

Pleiotropic Enhancers are Ubiquitous Regulatory Elements in the Human Genome

Ian Laiker¹ and Nicolás Frankel ^{1,2,*}

¹Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y Universidad de Buenos Aires (UBA), Buenos Aires 1428, Argentina

²Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales (FCEN), Universidad de Buenos Aires (UBA), Buenos Aires 1428, Argentina

*Corresponding author: E-mail: nfrankel@conicet.gov.ar.

Accepted: 09 May 2022

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. Historically, enhancers have been regarded as highly tissue-specific. However, recent evidence has demonstrated 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 analyzed active enhancers across human organs based on the analysis of both eRNA transcription (FANTOM5 consortium data sets) and chromatin architecture (ENCODE consortium data sets). 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.

Key words: human genome, non-coding DNA, gene regulation, CRE, enhancer, pleiotropy.

Significance

The human genome contains a vast number of regulatory elements, named enhancers, that control the tempo and mode of gene expression. Historically, enhancers have been regarded as genetic elements that are active in a single organ, but recent evidence has demonstrated 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 and O'Malley 2021). There are several features that distinguish

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

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 and 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 (Heintzman et al. 2009; Pennacchio et al. 2013). It has been shown that a large percentage of SNPs associated with human diseases fall within regions that are predicted to be enhancers (Corradin and Scacheri 2014), and, in the same vein, numerous cases in which mutations in enhancers cause disease have been reported (Smith and Shilatifard 2014). Although thousands of non-coding regions have been identified as putative enhancers, very little is known about the function of most of these predicted enhancers (Gasparini 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 (Nasser et al. 2021; Reilly et al. 2021). The fact that regulatory information within enhancer DNA can be dense and pleiotropic (Fuqua et al. 2020; Kvon et al. 2020; Le Poul et al. 2020; Xin et al. 2020) also illustrates regulatory complexity.

The idea that enhancers have strict tissue-specific activities (i.e., modular or context-specific activities) has been undermined by recent experimental data (Lonfat et al. 2014; Infante et al. 2015; Preger-Ben Noon et al. 2018; Lewis et al. 2019; Fuqua et al. 2020; Xin et al. 2020). Current evidence suggests that a large proportion of enhancers in animal genomes are active in multiple developmental contexts (Sabaris et al. 2019; Kittelmann et al. 2021). Furthermore, genomic analyses have shown that a significant number of enhancers in the human genome are pleiotropic (i.e., have roles in multiple organs and/or developmental stages) (Andersson et al. 2014; Moore et al. 2020). Although evidence shows that pleiotropy in human regulatory DNA is pervasive, a quantification of enhancer pleiotropy in the human genome is currently lacking. In this sense, it would be useful to have a reliable estimate of the percentage of pleiotropic enhancers in the human

genome and to explore the landscape of pleiotropy in each organ.

To assess the magnitude of pleiotropy in the regulatory genome, in this study we estimated the extent of enhancer reuse and analyzed the abundance of pleiotropic enhancers per organ. We show that pleiotropic enhancers are ubiquitous in the human genome and that the majority of 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.

Results

A Large Proportion of Predicted Human Enhancers are Pleiotropic

We first set out to estimate the number of pleiotropic enhancers in the human genome by using information from FANTOM5 (Andersson et al. 2014) and ENCODE (Dunham et al. 2012; Moore et al. 2020) projects. We grouped ENCODE and FANTOM biosamples into organs using UBERON anatomy ontology (Mungall et al. 2012) (see [supplementary tables S1 and S2, Supplementary Material](#) online). For FANTOM5 data, putative enhancers were defined as DNA regions transcribed bidirectionally above a certain threshold in at least one of 41 organs (see [supplementary table S1, Supplementary Material](#) online for a list of samples and organs). For ENCODE data, we defined putative enhancers as DNA regions with open chromatin that contain the epigenetic mark H3K27ac in at least one of 20 organs (see [supplementary table S2, Supplementary Material](#) online for a list of samples and organs). We observed a moderate degree of overlap between FANTOM5 and ENCODE enhancer sets; 40–60% of the FANTOM5 enhancers are contained within the larger ENCODE enhancer set in 16 common organs (testis is an exception, see [supplementary fig. S1, Supplementary Material](#) online). Also, we compared ENCODE enhancers with the “ABC enhancer set”, which was defined based on chromatin accessibility, the presence of the epigenetic mark H3K27ac and the interaction of the DNA region with core promoters (Fulco et al. 2019). We determined that, for 14 common organs, the overlap between our enhancer set predicted with ENCODE data and Activity By Contact (ABC) enhancers ranges between 43 and 83% ([supplementary fig. S1, Supplementary Material](#) online).

By intersecting genomic information from different organs we aimed to estimate the proportion of pleiotropic enhancers in the human genome. Hence, we considered that a pleiotropic FANTOM5 enhancer is a DNA region transcribed bidirectionally in two or more organs. For ENCODE data, we delimited consensus elements

considering the different organs in which an enhancer was predicted to be active, based on the presence of open chromatin and the H3K27ac mark (see details of the pipeline in [supplementary fig. S2, Supplementary Material](#) online and Materials and Methods, and [supplementary fig. S3, Supplementary Material](#) online for the size distribution of consensus elements). In this case, we define a pleiotropic enhancer as a consensus element active in two or more organs. We decided to generate our own enhancer set, instead of using cCRE-dELS enhancer elements (ENCODE 3 definitions), because this allowed us to group biosamples into organs and then compute enhancer pleiotropy by comparing enhancer usage among organs. However, to show that our conclusions are independent of enhancer definitions, we did use cCRE-dELS for additional calculations (see below).

