# A novel interpretable deep learning-based computational framework designed synthetic enhancers with broad cross-species activity

Zhaohong Li[1,2,#], Yuanyuan Zhang[1,2,#], Bo Peng[3,4,#], Shenghua Qin[1,2,#], Qian Zhang[5], Yun Chen[1,2] Choulin Chen[1,2], Yongzhou Bao[1,2], Yuqi Zhu[6], Yi Hong[6], Binghua Liu[7], Qian Liu[7], Lingna Xu[1,2], Xi Chen[8], Xinhao Ma[9], Hongyan Wang[7], Long Xie[1], Yilong Yao[10], Biao Deng[1,2], Jiaying Li[11], Baojun De[12], Yuting Chen[12], Jing Wang[1], Tian Li[13], Ranran Liu[14], Zhonglin Tang[10], Junwei Cao[12], Erwei Zuo[1], Chugang Mei[9], Fangjie Zhu[13], Changwei Shao[7], Guirong Wang[1], Tongjun Sun[6], Ningli Wang[11], Gang Liu[5], Jian-Quan Ni[3,4,15]*, Yuwen Liu[1,2,10]*

1 Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China.

2 Innovation Group of Pig Genome Design and Breeding, Research Centre for Animal Genome, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China.

3 School of Basic Medical Sciences, Tsinghua University, Beijing, 100084, China

4 State Key Laboratory of Molecular Oncology, Beijing, 100084, China

5 State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

6 Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

7 National Key Laboratory of Maricultural Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, Shandong 266071, China.

8 Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

9 Northwest A&F University, Yangling, Shaanxi, 712100, China.

10 Kunpeng Institute of Modern Agriculture at Foshan, Chinese Academy of Agricultural Sciences, Foshan, 528226, China.

11 Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, 100101, China.

41    12 College of Life Sciences, Inner Mongolia Agricultural University, Hohhot,
42    010018, China.
43    13 College of JUNCAO Science and Ecology, Haixia Institute of Science and
44    Technology, National Engineering Research Center of JUNCAO, Fujian Agriculture
45    and Forestry University (FAFU), Fuzhou, 0350002, China.
46    14 Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing,
47    100193, China.
48    15 SXMU-Tsinghua Collaborative Innovation Center for Frontier Medicine, Shanxi
49    Medical, Shanxi Medical University, Taiyuan, 030001, China.
50    [#] These authors contributed equally to this work.
51    *Correspondence:
52    Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key
53    Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics
54    Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China;
55    E-mail: liuyuwen@caas.cn
56    School of Basic Medical Sciences, Tsinghua University, Beijing, China; E-mail:
57    nijq@tsinghua.edu.cn
58
59    Running title: Zhaohong Li et al. / Designing the enhancers *in silico*
60

## Cell line culture

Various cell lines were employed in this study, including Human 293T cells (provided by Dr. Guoqiang Yi), A549 cells (supplied by Dr. Songhua Cai), ARPE-19 cells (gifted from Dr. Ningli Wang), mouse C2C12 cells (acquired from Dr. Tang Zhonglin), CHO cells (received from Su'yan Biotech), 3T3 cells (provided by Dr. Tang Zhonglin), chicken DF1 cells (received from Dr. Erwei Zuo), PSKM cells (supplied by Su'yan Biotech), and PK15 cells (received from Su'yan Biotech). The specified cell lines underwent cultivation in a medium comprising 90% DMEM (Giboc, C11995500BT), 10% serum (Gibco, 10270-106), and 1% penicillin/streptomycin (Gibco, 2441868) at 37°C with 5% $CO_2$. For HepG2 cells (obtained from Procell), the cultivation medium consists of 90% MEM medium (Procell, PM150410), 10% FBS (Gibco, 10270-106), and 1% penicillin/streptomycin (Gibco, 2441868). S2 cells (served by Su'yan Biotech) and SF9 cells (gifted from Dr. Guirong Wang) were cultured in a medium composed of 90% Schneider's medium (Thermo Fisher, 21720024), 10% serum (Gibco, 10270-106), and 1% penicillin/streptomycin (Gibco, 2441868) at 28°C. During S2's feeding，half of the conditioned medium was substituted with an equal volume of fresh medium. K562 cells (provided by Dr. Tang Zhonglin) and bovine kidney cells (obtained from KEPU Biology) were cultured in a medium consisting of 90% RPMI Medium (Giboc, C22400500BT), 10% serum (Gibco, 10270-106), and 1% penicillin/streptomycin (Gibco, 2441868) at 37°C with 5% $CO_2$. Cynoglossus semilaevis spermatogonium (provided by Dr. Changwei Shao) were suspended in a solution containing 8% Antibiotic-Antimycotic (Gibco, 15240-062) and 92% L15 medium (Solarbio, LA9510). Spermathecal tissues were dissected in an ultra-clean bench, digested with trypsin (Gibco, 25200072) for 10 minutes, and the tissue-cell mixture was transferred to 25 $cm^2$ cell culture flasks (Corning, 430639) with 4 mL of a medium consisting of 76% L15 medium (Solarbio, LA9510), 20% serum (Gibco, 10270-106), and 4% Antibiotic-Antimycotic (Gibco, 15240-062). Cells were seeded in 24-well plates (Corning, 3524) and incubated at 24°C.

## Cell line transfection

S2, SF9 cells: cells were cultured in 24-well plates (NEST, 702001) at a density of 2.5 × $10^5$ cells per well. For each transfection，we used 1.5 μl transfection Reagent (Qiagen, 301007) and 0.5 μg plasmid. In S2 and SF9 cells, pGL3_DSCP_Rluc plasmids were transfected into cells for renilla luciferase expression. The ratio of firefly luciferase expression plasmid to Renilla luciferase expression plasmid was 5:1.

K562 Cells: cells were inoculated at a density of $1.0 × 10^6$ cells/ml medium. For each transfection reaction, we used 100 μl transfection reagent, $1.0 × 10^6$ cells, and 2 μg plasmid. Transfection was performed using a 4D-Nucleofector® X Unit (Lonza, AAF-1003X) with the FF120 program, and P4 transfection reagent (Lonza, PBP4-00675).

ARPE-19 Cells: Transfection involved $1 × 10^6$ cells per reaction, 0.2 μg plasmid, and 20 μl P4 electrotransformation solution (Lonza, PBP4-00675). Transfection was carried out using a 4D-Nucleofector® X Unit (Lonza, AAF-1003X) with the CA137 program.

HepG2 Cells: The transfection system for a six-well plate included 8 μl jetPRIME (Polyplus, 101000046) per well, 4 μg plasmid, and 200 μl jetPRIME buffer (Polyplus, 101000046).

Human 293T, A549 cells, mouse C2C12, CHO, 3T3 cells, chicken DF1 cells, PSKM, PK15 cells, and bovine kidney bells: cultured at $2.0 × 10^5$ cells per well in a 24-well plate (NEST, 702001). Transfection experiments were performed with jetPRIME (Polyplus, 101000046) using 0.6 μg plasmid, 50 μl jetPRIME buffer (Polyplus, 101000046), and 2.5 μl jetPRIME per well.

Cynoglossus semilaevis Spermatogonium: Transfection experiments were conducted when cell confluence reached about 80%. For each well, 500 ng of plasmid was added, together with the transfection reagent Lipo8000 (Beyotime, C0533, 0.5 ml).

Pichia pastoris: 100 μL of Pichia pastoris competent was thawed on ice, plasmid was added, the competent was transferred to a pre-chilled 2 mm shock cup, the parameters of the shock converter were adjusted to 1515 V, 25 μF, 200 Φ, and the periphery of the shock cup was kept dry for electric shock. After the shock was completed, 1 mL of resuscitation solution (500 μL YPD + 500 μL 1 mol/L sorbitol) was mixed with the bacterial solution in the shock cup, and the bacterial solution was aspirated into an EP

119 tube, 28°C, 250 rpm/min shaker, and resuscitation was carried out for 3 hours. The

120 thawed bacterial solution was coated onto the required screening medium plate, and

121 monoclonal positive colonies were picked and preserved. In our Pichia pastoris

122 experiments, we employed a stable transformation method which results in the random

123 insertion of sequences into the genome. To ensure co-localization of the firefly and

124 Renilla luciferase reporters within the same genomic location and to minimize

125 variability in expression, both genes were cloned into a single plasmid. This strategy

126 provides more consistent and reliable normalization of luciferase activity compared to

127 separate plasmids.

## The sequence of plasmid

129 The sequence of E-pGL3-DSCP-luc (The uppercase bases are synthetic enhancer):

130 ttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtaga

131 taactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatt

132 tatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctatta

133 attgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgt

134 cacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaa

135 aagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgc

136 ataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgc

137 ggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgg

138 aaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaact

139 gatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataag

140 ggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggata

141 catatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgcgccctgta

142 gcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc

143 tttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggggctcccttagggttccgatt

144 tagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttt

145 ttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctat

146 tcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaaca

147 aaatattaacgcttacaatttgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgcta

148 ttacgccagcccaagctaccatgataagtaagtaatattaaggtacgggaggtacttggagcggccgcaataaaatatcttta

149 ttttcattacatctgtgtgttggttttttgtgtgaatcgatagtactaacatacgctctccatcaaaacaaaacgaaacaaaacaa

150 actagcaaaataggctgtccccagtgcaagtgcaggtgccagaacatttctctatcgataggtacTGATGTGACTC

151 ACATAGGCGGTGGCGTGATATGTTGTGACTCATTTCCCGGAAACGGATGA

152 CTAATGCCATATGTTATCAGTTTCCTGGAAATTTGATCACGCCATATTGTG

153 AAATCATGCGATTCCCGGATCACGTGACGGCCGGACGTGACAAGTATGAG

154 TCACTAAGTGGCGTGATCTTACGAATCACGTGATGGTCAATGTCACGTGA

155 TCGGCTGGTGAGTCAGCAATATCGTGTGATTCATTCgatctgcgatcgagctcgcccgg

156 ggatcgagcgcagcggtataaaagggcgcggggtggctgagagcatcagttgtgaatgaatgttcgagccgagcagac

157 gtgccgctgccttcgttaatatcctttgaataagccaactttgaatcacaagacgcataccaaacggcattccggtactgttgg

158    taaagccaccatggaagacgccaaaaacataaagaaaggcccggcgccattctatccgctggaagatggaaccgctgga

159    gagcaactgcataaggctatgaagagatacgccctggttcctggaacaattgcttttacagatgcacatatcgaggtggaca

160    tcacttacgctgagtacttcgaaatgtccgttcggttggcagaagctatgaaacgatatgggctgaatacaaatcacagaatc

161    gtcgtatgcagtgaaaactctcttcaattctttatgccggtgttgggcgcgttatttatcggagttgcagttgcgcccgcgaacg

162    acatttataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaaggggttgcaaa

163    aaattttgaacgtgcaaaaaaagctcccaatcatccaaaaaattattatcatggattctaaaacggattaccagggatttcagtc

164    gatgtacacgttcgtcacatctcatctacctcccggttttaatgaatacgatttttgtgccagagtccttcgatagggacaagaca

165    attgcactgatcatgaactcctctggatctactggtctgcctaaaggtgtcgctctgcctcatagaactgcctgcgtgagattct

166    cgcatgccagagatcctattttttggcaatcaaatcattccggatactgcgattttaagtgttgttccattccatcacggttttggaa

167    tgtttactacactcggatatttgatatgtggatttcgagtcgtcttaatgtatagatttgaagaagagctgtttctgaggagccttc

168    aggattacaagattcaaagtgcgctgctggtgccaaccctattctccttcttcgccaaaagcactctgattgacaaatacgattt

169    atctaatttacacgaaattgcttctggtggcgctcccctctctaaggaagtcggggaagcggttgccaagaggttccatctgc

170    caggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacacccgagggggatgataaaccgggc

171    gcggtcggtaaagttgttccattttttgaagcgaaggttgtggatctggataccgggaaaacgctgggcgttaatcaaagag

172    gcgaactgtgtgtgagaggtcctatgattatgtccggttatgtaaacaatccggaagcgaccaacgccttgattgacaagga

173    tggatggctacattctggagacatagcttactgggacgaagacgaacacttcttcatcgttgaccgcctgaagtctctgattaa

174    gtacaaaggctatcaggtggctcccgctgaattggaatccatcttgctccaacaccccaacatcttcgacgcaggtgtcgca

175    ggtcttcccgacgatgacgccggtgaacttcccgccgccgttgttgtttttggagcacggaaagacgatgacggaaaaaga

176    gatcgtggattacgtcgccagtcaagtaacaaccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaagtaccgaa

177    aggtcttaccggaaaactcgacgcaagaaaaatcagagagatcctcataaaggccaagaagggcggaaagatcgccgt

178    gtaattctagagtcggggcggccggccgcttcgagcagacatgataagatacattgatgagtttggacaaaccacaactag

179    aatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaac

180    aacaacaattgcattcattttatgtttcaggttcaggggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtg

181    gtaaaatcgataaggatccgtcgaccgatgcccttgagagccttcaacccagtcagctccttccggtgggcgcggggcat

182    gactatcgtcgccgcacttatgactgtcttctttatcatgcaactcgtaggacaggtgccggcagcgctcttccgcttcctcgc

183    tcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacag

184    aatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttg

185    ctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgac

186    aggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatac

187    ctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgct

188    ccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacc

189    cggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctaca

190    gagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttc

191    ggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattac

192    gcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaa

193    gggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtata

194    tatgagtaaacttggtctgacag

195    The sequence of E-CAG-luc (The uppercase bases are synthetic enhancer and CAG

196    promoter):

197    catgcaagctgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcat

198    aaccccttggggcggccgcttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtga

199    aaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaatt

200    gcattcattttatgtttcaggttcaggggggagatgtgggaggttttttaagcaagtaaaacctctacaaatgtggtaaaatcga

201    attttaacaaaatattaacgcttacaatttcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgcgg

202    atctgcgcagcaccatggcctgaaataacctctgaaagaggaacttggttaggtaccttaccggtggtgggagcggaggt

203    ggcggatcaggtggcggaggctccggagggattgaacaagatggattgcacgcaggttctccggccgcttgggtggag

204    aggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcg

205    cccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggc

206    cacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgc

207    cggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatac

208    gcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtc

209    ttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcat

210    gcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctgg

211    attcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagctt

212    ggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttct

213    tgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgatggccgca

214    ataaaatatctttattttcattacatctgtgtgttggttttttgtgtgaatcgatagcgataaggatcctctttgcgcttgcgttttccct

215    tgtccagatagcccagtagctgacattcatccggggtcagcaccgtttctgcggactggctttctacgtaatggtttcttagac

216    gtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattttttctaaatacattcaaatatgtatccgctcatga

217    gacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccttt

218    tttgcggcattttgccttcctgttttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacga

219    gtggggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcac

220    tttcaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctca

221    gaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgcc

222    ataaccatgagtgataacactgcggccaacttacttctgacaactatcggaggaccgaaggagctaaccgcttttttgcaca

223    acatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacacc

224    acgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaat

225    agactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatct

226    ggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctac

227    acgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggta

228    attcgaaatgaccgaccaagcgacgcccaaccggtatcagctcactcaaaggcggtaatacggttatccacagaatcagg

229    ggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtt

230    tttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactat

231    aaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc

232    ctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctg

233    ggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaaga

234    cacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttg

235    aagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaa

236    gagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcaga

237    aaaaaaggatttcaagaagatcctttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttg

238    gtcatgagattatcaaaaaggatcttcacctagatcctttttatagtccggaaatacaggaacgcacgctggatggcccttcgc

239    tgggatggtgaaaccatgaaaaatggcagcttcagtggattaagtggggggtaatgtggcctgtaccctctggttgcataggt

240    attcatacggttaaaatttatcaggcgcgattgcggcagttttttcgggtggtttgttgccattttttacctgtctgctgccgtgatcg

241    cgctgaacgcgtttttagcggtgcgtacaattaagggattatggtaaatccacttactgtctgccctcgtagccatcgagataaa

