Enhancers with tissue-specific activity are enriched in intronic regions

Beatrice Borsari[*1], Pablo Villegas-Mirón[*2], Sílvia Pérez-Lluch[1], Isabel Turpin[2], Hafid Laayouni[2,4], Alba Segarra-Casas[2], Jaume Bertranpetit[2], Roderic Guigó[1,3] and Sandra Acosta [Ŧ2,5]

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona, 08003, Catalonia, Spain

[2]Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Dr. Aiguader 88, Barcelona 08003, Catalonia, Spain

[3]Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, Barcelona 08003, Catalonia, Spain

[4]Bioinformatic Studies, ESCI-UPF, Pujades 1, 08003, Barcelona, Spain

[5]Dpt. Pathology and Experimental Therapeutics, Medical School, University of Barcelona, Feixa Llarga, 08907, L'Hospitalet de Llobregat, Barcelona, Spain

Running title:

Intronic enhancers lead tissue-specific regulation

Keywords:

enhancers, introns, gene regulation, tissue function, tissue patterning

[*]Beatrice Borsari and Pablo Villegas-Mirón contributed equally to this work.

[Ŧ]Correspondence should be addressed to E-mail: sandra.acosta@upf.edu; sandra.acosta@ub.edu (Sandra Acosta).

Abstract

Tissue function and homeostasis reflect the gene expression signature by which the combination of ubiquitous and tissue-specific genes contribute to the tissue maintenance and stimuli-responsive function. Enhancers are central to control this tissue-specific gene expression pattern. Here, we explore the correlation between the genomic location of enhancers and their role in tissue-specific gene expression. We found that enhancers showing tissue-specific activity are highly enriched in intronic regions and regulate the expression of genes involved in tissue-specific functions, while housekeeping genes are more often controlled by intergenic enhancers, common to many tissues. Notably, an intergenic-to-intronic active enhancers continuum is observed in the transition from developmental to adult stages: the most differentiated tissues present higher rates of intronic enhancers, while the lowest rates are observed in embryonic stem cells. Altogether, our results suggest that the genomic location of active enhancers is key for the tissue-specific control of gene expression.

Introduction

Multiple layers of molecular and cellular events tightly control the level, time and spatial distribution of expression of a particular gene. This wide range of mechanisms, known as gene regulation, defines tissue-specific gene expression signatures (Melé et al., 2015), which account for all the processes controlling the tissue function and maintenance, namely tissue homeostasis. Both the level and spatio-temporal pattern of expression of a gene are determined by a combination of regulatory elements (REs) controlling its transcriptional activation. Most genes contributing to tissue-specific expression signatures are actively transcribed in more than one tissue, but at different levels and with distinct patterns of expression in time and space, suggesting that the regulation of these genes is different across tissues. Nevertheless, approximately 10-20% of all genes are ubiquitously expressed

(*housekeeping genes*), and they are involved in basic cell maintenance functions (Pervouchine et al., 2015; Zabidi et al., 2015; Eisenberg and Levanon, 2013).

cis-REs (CREs) are distributed across the whole genome, and their histone signature correlates with the transcriptional control they exert over their target genes (Chen et al., 2019; Hawkins et al., 2010; Choukrallah et al., 2015). The activation of CREs depends on several epigenetic features, including combinations of different transcription factors' binding sites, and it is positively correlated with the H3K27ac histone modification signal (Heinz et al., 2015; Heintzman et al., 2007). Epigenetic features in specific tissues may change throughout the life-span of individuals. During development, embryos undergo morphological and functional changes. These changes shape cell fate and identity as a result of tightly regulated transcriptional programs, which in turn are intimately associated with CREs' activity and chromatin dynamics (Shlyueva et al., 2014; Bonev et al., 2017; Rand and Cedar, 2003; Gilbert et al., 2003).

Many key CREs known to regulate gene expression have been reported to locate in introns of their target genes (Ott et al., 2009; Kawase et al., 2011). However, it is unknown whether this is either a sporadic feature associated with certain types of genes - for instance long genes, such as *HBB* (also known as beta-globin) (Gillies et al., 1983) or *CFTR* (Ott et al., 2009) -, a common regulatory mechanism to most genes (Khandekar et al., 2007; Levine, 2010), or a pattern of biological significance. To delve into this question, we analyzed the genomic location of CREs across a panel of 70 adult and embryonic human cell types available from the Encyclopedia of DNA Elements (ENCODE) Project (The ENCODE Project Consortium, 2020).

Results

Enhancer-like regulatory elements define tissue-specific signatures

3

We leveraged the cell type-agnostic registry of candidate *cis*-Regulatory Elements (cCREs) generated for the human genome (hg19) by the ENCODE Project. We focused on the set of 991,173 cCREs classified as Enhancer-Like Signatures (ELSs), defined as DNase I hypersensitive sites supported by the H3K27ac epigenetic signal, and assessed their presence-absence patterns across 43 adult cell type-specific catalogues (Supplementary Table 1; see Methods). We first explored the data with multidimensional scaling (MDS), which uncovered tissue-specific presence-absence patterns (Supplementary Fig. 1A). Indeed, the separation of samples driven by ELSs' activity is comparable to the one obtained from the analysis of Genotype-Tissue Expression (GTEx) data (Melé et al., 2015), with blood and brain as the most diverging tissues. This suggests a correlation between gene regulatory mechanisms orchestrated by ELSs and tissue-specific gene expression patterns, which has been previously described (Pennacchio et al., 2007; Ernst et al., 2011).

