1 **Pleiotropic enhancers are ubiquitous regulatory elements in the human genome**

2

3 Authors: Ian Laiker[1], Nicolas Frankel[1,2*]

4

5 [1]Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Consejo Nacional de

6 Investigaciones Científicas y Técnicas (CONICET) y Universidad de Buenos Aires (UBA), Buenos Aires

7 1428, Argentina.

8 [2]Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales (FCEN),

9 Universidad de Buenos Aires (UBA), Buenos Aires 1428, Argentina.

10

11 *author for correspondence: nfrankel@conicet.gov.ar

1

Abstract

Enhancers are regulatory elements of genomes that determine spatio-temporal patterns of gene expression. The human genome contains a vast number of enhancers, which largely outnumber protein-coding genes. Classically, enhancers have been regarded as highly tissue-specific. However, recent evidence suggests that many enhancers are pleiotropic, with activity in multiple developmental contexts. Yet, the extent and impact of pleiotropy remain largely unexplored. In this study we predicted active enhancers across human organs based on the analysis of both eRNA transcription (FANTOM5 consortium datasets) and chromatin architecture (ENCODE consortium datasets). We show that pleiotropic enhancers are pervasive in the human genome and that most enhancers active in a particular organ are also active in other organs. In addition, our analysis suggests that the proportion of context-specific enhancers of a given organ is explained, at least in part, by the proportion of context-specific genes in that same organ. The notion that such a high proportion of human enhancers can be pleiotropic suggests that small regions of regulatory DNA contain abundant regulatory information and that these regions evolve under important evolutionary constraints.

Keywords: human genome, non-coding DNA, gene regulation, CRE, enhancer, pleiotropy

Significance statement

The human genome contains a vast number of regulatory elements, named enhancers, that control the tempo and mode of gene expression. Classically, enhancers have been regarded as genetic elements that are active in a single organ, but recent evidence suggests that many enhancers in animal genomes are active in multiple organs (i.e., are pleiotropic). Here we shed light on the architecture of non-coding human DNA by showing that a large percentage of the enhancers of the human genome are pleiotropic and that the majority of enhancers active in a particular organ are pleiotropic. This suggests that small regions of regulatory DNA may contain abundant regulatory information and that these regions evolve under important evolutionary constraints.

Introduction

Enhancers are regulatory elements of the genome, in general smaller than 1 kb, that determine spatio-temporal patterns of gene expression (Levine 2010). Enhancers are activated and repressed by the binding of transcription factors (TFs), which recognize specific motifs in the DNA named transcription factor binding sites (TFBSs)(Hill et al. 2021). Upon activation, enhancers interact with the basal transcriptional machinery at the core promoter, boosting mRNA synthesis of the target gene(s)(Panigrahi & O'Malley 2021). There are several features that distinguish active enhancers: (i) it has been shown that nucleosomes flanking active enhancers usually contain an acetyl group in lysine 27 of histone H3 (H3K27ac) (Creyghton et al. 2010) and a single methyl group in lysine 4 of histone H3 as well (H3K4me1)(Rada-Iglesias et al. 2011), which often coincides with the absence of the trimethylated form of H3 lysine 4 (H3K4me3)(Lidschreiber et al. 2021), (ii) it has been demonstrated that enhancer DNA may be transcribed bidirectionally, generating short capped non-polyadenylated RNAs ('eRNAs')(Sartorelli & Lauberth 2020), and (iii) it is typically observed that active enhancers are depleted of positioned nucleosomes, thus having a conformation that is known as 'open chromatin' (Klemm et al. 2019). Together, these features can be used to identify enhancers genome-wide (Kimura 2013; Kristjánsdóttir et al. 2020; Minnoye et al. 2021).

The human genome contains a vast number of enhancers, which largely outnumber protein-coding genes (Pennacchio et al. 2013; Heintzman et al. 2009). It has been shown that a large percentage of SNPs associated with human diseases fall within regions that are predicted to be enhancers (Corradin & Scacheri 2014), and, in the same vein, numerous cases in which mutations in enhancers cause disease have been reported (Smith & Shilatifard 2014). Although thousands of non-coding regions have been identified as putative enhancers (Gasperini et al. 2020). Indeed, the complexity of enhancer function in the human genome is just beginning to be comprehended. For example, recent studies uncovered that enhancers usually regulate multiple genes and that many enhancers bypass their nearest core-promoters to interact with distant core-promoters (Reilly et al. 2021; Nasser et al. 2021).

89   The fact that regulatory information within enhancer DNA can be dense and pleiotropic (Fuqua et al.

90   2020; Kvon et al. 2020; Xin et al. 2020; Le Poul et al. 2020) also illustrates regulatory complexity.

91   The classic conception in which enhancers have strict tissue-specific activities (i.e., modular

92   or context-specific activities) has been called into question by recent experimental data (Fuqua et al.

93   2020; Infante et al. 2015; Lewis et al. 2019; Lonfat et al. 2014; Preger-Ben Noon et al. 2018; Xin et al.

94   2020). Current evidence suggests that a large proportion of enhancers in vertebrate genomes are

95   active in multiple developmental contexts (Kittelmann et al. 2021; Sabarís et al. 2019), which

96   probably means that a significant number of enhancers in the human genome are pleiotropic (i.e.,

97   have roles in multiple organs and/or developmental stages). Although evidence suggest that

98   pleiotropy in regulatory DNA is pervasive, the extent of enhancer pleiotropy in the human genome

99   remains largely unexplored. In this sense, it would be useful to have a reliable estimate of the

100   percentage of pleiotropic enhancers in the human genome and to explore the landscape of

101   pleiotropy in each organ.

102   To assess the magnitude of pleiotropy in the regulatory genome, in this study we estimated

103   the extent of enhancer reuse and analyzed the abundance of pleiotropic enhancers per organ. We

104   show that pleiotropic enhancers are pervasive in the human genome and that the majority of

105   enhancers active in a particular organ are also active in other organs. In addition, our analysis

106   suggests that the proportion of context-specific enhancers of a given organ is explained, at least in

107   part, by the proportion of context-specific genes in that same organ.

108

109   **Results**

110

111   *A large proportion of predicted human enhancers are pleiotropic*

112

113   We first set out to estimate the number of pleiotropic enhancers in the human genome by using

114   information from the FANTOM5 (Andersson et al. 2014) and ENCODE (Dunham et al. 2012) projects.