242    ccgcagtactccggccacgatgcgtccggcgtaGAATTCCATCGCGTGATCGCGATCGGATG

243    ACTCGGCTGAGTCACCGGATGTGACTCAGCTCTGGTTCTATGATTCATGCA

244    GGAAATGACTCATTGAGTCATTCTTGGAAGTGGCGTGACTGTCTGGTGGC

245    GTGACTTCCTGGAATGTGACTCAGCGGAAGCTCATGACTCATTCCTGGAA

246 ATCGTACGATGCGAGTCACGTAGGCCTACGTGACGGTGAATCATACTACG
247 TGGCGTGATACGGAGGAAATGACTCagatctaccTCGAGGTGAGCCCCACGTTC
248 TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTAT
249 TTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGG
250 CGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGG
251 CGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTC
252 CTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGC
253 GCGGCGGGCGGGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCG
254 CCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCAC
255 AGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTG
256 GTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGG
257 CTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGT
258 GTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGCT
259 GTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGC
260 GAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCG
261 AGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGG
262 GGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCC
263 GAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGT
264 GGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGC
265 CGGGCGGGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCG
266 CGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGC
267 CATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
268 CAAATCTGTGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAG
269 CGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGG
270 GAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCT
271 CGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGGACGGGGCAGGG
272 CGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAAC
273 CATGTTCATGCCTTCTTCTTTTTCCTACAGacctgcagcccaagcttggcaatccggtactgttg
274 gtaaagccaccatggaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaagacgggaccgccg
275 gcgagcagctgcacaaagccatgaagcgctacgccctggtgcccggcaccatcgcctttaccgacgcacatatcgaggt
276 ggacattacctacgccgagtacttcgagatgagcgttcggctggcagaagctatgaagcgctatgggctgaatacaaacca
277 tcggatcgtggtgtgcagcgagaatagcttgcagttcttcatgcccgtgttgggtgccctgttcatcggtgtggctgtggccc
278 cagctaacgacatctacaacgagcgcgagctgctgaacagcatgggcatcagccagcccaccgtcgtattcgtgagcaa
279 gaaagggctgcaaaagatcctcaacgtgcaaaagaagctaccgatcatacaaaagatcatcatcatggatagcaagaccg
280 actaccagggcttccaaagcatgtacaccttcgtgacttcccatttgccacccggcttcaacgagtacgacttcgtgcccga
281 gagcttcgaccgggacaaaaccatcgccctgatcatgaacagtagtggcagtaccggattgcccaagggcgtagccctac
282 cgcaccgcaccgcttgtgtccgattcagtcatgcccgcgacccatcttcggcaaccagatcatccccgacaccgctatcct
283 cagcgtggtgccatttcaccacggcttcggcatgttcaccacgctgggctacttgatctgcggctttcgggtcgtgctcatgt
284 accgcttcgaggaggagctattcttgcgcagcttgcaagactataagattcaatctgccctgctggtgcccacactatttagct
285 tcttcgctaagagcactctcatcgacaagtacgacctaagcaacttgcacgagatcgccagcggcggggcgccgctcagc
286 aaggaggtaggtgaggccgtggccaaacgcttccacctaccaggcatccgccagggctacggcctgacagaaacaacc
287 agcgccattctgatcacccccgaaggggacgacaagcctggcgcagtaggcaaggtggtgcccttcttcgaggctaagg
288 tggtggacttggacaccggtaagacactgggtgtgaaccagcgcggcgagctgtgcgtccgtggccccatgatcatgag
289 cggctacgttaacaaccccgaggctacaaacgctctcatcgacaaggacggctggctgcacagcggcgacatcgcctac

290 tgggacgaggacgagcacttcttcatcgtggaccggctgaagagcctgatcaaatacaagggctaccaggtagccccag
291 ccgaactggagagcatcctgctgcaacaccccaacatcttcgacgccggggtcgccggcctgccccgacgacgatgccg
292 gcgagctgcccgccgcagtcgtcgtgctggaacacggtaaaaccatgaccgagaaggagatcgtggactatgtggccag
293 ccaggttacaaccgccaagaagctgcgcggtggtgttgtgttcgtggacgaggtgcctaaaggactgaccggcaagttgg
294 acgcccgcaagatccgcgagattctcattaaggccaagaagggcggcaagatcgccgtgtaattctagttgtttaaacgag
295 ctcgctagcctcgagtctagagtcgacctgcagg
296 The sequence of pGL3-DSCP-Rluc (The uppercase bases are Renilla luciferase reporter
297 gene):
298 ttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtaga
299 taactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatt
300 tatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctatta
301 attgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgt
302 cacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaa
303 aagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgc
304 ataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgc
305 ggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgg
306 aaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaact
307 gatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataag
308 ggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggata
309 catatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgcgccctgta
310 gcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc
311 tttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggggctccctttagggttccgatt
312 tagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttt
313 ttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctat
314 tcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaaca
315 aaatattaacgcttacaatttgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgcta
316 ttacgccagcccaagctaccatgataagtaagtaatattaaggtacgggaggtacttggagcggccgcaataaaatatcttta
317 ttttcattacatctgtgtgttggttttttgtgtgaatcgatagtactaacatacgctctccatcaaaacaaaacgaaacaaaacaa
318 actagcaaaataggctgtccccagtgcaagtgcaggtgccagaacatttctctatcgataggtacgatctgcgatcgagctc
319 gcccggggatcgagcgcagcggtataaaagggcgcggggtggctgagagcatcagttgtgaatgaatgttcgagccga
320 gcagacgtgccgctgccttcgttaatatcctttgaataagccaactttgaatcacaagacgcataccaaacggcattccggta
321 ctgttggtaaagccaccATGACTTCGAAAGTTTATGATCCAGAACAAAGGAAACGG
322 ATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGAATGTTCT
323 TGATTCATTTATTAATTATTATGATTCAGAAAAACATGCAGAAAATGCTGT
324 TATTTTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACATGTTGTG
325 CCACATATTGAGCCAGTAGCGCGGTGTATTATACCAGACCTTATTGGTATG
326 GGCAAATCAGGCAAATCTGGTAATGGTTCTTATAGGTTACTTGATCATTAC
327 AAATATCTTACTGCATGGTTTGAACTTCTTAATTTACCAAAGAAGATCATT
328 TTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTATGAG
329 CATCAAGATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGT
330 GATTGAATCATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGA
331 TCAAATCTGAAGAAGGAGAAAAAATGGTTTTGGAGAATAACTTCTTCGTG
332 GAAACCATGTTGCCATCAAAAATCATGAGAAAGTTAGAACCAGAAGAATT
333 TGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTCCAA

334    CATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGAC

335    GTTGTACAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGAT

336    TTACCAAAAATGTTTATTGAATCGGACCCAGGATTCTTTTCCAATGCTATT

337    GTTGAAGGTGCCAAGAAGTTTCCTAATACTGAATTTGTCAAAGTAAAAGG

338    TCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGAAAATATATCA

339    AATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAATAAttctagagtcggggcggcc

340    ggccgcttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaatgctttatt

341    tgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtt

342    tcaggttcaggggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtaaaatcgataaggatccgtcg

343    accgatgcccttgagagccttcaacccagtcagctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgac

344    tgtcttctttatcatgcaactcgtaggacaggtgccggcagcgctcttccgcttcctcgctcactgactcgctgcgctcggtcg

345    ttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggggataacgcaggaaag

346    aacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccc

347    ccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttc

348    cccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagc

349    gtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacc

350    ccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactg

351    gcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaacta

352    cggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgat

353    ccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaag

354    aagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaa

355    aaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacag

356    The sequence of E-HIS-RLuc-luc plasmid (The uppercase bases are synthetic enhancer):

357    ttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtaga

358    taactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatt

359    tatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctatta

360    attgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgt

361    cacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaa

362    aagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgc

363    ataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgc

364    ggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgg

365    aaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaact

366    gatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataag

367    ggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggata

368    catatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgcgccctgta

369    gcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc

370    tttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggggctccctttagggttccgatt

371    tagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttt

372    ttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctat

373    tcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaaca

374    aaatattaacgcttacaatttgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgcta

375    ttacgccagcccaagctaccatgataagtaagtaatattaaggtacgggaggtacttggagcggccgcgagctcgcccgg

376    ggatcgagcgcagcggtataaaagggcgcggggtggctgagagcatcagttgtgaatgaatgttcgagccgagcagac

377    gtgccgctgccttcgttaatatcctttgaataagccaactttgaatcacaagacgcataccaaacatgacttcgaaagtttatga

378    tccagaacaaaggaaacggatgataactggtccgcagtggtgggccagatgtaaacaaatgaatgttcttgattcatttatta

379    attattatgattcagaaaaacatgcagaaaatgctgttattttttttacatggtaacgcggcctcttcttatttatggcgacatgttgt

380    gccacatattgagccagtagcgcggtgtattataccagaccttattggtatgggcaaatcaggcaaatctggtaatggttctta

381    taggttacttgatcattacaaatatcttactgcatggtttgaacttcttaatttaccaaagaagatcatttttgtcggccatgattgg

382    ggtgcttgtttggcatttcattatagctatgagcatcaagataagatcaaagcaatagttcacgctgaaagtgtagtagatgtga

383    ttgaatcatgggatgaatggcctgatattgaagaagatattgcgttgatcaaatctgaagaaggagaaaaaatggttttggag

384    aataacttcttcgtggaaaccatgttgccatcaaaaatcatgagaaagttagaaccagaagaatttgcagcatatcttgaacca

385    ttcaaagagaaaggtgaagttcgtcgtccaacattatcatggcctcgtgaaatcccgttagtaaaaggtggtaaacctgacgt

386    tgtacaaattgttaggaattataatgcttatctacgtgcaagtgatgatttaccaaaaatgtttattgaatcggacccaggattctt

387    ttccaatgctattgttgaaggtgccaagaagtttcctaatactgaatttgtcaaagtaaaaggtcttcattttttcgcaagaagatg

388    cacctgatgaaatgggaaaatatatcaaatcgttcgttgagcgagttctcaaaaatgaacaataaagatctgcaataaaatatc

389    tttattttcattacatctgtgtgttggtttttttgtgtgaatcgatagtactaacatacgctctccatcaaaacaaaacgaaacaaaac

390    aaactagcaaaataggctgtccccagtgcaagtgcaggtgccagaacatttctctatcgataggtaCATCGCGTGA

391    TCGCGATCGGATGACTCGGCTGAGTCACCGGATGTGACTCAGCTCTGGTT

392    CTATGATTCATGCAGGAAATGACTCATTGAGTCATTCTTGGAAGTGGCGTG

393    ACTGTCTGGTGGCGTGACTTCCTGGAATGTGACTCAGCGGAAGCTCATGA

394    CTCATTCCTGGAAATCGTACGATGCGAGTCACGTAGGCCTACGTGACGGT

395    GAATCATACTACGTGGCGTGATACGGAGGAAATGACTCagatctgcgatcgagctcg

396    cccgggggatcgagcgcagcggtataaaagggcgcggggtggctgagagcatcagttgtgaatgaatgttcgagccgag

397    cagacgtgccgctgccttcgttaatatcctttgaataagccaactttgaatcacaagacgcataccaaacggcattccggtac

398    tgttggtaaagccaccatggaagacgccaaaaacataaagaaaggcccggcgccattctatccgctggaagatggaacc

399    gctggagagcaactgcataaggctatgaagagatacgccctggttcctggaacaattgcttttacagatgcacatatcgagg

400    tggacatcacttacgctgagtacttcgaaatgtccgttcggttggcagaagctatgaaacgatatgggctgaatacaaatcac

401    agaatcgtcgtatgcagtgaaaactctcttcaattctttatgccggtgttgggcgcgttatttatcggagttgcagttgcgcccg

402    cgaacgacatttataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaagggg

403    ttgcaaaaaattttgaacgtgcaaaaaaagctcccaatcatccaaaaaattattatcatggattctaaaacggattaccaggga

404    tttcagtcgatgtacacgttcgtcacatctcatctacctcccggttttaatgaatacgattttgtgccagagtccttcgatagggg

405    caagacaattgcactgatcatgaactcctctggatctactggtctgcctaaaggtgtcgctctgcctcatagaactgcctgcgt

406    gagattctcgcatgccagagatcctattttttggcaatcaaatcattccggatactgcgattttaagtgttgttccattccatcacg

407    gttttggaatgtttactacactcggatatttgatatgtggatttcgagtcgtcttaatgtatagatttgaagaagagctgtttctgag

408    gagccttcaggattacaagattcaaagtgcgctgctggtgccaaccctattctccttcttcgccaaaagcactctgattgacaa

409    atacgatttatctaatttacacgaaattgcttctggtggcgctcccctctctaaggaagtcggggaagcggttgccaagaggtt

410    ccatctgccaggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacacccgagggggatgataaa

411    ccgggcgcggtcggtaaagttgttccattttttgaagcgaaggttgtggatctggataccgggaaaacgctgggcgttaatc

412    aaagaggcgaactgtgtgtgagaggtcctatgattatgtccggttatgtaaacaatccggaagcgaccaacgccttgattga

413    caaggatggatggctacattctggagacatagcttactgggacgaagacgaacacttcttcatcgttgaccgcctgaagtct

414    ctgattaagtacaaaggctatcaggtggctcccgctgaattggaatccatcttgctccaacacccccaacatcttcgacgcag

415    gtgtcgcaggtcttcccgacgatgacgccggtgaacttcccgccgccgttgttgttttggagcacggaaagacgatgacgg

416    aaaaagagatcgtggattacgtcgccagtcaagtaacaaccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaa

417    gtaccgaaaggtcttaccggaaaactcgacgcaagaaaaatcagagagatcctcataaaggccaagaagggcggaaag

418    atcgccgtgtaattctagagtcggggcggccggccgcttcgagcagacatgataagatacattgatgagtttggacaaacc

419    acaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaa

420    caagttaacaacaacaattgcattcattttatgtttcaggttcaggggggaggtgtgggaggttttttaaagcaagtaaaacctct

421    acaaatgtggtaaaatcgataaggatccatgacatttcccttgctacctgcatacgcaagtgttgcagagtttgataattccttg

| 422 | agtttggtaggaaaagccgtgtttccctatgctgctgaccagctgcacaacctgatcaagttcactcaatcgactgagcttca |
|-----|-------|
| 423 | agttaatgtgcaagttgagtcatccgttacagaggaccaatttgaggagctgatcgacaacttgctcaagttgtacaataatg |
| 424 | gtatcaatgaagtgattttggacctagatttggcagaaagagttgtccaaaggatgatcccaggcgctagggttatctatagg |
| 425 | accctggttgataaagttgcatccttgcccgctaatgctagtatcgctgtgccttttcttctccactgggcgatttgaaaagtttc |
| 426 | actaatggcggtagtagaactgtttatgcttttttctgagaccgcaaagttggtagatgtgacttccactgttgcttctggtataat |
| 427 | ccccattattgatgctcggcaattgactactgaatacgaactttctgaagatgtcaaaaagttccctgtcagtgaaattttgttgg |
| 428 | cgtctttgactactgaccgccccgatggtctattcactactttggtggctgactcttctaattactcgttgggcctggtgtactcgt |
| 429 | ccaaaaagtctattccggaggctataaggacacaaactggagtctaccaatctcgtcgtcacggtttgtggtataaaggtgct |
| 430 | acatctggagcaactcaaaagttgctgggtatcgaattggattgtgatggagactgcttgaaatttgtggttgaacaaacagg |
| 431 | tgttggtttctgtcacttggaacgcacttcctgtttttggccaatcaaagggtcttagagccatggaagccaccttgtgggatcgt |
| 432 | aagagcaatgctccagaaggttcttataccaaacggttatttgacgacgaagtttttgttgaacgctaaaattagggaggaagc |
| 433 | tgatgaacttgcagaagctaaatccaaggaagatatagcctgggaatgtgctgacttattttattttgcattagttagatgtgcc |
| 434 | aagtacggtgtgacgttggacgaggtggagagaaacctggatatgaagtccctaaaggtcactagaaggaaaggagatg |
| 435 | ccaagccaggatacaccaaggaacaacctaaagaagaatccaaacctaaagaagtcccttctgaaggtcgtattgaattgt |
| 436 | gcaaaattgacgtttctaaggcctcctcacaagaaattgaagatgcccttcgtcgtcctatccagaaaacggaacagattatg |
| 437 | gaattagtcaaaccaattgtcgacaatgttcgtcaaaatggtgacaaagcccttttagaactaactgccaagtttgatggagtc |
| 438 | gctttgaagacacctgtgttagaagctcctttcccagaggaacttatgcaattgccagataacgttaagagagccattgatctc |
| 439 | tctatagataacgtcaggaaattccatgaagctcaactaacggagacgttgcaagttgagacttgccctggtgtagtctgctct |
| 440 | cgttttgcaagacctattgagaaagttggcctctatattcctggtggaaccgcaattctgccttccacttccctgatgctgggtg |
| 441 | ttcctgccaaagttgctggttgcaaagaaattgtttttgcatctccacctaagaaggatggtacccttaccccagaagtcatcta |
| 442 | cgttgcccacaaggttggtgctaagtgtatcgtgctagcaggaggcgcccaggcagtagctgctatggcttacggaacag |
| 443 | aaactgttcctaagtgtgacaaaatatttggtccaggaaaccagttcgttactgctgccaagatgatggttcaaaatgacacat |
| 444 | cagccctgtgtagtattgacatgcctgctgtgggccttctgaagttctagttattgctgataaatacgctgatccagatttcgttgcc |
| 445 | tcagaccttctgtctcaagctgaacatggtattgattcccaggtgattctgttggctgtcgatatgacagacaaggagcttgcc |
| 446 | agaattgaagatgctgttcacaaccaagctgtgcagttgccaagggttgaaattgtacgcaagtgtattgcacactctacaac |
| 447 | cctatcggttgcaacctacgagcaggctttggaaatgtccaatcagtacgctcctgaacacttgatcctgcaaatcgagaatg |
| 448 | cttcttcttatgttgatcaagtacaacacgctggatctgtgtttgttggtgcctactctccagagagttgtggagattactcctccg |
| 449 | gtaccaaccacactttgccaacgtacggatatgcccgtcaatacagcggagttaacactgcaaccttccagaagttcatcac |
| 450 | ttcacaagacgtaactcctgagggactgaaacatattggccaagcagtgatggatctggctgctgttgaaggtctagatgct |
| 451 | caccgcaatgctgttaaggttcgtatggagaaactgggacttatttaagtcgaccgatgcccttgagagccttcaacccagtc |
| 452 | agctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgactgtcttctttatcatgcaactcgtaggacaggt |
| 453 | gccggcagcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactc |
| 454 | aaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggc |
| 455 | caggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctc |
| 456 | aagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgt |
| 457 | tccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggta |
| 458 | tctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatcc |
| 459 | ggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcaga |
| 460 | gcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatct |
| 461 | gcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtg |
| 462 | gtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgc |
| 463 | tcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttaaattaaaaa |
| 464 | tgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacag |

