# Large-scale analysis of the integration of enhancer-enhancer signals by promoters

Miguel Martinez-Ara[1,2], Federico Comoglio[1,4], Bas van Steensel[1,2,3,*]

[1]Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, the Netherlands
[2]Oncode Institute, the Netherlands
[3]Division of Molecular Genetics, Netherlands Cancer Institute, Amsterdam, the Netherlands
[4]Present address: enGene Statistics GmbH, Basel, Switzerland

*correspondence: b.v.steensel@nki.nl

## ABSTRACT

Genes are often regulated by multiple enhancers. It is poorly understood how the individual enhancer activities are combined to control promoter activity. Anecdotal evidence has shown that enhancers can combine sub-additively, additively, synergistically, or redundantly. However, it is not clear which of these modes are more frequent in mammalian genomes. Here, we systematically tested how pairs of enhancers activate promoters using a three-way combinatorial reporter assay in mouse cells. By assaying about 69,000 enhancer-enhancer-promoter combinations we found that enhancer pairs generally combine near-additively. This behaviour was conserved across seven developmental promoters tested. Surprisingly, these promoters scale the enhancer signals in a non-linear manner that depends on promoter strength. A housekeeping promoter showed an overall different response to enhancer pairs, and a smaller dynamic range. Thus, our data indicate that enhancers mostly act additively, but promoters transform their collective effect non-linearly.

KEYWORDS:
Gene regulation; enhancers; promoters; combinatorial logic; massively parallel reporter assay; additivity; synergy; non-linear.

## INTRODUCTION

*Enhancers: background*. Enhancers are DNA elements, typically a few hundred base pairs long, that bind transcription factors (TFs) and boost the activity of genes *in cis* (Banerji et al., 1981). Over the past decades, many insights have been obtained into the molecular architecture of enhancers, the roles of TFs and their cofactors, the interplay of these factors with chromatin, and the diversity of mechanisms that govern the interactions of enhancers with promoters (S. Kim & Wysocka, 2023).

*MPRAs to probe enhancers.* In recent years, massively parallel reporter assays (MPRAs) have been developed to systematically dissect the sequence features that determine the functionality of regulatory elements by probing thousands or even millions of DNA elements in a multiplexed manner (Inoue & Ahituv, 2015; Klein et al., 2020; Gallego Romero & Lea, 2023). Among others, two-way combinatorial MPRA studies that tested a large number of enhancer - promoter pairs revealed a variable degree of intrinsic compatibility between enhancers and promoters (Zabidi et al., 2015; Bergman et al., 2022; Martinez-Ara et al., 2022), in line with earlier evidence from analyses of single genomic loci (Bertolino & Singh, 2002; Chang et al., 2004; Vakoc et al., 2005; Jing et al., 2008; Deng et al., 2012).

*Modes of enhancer interplay are still poorly understood.* Surveys of mammalian genomes suggest that there are many more enhancers than genes (ENCODE Project Consortium, 2012). Indeed, genes are frequently controlled by multiple enhancers (Shlyueva et al., 2014; Gasperini et al., 2020; S. Kim & Wysocka, 2023). Often, such enhancers are clustered and jointly regulate a nearby tissue-specific gene (Tuan et al., 1985; Grosveld et al., 1987; Tolhuis et al., 2002; Deng et al., 2012; Hnisz et al., 2013; Whyte et al., 2013; Hay et al., 2016; Schoenfelder & Fraser, 2019; Brosh et al., 2023). However, how multiple enhancers act together to control a single promoter is still poorly understood. Several studies have found that enhancers can combine either additively (Bothma et al., 2015; Lam et al., 2015; Dukler et al., 2016; Hay et al., 2016; Will et al., 2017; Lin et al., 2022), supra-additively (indicating synergy) (Bothma et al., 2015; Lam et al., 2015; Joo et al., 2016; Shin et al., 2016; Carleton et al., 2017; Thomas et al., 2021; Lin et al., 2022; Brosh et al., 2023) or sub-additively (indicating redundance) (Moorthy et al., 2017; Osterwalder et al., 2018). Nevertheless, all of these studies tested only small numbers of enhancers in one or a few genomic loci. Therefore, it is still unclear what the globally predominant mode of functional interaction is between enhancers, and how this may depend on the target promoter. For this, more scalable approaches are needed. Here, we report such an approach.

*Systematic approach.* To systematically assess enhancer-enhancer interplay and the collective effect on promoter activity, we designed a three-way combinatorial

MPRA based on our previous combinatorial reporters (Martinez-Ara et al., 2022). This assay allowed us to test tens of thousands of enhancer-enhancer-promoter combinations in a uniform setting (Inoue & Ahituv, 2015; van Arensbergen et al., 2017; Martinez-Ara et al., 2022). In order to test how different promoters might affect the enhancer interplay and their output, we not only varied enhancer pairs, but also the identity of the promoter in the three-way combinations.

*Results and significance*. By testing about 69,000 enhancer-enhancer-promoter combinations in mouse embryonic stem cells (mESCs) we found that pairs of enhancers mostly combine near-additively. Furthermore, promoter choice affects this enhancer-enhancer interplay, and promoters integrate the enhancer effects in a non-linear manner. Our three-way combinatorial approach provides further insights into two key aspects of gene regulation: how enhancers combine their effects, and how promoters integrate enhancer signals.

## RESULTS

### Construction of libraries/experimental design

*Triple combinations approach*. To systematically test how pairs of enhancers work together and in turn activate promoters, we implemented a three-way combinatorial approach in a massively parallel reporter assay (MPRA). We designed a cloning strategy that enabled us to construct libraries of tens of thousands of reporters that each contain a different enhancer-enhancer-promoter (EEP) combination (**Figure 1A-B**).

*Model system and promoter selection*. For these experiments we chose mESCs as a model. We selected eight promoters that are active in mESCs. Of these, seven are from tightly regulated genes that are involved in the control of the pluripotency state of the cells or their exit from this state (*Klf2, Sox2, Nanog, Otx2, Lefty1, Ffg5, Tbx3*) (Acampora et al., 2013; Dunn et al., 2014; D.-K. Kim et al., 2014; Thomas et al., 2021). In addition, we included the promoter of the housekeeping (Hounkpe et al., 2021) gene *Ap1m1*, which neighbours the *Klf2* gene. From each promoter we approximately included the −350- to +50-bp region around the transcription start site (TSS), because this includes the most information-rich part of functional promoters (Martinez-Ara et al., 2022; van Arensbergen et al., 2017).

*Clusters of enhancers selection.* We selected clusters of enhancers that are either known to regulate the above promoters (Blinka et al., 2016; Thomas et al., 2021; Zuin et al., 2022; Brosh et al., 2023), or are candidates for such function because of their proximity to the respective promoters (See methods and **Supplementary Table 1**) (Hnisz et al., 2013; Whyte et al., 2013). We also included putative enhancers from two

122 additional developmental genes from the core pluripotency network (*Pou5f1 and*
123 *Nodal*) (Dunn et al., 2014) that are located elsewhere in the genome.
124 *Definition and selection of single enhancers*. We defined single enhancers as
125 ~400 bp elements overlapping with DNase hypersensitive sites (DHSs) in clusters of
126 enhancers (**Fig 1C-D**), as previously described (Martinez-Ara et al., 2022). Some of
127 these enhancer clusters could be considered as single large enhancers (Long et al.,
128 2016). However, since DHSs have been shown to retain enhancer activity individually
129 (Barakat et al., 2018; Agrawal et al., 2021; Bergman et al., 2022; Martinez-Ara et al.,
130 2022) we decided to test the combinatorial contributions of these smaller elements of
131 an enhancer cluster. In total we selected 59 of such single enhancers. We also included
132 20 random sequences as negative controls (See Methods, **Supplementary Table 2**).
133 *Cloning of libraries*. We cloned each promoter separately into the reporter assay
134 vector. We then barcoded the 8 separate vectors (see Methods). Next, we randomly
135 combined enhancers and controls into pairs and cloned them into the eight different
136 barcoded vectors downstream of the promoters (**Fig1A and Fig S1A**). This generated
137 eight different EEP combinatorial libraries, one per promoter, with a total combinatorial
138 space of 49,928 (8 promoters x 79 enhancers and controls x 79 enhancers and controls)
139 or 199,712 if we consider the four possible orientation/position combinations of each
140 enhancer or control fragment (**Fig1A**). Of this space, the libraries covered 138,528
141 (69%) of the total combinatorial space, and after application of stringent quality filters
142 (see Methods) we were able to determine activities of 110,180 (55%) combinations
143 (**Supplementary table 3**).
144 *Explanation of library design.* The design of the libraries generates eight
145 matrices, one per promoter. These contain three types of combinations (**Figure 1B):**
146 control-control (CC), enhancer-control (EC, regardless of position), and enhancer-
147 enhancer (EE), which can be from the same or a different enhancer cluster. Control-
148 control combinations are necessary to measure the baseline activity of each promoter.
149 Enhancer-control combinations are used to estimate the effects of single enhancers. In
150 total, we measured 2,584 control-control combinations, 38,390 enhancer-control
151 combinations, and 69,206 enhancer-enhancer combinations.
152 *Library transfections*. We transfected each library separately into mESCs, with
153 three biological replicates performed on different days. RNA was extracted and
154 barcodes were counted by sequencing as previously described (Martinez-Ara et al.,
155 2022) both from RNA and from the plasmid libraries. We calculated the activity of each
156 barcode as the ratio between barcode counts in cDNA and plasmid DNA. We then
157 averaged barcode activities for each triple combination across a minimum of 5

158 barcodes. For these averaged activities, all developmental promoter libraries showed
159 high correlations between replicates (Pearson's R = 0.82 to 0.96) (**Figure S2**). The
160 housekeeping promoter library showed somewhat lower correlations between
161 replicates (Pearson's R = 0.74 to 0.78), which is most likely due to a smaller dynamic
162 range. For further analyses, we then averaged the biological replicates for each
163 promoter.
164
165 **Effects of single enhancers**
166 *Calculation of boost indices*. In order to quantify the activation by each enhancer we
167 first calculated the baseline activity of each promoter, defined as the median activity
168 across all control-control combinations. Then, for each individual enhancer-promoter
169 combination, we calculated the median activity across all enhancer-control
170 combinations. For each single enhancer we then calculated a boost index, which is the
171 $\log_2$ ratio of its median activity across enhancer-control combinations over the
172 promoter baseline across control-control combinations **Figure 2A**). We observed that
173 there was little position and orientation dependence of the enhancer effects (**Figure**
174 **S3**). We therefore averaged the boost indices of single enhancers across positions and
175 orientations for further analyses.
176 *Selected DHSs are active enhancers.* The single enhancer boost indices show
177 that the majority of the sequences that we selected indeed act as enhancers, although
178 the boost indices varied up to ~5 $\log_2$ units (**Figure 2A-B**). All tested enhancers showed
179 significant activation of at least one promoter (at an estimated false discovery rate
180 (FDR) < 0.01) (**Figure 2B, D**). However, the dynamic ranges and effects of each enhancer
181 vary per promoter, with the boost indices being particularly smaller for the
182 housekeeping promoter (*Ap1m1*) (**Figure 2A, C**). We conclude that most DHSs that we
183 selected from the enhancer clusters act individually as enhancers.
184 *Promoter selectivity of enhancers.* In our previous study of a large set of E-P
185 pairs we found that the majority of tested enhancers exhibited significant preference
186 for certain promoters (Martinez-Ara et al., 2022). Applying the same statistical analysis
187 to the current enhancer set confirmed this finding: the variability of enhancer effects
188 across promoters are generally larger than what is expected from experimental noise
189 (**Figure S4**). This indicates that the single enhancers tested here show significant
190 promoter selectivity. This analysis (**Figure S4**) also points to systematic differences in
191 promoter responses; this will be further highlighted below.
192
193 **Enhancers generally combine near-additively**

194 *Enhancer-enhancer combinations are generally stronger than single enhancers***.** We
195 then focused on the effects of enhancer-enhancer (EE) combinations. In the boost
196 index matrix that covers all combinations for a single promoter, we observed that
197 generally EE combinations were more active than EC combinations (**Figure 3A)**. Indeed,
198 quantitative analysis showed that EE combinations tend to induce stronger activation
199 than the single enhancers (EC) (**Figure 3B-C**), although some saturation appears to
200 occur with the strongest enhancers (**Figure 3C**).
201 *Additive model fits better than multiplicative model.* We then asked how the
202 effects of two single enhancers are integrated in the enhancer-enhancer (EE)
203 combinations. We considered two simple models: additivity and multiplicativity. For
204 each EE combination we calculated the expected effect according to each of these
205 models based on the single enhancer measurements, and then compared the observed
206 and expected activities (additive in **Figure 3D-E,** multiplicative in **Figure 3F,** see
207 Methods). Interestingly, for most of the promoters the additive model consistently
208 showed a better fit to the measured data than the multiplicative model (**Figure 3E-F,**
209 **Figure S5**). In particular, the multiplicative model tended to strongly overestimate
210 activities at the higher range of activities (most active enhancers) for most promoters.
211 When focusing on EE combinations for which the multiplicative and additive expected
212 values differ more than 0.5 $\log_2$ units (to limit random noise effects), we found that the
213 multiplicative model matched the observed values better for only 0-17% EE
214 combinations across the seven promoters. Only for the *Ap1m1* promoter the
215 multiplicative and additive models were nearly indistinguishable. This may be due to
216 the low dynamic range of activities observed with this promoter. We note that for the
217 *Fgf5* promoter in the low activity ranges the expected activities were consistently
218 underestimated for both models. It is likely that this is an inaccuracy due to the relative
219 sparseness of the *Fgf5* data (the baseline activity of this promoter was estimated from
220 only 23 measured control-control combinations, compared to 219-477 combinations
221 for the other promoters). In summary, we conclude that a simple additive model fits
222 the observed data better than a multiplicative model.
223
224 **Supra-additive activation is rare and promoter dependent.**
225 *Low frequency of supra-additive enhancer-enhancer interactions.* Most enhancer-
226 enhancer combinations show only minor deviations from the additive model (**Figure**
227 **S5**), as illustrated by the distributions of observed/expected ratios (**Figure 4A**). By
228 applying a simple error model to estimate the noise in the expected EE activities
229 according to the additive model (**Figure 3D**, see Methods) we found that 74.5% of the

230 measured EE activities lies within one standard deviation (SD) of the predicted
231 activities (**Figure 4A**) and 94% lies within two standard deviations. This underscores
232 that most of the tested enhancers combine near-additively, and that both sub-additive
233 and supra-additive effects are generally rare or weak.

