intensity at ND gene promoters (+/– 2kb from TSS). Shown are normalized Pol2 ChIP intensity at promoters (left panels) or gene body (right panels) regions in the indicated cell lines and the A-485 regulated gene class. A-485-regulated genes were classified based on A-485-induced nascent transcription regulation as defined in Fig. 1b. HD, highly downregulated (≥2-fold decreased); ID, intermediate downregulated (>1.5-2.0-fold decreased); SD, slightly downregulated (>1.2-1.5-fold decreased); and NC, not downregulated (≤1.2-fold decreased) after A-485 treatment. *** $p < 1e-50$, Mann–Whitney U test.

Extended Data Fig. 5



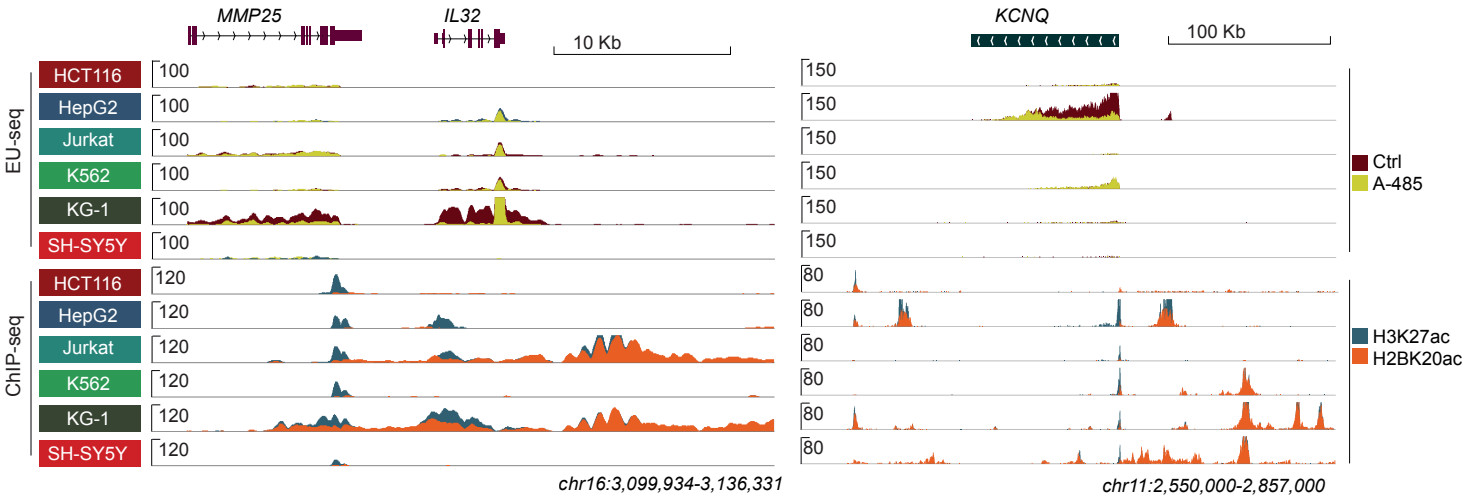
Extended Data Fig. 5. CBP/p300 inhibition downregulates enhancer proximal genes, and CTS shows greater enhancer dependence than HK genes. a, Shown are the number of H3K27ac and H2BK20ac ChIP-seq peaks identified, as well as the number of regions commonly marked with both H3K27ac and H2BK20ac, in the indicated cell lines. b, H2BK20ac intensity strongly relates to A-485-induced gene regulation in human cell lines. In the indicated cell lines, genes were classified based on their cell type-specificity as housekeeping (HK), cell-type-enriched (CTE), and cell-type-specific (CTE). A-485-downregulated genes were classified as defined in Fig 1b. Enhancers commonly marked with H2BK20ac and H3K27ac were ranked based on H2BK20ac intensity. Enhancers were grouped into Top10%, 25%, 50%, 75%, and All, based on H2BK20ac intensity. Enhancers were further grouped based on the distance from TSS of the nearest expressed genes. In each cell line, the bar charts show the relationship between cell-type-specificity class, enhancer rank category, enhancer distance, and A-485-induced gene regulation. Note that the fraction of A-485 downregulated genes progressively decreases with decreasing cell-type-specificity, decreasing enhancer strength, and increasing enhancer distance. In all cell-type-specificity categories, genes occurring in proximity to strong enhancers are more frequently downregulated by A-485 than genes occurring in proximity to weaker and distally located enhancers. c, Enhancer activity regulates a large fraction of mammalian genes. Fraction of A-485 regulated gene class among the specified cell-type-specificity categories in all mouse (left: ESC, EpiSC, NPC) and human (right: HCT116, K562, HepG2, Jurkat, KG-1, and SH-SY5Y) cell lines. The quantified genes were ranked based on their A-485-induced regulation, and assigned to the maximum downregulated category. For example, if a gene was downregulated ≥2-fold in one cell line but was regulated < 2-fold in other cell lines, the gene was assigned to the highest downregulation category, based on the maximum fold downregulation. The cumulative bar charts show the distribution of A-485 downregulated gene categories in the specified cell-type-specific groups. The numbers shown (top) indicate the number of non-redundant genes in each cell-type-specificity category, as well as the total number of non-redundant genes quantified. The 'Combined' category includes genes from CTS, CTE, and HK categories, as well as 'NC' genes that were not classified in any of these categories.

Extended Data Fig. 6



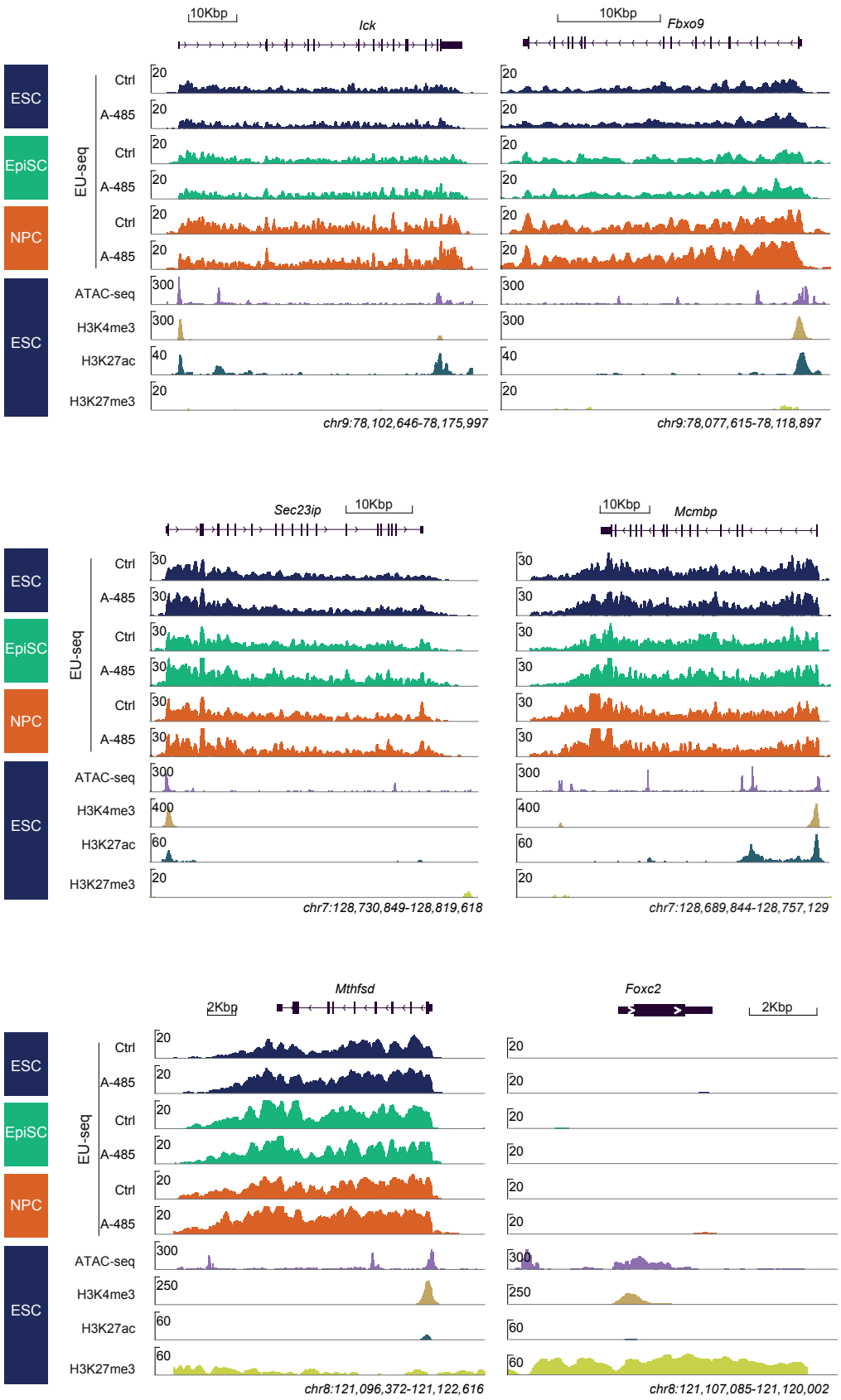
Extended Data Fig. 6. Gene-selective sensitivity to proximal enhancers. a, Representative examples of gene pairs that occur in proximity (<10kb) enhancers, but one of them is downregulated by A-485. In ESC and EpiSC, enhancer(s) is located between *Polr2e* and *Gpx4*. Only *Gpx4*, which is more closely positioned to the enhancer, is downregulated. Similarly, *Rpl3* and *Syngr1* occur in proximity to enhancers in ESC, but only *Syngr1*, which is more proximally positioned to the enhancers, is downregulated. Targeting of enhancers to *Gpx4* and *Syngr1* may explain the A-485 insensitivity of the other proximal genes. **b,** Examples of ubiquitously expressed (*Eif4a2* and *Ube2S*) and cell-type-specific (*Zfp936*) genes that appear insensitive to proximal enhancers. These genes represent exceptions whose enhancer insensitivity remains unexplained. **c,** 3D genome organization can promote preferential enhancer-promoter pairing, and genes with short (or none) intron sequences and stable mRNA can falsely appear as enhancer-insensitive in nascent transcription analyses. Through 3D genome organization, enhancers located between *MRPL23* and *H19* connect with and regulate *H19* and *IGF2*[41], which may explain the A-485-insensitivity of *MRPL23*. Because of the long mRNA half-life, in EU-seq analyses, the overall expression of H19 and *IGF2* is not downregulated after A-485. But a zoom-in view of the *IGF2* intron region unambiguously demonstrates its A-485-induced downregulation. Note that at a higher scale, TPM 2200, overall IGF2 expression does not appear regulated by A-485. At a lower scale, TPM 50, EU-seq signal at IGF2 introns is abolished in A-485 treated SH-SY5Y, showing that the gene is indeed downregulated.