## The sequence of synthetic enhancers

The optimized high activity developmental enhancers:

Generation:79_individual:1:
TGATGTGACTCACATAGGCGGTGGCGTGATATGTTGTGACTCATTTCCCGG
AAACGGATGACTAATGCCATATGTTATCAGTTTCCTGGAAATTTGATCACG
CCATATTGTGAAATCATGCGATTCCCGGATCACGTGACGGCCGGACGTGA
CAAGTATGAGTCACTAAGTGGCGTGATCTTACGAATCACGTGATGGTCAA
TGTCACGTGATCGGCTGGTGAGTCAGCAATATCGTGTGATTCATTC

Generation:79_individual:2:
TGATGTGACTCACATAGGCGGTGGCGTGATATGTTGTGAGTCACTTCCCG
GAAACGGATGACTAATGCCATATGTTATCAGTTTCCTGGAAATTGAATCA
CGCCATATTGTGAAATCATGCGATTCCCGGATCACGTGACGGCCGGACGT
GATGACTAATCAGCAAAAGTGGCGTGATCTCCGGAATCACGTGATGGTC
CATGTCACGTGATCGGCTGGTGAGTCAGCAATATCGTGTGATTCATTC

Generation:90_individual:1:
GCGTGATTATATCCGGTTTGTGGCGTGACTACAACCGGATGCATATCCGGT
CATCCGGTTATGTCACGCCACTGTCACGCCATTCCAGGAAATCATCCGATC
GATCACGCCACCGGATGATGGCGTGATCGAACATCCGATCACGCCATCCG
GAAATCGTACGATGTGGCGTGATCGCGTGATCTGGCGTGATTATTCCAGG
AATTGCGAATGAGTCATCACGCGACGTCATGACTCATAGGTGATAAG

Generation:90_individual:2:
CATCGCGTGATCGCGATCGGATGACTCGGCTGAGTCACCGGATGTGACTC
AGCTCTGGTTCTATGATTCATGCAGGAAATGACTCATTGAGTCATTCTTGG
AAGTGGCGTGACTGTCTGGTGGCGTGACTTCCTGGAATGTGACTCAGCGG
AAGCTCATGACTCATTCCTGGAAATCGTACGATGCGAGTCACGTAGGCCT
ACGTGACGGTGAATCATACTACGTGGCGTGATACGGAGGAAATGACTC

The optimized high activity housekeeping enhancers:

Generation:20_individual:1:
ACCGGCACACTAGAGCGACTAAGTCTGGGTGGCAACAGTCCGATACTTAC
AGCGTATGAGTGGCTGCGATGCGGCTAGAGATGGGCCGCGTTTTCCACTC
GATAGCATTCTAAGGCTTCCATCTCTAGGGCCGAAATCACGACACGGCAG
CTTGACGAGTCAGATGAGTCCGCGTATCGATATTCTCACCTCTATTAGGAA
AGCCATATGTGCTATCGATATGTTCATTAATCGCGGGCTTTTAGGGCT

Generation:35_individual:554:
CTCGCACGAGTTTGACACAACTCTATTCGTCCACCACTAGCCGCCCACATC
TCCGAGGCACGTCTATCGATAGCATCACTGACGCACTATCGATAGCCCAC
CGTAGTGGTTATCGAATATGCGGCTATCGAGACGTGAACTATCGATATGTT
CGAGATCTACCGCTAGCGTCCGTATCGATATTCTGATTGTGATGGACGCAT
CTGGTATCGATACAAACTATCGATAGGAAGAAAGACTGAACCGGTT

Generation:60_individual:677:
CTGGTAGAGAGCGAGCAGTTATCGATAGGTCCACCACTAGCCGCCCACAT
CTCCGAGGCACGTCTATCGATAGTGTACTATCGATAACTATCGATTGCTCC
GCCTATCGATAGACGTTGCCATCTCTAGCGGACGTGAACTATCGATACGT

508 ATCGATAGTTAGACGCGGACGACTATCGATTAGTTGACGTATCGATAGTA
509 TGCGCCTATCGATAGTGCCTATCGATAGGCAGAACGCGTATCGATAAT
510 Generation:79_individual:1426:
511 ACTTAGAGACATCGATAGTTATCGATACGTCCACCACTAGCCCATCTATCG
512 ATAGACTAACGTCTATCGATATATCGCCCGACGCACTATCGATAGGCTCC
513 GCCTATCGTCCGCGCGACTATCGATATGTATGCCCATCCCTAATGCGCGTA
514 TCGATAGTTATCGATACACGACTATCGATAGCTAGAACTATCGATAGTAT
515 GCGCCTATCGATAGTGCCTATCGATAGGCGGAACGCGTATCGATAAC
516 Generation:95_individual:99983:
517 ACTTCGTACTATCGATAGTTATCGATTGGCCCACCACTAGCCGCCCTATCG
518 ATAGCGGCACGTCTATCGATATATCGCCCGACGCACTATCGATAGTATCC
519 GCCCATCGATAGTATCGCCCATCTCTATGCGTCCCATCCCTAATGCCACTA
520 TCGATAGTTAGACGCACACGACTATCGATAGTTAGAACTATCGATAGTAT
521 CCGCCTATCGATAGCGACTATCGATAGCCTATCGATAGGTCGATAAC
522 Generation:99_individual:99988:
523 AGAACGTACCATCGATAGTTATCGATAGGTCCACCACTAGCCGCGACTAT
524 CGATTGTCCATGCCTATCGATATATCGCCCGACGCACTATCGATAGTATCC
525 GCCTATCGATAGAATCGCCCATCTCTATGCGTCCCATCCCTAATGCGCCTA
526 TCGATAGTTATCGATACGCGACTATCGATAGTTAGAACTATCGATAGTATG
527 CGCCTATCGATAGCGACTATCGATAGCCTATCGATAGGTCGATAGC
528 The optimized high activity housekeeping silencers:
529 Generation:60_individual:133:
530 CTGTCCGAGATACGGCTGGTGGTAAGCACGAGCCCAGCCGGATGCGCGGG
531 ACCGTCGCCCGGCTAGAAAGACCAGGTAAGTAACGCCCATGAAGGGCCG
532 CGGCAGGTAAGTTAGGACAGGTAAGTGCAAGGACAGGTAAGTCCTACCT
533 GCAGAGACTCCTTATCAAGGAGATAAACGTACTTACCTTACCAGGTACTT
534 TCCGGTCGCAGTACTTACCTACTTACCTATGTTCCATATCCGAGGCCATGG
535 Generation:79_individual:72:
536 GGTGATCTGAGAGCCCACAGGTAAGTTGAGCTCACAGGTAAGTAGCCCGC
537 AGGTAAGTCTGGGCTGTGCTCACTTACCTACGTATACGCAGTCCCAAGCA
538 GGTAAGTACAGGTAAGGTAAGTAGGTAAGTAAGTACTTACCTACCTACTT
539 ACCTTGATATCGTATACTTACCTACCTACGCACACTTGGCTTGACTTACCT
540 ACTTACCTATTTCCGGTCCGCGTTGGGGCCCGGCAGGTAAGTTAGTAA
541 Generation:79_individual:613:
542 CCCGGATGCCTGCGGCTGGTGGTAAGCACGAGCCCAGCCGGAAGCTCAGA
543 GGAAGTAGTTGCAGCCCCAGGTCAGGTAAGTAACCGCTGCCAACCCGCAG
544 GGCAGGTAAGTTAGGACAGGTAAGTGCACGGACAGGTAAGACAGGTAAG
545 TAGAGACTCCTTATCTTATCCTTATCTCTACTTACCTTACCAGTTACTTCCG
546 TCTTATCGTACTTACCTACTTACCTATCTTTCATATCCGGCCATAGCC
547 Generation:99_individual:99994:
548 CGAAGCATATCTGCCCACAGGTAAGTTGAGCTCACAGGTAAGTTGCCTGC
549 AGGTAAGTATGTCCTGCGCTCACTTACCTACGTATACGCAGAGGTAAGTA
550 GGTAAGTTCAGGTAAGGTAAGTAGGTAAGTAAGTACTTACCTAGATACTT

551 ACCTTGCGATGGCATACTTACCTACCTATCTCAGGGCTTACCTACTTACCT
552 ACTTACCTATTTCCGGCTCTGTGGTAAGTCTTGCGTATACTTACCAAC
553 Generation:99_individual:99956:
554 CCCGCCGAGATACGGCTGGTGGTAAGCACTCGCCCGGACGGAAGCTCCCC
555 AGTGGGACACGCAGCCCAAGGACAGGTAAGTACTTCCTGTCGGTAGGCAG
556 GTAAGGTAAGTAAGGACAGGTAAGTGCACGGACAGGTAAGTCAGGTAAG
557 TAGATAGTCCTTATCTTATCCTTATCTGTACTTACCTTACCAGGTACTTCCG
558 TCTTATCGTACTTACCTACTTACCTATCTTTCATATCCGGCCATAACC
559 The optimized high activity developmental enhancers with 3 fixed RESs (AgeI =
560 "ACCGGT", SalI = "GTCGAC", HindIII = "AAGCTT", marked in bold font):
561 Generation:80_individual:1:
562 AACTGACTCACTCGTCACGCCACTTCCTGTTCGTGGATGACTCATTCCAGA**CC**
563 **CCGGT**GGCGTGATGATTCATTTCCGGTGACTCATTTCCGGGATGGCGTGA
564 CCGGAAGTGACTCACAGATCACGTGATGGCGTGATCGGATGATCATCCG**G**
565 **TCGAC**ATCCGGTGACTCACTTCCGGTCGGAAGTGGCGTGATCACGCCAC**A**
566 **AGCTT**CCGGGTCTTAGTCATGCTGCTGAGTCATTTCCTGTTGAAAACC
567 Generation:90_individual:1:
568 TGAAATGATCACGTGATCACGTGATCGCGTGATCTTGCCTGAGTCATCCA
569 **ACCGGT**GAATCACTGGATGACTAATATGGCGTGATCGGAAATGACTCAG
570 ACCGGAAATCACGTGATGTGAGTCACAGATGGCGTGATCGGATCACGTGA
571 T**GTCGAC**CGGATGTGACTCACTGGATGACTCAACATGGCGTGATCATGAC
572 T**AAGCTT**CCGGTGACTCAGCATGGCGTGATCACGTGATTGTGATTCATTG
573 The optimized high activity housekeeping enhancers with 3 fixed RESs (AgeI =
574 "ACCGGT", SalI = "GTCGAC", HindIII = "AAGCTT", marked in bold font):
575 Generation:80_individual:1:
576 TACGCTGTATTTTTTTTTTTCGTTTCCGGTTATCCAAATAATTGGACGTGA**AC**
577 **CGGT**GGCGTGATGGCGTGGTTGACGTGTTATCGATTTATCGATATCGATA
578 AATATCGATGGCGTGATGACGTGTCGTTGATGACTAACGTCACGTAGT**GT**
579 **CGAC**GTGATTGACTCATTCCGATCACGTGATGTTTCTAAAGCGAAAGG**AA**
580 **GCTT**GAAAGGGGGTGGGGAGGGGGTGGTGAGGCAAAGGGATCACGAA
581 Generation:90_individual:1:
582 ATGTGACCCGACTATCGATTACTGAACCGCGCCTATCGCTTATCGACTGA**A**
583 **CCGGT**GCTGTTAGACGTGACTATCGATACTTAGCCTATCGATAGTTTGAG
584 ACTATCGATACACGAACCTATCGATAGTCAGCGCCGCGGACTATCGATA**G**
585 **TCGAC**AACTATCGATACGCACAACTATCGATACGCGAACTATCGATAGT**A**
586 **AGCTT**CCGCGTGACTATCGATAGCACGCTATCGATAGTCGCGACTGTC
587 The optimized AT-rich high activity developmental enhancers:
588 Generation:80_individual:1(GC:0.11):
589 TAATTTTATATATAAATTTTTTATGAATATTTAAATTTATAATAATTTATTA
590 TTTTTATATTTTTAAAATGATTCATAATAAATAAAAATAATCACATGATTT
591 TAATTTATGAATCATATCACGTGATGGCGTGATCGTATGATCATGATTAAT
592 AATAAAATTATGAATCATAAAATGACTAATACAAAATTCTTAGAAATTTA
593 TAAATTAAAATAAAATATAATTATTTATAAATTTTTTATATATTA
594 Generation:90_individual:1(GC:0.13):

595 AAAATTTATATAATATTTTTATATATATATAATTTTATATATAAATACATAT
596 ATAATTATGATTCATTTTATAAGTATGAATCATAGTCATATGACTATTATA
597 TGATTCATTTTTTGATGATTCACTTCTTGGAATGAATCATTTTTAATGATTC
598 ATATAATGATTCATAAAAAATTATGAATGATGAAATTCTTAGAATGAATC
599 ATTATAAATAATAAATTTAATAATTATGATGAAATATAATAAAT
600 The optimized AT-rich high activity housekeeping enhancers:
601 Generation:80_individual:1(GC:0.09):
602 AATTAATTTAATTATTTTTTTTTTTTTTTATTATTTTTTTTTATTTTTTTATGAT
603 GAAATATAAAAAAAAAAATATATCGATTAATTATCGATATAATAATTA
604 TCGATATATTTAGTATTAGTCATCATATGACTCATATATTTTAAATTAGTC
605 ATTAATATGATTCAAAAAAAATGACTCATATAATATTAAATTTTTTTTTAT
606 ATTTAATTTTTTTTATTATTAAAAAAAAAATTAAAAAATTTAA
607 Generation:90_individual:1(GC:0.08):
608 ATTATAATAATATTTTTATAAATTATTTTTTTTTTATATTAAAAATATATTT
609 AAAATAATATTTAATTTTATAATTAATTTTTAAAAAAAATTTAAATAAATT
610 ATCGATATATTTAGTATTTTATCGATATTATAAAATACTAAAAAAAATATC
611 GATATTTAGATTAATCGATATATCGATACATAAAAATATCGATATATCGA
612 TTATTATAACTATCGATATATTATTTTAAAATAATTTTTATTAT

## The sequence of CMV enhancer

614 CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
615 CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATA
616 GGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCA
617 CTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT
618 CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAT
619 GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
620 TG

## The sequence of Hr5 enhancer

622 CGCGTAAAACACAATCAAGTATGAGTCATAAGCTGATGTCATGTTTTGCA
623 CACGGCTCATAACCGAACTGGCTTTACGAGTAGAATTCTACTTGTAACGC
624 ACGATCAGTGGATGATGTCATTTGTTTTTCAAATCGAGATGATGTCATGTT
625 TTGCACACGGCTCATAAACTCGCTTTACGAGTAGAATTCTACGTGTAACGC
626 ACGATCGATTGATGAGTCATTTGTTTTGCAATATGATATCATACAATATGA
627 CTCATTTGTTTTTCAAAACCGAACTTGATTTACGGGTAGAATTCTACTTGT
628 AAAGCACAATCAAAAGATGATGTCATTTGTTTTTCAAAACTGAACTCGC
629 TTTACGAGTAGAATTCTACGTGTAAAACACAATCAAGAAATGATGTCATT
630 TGTTATAAAAATAAAGCTGATGTCATGTTTTGCACATGGCTCATAACTAA
631 ACTCGCTTTACGGGTAGAATTCTACGCG

## The sequence of UASE enhancer

633 AATAGAGGAACCGTTTACCTGTGGTTCCTATTGTGGCCTACTGTTACTAGC
634 TAGTGTAATACACCCTTGCCTCAGCTTTGCAAGTTGACAACTCAGCCAA

## Vector and DNA construction for luciferase transgenic Drosophila lines

The pVA-Enhancer-DSCP-Luciferase-SV40 vector is constructed based on the pVA-luciferase vector which is given from the Pro. Ni (Tsinghua University). Briefly, pVA-luciferase was cut with KpnI and SpeI and ligated with an amplified fragment of SV40 from pNP (gift from Pro. Ni in Tsinghua University, fwd: 5'-GGGGTACCGGATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGAC-3', rev: 5'- GCACTAGTTGGAACCAGACATGATAAGATACATTGATGAGTTTGGAC-3'). The resulting vector was cut with XbaI and BamHI and ligated with an amplified fragment from the following steps: Amplify the optimal enhancer sequence (fwd: 5'-CATCGCGTGATCGCGATCGG-3', rev:5'- GAGTCATTTCCTCCGTATCAC-3') and DSCP promoter sequence (fwd:5'-GATACGGAGGAAATGACTCGTCGACGAGCTCGCCC-3', rev:5'-CGGGATCCTTTGGTATGCGTCTTGTGATTCAAAGTTG-3') from the corresponding vector then use the overlap PCR process to amplify the enhancer-DSCP fragment. Then the pVA-Enhancer-DSCP-Luciferase-SV40 was constructed. The pVA-Enhancer-DSCP-Luciferase-SV40-Enhancer were constructed in the same way.