115   We grouped ENCODE and FANTOM biosamples into organs using UBERON anatomy ontology

116   (Mungall et al. 2012) (see tables S1 and S2). For FANTOM5 data, putative enhancers were defined as

117   DNA regions transcribed bidirectionally above a certain threshold in at least one of 41 organs (see

118   table S1 for a list of samples and organs). For ENCODE data, we defined putative enhancers as DNA

119   regions with open chromatin that contain the epigenetic mark H3K27ac in at least one of 20 organs

120   (see table S2 for a list of samples and organs). Although ENCODE and FANTOM5 predictions are

121   derived from different biosamples and disparate genomic analyses, 40-60% of the FANTOM5

122   enhancers are contained within the larger ENCODE enhancer set in 16 common organs (testis is an

123   exception, see Figure S1). Also, we compared ENCODE enhancers with the 'ABC enhancer set', which

124   was defined based on chromatin accessibility, the presence of the epigenetic mark H3K27ac and the

125   interaction of the DNA region with core promoters (Fulco et al. 2019). We determined that, for 14

126   common organs, the overlap between our enhancer set predicted with ENCODE data and ABC

127   enhancers ranges between 43% and 83% (Figure S1). Altogether, comparisons between different

128   enhancer sets suggest that most of our enhancer predictions based on ENCODE data are robust.

129   By intersecting genomic information from different organs we aimed to estimate the

130   proportion of pleiotropic enhancers in the human genome. Hence, we considered that a pleiotropic

131   FANTOM5 enhancer is a DNA region transcribed bidirectionally in two or more organs. For ENCODE

132   data, we delimited consensus elements considering the different organs in which an enhancer was

133   predicted to be active, based on the presence of open chromatin and the H3K27ac mark (see details

134   of the pipeline in figure S2 and Materials and Methods, and figure S3 for the size distribution of

135   consensus elements). In this case, we define a pleiotropic enhancer as a consensus element active in

136   two or more organs. To evaluate how the proportion of pleiotropic enhancers changes with the

137   number of organs considered we took combinations of organs from 2 to n out of the pool of ENCODE

138   organs (n=20) and FANTOM5 organs (n=41) and calculated the proportion of enhancers active in 2

139   or more organs. This approach allows us to comprehend how the proportion of pleiotropic

140   enhancers changes depending on the number of organs considered and, at the same time, to

141   validate our methods. We observed that the median of predicted pleiotropic enhancers increases

142   monotonically with the number of intersected organs, both for FANTOM5 and ENCODE data (Figure

143   1). Although enhancer sets predicted with ENCODE and FANTOM5 data are different (~50% of the

144   FANTOM5 enhancer set is contained within the larger ENCODE enhancer set and, as well, the
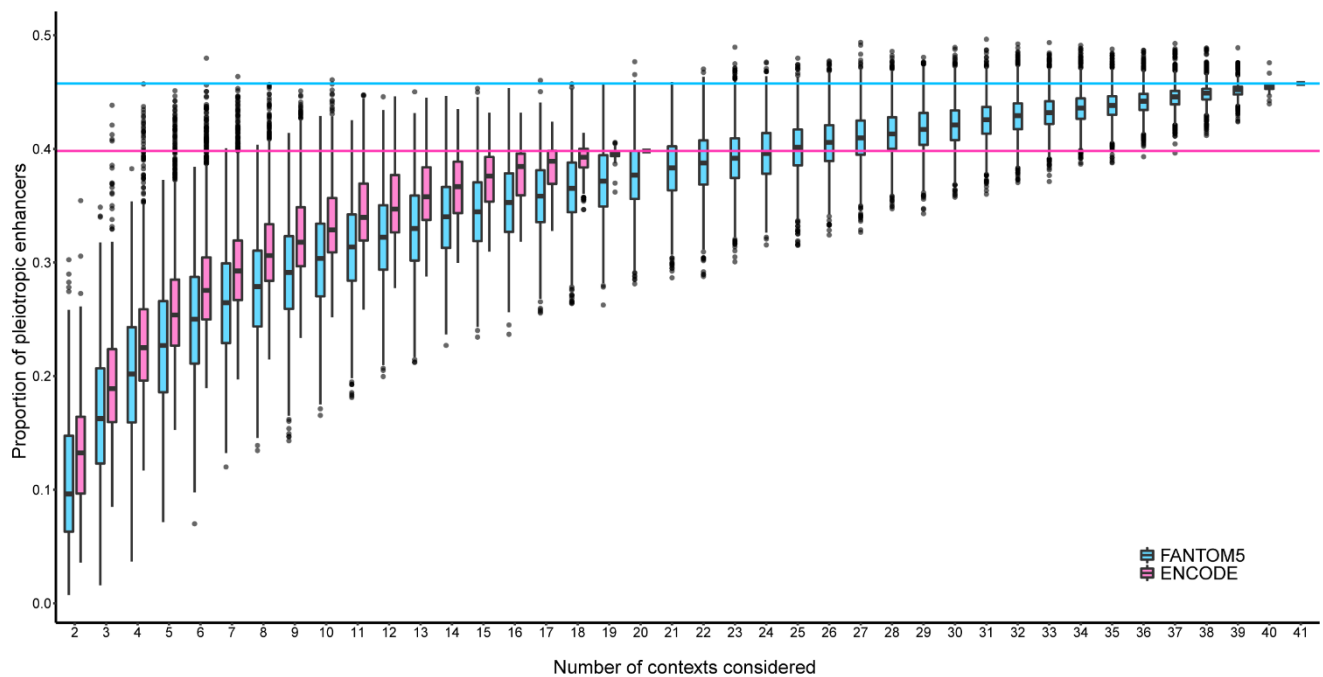
Figure 1. A large proportion of predicted human enhancers are pleiotropic. Proportion of pleiotropic enhancers as a function of the number of contexts considered in FANTOM5 (cyan) and ENCODE (magenta) enhancer sets. Boxplots indicate the median (horizontal line), interquartile range (box), observations within ± 1.5xinterquartile range (whiskers) and outliers (dots). The proportion of pleiotropic enhancers when considering all organs in a single comparison are marked with a magenta line (ENCODE) and a cyan line (FANTOM5).

FANTOM5 enhancer set represents enhancers active in more organs than the ENCODE set) both the shape and the asymptote of the two hypothetical curves are remarkably similar (Figure 1). Furthermore, the implication that more than 40% of the predicted enhancers in the human genome are pleiotropic (active in at least two organs) is noteworthy as well (Figure 1).