To evaluate how the proportion of pleiotropic enhancers changes with the number of organs considered we took combinations of organs from 2 to n out of the pool of ENCODE organs ($n=20$) and FANTOM5 organs ($n=41$) and calculated the proportion of enhancers active in 2 or more organs. This approach allows us to comprehend how the proportion of pleiotropic enhancers changes depending on the number of organs considered and, at the same time, to validate our methods. We observed that the median of predicted pleiotropic enhancers increases monotonically with the number of intersected organs,

both for FANTOM5 and ENCODE data ([fig. 1](#)). Although ENCODE and FANTOM5 enhancer sets are different (~50% of the FANTOM5 enhancer set is contained within the larger ENCODE enhancer set and, as well, the FANTOM5 enhancer set represents enhancers active in more organs than the ENCODE set) the shape of the two hypothetical curves are remarkably similar ([fig. 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 ([fig. 1](#)). We observed that the majority of FANTOM5 and ENCODE pleiotropic enhancers are active in 2–5 organs ([supplementary fig. S4, Supplementary Material](#) online). To investigate how enhancer pleiotropy would change with a deeper sampling of contexts, we fitted a log-linear model to the data. For FANTOM5 data the fitting was slightly better than for ENCODE data ($R_{\text{FANTOM5}}^2: 0.99$ vs. $R_{\text{ENCODE}}^2: 0.98$). When extrapolating to 80 organs (an estimate of the total number of organs in the human body), the percentage of pleiotropic enhancers would be ~52% for FANTOM5 and ~63% for ENCODE.

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

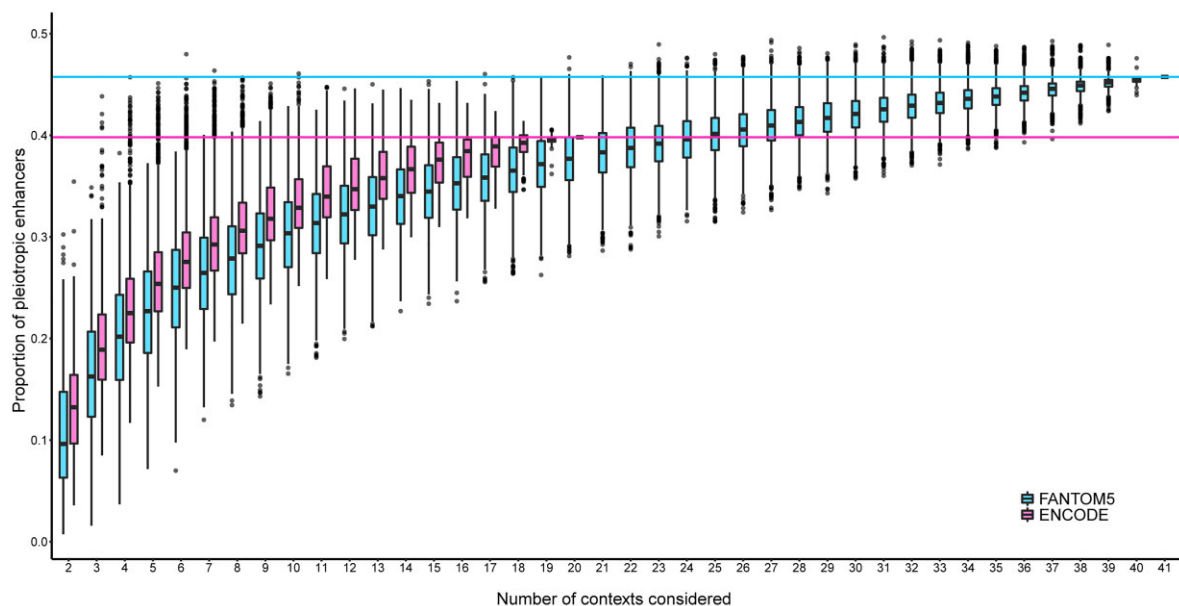


Fig. 1.—A large proportion of predicted human enhancers are pleiotropic. Proportion of pleiotropic enhancers as a function of the number of contexts considered in FANTOM5 (cyan) and ENCODE (magenta) enhancer sets. Boxplots indicate the median (horizontal line), interquartile range (box), observations within ± 1.5 interquartile 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).

in primary cell cultures by the FANTOM5 project (supplementary table S1, Supplementary Material online) (Andersson et al. 2014), combining the same cell types from different organs as a single context (see supplementary table S1, Supplementary Material online). Notably, when considering these 38 distinct cell types, the percentage of pleiotropic enhancers turns out to be 68%. We performed the same analysis using data from ENCODE 3 in 17 cell types (supplementary table S2, Supplementary Material online) and a different enhancer definition (cCRE-dELS defined by ENCODE 3, (Moore et al. 2020)). When considering all 17 cell types, the percentage of pleiotropic enhancers is 39%. In order to determine whether results from cells and organs are comparable, we contrasted the degree of pleiotropy of enhancers obtained with FANTOM5 cell types data with that calculated using FANTOM5 organ data (supplementary fig. S5, Supplementary Material online). We observed that the degree of pleiotropy of enhancers estimated with cell types data and organ data is remarkably similar (supplementary fig. S5, Supplementary Material online). Altogether, these data suggest that enhancer pleiotropy is mostly due to enhancers being active in different cell types of organs, rather than enhancers being active in the same cell type present in multiple organs. However, some of the cell lines that we considered distinct may be ontogenetically close and this may affect pleiotropy calculations, because cell types that derive from the same progenitors often have similar transcriptomes. Thus, ontogenetically close cell types will share a large repertoire of expressed genes (pleiotropic genes), and, certainly, this may contribute to a substantial part of enhancer pleiotropy. In fact, when we analyze the transcriptome and enhancer usage in FANTOM5 cell lines, we observe that cell lines with similar transcriptomes have more enhancers in common than cell lines with divergent transcriptomes (supplementary fig. S6, Supplementary Material online). These results indicate that ontogenetic proximity of cells influences enhancer pleiotropy.

Most Enhancers Active in a Particular Organ are Pleiotropic

Next, we sought to explore the abundance of pleiotropic enhancers in the different organs. To this end, we calculated the proportion of predicted pleiotropic enhancers per organ in the FANTOM5 and ENCODE data sets (fig. 2). We observed values ranging from 43.6% to 98.1% in FANTOM5 organs (fig. 2A, supplementary table S3, Supplementary Material online) and from 53.4% to 96.8% in ENCODE organs (fig. 2B, supplementary table S3, Supplementary Material online). These data suggest that most enhancers active in a given organ are also active in one or more other organs.

Indeed, most of the pleiotropic enhancers active in a given organ are active in more than two organs (fig. 2).