Extended Data Fig. 7



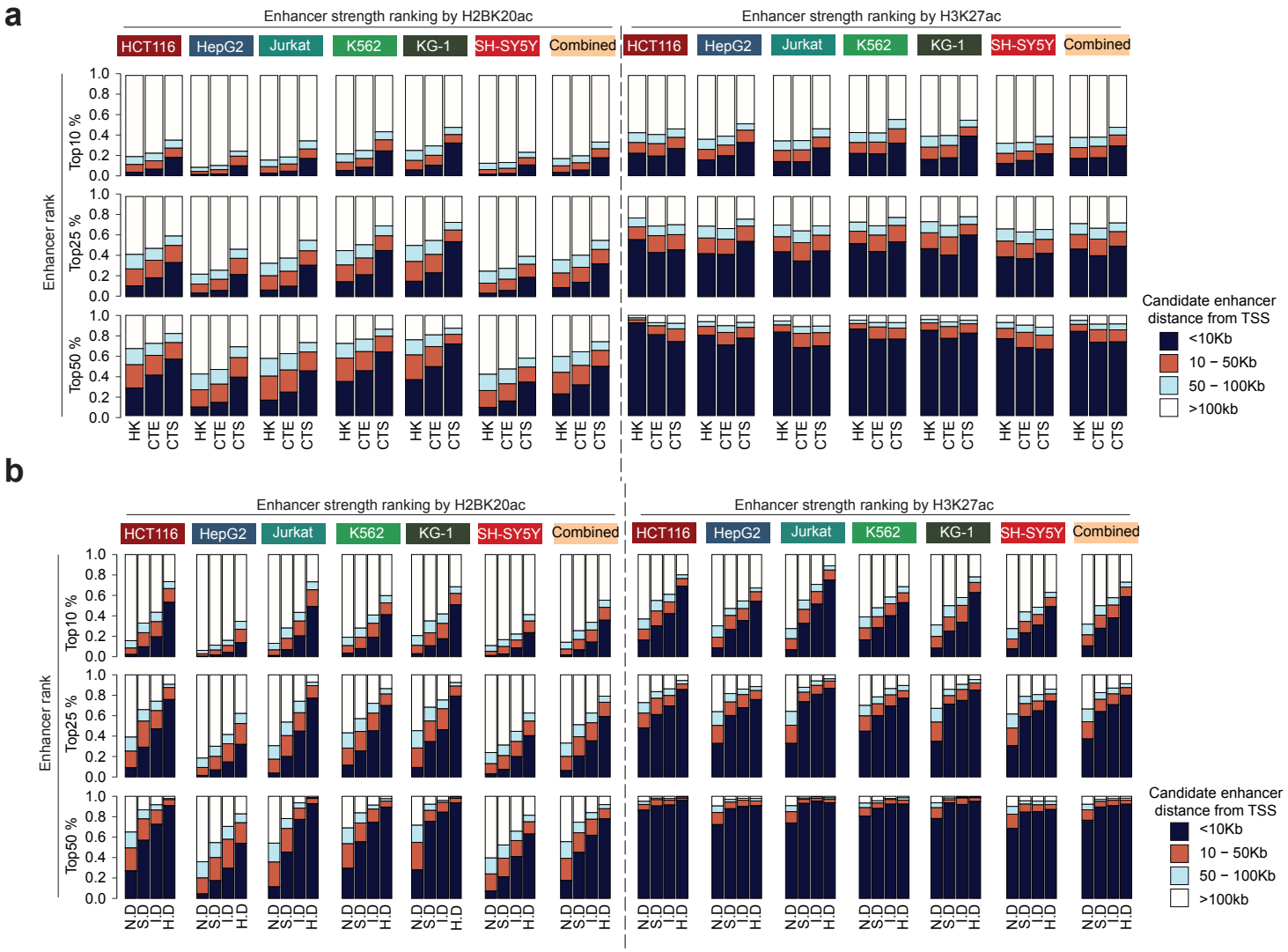
Extended Data Fig. 7. Cell-type-specific differences in regulation of genes with proximal enhancers. *IL32* harbors strong downstream enhancers in Jurkat and KG-1, as well as upstream enhancers in KG-1. However, *IL32* is more strongly expressed and downregulated by A-485 in KG-1 than in Jurkat. Similarly, *KCNQ* has proximal enhancer-like features in multiple cell lines, yet the gene is more strongly expressed and A-485 downregulated in HepG2. These examples indicate that *IL32* and *KCNQ* promoters are compatible with enhancers, and additional factors, such as 3D genome organization, presence of insulators, or accessibility of promoters, contribute to their cell-type-specific regulation by enhancers.

Extended Data Fig. 8



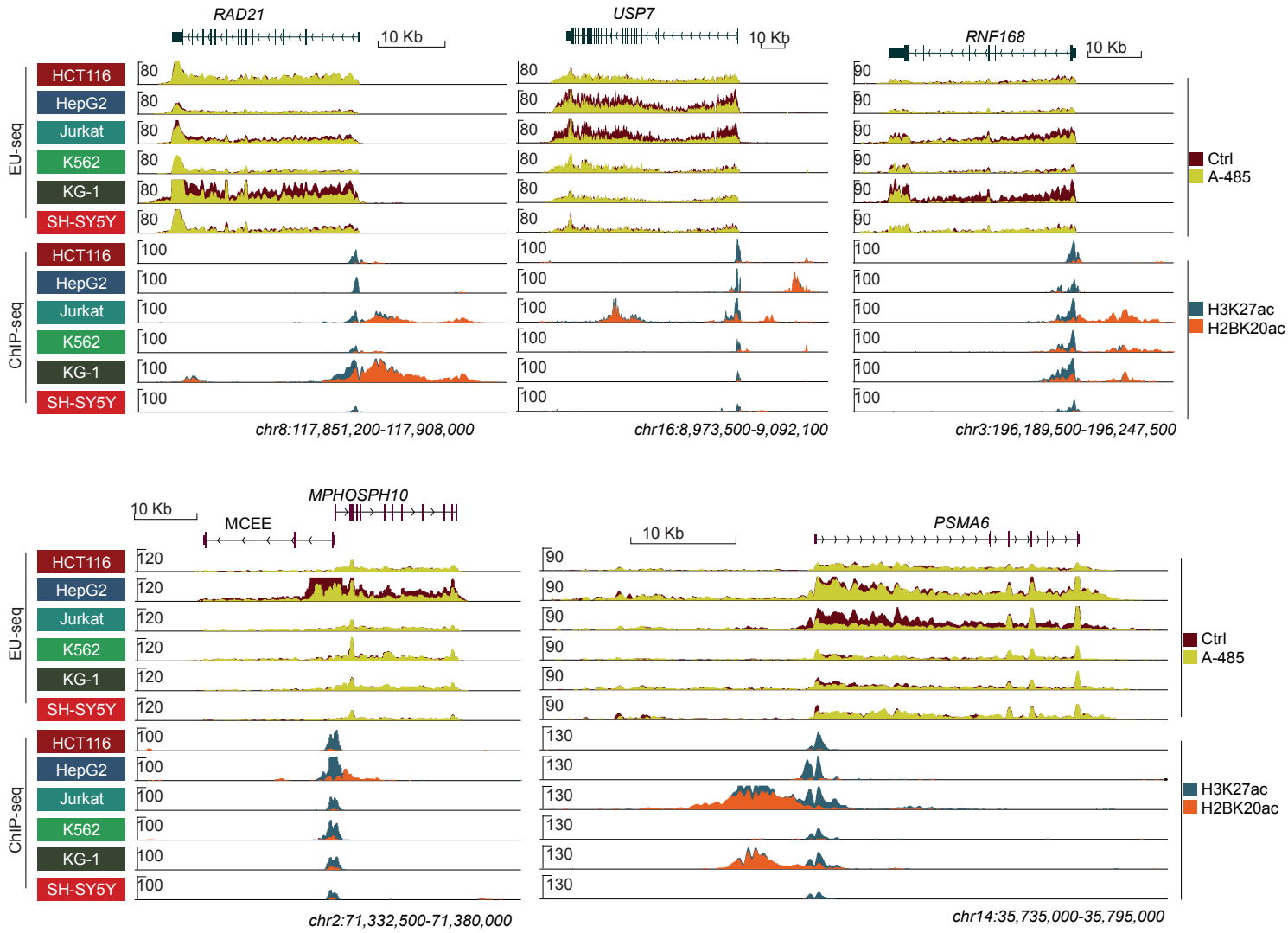
Extended Data Fig. 8. Expression of *Mcmbp, Ick, Fbox9, Sec23ip, Mthfsd,* and *Foxc2* in murine cell lines. Genome browser tracks show ubiquitous, enhancer-independent expression of *Mcmbp, Ick, Fbox9, Sec23ip, and Mthfsd,* in the indicated mouse cell lines. *Foxc2* is not expressed in any of the cell lines. *Foxc2* is marked with the repressive mark H3K27me3, but its promoter is accessible in ATAC-seq in mESC.

