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

Authors: Miguel Martinez-Ara[1], Federico Comoglio[1,2], Bas van Steensel[1,*]

[1]Division of Gene Regulation and Oncode Institute, Netherlands Cancer Institute, Amsterdam, the Netherlands
[2]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 approximately following a power-law, but the exponent of this response varies between promoters. 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/synergistically (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/redundantly (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

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

146 averaged activities, all developmental promoter libraries showed high correlations
147 between replicates (Pearson's R = 0.82 to 0.96) (Figure S2). The housekeeping promoter
148 library showed somewhat lower correlations between replicates (Pearson's R = 0.74 to
149 0.78), which is most likely due to a smaller dynamic range. For further analyses, we then
150 averaged the biological replicates for each promoter.
151
152 **Effects of single enhancers**
153 *Calculation of boost indices*. In order to quantify the activation by each enhancer we first
154 calculated the baseline activity of each promoter, defined as the median activity across
155 all control-control combinations. Then, for each individual enhancer-promoter
156 combination, we calculated the median activity across all enhancer-control
157 combinations. For each single enhancer we then calculated a boost index, which is the
158 $\log_2$ ratio of its median activity across enhancer-control combinations over the promoter
159 baseline across control-control combinations Figure 2A). We observed that there was
160 little position and orientation dependence of the enhancer effects (Figure S3). We
161 therefore averaged the boost indices of single enhancers across positions and
162 orientations for further analyses.
163 *Selected DHSs are active enhancers.* The single enhancer boost indices show that
164 the majority of the sequences that we selected indeed act as enhancers, although the
165 boost indices varied up to ~5 $\log_2$ units (Figure 2A-B). All tested enhancers showed
166 significant activation of at least one promoter (at an estimated false discovery rate (FDR)
167 < 0.01) (Figure 2B, D). However, the dynamic ranges and effects of each enhancer vary
168 per promoter, with the boost indices being particularly smaller for the housekeeping
169 promoter (*Ap1m1*) (Figure 2A, C). We conclude that most DHSs that we selected from
170 the enhancer clusters act individually as enhancers.
171 *Promoter selectivity of enhancers.* In our previous study of a large set of E-P pairs
172 we found that the majority of tested enhancers exhibited significant preference for
173 certain promoters (Martinez-Ara et al., 2022). Applying the same statistical analysis to
174 the current enhancer set confirmed this finding: the variability of enhancer effects across
175 promoters are generally larger than what is expected from experimental noise (Figure
176 S4). This indicates that the single enhancers tested here show significant promoter
177 selectivity.
178
179 **Enhancers generally combine near-additively**
180 *Enhancer-enhancer combinations are generally stronger than single enhancers***.** We then
181 focused on the effects of enhancer-enhancer (EE) combinations. In the boost index

182 matrix that covers all combinations for a single promoter, we observed that generally EE
183 combinations were more active than EC combinations (**Figure 3A)**. Indeed, quantitative
184 analysis showed that EE combinations tend to induce stronger activation than the single
185 enhancers (EC) (**Figure 3B-C**), although some saturation appears to occur with the
186 strongest enhancers (**Figure 3C**).

187 *Additive model fits better than multiplicative model.* We then asked how the
188 effects of two single enhancers are integrated in the enhancer-enhancer (EE)
189 combinations. We considered two simple models: additivity and multiplicativity. For
190 each EE combination we calculated the expected effect according to each of these
191 models based on the single enhancer measurements, and then compared the observed
192 and expected activities (additive in **Figure 3D-E,** multiplicative in **Figure 3F,** see
193 Methods). Interestingly, for most of the promoters the additive model consistently
194 showed a better fit to the measured data than the multiplicative model (**Figure 3E-F,**
195 **Figure S5**). In particular, the multiplicative model tended to strongly overestimate
196 activities at the higher range of activities (most active enhancers) for most promoters.
197 When focusing on EE combinations for which the multiplicative and additive expected
198 values differ more than 0.5 $\log_2$ units (to limit random noise effects), we found that the
199 multiplicative model matched the observed values better for only 0-17% EE
200 combinations across the seven promoters. Only for the *Ap1m1* promoter the
201 multiplicative and additive models were nearly indistinguishable. This may be due to the
202 low dynamic range of activities observed with this promoter. We note that for the *Fgf5*
203 promoter in the low activity ranges the expected activities were consistently
204 underestimated for both models. It is likely that this is an inaccuracy due to the relative
205 sparseness of the *Fgf5* data (the baseline activity of this promoter was estimated from
206 only 23 measured control-control combinations, compared to 219-477 combinations for
207 the other promoters). In summary, we conclude that a simple additive model fits the
208 observed data better than a multiplicative model.
209

210 **Supra-additive activation is rare and promoter dependent.**

211 *Low frequency of supra-additive enhancer-enhancer interactions.* Most enhancer-
212 enhancer combinations show only minor deviations from the additive model (**Figure S5**),
213 as illustrated by the distributions of observed/expected ratios (**Figure 4A**). By applying a
214 simple error model to estimate the noise in the expected EE activities according to the
215 additive model (**Figure 3D**, see Methods) we found that 74.5% of the measured EE
216 activities lies within one standard deviation (SD) of the predicted activities (**Figure 4A**)
217 and 94% lies within two standard deviations. This underscores that most of the tested

218 enhancers combine near-additively, and that both sub-additive and supra-additive
219 effects are generally rare or weak.

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

231 *EE synergies are relatively infrequent and often promoter-dependent***.** Next, we
232 investigated whether supra-additivity involves a specific subset of enhancers. Visual
233 inspection identified a few individual enhancers that exhibited supra-additive
234 interactions with most other enhancers, but often in the context of particular promoters.
235 An example is Nodal_E160 in combination with the *Fgf5* promoter (**Figure 4D**, left panel).
236 Other enhancers showed a diversity of supra-additive activities. Out of the 58 tested
237 enhancers 32.7% (19/58) showed consistent supra-additivity in at least one promoter
238 context (>0.5 $\log_2$(observed/expected), mean over all enhancers in at least one promoter
239 context), and only 10.3% (6/58) showed consistent supra-additivity in two or more
240 promoter contexts (**Figure 4C**). However, most enhancers behave like Sox2_E180 (**Figure**
241 **4D**, right panel), showing small deviations from the additive behaviour with most other
242 enhancers, and with all of the tested promoters (**Figure 4C**).

243 *No evidence for locus-specific EE supra-additivity*. We then asked whether
244 enhancer pairs derived from the same genomic locus have a higher propensity towards
245 supra-additivity than those not derived from the same locus. This could point towards
246 co-evolution of enhancers that jointly control a gene. However, we found that EE
247 combinations from the same locus showed on average not a higher observed/expected
248 ratio (according to the additive model) than EE combinations from different loci; in fact,
249 this ratio is even slightly lower (**Figure 4E)**. Thus, generally there appears to be no
250 preferential synergy among enhancers that are located in the same genomic locus.
251

252 **Enhancer effects scale non-linearly across developmental promoters**
253 *Enhancer effects correlate well across developmental promoters*. Prompted by our
254 observations that some supra-additive EE interactions appear to be more pronounced
255 in the context of specific promoters, we decided to systematically compare the boost
256 indices of all EE pairs between the eight promoters. Strikingly, all seven developmental
257 promoters showed strong pairwise correlations of their boost indices (Pearson's R =
258 0.79 to 0.98), whereas the housekeeping promoter (*Ap1m1*) correlated much less with
259 each of the developmental promoters (Pearson's R = 0.19 to 0.60; **Figure S6**).
260 *Slopes of promoter boost responses differ*. Although the developmental
261 promoters showed strong correlations, we were surprised to notice that the slopes of
262 these relationships often deviated from 1. To further explore this, we plotted the boost
263 indices of each promoter against the average boost indices across all eight promoters
264 (**Figure 5A**). All of the developmental promoters correlated strongly with this average
265 pattern (Pearson's R= 0.93 to 0.99) but this was not the case for the housekeeping
266 promoter (Pearson's R= 0.42). The slopes relative to the averaged promoters ranged
267 from 1.54-0.32, i.e., about 5-fold, across the developmental promoters. These results
268 imply that the developmental promoters differ in the scaling (i.e., steepness) of their
269 responses to the set of EE pairs, even though the pattern of their responses is highly
270 similar. Interestingly, the relative slopes negatively correlate with the baseline activity
271 of these promoters (**Figure 5B**). Thus, promoters with low intrinsic activity respond
272 more steeply to changes in combined EE activity than promoters with high intrinsic
273 activity.
274
275 **DISCUSSION**
276
277 *A large-scale survey of EEP combinations.* Here, we employed a MPRA approach to
278 measure three-way regulatory interactions in ~69,000 EEP combinations involving 8
279 promoters and 59 enhancers. This is of a much larger scale than previous efforts to
280 elucidate the combinatorial interplay among enhancers, which typically focused on a
281 handful of enhancers in single genomic loci (Bothma et al., 2015; Hay et al., 2016; Shin
282 et al., 2016; Dukler et al., 2016; Joo et al., 2016; Lam et al., 2015; Carleton et al., 2017;
283 Moorthy et al., 2017; Osterwalder et al., 2018; Thomas et al., 2021; Lin et al., 2022; Brosh
284 et al., 2023). These previous studies pointed to a diversity of functional enhancer-
285 enhancer (EE) interactions, including redundancy, additivity and synergy. Our study also
286 uncovered a broad spectrum of functional interactions, but simple near-additive effects
287 of EE combinations were predominant, while supra-additive interactions occurred at a
288 lower frequency and were dependent on the specific enhancer and promoter context.