## Transgenic Drosophila construction

The construction of luciferase transgenic flies is based on the previous protocol (20) with some modifications. Briefly, the pVA-Enhancer-DSCP-luciferase-SV40, and pVA-Enhancer-DSCP-luciferase-SV40-Enhancer were injected into the TB16 and TB18 to get the transgenic *Drosophila*.

## Luciferase assay *in vivo*

Luciferase activity was measured using the Steady-Glo Luciferase Assay Kit (Promega). A single adult fly (2-day-old) was collected and well-grinded in 100μL GLO lysis buffer for different genotype samples, each luciferase assay contained three independent samples. Samples were homogenized and centrifuged at 12,000xg for 5min, 30 μL supernatant was taken into the 96-well plates. Then add the 30 μL luciferin into the supernatant and well-mix. After incubation in the dark for 5min, luminescence was measured on a luminometer (Thermo Scientific, VAR-IOSKAN FLASH).

## Recombinase-based enhancer integration in human genome and enhancer activity measurement

The designed enhancers and the CMV enhancer were cloned into plasmids containing attB sites and a green fluorescent protein (GFP) marker. Each plasmid containing a designed enhancer was co-transfected at a 1:1 ratio with a plasmid expressing serine integrase into 293T landing pad cells, which possess attP integration sites (cell lines were sourced from our laboratory). Empty vectors were included as controls. After 10 days of culture, non-integrated plasmids were degraded, and GFP-expressing cells indicating successful integration were sorted using flow cytometry. Monoclonal cells obtained from the sorting process were seeded into 96-well plates and cultured in medium supplemented with 20% fetal bovine serum. Positive clones were expanded and imaged using a fluorescence microscope, and their average fluorescence intensity was quantified with ImageJ software (https://imagej.net/ij/).

## Training DREAM with human multi-cell line STARR-seq data

We collected STARR-seq data from the A549 (Accession: ENCSR895FD, 70,011 peaks), HCT116 (Accession: ENCSR064KUD, 43,718 peaks), and MCF7 (Accession: ENCSR547SBZ, 79,466 peaks) cell lines, all obtained from the ENCODE project (https://www.encodeproject.org/). These datasets share a common input DNA library (Accession: ENCSR547SBZ), effectively minimizing batch effects. The human genome was segmented using a step size of 300 bp and a window size of 400 bp. To broaden the activity spectrum of enhancers within the dataset, we randomly selected 420,000 non-peak regions (background regions) where input DNA read counts exceeded 20. Consequently, the dataset was comprised of peak regions from A549, HCT116, and MCF7 cell lines, along with randomly selected background regions. We used regions from chromosomes 21 and 22 as the validation set (7,995 regions) and test set (11,339 regions), respectively. Enhancer activity was quantified as the log2 fold change of RNA read counts mapped to genomic regions relative to the input DNA read counts. Except for modifying the output dimension of SENet's final layer to three, all

693    other hyperparameters were kept consistent. We retrained SENet on these datasets and

694    evaluated its performance using the hold-out test dataset from chromosome 22.

695

## Supplementary Figures

Supplementary Figure 1. The structure of the SENet

| input_1 | input: | [(None, 249, 4)] |
|---|---|---|
| InputLayer | output: | [(None, 249, 4)] |

| conv1d | input: | (None, 249, 4) |
|---|---|---|
| Conv1D | output: | (None, 243, 512) |

| batch_normalization | input: | (None, 243, 512) |
|---|---|---|
| BatchNormalization | output: | (None, 243, 512) |

| activation | input: | (None, 243, 512) |
|---|---|---|
| Activation | output: | (None, 243, 512) |

| average_pooling1d | input: | (None, 243, 512) |
|---|---|---|
| AveragePooling1D | output: | (None, 120, 512) |

| sequential | input: | (None, 120, 512) |
|---|---|---|
| Sequential | output: | (None, 120, 512) |

| sequential_3 | input: | (None, 120, 512) |
|---|---|---|
| Sequential | output: | (None, 120, 1024) |

| sequential_6 | input: | (None, 120, 1024) |
|---|---|---|
| Sequential | output: | (None, 60, 2048) |

| sequential_9 | input: | (None, 60, 2048) |
|---|---|---|
| Sequential | output: | (None, 30, 2048) |

| flatten | input: | (None, 30, 2048) |
|---|---|---|
| Flatten | output: | (None, 61440) |

| dropout_8 | input: | (None, 61440) |
|---|---|---|
| Dropout | output: | (None, 61440) |

| sequential_12 | input: | (None, 61440) |
|---|---|---|
| Sequential | output: | (None, 64) |

| sequential_13 | input: | (None, 61440) |
|---|---|---|
| Sequential | output: | (None, 64) |

| Dense_Dev | input: | (None, 64) |
|---|---|---|
| Dense | output: | (None, 1) |

| Dense_Hk | input: | (None, 64) |
|---|---|---|
| Dense | output: | (None, 1) |

The structure of the SENet.

702    Supplementary Figure 2. Performance evaluation of SENet's predictions in test set
703    excluding repeats.



704
705
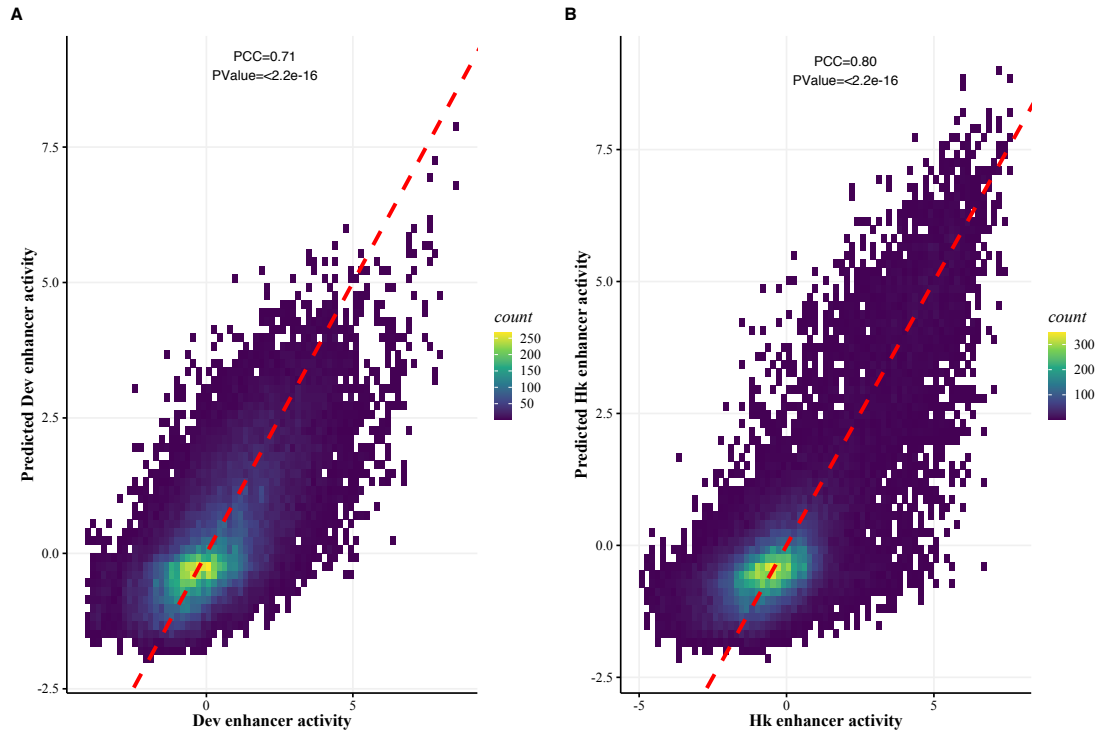
706    Scatter plots of SENet predicted vs. observed developmental (A) and housekeeping (B)
707    enhancer activity signal across 32,036 DNA sequences in the test set chromosome not
708    overlapping with repeats. The PCC is denoted for each comparison.
709

710  Supplementary Figure 3. Performance evaluation of SENet's predictions of chromatin
711  accessible regions in test chromosome (23,506 DNA accessible regions).
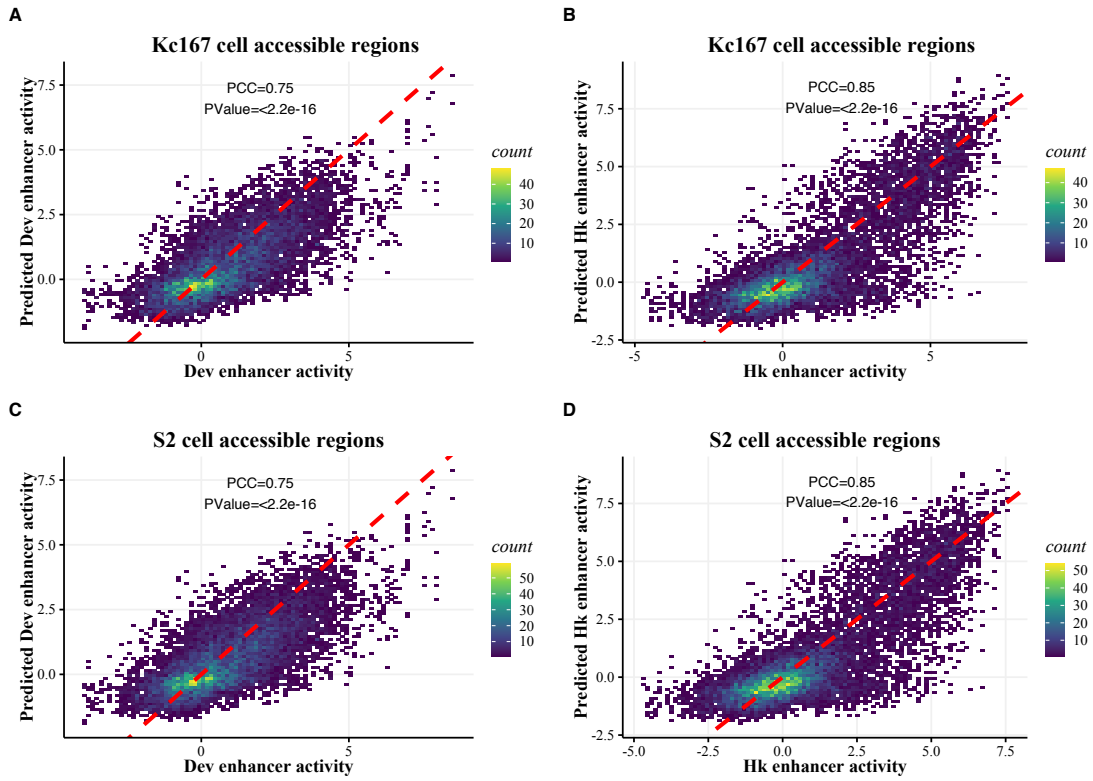


712
713  Scatter plots of SENet predicted vs. observed developmental (A, C) and housekeeping
714  (B, D) enhancer activity signal across DNA sequences in the test set chromosome
715  overlapping with accessible region of Kc167 cell (A, B; 10746 DNA sequences) and
716  S2 cell (C, D; 12760 DNA sequences). The PCC is denoted for each comparison.
717

718    Supplementary Figure 4. Performance evaluation of SENet's predictions of synthetic
719    enhancers in *Drosophila* S2 cells
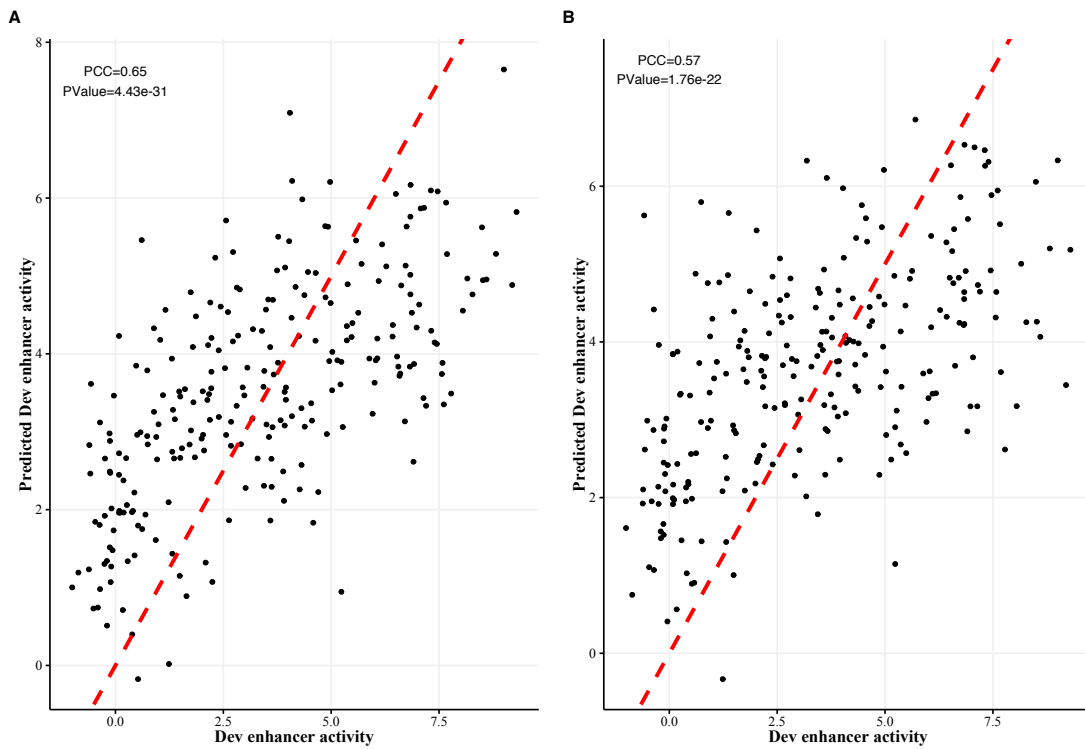


720

721    (A) Scatter plot of SENet predicted vs. experimentally measured activity of synthetic
722    enhancers in *Drosophila* S2 cells. (B) The Scatter plot of DeepSTARR predicted vs.
723    the same experimentally measured activity of synthetic enhancers in *Drosophila* S2
724    cells. The synthetic enhancers activity data sourced from the de Almeida *et a*l. (1) The
725    PCC is denoted for each comparison.
726

727    Supplementary Figure 5. Dataset distribution and model performance evaluation using
728    10-fold cross validation
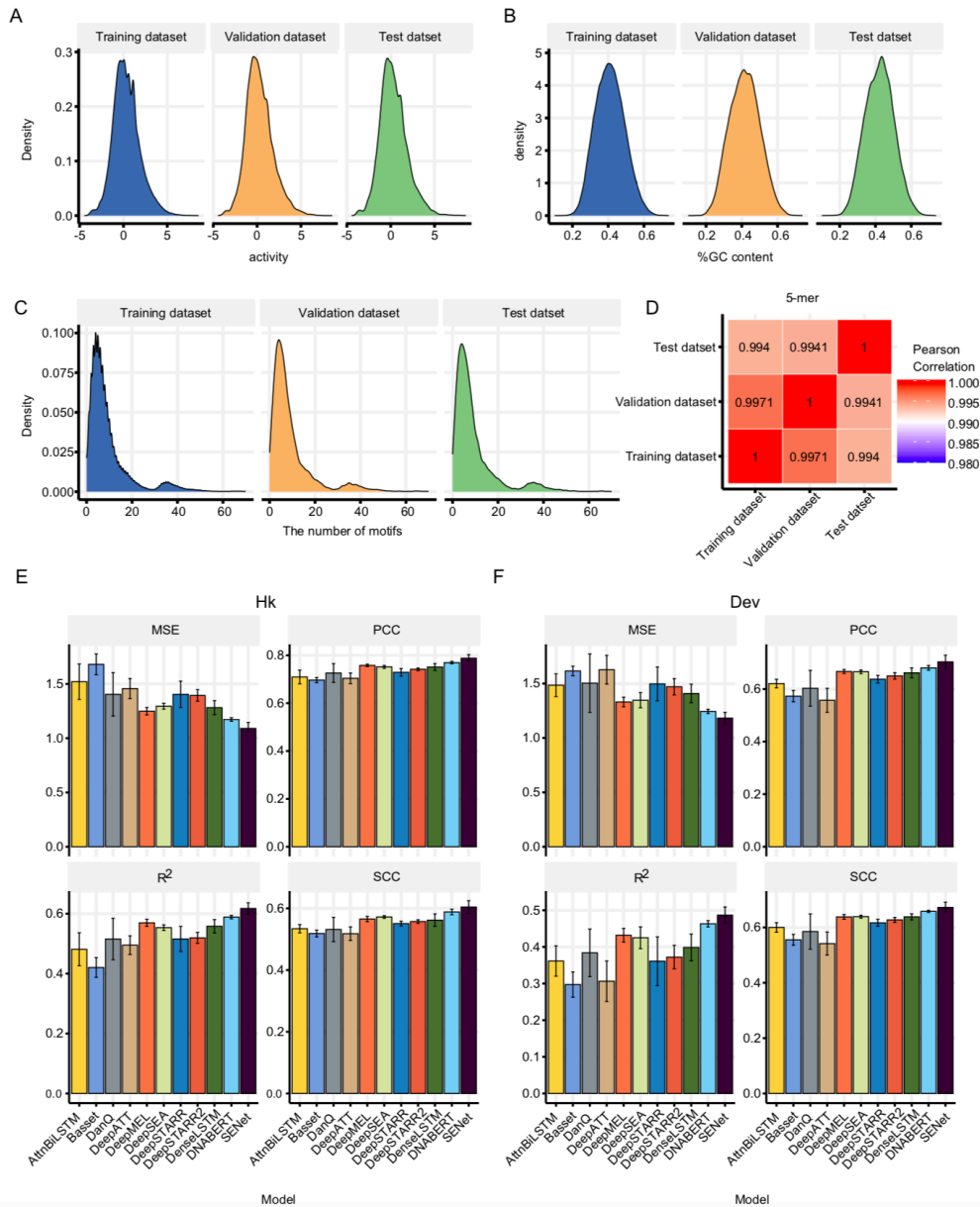


729
730    (A), (B), and (C) display the distributions of sequence activity, GC content, and the
731    number of transcription factor (TF) motifs per sequence in the training, validation, and
732    test datasets, respectively. Panel (D) illustrates the 5-mer frequency similarities among
733    sequences across these datasets. (E) and (F) present the 10-fold cross-validation results
734    for 11 models, including SENet, on an independent test dataset, with performance
735    metrics represented as bar graphs for Mean Squared Error (MSE), Pearson Correlation
736    Coefficient (PCC), Coefficient of Determination (R²), and Spearman Correlation
737    Coefficient (SCC).