Extended Data Fig. 9



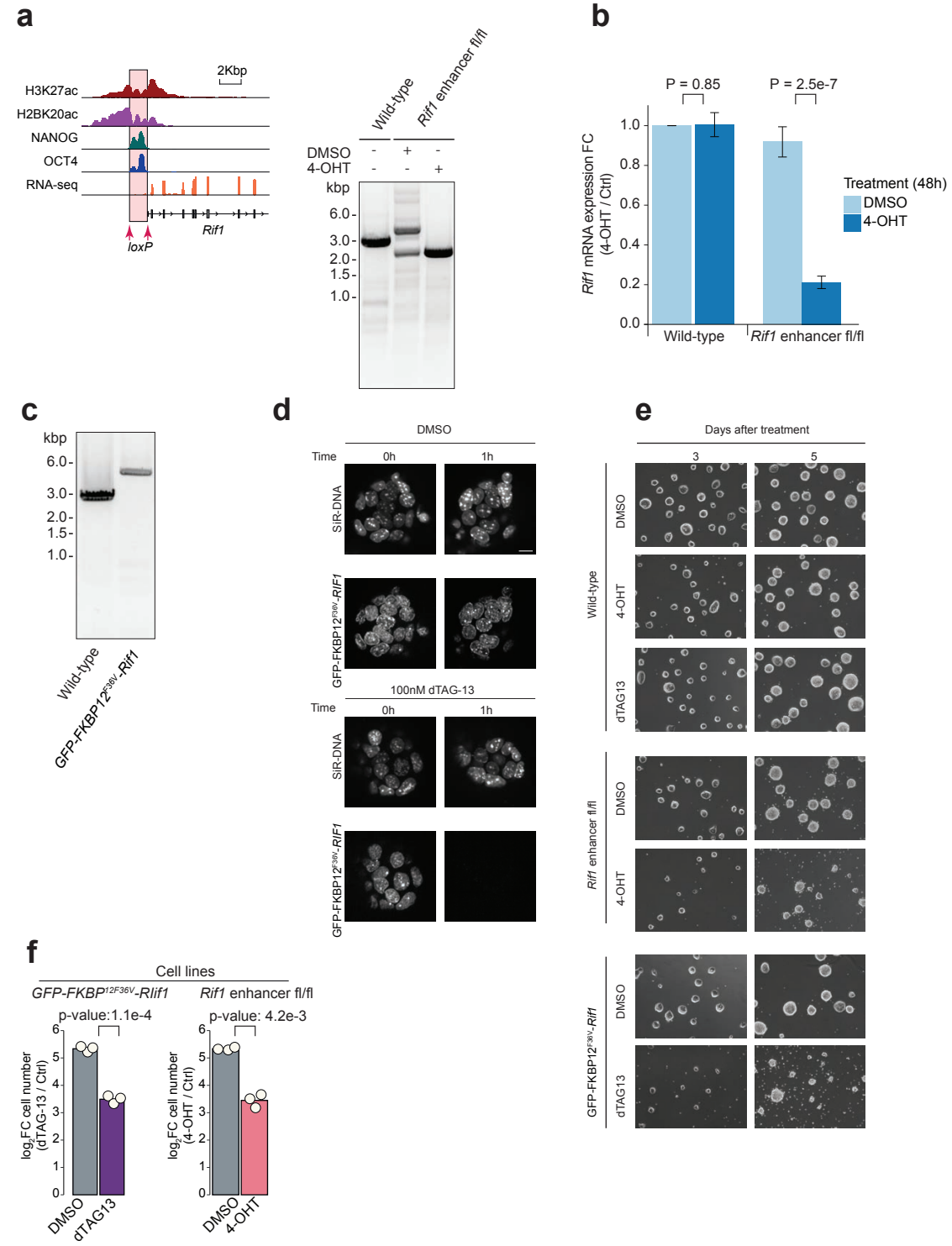
Extended Data Fig. 9. Enhancers are preferentially positioned in proximity to A-485 downregulated genes, enhancer-dependent gene regulation is associated with enhancer strength, distance, and cell-type specificity of genes. a, Genes expressed in the indicated human cell lines were classified based on their cell type-specificity (HK, CTE, CTS) as defined **Fig. 1a (a)** or A-485-induced downregulation (ND, SD, ID, HD) as defined in **Fig1b (b)**. Enhancer strength was determined using H2BK20ac (left panels) or H3K27ac (right panels) intensity, as defined in Fig. 2. Based on the acetylation mark intensity, the Top50% of enhancers were ranked grouped into Top10%, 25%, and 50% categories. For calculating H2BK20ac-based enhancer strength, candidate enhancer regions commonly marked with H3K27ac and H2BK20ac were used in these analyses. Genes were further grouped based on the distance between the enhancer and the nearest TSS. The bar charts show the relationship between cell-type-specificity, enhancer rank category, and enhancer-TSS distance. Note that, in comparison to HK genes, a greater fraction of CTS genes harbor strong enhancers in their proximity (<10kb). Notably, unlike H2BK20ac intensity-based enhancer ranking, H3K27ac intensity-based enhancer ranking shows a poor relationship with cell-type-specificity, enhancer strength, and distance. **b,** Relationship between H2BK20ac (left panels) and H3K27ac (right panels) intensity-based enhancer ranking and A-485-induced gene regulation. In the indicated human cell lines, genes were classified based on A-485-induced downregulation (ND, SD, ID, HD) as defined in **Fig. 1b**. Enhancer strength was defined by the H2BK20ac or H3K27ac ChIP signal, Top50% enhancers were rank-ordered, and grouped into the indicated percentile categories. Candidate enhancers commonly marked with H3K27ac and H2BK20ac were used in these analyses. Genes were further grouped based on the distance between the enhancer and the nearest TSS. Within different groups, the fraction of genes in A-485 regulated categories is shown. A much greater fraction of HD genes harbors top-ranked H2BK20ac marked enhancers than ND genes, explaining why HD genes are more prominently downregulated by A-485 than ND genes. Of note, the relationship between H3K27ac intensity-based enhancer ranking and gene regulation rapidly deteriorates and in the Top50% rank category, H3K27ac-based ranking provides minimal distinction among different A-485 regulated gene classes.

Extended Data Fig. 10



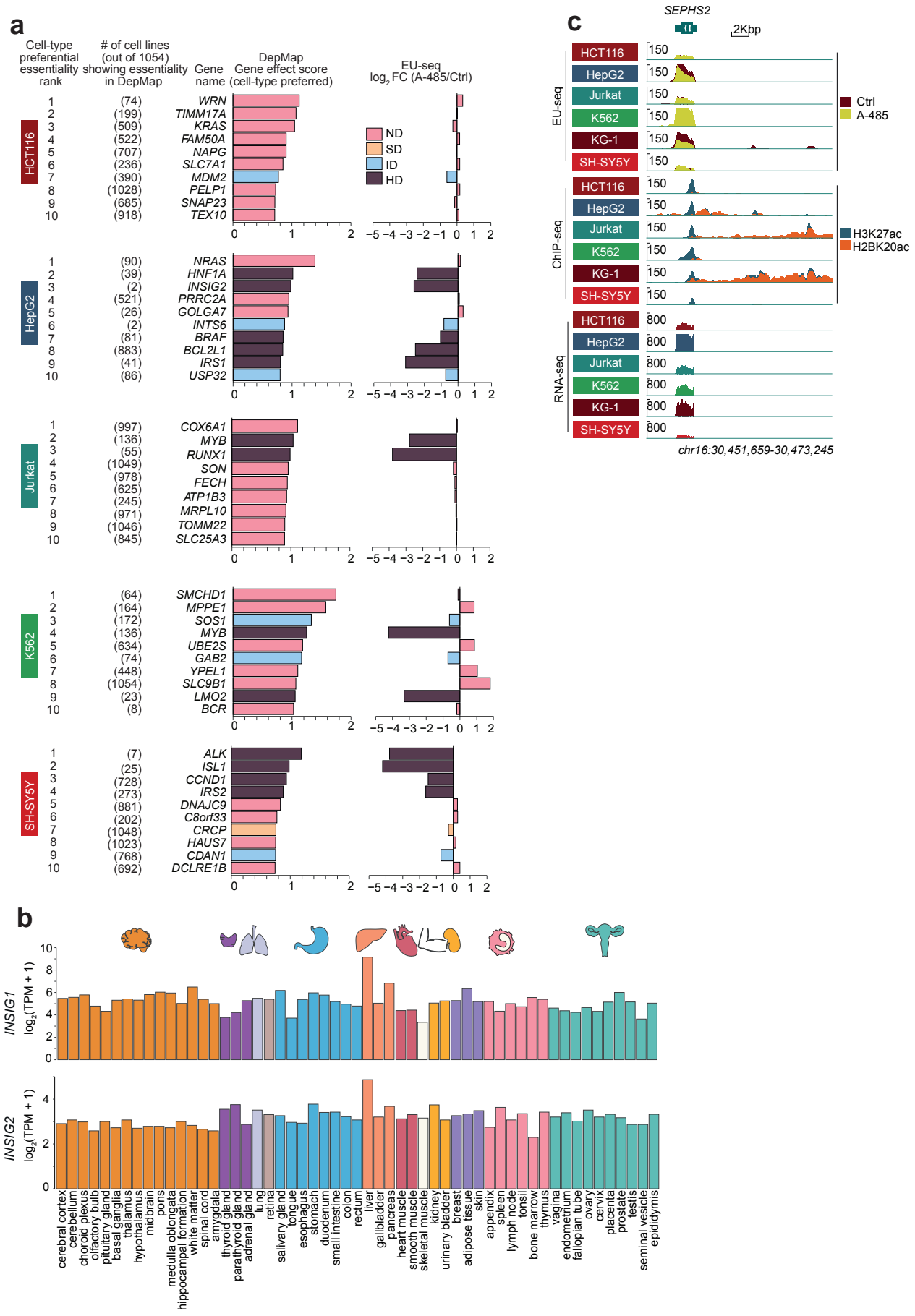
Extended Data Fig. 10. Enhancer-dependent quantitative boosting of housekeeping genes. Genome browser view of several housekeeping genes (*RAD21, USP7, RNF168, MPHOSPH10, and PSMA6*) showing cell-type-specific transcriptional boosting by enhancers. Note that cell line-specific enhancer-dependent boosting is associated with the occurrence of proximal enhancers marked with H2BK20ac.

Extended Data Fig. 11



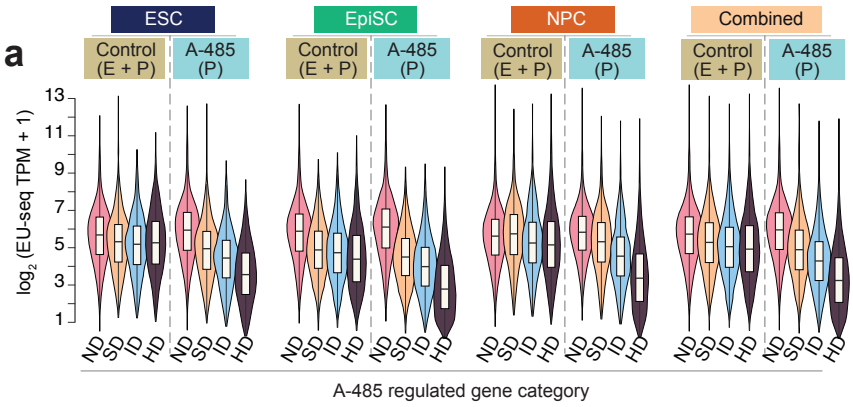
Extended Data Fig. 11. Enhancer-dependent RIF1 boosting is functionally relevant. a-b, Deletion of the upstream (-1kb) enhancer decreases *Rif1* expression in mESC. Using CRISPR, the *Rif1* upstream enhancer was flanked by loxP sites in mESC expressing Cre-ERT2 **(a)**. The enhancer was conditionally deleted by treating cells with 4-OHT (100nM, 48h) **(b)**. The bar charts show *Rif1* expression wild-type and Rif1 enhancer fl/fl cells, treated with DMSO or 4-OHT (100nM, 48h). **c-d,** Generation of GFP-FKBP12^{F36V}-RIF1 expressing mESC **(c)**, and validation of GFP-FKBP12^{F36V}-RIF1 degradation (dTAG-13, 100nM, 1h) **(d)**. **e-f,** Rif1 enhancer deletion impairs mESC proliferation. Rif1 enhancer fl/fl cells were treated with 4-OHT (100nM) to conditionally delete the enhancer, and GFP-FKBP12F36V-RIF1 cells were treated with dTAG13 (100nM) to conditionally deplete RIF1 protein. Cells were imaged 3 and 5 days after the treatment **(e)**, and the number of live and dead cells was counted on day 3 **(f)**. p-values: two-sided t-test.