We observed that the proportion of active ELSs located in intergenic regions increases with the number of samples in which ELSs are active (Fig. 1A), suggesting an unexpected role for the genomic location of ELSs. Thus, to untangle the relationship between the genomic location and cell-type specificity of ELSs, we selected a subset of 33 samples that formed 9 main tissue groups, supported by both hierarchical clustering and MDS proximity: brain, iPSCs, blood, digestive system, intestinal mucosa, fibro/myoblasts, aorta, skeletal/cardiac muscle and smooth muscle (Figs. 1B-C; Supplementary Table 1, *Samples' Cluster*). Tissues represented by only one sample (ovary, thyroid gland, lung, esophagus, spleen), or samples that do not cluster consistently with their tissue of origin and function (endocrine pancreas, liver, right lobe of liver, gastrocnemius medialis, bipolar neuron), were not included in the subsequent analyses (Supplementary Table 1; see Methods).

The fact that tissue-specific enhancer signatures contribute to the *ad hoc* tissues' functional clustering suggests a direct link between ELSs' activity and the regulation of tissue-specific

functions (Fig. 1C). Thus, we set out to characterize tissue-specific enhancer signatures and to compare them with regulatory mechanisms that are common, i.e. shared among most tissues. Tissue-specific ELSs (Ts ELSs) were defined as those ELSs active in ≥ 80\% of the samples within a given cluster and in no more than one sample outside the cluster (Supplementary Table 2; see Methods). For clusters with limited sample number (≤ 3), we required Ts ELSs to be active exclusively within the corresponding tissue cluster (see Methods). The overlap of Ts ELSs with samples from other clusters (Fig. 1D) is consistent with the samples' MDS proximity observed in Fig. 1C, suggesting a functional relevance of the genes regulated by shared ELSs. In addition, we identified a set of 555 ELSs active in 95% of the 33 samples, herein named as common ELSs (Supplementary Table 2).

The genomic location of regulatory elements correlates with their tissue-homeostatic functions

We next explored the genomic location of the sets of common and Ts ELSs. While common ELSs are preferentially located in intergenic regions (58%, Fig. 2A), the majority of aorta, muscle- and brain-specific ELSs fall inside introns (between 63 and 74%; Fig. 2A). These significant differences in genomic distribution between tissue-specific and common regulatory elements (Supplementary Table 3) are consistent with our initial observation of a high sharing rate of intergenic ELSs across samples (Fig. 1A). In contrast, the iPSCs, fibro/myoblasts, mucosa, digestive and blood clusters - which comprise undifferentiated, non-specialized, highly proliferative or more heterogeneous cell types - show a more even distribution of Ts ELSs between intergenic and intronic regions (Fig. 2A). Overall, we observed a limited abundance of exonic ELSs (Fig. 2A, Supplementary Tables 3 and 4).

Genes harboring Ts ELSs may present distinctive features, including differences in gene and intron length. To rule out any bias in our analyses, we compared these features between genes hosting common and Ts ELSs. While the number of introns per hosting gene is comparable

across groups (Kruskal-Wallis *p*-value test = 0.08), we reported significant differences in gene and median intron length amongst tissues (Kruskal-Wallis *p*-value test < 2.2×10^{-16}; Supplementary Fig. 1B). Nevertheless, we did not observe a correlation between such differences and the presence of intronic ELSs (Supplementary Fig. 1B). Across all tissues, most of the intronic Ts ELSs are located further than 5 kb from annotated TSSs (Supplementary Fig. 1C), and do not show chromatin marking typical of promoters (see Methods section "Tissue-specific and common ELSs").

We subsequently explored whether the genes harboring tissue-specific intronic ELSs perform functions associated with maintenance of tissue homeostasis and response to stimuli. Indeed, the enrichment of Gene Ontology (GO) terms associated with tissue-specific cellular components is consistent with the ELSs' tissue identity (Supplementary Table 5). For instance, genes hosting brain-specific ELSs perform functions associated with synapses and axons, while in the case of muscle and blood we found significant terms related to sarcolemma, actin cytoskeleton and contractile fibers, and immunological synapses and cell membranes, respectively. Conversely, genes harboring common ELSs reported terms related to ordinary cell functions and membrane composition (Supplementary Table 5). Although this suggests an implication of intronic ELSs in tissue-specific functions, likely through tissue-specific gene regulation mechanisms, there is no proven evidence of intronic ELSs being direct regulators of their host genes. To identify genes targeted by Ts ELSs, we integrated our ELS analysis with the catalogue of expression Quantitative Trait Loci (eQTLs) provided by the Genotype-Tissue Expression (GTEx) Project (The GTEx Consortium, 2017). eQTLs provide functional information about the changes of expression associated with human variants. We leveraged eQTLs located in both intronic and intergenic ELSs to identify their target genes. Among the 48,555 common and Ts ELSs, 6,349 overlap with a significantly associated eQTL-eGene pair, hereafter referred to as eQTL-ELSs. The proportion of eQTL-ELSs is similar among the tissue samples represented in the GTEx sampling collection, ranging between 10 and 25% (Fig. 2B). In all annotated tissues, gene regulation driven by eQTL-ELSs occurs predominantly in the