Given that 17 organs are shared between FANTOM5 and ENCODE datasets, we compared the estimations of the proportion of predicted pleiotropic enhancers for both analyses. We observed that they are notably concordant, especially considering that biosamples are derived from different parts of the organ and different individuals, and that enhancers were predicted through distinct methods (see supplementary table S3, Supplementary Material online). For most comparisons the difference in the estimation of the proportion of pleiotropic enhancers is smaller than 10%. However, testis is an outlier, with a difference of more than 20%. Despite differences in the estimation of the proportion of pleiotropic enhancers for testis, both FANTOM5 and ENCODE data indicate that this organ is rich in context-specific enhancers (fig. 2). Since it has been reported that testis is the organ with the most context-specific genes (Djureinovic et al. 2014; Zhu et al. 2016; Pineau et al. 2019), we reasoned that a possible cause for the high number of context-specific enhancers in the testis could be the high number of testis-specific genes and, moreover, that there could be a relationship between the number of context-specific genes and the abundance of context-specific enhancers per organ. To test this hypothesis, we computed the number of organ-specific genes in ENCODE and FANTOM5 organs (with RNA-seq and transcribed TSSs data respectively, see Materials and Methods for details) and, subsequently, we looked for a possible correlation between the proportion of organ-specific genes and the proportion of organ-specific enhancers. Remarkably, for both FANTOM5 and ENCODE organs, we found a positive correlation (Pearson's $r=0.653$, $P<0.001$ and Pearson's $r=0.647$, $P<0.001$, respectively) between the proportion of context-specific genes and the proportion of context-specific enhancers (fig. 3). A positive correlation is also observed when using data from FANTOM5 cell types (Pearson's $r=0.71$, $P<0.001$). Accordingly, when we analyze the predicted target genes of some ENCODE enhancers ($n=50,090$, see Materials and Methods), we observe that context-specific genes are more likely to be regulated by context-specific enhancers than pleiotropic genes [Fisher exact test, odds ratio 95% CI = (1.7–2.2), $P<2.2e-16$].

Pleiotropic Enhancers are Enriched in SNPs with High Regulatory Potential

It has been shown that pleiotropic enhancers are larger than context-specific enhancers (Fong and Capra 2021; Singh and Yi 2021) and that they have more density of TF binding motifs (Fish et al. 2017; Singh and Yi 2021). We reasoned that a way to provide a validation for our pleiotropy analysis was to explore a possible connection

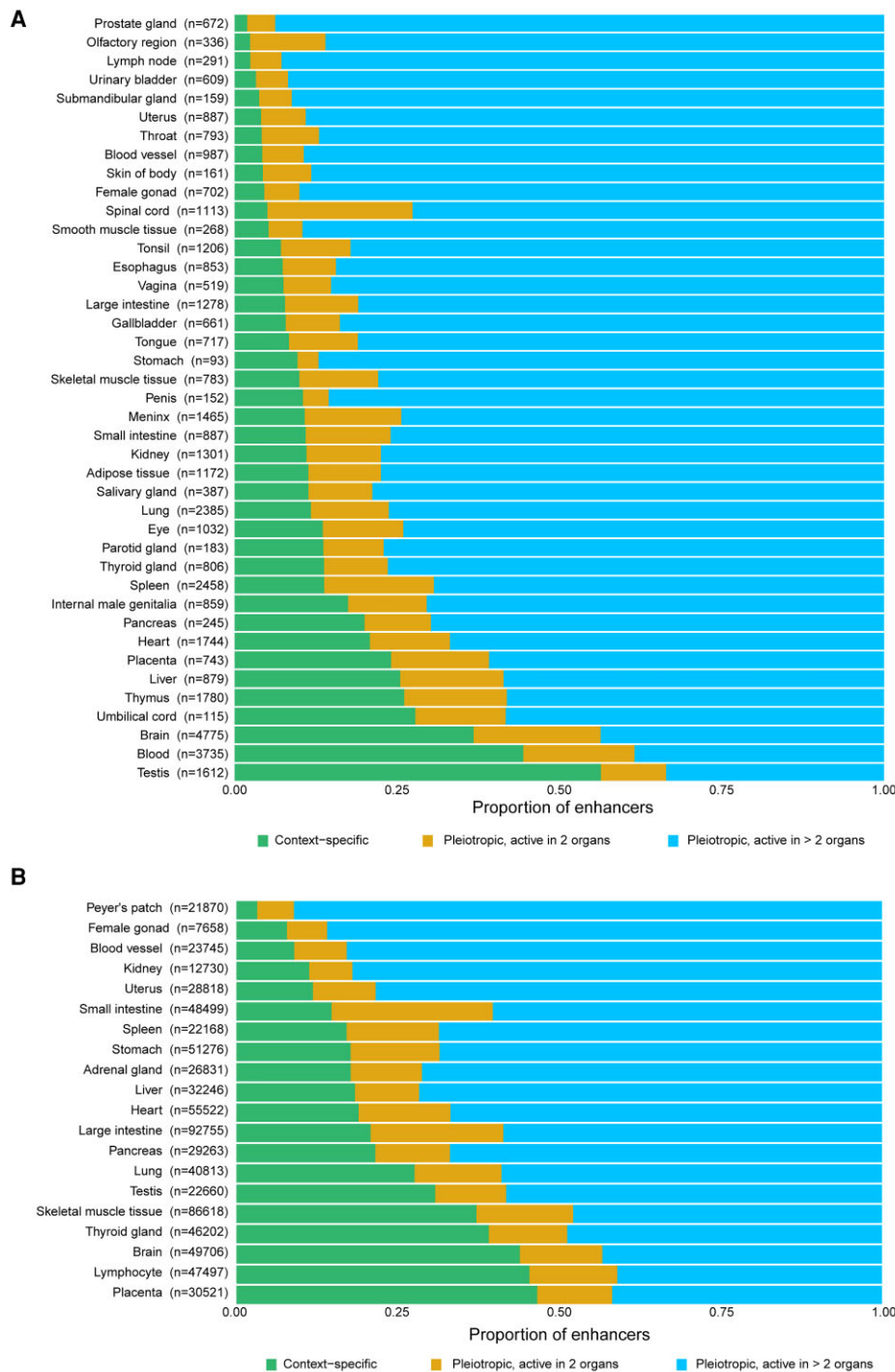


Fig. 2.—The majority of active enhancers in a particular organ are pleiotropic. Proportion of context-specific enhancers (green), pleiotropic enhancers active in two organs (orange) and pleiotropic enhancers active in more than two organs (blue) in FANTOM5 (A) and ENCODE (B) organs. The number of predicted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

between the degree of pleiotropy of enhancers and the abundance of genetic variants related to gene regulation. Thus, we interrogated the Regulome database (Boyle et al. 2012), which provides a means to predict the

regulatory potential of SNPs across the genome. In this approach, SNPs are given a score based on their likelihood of being variants with regulatory function (see Materials and Methods for details). Hence, we quantified the enrichment