234 *Supra-additivity occurs more frequently with weak promoters*. Nevertheless,
235 **Figure 4A** suggests that deviations from the additive model may vary between
236 promoters. The *Sox2*, *Lefty1*, *Tbx3* and *Fgf5* promoters showed a relatively large
237 proportion of EE pairs with supra-additive effects, while the *Ap1m1* promoter exhibited
238 the lowest frequency of supra-additivity. Interestingly, the proportion of supra-
239 additivity correlates inversely with promoter baseline expression (**Figure 4B**, left panel).
240 Thus, lowly active promoters allow more supra-additive EE interactions, although these
241 remain a minority (<40% of EE pairs). In contrast, the proportion of sub-additive effects
242 (i.e., >1 SD below expected activity) is very similar across promoters (**Figure 4B right**).
243 This latter observation suggests that there is no strong saturation effect that could
244 explain the decrease in number of supra-additive interactions as promoter activity
245 increases.

246 *EE supra-additive effects are relatively infrequent and often promoter-*
247 *dependent***.** Next, we investigated whether supra-additivity involves a specific subset of
248 enhancers. Visual inspection identified a few individual enhancers that exhibited supra-
249 additive interactions with most other enhancers, but often in the context of particular
250 promoters. An example is Nodal_E160 in combination with the *Fgf5* promoter (**Figure**
251 **4D**, left panel). Other enhancers showed a diversity of supra-additive activities. Out of
252 the 58 tested enhancers 32.7% (19/58) showed consistent supra-additivity in at least
253 one promoter context (>0.5 $\log_2$(observed/expected), mean over all enhancers in at
254 least one promoter context), and only 10.3% (6/58) showed consistent supra-additivity
255 in two or more promoter contexts (**Figure 4C**). However, most enhancers behave like
256 Sox2_E180 (**Figure 4D**, right panel), showing small deviations from the additive
257 behaviour with most other enhancers, and with all of the tested promoters (**Figure 4C**).

258 *No evidence for locus-specific EE supra-additivity*. We then asked whether
259 enhancer pairs derived from the same genomic locus have a higher propensity towards
260 supra-additivity than those not derived from the same locus. This could point towards
261 co-evolution of enhancers that jointly control a gene. However, we found that EE
262 combinations from the same locus showed on average not a higher observed/expected
263 ratio (according to the additive model) than EE combinations from different loci; in fact,
264 this ratio is even slightly lower (**Figure 4E)**. Thus, generally there appears to be no
265 preferential synergy among enhancers that are located in the same genomic locus.

266

267 **Enhancer effects scale non-linearly across developmental promoters**

268 *Enhancer effects correlate well across developmental promoters*. Prompted by our
269 observations that some supra-additive EE interactions appear to be more pronounced
270 in the context of specific promoters, we decided to systematically compare the boost
271 indices of all EE pairs between the eight promoters. Strikingly, all seven developmental
272 promoters showed strong pairwise correlations of their boost indices (Pearson's R =
273 0.79 to 0.98), whereas the housekeeping promoter (*Ap1m1*) correlated much less with
274 each of the developmental promoters (Pearson's R = 0.19 to 0.60; **Figure S6**).

275 *Slopes of promoter boost responses differ*. Although the developmental
276 promoters showed strong correlations, we were surprised to notice that the slopes of
277 these relationships often deviated from 1. To further explore this, we plotted the boost
278 indices of each promoter against the average boost indices across all eight promoters
279 (**Figure 5A**). All of the developmental promoters correlated strongly with this average
280 pattern (Pearson's R= 0.93 to 0.99) but this was not the case for the housekeeping
281 promoter (Pearson's R= 0.42). The slopes relative to the averaged promoters ranged
282 from 1.54-0.32, i.e., about 5-fold, across the developmental promoters. These results
283 imply that the developmental promoters differ in the scaling (i.e., steepness) of their
284 responses to the set of EE pairs, even though the pattern of their responses is highly
285 similar. Interestingly, the relative slopes negatively correlate with the baseline activity
286 of these promoters (**Figure 5B**). Thus, promoters with low intrinsic activity respond
287 more steeply to changes in combined EE activity than promoters with high intrinsic
288 activity.

289 *Promoter response differences also occur with single enhancers.* Finally, we
290 asked whether these differential responses of promoters require EE pairs, or can also
291 be observed with single E elements. Indeed, this is the case (**Figure S7**).

292

293 **DISCUSSION**

294

295 *A large-scale survey of EEP combinations.* Here, we employed a MPRA approach to
296 measure three-way regulatory interactions in ~69,000 EEP combinations involving 8
297 promoters and 59 enhancers. This is of a much larger scale than previous efforts to
298 elucidate the combinatorial interplay among enhancers, which typically focused on a
299 handful of enhancers in single genomic loci (Bothma et al., 2015; Hay et al., 2016; Shin
300 et al., 2016; Dukler et al., 2016; Joo et al., 2016; Lam et al., 2015; Carleton et al., 2017;
301 Moorthy et al., 2017; Osterwalder et al., 2018; Thomas et al., 2021; Lin et al., 2022;
302 Brosh et al., 2023). These previous studies pointed to a diversity of functional

303 enhancer-enhancer (EE) interactions, including redundancy, additivity and synergy. Our
304 study also uncovered a broad spectrum of functional interactions, but simple near-
305 additive effects of EE combinations were predominant, while supra-additive
306 interactions occurred at a lower frequency and were dependent on the specific
307 enhancer and promoter context.

308 *Promoters "interpret" and scale EE interplay.* An important finding is that
309 promoters can have substantial impact on supra-additivity of enhancer-enhancer
310 interactions. Promoters may thus be seen as "interpreters" of EE combinations. This
311 underscores the importance of studying triple combinations of elements, and perhaps
312 in the future even more complex settings. A salient finding is that the different
313 promoters show different degrees of responsiveness to EE pairs, with the stronger
314 promoters being boosted less than the weaker promoters. The responses of the
315 promoters follow roughly a power-law function of which the exponent is dependent on
316 the promoter (**Figures 5A and S6**, which are on log-log scale). Because it is not possible
317 to define an absolute measure of enhancer strength, we cannot determine the
318 absolute values of the exponents of the promoter power-law responses. This would be
319 of interest because exponents <1 would lead to "diminishing returns" with higher EE
320 activities, while exponents >1 would correspond to increasing non-linear amplification
321 with higher EE activities. The apparent saturation observable at high enhancer
322 activities in **Figure 3C** suggests that the exponents are most likely below 1. Our
323 findings resemble those of an earlier study (Hong & Cohen, 2022) indicating that
324 promoter activities are nonlinearly scaled by the genomic environment, with the
325 steepness of the scaling being negatively correlated with promoter strength.

326 *Episomal MPRAs vs genomic context*. The reporters in our MPRA were not
327 integrated into the genome. On the one hand, this ensures that all EEP combinations
328 are tested under similar conditions, without confounding effects of variable chromatin
329 contexts. On the other hand, we cannot rule out that the results are not fully
330 representative of a natural chromatin environment. However, other studies have
331 indicated that transiently transfected reporters generally behave quite similar, although
332 not identically, to integrated reporters (Inoue & Ahituv, 2015; Klein et al., 2020), and
333 another study indicated that chromatin context can scale transcriptional activity of
334 inserted reporters in a simple linear manner that is largely independent of the inserted
335 sequence (Maricque et al., 2018). Some of the E elements that we tested were
336 previously

337 *Results may be biased by model system.* We cannot rule out that our results are
338 biased by our selection of enhancers and promoters, which was mostly focused on

339 genes linked to pluripotency in mouse ESCs. We may have missed more prominent
340 synergistic or other non-additive effects that may arise in different cell types or among
341 different classes of enhancers and promoters. However, the protocol that we
342 developed for three-way MPRAs should be generally applicable to any set of enhancers
343 and promoters.

344 *Spacing and context of elements*. It should also be considered that in our MPRA
345 the enhancer and promoter elements are placed in fixed positions and close to one
346 another, which does not recapitulate their relative positioning in the genome. Genomic
347 distance and chromatin contexts could act as thresholds for activation, or turn additive
348 interactions into non-linear relationships. For instance, the ability of a *Sox2* enhancer to
349 activate its cognate promoter was found to depend on their distance, following a non-
350 linear decay function (Zuin et al., 2022). The rules for EE interplay may depend on the
351 EE distance or genomic context, as suggested by a recent report indicating that distant
352 pairs of enhancers are more often synergistic than closely spaced enhancers, which
353 tend to act additively (Lin et al., 2022). A recent study reported synergies between E
354 elements in the *Sox2* Locus Control Region (a large compound enhancer region) in the
355 native genomic context (Brosh et al., 2023) We did not observe such synergies, but the
356 data of that study are not directly comparable to ours due to differences in the size of
357 the elements tested. In the native *Fgf5* locus, several of the E elements exhibited some
358 degree of synergy in differentiated cells (Thomas et al., 2021). However, in mESCs this
359 synergy was not detected, which is in line our results for the same E elements from the
360 *Fgf5* locus. Possibly, the default mode of enhancer-enhancer interplay is additive, while
361 the precise chromatin context, element spacing and cellular state may enable synergy
362 in some genomic loci.

363 *Enhancer - promoter compatibility re-interpreted.* In our previous MPRA study
364 (Martinez-Ara et al., 2022) we found evidence for selectivity of enhancer - promoter
365 functional interactions. However, due to the limited number of enhancer-promoter
366 combinations tested we could not fully disentangle what caused these differences in
367 promoter responses. The data from the eight promoters indicate that there are two
368 intermingled factors that account for differences between promoters. First, distinct
369 classes of promoters show different specificities, as illustrated by the poor correlations
370 between the *Ap1m1* housekeeping promoter and the developmental promoters.
371 Second, promoters scale the input they receive from enhancer pairs differently, as
372 discussed above. It is probable that the compatibility differences we observed in our
373 previous study (Martinez-Ara et al., 2022) are a mix of these two phenomena: (i)
374 promoters that respond differently to enhancers because of specificity, and (ii)

375 promoters that respond differently to enhancers because of differences in
376 responsiveness caused by their intrinsic strength.
377 *Relevance of findings and context.* The findings presented here offer insights
378 into two aspects of the gene regulatory process – interplay between enhancers, and
379 promoter integration of the resulting signals – in a more quantitative manner. The
380 near-additivity of enhancers will help us understand how different enhancers may be
381 more or less relevant in different loci. The non-linearity of promoter responsiveness
382 and the effects of promoters on enhancer-enhancer interactions illustrates how some
383 promoters may be more sensitive to activating signals. Thus, it will be useful to
384 quantify promoter responsiveness genome-wide to understand which genes may be
385 more or less sensitive than others to environmental cues or to non-coding mutations.
386 These aspects may be especially relevant for developmentally regulated transcription
387 factors and their target genes, like the ones here studied, as they can be very sensitive
388 to dosage changes (Naqvi et al., 2023).
389

390 ## METHODS

391

392 **Cell culture**

393 We used E14tg2a male mouse embryonic stem cells (ATCC CRL-1821) cultured in 2i+LIF
394 for all experiments. As described {Martinez-Ara,2022,35594855}, we prepared 2i+LIF
395 media according to the 4DN nucleome protocol
396 (https://data.4dnucleome.org/protocols/cb03c0c6-4ba6-4bbe-9210-c430ee4fdb2c/).
397 The reagents used were Neurobasal medium (#21103-049, Gibco), DMEM-F12medium
398 (#11320-033, Gibco), BSA (#15260-037; Gibco), N27 (#17504-044; Gibco), B2 (#17502-
399 048; Gibco), LIF(#ESG1107; Sigma-Aldrich), CHIR-99021 (#HY-10182; MedChemExpress)
400 and PD0325901 (#HY-10254; MedChemExpress), monothioglycerol(#M6145-25ML;
401 Sigma) and glutamine (#25030-081, Gibco). Mycoplasma contamination was ruled out
402 by monthly tests (#LT07-318; Lonza).
403

404 **Selection of promoters, enhancers and controls.**

405 For the construction of the libraries we selected seven developmental genes from the
406 mESC pluripotency regulatory network (Acampora et al., 2013; Dunn et al., 2014; D.-K.
407 Kim et al., 2014; Thomas et al., 2021), *Klf2, Sox2, Nanog, Otx2, Lefty1, Ffg5, Tbx3;* and
408 the housekeeping gene *Ap1m1* (Hounkpe et al., 2021), which is neighbouring the Klf2
409 gene**.** The promoters of these genes were identified based on transcript and TSS
410 annotation (Frankish et al., 2019) and overlaps with DNAse hypersensitivity sites (DHS)