Given that organs are comprised of different cell types and that some cell types are present in multiple organs (e.g., fibroblasts and endothelial cells), the high proportion of pleiotropic enhancers calculated with bulk organ data could be a mere consequence of considering enhancers that are active in cell types common to many organs. To rule out this possibility, we analyzed enhancers identified in primary cell cultures by the FANTOM5 project (Table S1) (Andersson et al. 2014), combining the same cell types from different organs as a single context (see Table S1). Notably, when considering these 39 distinct cell types, the percentage of pleiotropic enhancers turns out to be 68%. We performed the same analysis (Table S2) using distal enhancers (cCRE-dELS) defined by the ENCODE consortium in 17 cell types (Moore et al. 2020). When considering all 17 cell types, the

7

165 percentage of pleiotropic enhancers is 39%. In order to determine whether results from cells and

166 organs are comparable, we contrasted the degree of pleiotropy of enhancers obtained with

167 FANTOM5 cell types data with that calculated using FANTOM5 organ data (Figure S4). We observed

168 that the degree of pleiotropy of enhancers estimated with cell types data and organ data is

169 remarkably similar (Figure S4). Altogether, these data suggest that enhancer pleiotropy is mostly due

170 to enhancers being active in different cell types of organs, rather than enhancers being active in the

171 same cell type present in multiple organs.

172

173   *Most enhancers active in a particular organ are pleiotropic*

174

175 Next, we sought to explore the abundance of pleiotropic enhancers in the different organs. To this

176 end, we calculated the proportion of predicted pleiotropic enhancers per organ in the FANTOM5 and

177 ENCODE data sets (Figure 2). We observed values ranging from 43.6% to 98.1% in FANTOM5 organs

178 (Figure 2A, table S3) and from 53.4% to 96.8% in ENCODE organs (Figure 2B, table S3). These data

179 suggest that most enhancers active in a given organ are also active in one or more other organs.

180 Indeed, most of the pleiotropic enhancers active in a given organ are active in more than two organs

181 (Figure 2).

182

183

184

185

186

187

188

189

190

191

192
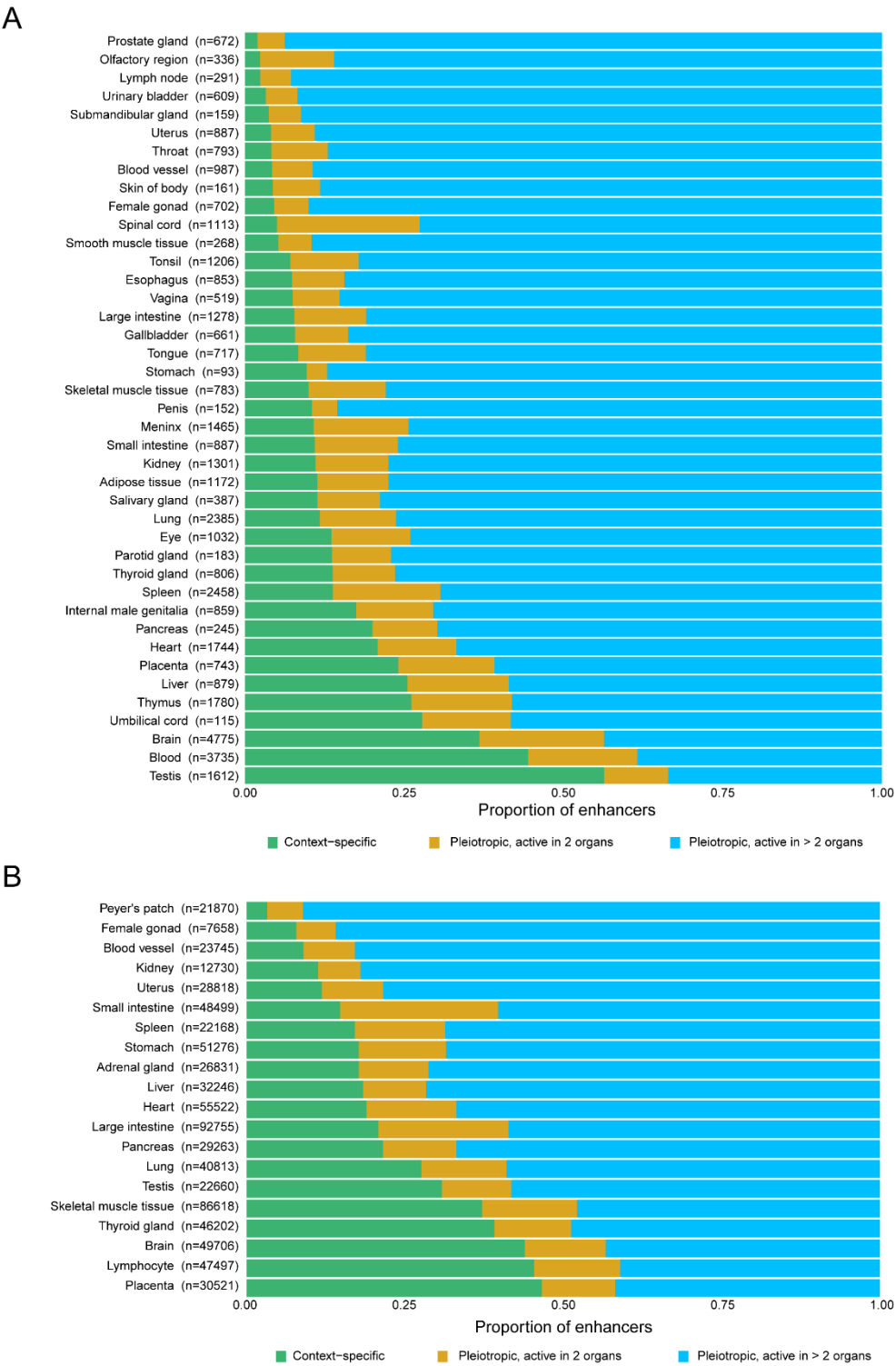
Figure 2. The majority of active enhancers in a particular organ are pleiotropic. Proportion of context-specific enhancers (green), pleiotropic enhancers active in 2 organs (orange) and pleiotropic enhancers active in more than 2 organs (blue) in FANTOM5 (A) and ENCODE (B) organs. The number of predicted enhancers per organ is indicated beside the name of each organ.

201   Given that 17 organs are shared between FANTOM5 and ENCODE datasets, we compared the

202   estimations of the proportion of predicted pleiotropic enhancers for both analyses. We observed that

203   they are notably concordant, especially considering that biosamples are derived from different parts

204   of the organ and different individuals, and that enhancers were predicted through distinct methods

205   (see Table S3). For most comparisons the difference in the estimation of the proportion of pleiotropic

206   enhancers is smaller than 10%. However, testis is an outlier, with a difference of more than 20%.