tissue where the ELS is specifically active (Fig. 2C). In line with the above-mentioned results (Fig. 2A), highly specialized tissues such as brain and muscle show the highest proportion of intronic *vs.* intergenic ELSs hosting eQTLs detected in the corresponding tissue (Figs. 2B-C). Conversely, common eQTL-ELSs are more frequently located in intergenic elements (32% *vs.* 62%) (Fig. 2C). GO enrichment analysis on the sets of target genes associated with intronic and intergenic eQTL-ELSs shows a clear prevalence of tissue-specific terms for those genes targeted by intronic rather than intergenic eQTL-ELSs - for instance, skeletal/cardiac muscle: carbohydrate and amino acid metabolism; brain: cell projection and microtubule cytoskeleton organization (Supplementary Table 6). In contrast, genes associated with common eQTL-ELSs (either intronic or intergenic) do not show any significantly enriched term. Altogether, these results suggest that intronic eQTL-ELSs are involved in the regulation of genes associated with tissue-specific functions, while intergenic ELSs are more devoted to tissue homeostatic processes.

Target genes of intronic ELSs identified by Hi-C regulate tissue-specific functions

The interaction between ELSs and promoters is central for the onset of gene expression. These types of interactions are defined in each tissue, and can be identified genome-wide through Hi-C-seq. Here, we explored ELS-promoter interactions reported by published Hi-C datasets in relevant tissues, identifying Ts ELS-target genes, and thus improving the annotations of ELSs-target genes with respect to the eQTL analysis (Figs. 2D-E) (Jung et al., 2019; Lu et al., 2020; Mifsud et al., 2015). This approach allowed us to observe that most of the target genes are regulated by multiple ELSs (Supplementary Fig. 2A). As in the case of eQTL-ELSs, brain and muscle show the highest proportion of intronic *vs.* intergenic ELSs intersecting Hi-C interacting fragments in the corresponding tissue, while common Hi-C-ELSs are enriched in intergenic regions (Fig. 2E). The GO enrichment analysis reported an increase in relevant terms involved in tissue-specific functional roles as well. Of note, intronic Hi-C-ELSs show stronger enrichment in tissue-specific terms (skeletal/cardiac muscle: I band and

Z disc components; brain: pre/postsynaptic assembly and organization; aorta: regulation of smooth muscle cell migration and proliferation), while we observed broader functionality from intergenic ELSs' interactions (brain: choline catabolic process and copper ion homeostasis, amongst others) (Supplementary Table 7). Moreover, common Hi-C-ELSs appear to target genes that are enriched in housekeeping functions, such as cell adhesion and nucleosome organization (Supplementary Table 7). Overall, these results on ELS-promoter interactions further support the fact that intronic ELSs regulate genes controlling tissue-specific functions, while intergenic ELSs are more devoted to tissue homeostatic processes.

Intronic ELSs regulate the expression of hosting and non-hosting genes

Next, we wanted to understand the relationship between tissue-specific intronic ELSs and their host genes. To do so, we analyzed the expression patterns of genes targeted by Hi-C ELSs, as a proxy for direct regulation. The proportion of intronic Hi-C-ELSs targeting their host genes is comparable among most groups of samples and ranges between 45 and 65%, with the exception of muscle and blood which show lower values (Fig. 3A). We compared the expression patterns of Hi-C-ELSs' target genes, considering the type of ELS regulating them (*intergenic, intronic host* - i.e. an ELS targeting its host gene -, *intronic non-host* - i.e. an ELS targeting a gene that is not its host gene). Genes regulated by intronic host ELSs exhibit expression patterns that better recapitulate tissue identity (the relevant tissue clusters, in almost all the cases, separately from the other groups), while hierarchical clustering of genes regulated by intronic non-host ELSs does not efficiently discriminate tissue-specific patterns (Fig. 3B and Supplementary Fig. 2B). Genes regulated by intronic host ELSs are associated with tissue-specific functions (Supplementary Table 8), in particular synaptic vesicle clustering and active zone organization for brain (e.g. *PCDH17*), regulation of cell division and establishment of cell polarity for fibroblasts (e.g. *TGFB2*), cardiac myofibril assembly and muscle fiber development for skeletal/cardiac muscle (e.g. *MEF2A*), and regulation of smooth muscle cell migration for aorta (e.g. *DOCK5*). On the contrary, those genes targeted by intronic

non-host ELSs are involved in homeostatic functions not uniquely associated with the relevant tissue, suggesting that they are not expressed in a tissue-specific manner, but are nevertheless regulated by tissue-specific enhancers. For instance, brain and aorta present significant terms related to protein monoubiquitination (e.g. *PDCD6*), and cellular response to endogenous stimulus (e.g. *TNC*), respectively (Fig. 3C). Overall, this indicates that the intronic location of regulatory elements cannot be associated exclusively with the regulation of the host gene. Furthermore, the identification of a large proportion of intronic non-host ELSs suggests that the intronic location may be, in a particular tissue, advantageous for the establishment and maintenance of gene expression programs, including non-tissue-specific events.