411 (Joshi et al., 2015) as previously described (Martinez-Ara et al., 2022). We then
412 selected sequences that approximately cover -350 to 50bp around the TSS.
413 The clusters of enhancers were selected in as follows: 1. Clusters of enhancers
414 that were known to regulate the developmental genes (*Nanog, Sox2, Fgf5* clusters)
415 (Blinka et al., 2016; Brosh et al., 2023; Thomas et al., 2021; Zuin et al., 2022); 2.
416 Putative clusters of enhancers in proximity of one of the developmental genes and that
417 overlapped with previously defined super-enhancers (*Lefty1, Tbx3, Otx2* clusters)
418 (Hnisz et al., 2013; Whyte et al., 2013); 3. A cluster of putative enhancers previously
419 identified by us (Martinez-Ara et al., 2022) to act as enhancers in an MPRA (*Klf2*
420 cluster). 4. Clusters of enhancers identified as putative enhancers of other mESC
421 pluripotency genes for which we did not include their promoters (*Pou5f1* and *Nodal*
422 clusters) and overlapping with previously defined super-enhancers (Hnisz et al., 2013,
423 2015; Whyte et al., 2013).
424 Clusters of enhancers were divided into single enhancers based on DNAse
425 hypersensitivity sites (DHS) (Joshi et al., 2015). These single enhancers were then
426 resized to ~450bp from the center of DNAse hypersensitivity peaks as previously
427 described (Martinez-Ara et al., 2022).
428 For the library construction we included 20 450 bp long randomized control
429 sequences as previously described (Martinez-Ara et al., 2022). Fifteen of these were
430 based on five DHS peak sequences that were randomly scrambled. The other five
431 controls were randomly generated sequences with the similar GC content as the DHS
432 peak sequences. These control sequences were ordered as synthetic DNA.
433 **Supplementary Table 2** contains the sequences of these controls**.**
434 For single enhancers and promoters we designed PCR primers against the 50 bp
435 ends of each selected element using BatchPrimer3 v1.0 (You et al., 2008). This yielded
436 PCR products of ~400 bp for each element. **Supplementary Table 1** lists the
437 coordinates of all individual enhancers and promoters. **Supplementary Table 4** lists
438 primer sequences.
439
440 **EEP Library generation**
441 The vector for library construction was previously used for the downstream reporter
442 assay in our previous study (Martinez-Ara et al., 2022). This vector is based on a
443 pSMART backbone. It contains a green fluorescent protein (GFP) open reading frame
444 followed by a barcode, and a psiCheck polyadenylation signal (PAS) introduced during
445 barcoding, followed by the cloning site for inserts and a triple polyadenylation site
446 (SV40+bGH+psiCheckPAS).

447 Each of the selected eight promoters were amplified by PCR and individually
448 inserted by Gibson assembly (#E2611S; New England Biolabs) into the reporter vector.
449 Each construct was transformed into standard 5-alpha competent bacteria (#C2987;
450 NEB) grown overnight in in 500 ml of standard Luria Broth (LB) with 50 mg/ml of
451 kanamycin and purified. Then each promoter sequence was verified by Sanger
452 sequencing.
453 Each of the promoter vectors was then barcoded. Similar to what we described
454 previously (Martinez-Ara et al., 2022), we digested 10 µg of each vector with AvrII
455 (#ER1561; Thermo Fischer) and XcmI (#R0533; NEB) and performed a gel-purification.
456 Barcodes were amplified in 10 separate 100ul PCR reactions using 5 µl of 10 mM
457 primer 275JvA, 5 ml of 10 mM primer 465JvA and 1 µl of 0.1 mM template 274JvA (see
458 **Supplementary Table 4** for sequences). 14 PCR cycles were performed using MyTaq
459 Red Mix (#BIO-25043; Bioline), yielding 30 µg of barcodes. We purified barcodes by
460 phenol-chloroform extraction and isopropanol precipitation. Barcodes were digested
461 overnight with 80 units of NheI (#R0131S; NEB) and purified by magnetic bead
462 purification (#CPCR-0050; CleanNA). Each promoter vector and barcodes were then
463 ligated in one 100 µl reaction. In each reaction we used 3 µg of digested vector and 2.7
464 µg digested barcodes, 20 units NheI (#R0131S; NEB), 20 units AvrII, 10 ml of 103
465 CutSmart buffer, 10 µl of 10 mM ATP, 10 units T4 DNA ligase (#10799009001 Roche).
466 We performed a cycle-ligation of six cycles (10 min at 22 °C and 10 min at 37 °C),
467 followed by 20 min heat-inactivation at 80 °C. Linear barcoded vectors were purified by
468 magnetic beads and digested for 3h with 40 units of XcmI (#R0533S; NEB). Finally
469 barcoded vectors were purified again by gel-purification.
470 Single enhancers were amplified by PCR from E14tg2a mESC genomic DNA, then
471 purified with magnetic beads (#CPCR-0050, Clean NA) and quantified by Qubit using
472 dsDNA High sensitivity Qubit kit (#Q33231; invitrogen). Single enhancers and random
473 controls were combined in approximately equimolar manner. This pool of elements
474 was end-repaired using End-It DNA End-Repair Kit (#ER0720; Epicentre) and self-ligated
475 using Fast-link ligase (LK0750H; Lucigen). Duplets of 800bp-1000bp were excised from
476 agarose gel, purified (BIO-52059; Bioline), and A-tailed using Klenow HC 3´ → 5´ exo
477 (#M0212L; NEB).
478 The pool of duplets was ligated to each of the barcoded promoter vectors using
479 Takara ligation kit version 2.1 (#6022; Takara). Ligation products were purified using
480 magnetic bead purification and 2 µl of the ligation were electroporated into 20 µl of
481 electrocompetent e. cloni 10G supreme (#60081-1; Lucigen). To estimate the
482 approximate complexity of the resulting libraries, dilutions of the electroporated

483 bacteria were plated for overnight culture and colonies were counted. We aimed at a
484 minimum of 2 million complexity per library in order to have sufficient expected
485 representation of all possible combinations.
486 When the estimated complexity of the libraries was not high enough, we
487 amplified the ligation products with primers aligning to the vector. This generated PCR
488 products consisting of the duplets pool and a vector overhang. These PCR products
489 were used for Gibson assembly with the barcoded promoters (#E2611S; NEB). Gibson
490 assembly reactions were purified with magnetic beads and 2 µl of the reaction were
491 electroporated into 20 µl of electrocompetent e. cloni 10G supreme (#60081-1;
492 Lucigen).
493 Each library was grown overnight in 500 ml of standard Luria Broth (LB) with 50
494 mg/ml of kanamycin and purified using a maxiprep kit (K210016, Invitrogen).
495
496 **Characterization of libraries by iPCR and sequencing**
497 Barcode to insert (enhancer and control duplets) combinations were identified by
498 inverse PCR (iPCR) and Illumina sequencing as described before(van Arensbergen et al.,
499 2017)**.** In brief, each library was digested overnight with I-CeuI (#R0699S, NEB),
500 barcode-insert fragments were then circularized, remaining linear fragments digested,
501 and barcode-insert fragments were linearized again with I-SceI (#R0694S; NEB).
502 Barcode-insert fragments were then amplified by PCR with Illumina adapters and
503 sequenced on an Illumina NextSeq 550 platform using 150 bp paired-end sequencing.
504 Each library was processed separately and mixed together for sequencing.
505
506 **Transfection of libraries**
507 Each EEP library was transfected separately into E14tg2a mESCs. Per library 20 million
508 cells were nucleofected (5µgs of library and 5million cells per cuvette, 5 cuvettes) using
509 Amaxa nucleofector II, program A-30, and Mouse Embryonic Stem Cell NucleofectorTM
510 Kit (#VPH-1001, Lonza). Three biological replicates were performed per library on
511 different days. Cells were collected 24 hours after transfection in 5 ml of TRIsure (#BIO-
512 38032; Bioline) and frozen at -80 °C until further processing.
513
514 **cDNA sequencing**
515 RNA was extracted and prepared for sequencing as described for the Downstream
516 Assay in our previous study (Martinez-Ara et al., 2022). From each sample, the aqueous
517 phase containing the total RNA of the TRIsure solution was extracted and purified on
518 an RNA extraction column (#K0732, Thermo Scientific). Total RNA was digested with 10

519 units of DNase I (#04716728001; Roche) for 30 minutes in 10-8 10 µl reactions
520 containing 5 µg of RNA. DNase I was inactivated by addition of 1 µl of 25 mM EDTA and
521 incubation at 70°C for 10 min.
522 cDNA was produced in 20 µl reactions (1 per 10 µDNAse reaction) using Maxima
523 Reverse transcriptase (#EP0743; ThermoFisher Scientific) and a gene specific (targeting
524 the GFP reporter) primer (JvA304, see **Supplementary table 4** for sequence). dNTPs
525 and the primer were mixed with the DNAse digested RNA and incubated at 65°C for 5
526 minutes. The RT buffer, enzyme and RNAse inhibitor were added and the reaction was
527 incubated at 50 °C for 1 hour. cDNA was then amplified by two nested PCRs to make it
528 strand specific. The first PCR reaction uses (index variants of 285JvA containing the S2,
529 index and p7 adaptor) and 305JvA (targeting the adapter introduced by 304JvA). Each
530 20 µl RT reaction was amplified in a 100µl PCR reaction with MyTaq Red mix (#BIO-
531 25043; Bioline). The second PCR reaction was performed using 10 µl of the product of
532 the previous reaction in a 100 µl reaction using the same index variant primer and
533 index variants of 437JvA (containing the S1, index and p5 adaptor) (see **Supplementary**
534 **table 4** for primer sequences). The first PCR was run for 10 cycles and the second one
535 for 8 cycles using the recommended Mytaq Red mix conditions (for both PCRs: 1 min
536 96 °C, then each cycle 15 s 96 °C, 15 s 60 °C, 15 s 72 °C). PCR products were purified
537 with magnetic beads and mixed for sequencing in an Illumina NovaSeq 6000 platform
538 using 100bp single end reads.
539
540 **Plasmid DNA (pDNA) sequencing**
541 Plasmid libraries were processed for sequencing as described (Martinez-Ara et al.,
542 2022) but using dual indexing for sequencing on a NovaSeq 6000 platform with 100 bp
543 single end reads. 1 µg of each plasmid was digested with I-SceI in order to linearise the
544 plasmid. Barcodes were amplified by PCR for 9 cycles from 50 ng of material. Primers
545 and reaction conditions were the same as in the amplification of cDNA. PCR products
546 were purified using magnetic beads and mixed for sequencing.
547
548 ## COMPUTATIONAL METHODS
549 cDNA and pDNA Data pre-processing and linking of barcodes to duplets was performed
550 using a custom snakemake pipeline (Köster & Rahmann, 2012). All other data analyses
551 and quantifications were performed in R (R Core Team, 2022). Figures were generated
552 using ggplot2 (Wickham, 2016). The snakemake pipeline and all the scripts used in this
553 publication are available in a GitHub repository
554 (https://github.com/vansteensellab/EEPCombinations).

555

556 **Linking Barcodes to duplet inserts**

557 The custom pipeline for linking barcodes to duplet inserts was previously described
558 {Martinez-ara,2022, 35594855}. In brief, iPCR reads from each library were locally
559 aligned using bowtie (version 2.3.4) (Langmead & Salzberg, 2012) with very sensitive
560 parameters (–very-sensitive-local) on a custom bowtie genome. This custom Bowtie
561 genome consists of virtual chromosomes that correspond to each of the enhancer and
562 control sequences used to generate the combinatorial libraries. Bam files were
563 processed by a custom Python script (Van Rossum & Drake, 2009). This script extracts
564 from read 1 the barcode sequence, and the identity and orientation of the DNA
565 fragment in position 1. From read 2 it extracts the identity and orientation of the DNA
566 fragment in position 2. Finally, barcodes were clustered using Starcode (version 1.1)
567 (Zorita et al., 2015) to remove PCR and sequencing errors.

568

569 **cDNA and pDNA data pre-processing**

570 For each cDNA and pDNA replicate of each library, barcodes were extracted from
571 single-end reads using a custom Python script that matches the constant region after
572 the barcode. Barcodes were clustered using Starcode (version 1.1) (Zorita et al., 2015)
573 to remove errors from sequencing and counts were summarized.

574

575 **cDNA and pDNA data post-processing**

576 For each library cDNA and pDNA barcodes were matched to iPCR barcodes. Any
577 barcode assigned to multiple fragment identities was removed from the data. Per
578 cDNA/pDNA replicate barcode counts were normalized to the total number of barcode
579 counts. Then activity was calculated per barcode and cDNA replicate as the
580 cDNA:pDNA normalized counts ratio. Per individual duplet combination these
581 normalized activity ratios were averaged across barcodes with a minimum requirement
582 of five barcodes and eight pDNA counts per barcode. The mean activity across the
583 three replicates was calculated as the geometric mean.