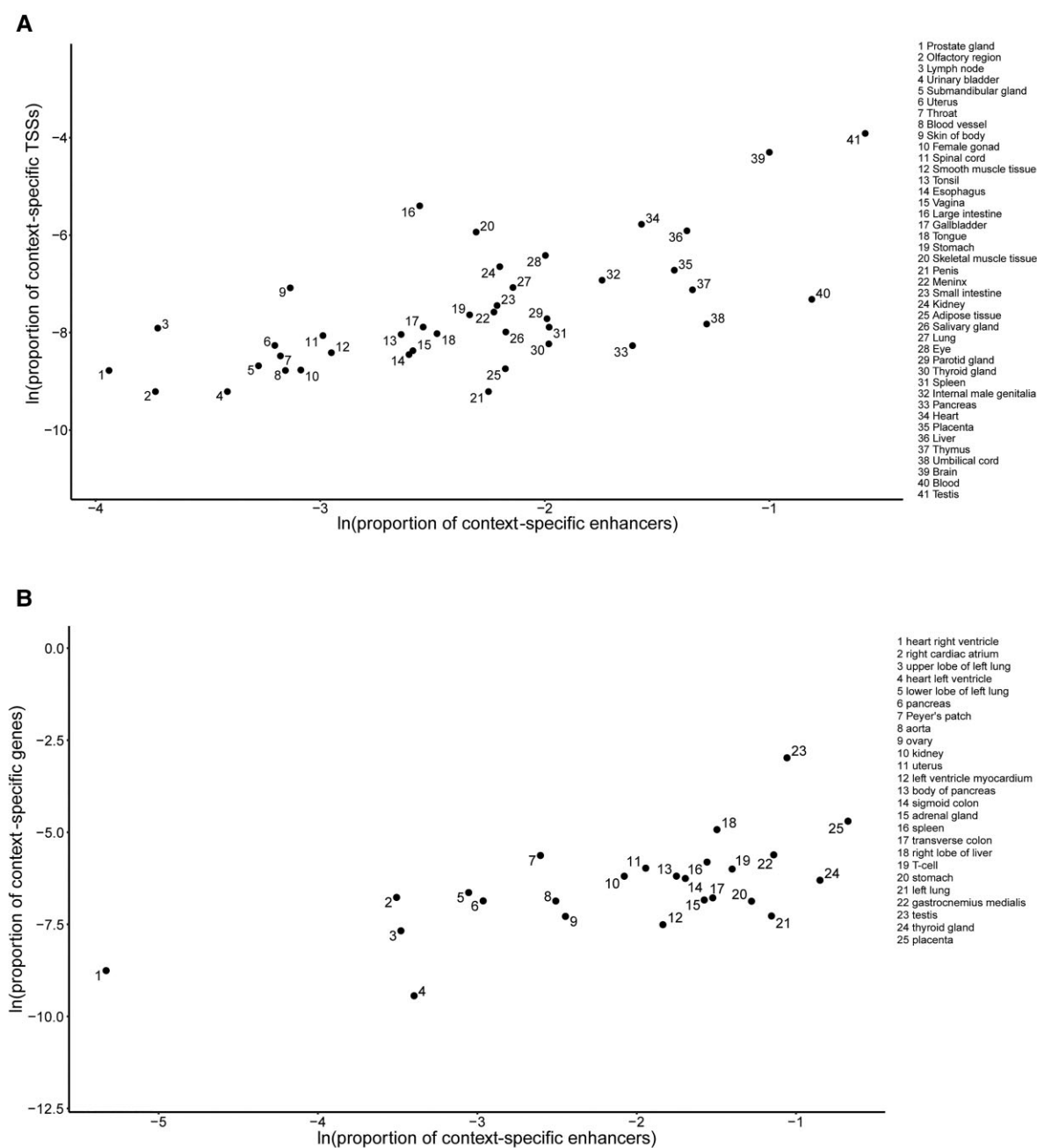


FIG. 3.—The proportion of context-specific enhancers is positively correlated with the proportion of context-specific genes. (A) Relationship between the proportion of context-specific enhancers and context-specific TSSs for the FANTOM5 data set. Active TSSs were defined in each organ by comparing their 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).

of SNPs with predicted regulatory activity within categories that group enhancers based on the number of contexts in which they are active (context-specific enhancers comprise Category 1 and pleiotropic enhancers comprise categories 2–20). We noted that there is a positive relationship between the degree of pleiotropy of enhancers and the level

of enrichment in SNPs with predicted regulatory function (fig. 4A). When using only SNPs in the most stringent category of Regulome database [Category 1a: SNPs that (i) are eQTLs, (ii) match TF motifs, (iii) match DNase footprints, (iv) lie within DNase peaks, and (v) lie within TF binding peaks] the positive relationship holds, and the