289 *Promoters "interpret" and scale EE interplay.* An important finding is that
290 promoters can have substantial impact on supra-additivity of enhancer-enhancer
291 interactions. Promoters may thus be seen as "interpreters" of EE combinations. This
292 underscores the importance of studying triple combinations of elements, and perhaps
293 in the future even more complex settings. A salient finding is that the different promoters
294 show different degrees of responsiveness to EE pairs. The responses of the promoters
295 follow roughly a power-law function of which the exponent is dependent on the
296 promoter (**Figures 5A and S6**, which are on log-log scale). Because it is not possible to
297 define an absolute measure of enhancer strength, we cannot determine the absolute
298 values of the exponents of the promoter power-law responses. This would be of interest
299 because exponents <1 would lead to "diminishing returns" with higher EE activities,
300 while exponents >1 would correspond to increasing non-linear amplification with higher
301 EE activities. The apparent saturation observable at high enhancer activities in **Figure 3C**
302 suggests that the exponents are most likely below 1.
303 *Episomal MPRAs vs genomic context*. The reporters in our MPRA were not
304 integrated into the genome. On the one hand, this ensures that all EEP combinations are
305 tested under similar conditions, without confounding effects of variable chromatin
306 contexts. On the other hand, we cannot rule out that the results are not fully
307 representative of a natural chromatin environment. However, other studies have
308 indicated that transiently transfected reporters generally behave quite similar, although
309 not identically, to integrated reporters (Inoue & Ahituv, 2015; Klein et al., 2020), and
310 another study indicated that chromatin context can scale transcriptional activity of
311 inserted reporters in a simple linear manner that is largely independent of the inserted
312 sequence (Maricque et al., 2018).
313 *Results may be biased by model system.* We cannot rule out that our results are
314 biased by our selection of enhancers and promoters, which was mostly focused on genes
315 linked to pluripotency in mouse ESCs. We may have missed more prominent synergistic
316 or other non-additive effects that may arise in different cell types or among different
317 classes of enhancers and promoters. However, the protocol that we developed for three-
318 way MPRAs should be generally applicable to any set of enhancers and promoters.
319 *Spacing of elements*. It should also be considered that in our MPRA the enhancer
320 and promoter elements are placed in fixed positions and close to one another, which
321 does not recapitulate their relative positioning in the genome. Genomic distance and
322 chromatin contexts could act as thresholds for activation, or turn additive interactions
323 into non-linear relationships. For instance, the ability of a *Sox2* enhancer to activate its
324 cognate promoter was found to depend on their distance, following approximately a

power law (Zuin et al., 2022). The rules for EE interplay may also depend on the EE distance or genomic context, as suggested by a recent report indicating that distant pairs of enhancers are more often synergistic than closely spaced enhancers, which tend to act additively (Lin et al., 2022).

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

Relevance of findings and context. The findings presented here offer insights into two aspects of the gene regulatory process – interplay between enhancers, and promoter integration of the resulting signals – in a more quantitative manner. The near-additivity of enhancers will help us understand how different enhancers may be more or less relevant in different loci. The non-linearity of promoter responsiveness and the effects of promoters on enhancer-enhancer interactions illustrates how some promoters may be more sensitive to activating signals. Thus, it will be useful to quantify promoter responsiveness genome-wide to understand which genes may be more or less sensitive than others to environmental cues or to non-coding mutations. These aspects may be especially relevant for developmentally regulated transcription factors and their target genes, like the ones here studied, as they can be very sensitive to dosage changes (Naqvi et al., 2023).

METHODS

Cell culture

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

Selection of promoters, enhancers and controls.

For the construction of the libraries we selected seven developmental genes from the mESC pluripotency regulatory network (Acampora et al., 2013; Dunn et al., 2014; D.-K. Kim et al., 2014; Thomas et al., 2021), *Klf2, Sox2, Nanog, Otx2, Lefty1, Ffg5, Tbx3;* and the housekeeping gene *Ap1m1* (Hounkpe et al., 2021), which is neighbouring the Klf2 gene**.** The promoters of these genes were identified based on transcript and TSS annotation (Frankish et al., 2019) and overlaps with DNAse hypersensitivity sites (DHS) (Joshi et al., 2015) as previously described (Martinez-Ara et al., 2022). We then selected sequences that approximately cover -350 to 50bp around the TSS.

The clusters of enhancers were selected in as follows: 1. Clusters of enhancers that were known to regulate the developmental genes (*Nanog, Sox2, Fgf5* clusters) (Blinka et al., 2016; Brosh et al., 2023; Thomas et al., 2021; Zuin et al., 2022); 2. Putative clusters of enhancers in proximity of one of the developmental genes and that overlapped with previously defined super-enhancers (*Lefty1, Tbx3, Otx2* clusters) (Hnisz et al., 2013; Whyte et al., 2013); 3. A cluster of putative enhancers previously identified by us (Martinez-Ara et al., 2022) to act as enhancers in an MPRA (*Klf2* cluster). 4. Clusters of enhancers identified as putative enhancers of other mESC pluripotency genes for which we did not include their promoters (*Pou5f1* and *Nodal* clusters) and overlapping with previously defined super-enhancers (Hnisz et al., 2013, 2015; Whyte et al., 2013).

Clusters of enhancers were divided into single enhancers based on DNAse hypersensitivity sites (DHS) (Joshi et al., 2015). These single enhancers were then resized to ~450bp from the center of DNAse hypersensitivity peaks as previously described (Martinez-Ara et al., 2022).

390 For the library construction we included 20 450 bp long randomized control
391 sequences as previously described (Martinez-Ara et al., 2022). Fifteen of these were
392 based on five DHS peak sequences that were randomly scrambled. The other five
393 controls were randomly generated sequences with the similar GC content as the DHS
394 peak sequences. These control sequences were ordered as synthetic DNA.
395 **Supplementary Table 2** contains the sequences of these controls**.**
396 For single enhancers and promoters we designed PCR primers against the 50 bp
397 ends of each selected element using BatchPrimer3 v1.0 (You et al., 2008). This yielded
398 PCR products of ~400 bp for each element. **Supplementary Table 1** lists the coordinates
399 of all individual enhancers and promoters. **Supplementary Table 4** lists primer
400 sequences.
401
402 **EEP Library generation**
403 The vector for library construction was previously used for the downstream reporter
404 assay in our previous study (Martinez-Ara et al., 2022). This vector is based on a pSMART
405 backbone. It contains a green fluorescent protein (GFP) open reading frame followed by
406 a barcode, and a psiCheck polyadenylation signal (PAS) introduced during barcoding,
407 followed by the cloning site for inserts and a triple polyadenylation site
408 (SV40+bGH+psiCheckPAS).
409 Each of the selected eight promoters were amplified by PCR and individually
410 inserted by Gibson assembly (#E2611S; New England Biolabs) into the reporter vector.
411 Each construct was transformed into standard 5-alpha competent bacteria (#C2987;
412 NEB) grown overnight in in 500 ml of standard Luria Broth (LB) with 50 mg/ml of
413 kanamycin and purified. Then each promoter sequence was verified by Sanger
414 sequencing.
415 Each of the promoter vectors was then barcoded. Similar to what we described
416 previously (Martinez-Ara et al., 2022)**,** we digested 10 µg of each vector with AvrII
417 (#ER1561; Thermo Fischer) and XcmI (#R0533; NEB) and performed a gel-purification.
418 Barcodes were amplified in 10 separate 100ul PCR reactions using 5 µl of 10 mM primer
419 275JvA, 5 ml of 10 mM primer 465JvA and 1 µl of 0.1 mM template 274JvA (see
420 **Supplementary Table 4** for sequences). 14 PCR cycles were performed using MyTaq Red
421 Mix (#BIO-25043; Bioline), yielding 30 µg of barcodes. We purified barcodes by phenol-
422 chloroform extraction and isopropanol precipitation. Barcodes were digested overnight
423 with 80 units of NheI (#R0131S; NEB) and purified by magnetic bead purification (#CPCR-
424 0050; CleanNA). Each promoter vector and barcodes were then ligated in one 100 µl
425 reaction. In each reaction we used 3 µg of digested vector and 2.7 µg digested barcodes,

426 20 units NheI (#R0131S; NEB), 20 units AvrII, 10 ml of 103 CutSmart buffer, 10 µl of 10
427 mM ATP, 10 units T4 DNA ligase (#10799009001 Roche). We performed a cycle-ligation
428 of six cycles (10 min at 22 °C and 10 min at 37 °C), followed by 20 min heat-inactivation
429 at 80 °C. Linear barcoded vectors were purified by magnetic beads and digested for 3h
430 with 40 units of XcmI (#R0533S; NEB). Finally barcoded vectors were purified again by
431 gel-purification.
432 Single enhancers were amplified by PCR from E14tg2a mESC genomic DNA, then
433 purified with magnetic beads (#CPCR-0050, Clean NA) and quantified by Qubit using
434 dsDNA High sensitivity Qubit kit (#Q33231; invitrogen). Single enhancers and random
435 controls were combined in approximately equimolar manner. This pool of elements was
436 end-repaired using End-It DNA End-Repair Kit (#ER0720; Epicentre) and self-ligated using
437 Fast-link ligase (LK0750H; Lucigen). Duplets of 800bp-1000bp were excised from agarose
438 gel, purified (BIO-52059; Bioline), and A-tailed using Klenow HC 3´ → 5´ exo (#M0212L;
439 NEB).
440 The pool of duplets was ligated to each of the barcoded promoter vectors using
441 Takara ligation kit version 2.1 (#6022; Takara). Ligation products were purified using
442 magnetic bead purification and 2 µl of the ligation were electroporated into 20 µl of
443 electrocompetent e. cloni 10G supreme (#60081-1; Lucigen). To estimate the
444 approximate complexity of the resulting libraries, dilutions of the electroporated
445 bacteria were plated for overnight culture and colonies were counted. We aimed at a
446 minimum of 2 million complexity per library in order to have sufficient expected
447 representation of all possible combinations.
448 When the estimated complexity of the libraries was not high enough, we
449 amplified the ligation products with primers aligning to the vector. This generated PCR
450 products consisting of the duplets pool and a vector overhang. These PCR products were
451 used for Gibson assembly with the barcoded promoters (#E2611S; NEB). Gibson
452 assembly reactions were purified with magnetic beads and 2 µl of the reaction were
453 electroporated into 20 µl of electrocompetent e. cloni 10G supreme (#60081-1; Lucigen).
454 Each library was grown overnight in 500 ml of standard Luria Broth (LB) with 50
455 mg/ml of kanamycin and purified using a maxiprep kit (K210016, Invitrogen).
456
457 **Characterization of libraries by iPCR and sequencing**
458 Barcode to insert (enhancer and control duplets) combinations were identified by
459 inverse PCR (iPCR) and Illumina sequencing as described before(van Arensbergen et al.,
460 2017)**.** In brief, each library was digested overnight with I-CeuI (#R0699S, NEB), barcode-
461 insert fragments were then circularized, remaining linear fragments digested, and