584

585 **Calculation of boost indices**

586 In our previous study (Martinez-Ara et al., 2022) we observed that it was critical to use
587 control sequences to estimate the baseline activity of the promoters. For each
588 promoter we used all of the control-control combinations to estimate its baseline
589 activity. We calculated this as the median activity across all control-control
590 combinations. Then the boost index of an enhancer-enhancer or enhancer-control

591 (single enhancer boost index) combination was calculated as the $\log_2$ ratio between the
592 observed activity of this combination and the baseline activity of the promoter. Except
593 in **Supplementary Figure 3**, boost indices were averaged across positions and
594 orientations. In **Figure 2a** boost indices were averaged across all enhancer-control
595 combinations for each enhancer. In **Figure 3a** boost indices were averaged across
596 orientations but not positions.
597
598 **Selectivity of promoters**
599 To analyse the selectivity of promoters on single enhancers (enhancer-control-
600 promoter combinations) we used a Welch F-test due to the heteroscedasticity of the
601 data. The test was performed on each of the populations of enhancer-control-
602 promoter combinations. First the data was confirmed to be approximately normal by a
603 Sapiro test, which confirmed that 94.1% of the 439 single enhancers + promoter
604 combinations (that could be tested) show normal distributions (at a 1% FDR).
605
606 **Calculation of additive and multiplicative expected activities**
607 To calculate the expected additive and multiplicative activities we first had to
608 determine three values. The baseline activity of the promoter ($A_{CCP}$) was measured as
609 the median activity across all control-control combinations for each promoter. The
610 single enhancer activity estimates ($A_{E1CP}$ and $A_{CE2P}$, enhancer in position1 and position 2
611 respectively) were calculated as the median activity of all enhancer-control
612 combinations where the enhancer was found in either position 1 ($A_{E1CP}$) or position 2
613 ($A_{CE2P}$). Therefore
614
615 $A_{CCP} = \text{median}\{A_{C1C2P}, A_{C1C3P}, \ldots, A_{CiCjP}\}$
616 $A_{E1CP} = \text{median}\{A_{E1C1P}, A_{E1C2P}, \ldots, A_{E1CiP}\}$
617 $A_{CE2P} = \text{median}\{A_{C1E2P}, A_{C2E2P}, \ldots, A_{CiE2P}\}$
618
619 Then for each enhancer-enhancer-promoter combination the expected additive activity
620 was calculated as the sum of the estimated activities of the single enhancers in each
621 position minus the promoter baseline activity ($A_{E1CP} + A_{CE2P} - A_{CCP}$). The multiplicative
622 expected activity was calculated as the product of the estimated activities of the single
623 enhancers in each position divided by the promoter baseline activity (($A_{E1CP}$ *
624 $A_{CE2P}$)/$A_{CCP}$). For the expected additive activity, the standard deviation of the estimate
625 was propagated as the square root of the sum of the variances of the single
626 measurements. Therefore

Expected Additive Activity = ($A_{E1CP} + A_{CE2P} - A_{CCP}$)

Expected Multiplicative Activity = ($A_{E1CP} * A_{CE2P}$)/$A_{CCP}$

The additive expected activities were then used to calculate $log_2$(observed/expected) ratios in order to find supra-additive behaviours between enhancers. To find trends the $log_2$ ratios were then averaged for each enhancer across enhancer-enhancer pairs.

**Non-linear Promoter responses.**

For every enhancer-enhancer combination present in all eight libraries we averaged boost indices across the eight promoters. This average enhancer effect was then used to calculate the slope between the Average boost index and the observed boost index for each of the promoters. The slopes were calculated using the lm function in R.

## DATA AVAILABILITY

Raw sequencing data and processed data are available at Gene expression Omnibus (GEO): GSE240586**.** Publicly available datasets that were used are listed in and are available at https://osf.io/yuc4e/. Lab journal records are available at https://osf.io/yuc4e/. Code of data processing pipelines and analysis scripts are available at https://github.com/vansteensellab/EEPCombinations.

## AUTHOR CONTRIBUTIONS

M.M.A., F.C., and B.v.S. designed the study. M.M.A. developed computational and experimental methods, performed analyses and experiments. B.v.S. and M.M.A. wrote the manuscript, with input from F.C. B.v.S. supervised the study.

## ACKNOWLEDGEMENTS

We thank the NKI Genomics, and Research High Performing Computing facilities for technical support, and members of our laboratory for helpful discussions and comments. Funded by the European Union (European Research Council, RE_LOCATE, 101054449). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. Research at the Netherlands Cancer Institute is supported by an

institutional grant of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport. The Oncode Institute is partially funded by the Dutch Cancer Society

## DECLARATION OF INTERESTS

F.C. is a co-founder of enGene Statistics GmbH.

## REFERENCES


Acampora, D., Di Giovannantonio, L. G., & Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development (Cambridge, England)*, *140*(1), 43–55. https://doi.org/10.1242/dev.085290

Agrawal, P., Blinka, S., Pulakanti, K., Reimer, M. H., Stelloh, C., Meyer, A. E., & Rao, S. (2021). Genome editing demonstrates that the -5 kb Nanog enhancer regulates Nanog expression by modulating RNAPII initiation and/or recruitment. *The Journal of Biological Chemistry*, *296*, 100189. https://doi.org/10.1074/jbc.RA120.015152

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2 Pt 1), 299–308. https://doi.org/10.1016/0092-8674(81)90413-x

Barakat, T. S., Halbritter, F., Zhang, M., Rendeiro, A. F., Perenthaler, E., Bock, C., & Chambers, I. (2018). Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell*, *23*(2), 276-288.e8. https://doi.org/10.1016/j.stem.2018.06.014

Bergman, D. T., Jones, T. R., Liu, V., Ray, J., Jagoda, E., Siraj, L., Kang, H. Y., Nasser, J., Kane, M., Rios, A., Nguyen, T. H., Grossman, S. R., Fulco, C. P., Lander, E. S., & Engreitz, J. M. (2022). Compatibility rules of human enhancer and promoter sequences. *Nature*, *607*(7917), 176–184. https://doi.org/10.1038/s41586-022-04877-w

Bertolino, E., & Singh, H. (2002). POU/TBP cooperativity: A mechanism for enhancer action from a distance. *Molecular Cell*, *10*(2), 397–407. https://doi.org/10.1016/s1097-2765(02)00597-x

Blinka, S., Reimer, M. H., Pulakanti, K., & Rao, S. (2016). Super-Enhancers at the Nanog Locus Differentially Regulate Neighboring Pluripotency-Associated Genes. *Cell Reports*, *17*(1), 19–28. https://doi.org/10.1016/j.celrep.2016.09.002

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T., & Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *eLife*, *4*, e07956. https://doi.org/10.7554/eLife.07956

735 Brosh, R., Coelho, C., Ribeiro-Dos-Santos, A. M., Ellis, G., Hogan, M. S., Ashe, H. J.,
736 Somogyi, N., Ordoñez, R., Luther, R. D., Huang, E., Boeke, J. D., & Maurano, M.
737 T. (2023). Synthetic regulatory genomics uncovers enhancer context
738 dependence at the Sox2 locus. *Molecular Cell*, *83*(7), 1140-1152.e7.
739 https://doi.org/10.1016/j.molcel.2023.02.027
740 Carleton, J. B., Berrett, K. C., & Gertz, J. (2017). Multiplex Enhancer Interference
741 Reveals Collaborative Control of Gene Regulation by Estrogen Receptor α-
742 Bound Enhancers. *Cell Systems*, *5*(4), 333-344.e5.
743 https://doi.org/10.1016/j.cels.2017.08.011
744 Chang, T. H.-T., Primig, M., Hadchouel, J., Tajbakhsh, S., Rocancourt, D., Fernandez, A.,
745 Kappler, R., Scherthan, H., & Buckingham, M. (2004). An enhancer directs
746 differential expression of the linked Mrf4 and Myf5 myogenic regulatory genes
747 in the mouse. *Developmental Biology*, *269*(2), 595–608.
748 https://doi.org/10.1016/j.ydbio.2004.02.013
749 Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., & Blobel, G. A.
750 (2012). Controlling long-range genomic interactions at a native locus by
751 targeted tethering of a looping factor. *Cell*, *149*(6), 1233–1244.
752 https://doi.org/10.1016/j.cell.2012.03.051
753 Dukler, N., Gulko, B., Huang, Y.-F., & Siepel, A. (2016). Is a super-enhancer greater than
754 the sum of its parts? *Nature Genetics*, *49*(1), 2–3.
755 https://doi.org/10.1038/ng.3759
756 Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., & Smith, A. G. (2014). Defining an
757 essential transcription factor program for naïve pluripotency. *Science (New*
758 *York, N.Y.)*, *344*(6188), 1156–1160. https://doi.org/10.1126/science.1248882
759 ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in
760 the human genome. *Nature*, *489*(7414), 57–74.
761 https://doi.org/10.1038/nature11247
762 Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J.,
763 Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A.,
764 Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S.,
765 Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human
766 and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766–D773.
767 https://doi.org/10.1093/nar/gky955
768 Gallego Romero, I., & Lea, A. J. (2023). Leveraging massively parallel reporter assays for
769 evolutionary questions. *Genome Biology*, *24*(1), 26.
770 https://doi.org/10.1186/s13059-023-02856-6

Gasperini, M., Tome, J. M., & Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews. Genetics*, *21*(5), 292–310. https://doi.org/10.1038/s41576-019-0209-0

Grosveld, F., van Assendelft, G. B., Greaves, D. R., & Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, *51*(6), 975–985. https://doi.org/10.1016/0092-8674(87)90584-8

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L., Kassouf, M. T., Marieke Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., Telenius, J., Williams, R., Rode, C., Li, P.-S., Pennacchio, L. A., Sloane-Stanley, J. A., Ayyub, H., Butler, S., Sauka-Spengler, T., … Higgs, D. R. (2016). Genetic dissection of the α-globin super-enhancer in vivo. *Nature Genetics*, *48*(8), 895–903. https://doi.org/10.1038/ng.3605

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, *155*(4), 934–947. https://doi.org/10.1016/j.cell.2013.09.053

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E., & Young, R. A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Molecular Cell*, *58*(2), 362–370. https://doi.org/10.1016/j.molcel.2015.02.014

Hong, C. K. Y., & Cohen, B. A. (2022). Genomic environments scale the activities of diverse core promoters. *Genome Research*, *32*(1), 85–96. https://doi.org/10.1101/gr.276025.121

Hounkpe, B. W., Chenou, F., de Lima, F., & De Paula, E. V. (2021). HRT Atlas v1.0 database: Redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Research*, *49*(D1), D947–D955. https://doi.org/10.1093/nar/gkaa609

Inoue, F., & Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics*, *106*(3), 159–164. https://doi.org/10.1016/j.ygeno.2015.06.005

Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X., & Blobel, G. A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Molecular Cell*, *29*(2), 232–242. https://doi.org/10.1016/j.molcel.2007.11.020

Joo, J.-Y., Schaukowitch, K., Farbiak, L., Kilaru, G., & Kim, T.-K. (2016). Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nature Neuroscience*,

807 *19*(1), 75–83. https://doi.org/10.1038/nn.4170
808 Joshi, O., Wang, S.-Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P. J., Habibi, E., Shaik,
809 J., Saeed, S., Handoko, L., Richmond, T., Spivakov, M., Burgess, D., &
810 Stunnenberg, H. G. (2015). Dynamic Reorganization of Extremely Long-Range
811 Promoter-Promoter Interactions between Two States of Pluripotency. *Cell Stem*
812 *Cell*, *17*(6), 748–757. https://doi.org/10.1016/j.stem.2015.11.010
813 Kim, D.-K., Cha, Y., Ahn, H.-J., Kim, G., & Park, K.-S. (2014). Lefty1 and lefty2 control the
814 balance between self-renewal and pluripotent differentiation of mouse
815 embryonic stem cells. *Stem Cells and Development*, *23*(5), 457–466.
816 https://doi.org/10.1089/scd.2013.0220
817 Kim, S., & Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory
818 code. *Molecular Cell*, *83*(3), 373–392.
819 https://doi.org/10.1016/j.molcel.2022.12.032
820 Klein, J. C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., Ahituv, N., &
821 Shendure, J. (2020). A systematic evaluation of the design and context
822 dependencies of massively parallel reporter assays. *Nature Methods*, *17*(11),
823 1083–1091. https://doi.org/10.1038/s41592-020-0965-y
824 Köster, J., & Rahmann, S. (2012). Snakemake—A scalable bioinformatics workflow
825 engine. *Bioinformatics (Oxford, England)*, *28*(19), 2520–2522.
826 https://doi.org/10.1093/bioinformatics/bts480
827 Lam, D. D., de Souza, F. S. J., Nasif, S., Yamashita, M., López-Leal, R., Otero-Corchon, V.,
828 Meece, K., Sampath, H., Mercer, A. J., Wardlaw, S. L., Rubinstein, M., & Low, M.
829 J. (2015). Partially redundant enhancers cooperatively maintain Mammalian
830 pomc expression above a critical functional threshold. *PLoS Genetics*, *11*(2),
831 e1004935. https://doi.org/10.1371/journal.pgen.1004935
832 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.
833 *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
834 Lin, X., Liu, Y., Liu, S., Zhu, X., Wu, L., Zhu, Y., Zhao, D., Xu, X., Chemparathy, A., Wang,
835 H., Cao, Y., Nakamura, M., Noordermeer, J. N., La Russa, M., Wong, W. H., Zhao,
836 K., & Qi, L. S. (2022). Nested epistasis enhancer networks for robust genome
837 regulation. *Science (New York, N.Y.)*, *377*(6610), 1077–1085.
838 https://doi.org/10.1126/science.abk3512
839 Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes:
840 Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170–
841 1187. https://doi.org/10.1016/j.cell.2016.09.018
842 Maricque, B. B., Chaudhari, H. G., & Cohen, B. A. (2018). A massively parallel reporter

assay dissects the influence of chromatin structure on cis-regulatory activity. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4285

Martinez-Ara, M., Comoglio, F., van Arensbergen, J., & van Steensel, B. (2022). Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Molecular Cell*, *82*(13), 2519-2531.e6. https://doi.org/10.1016/j.molcel.2022.04.009

Moorthy, S. D., Davidson, S., Shchuka, V. M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N. N., & Mitchell, J. A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Research*, *27*(2), 246–258. https://doi.org/10.1101/gr.210930.116

Naqvi, S., Kim, S., Hoskens, H., Matthews, H. S., Spritz, R. A., Klein, O. D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J. K., & Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nature Genetics*, *55*(5), 841–851. https://doi.org/10.1038/s41588-023-01366-2

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., Kato, M., Garvin, T. H., Pham, Q. T., Harrington, A. N., Akiyama, J. A., Afzal, V., Lopez-Rios, J., Dickel, D. E., Visel, A., & Pennacchio, L. A. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, *554*(7691), 239–243. https://doi.org/10.1038/nature25461

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews. Genetics*, *20*(8), 437–455. https://doi.org/10.1038/s41576-019-0128-0