738    Supplementary Figure 6. Some functional motifs are capture by different filters
739    recurrently, which implies that the filters are highly redundant
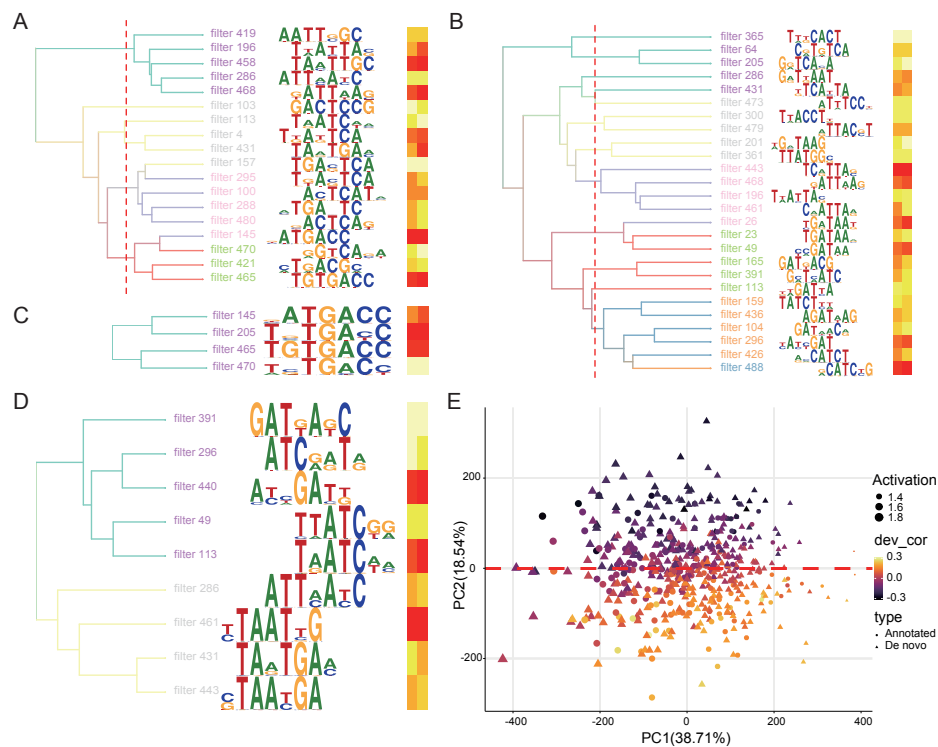


740
741    The filters in the first convolutional layer of SENet can recover the kay (A), GATAd
742    (B), Ohler1 (C), and DRE (D) motifs associated to enhancer activity. The heatmap
743    indicates the influence on the predictions of developmental and housekeeping enhancer
744    activity, respectively. (E) Discovered DNA motifs associated with enhancer activity.
745    Clustering of 512 motifs discovered by DREAM framework. Shown are the first two
746    principal components of the motif occurrence frequencies in sequence windows
747    (activity). Triangles represent the *de novo* motifs and dots denote motifs with significant
748    (FDR <0.05) similarity to annotated motifs in the databases. Marker size indicates the
749    average activity; the estimated motif effect on housekeeping enhancer activity is shown
750    by colour.
751

754
755

Six attributes of the *de novo* motifs and annotated motifs learned by SENet are
presented, including (A) information content, (B) activation of filters in the initial
convolutional layer of SENet, association with developmental enhancers (C), and
housekeeping enhancers (D). Additionally, the PhastCons scores within motif instances
of developmental enhancers (E) and housekeeping enhancers (F) are displayed (one-
sided Wilcoxon rank-sum test).
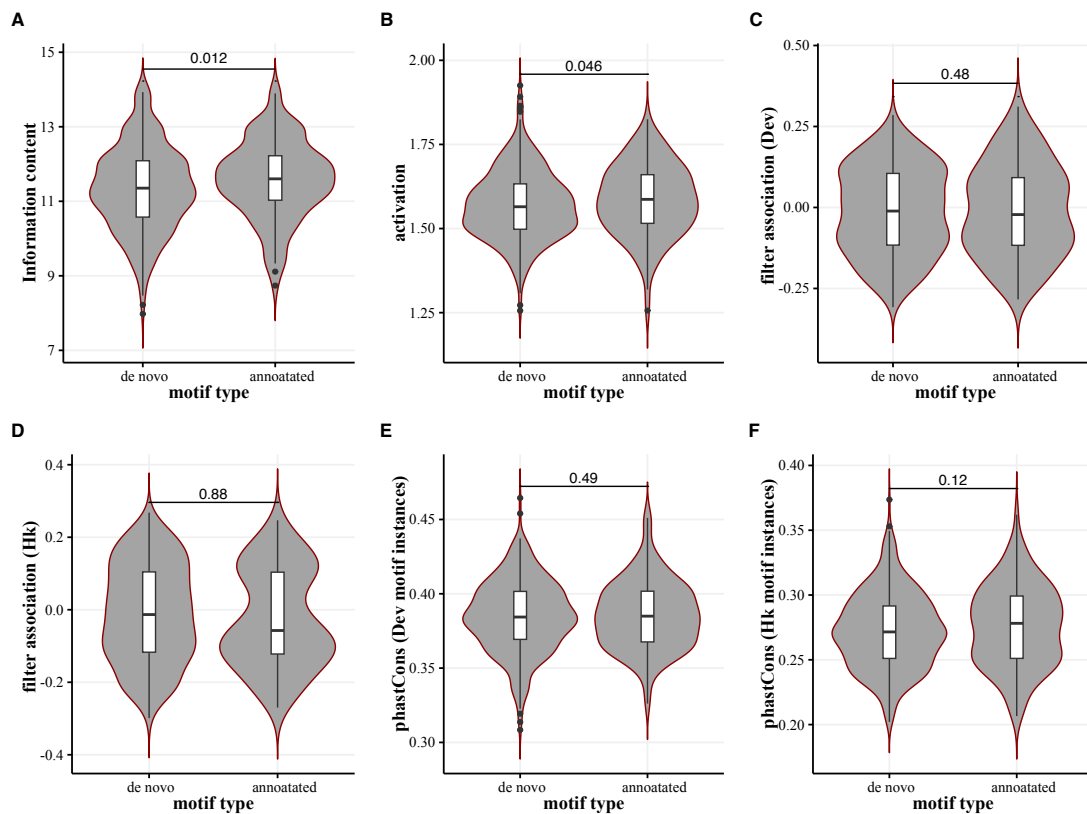
756
757
758
759
760
761
762
763

Supplementary Figure 8. SENet learned motifs can predict enhancer activity, matching
or surpassing known TF motifs.
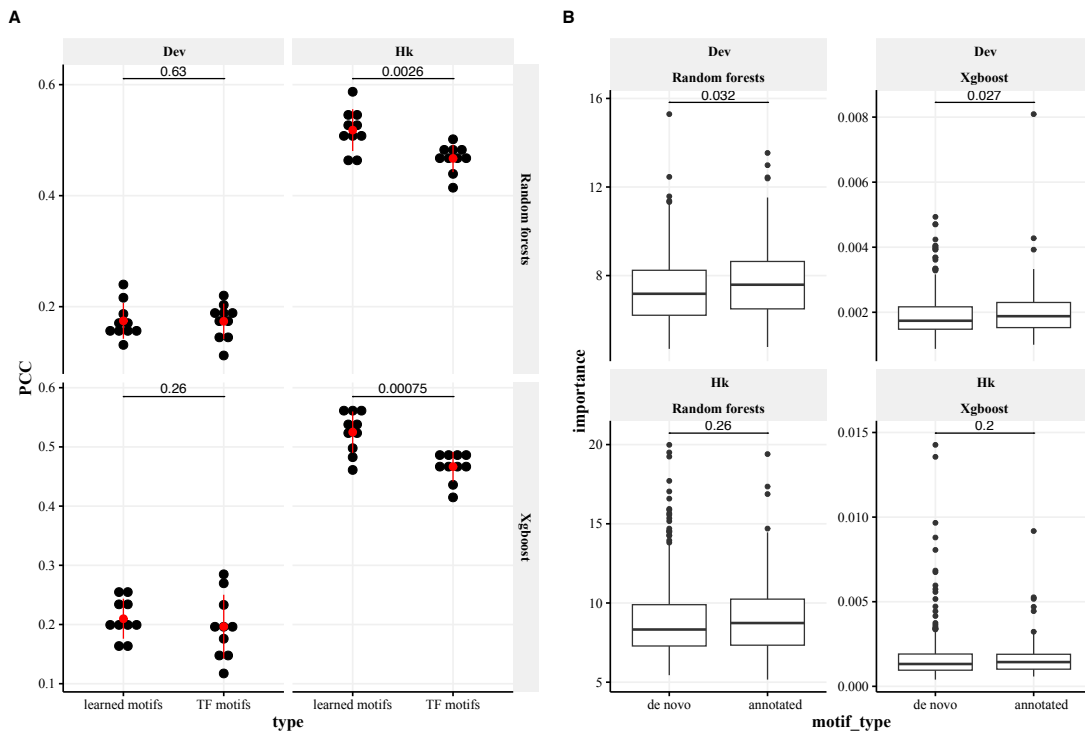


(A) Random forests and XGBoost models were constructed using SENet-learned motifs
and known TF motifs from the JASPAR database as features to predict the activity of
development and housekeeping enhancers, respectively. The ten-fold cross-validation
results are presented in the figure, showcasing that SENet-learned motifs exhibit the
capability to predict enhancer activity, either matching or surpassing the predictive
power of known TF motifs. The red dots represent the mean PCC. (B) The feature
importance for the annotated motifs (can match with the JASPAR database) and *de
novo* motif (one-sided Wilcoxon rank-sum test).

776 Supplementary Figure 9. Two complementary metrics, the activation and the influence
777 of filters reflect the evolutionary constraints of the matched DNA sequences and
778 recovered TFs, respectively.
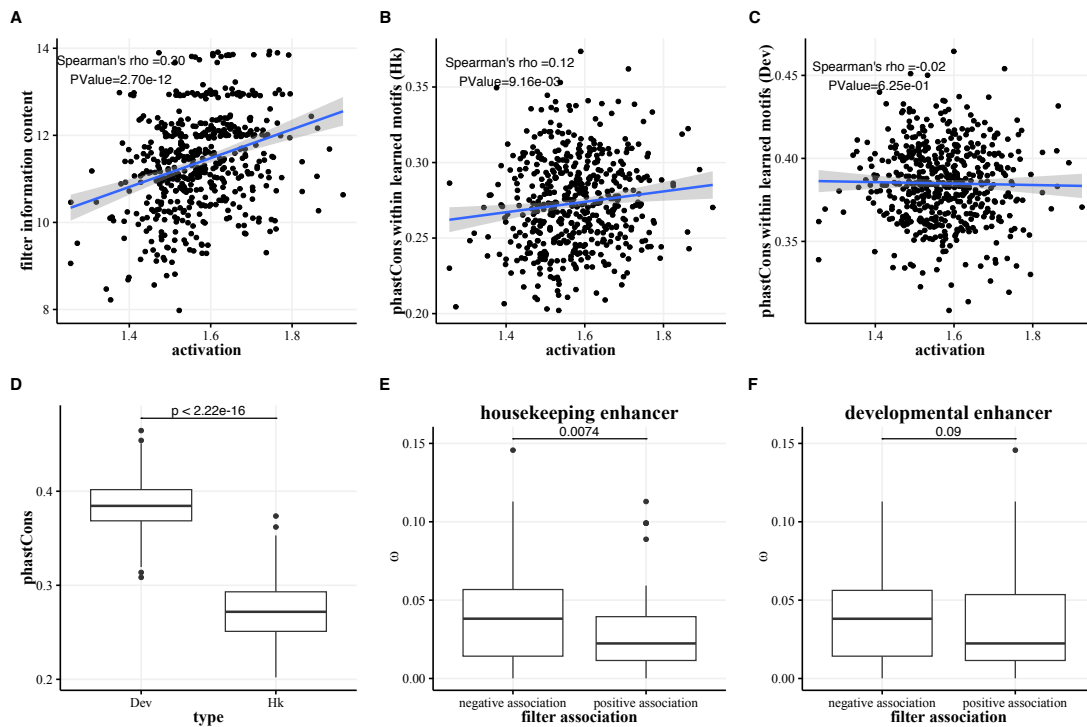


779

780 (A) Filters exhibiting higher activation in the first convolutional layer of SENet
781 correspond to motifs with elevated information content. (B) Depicts the motif instances
782 (matching the filters) exhibiting increased activation within housekeeping enhancers,
783 which show greater evolutionary conservation (15-way PhastCons score, obtained from
784 https://hgdownload.cse.ucsc.edu/goldenPath/dm3/phastCons15way/). However, no
785 statistically significant correlation was detected in developmental enhancers (C). (D)
786 Developmental enhancers exhibit greater evolutionary conservation compared to
787 housekeeping enhancers (one-sided Wilcoxon rank-sum test). (E) and (F) Illustrate
788 filters with higher influence on activity prediction capturing TFs with lower
789 evolutionary rates (one-sided Wilcoxon rank-sum test). The blue line represents linear
790 regression between x and y variables for the corresponding figures, with the grey region
791 indicating the 95% confidence interval. The PAML 4.9 (2) was used to calculate
792 evolutionary rates (dN/dS,) of captured TFs based on their orthologous genes from the
793 same six Drosophila species used to calculate dN/dS previously (3).

Supplementary Figure 10. Correlation of DNA sequence motifs
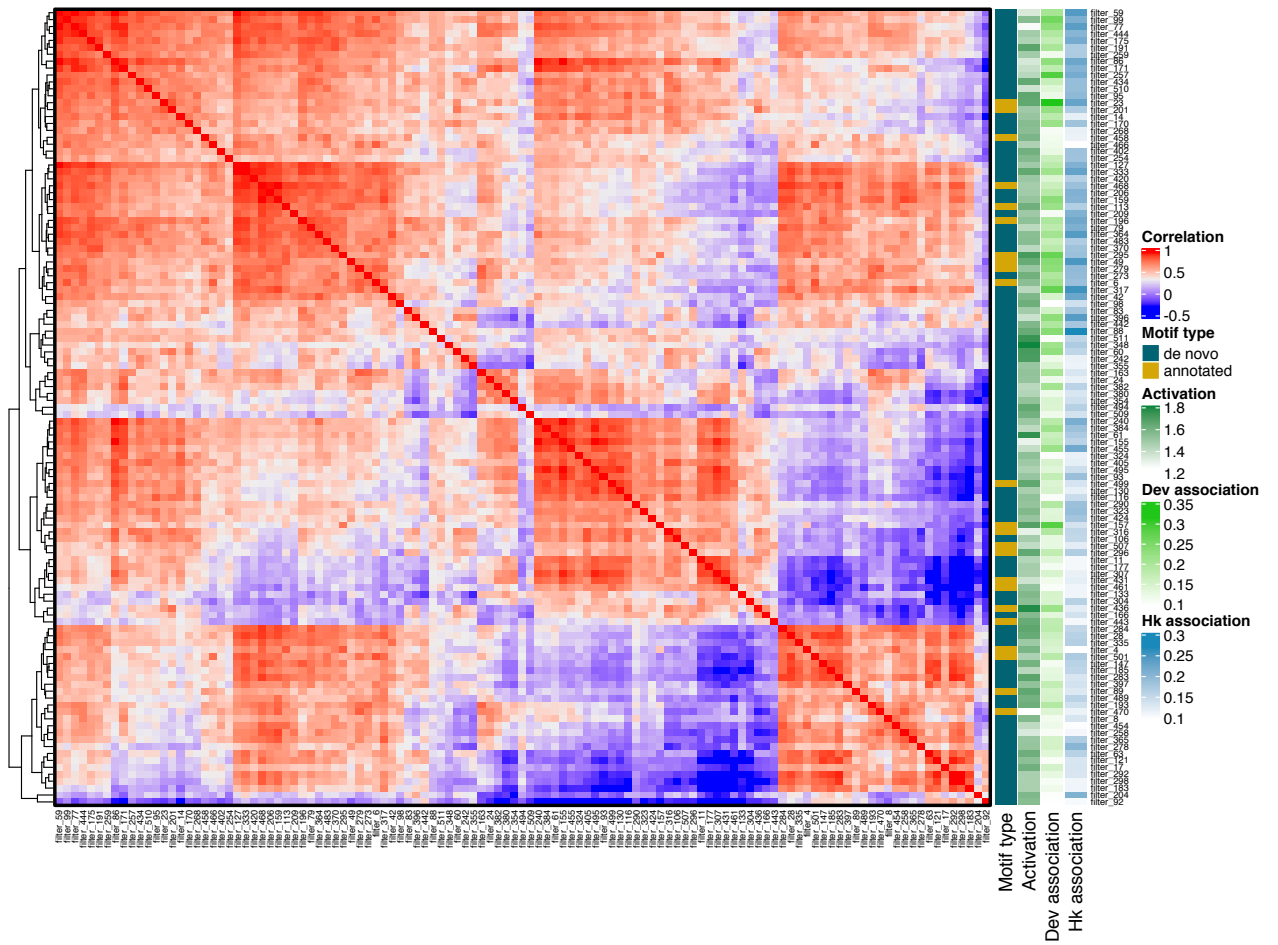


795

796 Pairwise Pearson correlation coefficients between motif activities, showing clusters of
797 motifs with large association with enhancer activities (association > 0.1).