Extended Data Fig. 12



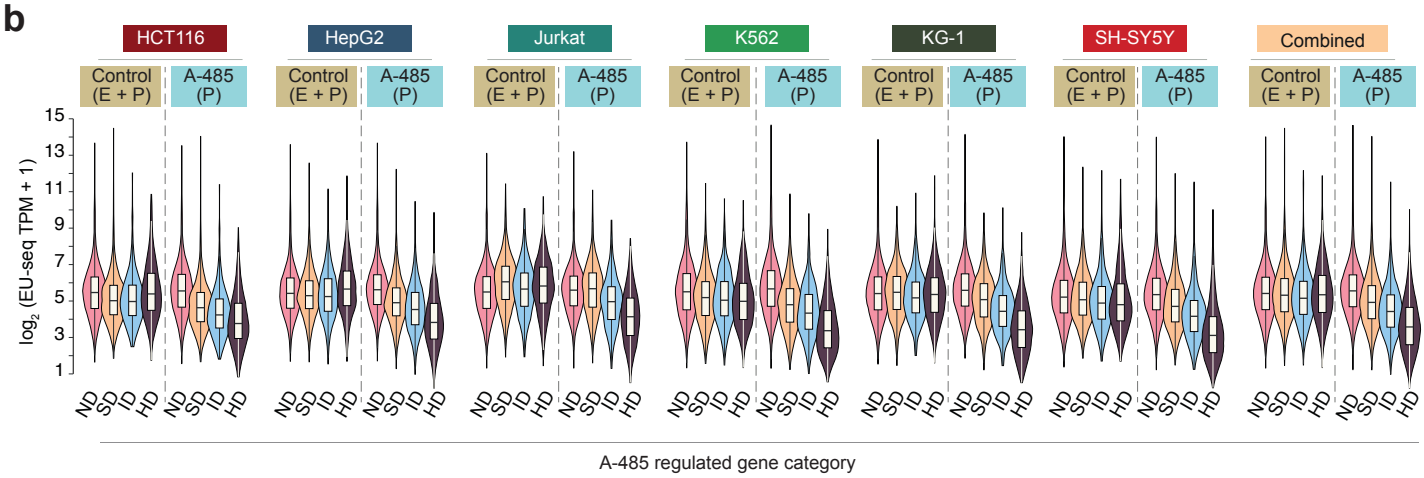
Extended Data Fig. 12. Expression profiles of *INSIG1* and *INSIG2,* and cell-type preferential essentiality of enhancer-regulated genes. a, Some enhancer-regulated genes score among the top cancer cell line preferential essential genes. The figure shows the Top10 cell-type preferential essential genes in the indicated cell cancer lines. For each gene, cell-type preferential essentiality rank, number of cell lines showing gene essentiality (out of 1054 cell lines), DepMap gene effect score, and A-485-induced nascent transcription downregulation in the respective cell lines, are indicated. Except for HCT116, all cell lines include 2-4 genes that are strongly enhancer-regulated and score among the Top10 essential genes in the respective cell lines. Most notably, *INSIG2* is essential in just 2 (out of 1054) cell lines. Cell-type preferential essentiality data were obtained from the DepMap CRISPR-based genome-wide essentiality database[45]. **b,** Shown are mRNA expression levels of INSIG1 and *INSIG2* across the indicated human tissues. Gene expression data was obtained from the human protein atlas[36]. **c,** Genome browser tracks showing the expression and A-485-induced regulation of *SEPHS2* in the indicated cell lines.

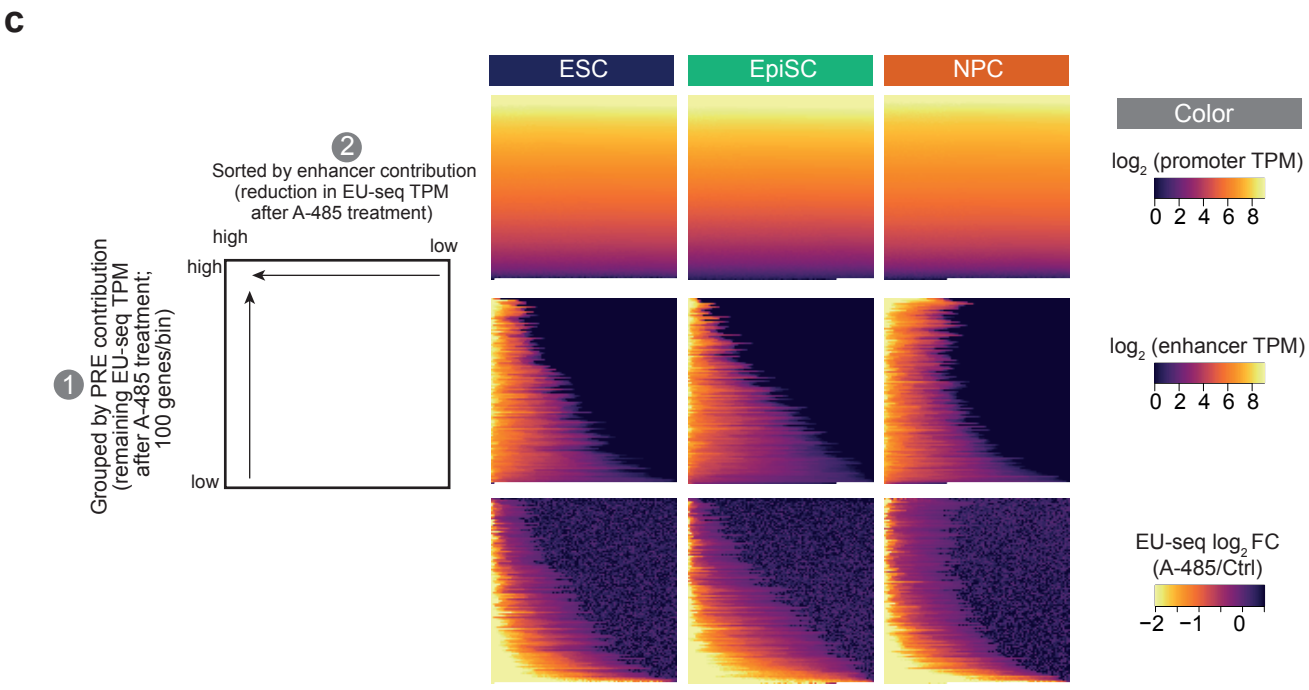
Extended Data Fig. 13



Control (E + P):
Enhancer + PRE-driven transcription output
(EU-seq TPM in control cells without A-485 treatment)

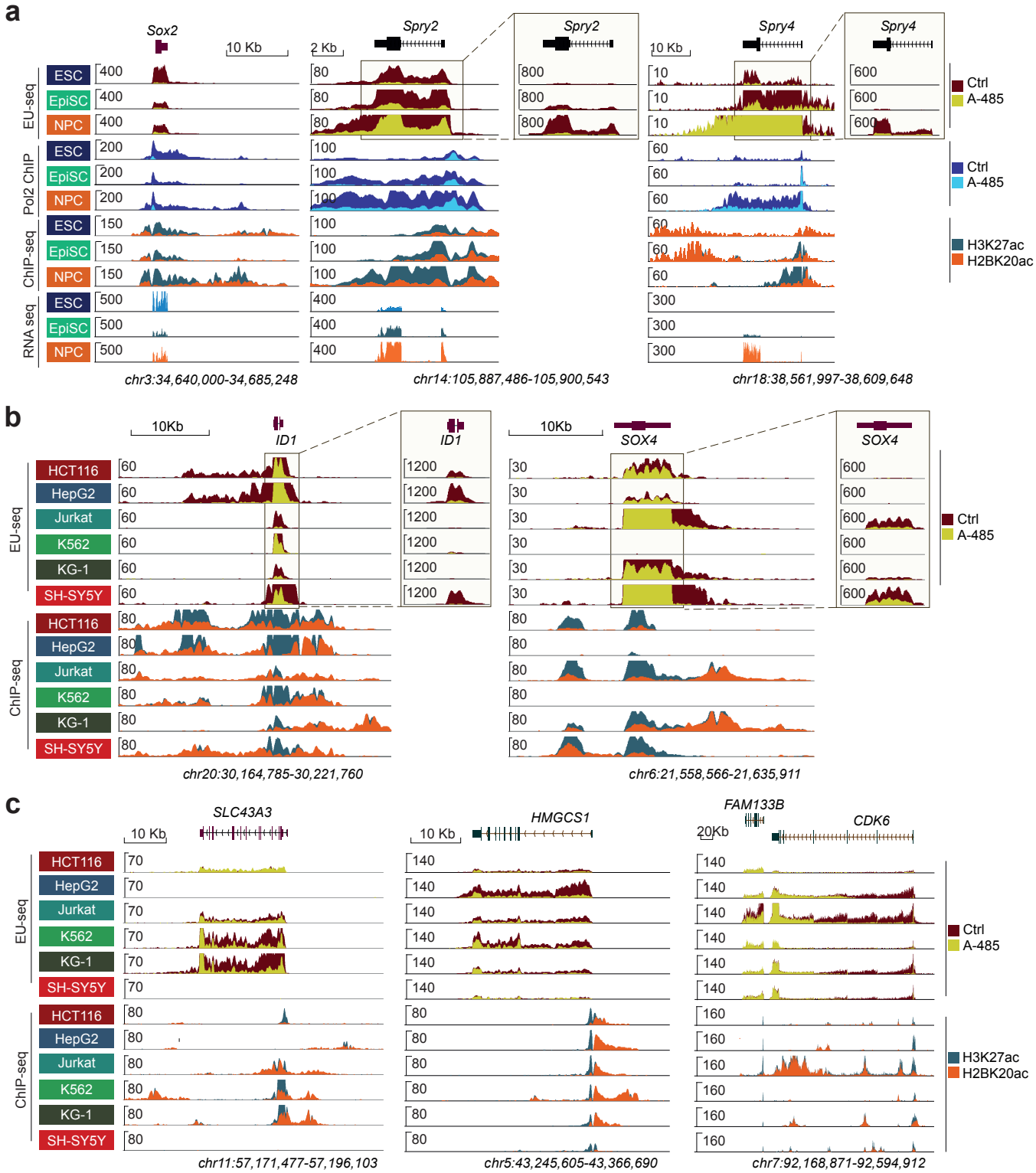
A-485 (P):
PRE-driven transcription output
(Remaining EU-seq TPM in cells treated with A-485 (1h))