Shin, H. Y., Willi, M., HyunYoo, K., Zeng, X., Wang, C., Metser, G., & Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nature Genetics*, *48*(8), 904–911. https://doi.org/10.1038/ng.3606

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews. Genetics*, *15*(4), 272–286. https://doi.org/10.1038/nrg3682

Thomas, H. F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., Penz, T., Bock, C., Leeb, M., Halbritter, F., Wysocka, J., & Buecker, C. (2021). Temporal

879 dissection of an enhancer cluster reveals distinct temporal and functional
880 contributions of individual elements. *Molecular Cell*, *81*(5), 969-982.e13.
881 https://doi.org/10.1016/j.molcel.2020.12.047
882 Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and
883 interaction between hypersensitive sites in the active beta-globin locus.
884 *Molecular Cell*, *10*(6), 1453–1465. https://doi.org/10.1016/s1097-
885 2765(02)00781-5
886 Tuan, D., Solomon, W., Li, Q., & London, I. M. (1985). The 'beta-like-globin' gene
887 domain in human erythroid cells. *Proceedings of the National Academy of*
888 *Sciences of the United States of America*, *82*(19), 6384–6388.
889 https://doi.org/10.1073/pnas.82.19.6384
890 Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss,
891 M. J., Dekker, J., & Blobel, G. A. (2005). Proximity among distant regulatory
892 elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular Cell*,
893 *17*(3), 453–462. https://doi.org/10.1016/j.molcel.2004.12.028
894 van Arensbergen, J., FitzPatrick, V. D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker,
895 H. J., & van Steensel, B. (2017). Genome-wide mapping of autonomous
896 promoter activity in human cells. *Nature Biotechnology*, *35*(2), 145–153.
897 https://doi.org/10.1038/nbt.3754
898 Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.
899 Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P.
900 B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator
901 establish super-enhancers at key cell identity genes. *Cell*, *153*(2), 307–319.
902 https://doi.org/10.1016/j.cell.2013.03.035
903 Wickham, H. (2016). *Ggplot2 Elegant Graphics for Data Analysis*. Springer-Verlag New
904 York. https://ggplot2.tidyverse.org
905 Will, A. J., Cova, G., Osterwalder, M., Chan, W.-L., Wittler, L., Brieske, N., Heinrich, V.,
906 de Villartay, J.-P., Vingron, M., Klopocki, E., Visel, A., Lupiáñez, D. G., &
907 Mundlos, S. (2017). Composition and dosage of a multipartite enhancer cluster
908 control developmental expression of Ihh (Indian hedgehog). *Nature Genetics*,
909 *49*(10), 1539–1545. https://doi.org/10.1038/ng.3939
910 You, F. M., Huo, N., Gu, Y. Q., Luo, M.-C., Ma, Y., Hane, D., Lazo, G. R., Dvorak, J., &
911 Anderson, O. D. (2008). BatchPrimer3: A high throughput web application for
912 PCR and sequencing primer design. *BMC Bioinformatics*, *9*, 253.
913 https://doi.org/10.1186/1471-2105-9-253
914 Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A.

(2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, *518*(7540), 556–559. https://doi.org/10.1038/nature13994

Zorita, E., Cuscó, P., & Filion, G. J. (2015). Starcode: Sequence clustering based on all-pairs search. *Bioinformatics (Oxford, England)*, *31*(12), 1913–1919. https://doi.org/10.1093/bioinformatics/btv053

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., & Giorgetti, L. (2022). Nonlinear control of transcription through enhancer-promoter interactions. *Nature*, *604*(7906), 571–577. https://doi.org/10.1038/s41586-022-04570-y

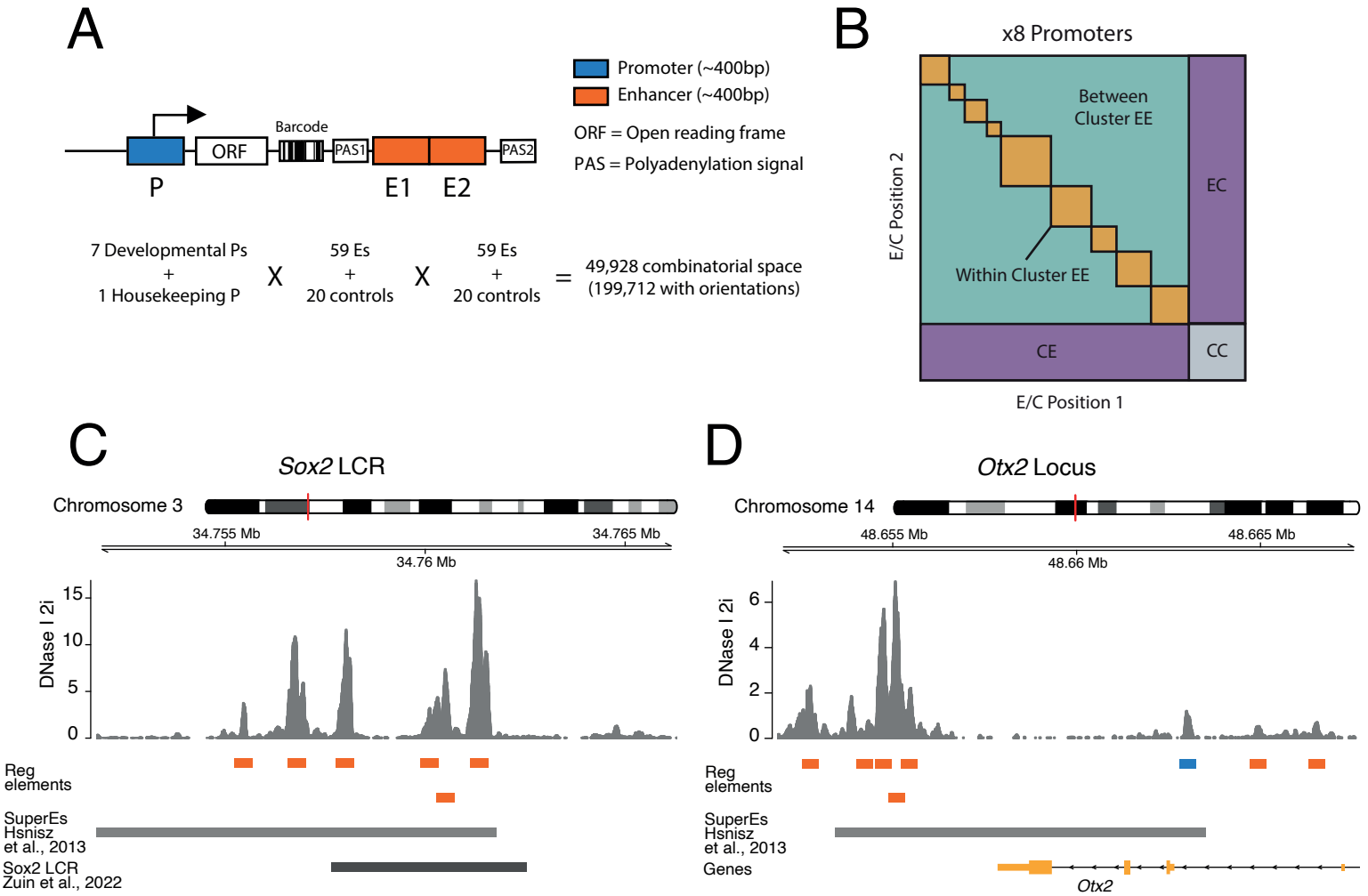



**Figure 1.** Schematic of three-way combinatorial approach. **A)** Three-way combinatorial Massively parallel reporter assay (MPRA) design to test enhancer-enhancer-promoter combinations. 8 barcoded reporter assay libraries, 1 per promoter, were constructed. Pairs of DNA elements (enhancers and scrambled control sequences) were inserted after barcoded reporter. The enhancers and controls can be placed in both orientations in either the enhancer position 1 (E1) or enhancer position 2 (E2). **B)** The design of the library yields 8 matrices that contain control-control combinations (CC), enhancer-control combinations (EC and CE) and enhancer-enhancer combinations (EE). **C)** and **D)** 2 example loci, *Sox2* LCR (locus control region) (**C**) and *Otx2* (**D**) from where we selected enhancers to test in the reporter libraries. Enhancers (orange) were defined around DNAse I hypersensitivty sites from mouse embryonic stem cells (mESC). Promoters (Blue) were chosen according to TSS annotatition and mESC DNAse I hypersensitivty sites. DNAse I data is publicly available from Joshi et al., 2015.

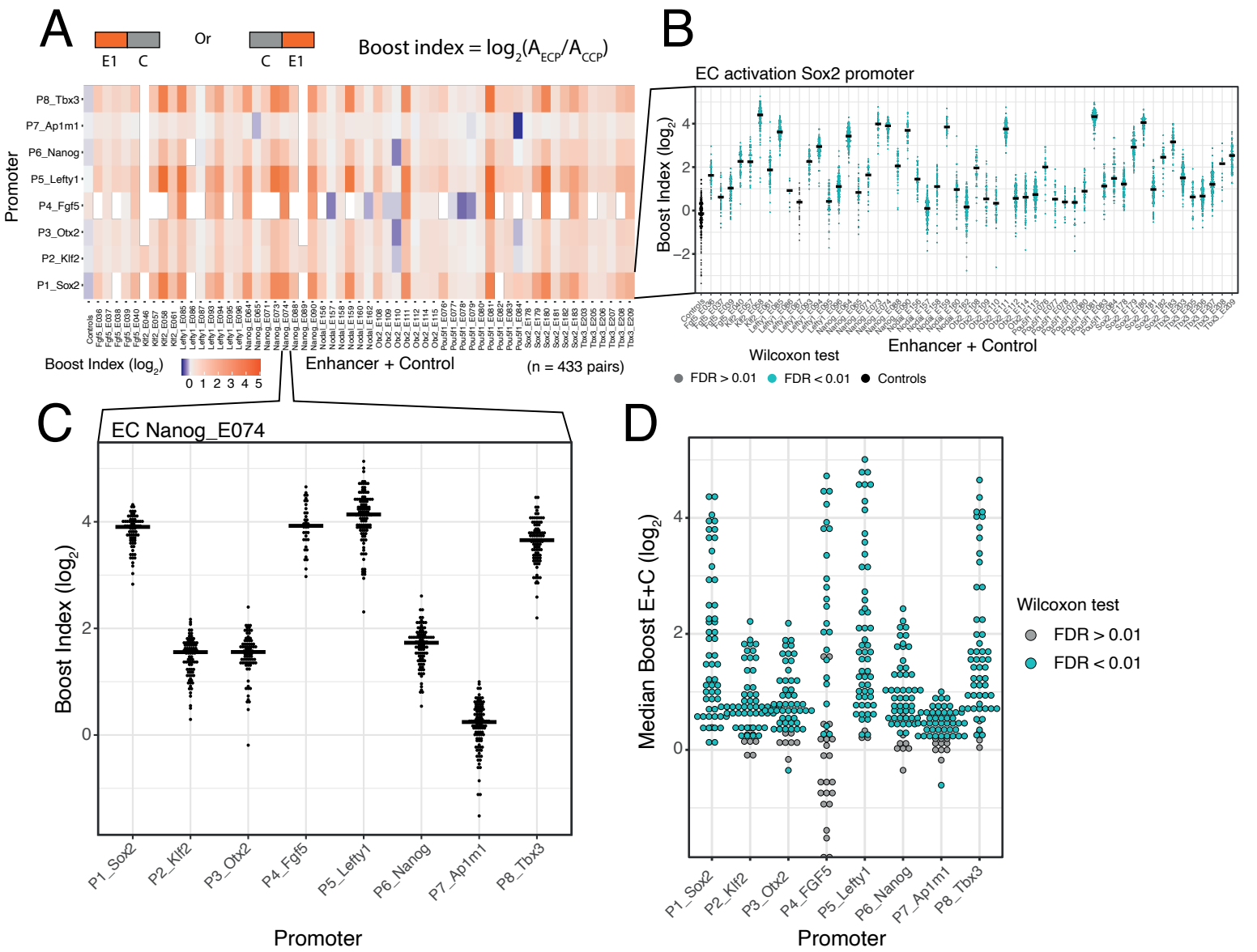



**Figure 2.** Effects of single enhancers across promoters. **A)** Enhancer-promoter median boost index matrix of single enhancers**.** For each single enhancer a median boost index (see figure and methods, $\log_2$(Activity_Combination/Activity_control-control_baseline)) was calculated across all enhancer-control combinations for that particular enhancer in any position and orientation, the baseline is the median activity of the promoter across all control-control combinations. For each control-control combination a boost index was calculated as the activity of the control-control combination over the median activity of all controls. A median control-control boost index was calculated based on this. Color coding of the matrix corresponds to the median boost indexes, white spaces are missing data. **B)** Distribution of boost indices for all enhancer+control combinations for the *Sox2* promoter. Leftmost column corresponds to the boost index distribution for all control-control combinations for the *Sox2* promoter. Each dot represents one enhancer+control combination. Horizontal lines correspond to the median of each distribution (same median as represented in the matrix in **A**). **C)** Distribution of boost indices for all enhancer+control combinations for the enhancer Nanog_E074 across all promoters. Each dot represents one enhancer+control combination. Horizontal lines correspond to the median of each distribution. **D)** Distribution of median boost indices for each single enhancer across promoters. Each dot represents the median of an enhancer across all enhancer+control combinations for that enhancers (same median as represented in the matrix in **A**). Coloring in B and D represents the significance at a 1% FDR for a Wilcoxon test comparing boost index distributions between enhancer+control combinations and the controls for each enhancer.