798

799

800 Supplementary Figure 11. The comparison of the DNA features learned by SENet and
801 DeepSTARR.
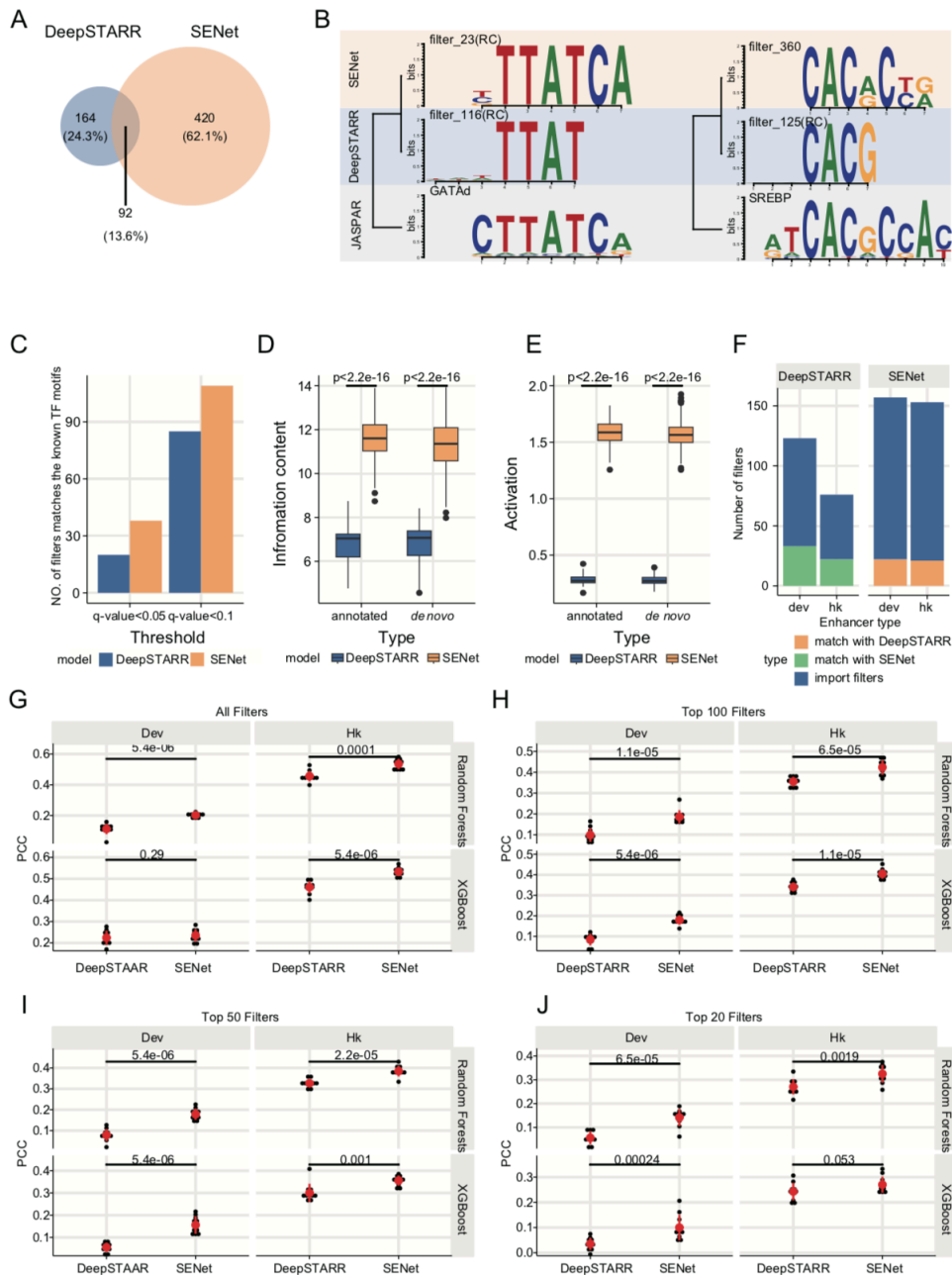


802

803 (A) Venn diagram showing the number of shared DNA motifs captured by the first
804 convolution layer filters of SENet and DeepSTARR (q-value < 0.1). (B) The first
805 convolution layer filters of SENet and DeepSTARR capturing the DNA motifs of
806 GATAd and SREBP. (C) The number of unknown TF motifs captured by filters in
807 SENet and DeepSTARR under different thresholds. (D) and (E) The information
808 content and activation value of filters in SENet and DeepSTARR, respectively. (F) The
809 number of filters with importance greater than 0.1 for predicting developmental and
810 housekeeping enhancers in SENet and DeepSTARR. (G-J) Random forest and
811 XGBoost models constructed using the DNA features captured by filters in SENet and

812    DeepSTARR for predicting the activity of developmental and housekeeping enhancers.
813    "All filters", "top 100 filters", "top 50 filters", and "top 20 filters" denote models
814    constructed using all filters and the top 100, top 50, and top 20 filters ranked by
815    importance, respectively.
816

Supplementary Figure 12. *In silico* analysis reveals positional effects of key motifs in
housekeeping enhancers, distance-dependent TF motif epistasis, and contributions of
TF motif-related features to enhancer activity
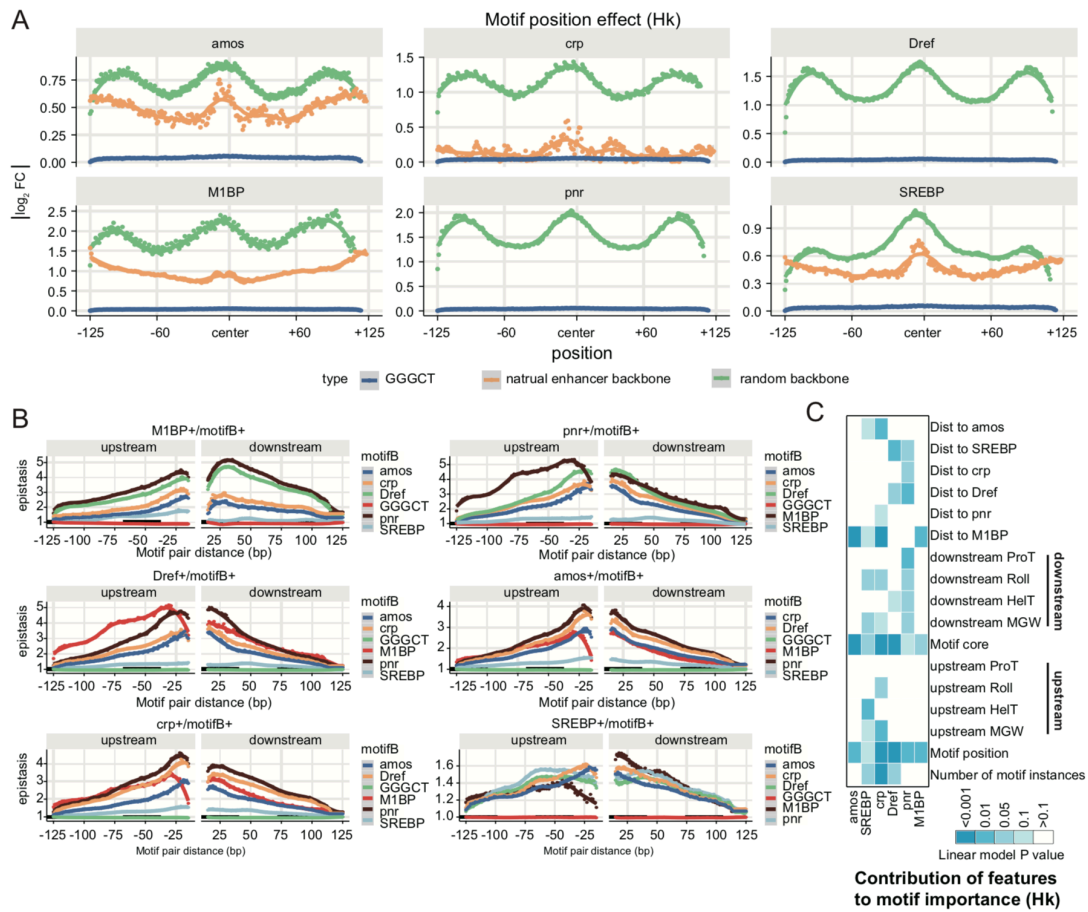


(A) Positional effects of the top six TF motifs in housekeeping enhancers. Green lines
represent the "Random Backbone Sequences" strategy, orange lines indicate the
"Natural Enhancers" strategy, and blue lines denote the negative control "GGGCT".
Panels without orange lines indicate the absence of natural enhancers containing only
a single instance of the specific TF motif, thus preventing the use of the "Natural
Enhancers" strategy (see Methods). (B) Epistasis effects between TF motifs as a
function of the relative distance between motifs. The first motif in the title is fixed at
the center of the backbone, while the second motif (motif B) is computationally moved
(color-coded). The "GGGCT" motif serves as a negative control (see Methods). Dashed
line indicates an additive effect. (C) Contributions of TF motif-related features to
housekeeping enhancer activity. For each TF motif (each row), multiple linear
regression models were constructed using the number of motif instances, the distance
from the enhancer center, the binding strength of the TF (motif core, with -log(binding
probability) as a proxy), DNA shape scores of the flanking sequences, and the relative
distances between key motifs. The P-Value of each motif feature from these models
indicate the significance of each motif feature's contribution to enhancer activity.

Supplementary Figure 13. *In silico* analysis of higher-order motif combination effects
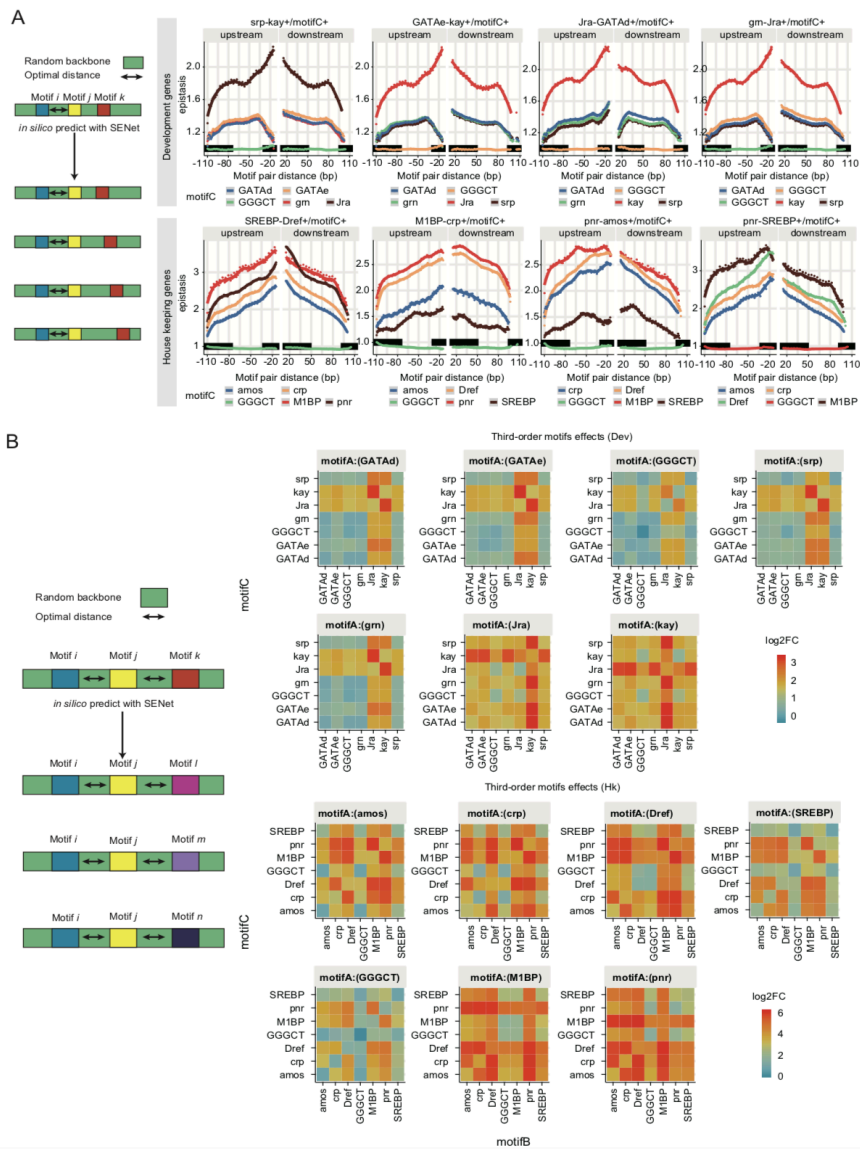
840     (A) Epistasis effects of three TF motifs as a function of the relative distance between
841     the third motif and the two fixed motifs at the backbone sequence center. Each panel
842     title indicates the fixed motifs (motifA-motifB) at the center of the backbone, while
843     motifC (shown in different colors) is computationally moved to calculate the epistasis
844     effect at distance d (see Methods). The "GGGCT" motif serves as a negative. The
845     dashed line indicates an additive effect. (B) Combination effects of three TF motifs in
846     developmental and housekeeping enhancers. MotifA, motifB, and motifC are fixed at
847     the center of the backbone, maintaining optimal relative distances. The combination
848     effect is defined as the fold change in predicted enhancer activity with motifs embedded
849     compared to the backbone's enhancer activity (see Methods).
850

851     Supplementary Figure 14. Impact of initial population size on DREAM-optimized
852     enhancer activity
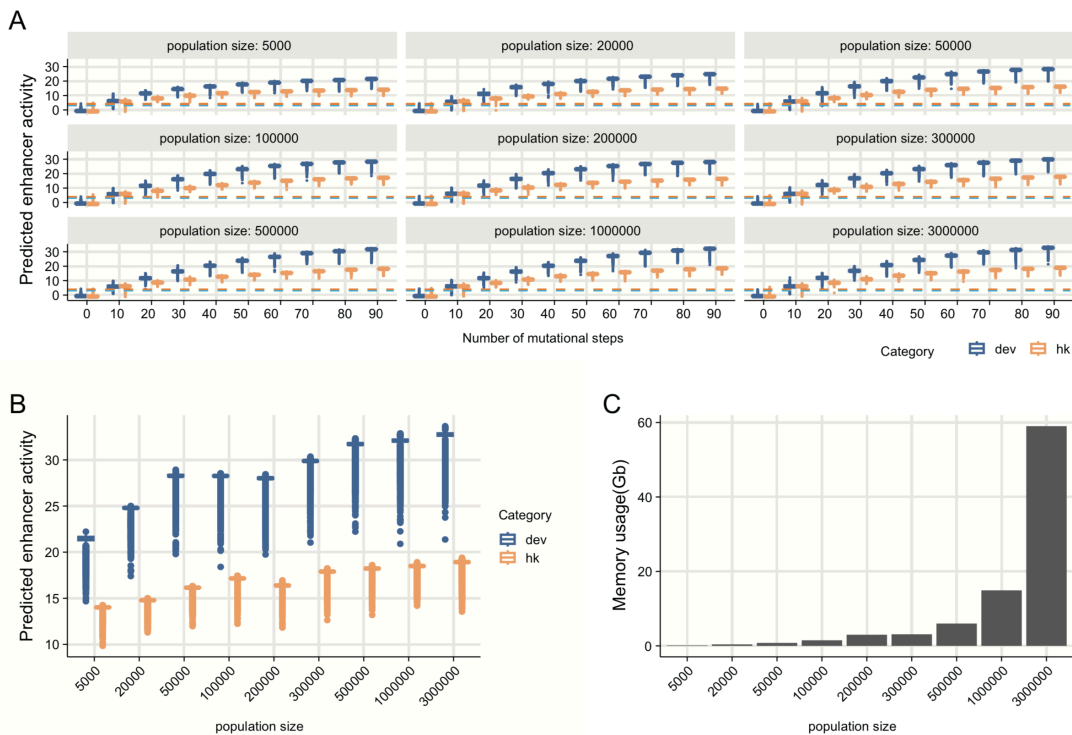