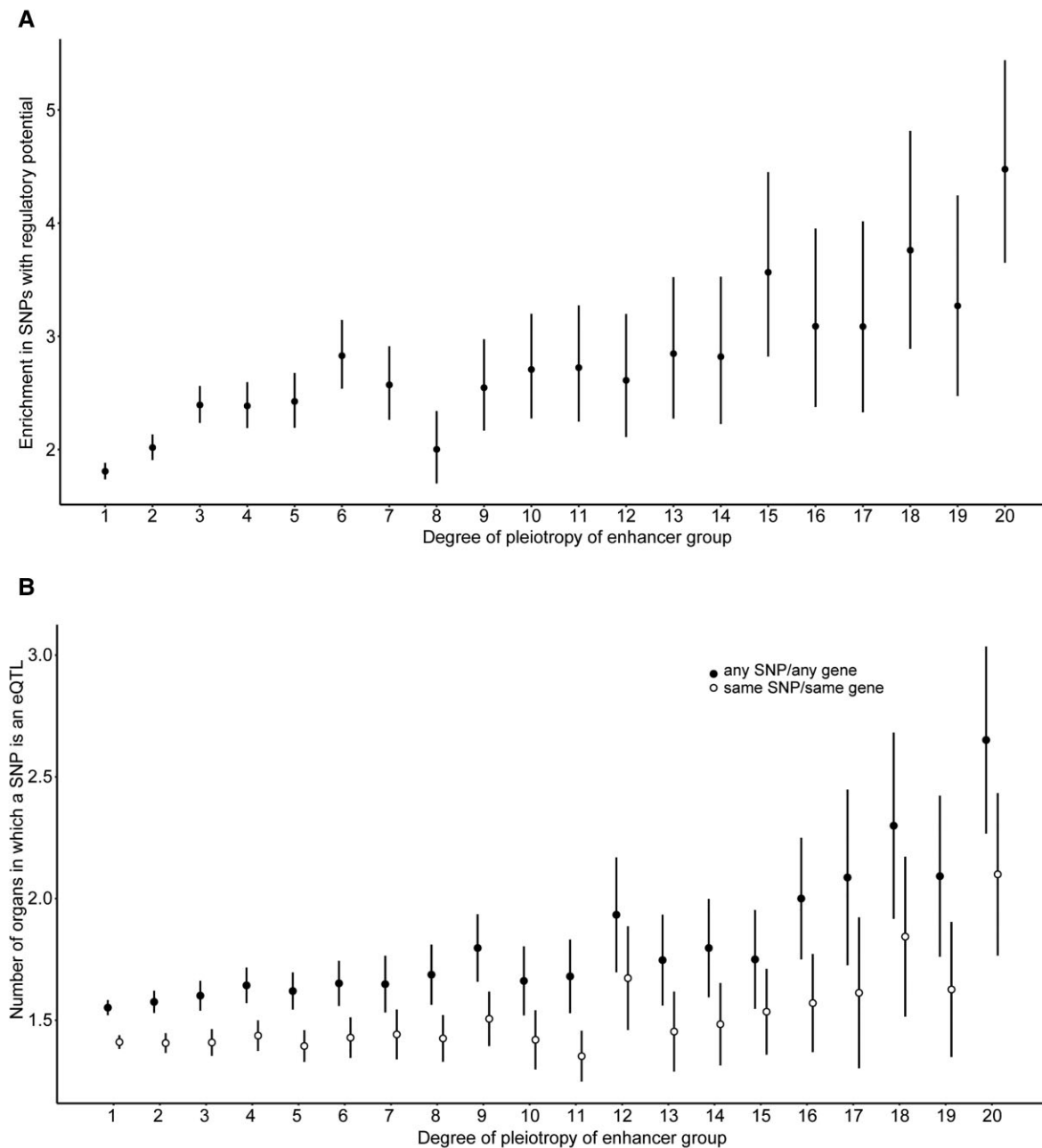


Fig. 4.—Pleiotropic enhancers are enriched in SNPs with high regulatory potential. (A) The degree of pleiotropy of ENCODE enhancers is positively correlated with the enrichment in SNPs with high regulatory potential (Category 1 SNPs of Regulome DB). Black circles represent the odds ratio and bars represent the 95% CI. Enrichments (odds ratios) were calculated using common SNPs as background (see Materials and Methods for details). (B) The number of organs in which ENCODE enhancer SNPs are associated with changes in gene expression is positively correlated with the degree of pleiotropy of enhancers. Open circles indicate the mean number of organs in which a single SNPs is associated with changes in the expression of the same gene. Black circles represent the mean number of organs in which enhancers SNPs are associated with changes in the expression of any gene [considers cases in which (i) a single SNP regulates different genes in different organs, (ii) the same gene is regulated by different SNPs of the same enhancer in different organs, or (iii) when different SNPs of the same enhancer regulate different genes in different organs]. Bars represent the 95% CI.

enrichment becomes even larger (supplementary fig. S7, Supplementary Material online).

We also wanted to determine how enhancer pleiotropy affects variation in gene expression. In this sense, an

obvious prediction would be that pleiotropic enhancers will influence gene expression in more organs than context-specific enhancers. Using the GTEx database (Lonsdale et al. 2013), we searched for fine-mapped *cis*-eQTLs among

SNPs that fall within ENCODE enhancers with different degrees of pleiotropy. This analysis showed a clear trend indicating that the more pleiotropic the enhancer, the higher the number of organs in which its SNPs are eQTLs (i.e., the number of organs in which its SNPs are associated with changes in gene expression) (fig. 4B).

Discussion

In this study, we surveyed enhancer activity across human organs by analyzing different types of genomic data. We used enhancer predictions based on either the analysis of eRNA transcription (FANTOM5 enhancers) or the analysis of chromatin architecture (ENCODE enhancers). 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 did not analyze the entirety of human organs at all developmental stages (data is only available for some adult organs and a few fetal organs) 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). Future analyses using single-cell data may help unravel the contribution of the different cell types of organs to pleiotropy estimations.

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. However, we cannot rule out the possibility that the correlation between the proportion of context-specific enhancers and context-specific genes is induced by the developmental trajectories of sampled contexts. Further analyses accounting for the developmental relationships between contexts would be needed to exclude this possibility.

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 evolutionary constraints, as was previously observed for some mammalian (Hiller et al. 2012; Fish et al. 2017) and human (Radke et al. 2021) pleiotropic enhancers.

Previous studies determined that pleiotropic enhancers have more density and diversity of TF binding motifs (Fish et al. 2017; Singh and Yi 2021). Clearly, this pattern makes sense, since it is logical to expect an increase in the number of TFBSs when more expression patterns are encoded within an enhancer. Investigating the possible regulatory function of SNPs within our enhancer pleiotropy categories, we found that the number of SNPs with regulatory potential increases with the degree of pleiotropy of the enhancer, implying that the more pleiotropic an enhancer is, the more regulatory information that enhancer contains. Nevertheless, we cannot rule out the possibility that this pattern of enrichment is simply caused by SNPs in pleiotropic enhancers having more chances to be detected in experimental essays. We also found that the number of organs in which enhancer SNPs are eQTLs escalates with the degree of pleiotropy of the enhancer. This finding suggests that changes in the activity of pleiotropic enhancers will have greater phenotypic consequences than changes in the activity of context-specific enhancers.