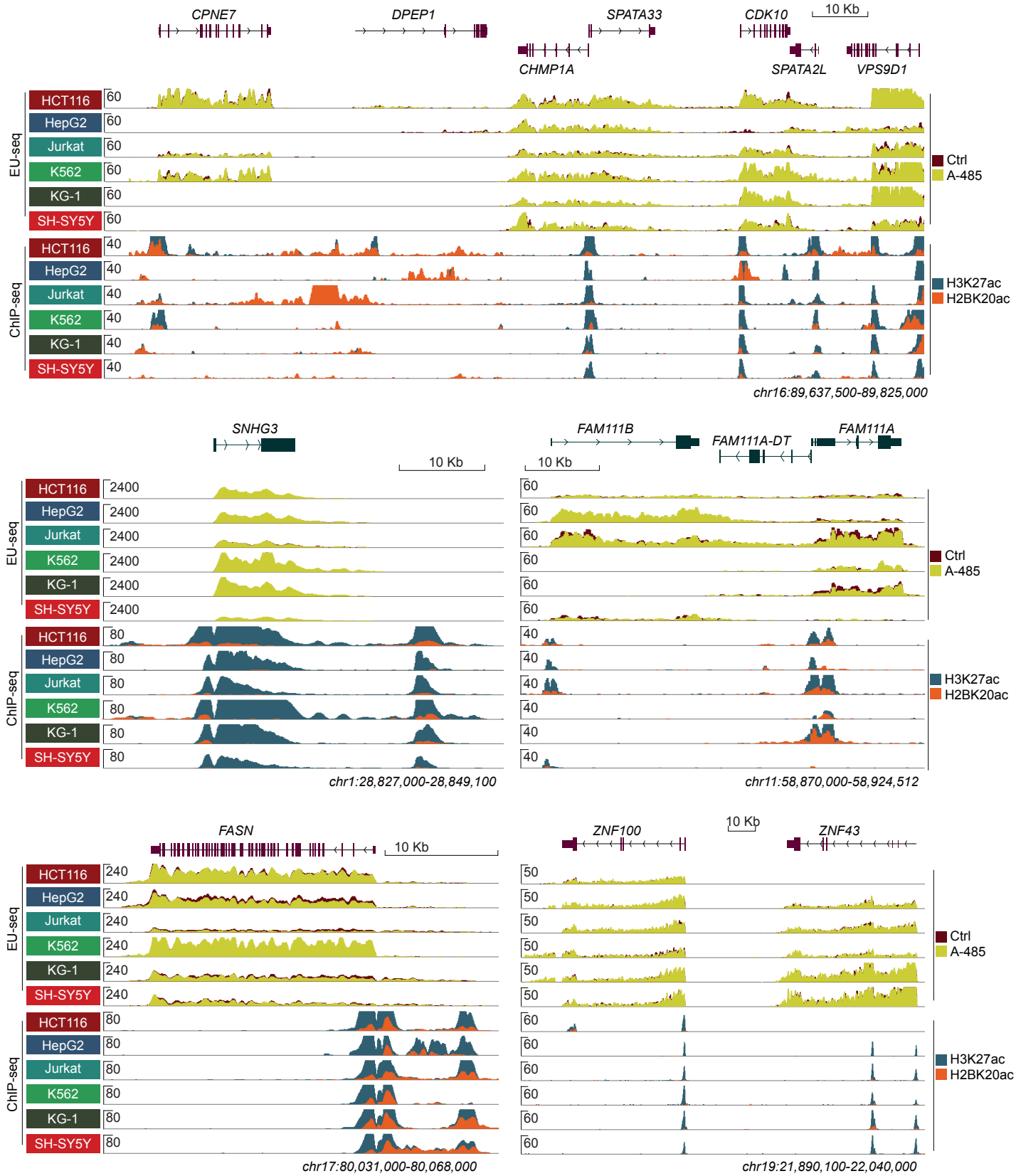
Extended Data Fig. 13. Enhancers and PREs, individually or in combination, drive gene expression over a large dynamic range. a-b, Nascent transcription levels in mouse **(a)** and human **(b)** cell lines. Shown are EU-seq TPM values for the indicated A-485 regulated gene class (ND, SD, ID, and HD), in the specified cell lines. In the untreated (Control) condition, gene expression is driven by the combined strength of enhancer (E) and PRE (P). After A-485 treatment, the contribution of enhancers to gene activation is ablated, and the remaining nascent RNA expression levels indicate promoter output sustained by PRE activity. In the control condition, gene expression levels in ND, SD, ID, and HD categories are roughly similar, albeit with some variability across cell lines. In each category, gene expression values vary by more than 3 orders of magnitude. After A-485 treatment, SD and ID genes are partially downregulated, and their expression is sustained at a much higher level than that of HD genes. Because CBP/p300 activity is globally inhibited, the remaining gene expression of SD and ID genes is sustained enhancer-independently. This shows that a large portion of genes is only partially activated by enhancers, and upon CBP/p300 inhibition, their remaining expression level is sustained by PRE activity. **c,** Absolute and relative contribution of enhancers differs for different genes. Gene expression in ESC, EpiSC, and NPC was quantified using nascent transcription analyses (EU-seq, TPM). Contribution of enhancer and PRE to gene expression was determined as described above. Genes were ranked from lowest to highest expression based on the PRE-driven gene expression level (i.e. remaining gene expression level after A-485 treatment), and binned (100 genes in each raw). Within each bin, genes were sorted based on the contribution by enhancers. Contribution of enhancer to gene expression is determined by the reduction in EU-seq TPM after A-485 treatment. Note that genes with relatively low contribution of PREs, show greater fold-change after A-485 treatment, indicating a relatively greater contribution of enhancers to genes with weaker PREs.

Extended Data Fig. 14



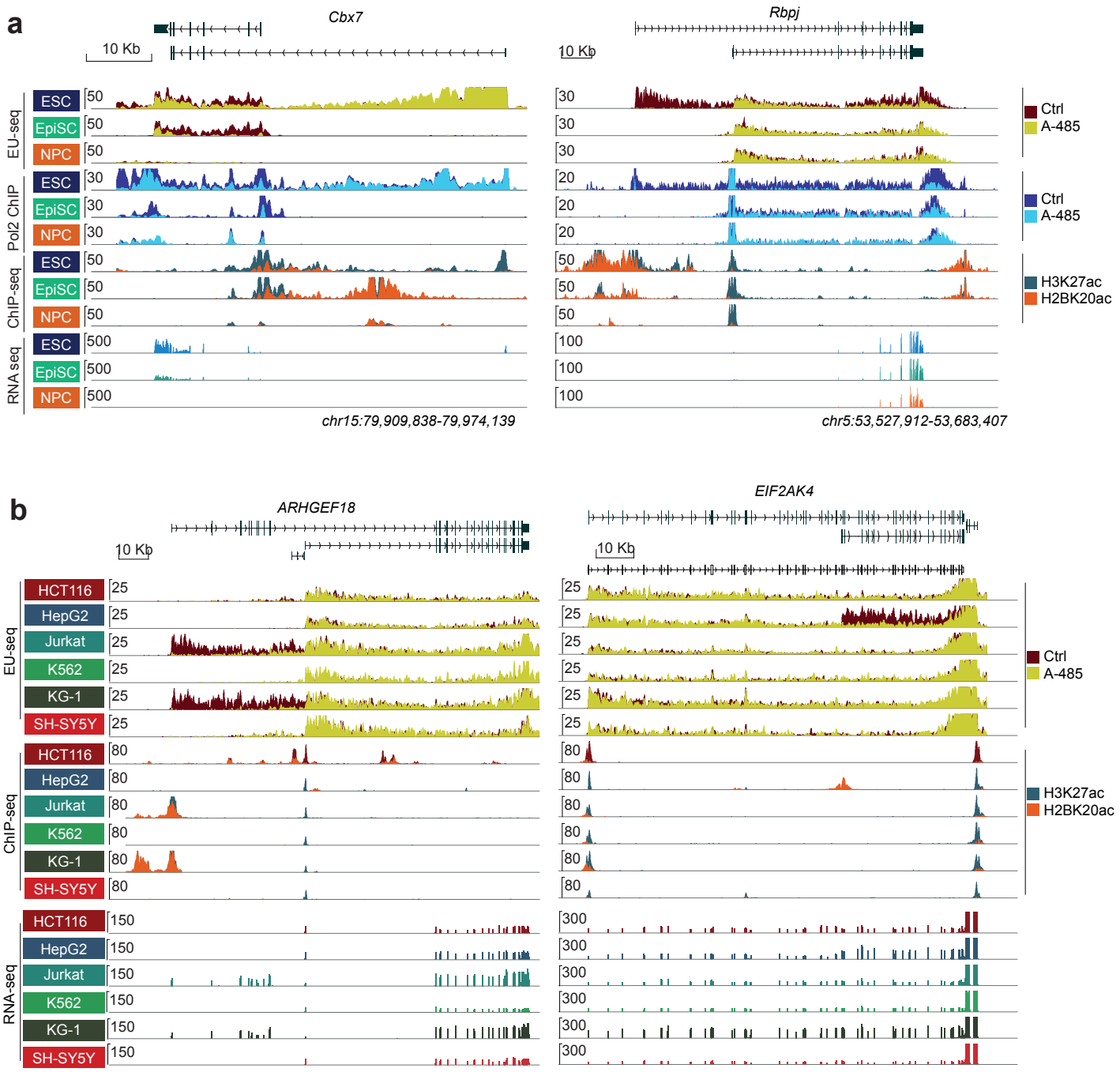
Extended Data Fig. 14. Enhancer strength is variable across different cell types. a-c, Shown are tracks of representative genes exhibiting very high variation in enhancer-dependent gene expression in mouse **(a)** and human cells **(b-c).** Note that genes are highly downregulated in most of the cell lines in which they are expressed, yet their expression levels vary greatly across the cell lines, demonstrating that gene-specific enhancer strength is highly variable across cell lines. Because of a very high dynamic range of expression, *Spry2, Spry4, ID1,* and *SOX4* expression is shown at two different scales. *FAM113B,* located downstream of *CDK6,* is only boosted by an enhancer in Jurkat, likely through the upstream enhancer located in this cell line.