853
854     (A) Distribution of predicted enhancer activity (y-axis) for developmental and
855     housekeeping enhancers across mutation steps (x-axis) under varying initial population
856     sizes. The blue and yellow dashed lines represent the strongest activity measured in
857     Drosophila S2 cells for developmental and housekeeping enhancers, respectively. (B)
858     Final optimized enhancer activity distribution under different initial population sizes.
859     (C) Computational memory requirements for DREAM at different initial population
860     sizes. Boxes indicate the 25th, 50th, and 75th percentile values, with whiskers
861     extending to the most extreme data points within 1.5 times the interquartile range from
862     the edges of the boxes.
863

864    Supplementary Figure 15. Dynamic changes in sequence properties of designed
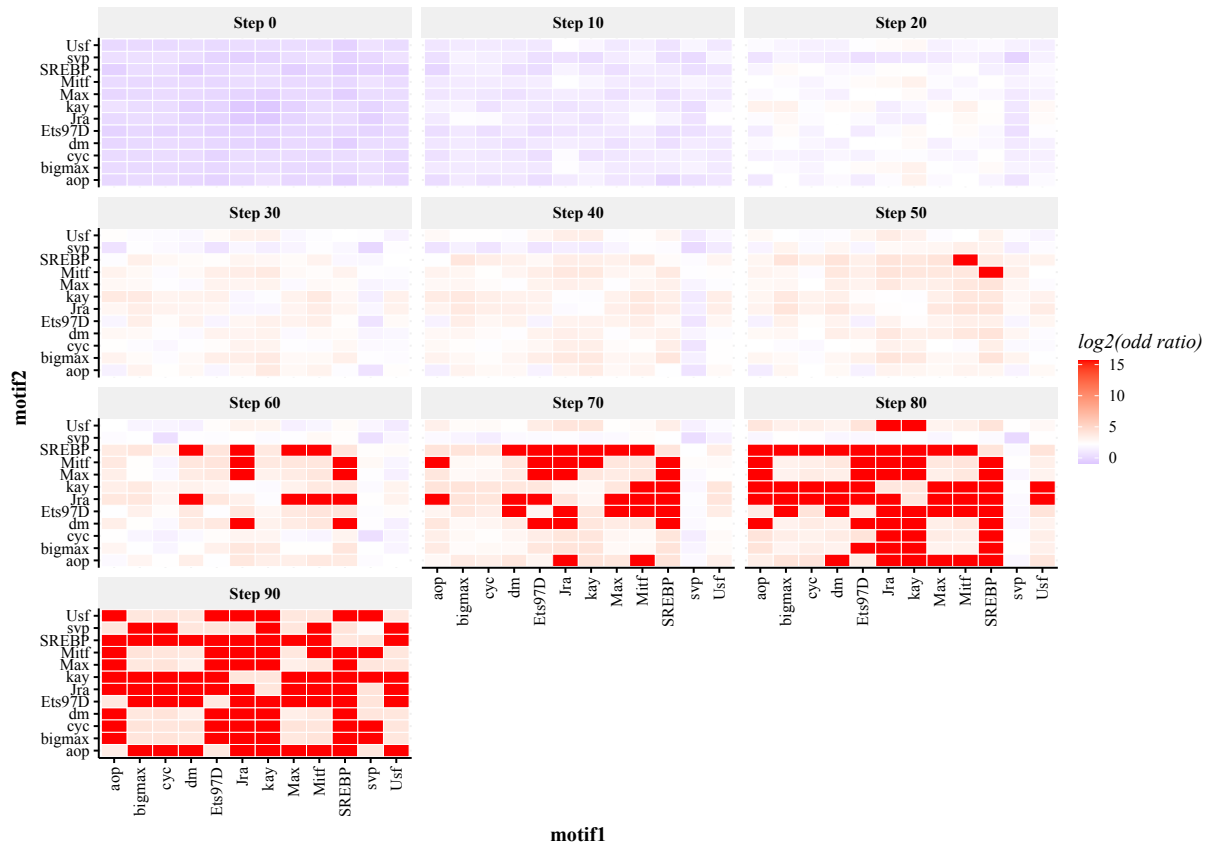865    enhancers during the *in silico* optimization trajectory
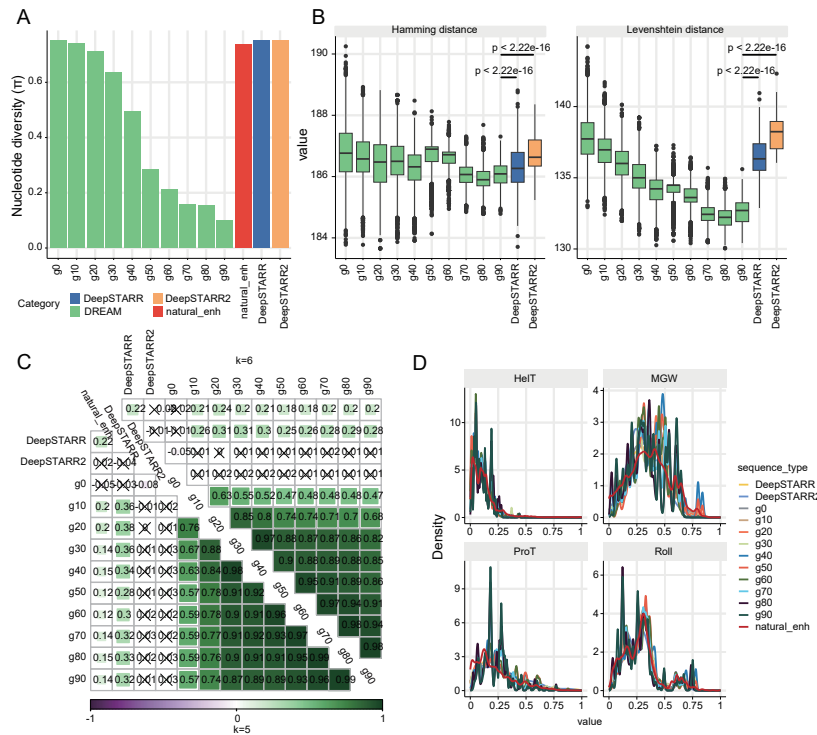


866
867    (A) The dynamic changes in nucleotide diversity($\pi$) within the population during the
868    sequence optimization process, with natural enhancers in the Drosophila genome, and
869    enhancers designed by DeepSTARR and DeepSTARR2 serving as references. The
870    letter "g" denotes "Generation". (B) The dynamic changes in sequence distance to
871    natural enhancers during the sequence optimization process, measured using Hamming
872    distance and Levenshtein distance. The sequence similarities of designs by
873    DeepSTARR and DeepSTARR2 with natural enhancers are provided as references. (C)
874    Heatmaps illustrating the Pearson correlation coefficient (PCC) of 5-mer and 6-mer
875    frequencies between sequences designed by DREAM, DeepSTARR, and
876    DeepSTARR2, and natural enhancers. Pairs with a P-Value > 0.05 are marked with an
877    "×". (D) Distributions of the predicted scores for the four DNA shapes (minor groove
878    width (MGW), roll (Roll), propeller twist (ProT), and helix twist (HelT)) for the 10 bp
879    flanking regions of key TF motifs during the sequence optimization process by
880    DREAM. The Corresponding predicted scores in natural enhancers, DeepSTARR, and
881    DeepSTARR2 are provided as references.
882

Supplementary Figure 16. The dynamic co-occurrence pattern of known TFs captured
884    by the Top10 filters during the *in silico* optimization trajectory.
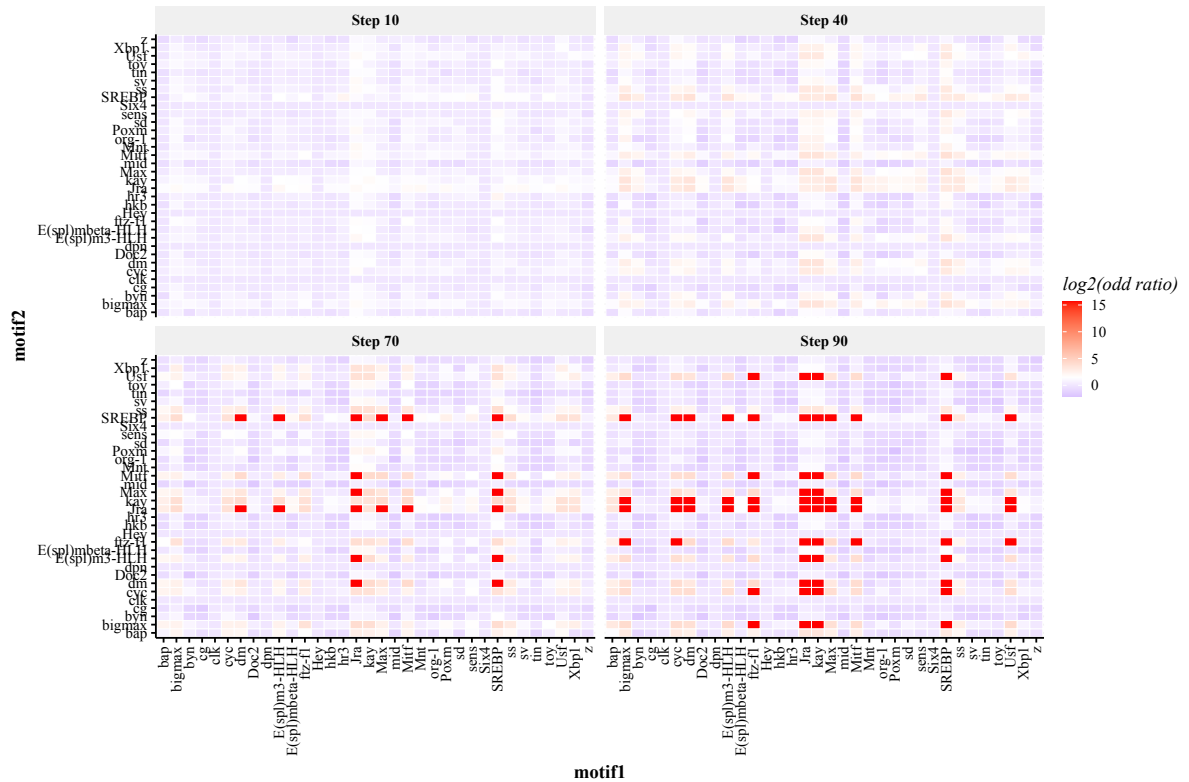


885
886    The heatmap depicts the dynamic enrichment of known motif pairs captured by the Top
887    10 filters throughout the *in silico* optimization trajectory. For simplicity, samples from
888    the sequence population were taken at the 10th, 40th, 70th, and 90th evolutionary steps.
889    The odds ratio was calculated using a one-sided Fisher's exact test, with the *in silico*
890    initial sequence population serving as the background.
891

Supplementary Figure 17. The dynamic co-occurrence pattern of known TFs enriched
the final optimized enhancers during the *in silico* optimization trajectory.



The heatmap illustrates the dynamic enrichment of known motif pairs in the final
optimized sequences during the *in silico* optimization trajectory. For simplicity, only
representative motifs with a Fisher's odds ratio greater than 4 are displayed. The odds
ratio was calculated using a one-sided Fisher's exact test, with natural enhancers in the
*Drosophila* genome as the background.

Supplementary Figure 18. By customizing the fitness function, the DREAM framework can design silencers or simultaneously optimize the sequence properties of enhancers
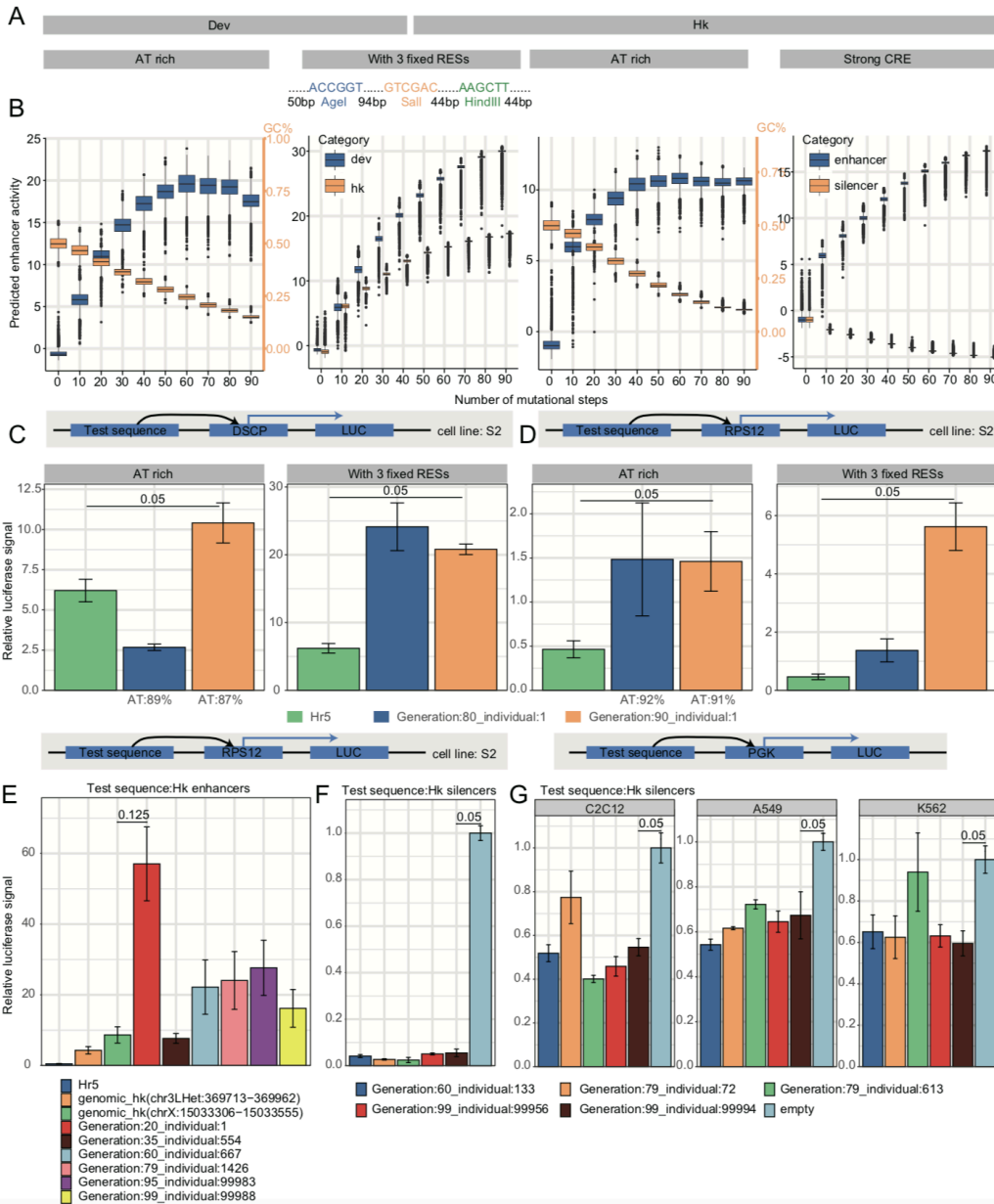
(A) Illustration of CREs design tasks: 1. "AT rich + strong activity" developmental and housekeeping enhancers, 2. developmental and housekeeping enhancers with user-specified restriction enzyme sites ("with 3 fixed RESs"), and 3. strong housekeeping silencers and enhancers (see Methods for details). (B) Distribution of predicted activity for developmental and housekeeping CREs during the *in silico* optimization process (y-axis). The x-axis represents mutation steps. In the "AT rich + strong activity" panel, the second y-axis indicates the GC content of sequences. Blue and orange boxes represent enhancer activity and GC content, respectively. The DREAM framework was used to simultaneously increase both the AT content and the activity of developmental and housekeeping enhancers. In the "with 3 fixed RESs" panel, blue and orange boxes represent developmental and housekeeping enhancers, respectively. In the "strong housekeeping silencers/enhancers" panel, blue and orange boxes represent

916     housekeeping enhancers and housekeeping silencers, respectively. In all panels, boxes
917     denote the 25th, 50th, and 75th percentile values, with whiskers indicating the
918     outermost points within 1.5 times the interquartile range from the box edges. (C) and
919     (D) Validation of synthetic CREs designed for the "AT rich + strong activity" and "with
920     3 fixed RESs" tasks in *Drosophila* S2 cells using luciferase reporter assays, with the
921     Hr5 enhancer as the control. (E) and (F) Validation of housekeeping enhancers (E) and
922     housekeeping silencers (F) in *Drosophila* S2 cells using luciferase reporter assays. In
923     validating synthetic housekeeping enhancers, the two strongest housekeeping
924     enhancers (chr3LHet:369713-369962 and chrX:15033306-15033555) from the
925     *Drosophila* genome and the Hr5 enhancer were used as controls. For synthetic
926     housekeeping silencers, the empty vector was used as a control. (G) Validation of
927     housekeeping silencers in C2C12 (mouse), A549 (human), and K562 (human) cell lines
928     using luciferase reporter assays (using PGK promoter), with the empty vector as the
929     control. All luciferase values are normalized to Renilla luciferase activity. Error bars
930     represent the standard error of the mean (n = 3 biological replicates; one-sided
931     Wilcoxon rank-sum test).
932