Although our data indicate that pleiotropic enhancers regulate gene expression in multiple organs, they do not enlighten the mechanisms underlying pleiotropic function. Pleiotropic enhancers may harbor distinct sets of TFBSs for driving gene expression in different organs. Alternatively, the same TFBSs within pleiotropic enhancers might be needed to regulate gene expression in multiple organs (i.e., pleiotropic enhancers bearing pleiotropic TFBSs). Indeed, a recent study showed that pleiotropic SNPs regulate the expression of *IRX3/IRX5* genes in adipose and brain tissue of humans and mice (Sobreira et al. 2021). Furthermore, a comparative analysis of the genetic architecture of hundreds of human complex traits also provides evidence for the existence of pleiotropic SNPs in enhancers of the human genome (Watanabe et al. 2019). Undoubtedly, this and other issues related to the structure and function of pleiotropic enhancers remain to be studied in detail.

Materials and Methods

The FANTOM5 Enhancer Set

The FANTOM5 enhancer set was downloaded as a binary matrix from <https://enhancer.binf.ku.dk/presets>. Of all the available biosamples, 139 were mapped to 41 organs (supplementary table S1, Supplementary Material online)

based on UBERON ontology terms, as described in Andersson et al. (2014). The FANTOM5 enhancers were identified in each biosample as bidirectional transcribed loci that are expressed significantly above genomic background and that are distal from known TSSs and exons of protein-coding genes. These elements were subsequently filtered to remove directionally biased elements (those with a higher level of transcription in one of the DNA strands), which often correspond to TSSs of protein-coding genes. We eliminated from this set of consensus elements those that overlap with regions that are in the ENCODE blacklist of the hg19 assembly (<https://github.com/Boyle-Lab/Blacklist>). In total, we kept 14,814 putative enhancer elements. A subset of these elements has been validated with reporter constructs in HeLa cells (Andersson et al. 2014). Enhancers were considered active in an organ if they were active in at least one of the biosamples that comprise the organ. Enhancers included in the pleiotropy category n were those predicted to be active in n organs. The size distribution of FANTOM5 enhancers is shown in [supplementary figure S2, Supplementary Material](#) online.

Generation of the ENCODE Enhancer Set and Definition of Pleiotropic Enhancers

For generating the ENCODE enhancer set we used high quality DNase-seq and H3K27ac ChIP-seq data from 37 tissues from the ENCODE Project. We downloaded alignment files in bam format for DNase-seq data (see [supplementary table S4, Supplementary Material](#) online for details) and peak files in bed narrowPeak format for H3K27ac ChIP-seq data (see [supplementary table S5, Supplementary Material](#) online for details) from <https://www.encodeproject.org>. For each DNase-seq experiment, we called peaks with MACS2 (Zhang et al. 2008) with `-min-size 50 -q 0.01 -f BAMPE` for paired-end experiments and `-min-size 50 -q 0.01 -f BAM -extsize 200` for single-end experiments. Since the precise location and width of a given accessibility peak may vary between tissues and technical replicates of the same tissue, we developed a method for merging elements from different tissues by searching for overlaps between the summits of accessible elements rather than overlaps between whole elements. To achieve this, we constructed a “summit confidence interval” for each element in each tissue by considering variation in the position of summits between replicates (see [supplementary fig. S2, Supplementary Material](#) online). First, for tissues with DNase-seq replicates, we estimated the signal-to-noise ratio (SNR) for each replicate by calculating the proportion of reads inside peaks (analogous to the Fraction of Reads in Peaks score for ATAC-seq) with featureCounts (Liao et al. 2014) using the following parameters: `-p -countReadPairs` for paired-end experiments and default parameters for single-end experiments. We then intersected the highest SNR replicate MACS2 called

peaks with the available H3K27ac peaks for the same tissue (when multiple replicates of H3K27ac ChIP-seq experiments were available, we used only those H3K27ac peaks represented in more than half of the replicates). We then calculated the variation in the position of the summit of the same element in different replicates by generating a summit confidence interval (see [supplementary fig. S2, Supplementary Material](#) online) by using a custom-made R script. DNase-H3K27ac peaks only present in the highest SNR file were not included in subsequent analyses. Summits of tissues with only one DNase-seq replicate were only considered if the element overlapped H3K27ac peaks in that same tissue. By overlapping summit confidence intervals between tissues, we created “summit clusters” (see [supplementary fig. S2, Supplementary Material](#) online). Since we did not limit the size of summit confidence intervals, it may be argued that differences in the length of confidence intervals between context-specific and pleiotropic enhancers may bias the demarcation of enhancers (the formation of summit clusters). However, when we set various limits for the size of summit confidence intervals we do not see differences in our subsequent analyses (data not shown). We created consensus elements by merging all accessible elements (MACS2 peaks) that contribute to each summit cluster (see [supplementary fig. S2, Supplementary Material](#) online). Consensus elements overlapping by more than 80% of the sequence were merged. The 37 ENCODE tissues were mapped to 19 UBERON organs and one cell ontology category (Lymphocyte; CL:0000542) ([supplementary table S2, Supplementary Material](#) online). Enhancers active in an organ are consensus elements predicted to be active in at least one of the tissues that comprise that organ. With these data we constructed a binary matrix in which rows represent consensus elements and columns represent organs. Pleiotropic enhancers were defined as consensus elements predicted to be active in 2 or more organs. The code generated to create the consensus elements from DNase-seq and H3K27ac ChIP-seq data can be found at https://github.com/frankel-lab/human_pleiotropic_enhancers. We eliminated from this set of consensus elements those that overlap with regions that are in the ENCODE blacklist of the hg38 assembly (<https://github.com/Boyle-Lab/Blacklist>), those that are located at less than 500 bp from annotated transcription start sites and those that overlap exons (including putative enhancers that overlap exons does not affect the analyses, data not shown). The final set of putative enhancers consisted of 334,189 elements.

Comparisons between FANTOM5, ENCODE, and ABC Enhancer Sets

The ABC enhancer set was downloaded from <https://www.engreitzlab.org/resources>. We mapped the biological samples to the corresponding UBERON organs ([supplementary](#)

table S6, Supplementary Material online) in order to compare with FANTOM5 and ENCODE enhancers. ABC enhancer regions overlapping by more than 80% were merged with bedtools (Quinlan and Hall 2010). In total, we used 176,220 ABC elements. FANTOM5 and ABC enhancers were lifted to the hg38 human genome assembly with UCSC liftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) for comparing with the coordinates of the ENCODE enhancers. The overlap fraction was calculated by intersecting enhancers active in each organ (comparing two enhancers sets) and then dividing by the size of the smaller set (supplementary table S7, Supplementary Material online).