The enrichment of transcription factor binding sites in Ts ELSs is independent of their genomic location

The activation of ELSs is a dynamic process depending, amongst other factors, on its accessible chromatin to be bound by transcription factors (TFs). Thus, tissue-specific gene expression programs may be controlled by the underlying signature of TFs-ELSs pairing (Schmitt et al., 2016). We next wondered whether the specific distribution of ELSs, i.e. intronic *vs.* intergenic, is associated with a different transcription factor binding site (TFBS) signature that could account for their tissue-specific activity. To this purpose we explored, with HOMER (Heinz et al., 2010), the enrichment of TFBSs independently for intronic and intergenic ELSs. Indeed, a distinct TFBS signature for each tissue in both intronic and intergenic ELSs can be observed (Fig. 4A), supporting our previous results that Ts ELSs significantly contribute to the regulation of tissue-specific functions. The number of enriched TFBSs in intronic regions is higher in highly specialized tissues such as brain and muscle, and shows no overlap with TFBSs found in intergenic ELSs. The opposite picture is observed in common ELSs, with higher enrichment of TFBSs in intergenic ELSs. An intermediate pattern is observed for highly proliferative tissues such as iPSCs, fibroblasts, mucosa and blood, in which the amount of enriched TFBSs is similar between intronic and intergenic ELSs (Fig. 4A and Supplementary

Table 9). Amongst the TFBSs enriched in Ts intronic and intergenic ELSs, we found well-known TFs associated with tissue-specific homeostatic events, such as RUNX2 in blood controlling adult endothelial hemogenesis (Lis et al., 2017), and SOX4 and SOX8 in brain controlling adult neural differentiation (Chen et al., 2015). POU5F1 (previously known as OCT4) is required for iPSCs. Still, with the exception of those TFs associated with enriched TFBSs in iPSCs, most other TFs are widely expressed across tissues (Fig. 4B). This distinct iPSCs' TF-ELS binding potential is supported by previous data indicating that iPSCs share their epigenetic signature with early developmental stages rather than with the original tissue prior to reprogramming. Overall, the TFBSs enrichment, rather than the TFs' gene expression patterns, is the most variable feature between intronic and intergenic ELSs and amongst tissues.

Dynamic location of ELSs throughout embryonic development and maturation

Throughout embryonic development, tissues mature to fully reach their functional capacity in adulthood, giving rise to several tissue-specific homeostatic features that vary among different tissues. For instance, blood comprises a wide number of cell types characterized by heterogeneous functions and high turnover. On the opposite side, we find highly specialized tissues such as muscle, that are formed by fewer cell types, mainly dedicated to the same function and with limited cell division capacity. During development, tissues share features of basic homeostasis, proliferation and plasticity, but they are also already patterned to perform their adult functions. Still, whether the regulatory features of a given adult tissue are reminiscent of their developmental lineage remains largely unknown. To answer this question, we assessed the activity and the intronic location of the 991,173 cell type-agnostic ELSs across 27 embryonic samples (Fig. 5A and Supplementary Table 10). MDS analysis highlighted three main groups of embryonic samples: stem cells (ESCs), neural progenitors, and a larger group of more differentiated cell types (Fig. 5B; Supplementary Table 10, *Samples' Group*). The three groups of samples are associated with 3,112, 784 and 1,166

specific ELSs, respectively (Supplementary Table 11). Although the majority of these ELSs are active only within the corresponding cluster, we reported that 26% of the neural progenitors-specific ELSs are also active in one ESCs sample (Supplementary Fig. 3A). On the contrary, we identified only 94 ELSs common to all embryonic samples (Supplementary Table 11). The proportion of specific intronic ELSs is higher for neural progenitors and differentiated tissues, compared to ESC-specific and common ELSs (Fig. 5C), but lower with respect to clusters of adult tissues such as aorta, muscle and brain (Fig. 2A). As in the case of adult samples, we observed a limited abundance of exonic ELSs (Fig. 5C, Supplementary Tables 12-13), while we could not find significant associations between the frequency of group-specific intronic ELSs and features of gene and intron length (Supplementary Fig. 3B). As for adult samples, most of these group-specific intronic ELSs are located further than 5 kb from annotated TSSs (Supplementary Fig. 3C).

Next, we wanted to validate the dynamics of intronic *vs.* intergenic ELSs active throughout development, using brain development as a paradigm (Supplementary Fig. 4A). To this purpose, we identified active ELSs (ChIP-seq H3K27ac+/H3K4me3- peaks) in human ESCs, and hESC-derived NPCs and neurons, and assessed their degree of overlap with ENCODE ELSs. Active ELSs identified by ChIP-seq in ESCs, NPCs and neurons overlap with ENCODE ELSs specific to ESCs (86%), embryonic neural progenitors (40%) and adult brain (53%) samples, respectively. In particular, the proportion of active intronic ELSs increases with the degree of differentiation of the samples (55% in ESCs, 64% in NPCs and 68% in neurons) (Fig. 5D), validating the observed correlation between active Ts ELSs and their intronic location. We observed a high overlap (86% to 98%) between ENCODE common embryonic ELSs and H3K27ac ChIP-seq peaks detected during the hESC-differentiation, including known ELSs for housekeeping genes, such as *ACTB* (Supplementary Fig. 4B). The expression of genes regulated by individual candidate ELSs (Supplementary Figs. 4C-E) is, in most of cases, consistent with the activity of the ELS, being active either in a tissue-specific manner in ESCs or neurons, or in all three differentiation stages (Supplementary Table 14).