Extended Data Fig. 15



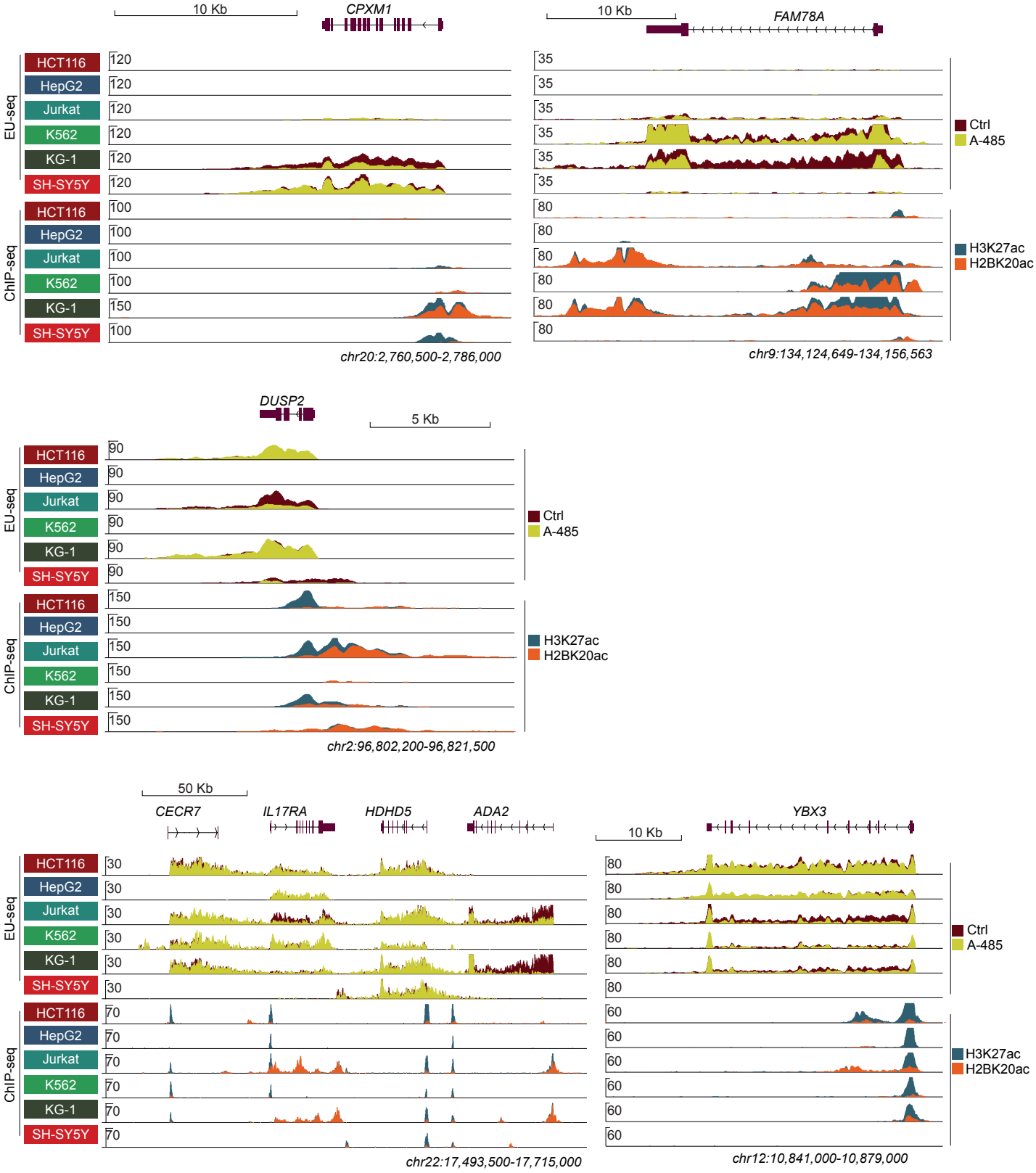
Extended Data Fig. 15. Cell-type-specific quantitative gene expression variation through alteration in PRE strength. Genome browser view of representative human genes showing cell-type-specific quantitative expression variation. Note that genes are expressed to variable levels in diverse cell lines, but they show no or only minor downregulation by A-485, demonstrating that their expression is quantitatively adjusted enhancer independently, likely through alteration of PRE strength.

Extended Data Fig. 16



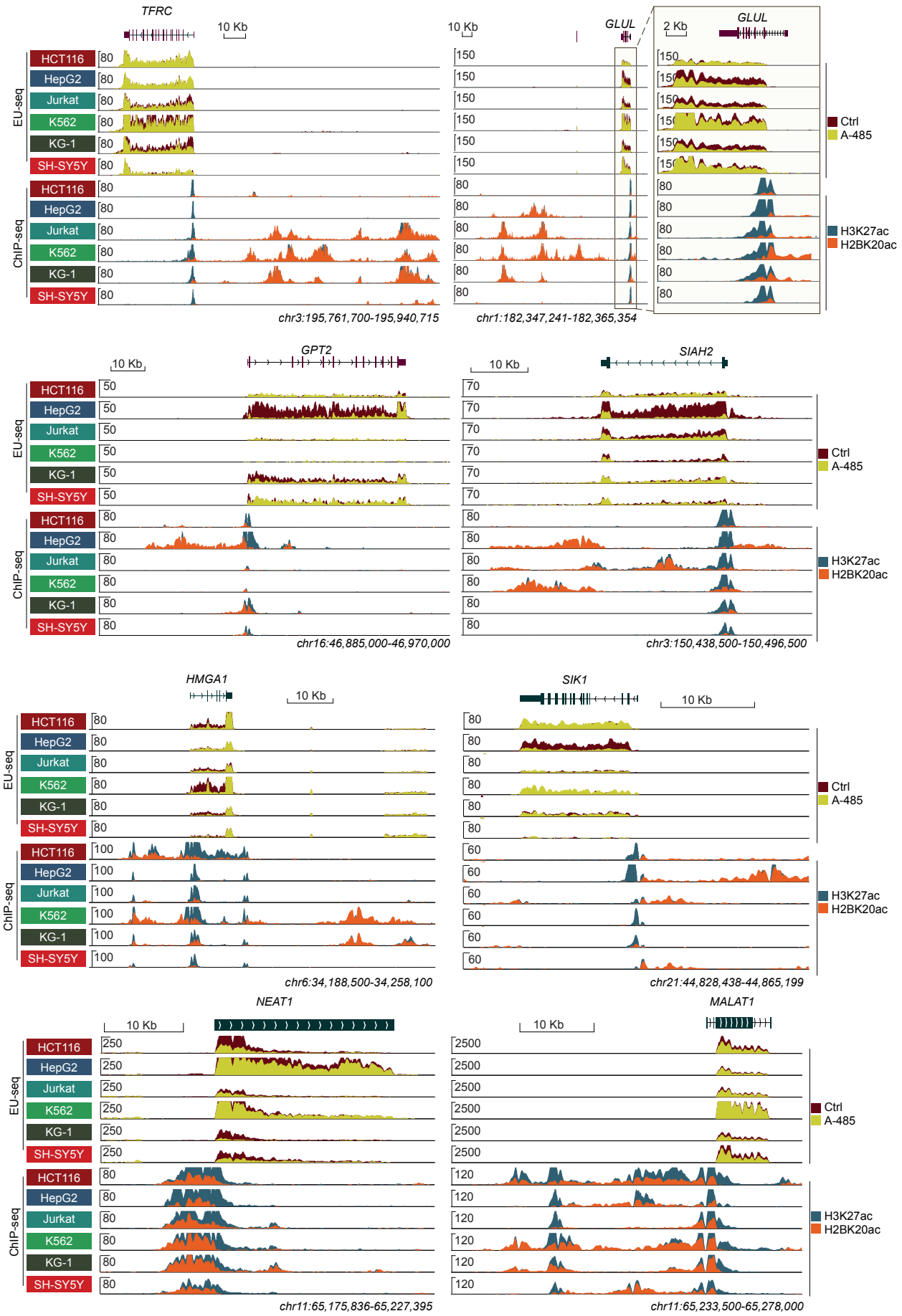
Extended Data Fig. 16. Enhancers and PREs differentially activate alternative promoters of the same genes. a, Genome browser view of representative human genes showing differential activation of alternative promoters by enhancers and PREs. *ARHGEF18* and *EIF2AK4* are expressed in all cell lines, but each of them expresses cell-type-specific isoforms. The promoter of the longer isoform of *ARHGEF18* is activated by an enhancer, whereas the promoter of the shorter isoform is activated by a PRE. Conversely, the promoter of the longer isoform of *EIF2AK4* is activated by PRE, whereas the shorter isoform is activated by an enhancer in HepG2. **b,** Differential activation of alternative promoters of the same genes in mouse cell lines. In ESC, *Cbx7* is expressed from two alternative promoters, whereas only the shorter isoform is expressed in EpiSC. The ESC-specific longer isoform of *Cbx7* is transcribed enhancer-independently by PRE, whereas the shorter isoform is activated by enhancers in both the cell lines. Of note, the remaining *Cbx7* expression in A-485 treated ESC likely reflects the EU-seq signal from the longer isoform of *Cbx7*. *Rbpj* is also expressed by two alternative promoters in ESC, but EpiSC and NPC only express the shorter isoform. While the promoter of the ESC-specific longer isoform is activated by an enhancer, the promoter of the shorter isoform is activated by PRE in all cell lines.