462 barcode-insert fragments were linearized again with I-SceI (#R0694S; NEB). Barcode-
463 insert fragments were then amplified by PCR with Illumina adapters and sequenced on
464 an Illumina NextSeq 550 platform using 150 bp paired-end sequencing. Each library was
465 processed separately and mixed together for sequencing.
466
467 **Transfection of libraries**
468 Each EEP library was transfected separately into E14tg2a mESCs. Per library 20 million
469 cells were nucleofected (5µgs of library and 5million cells per cuvette, 5 cuvettes) using
470 Amaxa nucleofector II, program A-30, and Mouse Embryonic Stem Cell NucleofectorTM
471 Kit (#VPH-1001, Lonza). Three biological replicates were performed per library on
472 different days. Cells were collected 24 hours after transfection in 5 ml of TRIsure (#BIO-
473 38032; Bioline) and frozen at -80 °C until further processing.
474
475 **cDNA sequencing**
476 RNA was extracted and prepared for sequencing as described for the Downstream Assay
477 in our previous study (Martinez-Ara et al., 2022). From each sample, the aqueous phase
478 containing the total RNA of the TRIsure solution was extracted and purified on an RNA
479 extraction column (#K0732, Thermo Scientific). Total RNA was digested with 10 units of
480 DNase I (#04716728001; Roche) for 30 minutes in 10-8 10 µl reactions containing 5 µg
481 of RNA. DNase I was inactivated by addition of 1 µl of 25 mM EDTA and incubation at
482 70°C for 10 min.
483 cDNA was produced in 20 µl reactions (1 per 10 µDNAse reaction) using Maxima
484 Reverse transcriptase (#EP0743; ThermoFisher Scientific) and a gene specific (targeting
485 the GFP reporter) primer (JvA304, see **Supplementary table 4** for sequence). dNTPs and
486 the primer were mixed with the DNAse digested RNA and incubated at 65°C for 5
487 minutes. The RT buffer, enzyme and RNAse inhibitor were added and the reaction was
488 incubated at 50 °C for 1 hour. cDNA was then amplified by two nested PCRs to make it
489 strand specific. The first PCR reaction uses (index variants of 285JvA containing the S2,
490 index and p7 adaptor) and 305JvA (targeting the adapter introduced by 304JvA). Each 20
491 µl RT reaction was amplified in a 100µl PCR reaction with MyTaq Red mix (#BIO-25043;
492 Bioline). The second PCR reaction was performed using 10 µl of the product of the
493 previous reaction in a 100 µl reaction using the same index variant primer and index
494 variants of 437JvA (containing the S1, index and p5 adaptor) (see **Supplementary table**
495 **4** for primer sequences). The first PCR was run for 10 cycles and the second one for 8
496 cycles using the recommended Mytaq Red mix conditions (for both PCRs: 1 min 96 °C,
497 then each cycle 15 s 96 °C, 15 s 60 °C, 15 s 72 °C). PCR products were purified with

magnetic beads and mixed for sequencing in an Illumina NovaSeq 6000 platform using 100bp single end reads.

Plasmid DNA (pDNA) sequencing

Plasmid libraries were processed for sequencing as described (Martinez-Ara et al., 2022) but using dual indexing for sequencing on a NovaSeq 6000 platform with 100 bp single end reads. 1 µg of each plasmid was digested with I-SceI in order to linearise the plasmid. Barcodes were amplified by PCR for 9 cycles from 50 ng of material. Primers and reaction conditions were the same as in the amplification of cDNA. PCR products were purified using magnetic beads and mixed for sequencing.

COMPUTATIONAL METHODS

cDNA and pDNA Data pre-processing and linking of barcodes to duplets was performed using a custom snakemake pipeline (Köster & Rahmann, 2012). All other data analyses and quantifications were performed in R (R Core Team, 2022). Figures were generated using ggplot2 (Wickham, 2016). The snakemake pipeline and all the scripts used in this publication are available in a GitHub repository (https://github.com/vansteensellab/EEPCombinations).

Linking Barcodes to duplet inserts

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

cDNA and pDNA data pre-processing

For each cDNA and pDNA replicate of each library, barcodes were extracted from single-end reads using a custom Python script that matches the constant region after the

533 barcode. Barcodes were clustered using Starcode (version 1.1) (Zorita et al., 2015) to
534 remove errors from sequencing and counts were summarized.
535
536 **cDNA and pDNA data post-processing**
537 For each library cDNA and pDNA barcodes were matched to iPCR barcodes. Any barcode
538 assigned to multiple fragment identities was removed from the data. Per cDNA/pDNA
539 replicate barcode counts were normalized to the total number of barcode counts. Then
540 activity was calculated per barcode and cDNA replicate as the cDNA:pDNA normalized
541 counts ratio. Per individual duplet combination these normalized activity ratios were
542 averaged across barcodes with a minimum requirement of five barcodes and eight pDNA
543 counts per barcode. The mean activity across the three replicates was calculated as the
544 geometric mean.
545
546 **Calculation of boost indices**
547 In our previous study (Martinez-Ara et al., 2022) we observed that it was critical to use
548 control sequences to estimate the baseline activity of the promoters. For each promoter
549 we used all of the control-control combinations to estimate its baseline activity. We
550 calculated this as the median activity across all control-control combinations. Then the
551 boost index of an enhancer-enhancer or enhancer-control (single enhancer boost index)
552 combination was calculated as the $\log_2$ ratio between the observed activity of this
553 combination and the baseline activity of the promoter. Except in **Supplementary Figure**
554 **3**, boost indices were averaged across positions and orientations. In **Figure 2a** boost
555 indices were averaged across all enhancer-control combinations for each enhancer. In
556 **Figure 3a** boost indices were averaged across orientations but not positions.
557
558 **Calculation of additive and multiplicative expected activities**
559 To calculate the expected additive and multiplicative activities we first had to determine
560 three values. The baseline activity of the promoter (A_{CCP}) was measured as the median
561 activity across all control-control combinations for each promoter. The single enhancer
562 activity estimates (A_{E1CP} and A_{CE2P}, enhancer in position1 and position 2 respectively)
563 were calculated as the median activity of all enhancer-control combinations where the
564 enhancer was found in either position 1 (A_{E1CP}) or position 2 (A_{CE2P}). Therefore
565
566 $A_{CCP} = \text{median}\{A_{C1C2P}, A_{C1C3P}, \ldots, A_{CiCjP}\}$
567 $A_{E1CP} = \text{median}\{A_{E1C1P}, A_{E1C2P}, \ldots, A_{E1CiP}\}$
568 $A_{CE2P} = \text{median}\{A_{C1E2P}, A_{C2E2P}, \ldots, A_{CiE2P}\}$

Then for each enhancer-enhancer-promoter combination the expected additive activity was calculated as the sum of the estimated activities of the single enhancers in each position minus the promoter baseline activity ($A_{E1CP} + A_{CE2P} – A_{CCP}$). The multiplicative expected activity was calculated as the product of the estimated activities of the single enhancers in each position divided by the promoter baseline activity ((A_{E1CP} * A_{CE2P})/A_{CCP}). For the expected additive activity, the standard deviation of the estimate was propagated as the square root of the sum of the variances of the single 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 cooperative 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 (DOI: ###).