Although a small fraction of common ELSs is marked by H3K4me3 in ESCs, NPCs and neurons (Supplementary Fig. 4B), the corresponding H3K4me3 signal is comparatively lower than the H3K4me3 level observed at promoter regions (Supplementary Fig. 4F). When analyzing the genes hosting developmental group-specific intronic ELSs, we observed that they are enriched in functions consistent with the corresponding adult tissue (Supplementary Table 15). For instance, the ones hosting neural progenitors-specific ELSs are enriched in neural development-related terms, such as axonogenesis and dendritic spine organization. On the contrary, genes hosting developmental common ELSs are enriched in protein complexes like nBAF and SWI/SNF, known developmental chromatin remodelers (Alver et al., 2017).

Lastly, in an attempt to define the amount of regulatory activity shared by embryonic and adult samples as an indicator of the reminiscent embryonic function in adult tissue homeostasis, we computed, for specific and common embryonic ELSs, the number of adult tissues in which they are found active. As expected, whereas ELSs specific to ESCs and neural progenitors are active in a limited set of adult samples, embryonic differentiated tissues report a higher degree of shared regulatory activity with adult cell types (Supplementary Fig. 5). Moreover, ELSs active in all embryonic samples (common) are also active in the majority of adult samples. Overall, these results show that the genomic location of ELSs is dynamic throughout development, and shifts towards an intronic localization during tissue maturation.

Discussion

In this study, we show the central role of intronic Enhancer-Like Signatures (ELSs) in the control of tissue-specific expression programs. Since Heitz described in 1928 (Heitz, 1928) euchromatin as transcription-permissive chromosomal regions enriched in genes, and heterochromatin as inactive or passive chromatin regions, this dual definition has been shaped

throughout the years but it still remains vastly correct (De Laat and Duboule, 2013; DeMare et al., 2013; Ernst and Kellis, 2010). Intergenic regions are often regulatorily silenced, and this happens more frequently in adult than embryonic tissues (Heinz et al., 2015). The ENCODE project reports that about half of the ELSs are intergenic, and 38% are intronic (ENCODE SCREEN Portal: https://screen-v10.wenglab.org/, section "About"). In our study, we describe an enrichment of intronic ELSs in the most specialized tissues. These elements regulate genes involved in tissue-specific functions, suggesting an important role for the genomic location of ELSs. On the contrary, in less specialized adult tissues and embryonic samples, ELSs are less frequently found in intronic elements, suggesting that the maturation and tissue commitment correlates with the ELSs' distribution across the whole genome. One could hypothesize that the enriched presence of intronic ELSs is advantageous for the control of the gene expression signature of a particular tissue, for instance granting ELSs accessibility in open DNA regions (genes) and avoiding their leaky activity. In line with this, active transcription and nascent RNA have been recently associated with the maintenance of open chromatin (Hilbert et al., 2021), a process that can be advantageous to the presence of intronic ELSs in actively transcribed genes. Introns have long been observed as gene expression regulators through different mechanisms (Rose, 2019; Chorev and Carmel, 2012; Shaul, 2017). Specifically, introns' regulatory potential has been associated with the regulation of the host gene's expression in several different ways, often related to alternative splicing, intron retention (Jacob and Smith, 2017), non-sense mediated decay (Lewis et al., 2003), and to the control of transcription initiation via recruitment of RNA Polymerase II, likely as alternative promoters (Bieberstein et al., 2012; Kowalczyk et al., 2012). However, here we found that, in most tissues, about half of the ELSs located in introns do not regulate the expression of the host gene, but of genes involved in important tissue homeostatic functions, whose expression is not restricted to that particular tissue. This is important regulatory information, since it disentangles the presence of intronic ELSs from the regulation of the host gene, opening new opportunities to identify the regulatory mechanisms controlling tissue-specific gene expression. Overall, our results suggest that the genomic distribution of tissue-specific active

ELSs is not stochastic, and mainly overlaps with intronic elements. The opposite happens to active ELSs common to all tissues, which are instead enriched in intergenic regions. These results suggest that intronic enhancers play a role in the regulation of gene expression in a tissue-specific manner.

Methods

The ENCODE registry of candidate *cis*-Regulatory Elements

The cell type-agnostic registry of human candidate *cis*-Regulatory Elements (cCREs) available from the ENCODE portal corresponds to a subset of 1,310,152 representative DNase hypersensitivity sites (rDHSs) in the human genome with epigenetic activity further supported by histone modification (H3K4me3 and H3K27ac) or CTCF-binding data (https://screen-v10.wenglab.org/; section "About"). It comprises 991,173 Enhancer-Like Signatures (ELS), 254,880 Promoter-Like Signatures (PLS), and 64,099 CTCF-only Signatures. In addition, cell type-specific catalogues are provided for those cell types with available DNase and ChIP-seq ENCODE data.

Selection of cCREs with enhancer-like signature (ELS) across human samples

We downloaded the set of 1,310,152 cell type-agnostic cCREs for human assembly 19 (hg19) from the ENCODE SCREEN webpage (https://screen-v10.wenglab.org/; file ID: ENCFF788SJC). From the ENCODE portal (www.encodeproject.org/matrix/?type=Annotation&encyclopedia_version=ENCODE+v4&annotation_type=candidate+Cis-Regulatory+Elements&assembly=hg19), we retrieved cell type-specific registries of cCREs for 43 adult and 27 embryonic human samples with available DNase data and ChIP-seq H3K4me3 and H3K27ac data. The ENCODE File Identifiers for the

adult and embryonic datasets are reported in Supplementary Table 1 and Supplementary Table 8, respectively. No significant changes are expected upon realignment to GRCh38, since main improvements with respect to hg19 have been made in the representation of so-called alternate haplotypes, with a small impact on the definition of genic and intergenic regions (Church et al., 2015). We focused on the 991,173 cell type-agnostic cCREs with ELS activity, and generated a binary table in which we assessed, for a given cCRE, the presence/absence of ELS activity annotation (column 9 = "255, 205, 0") in each of the 43 adult and 27 embryonic samples. A binary distance matrix between all pairs of adult samples was used to perform multidimensional scaling (MDS) in three dimensions. This resulted in the selection of 33 adult samples, which form 9 tissue groups well supported by hierarchical clustering (Figs. 1B-C) The same procedure was applied, independently, to the embryonic samples. In this case, IMR-90, mesendoderm, mesodermal cell, endodermal cell and ectodermal cell samples were not included in subsequent analyses.