207   Despite differences in the estimation of the proportion of pleiotropic enhancers for testis, both

208   FANTOM5 and ENCODE data indicate that this organ is rich in context-specific enhancers (Figure 2).

209   Since it has been reported that testis is the organ with the most context-specific genes (Djureinovic

210   et al. 2014; Pineau et al. 2019; Zhu et al. 2016), we reasoned that a possible cause for the high number

211   of context-specific enhancers in the testis could be the high number of testis-specific genes and,

212   moreover, that there could be a relationship between the number of context-specific genes and the

213   abundance of context-specific enhancers per organ. To test this hypothesis, we computed the

214   number of organ-specific genes in ENCODE and FANTOM5 organs (with RNA-seq and transcribed

215   TSSs data respectively, see Materials and Methods for details) and, subsequently, we looked for a

216   possible correlation between the proportion of organ-specific genes and the proportion of organ-

217   specific enhancers. Remarkably, for both FANTOM5 and ENCODE organs, we found a positive

218   correlation (Pearson's r=0.653, p<0.001 and Pearson's r=0.647, p<0.001, respectively) between the

219   proportion of context-specific genes and the proportion of context-specific enhancers (Figure 3). A

220   positive correlation is also observed when using data from FANTOM5 cell types (Pearson's r=0.71,

221   p<0.001).

222

223

224
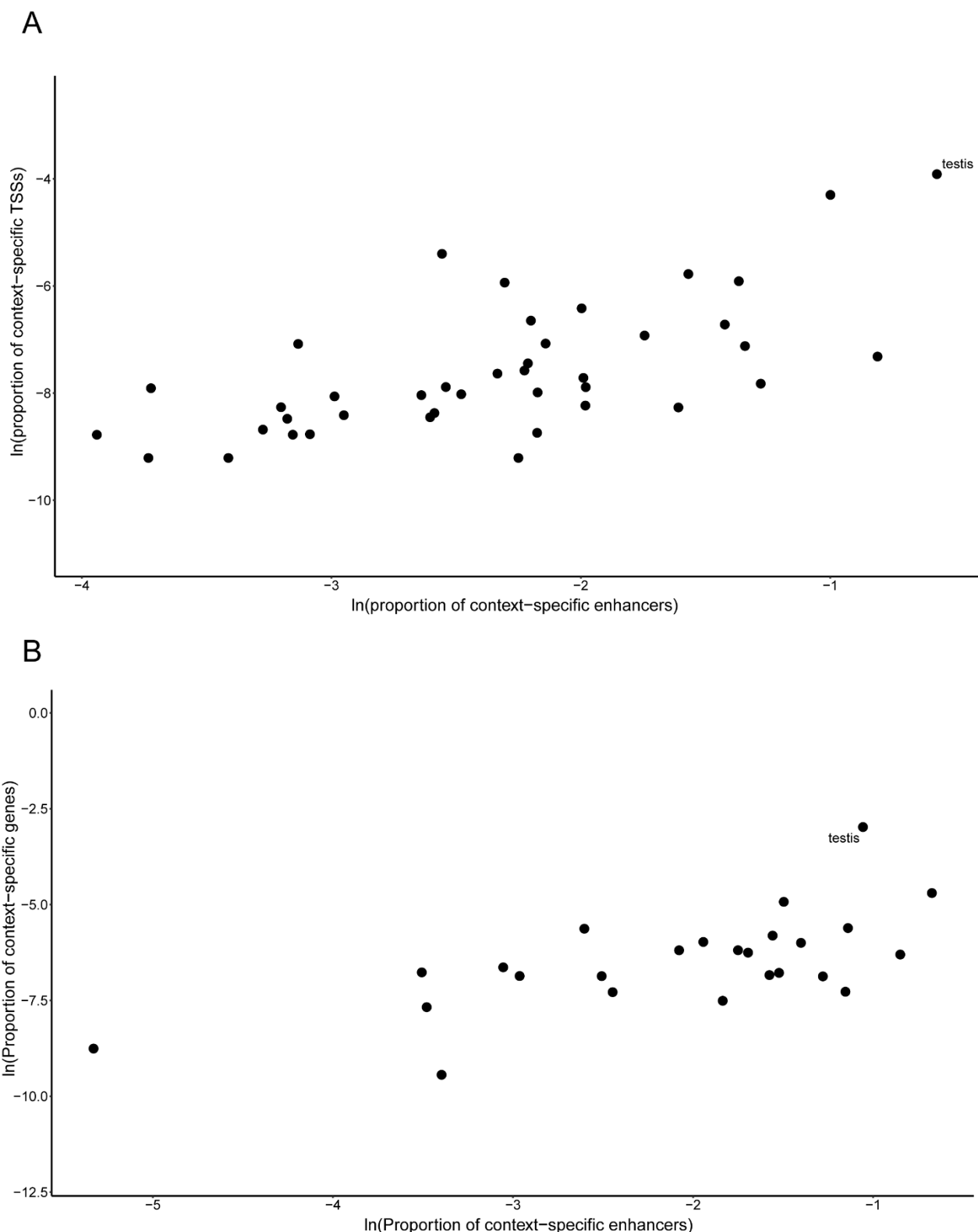
Figure 3. The proportion of context-specific enhancers is positively correlated with the proportion of context-specific genes. (A) Relationship between the proportion of context-specific enhancers and context-specific TSSs for the FANTOM5 dataset. Active TSSs were defined in each organ by comparing transcription levels against genomic background (see Materials and Methods for details). (B) Relationship between the proportion of context-specific enhancers and context-specific genes for the ENCODE dataset. An expression threshold was used to define expressed genes in each organ (see Materials and Methods for details).

225

226

227

228

229

230

231

232

11

233 *Pleiotropic enhancers are enriched in SNPs with high regulatory potential*

234

235 Bearing in mind the fact that pleiotropic enhancers have more density of TF binding motifs (Fish et

236 al. 2017; Singh & Yi 2021), we reasoned that a way to provide a functional validation for the

237 pleiotropic function of enhancers was to analyze a possible connection between the degree of

238 pleiotropy of enhancers and the abundance of genetic variants related to gene regulation. Thus, we

239 interrogated the Regulome database (Boyle et al. 2012), which provides a means to predict the

240 regulatory potential of SNPs across the genome. In this approach, SNPs are given a score based on

241 their likelihood of being variants with regulatory function (see Materials and Methods for details).