Calculation of the Percentage of Pleiotropic Enhancers

To evaluate how the proportion of pleiotropic enhancers changes with the number of organs considered we took all possible combinations of organs from 2 to n , where n is the total number of organs (20 for the ENCODE enhancers and 41 for the FANTOM5 enhancers) and calculated the proportion of enhancers active in two or more organs. For k organs, there exists n choose k combinations of organs. For k values with more than 2,000 combinations, we estimated the distribution of the proportion of pleiotropic enhancers with 2,000 randomly chosen sets of size k (without repetition).

We calculated the proportion of pleiotropic enhancers (cCRE-dELS) for 17 cell types (supplementary table S2, Supplementary Material online). The ENCODE cCRE-dELS ($n = 290,876$) in these cells were defined by Moore et al (2020). We explicitly excluded from the analysis data stemming from cancer derived cell lines. We also calculated the proportion of pleiotropic enhancers in 38 cell types (69 primary cell cultures were combined to form 38 different cell types, see supplementary table S1, Supplementary Material online) (Andersson et al. 2014). For comparing the degree of pleiotropy of enhancers in cell types and organs we analyzed the 11,400 enhancers that were detected as active in both cell types and organs (Andersson et al. 2014).

Transcriptome Similarity and Enhancer Usage in FANTOM5 Cell Lines

To estimate similarities in expression patterns (a proxy to ontogenetic proximity) between cell types we used the euclidean genetic distance. We downloaded a TPM normalized RNA-seq matrix for all FANTOM5 cell types from https://fantom.gsc.riken.jp/5/datafiles/latest/extra/gene_level_expression/. We mapped closely related FANTOM5 cell types into 38 final cell types as previously done (see supplementary table S1, Supplementary Material online). The median TPMs of the cell types that comprise the final cell type were used as expression levels. For all pairs of final cell types we computed the euclidean distance by log

transforming the TPM values and adding a pseudocount of $1e-5$ to avoid taking the logarithm of zero. We then computed a distance matrix with the `dist` function of R. Hierarchical clustering using this matrix resulted in a biologically meaningful grouping. For comparing enhancer usage between final cell types we used the overlap coefficient, defined as the number of common active enhancers between two cell types divided by the number of active enhancers in the smallest set.

Correlation between the Proportion of Organ-Specific Enhancers and Organ-Specific TSSs/Transcripts

For FANTOM5 enhancers, we used the same CAGE-seq libraries that were used to identify enhancers to estimate the proportion of context-specific TSSs (reads that correspond to first exons of genes). We used the RefSeq curated gene set (`ncbiRefSeqCurated`, assembly hg19) from the UCSC site (<https://genome.ucsc.edu/>). We extended each TSS 100 bp in both directions and merged overlapping elements that were on the same DNA strand with bedtools (Quinlan and Hall 2010) (we obtained similar results without merging, data not shown). We obtained a total of 25,797 extended TSSs. For each biosample, we counted CAGE-seq reads overlapping extended TSSs in a strand-specific manner with `featureCounts` and compared the counts of each TSS with the count distribution of 1,477,250 random genomic elements of the same size (we excluded exons and TSSs from GRCh38 and all the ENCODE and FANTOM5 enhancers from this random genomic elements) generated with bedtools. We computed an empirical P value for each TSS in each biosample as the fraction of random elements with the same or greater number of reads than the TSS in that biosample. We corrected these empirical P values with the Benjamini-Hochberg procedure and considered that a TSS is active in a biosample if its corrected P value was 0.01 or less (we obtained similar results using thresholds of $P < 0.005$ or $P < 0.001$; data not shown).

For ENCODE enhancers we used ENCODE RNA-seq data available for 25 of the 37 tissues that were used to create the enhancer set (supplementary table S8, Supplementary Material online). We downloaded FPKM normalized counts and only kept genes with protein products annotated by the HUGO Gene Nomenclature Committee (19,167 genes). When replicates of the same tissues were available, we used the median FPKM for each gene. To use a uniform threshold across tissues to decide when a gene is expressed or not, we used the `zFPKM` R package (Hart et al. 2013) to normalize FPKM. Genes were considered to be expressed if their `zFPKM` was above -2 (we obtained similar results using -3.5 , -3 , -2.5 , -1.0 and 1 as thresholds; data not shown). We regard a gene as context-specific when the gene is expressed in a single organ (for both the FANTOM5 and ENCODE analyses).

Gene Pleiotropy and Enhancer Pleiotropy

We used the ABC predictions of enhancer-gene links (Fulco et al. 2019). The ABC model assigns a score to enhancer-gene pairs for different organs and cell types, and it is known to outperform other predictions, such as distance to the TSS, in this task. Thus, to find the most likely target gene for a given enhancer, we took the intersection of ENCODE and ABC enhancers in each organ (enhancers that overlapped by at least 95% of their extension) and considered that the target gene of an enhancer is the gene with the maximum ABC score in each organ. ABC organs were mapped to UBERON terms as previously described (supplementary table S6, Supplementary Material online). To test for the statistical significance of the association between context-specific enhancers and context-specific genes we generated a 2×2 contingency table with the number of interactions between context-specific/pleiotropic enhancers and context-specific/pleiotropic genes. The genes that have a zFPKM value > -2 in a single organ are considered as context-specific genes.