Intersection of ELSs with genes, introns, exons and intergenic regions

Genes, exons and introns' coordinates were obtained from GENCODE v19 annotation (https://www.gencodegenes.org/human/release_19.html). The overlap between ELSs and genes, exons and introns was computed using BEDTools intersectBed v2.27.1 (Quinlan and Hall, 2010). The proportions of ELSs overlapping intronic segments (Figs. 2A, 5C) also include a limited set of ELSs overlapping both intronic and exonic regions. On the other hand, we defined as exonic ELSs those intersecting exclusively exonic regions (Figs. 2A, 5C). The overlap of ELSs with intergenic regions was obtained by intersecting the former with the genes' coordinates using the BEDTools intersectBed option -v.

Tissue-specific and common ELSs

Tissue-specific ELSs are ELSs active (see Methods section *Selection of cCREs with enhancer-like signature (ELS) across human samples*) in ≥ 80\% of the samples within a given group of samples (blood = 4/5; skeletal/cardiac muscle = 3/4; smooth muscle = 3/4; brain = 6/7; stem cells = 5/6; neural progenitors = 5/6; differentiated tissues = 8/10). Because of the small sample size, we required iPSCs-, fibro/myoblasts-, digestive-, mucosa- and aorta-specific ELSs to be active in 100% of the samples (either 2/2 or 3/3). In addition, Ts ELSs are active in 0 (iPSCs, fibro/myoblasts, digestive, mucosa and aorta) or at most 1 (all other groups) outer samples (i.e. samples outside of the considered group). Common adult and embryonic ELSs are ELSs active in 95% and 100% of the samples, respectively (i.e. 31/33 and 22/22). To rule out indirect effects of ELS activity related to promoter regions, we discarded common and Ts ELSs overlapping any annotated Transcription Start Site (TSS, ± 2kb) in GENCODE v19. We further computed, for every tissue-specific intronic ELS, the minimum distance from any annotated TSS (Supplementary Figs. 1C and 3C): most of these ELSs are located more than 5 kb from TSSs. We also controlled our sets of Ts ELSs for the presence of potential alternative promoters, leveraging every adult and embryonic sample with available H3K4me3 ChIP-seq experiments from ENCODE (Supplementary Fig. 6 and Supplementary Tables 16-17). More specifically, we computed the proportion of intronic and intergenic Ts ELSs showing peaks of H3K4me3 (Supplementary Figs. 6A-B). Overall, we did not observe differences in the proportion of marked regions between intronic and intergenic ELSs. In the case of marked ELSs, we compared, for both adult and embryo samples, their aggregated H3K4me3 signal (expanding +/- 5 kb from the center of the ELS) to the signal detected at marked annotated TSSs (+/- 2 kb; for some examples, see Supplementary Figs. 6C-D, 7).

Assessing enhancer regulatory activity with GTEx eQTL-eGene significant pairs

ELSs were annotated using the GTEx v7 (The GTEx Consortium, 2017) significant variant-gene pairs from 46 different tissues (number of samples with genotype ≥ 70), available on the GTEx portal (www.gtexportal.org). Only single-tissue eQTL-eGene associations with a qval ≤

0.05 were used. Similar GTEx tissues were grouped in unique categories in order to consider the most complete catalogue of eQTL-eGene pairs per group of samples. These categories were named as follows: fibroblasts (Skin Not Sun Exposed Suprapubic, Cells Transformed Fibroblasts), blood (Whole Blood, Spleen), skeletal/cardiac muscle (Skeletal Muscle, Heart Atrial Appendage, Heart Left Ventricle), brain subregions (all brain subregions, Pituitary Gland, Nerve Tibial), Aorta (Artery Aorta), smooth muscle (Artery Coronary, Artery Tibial), digestive (Liver, Pancreas, Small Intestine Terminal Ileum, Stomach, Colon Sigmoid, Colon Transverse, Esophagus Gastroesophageal Junction, Esophagus Muscularis, Adipose Subcutaneous, Adipose Visceral Omentum), mucosa (Esophagus Mucosa), gland (Adrenal Gland, Thyroid, Minor Salivary Gland), breast (Breast Mammary Tissue), lung (Lung), sexual tissues (Ovary, Prostate, Testis, Uterus, Vagina). BEDTools (Quinlan and Hall, 2010) was used to intersect the Ts ELSs' coordinates with the *cis*-eQTLs' positions in the considered genomic locations (intronic and intergenic). We kept all eQTL-eGene pairs that were found significantly associated with the matching eQTL-ELS's tissue category (muscle skeletal/cardiac, muscle smooth, fibro/myoblast, digestive, mucosa, brain, blood, aorta). In the case of iPSCs-specific and common ELSs, we considered those eQTL-eGene pairs that were significantly reported in at least 50% of all the tissues. The resulting intersected ELSs were considered as being responsible for the regulation of the associated eGene. The functional enrichment of the ELSs' target genes was performed by the online utility WebGestalt (Liao et al., 2019).