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

Brosh, R., Coelho, C., Ribeiro-Dos-Santos, A. M., Ellis, G., Hogan, M. S., Ashe, H. J., Somogyi, N., Ordoñez, R., Luther, R. D., Huang, E., Boeke, J. D., & Maurano, M. T. (2023). Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. *Molecular Cell*, *83*(7), 1140-1152.e7. https://doi.org/10.1016/j.molcel.2023.02.027

Carleton, J. B., Berrett, K. C., & Gertz, J. (2017). Multiplex Enhancer Interference Reveals Collaborative Control of Gene Regulation by Estrogen Receptor α-Bound Enhancers. *Cell Systems*, *5*(4), 333-344.e5. https://doi.org/10.1016/j.cels.2017.08.011

Chang, T. H.-T., Primig, M., Hadchouel, J., Tajbakhsh, S., Rocancourt, D., Fernandez, A., Kappler, R., Scherthan, H., & Buckingham, M. (2004). An enhancer directs differential expression of the linked Mrf4 and Myf5 myogenic regulatory genes in the mouse. *Developmental Biology*, *269*(2), 595–608. https://doi.org/10.1016/j.ydbio.2004.02.013

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, *149*(6), 1233–1244. https://doi.org/10.1016/j.cell.2012.03.051

Dukler, N., Gulko, B., Huang, Y.-F., & Siepel, A. (2016). Is a super-enhancer greater than the sum of its parts? *Nature Genetics*, *49*(1), 2–3. https://doi.org/10.1038/ng.3759

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., & Smith, A. G. (2014). Defining an essential transcription factor program for naïve pluripotency. *Science (New York, N.Y.)*, *344*(6188), 1156–1160. https://doi.org/10.1126/science.1248882

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766–D773. https://doi.org/10.1093/nar/gky955

Gallego Romero, I., & Lea, A. J. (2023). Leveraging massively parallel reporter assays for evolutionary questions. *Genome Biology*, *24*(1), 26. 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

718 Grosveld, F., van Assendelft, G. B., Greaves, D. R., & Kollias, G. (1987). Position-
719 independent, high-level expression of the human beta-globin gene in transgenic
720 mice. *Cell*, *51*(6), 975–985. https://doi.org/10.1016/0092-8674(87)90584-8
721 Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L., Kassouf, M.
722 T., Marieke Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., Telenius, J., Williams, R.,
723 Rode, C., Li, P.-S., Pennacchio, L. A., Sloane-Stanley, J. A., Ayyub, H., Butler, S.,
724 Sauka-Spengler, T., … Higgs, D. R. (2016). Genetic dissection of the α-globin
725 super-enhancer in vivo. *Nature Genetics*, *48*(8), 895–903.
726 https://doi.org/10.1038/ng.3605
727 Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., &
728 Young, R. A. (2013). Super-enhancers in the control of cell identity and disease.
729 *Cell*, *155*(4), 934–947. https://doi.org/10.1016/j.cell.2013.09.053
730 Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E.,
731 & Young, R. A. (2015). Convergence of developmental and oncogenic signaling
732 pathways at transcriptional super-enhancers. *Molecular Cell*, *58*(2), 362–370.
733 https://doi.org/10.1016/j.molcel.2015.02.014
734 Hounkpe, B. W., Chenou, F., de Lima, F., & De Paula, E. V. (2021). HRT Atlas v1.0
735 database: Redefining human and mouse housekeeping genes and candidate
736 reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids*
737 *Research*, *49*(D1), D947–D955. https://doi.org/10.1093/nar/gkaa609
738 Inoue, F., & Ahituv, N. (2015). Decoding enhancers using massively parallel reporter
739 assays. *Genomics*, *106*(3), 159–164.
740 https://doi.org/10.1016/j.ygeno.2015.06.005
741 Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X., & Blobel, G. A. (2008).
742 Exchange of GATA factors mediates transitions in looped chromatin organization
743 at a developmentally regulated gene locus. *Molecular Cell*, *29*(2), 232–242.
744 https://doi.org/10.1016/j.molcel.2007.11.020
745 Joo, J.-Y., Schaukowitch, K., Farbiak, L., Kilaru, G., & Kim, T.-K. (2016). Stimulus-specific
746 combinatorial functionality of neuronal c-fos enhancers. *Nature Neuroscience*,
747 *19*(1), 75–83. https://doi.org/10.1038/nn.4170
748 Joshi, O., Wang, S.-Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P. J., Habibi, E., Shaik, J.,
749 Saeed, S., Handoko, L., Richmond, T., Spivakov, M., Burgess, D., & Stunnenberg,
750 H. G. (2015). Dynamic Reorganization of Extremely Long-Range Promoter-
751 Promoter Interactions between Two States of Pluripotency. *Cell Stem Cell*, *17*(6),
752 748–757. https://doi.org/10.1016/j.stem.2015.11.010

Kim, D.-K., Cha, Y., Ahn, H.-J., Kim, G., & Park, K.-S. (2014). Lefty1 and lefty2 control the balance between self-renewal and pluripotent differentiation of mouse embryonic stem cells. *Stem Cells and Development*, *23*(5), 457–466. https://doi.org/10.1089/scd.2013.0220

Kim, S., & Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. *Molecular Cell*, *83*(3), 373–392. https://doi.org/10.1016/j.molcel.2022.12.032

Klein, J. C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., Ahituv, N., & Shendure, J. (2020). A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nature Methods*, *17*(11), 1083–1091. https://doi.org/10.1038/s41592-020-0965-y

Köster, J., & Rahmann, S. (2012). Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics (Oxford, England)*, *28*(19), 2520–2522. https://doi.org/10.1093/bioinformatics/bts480

Lam, D. D., de Souza, F. S. J., Nasif, S., Yamashita, M., López-Leal, R., Otero-Corchon, V., Meece, K., Sampath, H., Mercer, A. J., Wardlaw, S. L., Rubinstein, M., & Low, M. J. (2015). Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold. *PLoS Genetics*, *11*(2), e1004935. https://doi.org/10.1371/journal.pgen.1004935

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lin, X., Liu, Y., Liu, S., Zhu, X., Wu, L., Zhu, Y., Zhao, D., Xu, X., Chemparathy, A., Wang, H., Cao, Y., Nakamura, M., Noordermeer, J. N., La Russa, M., Wong, W. H., Zhao, K., & Qi, L. S. (2022). Nested epistasis enhancer networks for robust genome regulation. *Science (New York, N.Y.)*, *377*(6610), 1077–1085. https://doi.org/10.1126/science.abk3512

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018

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 dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Molecular Cell*, *81*(5), 969-982.e13. https://doi.org/10.1016/j.molcel.2020.12.047

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*, *10*(6), 1453–1465. https://doi.org/10.1016/s1097-2765(02)00781-5

Tuan, D., Solomon, W., Li, Q., & London, I. M. (1985). The 'beta-like-globin' gene domain in human erythroid cells. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(19), 6384–6388. https://doi.org/10.1073/pnas.82.19.6384

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss, M. J., Dekker, J., & Blobel, G. A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular Cell*, *17*(3), 453–462. https://doi.org/10.1016/j.molcel.2004.12.028

van Arensbergen, J., FitzPatrick, V. D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H. J., & van Steensel, B. (2017). Genome-wide mapping of autonomous promoter activity in human cells. *Nature Biotechnology*, *35*(2), 145–153. https://doi.org/10.1038/nbt.3754

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Wickham, H. (2016). *Ggplot2 Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Will, A. J., Cova, G., Osterwalder, M., Chan, W.-L., Wittler, L., Brieske, N., Heinrich, V., de Villartay, J.-P., Vingron, M., Klopocki, E., Visel, A., Lupiáñez, D. G., & Mundlos, S. (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nature Genetics*, *49*(10), 1539–1545. https://doi.org/10.1038/ng.3939

You, F. M., Huo, N., Gu, Y. Q., Luo, M.-C., Ma, Y., Hane, D., Lazo, G. R., Dvorak, J., & Anderson, O. D. (2008). BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*, *9*, 253. https://doi.org/10.1186/1471-2105-9-253

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

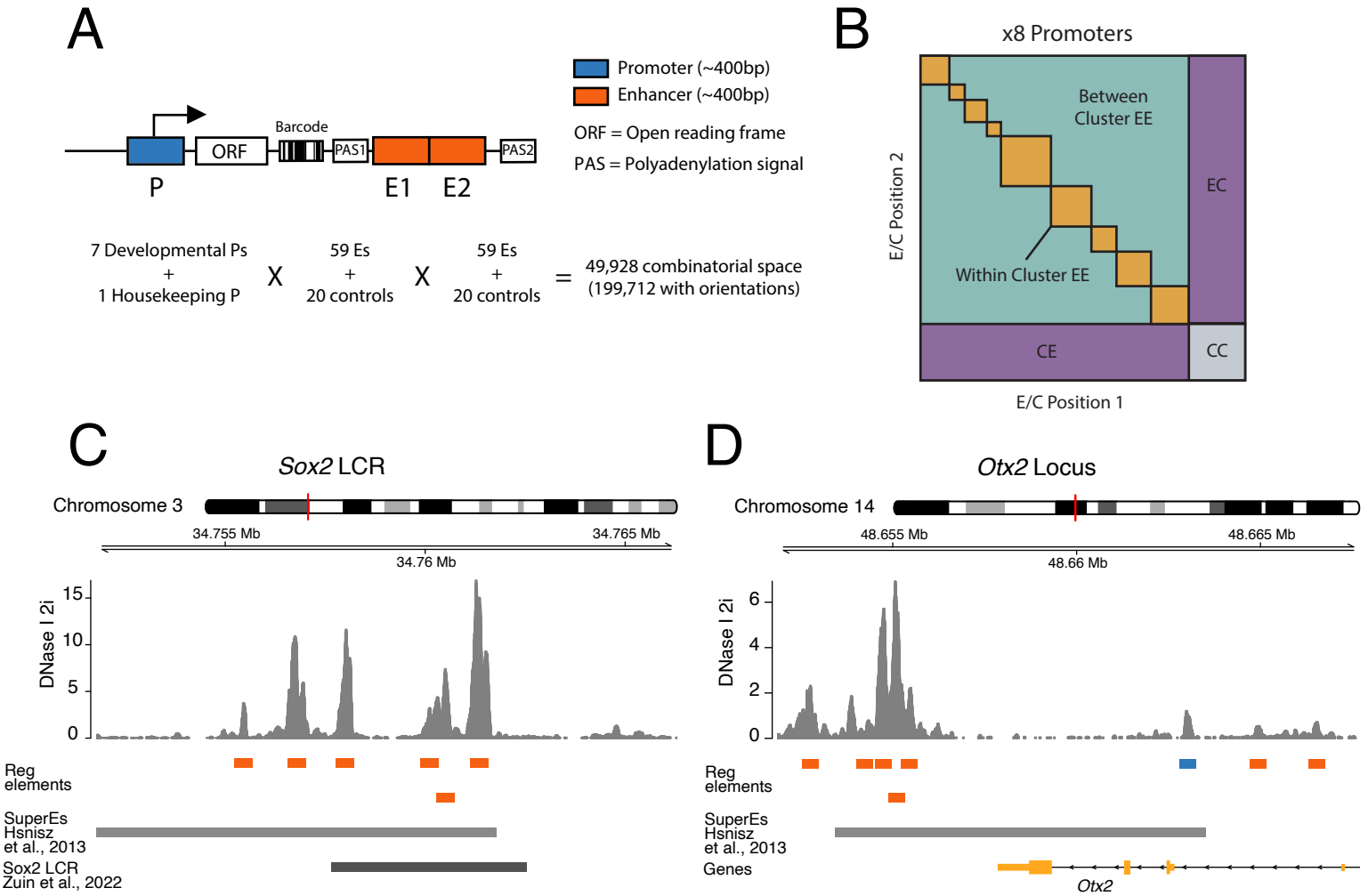


Figure 1. Schematic of three-way combinatorial approach. **A)** Three-way combinatorial Massively parallel reporter assay (MPRA) design to test enhancer-enhancer-promoter combinations. 8 barcoded reporter assay libraries, 1 per promoter, were constructed. Pairs of DNA elements (enhancers and scrambled 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.