242 Hence, we quantified the enrichment of SNPs with predicted regulatory activity that fall within

243 enhancers in different categories of pleiotropy. We noted that there is a positive relationship

244 between the degree of pleiotropy of enhancers and the level of enrichment in SNPs with predicted

245 regulatory function (Figure 4). When using only SNPs in the most stringent category of Regulome

246 database (category 1a: SNPs that (i) are eQTLs, (ii) match TF motifs, (iii) match DNase footprints, (iv)

247 lie within DNase peaks and (v) lie within TF binding peaks) the positive relationship holds, and the

248 enrichment becomes even larger (Figure S5).
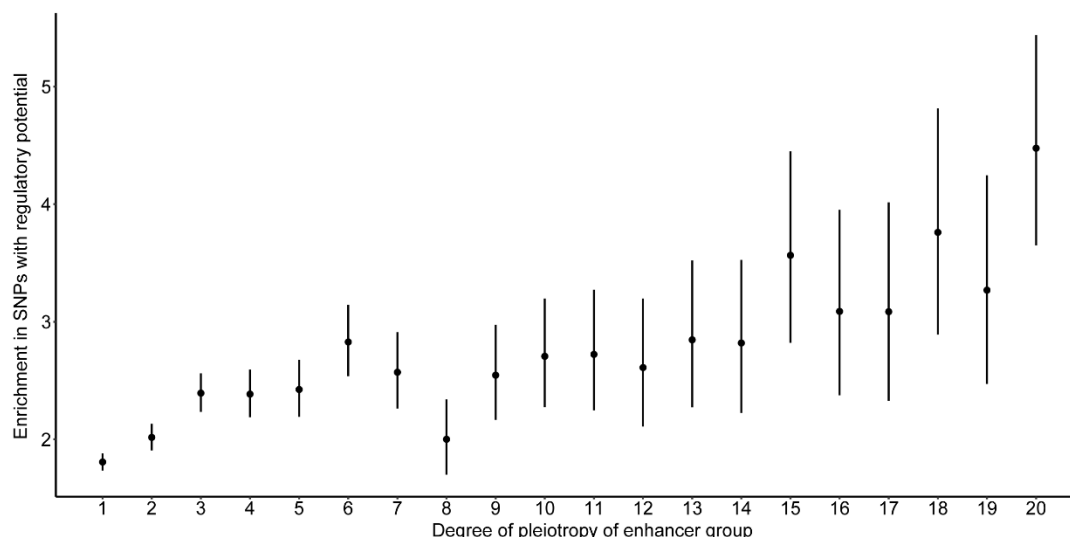
249

250 **Figure 4. Pleiotropic enhancers are enriched in SNPs with high regulatory potential.** The degree of pleiotropy of

251 ENCODE enhancers is positively correlated with the enrichment in SNPs with high regulatory potential (category 1 SNPs

252 of Regulome DB). Black circles represent the odds ratio and bars represent the 95% CI. Enrichments (odds-ratios) were

253 calculated using common SNPs as background (see Materials and Methods for details).

254

Discussion

In this study, we predicted active enhancers across human organs by analyzing different types of genomic data. Enhancer predictions were based on either the analysis of eRNA transcription (FANTOM5 enhancers) or the analysis of chromatin architecture (ENCODE enhancers). Remarkably, although these two enhancer sets are different and are derived from different data types, their make-ups are highly similar, which provides a solid basis for our conclusions. Our data suggest that more than 40% of the enhancers in the human genome are pleiotropic. We argue that 40% is a minimum estimation because we mostly considered adult organs (we included just a few fetal biosamples) and because we cannot identify enhancers that are active in a single organ but regulating more than one gene (these context-specific enhancers are pleiotropic as well). However, it is also conceivable that using bulk data from organs imposes a bias in the identification of pleiotropic enhancers. To evaluate this possibility, we calculated the percentage of pleiotropic enhancers using primary cells instead of organs. Notably, the percentage of pleiotropic enhancers calculated with cells is similar, or even larger, than the percentage obtained with organs. Hence, it is unlikely that the estimation of the percentage of pleiotropic enhancers in organs is inflated by the contribution of the same cell types present in multiple organs (e.g., fibroblasts).

A further analysis of the abundance of context-specific versus pleiotropic enhancers per organ uncovered that there is a wide variability in the proportion of context-specific enhancers per organ, and that the majority of active enhancers in a particular organ are pleiotropic. Furthermore, when trying to comprehend the variability in the proportion of context-specific enhancers among organs, we showed that this proportion might be partly explained by the occurrence of context-specific genes in that same organ.

Altogether, the analysis of ENCODE and FANTOM5 enhancers suggests that pleiotropic enhancers are ubiquitous regulatory elements in the human genome. The notion that such a high proportion of human enhancers can be pleiotropic suggests that small regions of regulatory DNA may contain abundant regulatory information and that these regions evolve under important

282   evolutionary constraints, as was previously observed for some mammalian (Fish et al. 2017; Hiller et

283   al. 2012) and human (Radke et al. 2021) pleiotropic enhancers.

284   Previous studies determined that pleiotropic enhancers have more density and diversity of

285   TF binding motifs (Fish et al. 2017; Singh & Yi 2021). Clearly, this pattern makes sense, since it is logical

286   to expect an increase in the number of TFBSs when more expression patterns are encoded within an

287   enhancer. Investigating the possible regulatory function of SNPs within our enhancer pleiotropy

288   categories, we found that the number of SNPs with regulatory potential increases with the degree of

289   pleiotropy of the enhancer, implying that the more pleiotropic an enhancer is, the more regulatory

290   information that enhancer contains. This finding confirms the suitability of our pleiotropy categories.

291   Although our data indicate that pleiotropic enhancers regulate gene expression in multiple

292   organs, they do not enlighten the mechanisms underlying pleiotropic function. Pleiotropic

293   enhancers may harbor distinct sets of TFBSs for driving gene expression in different organs.

294   Alternatively, the same TFBSs within pleiotropic enhancers might be needed to regulate gene

295   expression in multiple organs (i.e., pleiotropic enhancers bearing pleiotropic TFBSs). Indeed, a recent

296   study showed that pleiotropic SNPs regulate the expression of *IRX3/IRX5* genes in adipose and brain

297   tissue of humans and mice (Sobreira et al. 2021). Furthermore, a comparative analysis of the genetic

298   architecture of hundreds of human complex traits also provides evidence for the existence of

299   pleiotropic SNPs in enhancers of the human genome (Watanabe et al. 2019). Undoubtedly, this and

300   other issues related to the structure and function of pleiotropic enhancers remain to be studied in

301   detail.