Assessing enhancer regulatory activity with Hi-C-based significant fragment pairs from loop contacts

ELSs were also annotated using significant Hi-C-based interacting fragment pairs from three independent datasets (Jung et al., 2019; Lu et al., 2020; Mifsud et al., 2015). Different primary tissue and cell line samples were used to annotate each of the Ts ELSs categories in our study, except for smooth muscle, for which no Hi-C samples were found. As for the GTEx samples' groups in the previous section, we grouped the Hi-C samples in unique categories

in order to consider the most complete catalogue of Hi-C fragment pairs per group of samples. These categories were named as follows: skeletal/cardiac muscle (Right ventricle (RV), Right heart atrium (RA3), Psoas (PO3), left ventricle (LV)), fibro/myoblasts (Fibroblast cells (IMR-90)), brain (Hippocampus, dorsolateral prefrontal cortex, cortex adult, Neuron), blood (GM12878+GM19240 lymphoblastoid cell line, CD34, GM12878), iPSCs (iPSCs), aorta (Aorta), mucosa (Sigmoid Colon), digestive (Pancreas, Gastric tissue). In order to identify the significant ELS-gene pairs, BEDTools (Quinlan and Hall, 2010) was used to intersect the Hi-C fragment coordinates with our ELSs associated with the different genomic locations (intronic and intergenic). In those cases in which the other fragment did not belong to any other ELS, we intersected them with the GENCODE annotation (v19), inferring in this way the target genes of these ELSs. As for the eQTL annotation, only the Hi-C-based ELS-gene interactions associated with the matching Hi-C-ELSs' tissue category were kept (iPSC, skeletal/cardiac muscle, fibro/myoblast, digestive, brain, blood, aorta). Mucosa- and smooth muscle-Ts ELSs were removed from the analysis due to the lack of intersection with significant fragment pairs and Hi-C sample tissues, respectively. In the case of common ELSs, we considered the ELS-gene pairs reported in at least 50% of all the Hi-C tissue samples. After the annotation of our ELSs, we ended up with a collection of enhancer-gene interactions where the target gene was considered as being regulated by the interacting ELS. In order to define the sets of intronic Host/Non-Host ELSs in Fig. 3A, we identified the ELSs' target genes that are also the host gene of that ELS. If a particular ELS presents among their target genes also its own host gene, then that ELS was classified as Host, if none of the target genes is hosting the ELS, then that element was classified as Non-Host. When considering the interactions ELS-gene in Fig. 3B and Supplementary Fig. 2B, we defined an interaction as Host if the target gene is hosting that ELS, otherwise if the same ELS is targeting a gene that is not hosting the element, that interaction is classified as Non-Host. The target gene expression values were obtained from the GTEx expression data (v7), and Z-score normalized across the different GTEx tissue categories. The hierarchical clustering analyses of the Host/Non-Host target genes and GTEx tissue categories were performed with the R function hclust. The functional enrichment

analyses on the ELSs' target genes and Host/Non-Host target genes were performed with the online utility WebGestalt (Liao et al., 2019).

cis-Regulatory Elements and Transcription Factor Binding Sites

Transcription factor binding sites (TFBSs) were predicted by using the motif discovery software HOMER (Heinz et al., 2010). This program performs a differential motif discovery by taking two sets of genomic regions (findMotifGenome.pl script) and identifying the motifs that are enriched in one set of sequences relative to a background list of regions. We analyzed the Ts ELSs' binding motifs by considering the ELS regions from all the other tissues as background. We searched for 6-mer and 7-mer length motifs as a way to focus on enriched core motif sequences and avoid redundancy from longer motifs with similar functions. A hypergeometric test and FDR correction were applied for the motif enrichment. Only significantly enriched motifs were considered in the subsequent analyses. The functionality of the predicted TFBSs was assessed by analyzing the tissue-specific expression of the transcription factors that bind to them. GTEx expression data (v7) was analyzed for those transcription factors whose TFBSs were reported as significant by HOMER in all tissues and genomic locations. In the gene expression analysis, some transcription factors were removed due to the lack of expression data. *Z*-score normalization was performed across the different GTEx tissue categories in all transcription factors.

ChIP-seq data generation and processing

ChIP-seq was performed in hESC line H9 (WiCell), hESC-derived neural progenitors (NPCs) and neurons. hESCs were maintained in culture in mTESR (Stem Cell Technologies), and NPCs and neurons were obtained upon cerebral organoid differentiation (Lancaster and Knoblich, 2014). Briefly, 9000 H9 hESCs were seeded in a low attachment 96-well (Corning) with Rock Inhibitor in mTESR. After 6 days, neuroepithelium differentiation was triggered using