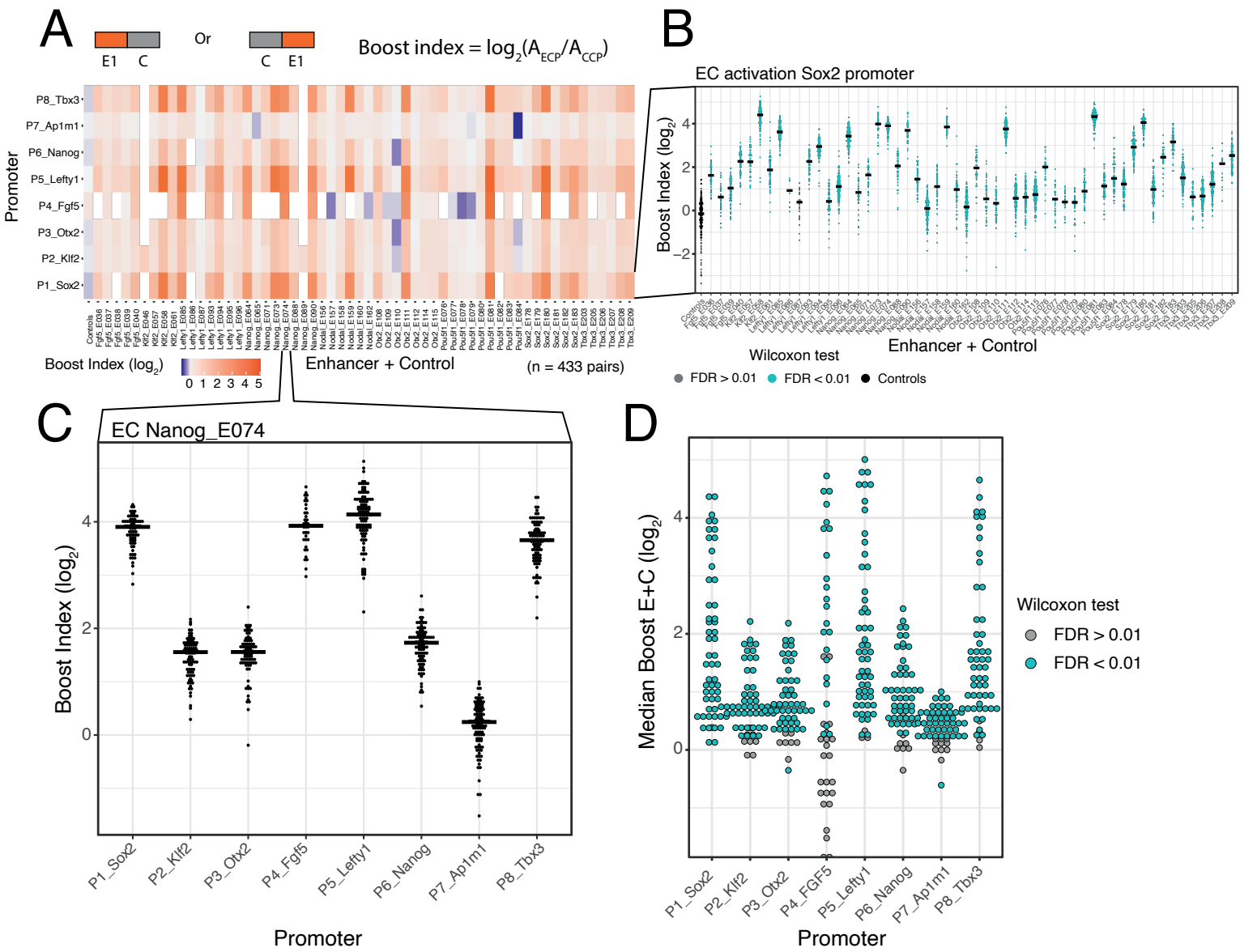


Figure 2. Effects of single enhancers across promoters. **A)** Enhancer-promoter median boost index matrix of single enhacers**.** For each single enhancer a median boost index (see figure and methods, $\log_2$(Activity_Combination/control-control baseline)) was calculated across all enhancer-control combinations for that particular enhancer in any position and orientation, the baseline is the median activity of the promoter across al control-control combinations. For controls a median boost index was calculated across all controls over the median activity of all controls. 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 a single enhancer enhancer+control combinations and the controls.

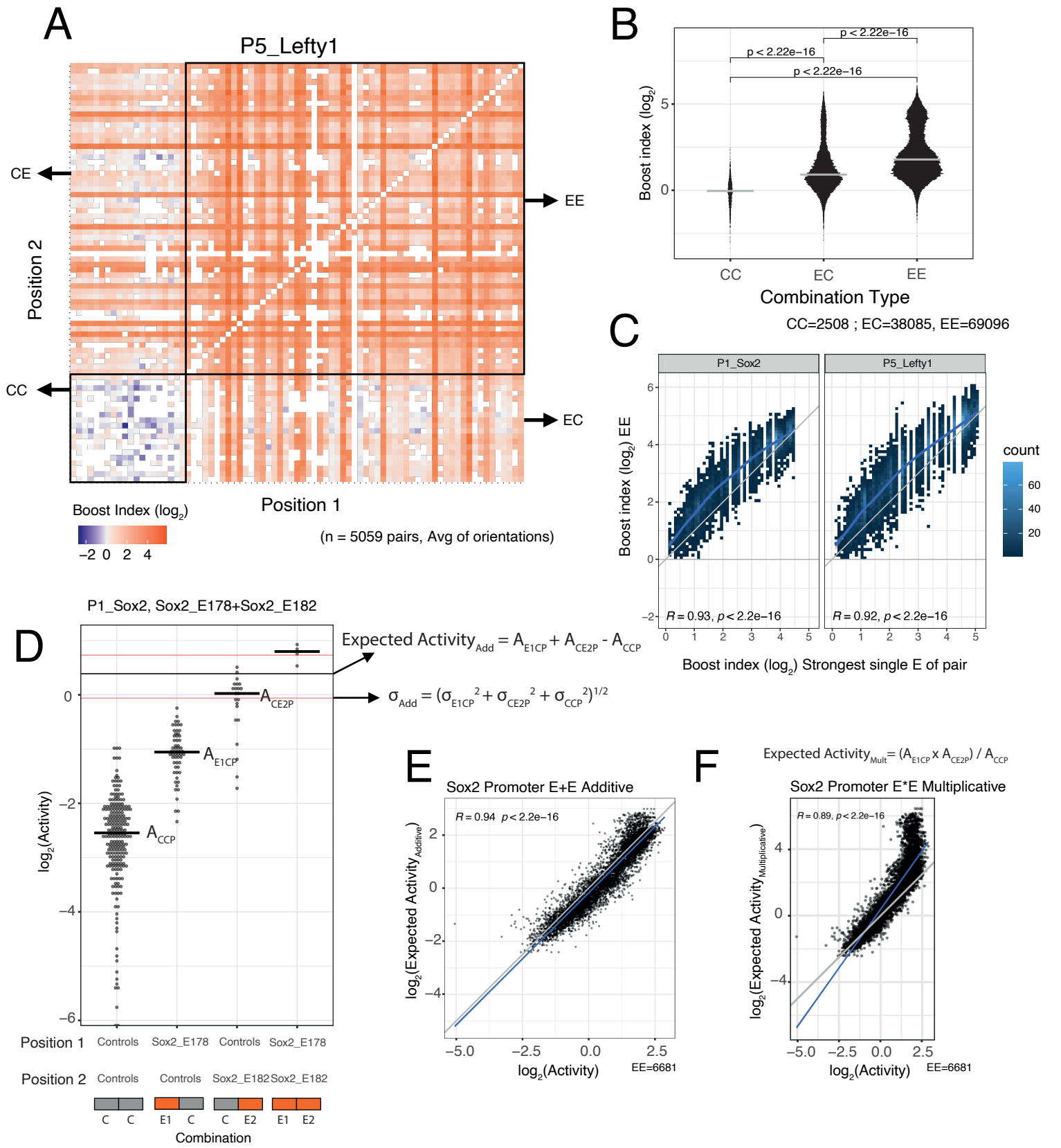


Figure 3. Effects of enhancer-enhancer combinations. **A)** Fragment-fragment combinatorial boost index matrix for the *Lefty1* promoter. Each square represents 1 control-control (CC), enhancer-control (EC or CE) or enhancer-enhancer (EE) combination. Color coding corresponds to the average boost index 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.