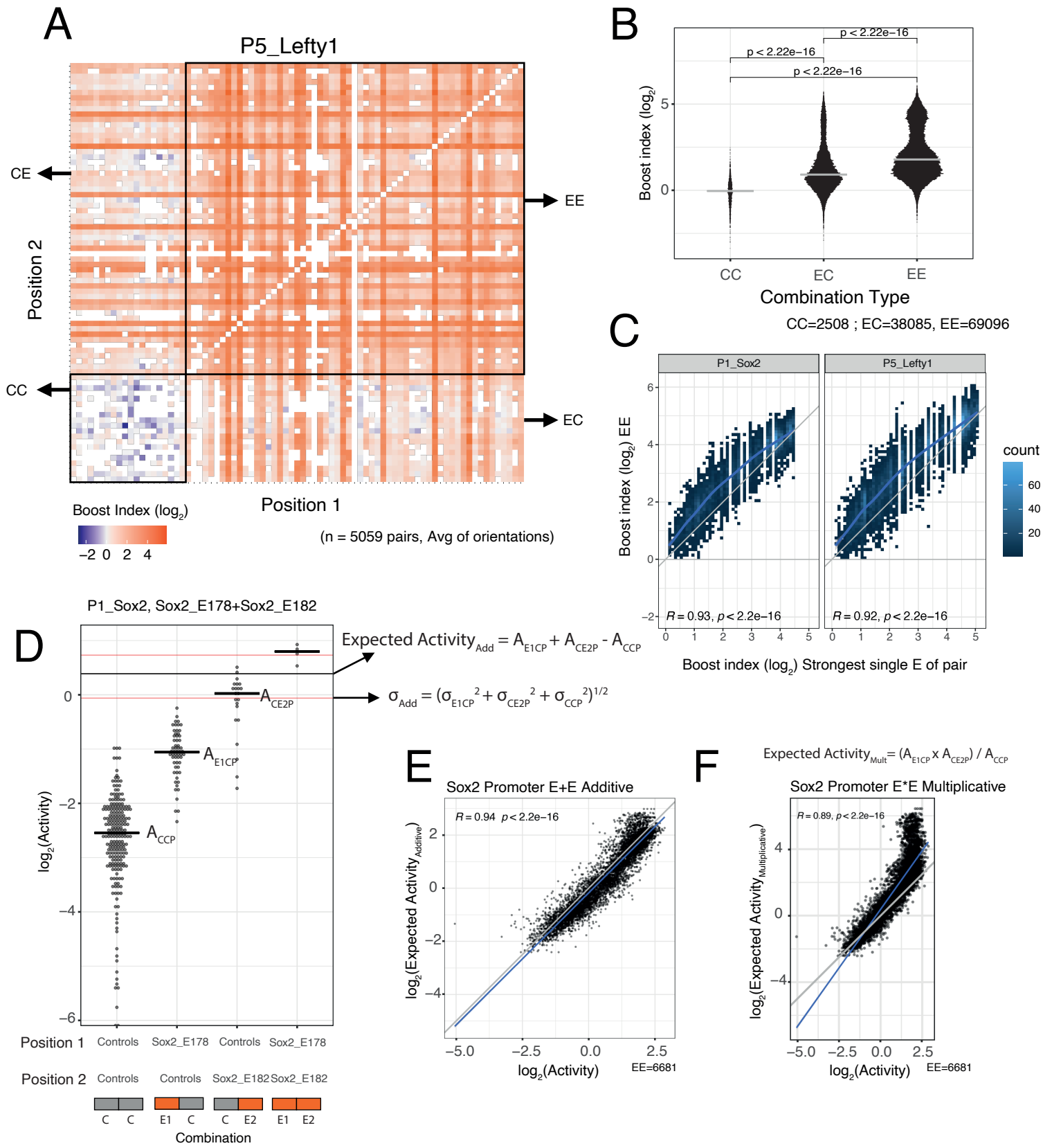



**Figure 3.** Effects of enhancer-enhancer combinations. **A)** Fragment-fragment combinatorial boost index matrix for the *Lefty1* promoter. Each square represents 1 control-control (CC), enhancer-control (EC or CE) or enhancer-enhancer (EE) combination. Color coding corresponds to the average boost index for each combination across all orientations measured over the median control-control baseline. **B)** Boost index distributions across all 8 promoters for each combination type, control-control (CC), enhancer-control (EC regardless of position) or enhancer-enhancer (EE) combinations. P-values correspond to the result of a Wilcoxon test. **C)** Relationship between observed boost index for each EE combination and the observed boost index of the strongest single enhancer of the pair for *Sox2* and *Lefty1* promoters. Blue lines represent the LOESS fit of the data. **D)** Observed and expected additive activities for the Sox2_E178+Sox2_E182 combination with the *Sox2* promoter and the individual activities of each of the elements. Each column represents the observed activities for the control-contol combinations, the enhancer-control combinations and the enhancer-enhancer combinations. The horizontal bars represent the median of each distribution. The horizontal black line represents the expected additive activity of the enhancer-enhancer combination as calculated by the formula in the panel in the linear space. The horizontal red lines represents the propagated standard deviations of the expected additive activity of the enhancer-enhancer combinationas as calculated by the formula in the panel. **E and F)** Relationship between observed and expected activities (additive in **E**, multiplicative in **F**) for all enhancer-enhancer combinations for the *Sox2* promoter. The blue lines represent the linear fit of the data. Grey diagonal line is the x=y identity line. In all panels R represents Pearson's correlation. Expected activities are calculated in the linear space and then plotted in the $\log_2$ space.

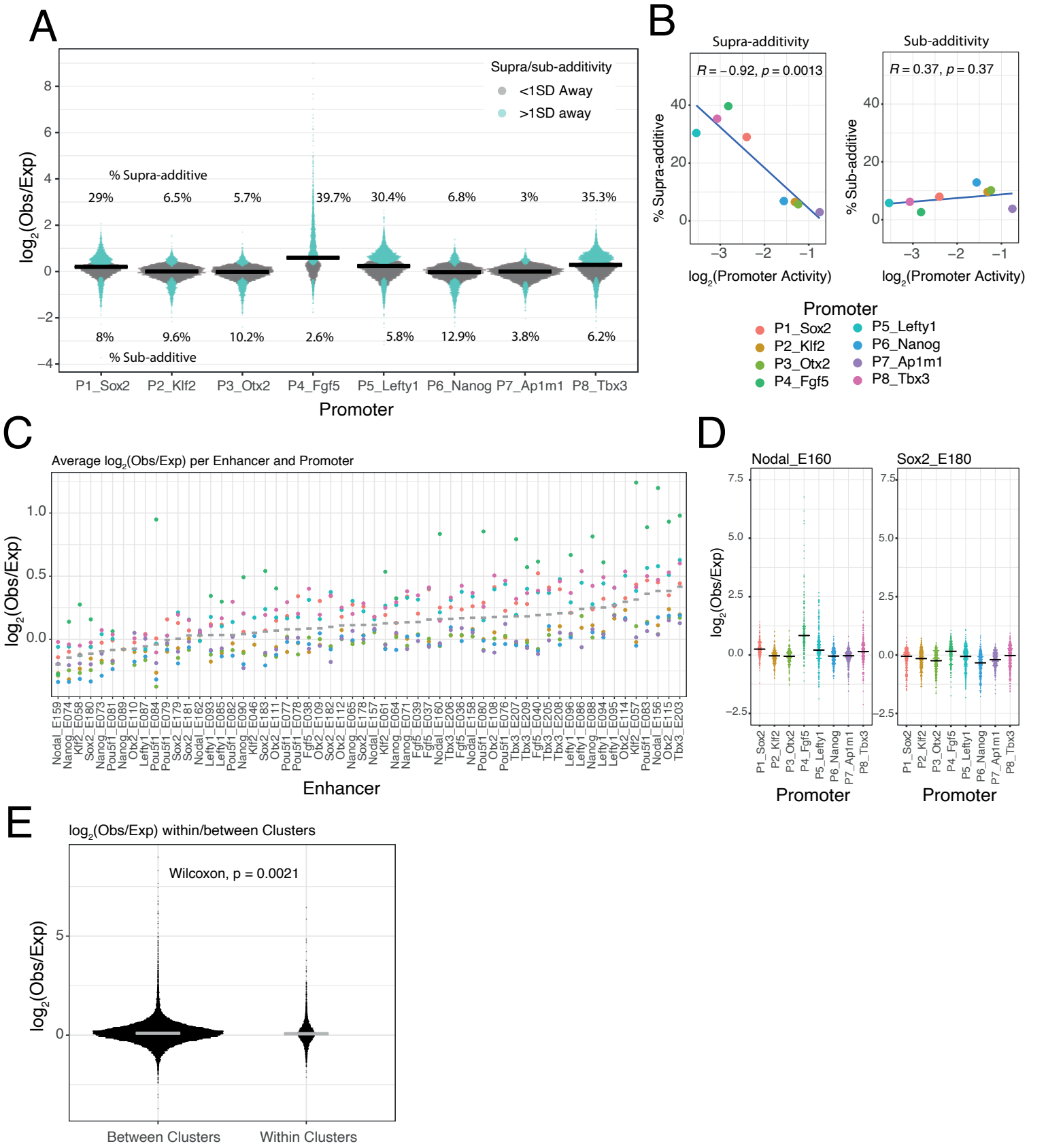



**Figure 4.** Supra and sub-additive behaviours of enhancer combinations. **A)** Distributions of $\log_2$ observed activities over expected additive activities ratios of enhancer-enhancer combinations across promoters. Colored in turquoise are supra and sub-additive combinations for which the observed activity is more than 1 standard deviation away from the expected activity. Horizontal bars represent the median of each distribution. Numbers on the top part are the percentage of supra-additive combinations for each promoter. Numbers on the lower part are the percentage of sub-additive combinations for each promoter. **B)** Relationship between the percentages of supra and sub-additive enhancer-enhancer combinations and promoter control-control baselines. Blue lines are the linear fit of the data. R is Pearson's correlation.**C)** Average supra or sub-additive behaviour of each single enhancer across enhancer-enhancer combinations for each promoter. Each dot represents the median $\log_2$ observed over expected for all enhancer-enhancer combinations of a single enhancer and a particular promoter. Grey bars represent the median of each distribution. **D)** For 2 example enhancers, distribution of $\log_2$ observed over expected ratios for combinations of that enhancer with any other enhancer and promoter. Horizontal black bars represent the median of the distribution. E) Distribution of $\log_2$ observed over expected ratios for enhancer-enhancer pairs from the same enhancer cluster (within clusters) or from different enhancer clusters (between clusters).The p-value results from comparing both distributions using a Wilcoxon test. Horizontal grey bars represent the median of each distibution. In all panels Obs/exp refers to observed activity over expected additive activity.