302

303   **Materials and Methods**

304

305   *The FANTOM5 enhancer set*

306   The   FANTOM5   enhancer   set   was   downloaded   as   a   binary   matrix   from

307   https://enhancer.binf.ku.dk/presets. Of all the available biosamples, 139 were mapped to 41 organs

308   (Table S1) based on UBERON ontology terms, as described in Andersson et al. 2014. The FANTOM5

309   enhancers were identified in each biosample as bidirectional transcribed loci that are expressed

14

310   significantly above genomic background and that are distal from known TSSs and exons of protein

311   coding genes. These elements were subsequently filtered to remove directionally biased elements

312   (those with a higher level of transcription in one of the DNA strands), which often correspond to TSSs

313   of protein-coding genes. We eliminated from this set of consensus elements those that overlap with

314   regions that are in the ENCODE blacklist of the hg19 assembly (https://github.com/Boyle-

315   Lab/Blacklist). In total, we kept 14814 putative enhancer elements. A subset of these elements has

316   been validated with reporter constructs in HeLa cells (Andersson et al. 2014). Enhancers were

317   considered active in an organ if they were active in at least one of the biosamples that comprise the

318   organ. Enhancers included in the pleiotropy category n were those predicted to be active in n organs.

319   The size distribution of FANTOM5 enhancers is shown in figure S2.

320

321   *Generation of the ENCODE enhancer set and definition of pleiotropic enhancers*

322   For generating the ENCODE enhancer set we used high quality DNase-seq and H3K27ac ChIP-seq

323   data from 37 tissues from the ENCODE Project. We downloaded alignment files in bam format for

324   DNase-seq data (see table S4 for details) and peak files in bed narrowPeak format for H3K27ac ChIP-

325   seq data (see table S5 for details) from https://www.encodeproject.org. For each DNase-seq

326   experiment, we called peaks with MACS2 (Zhang et al. 2008) with --min-size 50 -q 0.01 -f BAMPE for

327   paired-end experiments and --min-size 50 -q 0.01 -f BAM --extsize 200 for single-end experiments.

328   Since the precise location and width of a given accessibility peak may vary between tissues and

329   technical replicates of the same tissue, we developed a method for merging elements from different

330   tissues by searching for overlaps between the summits of accessible elements rather than overlaps

331   between whole elements. To achieve this, we constructed summit confidence intervals for each

332   element in each tissue by using the variability in the position of summits of the same accessible

333   element on the different replicates. Firstly, for each tissue with replicates we estimated the signal-to-

334   noise ratio (SNR) by calculating the proportion of reads inside peaks (analogous to the Fraction of

335   Reads in Peaks score for ATAC-seq) with featureCounts (Liao et al. 2014) using the following

336   parameters: --p --countReadPairs for paired-end experiments and default parameters for single-end

337   experiments. We then intersected the highest SNR experiment MACS2 called peaks with the available

15

338   H3K27ac peaks for the same tissue (when multiple replicates of H3K27ac ChIP-seq experiments

339   where available, we used only those H3K27ac peaks represented in more than half of the replicates).

340   We then intersected these DNase-H3K27ac peaks with the DNase-seq peaks of the other replicates

341   with bedmap (Neph et al. 2012) using parameters --fraction-either 0.5 and created a gappedPeak file

342   with a custom-made R script that represents the positions of the summits of the same element in the

343   different replicates. DNase-H3K27ac peaks only present in the highest SNR file were not included in

344   subsequent analyses. We called intervals in gappedPeak files 'summit confidence intervals'. Summits

345   of tissues with only one DNase-seq were only considered if the element overlapped with H3K27ac

346   peaks in that same tissue. By overlapping the summit confidence intervals between tissues, we

347   created 'summit clusters' present in multiple tissues. We created consensus elements by merging all

348   the accessible elements (MACS2 peaks) that contribute to each summit cluster. Consensus elements

349   overlapping by more than 80% of the sequence were merged. The 37 ENCODE tissues were mapped

350   to 19 UBERON organs and one cell ontology category (Lymphocyte; CL:0000542) (Table S2).

351   Enhancers active in an organ are consensus elements predicted to be active in at least one of the

352   tissues that comprise that organ. With these data we constructed a binary matrix in which rows

353   represent consensus elements and columns represent organs. Pleiotropic enhancers were defined

354   as consensus elements predicted to be active in 2 or more organs. The code generated to create the

355   consensus elements from DNase-seq and H3K27ac ChIP-seq data can be found at

356   https://github.com/frankel-lab/human_pleiotropic_enhancers. We eliminated from this set of

357   consensus elements those that overlap with regions that are in the ENCODE blacklist of the hg38

358   assembly (https://github.com/Boyle-Lab/Blacklist), those that are located at less than 500 bp from

359   annotated transcription start sites and those that overlap exons (including putative enhancers that

360   overlap exons does not affect the analyses, data not shown). The final set of putative enhancers

361   consisted of 334189 elements.

362

363   *Comparisons between FANTOM5, ENCODE and ABC enhancer sets*

364   The ABC enhancer set was downloaded from https://www.engreitzlab.org/resources. We mapped

365   the biological samples to the corresponding UBERON organs (Table S6) in order to compare with

366 FANTOM5 and ENCODE enhancers. ABC enhancer regions overlapping by more than 80% were

367 merged with bedtools (Quinlan & Hall 2010). In total, we used 176220 ABC elements. FANTOM5 and

368 ABC enhancers were lifted to the hg38 human genome assembly with UCSC liftOver

369 (https://genome.ucsc.edu/cgi-bin/hgLiftOver) for comparing with the coordinates of the ENCODE

370 enhancers. The overlap fraction was calculated by intersecting enhancers active in each organ

371 (comparing two enhancers sets) and then dividing by the size of the smaller set (Table S7).

372

373 *Calculation of the percentage of pleiotropic enhancers*

374 To evaluate how the proportion of pleiotropic enhancers changes with the number of organs

375 considered we took all possible combinations of organs from 2 to n, where n is the total number of

376 organs (20 for the ENCODE enhancers and 41 for the FANTOM5 enhancers) and calculated the

377 proportion of enhancers active in 2 or more organs. For k organs, there exists n choose k

378 combinations of organs. For k values with more than 2000 combinations, we estimated the

379 distribution of the proportion of pleiotropic enhancers with 2000 randomly chosen sets of size k

380 (without repetition).

381 We calculated the proportion of pleiotropic enhancers (cCRE-dELS) for 17 cell types (Table S2). The

382 ENCODE cCRE-dELS (n= 290876) in these cells were defined by Moore et al (2020). We explicitly

383 excluded from the analysis data stemming from cancer derived cell lines. We also calculated the

384 proportion of pleiotropic enhancers in 39 cell types (69 primary cell cultures were combined to form

385 39 different cell types, see table S1)(Andersson et al. 2014). For comparing the degree of pleiotropy

386 of enhancers in cell types and organs we analyzed the 11400 enhancers that were detected as active

387 in both cell types and organs (Andersson et al. 2014).