Extended Data Fig. 17



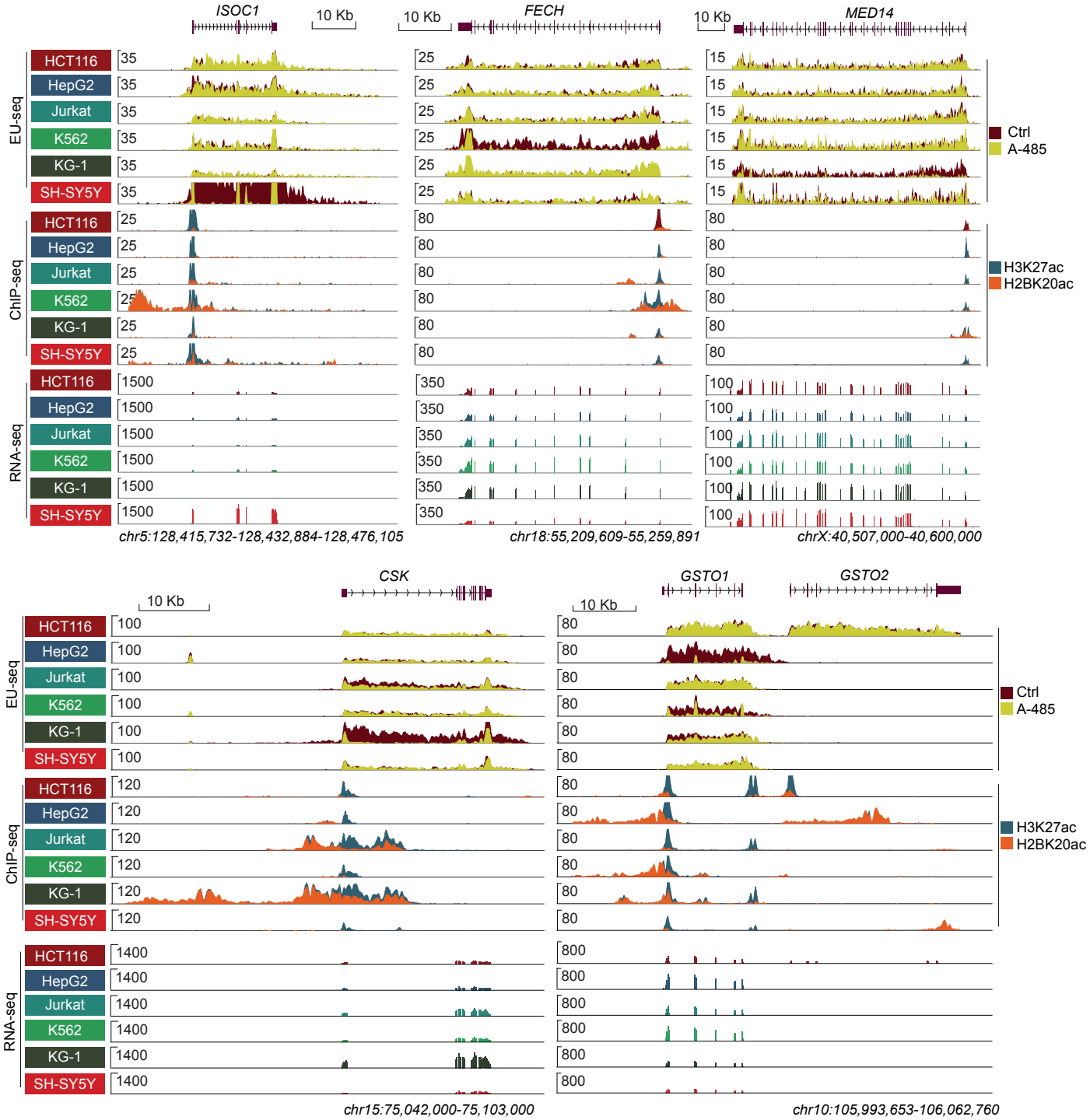
Extended Data Fig. 17. Combined strengths of enhancers and PREs afford quantitative expression variation of cell-type-specific genes. Genome browser view of representative human cell-type-specific genes showing quantitative expression variation through cell-type-specific tuning of the strength of enhancers, PREs, or both. Because of the variable strength of PREs and enhancers, expression of the genes is downregulated to a variable extent by A-485. There is a notable agreement between the occurrence of proximal H2BK20ac marked enhancers and A-485 sensitivity.

Extended Data Fig. 18



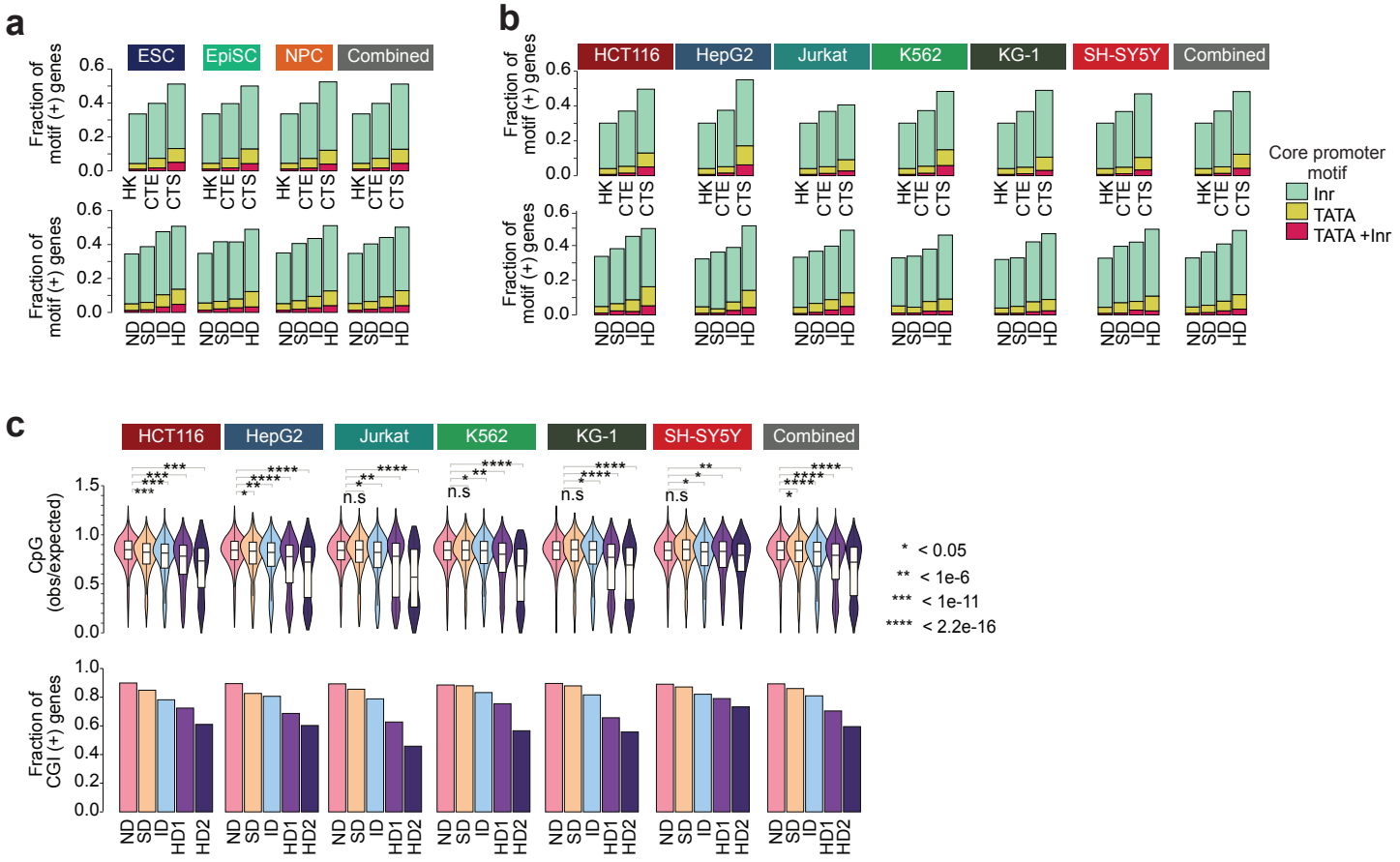
Extended Data Fig. 18. Tunable and integrated strengths of enhancers and PREs afford cell-type-specific quantitative expression variation of widely expressed genes. Genome browser view of representative human genes that are expressed in all the cell lines analyzed, yet show large quantitative expression variation. The expression of these genes is apparently altered through cell-type-specific tuning of the strength of enhancers, PREs, or both. The top and middle panels show genes (*TFRC, GLUL, GPT2, SIHA2, HMGA1, SIK1*) that are expressed at relatively low-to-modest levels. Their cell-type-specific quantitative variation and cell-type-specific regulation by enhancers are apparent in the reduction of nascent transcription in the A-485 treated condition. The bottom panel shows genes (*MALAT1*, and *NEAT1*) that are expressed at a high level. Enhancers also appear to contribute to their expression, but depending on the strength of PREs in individual cell lines, the measurable impact of enhancers is highly variable, and in some cell lines it is barely noticeable. This shows that in genes that harbor very strong PREs, the measurable relative impact of weak and modest enhancers can be extremely small.

Extended Data Fig. 19



Extended Data Fig. 19. PREs and enhancers can interchangeably activate the same genes in different cell types. Genome browser tracks showing cell-type-specific, promoter-to-enhancer type gene activation switching. A-485 treatment abrogates *ISOC1* expression in SH-SY5Y cells, yet has no appreciable impact on any of the other cell lines. Expression of *FECH* and *MED14* is almost abrogated after CBP/p300 inhibition in K562 and KG-1, respectively, yet in the other cell lines, the genes remain unaffected by the inhibitor. Notably, the expression of *FECH* and *MED14* is similar in all cell lines, but the type of regulation differs among the cell lines. *CSK* is partially downregulated by A-485 in Jurkat, its expression is ablated in KG-1, but the expression is unaffected in the other four cell lines. Similarly, *GSTO1* expression is abrogated by A-485 in HepG2, remains unaffected in HCT116, K562, and SH-SY5Y, and is partially downregulated in KG-1 and Jurkat. Of note, the proximally occurring gene, *GSTO2*, is cell-type-specifically expressed, yet not regulated by A-485. Remarkably, all genes showing PRE-to-enhancer switching are marked with the strongest H2BK20ac signal, suggesting that gene regulation switching occurs as a consequence of the cell-type-specific acquisition of enhancers.