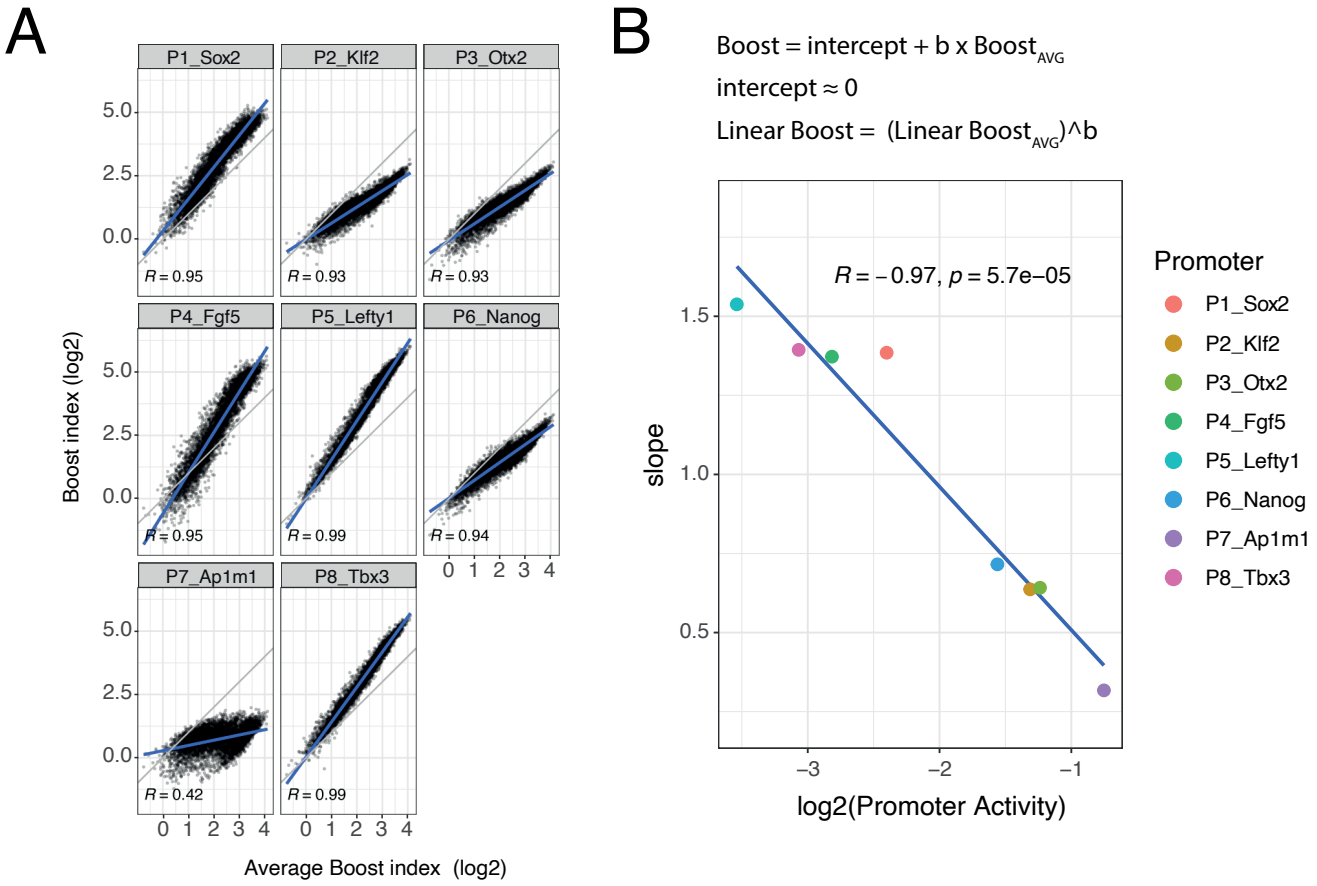
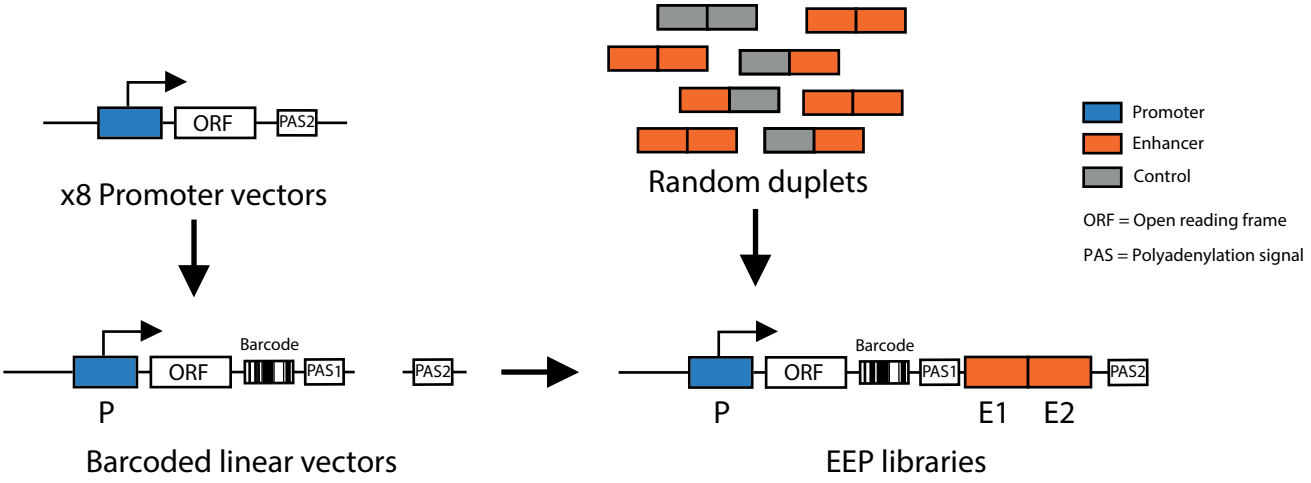
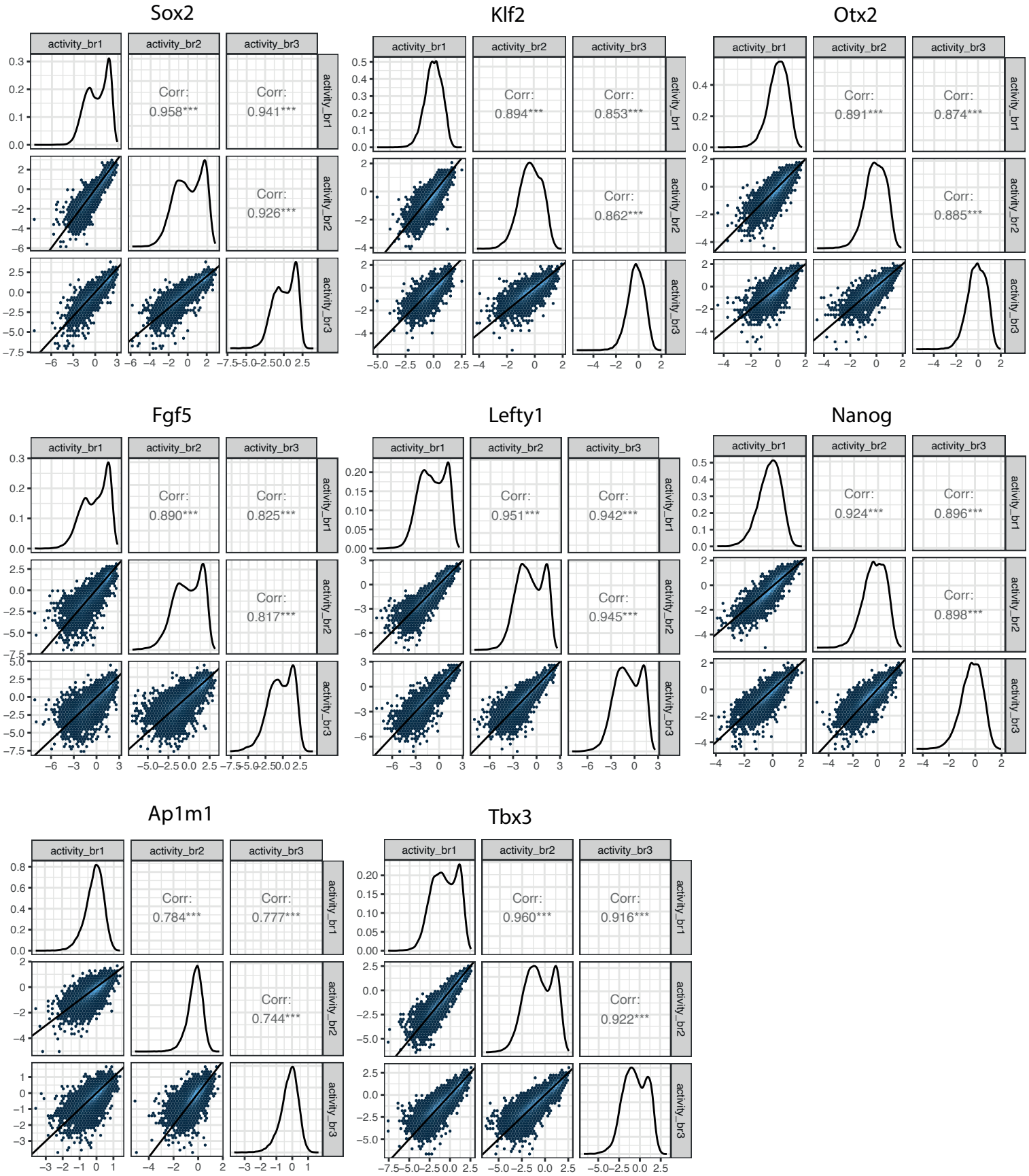


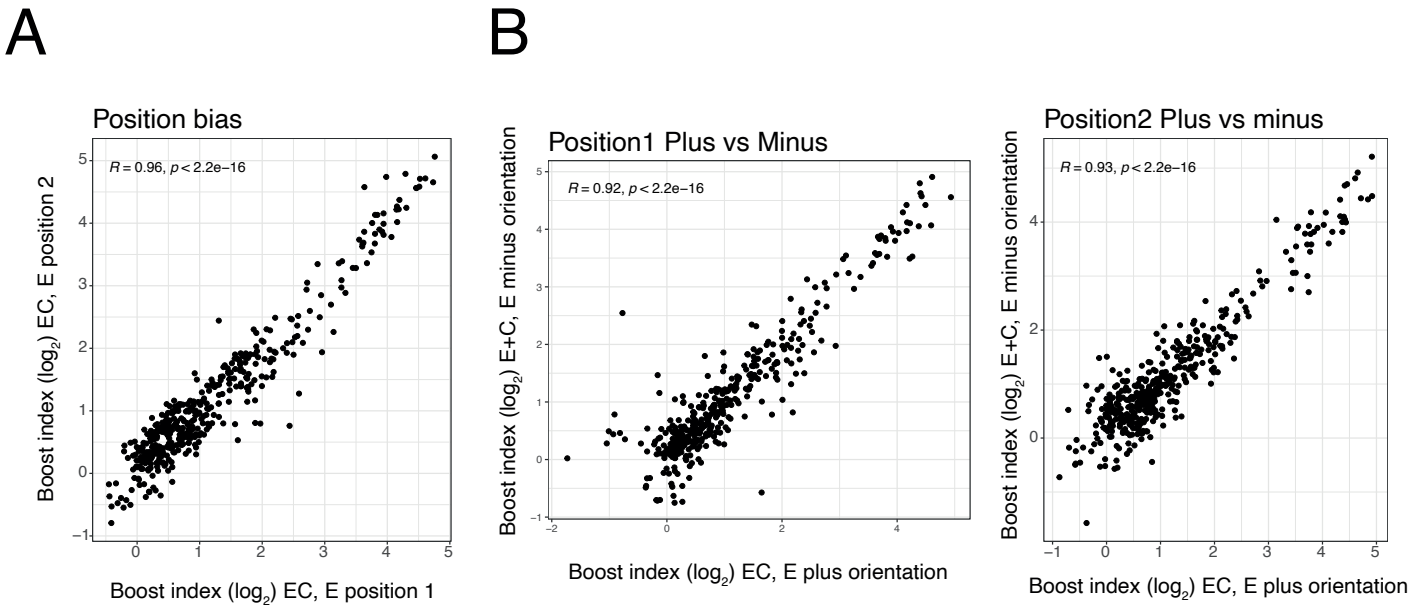

**Figure 5.** Non-linear responses of promoters to enhancer-enhancer combinations. **A)** Relationship between observed boost indices and average boost index across promoters for all shared enhancer-enhancer combinations. Blue lines represent the linear fit of the data. **B)** Relationship between the slopes extracted from the linear fits in **A** and the baseline promoter activities derived from the control-control combinations. The formulae depict the relationship between the average boost indices and the observed boost indices of each promoter through the extracted slopes. For both panels R is Pearson's correlation.



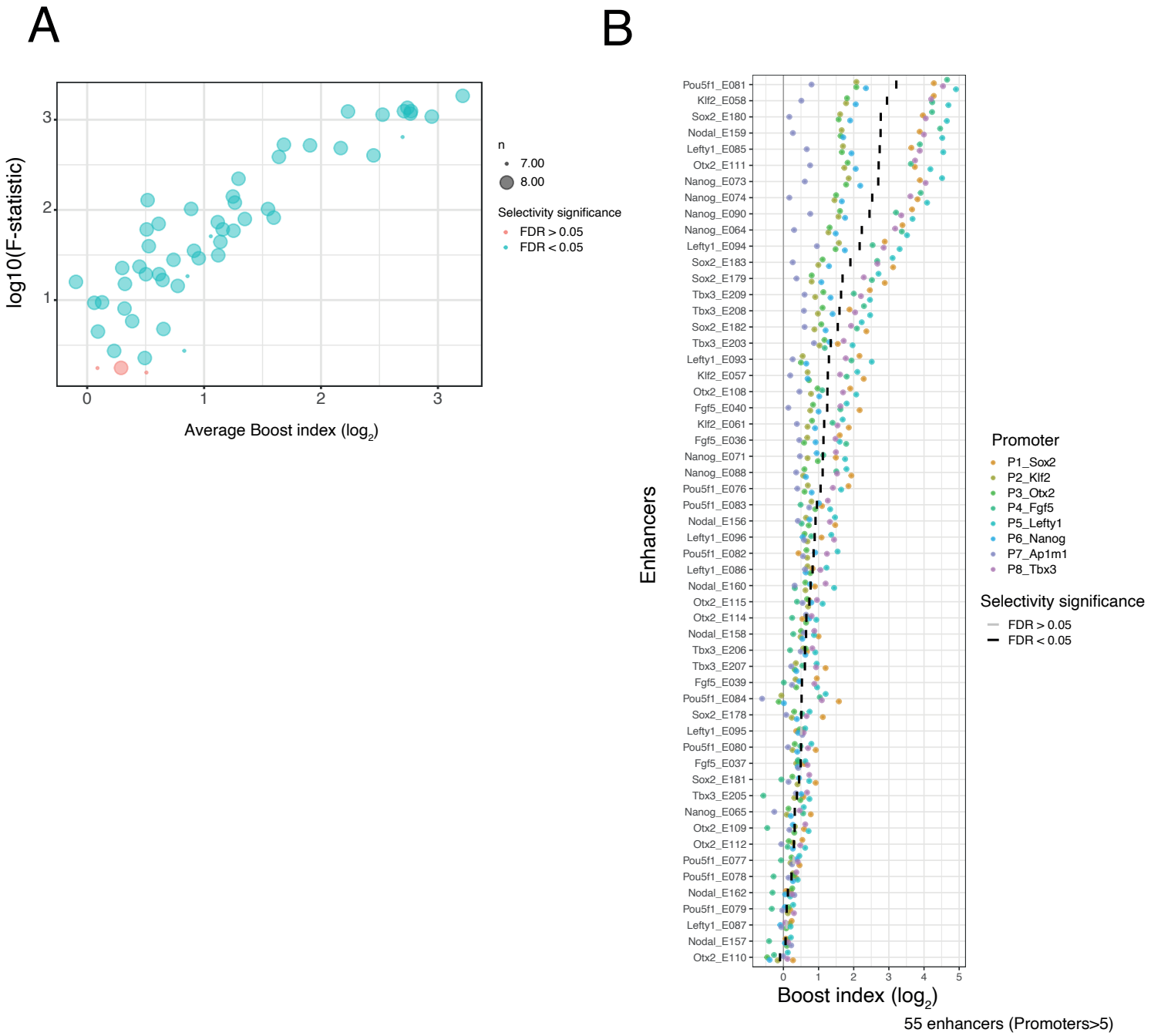

Supplementary Figure 1. Schematic of cloning strategy to generate enhancer-enhancer-promoter libraries. First the 8 different promoters were cloned into a reporter vector. Each of the 8 vectors was barcoded and linearised, random enhancer-enhancer, enhancer-control and control-control combinations were cloned downstream of the barcoded reporter.