388

389 *Correlation between the proportion of organ-specific enhancers and organ-specific TSSs/transcripts*

390 For FANTOM5 enhancers, we used the same CAGE-seq libraries that were used to identify enhancers

391 to estimate the proportion of context-specific TSSs (reads that correspond to first exons of genes).

392 We used the RefSeq curated gene set (ncbiRefSeqCurated, assembly hg19) from the UCSC site

393 (https://genome.ucsc.edu/). We extended each TSS 100 bp in both directions and merged

17

394   overlapping elements that were on the same DNA strand with bedtools (Quinlan & Hall 2010) (we

395   obtained similar results without merging, data not shown). We obtained a total of 25797 extended

396   TSSs. For each biosample, we counted CAGE-seq reads overlapping extended TSSs in a strand-

397   specific manner with featureCounts and compared the counts of each TSS with the count

398   distribution of 1477250 random genomic elements of the same size (we excluded exons and TSSs

399   from GRCh38 and all the ENCODE and FANTOM5 enhancers from this random genomic elements)

400   generated with bedtools. We computed an empirical p value for each TSS in each biosample as the

401   fraction of random elements with the same or greater number of reads than the TSS in that

402   biosample. We corrected these empirical p values with the Benjamini-Hochberg procedure and

403   considered that a TSS is active in a biosample if its corrected p value was 0.01 or less (we obtained

404   similar results using thresholds of $p<0.005$ or $p<0.001$; data not shown).

405   For ENCODE enhancers we used ENCODE RNA-seq data available for 25 of the 37 tissues that were

406   used to create the enhancer set (Table S8). We downloaded FPKM normalized counts and only kept

407   genes with protein products annotated by the HUGO Gene Nomenclature Committee (19167 genes).

408   When replicates of the same tissues were available, we used the median FPKM for each gene. In

409   order to use a uniform threshold across tissues to decide when a gene is expressed or not, we used

410   the zFPKM R package (Hart et al. 2013) to normalize FPKM. Genes were considered to be expressed

411   if their zFPKM was above -2 (we obtained similar results using -3.5, -3, -2.5, -1,0 and 1 as thresholds;

412   data not shown). We regard a gene as context-specific when the gene is expressed in a single organ

413   (for both the FANTOM5 and ENCODE analyses).

414

415   *RegulomeDB SNP analysis*

416   SNPs with regulatory potential were downloaded from the RegulomeDB site

417   (https://regulomedb.org/files/TSTFF344324). We lifted the hg19 positions to the hg38 assembly

418   using UCSC liftOver. For this analysis we considered SNPs in category 1 (SNPs that are eQTLs and that

419   have at least one additional source of evidence for being regulatory) and the most stringent category

420   of RegulomeDB, which is category 1a (Boyle et al. 2012). Odds ratios were calculated for each

421   pleiotropy category as the RegulomeDB SNPs contained in the category divided by the non-

422 RegulomeDB SNPs (1000 Genomes SNPs with a MAF > 0.05 in Europeans) in that same category

423 divided by the same ratio for the rest of the genome. In total we considered 71263 SNPs in category

424 1 and 978 SNPs in category 1a, and 7894464 common variants. All SNP count operations were

425 performed with bedmap.

426

427 **Acknowledgments**

428

432

433 **Competing interests**

434 The authors have no competing interests to declare

435

436 **Data availability statement**

437 The data underlying this article are available in the article and in its online supplementary material.

438

439 **References**

440

441 Andersson R et al. 2014. An atlas of active enhancers across human cell types and tissues. Nature.

442 507:455–461. doi: 10.1038/nature12787.

443 Boyle AP et al. 2012. Annotation of functional variation in personal genomes using RegulomeDB.

444 Genome Res. 22:1790–1797. doi: 10.1101/gr.137323.112.

445 Corradin O, Scacheri PC. 2014. Enhancer variants: Evaluating functions in common disease.

446 Genome Med. 6:1–14. doi: 10.1186/s13073-014-0085-3.

447 Creyghton MP et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts

448 developmental state. Proc. Natl. Acad. Sci. U. S. A. 107:21931–21936. doi: 10.1073/pnas.1016071107.

449 Djureinovic D et al. 2014. The human testis-specific proteome defined by transcriptomics and

450   antibody-based profiling. Mol. Hum. Reprod. 20:476–488. doi: 10.1093/molehr/gau018.

451   Dunham I et al. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature.

452   489:57–74. doi: 10.1038/nature11247.

453   Fish A, Chen L, Capra JA. 2017. Gene regulatory enhancers with evolutionarily conserved activity

454   aremore pleiotropic than those with species-specific activity. Genome Biol. Evol. 9:2615–2625. doi:

455   10.1093/gbe/evx194.

456   Fulco CP et al. 2019. Activity-by-contact model of enhancer–promoter regulation from thousands

457   of CRISPR perturbations. Nat. Genet. 51:1664–1669. doi: 10.1038/s41588-019-0538-0.

458   Fuqua T et al. 2020. Dense and pleiotropic regulatory information in a developmental enhancer.

459   Nature. 587:235–239. doi: 10.1038/s41586-020-2816-5.

460   Gasperini M, Tome JM, Shendure J. 2020. Towards a comprehensive catalogue of validated and

461   target-linked human enhancers. Nat. Rev. Genet. 21:292–310. doi: 10.1038/s41576-019-0209-0.

462   Hart T, HK K, S L, K P, DR S. 2013. Finding the active genes in deep RNA-seq gene expression studies.

463   BMC Genomics. 14. doi: 10.1186/1471-2164-14-778.

464   Heintzman ND et al. 2009. Histone modifications at human enhancers reflect global cell-type-

465   specific gene expression. Nature. 459:108–112. doi: 10.1038/nature07829.

466   Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating

467   variation in gene expression. Nat. Rev. Genet. 22:203–215. doi: 10.1038/s41576-020-00304-w.

468   Hiller M, Schaar B, Bejerano G. 2012. Hundreds of conserved non-coding genomic regions are

469   independently lost in mammals. Nucleic Acids Res. 40:11463–11476. doi: 10.1093/NAR/GKS905.

470   Infante CR et al. 2015. Shared Enhancer Activity in the Limbs and Phallus and Functional

471   Divergence of a Limb-Genital cis-Regulatory Element in Snakes. Dev. Cell. 35:107–119. doi:

472   10.1016/j.devcel.2015.09.003.

473   Kimura H. 2013. Histone modifications for human epigenome analysis. J. Hum. Genet. 58:439–445.

474   doi: 10.1038/jhg.2013.66.

475   Kittelmann S, Preger-Ben Noon E, McGregor AP, Frankel N. 2021. A complex gene regulatory

476   architecture underlies the development and evolution of cuticle morphology in Drosophila. Curr.