RegulomeDB and cis-eQTL SNP Analyses

SNPs with regulatory potential were downloaded from the RegulomeDB site (<https://regulomedb.org/regulome-search/>). We lifted the hg19 positions to the hg38 assembly using UCSC liftOver. For this analysis we considered SNPs in Category 1 (SNPs that are eQTLs and that have at least one additional source of evidence for being regulatory) and the most stringent category of RegulomeDB, which is category 1a (Boyle et al. 2012). Odds ratios were calculated for each pleiotropy category as the RegulomeDB SNPs contained in the category divided by the non-RegulomeDB SNPs (1,000 Genomes SNPs with a MAF > 0.05 in Europeans) in that same category divided by the same ratio for the rest of the genome. In total we considered 71,263 SNPs in Category 1 and 978 SNPs in Category 1a, and 7,894,464 common variants. All SNP count operations were performed with bedmap (Neph et al. 2012). Fine-mapped cis-eQTLs were downloaded from the GTEx Portal (GTEx Analysis V8, <https://gtexportal.org/home/>). We used 95% credible set variants with a posterior inclusion probability ≥ 0.1 in the HighConfidenceVariants file. GTEx biosamples were mapped to 18 organs (of the 20 ENCODE organs) using UBERON ontology (supplementary table S9, Supplementary Material online). We obtained similar results when standardizing for enhancer size (enhancer size was fixed to 500 bp).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Justin Crocker, Hernan Dopazo, and members of the Frankel lab for helpful comments on the manuscript. This work was supported by grants PICT 2016-0414 and PICT 2019-00497 (ANPCyT, Argentina) to N.F.

Data Availability

The data underlying this article are available in the article and in its online [supplementary material](#).

Literature Cited

- Andersson R, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507:455–461.
- Boyle AP, et al. 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22:1790–1797.
- Corradin O, Scacheri PC. 2014. Enhancer variants: evaluating functions in common disease. *Genome Med.* 6:85.
- Creyghton MP, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 107:21931–21936.
- Djureinovic D, et al. 2014. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol Hum Reprod.* 20:476–488.
- Dunham I, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.
- Fish A, Chen L, Capra JA. 2017. Gene regulatory enhancers with evolutionarily conserved activity are more pleiotropic than those with species-specific activity. *Genome Biol Evol.* 9:2615–2625.
- Fong SL, Capra JA. 2021. Modeling the evolutionary architectures of transcribed human enhancer sequences reveals distinct origins, functions, and associations with human trait variation. *Mol Biol Evol.* 38:3681–3696.
- Fulco CP, et al. 2019. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat Genet.* 51:1664–1669.
- Fuqua T, et al. 2020. Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* 587:235–239.
- Gasparini M, Tome JM, Shendure J. 2020. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet.* 21:292–310.
- Hart T, Komori H, LaMere S, Podshivalova K, Salomon DR. 2013. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics.* 14:778.
- Heintzman ND, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459:108–112.
- Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet.* 22:203–215.
- Hiller M, Schaar B, Bejerano G. 2012. Hundreds of conserved non-coding genomic regions are independently lost in mammals. *Nucleic Acids Res.* 40:11463–11476.
- Infante CR, et al. 2015. Shared enhancer activity in the limbs and phallus and functional divergence of a limb-genital cis-regulatory element in snakes. *Dev Cell.* 35:107–119.
- Kimura H. 2013. Histone modifications for human epigenome analysis. *J Hum Genet.* 58:439–445.
- Kittelmann S, Preger-Ben Noon E, McGregor AP, Frankel N. 2021. A complex gene regulatory architecture underlies the development

- and evolution of cuticle morphology in *Drosophila*. *Curr Opin Genet Dev*. 69:21–27.
- Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*. 20:207–220.
- Kristjánsdóttir K, Dziubek A, Kang HM, Kwak H. 2020. Population-scale study of eRNA transcription reveals bipartite functional enhancer architecture. *Nat Commun*. 11:5963.
- Kvon EZ, et al. 2020. Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. *Cell* 180: 1262–1271.e15.
- Le Poul Y, et al. 2020. Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Sci Adv*. 6:eabe2955.
- Levine M. 2010. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 20:R754–R763.
- Lewis JJ, et al. 2019. Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proc Natl Acad Sci U S A*. 116:24174–24183.
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.
- Lidschreiber K, et al. 2021. Transcriptionally active enhancers in human cancer cells. *Mol Syst Biol*. 17:e9873.
- Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D. 2014. Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* 346:1004–1006.
- Lonsdale J, et al. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 45:580–585.
- Minnoye L, et al. 2021. Chromatin accessibility profiling methods. *Nat Rev Dis Primers*. 1:10.
- Moore JE, et al. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583:699–710.
- Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, Haendel MA. 2012. Uberon, an integrative multi-species anatomy ontology. *Genome Biol*. 13:R5.
- Nasser J, et al. 2021. Genome-wide enhancer maps link risk variants to disease genes. *Nature* 593:238–243.
- Neph S, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28:1919–1920.
- Panigrahi A, O'Malley BW. 2021. Mechanisms of enhancer action: the known and the unknown. *Genome Biol*. 22:108.
- Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. 2013. Enhancers: five essential questions. *Nat Rev Genet*. 14: 288–295.
- Pineau C, et al. 2019. Cell type-specific expression of testis elevated genes based on transcriptomics and antibody-based proteomics. *J. Proteome Res*. 18:4215–4230.
- Preger-Ben Noon E, et al. 2018. Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function. *Cell Rep*. 22:3021–3031.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- Rada-Iglesias A, et al. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279–285.
- Radke DW, et al. 2021. Purifying selection on noncoding deletions of human regulatory loci detected using their cellular pleiotropy. *Genome Res*. 31:935–946.
- Reilly SK, et al. 2021. Direct characterization of cis-regulatory elements and functional dissection of complex genetic associations using HCR–FlowFISH. *Nat Genet*. 53:1166–1176.
- Sabaris G, Laiker I, Preger-Ben Noon E, Frankel N. 2019. Actors with multiple roles: pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet*. 35:423–433.
- Sartorelli V, Lauberth SM. 2020. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol*. 27:521–528.
- Singh D, Yi S. 2021. Enhancer pleiotropy, gene expression, and the architecture of human enhancer-gene interactions. *Mol Biol Evol* 38:3898–3909.
- Smith E, Shilatifard A. 2014. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol*. 21:210–219.
- Sobreira DR, et al. 2021. Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5. *Science* 372: 1085–1091.
- Watanabe K, et al. 2019. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 51:1339–1348.
- Xin Y, et al. 2020. Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility. *Proc Natl Acad Sci U S A*. 117: 20636–20644.
- Zhang Y, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 9:R137.
- Zhu J, et al. 2016. Identification of tissue-specific protein-coding and noncoding transcripts across 14 human tissues using RNA-seq. *Sci Rep*. 6:28400.

Associate editor: Selene Fernández Valverde