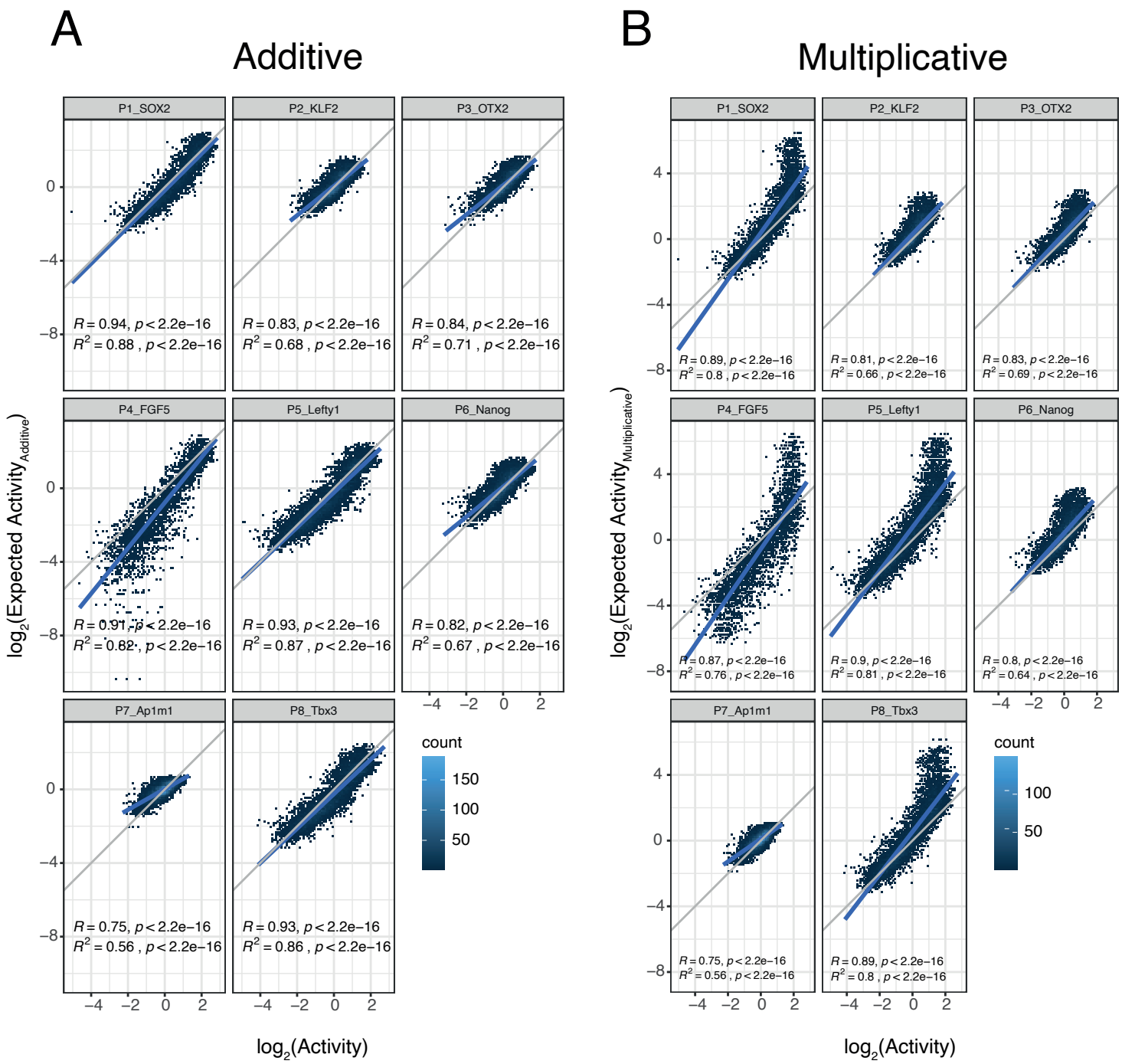

**Supplementary Figure 2.** Reproducibility of experimental data. matrix of replicate correlations for all 8 EEP libraries. Lower left panels represent the 2 dimensional density plots of replicate-replicate activity correlations. Middle panels represent the 1 dimensional density plot of each replicate. Upper right panels are the Pearson's correlation of the mirror lower left panels.



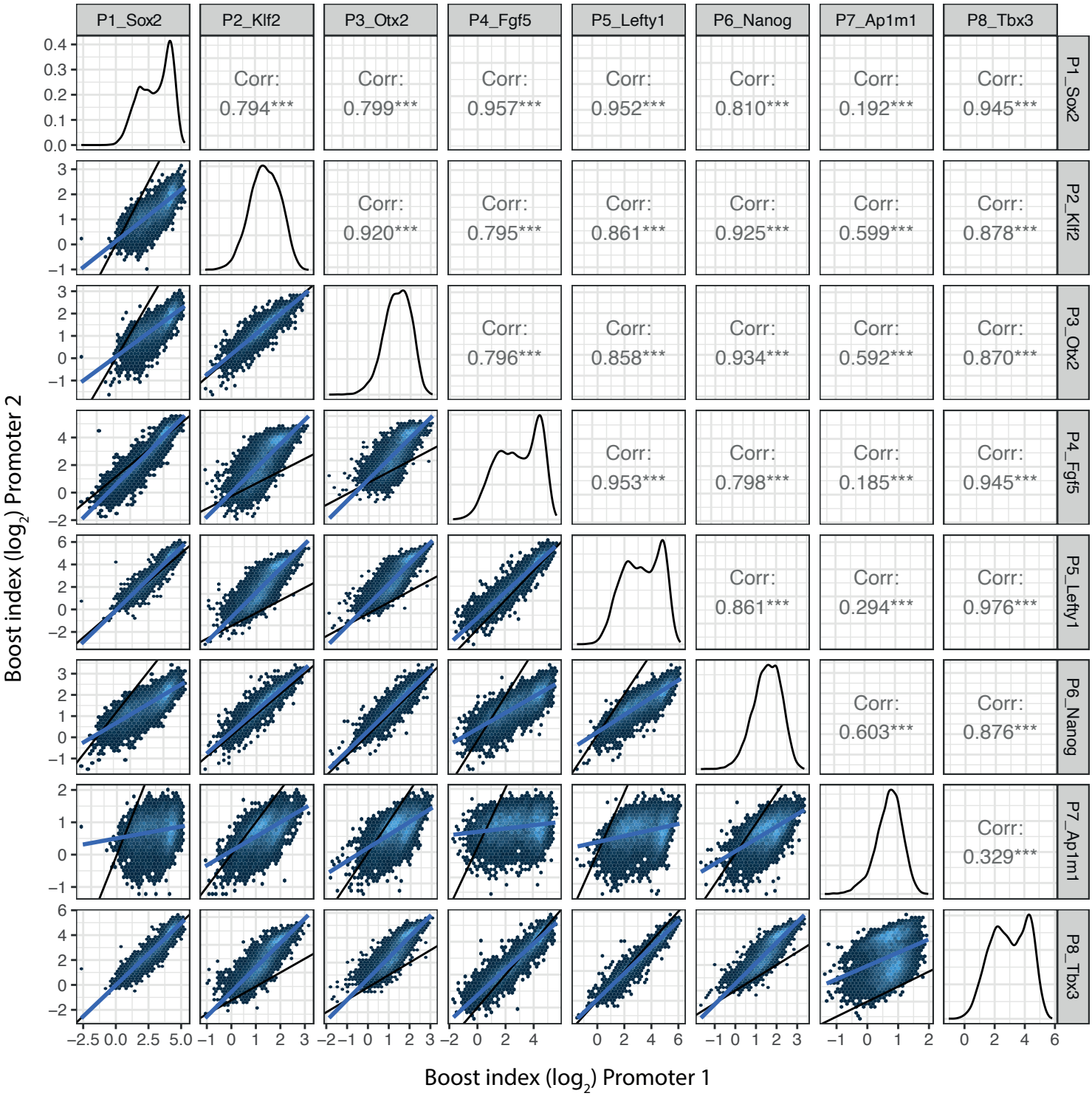

**Supplementary Figure 3.** Position and orientation bias. **A)** Relationship between single enhancer boost indices across all EC combinations in position 1 versus in position 2 (See Figure 1). **B)** Relationship between single enhancer boost indices across all EC combinations plus versus minus orientation, in position 1(left) and in position 2 (right). For both panels each dot is one enhancer with one promoter. R is Pearson's correlation.




**Supplementary Figure 4.** Selectivity of single enhancers. **A)** Relationship between average boost index of each single enhancer across all promoters and the F-statistic of a Welch F-test for that enhancer across all promoters. Colors indicate significance at a 5% FDR. **B)** Distributions of single enhancer boost indices for each enhancer across promoters. Each dot is one enhancer with one promoter and the boost index is the average boost index across all enhancer-control combinations. Colors of each point represents promoter identity. Vertical bars indicate the median boost index, and their grey or black shading indicates significance according to the Welch F-test at a 5% FDR.




**Supplementary Figure 5.** Additivity versus multiplicativity for all promoters. Relationship between observed and expected activities (additive in **A**, multuplicative in **B**) for all enhancer-enhancer combinations for the all promoters. The blue lines represent the linear fit of the data. In all panels R and $R^2$ are based on Pearson's correlation.




**Supplementary Figure 6.** Promoter promoter Boost index correlations for all shared enhancer-enhancer combinations. Lower left panels represent the 2 dimensional density plots of promoter-promoter boost index correlations. Middle panels represent the 1 dismensional density plot of each promoter. Upper right panels are the Pearson's correlation of the mirror lower left panels. Blue lines are the linear fit of the data.

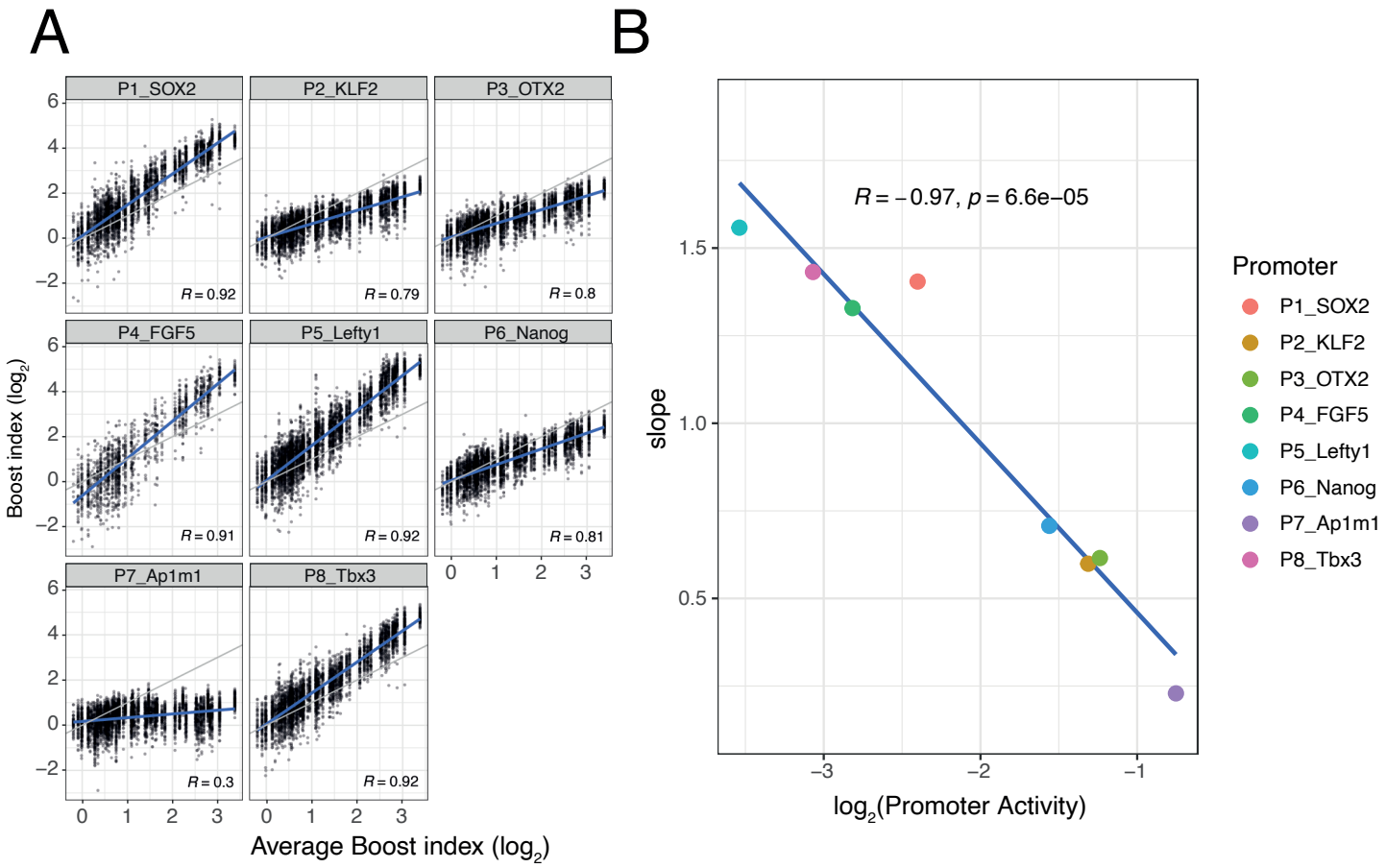



**Figure S7.** (related to figure 5) Non-linear responses of promoters to enhancer-control (single enhancers) combinations. **A)** Relationship between observed boost indices of single enhancers and average boost index across promoters for all shared single enhancer-control combinations. Blue lines represent the linear fit of the data. **B)** Relationship between the slopes extracted from the linear fits in **A** and the baseline promoter activities derived from the control-control combinations. For both panels R is Pearson's correlation.