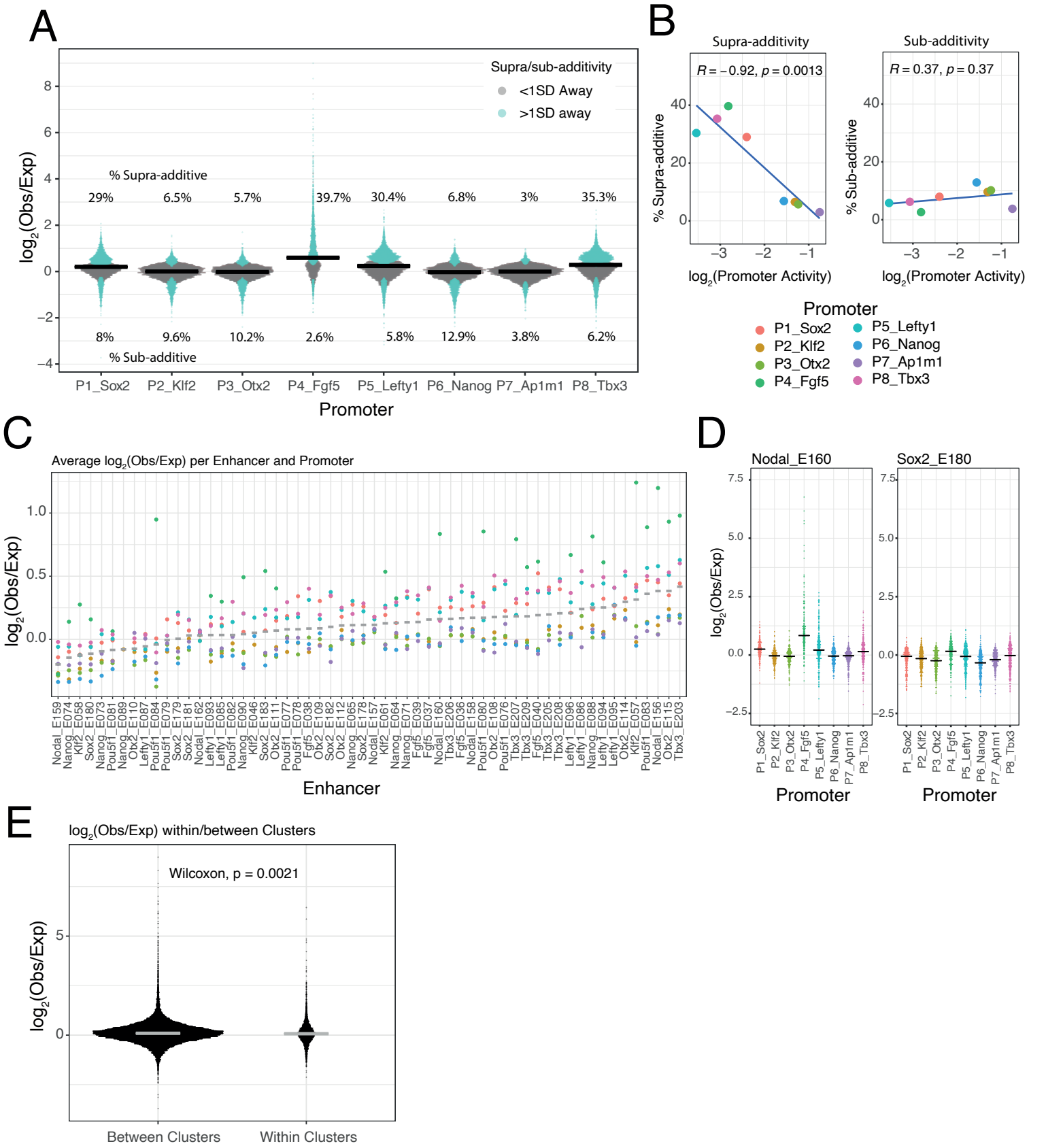


Figure 4. Supra and sub-additive behaviours of enhancer combinations. **A)** Distributions of $\log_2$ observed activities over expected additive activities ratios of enhancer-enhancer combinations across promoters. Colored in turquoise are supra and sub-additive combinations for which the observed activity 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.

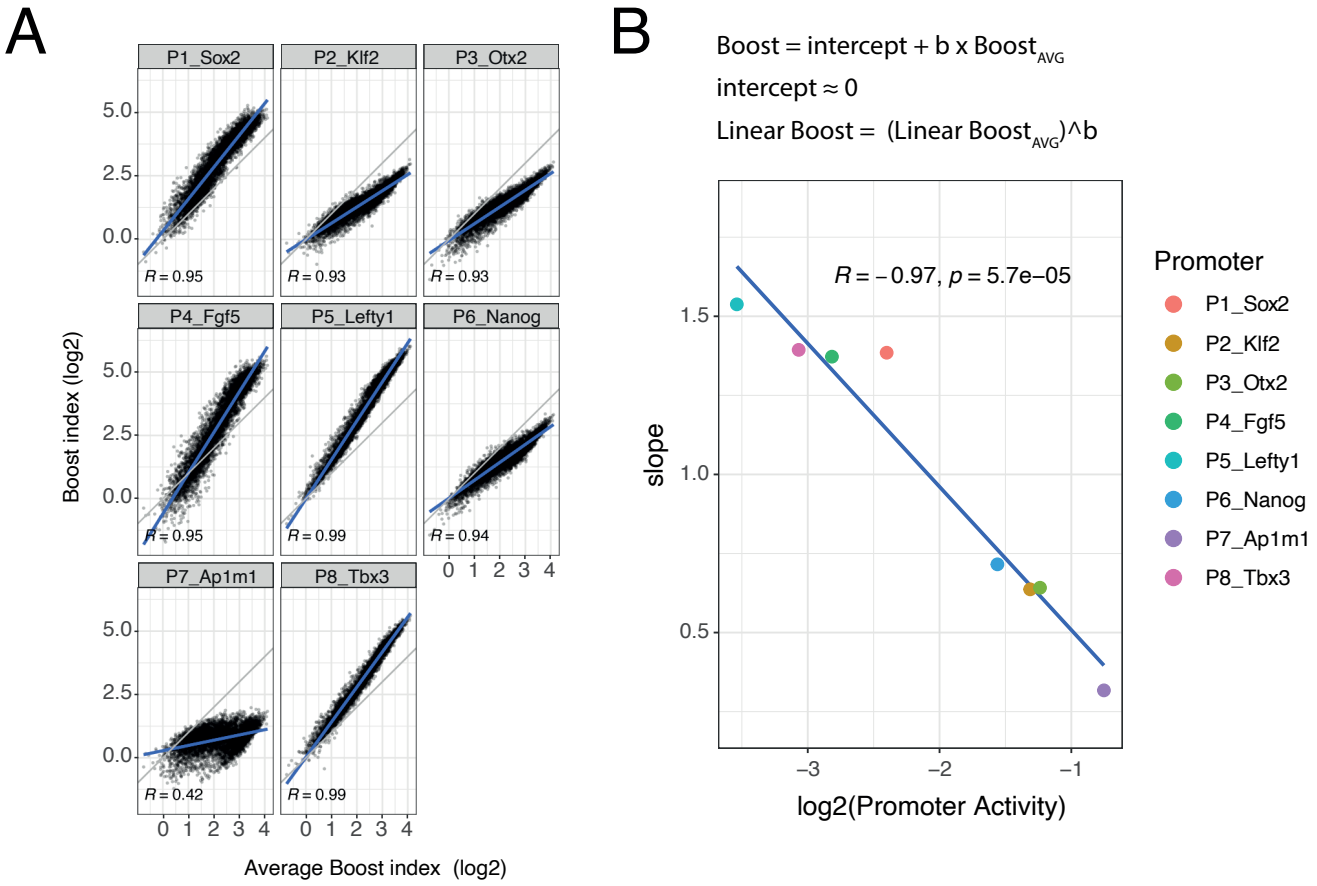
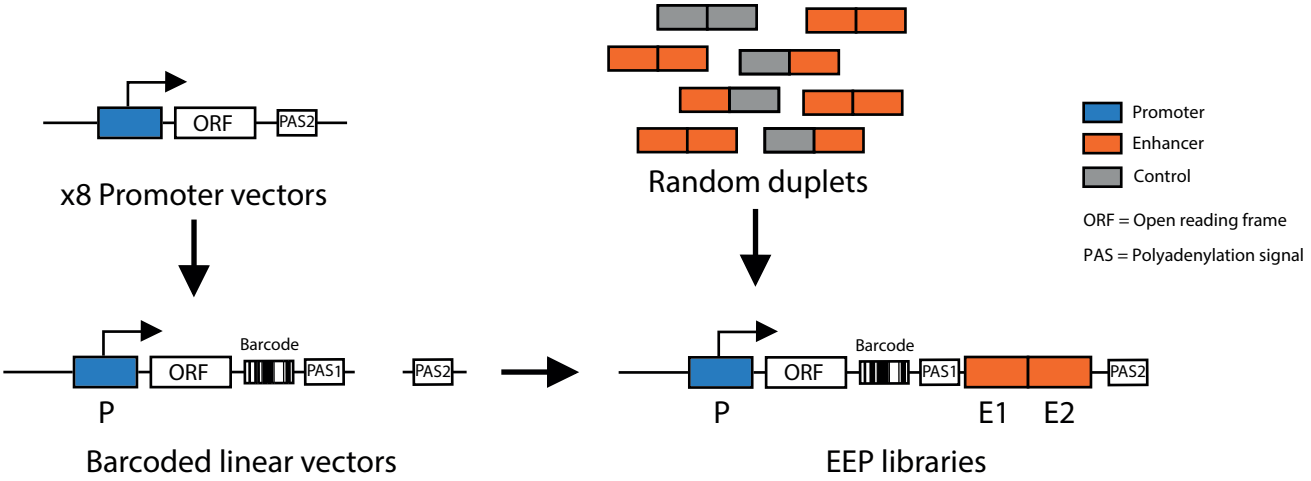
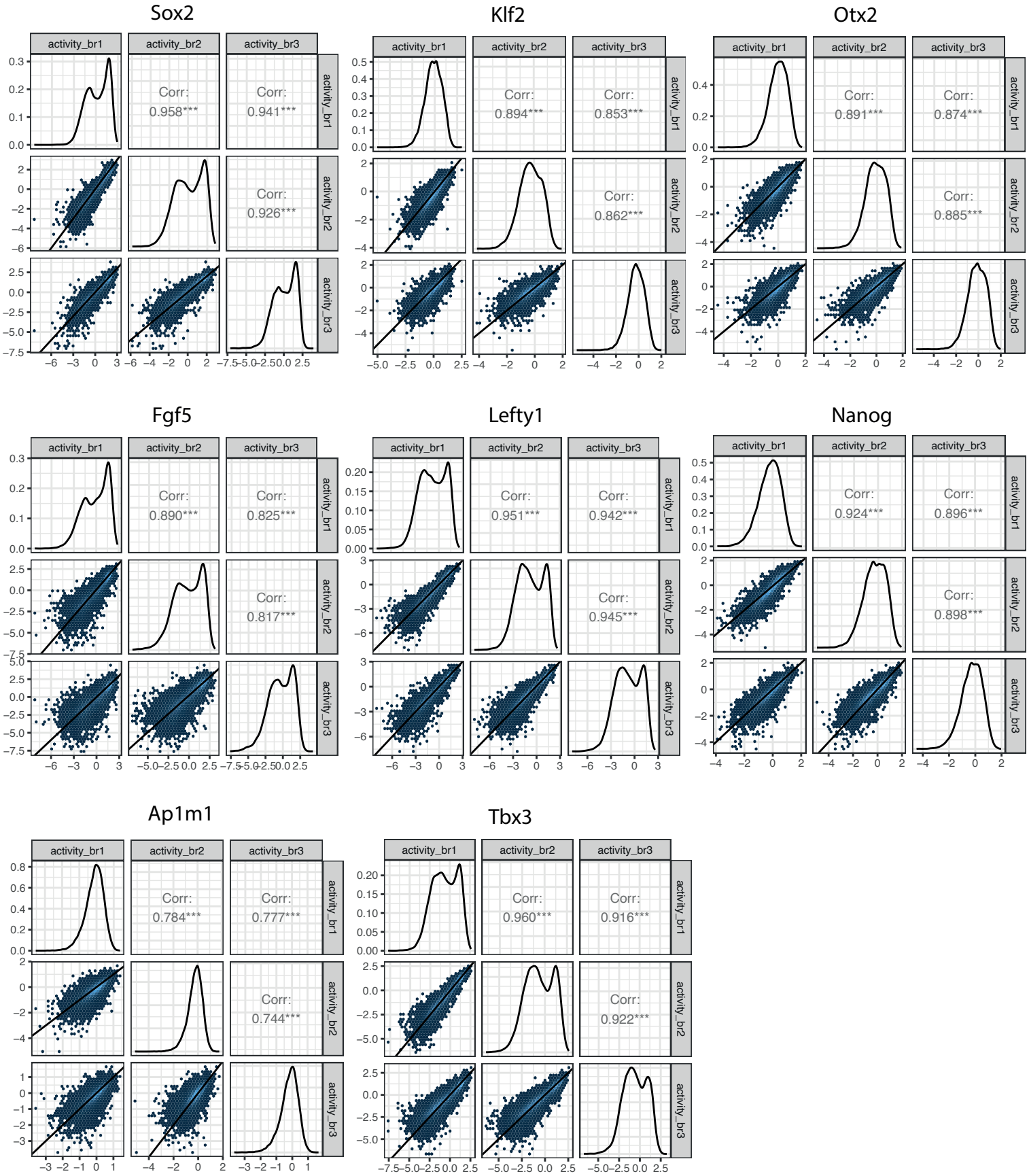