933     Supplementary Figure 19. The TF motifs in optimized housekeeping silencers using
934     DREAM.
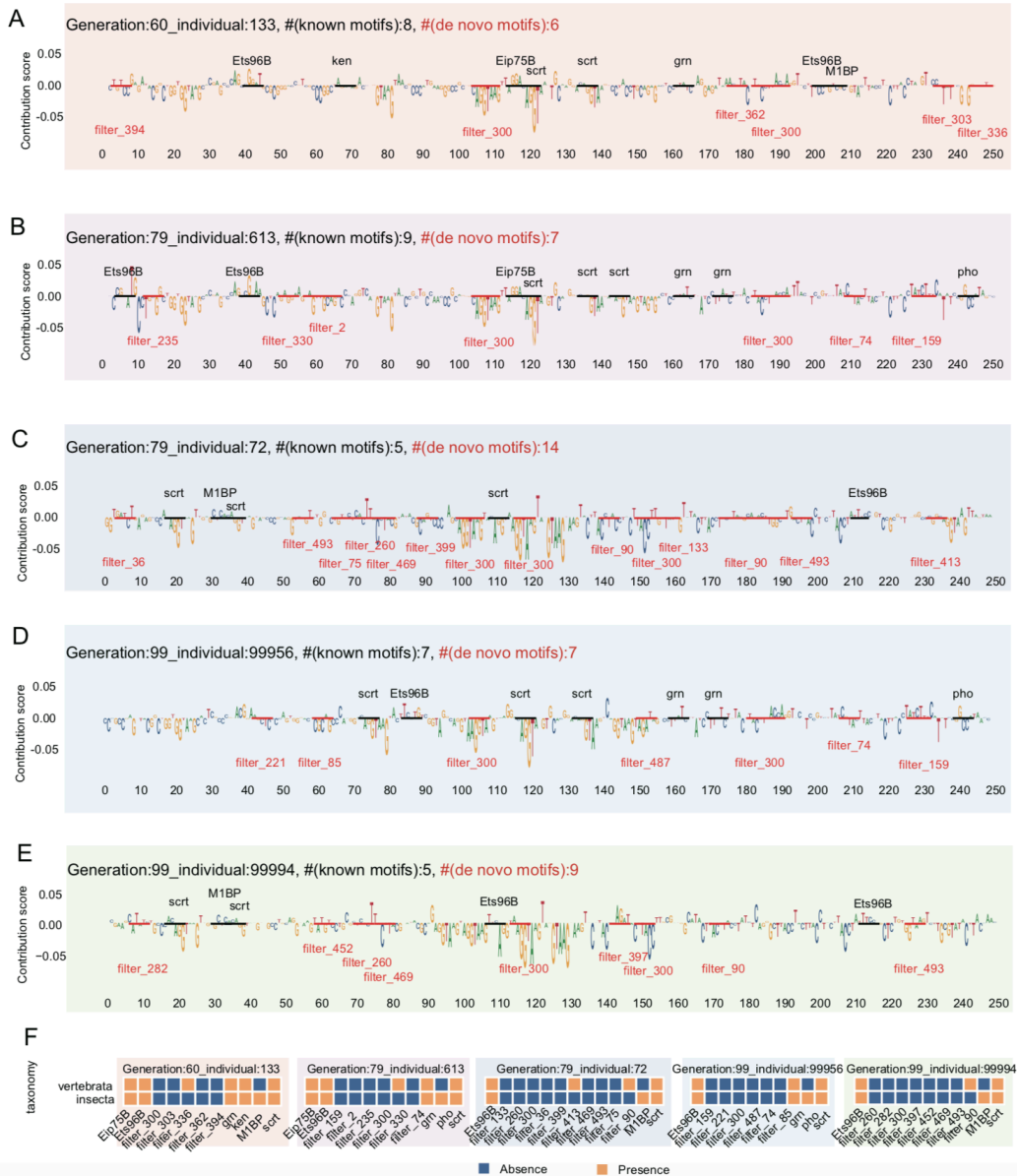


935
936     (A)-(E) Nucleotide contribution scores for the optimized housekeeping silencers
937     derived from the SENet using DeepExplainer. Instances of motifs identified by
938     DREAM are emphasized, with known motifs indicated in black and *de novo* motifs
939     marked in red. The number of known motifs (#(known motifs)) and the number of de
940     novo motifs (#(*de novo* motifs)) are also marked. (F) The colored matrices illustrate the
941     presence or absence of TF motifs (x-axis) in the corresponding taxonomy (y-axis), with
942     blue indicating absence and orange indicating presence of the TF in the respective
943     taxonomy.
944

Supplementary Figure 20. Designing medium strength enhancers with DREAM.
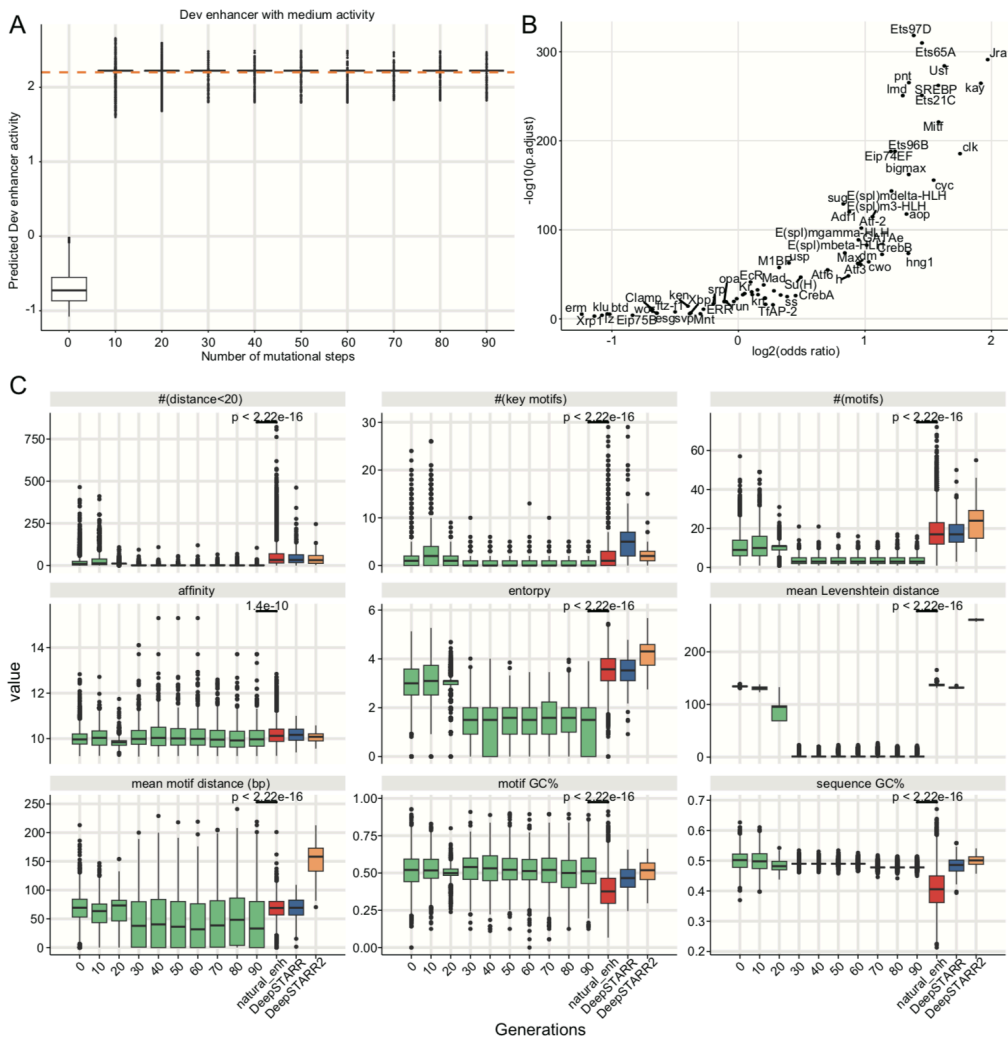


946
947 (A) Distribution of predicted enhancer activity (y-axis) for developmental enhancers at
948 various mutational steps (x-axis) during the *in silico* optimization trajectory favoring
949 medium activity. The boxes denote the 25th, 50th, and 75th percentile values, while the
950 whiskers extend to the furthest point within 1.5 times the interquartile range from the
951 boxes' edges. The yellow dashed line indicates the mean activity of Drosophila
952 developmental enhancers measured in S2 cells (2.19). (B) TF motifs over-represented
953 in designed enhancers with medium strength compared to genomic noncoding regions
954 of non-enhancers (background). The x-axis represents the log2-transformed one-sided
955 Fisher's exact test odds ratio, and the y-axis represents the corresponding FDR-
956 corrected significance. (C) This panel illustrates various characteristics, including the
957 count of TF motif pairs with distances less than 20 bp (#(distance<20)), the average
958 number of (key) TF motifs (#(motifs)), TF binding affinity quantified by -log(binding
959 probability), diversity of TF motifs evaluated through entropy, sequence diversity
960 quantified by Levenshtein distance, the mean distance between TF motifs, and GC
961 content of binding TF motifs (motif GC%) and sequences (sequence GC%) throughout
962 the in silico optimization trajectory. The x-axis outlines the steps of the *in silico*
963 optimization process, and the results are compared with natural enhancers in the

964    *Drosophila* genome as well as synthetic enhancers designed by DeepSTARR and
965    DeepSTARR2 (one-sided Wilcoxon rank-sum test).
966

967    Supplementary Figure 21. DREAM-optimized enhancer displayed stronger ability to
968    stimulate the expression of CAG promoter and CMV promoter.
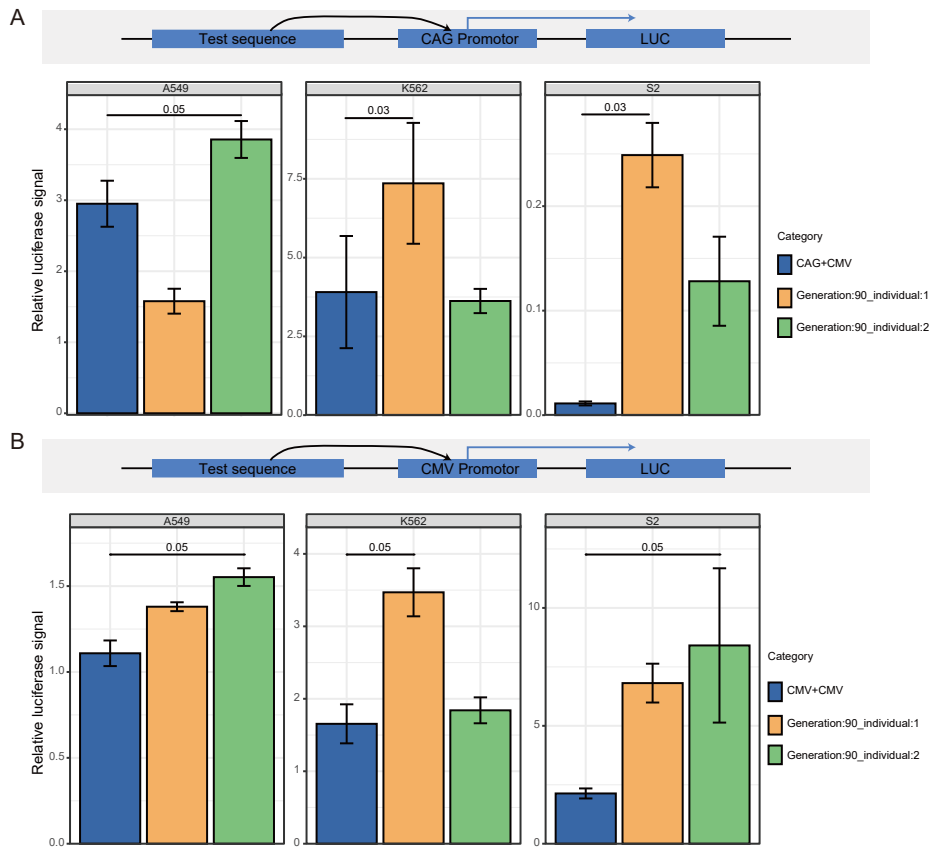


969

970    Quantitative comparison of luciferase assay activity levels for the synthetic enhancers
971    + CAG promoter (A) and synthetic enhancers + CMV promoter (B) in A549, K562 and
972    S2 cell lines. The luciferase values are normalized with the signal of Renilla luciferase.
973    Error bars: Standard error of the mean (n = 3 biological replicates; one-sided Wilcoxon
974    rank-sum test).

975

Supplementary Figure 22. Validation of the designed enhancer activity in the
endogenous chromatin context.
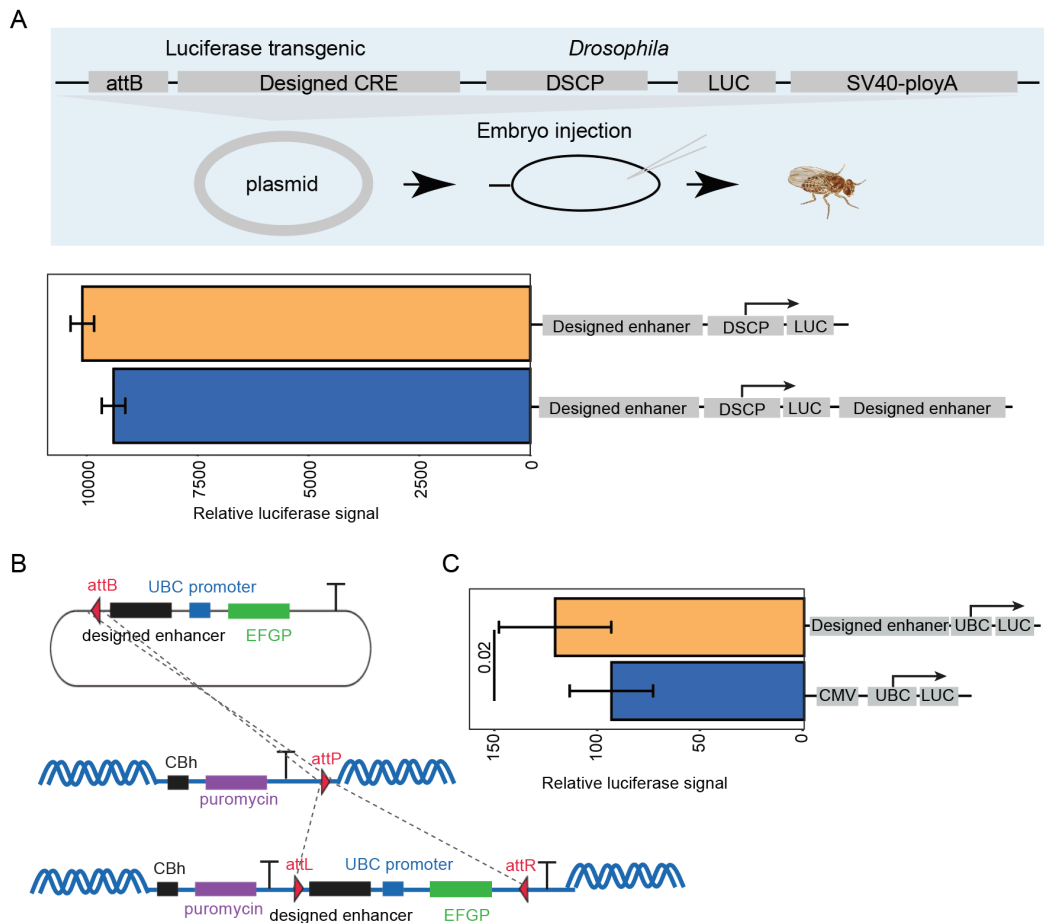


(A) Using the attB/attP site-specific recombination system, luciferase transgenic *Drosophila* lines were constructed to validate the activity of the designed enhancer (Generation:90_individual:2). The designed enhancer was inserted both upstream and downstream of the DSCP promoter. (B) Schematic illustration of the integration of the designed enhancer into the human genome (293T cell line) using the attB/attP site-specific recombination system. (C) Validation of the activity of the CMV enhancer and the designed enhancer (Generation:79_individual:2) within the endogenous chromatin context of human 293T cells, achieved through recombinase-mediated integration. All luciferase values were normalized against the activity of an empty vector containing only the corresponding promoter. Error bars represent the standard error of the mean (n = 3 biological replicates; one-sided T-test).

991

## Reference

993     1.     de Almeida, B.P., Reiter, F., Pagani, M. and Stark, A. (2022)
994     DeepSTARR predicts enhancer activity from DNA sequence and enables the
995     de novo design of synthetic enhancers. *Nat. Genet.*, **54**, 613-624.
996     2.     Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum
997     Likelihood. *Mol. Biol. Evol.*, **24**, 1586-1591.
998     3.     Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B.,
999     Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N. *et al.*
1000     (2007) Evolution of genes and genomes on the Drosophila phylogeny. *Nature*,
1001     **450**, 203-218.

1002