477   Opin. Genet. Dev. 69:21–27. doi: 10.1016/j.gde.2021.01.003.

478   Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome.

479   Nat. Rev. Genet. 20:207–220. doi: 10.1038/s41576-018-0089-8.

480   Kristjánsdóttir K, Dziubek A, Kang HM, Kwak H. 2020. Population-scale study of eRNA transcription

481   reveals bipartite functional enhancer architecture. Nat. Commun. 11:1–12. doi: 10.1038/s41467-

482   020-19829-z.

483   Kvon EZ et al. 2020. Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human

484   Enhancer Variants. Cell. 180:1262-1271.e15. doi: 10.1016/j.cell.2020.02.031.

485   Levine M. 2010. Transcriptional enhancers in animal development and evolution. Curr. Biol.

486   20:R754-63. doi: 10.1016/j.cub.2010.06.070.

487   Lewis JJ et al. 2019. Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing

488   pattern mimicry. Proc. Natl. Acad. Sci. U. S. A. 116:24174–24183. doi: 10.1073/pnas.1907068116.

489   Liao Y, GK S, W S. 2014. featureCounts: an efficient general purpose program for assigning sequence

490   reads to genomic features. Bioinformatics. 30:923–930. doi: 10.1093/BIOINFORMATICS/BTT656.

491   Lidschreiber K et al. 2021. Transcriptionally active enhancers in human cancer cells. Mol. Syst. Biol.

492   17. doi: 10.15252/msb.20209873.

493   Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D. 2014. Convergent evolution of complex

494   regulatory landscapes and pleiotropy at Hox loci. Science (80-.). 346:1004–1006. doi:

495   10.1126/science.1257493.

496   Minnoye L et al. 2021. Chromatin accessibility profiling methods. Nat. Rev. Methods Prim. 1:1–24.

497   doi: 10.1038/s43586-020-00008-9.

498   Moore JE et al. 2020. Expanded encyclopaedias of DNA elements in the human and mouse

499   genomes. Nat. 2020 5837818. 583:699–710. doi: 10.1038/s41586-020-2493-4.

500   Mungall CJ, C T, GV G, SE L, MA H. 2012. Uberon, an integrative multi-species anatomy ontology.

501   Genome Biol. 13. doi: 10.1186/GB-2012-13-1-R5.

502   Nasser J et al. 2021. Genome-wide enhancer maps link risk variants to disease genes. Nat. 2021

503   5937858. 593:238–243. doi: 10.1038/s41586-021-03446-x.

504   Neph S et al. 2012. BEDOPS: high-performance genomic feature operations. Bioinformatics.

505   28:1919. doi: 10.1093/BIOINFORMATICS/BTS277.

506 Panigrahi A, O'Malley BW. 2021. Mechanisms of enhancer action: the known and the unknown.

507 Genome Biol. 22:1–30. doi: 10.1186/s13059-021-02322-1.

508 Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. 2013. Enhancers: Five essential

509 questions. Nat. Rev. Genet. 14:288–295. doi: 10.1038/nrg3458.

510 Pineau C et al. 2019. Cell Type-Specific Expression of Testis Elevated Genes Based on

511 Transcriptomics and Antibody-Based Proteomics. J. Proteome Res. 18:4215–4230. doi:

512 10.1021/acs.jproteome.9b00351.

513 Le Poul Y et al. 2020. Regulatory encoding of quantitative variation in spatial activity of a

514 Drosophila enhancer. Sci. Adv. 6:eabe2955. doi: 10.1126/sciadv.abe2955.

515 Preger-Ben Noon E et al. 2018. Comprehensive Analysis of a cis-Regulatory Region Reveals

516 Pleiotropy in Enhancer Function. Cell Rep. 22:3021–3031. doi: 10.1016/j.celrep.2018.02.073.

517 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.

518 Bioinformatics. 26:841–842. doi: 10.1093/BIOINFORMATICS/BTQ033.

519 Rada-Iglesias A et al. 2011. A unique chromatin signature uncovers early developmental enhancers

520 in humans. Nature. 470:279–285. doi: 10.1038/nature09692.

521 Radke DW et al. 2021. Purifying selection on noncoding deletions of human regulatory loci

522 detected using their cellular pleiotropy. Genome Res. 31:935–946. doi: 10.1101/GR.275263.121.

523 Reilly SK et al. 2021. Direct characterization of cis-regulatory elements and functional dissection of

524 complex genetic associations using HCR–FlowFISH. Nat. Genet. 2021 538. 53:1166–1176. doi:

525 10.1038/s41588-021-00900-4.

526 Sabarís G, Laiker I, Preger-Ben Noon E, Frankel N. 2019. Actors with Multiple Roles: Pleiotropic

527 Enhancers and the Paradigm of Enhancer Modularity. Trends Genet. 35:423–433. doi:

528 10.1016/j.tig.2019.03.006.

529 Sartorelli V, Lauberth SM. 2020. Enhancer RNAs are an important regulatory layer of the epigenome.

530 Nat. Struct. Mol. Biol. 27:521–528. doi: 10.1038/s41594-020-0446-0.

531 Singh D, Yi S. 2021. Enhancer Pleiotropy, Gene Expression, and the Architecture of Human

532 Enhancer-Gene Interactions. Mol. Biol. Evol. 38:3898–3909. doi: 10.1093/MOLBEV/MSAB085.

533 Smith E, Shilatifard A. 2014. Enhancer biology and enhanceropathies. Nat. Struct. Mol. Biol. 21:210–

534   219. doi: 10.1038/nsmb.2784.

535   Sobreira DR et al. 2021. Extensive pleiotropism and allelic heterogeneity mediate metabolic effects

536   of IRX3 and IRX5. Science (80-.). 372:1085–1091. doi: 10.1126/SCIENCE.ABF1008.

537   Watanabe K et al. 2019. A global overview of pleiotropy and genetic architecture in complex traits.

538   Nat. Genet. 2019 519. 51:1339–1348. doi: 10.1038/s41588-019-0481-0.

539   Xin Y et al. 2020. Ancestral and derived transcriptional enhancers share regulatory sequence and a

540   pleiotropic site affecting chromatin accessibility. Proc. Natl. Acad. Sci. U. S. A. 117:20636–20644. doi:

541   10.1073/pnas.2004003117.

542   Zhang Y et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008 99. 9:1–9. doi:

543   10.1186/GB-2008-9-9-R137.

544   Zhu J et al. 2016. Identification of Tissue-Specific Protein-Coding and Noncoding Transcripts across

545   14 Human Tissues Using RNA-seq. Sci. Rep. 6:1–11. doi: 10.1038/srep28400.