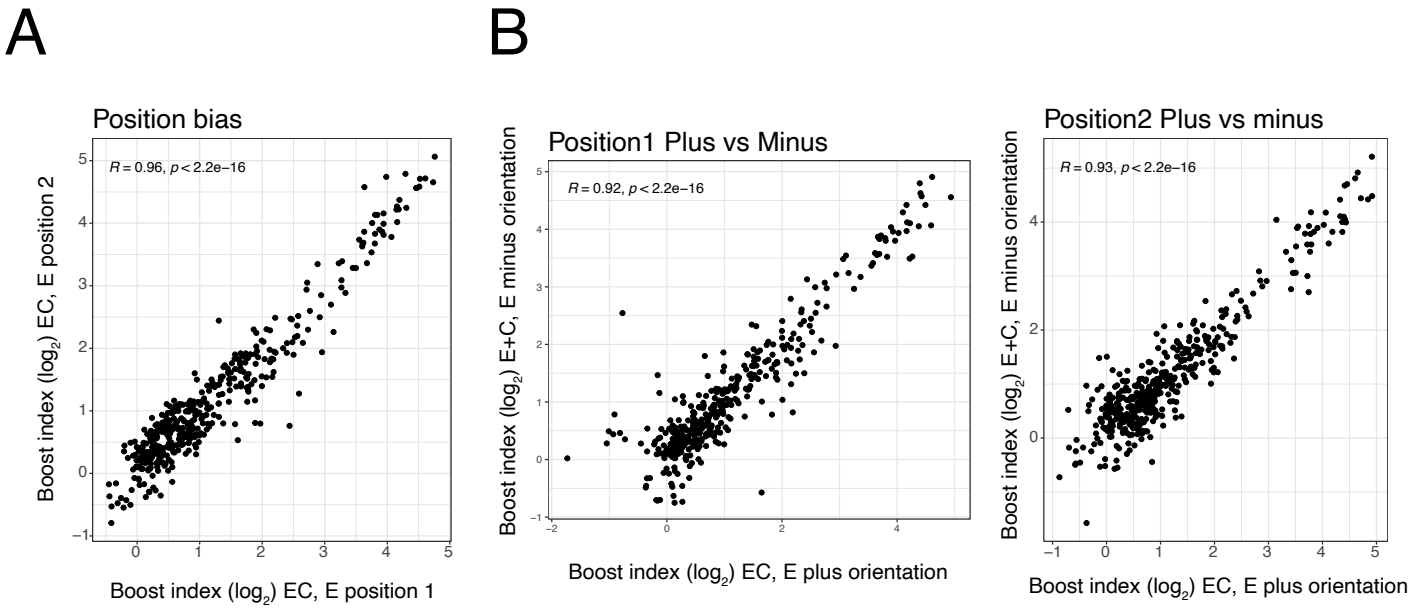
Figure 5. Non-linear responses of promoters to enhancer-enhancer combinations. **A)** Relationship between observed boost indices and average boost index across promoters for all shared enhancer-enhancer combinations. Blue lines represent the linear fit of the data. **B)** Relationship between the slopes 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.



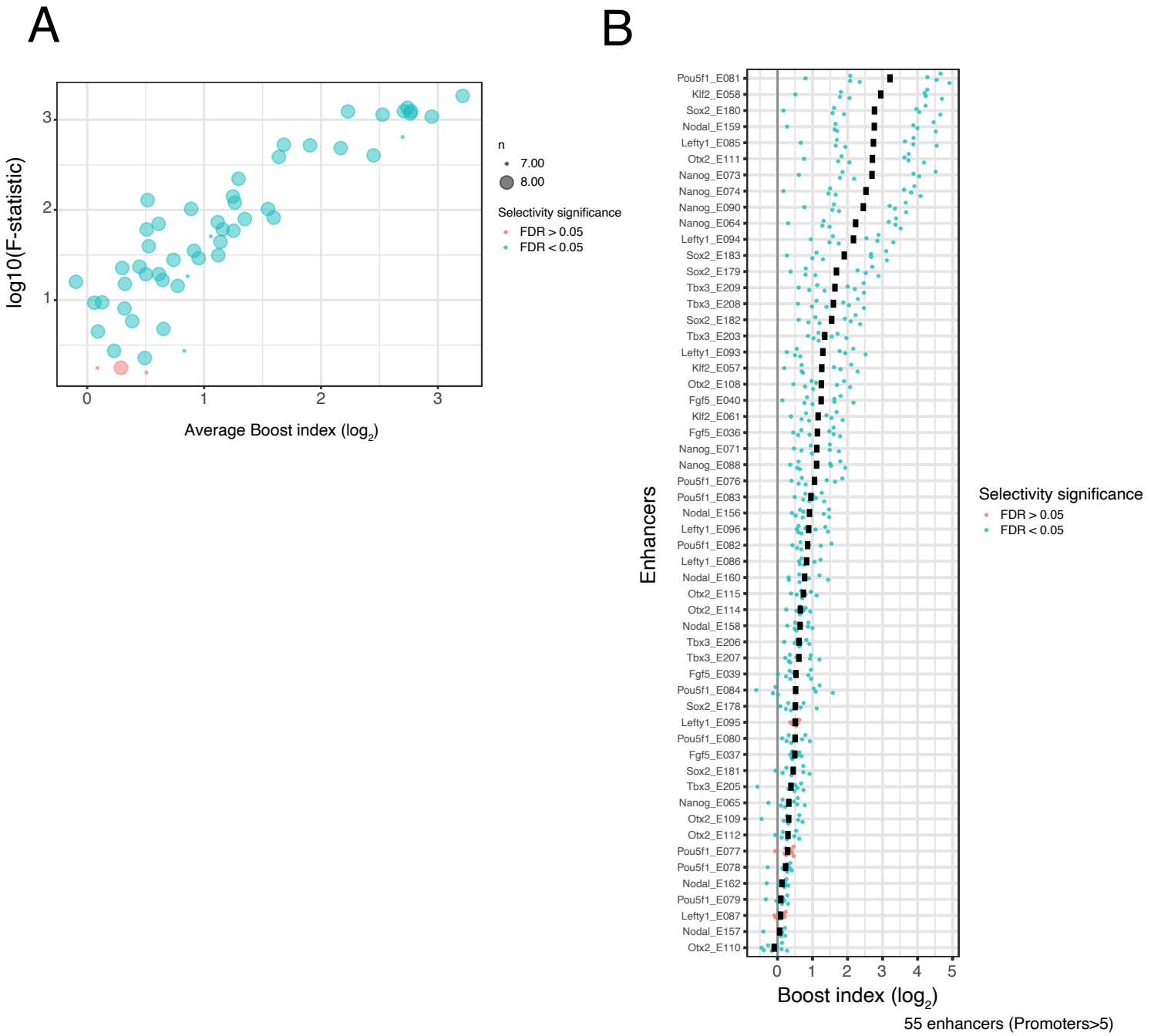
Supplementary Figure 1. Schematic of cloning strategy to generate enhancer-enhancer-promoter libraries. First the 8 different promoters were cloned into a reporter vector. Each of the 8 vectors was barcoded and linearised, random enhancer-enhancer, enhancer-control and control-control combinations were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