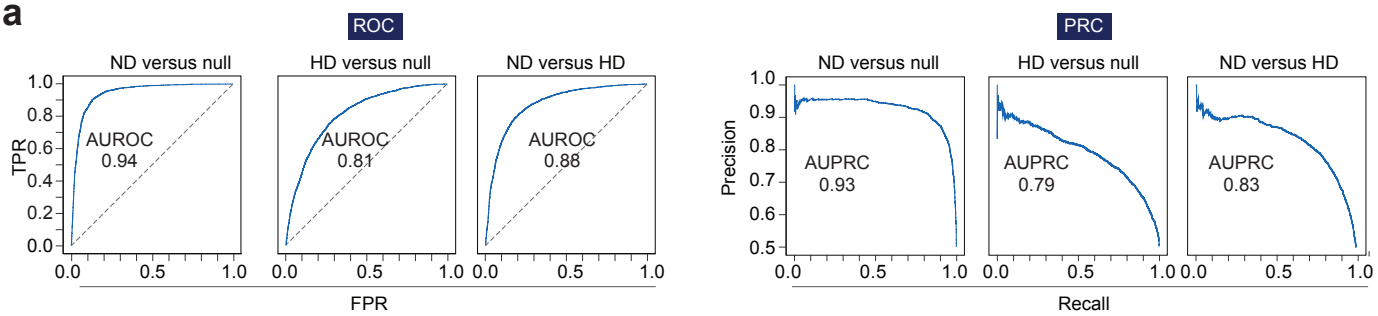
Extended Data Fig. 20



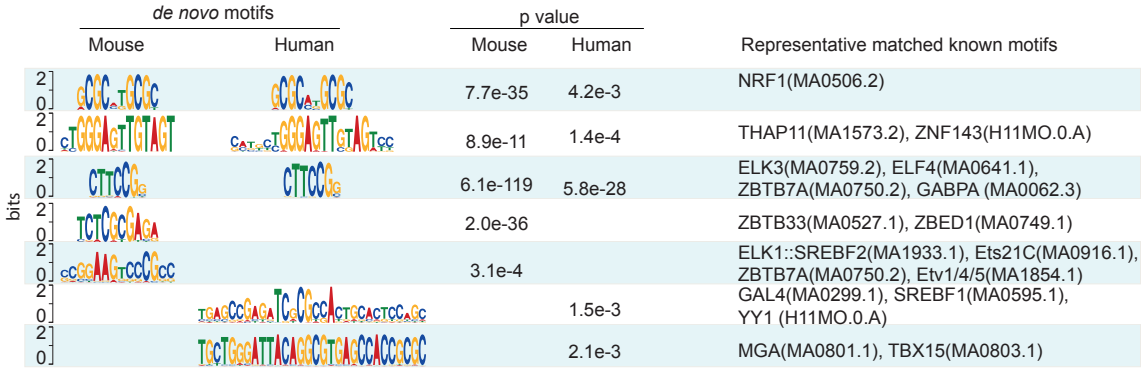
Extended Data Fig. 20. A greater fraction of enhancer-dependent promoters harbor TATA box, and autonomous promoters contain higher CGI density. a-b, Fraction of promoters containing the indicated core promoter motifs, in the specified cell lines. Genes were grouped based on cell-type-specificity **(a)** or A-485 downregulated gene **(b)** categories, and the fraction of genes harboring the indicated core promoter motifs are shown. Across different cell-type-specificity and A-485 regulated gene categories, the fraction of Inr motif-bearing promoters is relatively similar. The fraction of genes with TATA-box increases with increasing cell-type-specificity (HK<CTE<CTS) and with increasing A-485-induced gene down-regulation (ND<SD<ID<HD). However, overall, in all categories, only a small fraction of genes harbor a TATA box. c, Shown is the observed to expected (O/E) ratio of CpG dinucleotides (top panel) and the fraction of CpG island (CGI) positive promoters (bottom panel) and in the indicated A-485 regulated gene categories and specified cell lines. CpG O/E ratio was calculated using the sequence of ±250bp from TSS. CGI positive genes are defined as GC content > 50 % and CpG O/E ratio > 0.6 in the ±250bp regions from TSS. * $p < 0.05$, ** $p < 1e-6$, *** $p < 1e-11$, **** $p < 2.2e-16$, Mann–Whitney U test.

Extended Data Fig. 21

a



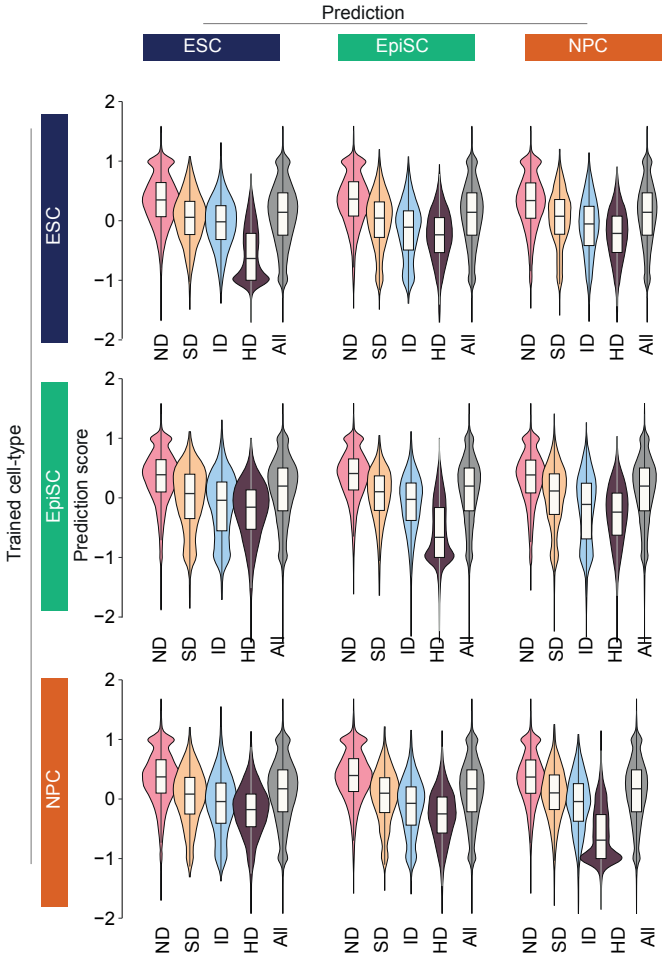
b



de novo motifs		p value		
Mouse	Human	Mouse	Human	Representative matched known motifs
		7.7e-35	4.2e-3	NRF1(MA0506.2)
		8.9e-11	1.4e-4	THAP11(MA1573.2), ZNF143(H11MO.0.A)
		6.1e-119	5.8e-28	ELK3(MA0759.2), ELF4(MA0641.1), ZBTB7A(MA0750.2), GABPA (MA0062.3)
		2.0e-36		ZBTB33(MA0527.1), ZBED1(MA0749.1)
		3.1e-4		ELK1::SREBF2(MA1933.1), Ets21C(MA0916.1), ZBTB7A(MA0750.2), Etv1/4/5(MA1854.1)
			1.5e-3	GAL4(MA0299.1), SREBF1(MA0595.1), YY1 (H11MO.0.A)
			2.1e-3	MGA(MA0801.1), TBX15(MA0803.1)

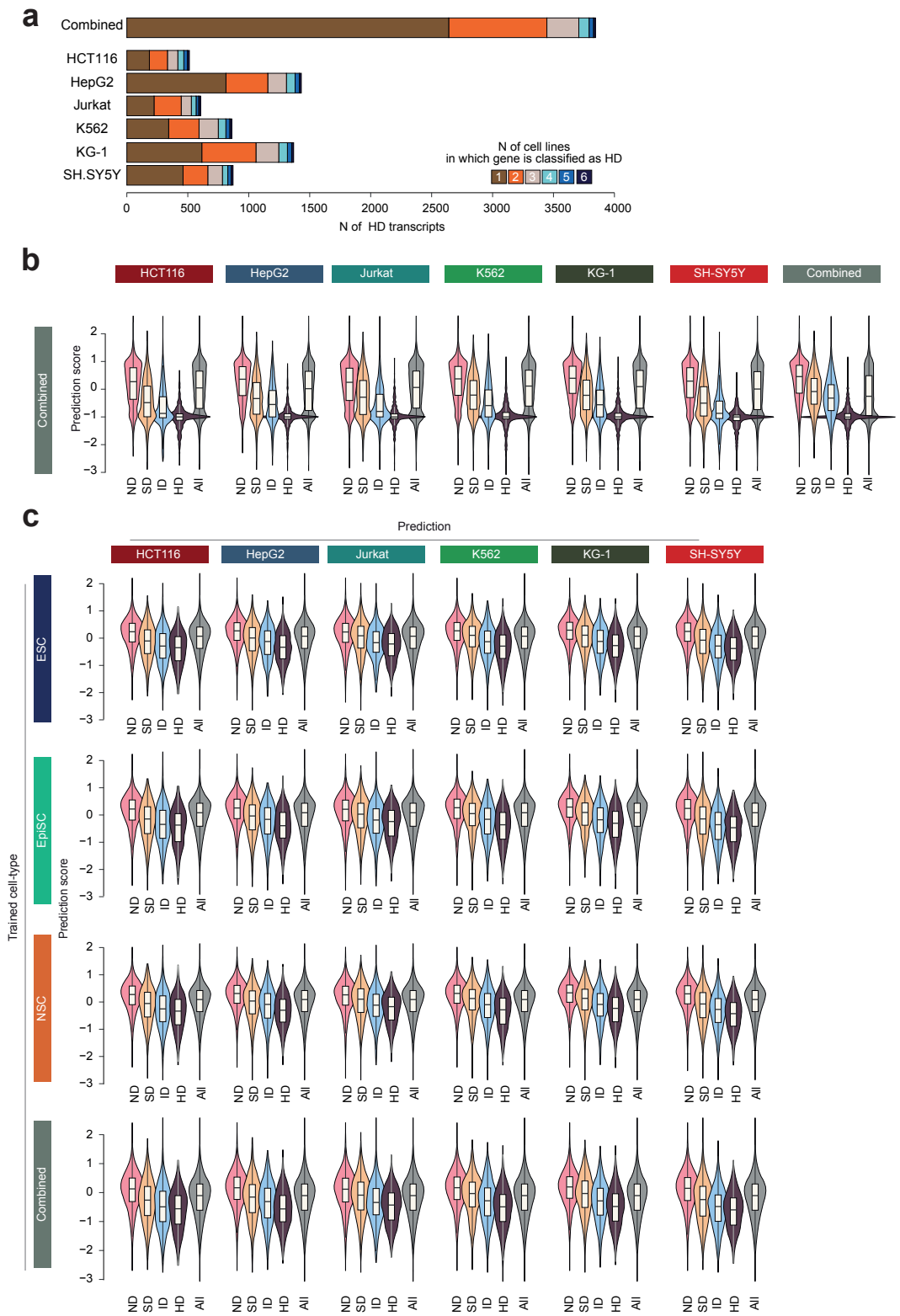
Extended Data Fig. 21. PRE sequences discriminate between autonomous and enhancer-dependent promoters. a, Gapped k-mer support vector machine (SVM)-based kernels (gkm-SVM) can discriminate ND and HD promoters from each other, and null DNA sequences. The receiver operating characteristic curve (ROC, top panel) and precision-recall curve (PRC, bottom panel) show the accuracy of the gkm-SVM classifier for discriminating ND and HD promoters from null sequences, as well as for discriminating ND promoters from HD promoters. Notably, ND promoters are discriminated from null sequences with greater accuracy than HD promoters. ND promoters can be discriminated from HD promoters with similar accuracy as for HD promoters to null sequences. gkm-SVM kernels were trained using DNA sequences, upstream 500bp from TSS of ND (n= 3258) and HD (n= 3154) gene promoters from 6 human cell lines, or randomly selected repeat- and GC-matching sequences from the same genome (see methods for details). AUROC and AUPRC are shown. **b,** Shown are the de novo sequence motifs represented in the importance scores calculated by gkm-SVM discriminating ND promoters from HD promoters. Representative transcription factors with matching DNA binding motifs are shown. The enrichment of each motif in ND promoter over HD promoter (upstream 500bp from TSS) is tested using a two-sided fisher test and the calculated p values are shown.

Extended Data Fig. 22



Extended Data Fig. 22. Promoter sequence difference discriminate enhancer-dependence of genes. The classifier was trained using ND and HD genes exclusively downregulated in one of the indicated cell lines and performance was tested on the same cell line, as well as in the other two held-out cell lines.

Extended Data Fig. 23



Extended Data Fig. 23. Mouse and human ND and HD promoters share sequence features.
a, The overlap of human HD genes between different cell-types. The bar chart shows the number of HD transcripts on each human cell lines and the number of overlapping with other cell lines are shown with individual colors. **b-c,** A gkm SVM model was trained using ND and HD genes as positive and negative sets, respectively. For training the model in human cells **(b)**, ND and HD genes combined from all the human cell lines were used. Separately, a model was trained using ND and HD genes from one of the indicated mouse cell lines **(c)**. For both the models, trained in human or mouse cell lines, performance was tested on each of the indicated human cell lines.