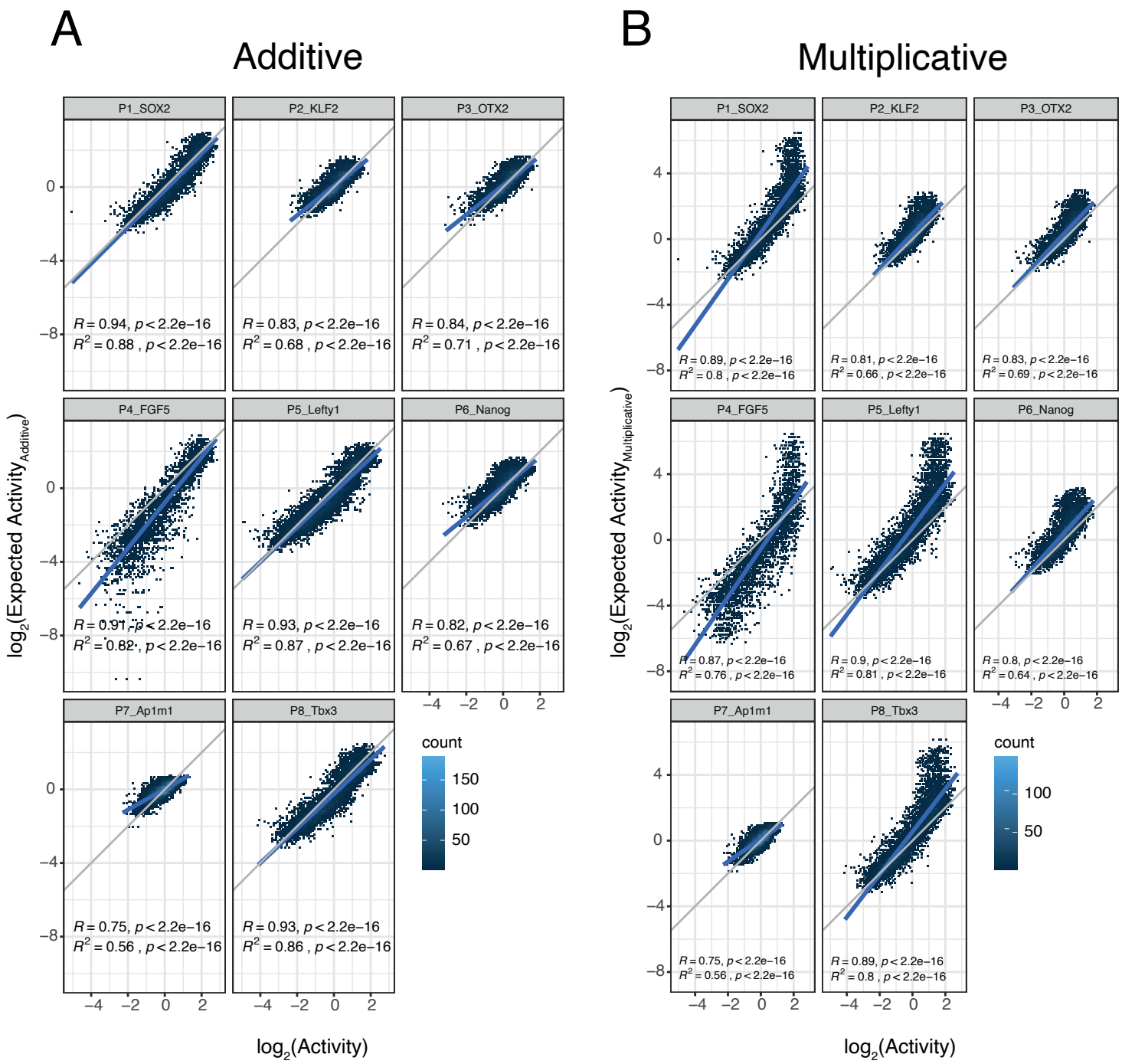
Supplementary Figure 2. Reproducibility of experimental data. matrix of replicate correlations for all 8 EEP libraries. Lower left panels represent the 2 dimensional density plots of replicate-replicate activity correlations. Middle panels represent the 1 dimensional density plot of each replicate. Upper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



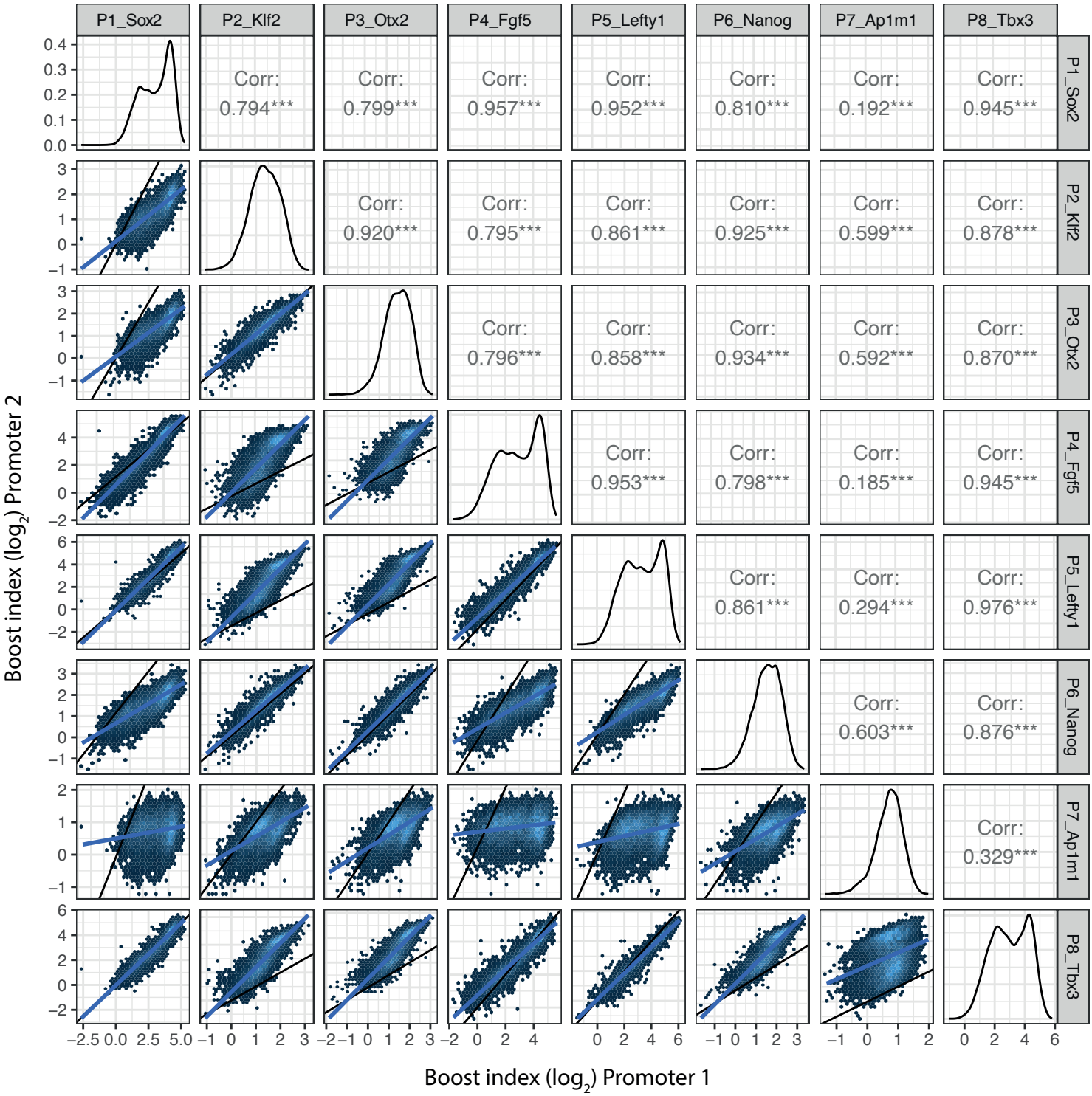
Supplementary Figure 3. Position and orientation bias. **A)** Relationship between single enhancer boost indices across all EC combinations in position 1 versus in position 2 (See Figure 1). **B)** Relationship between single enhancer boost indices across all EC combinations plus versus minus orientation, 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 indicate significance of the Welch F-test at a 5% FDR. Vertical black bars represent the median of the distribution.



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.