Induction media for another 6-8 days until the neuroepithelium was detectable and subsequently transferred to the neural expansion in Matrigel (Corning). Organoids were disaggregated at day 30 post-differentiation and maintained in neural differentiation media (common N2B27) supplemented with 20ng/ml of each EGF (Ref: PHG0315; Thermo Fisher Scientific) and FGF2 (Ref:100-18B; Peprotech) to obtain 2D NPC monolayer. NPCs were harvested after 2 passages. Neurons were terminally differentiated in maturation media (N2B27) for 3 more weeks. Cells were harvested with Cell Dissociation Solution (Stem Cell Technologies) and kept at -80° C. DNA was crosslinked with formaldehyde for 10 minutes at room temperature. Fixation was stopped by incubating with PBS / 0.1 % Triton X-100 / 0.125 M glycine for 5 minutes at room temperature and chromatin was fragmented in a Q-sonica sonicator (15 minutes constant sonication at 40% Amplitude). H3K27ac (Active Motif reference 39336) and H3K4me3 (Active Motif reference 39916) antibodies were used for immunoprecipitation following the protocol previously described (Pérez-Lluch et al., 2015). ChIP libraries were performed following Illumina procedures. Libraries were quantified by Qubit (Thermo Fisher Scientific) and visualized in a Fragment Analyzer (Agilent) previous to sequencing. Sequencing was performed in an Illumina NextSeq 500, single-end run, following the instructions of the manufacturer.

Data was processed using the *ChIP-nf* (https://github.com/guigolab/chip-nf) Nextflow (DI Tommaso et al., 2017) pipeline. Input samples were down-sampled to a number of reads comparable to the ChIP samples with the tool seqtk (https://github.com/lh3/seqtk). ChIP-seq reads were aligned to the human genome assembly (GRCh37) using the GEM (Marco-Sola et al., 2012) mapping software, allowing up to two mismatches. Only alignments for reads mapping to ten or fewer loci were reported. Duplicated reads were removed using Picard (http://broadinstitute.github.io/picard/). Peak calling was performed using Zerone (Cuscó and Filion, 2016) with replicates handled internally. Pile-up signal from bigWig files was obtained running MACS2 (Zhang et al., 2008) on individual replicates. No shifting model was built. Instead, fragment length was defined for each experiment and used to extend each read

towards the 3' end (using the --extsize option). Pile-up signal was normalized by scaling larger samples to smaller samples (using the default for the --scale-to option) and adjusting signal per million reads (enabling the --SPMR option). To calculate the proportion of ELSs described in Fig. 5D, only active candidate ELSs (H3K27ac+/H3K4me3-) overlapping ENCODE tissue-specific ELSs in the matched ENCODE biosamples were considered (i.e. Neuron ELSs overlapping ENCODE adult brain-specific ELSs; NPCs ELSs overlapping ENCODE neural progenitors-specific ELSs; ESCs ELSs overlapping ENCODE ESCs-specific ELSs).

Gene expression analysis

To validate gene expression regulation, target genes regulated by intronic or intergenic ELSs were selected based on the following criteria (see also Supplementary Table 14): i) controlled by a single ENCODE ELS in adult samples, either brain-specific or common (column "Tissue"), ii) showing H3K27ac+/H3K4me3- peaks in the relevant cell's ChIP-seq validation (column "Peak ChIP-seq"), iii) not overlapping with exons.

RNA was obtained from hESCs, NPCs and neuron pellets used for ChIP-seq. Retrotranscription was performed using SuperScript III retrotranscriptase. qPCR was performed in 10 ng cDNA with the Roche SYBR Green Master Mix. Primers used for qPCR are reported in Supplementary Table 14. Gene expression is reported following the relative expression of the DDCt method. *GAPDH* and *ACTB* were used as reference genes. *ACTB* gene expression showed more stability throughout the differentiation process and, therefore, it was used as the reference gene for the analysis.

Statistical analyses and visualization

All statistical analyses and visualization plots were performed using the R language for statistical computation and graphics (R Core Team, 2017) (http://www.R-project.org/).

Data access

All raw and processed sequencing data generated from this study have been submitted to ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-10595.

Acknowledgments

Author Contributions: S.A., B.B, and P.V-M. designed the study, analyzed the data and wrote the manuscript with feedback from all the authors. S.A. and S.P-L. performed ChIP-seq experiments, I.T. generated the differentiations for the ChIP-seq. H.L. and A.S-C. performed some of the bioinformatic analyses. J.B. and R.G. contributed to the manuscript edition. S.A. supervised the study.

Competing interest statement

The authors declare no competing interests.

Figures Legends

Fig. 1. Active enhancers define tissue identity. **A**: Highly-shared ELSs are more frequently located in intergenic regions. The scatter plot represents the proportion of intergenic ELSs active in increasing numbers of human adult samples. Error bars represent the 95% confidence interval. **B**: Samples' clustering defined by ELSs' presence-absence patterns. The heatmap depicts the binary distance between any pair of samples, based on the activity of 921,166 distal ELSs (+/- 2 kb from any annotated TSS). The correspondence between samples and numbers is reported in Supplementary Table 1 in Supplementary_File.pdf. **C**: MDS distribution of human adult samples defined by ELSs' activity. Analogous representation to Supplementary Fig. 1A in Supplementary_File.pdf for the subset of 33 selected adult human samples. **D**: Tissue-specific ELSs. The barplot represents the type of samples found within sets of brain-, blood- and muscle-specific ELSs. Most tissue-specific ELSs are only active in the samples of the corresponding cluster ("within-cluster", *black*), but a few of them may be active in at most one outer sample (i.e. a sample that does not belong to the tissue cluster, *coloured*). IPSCs-, fibro/myoblasts-, digestive-, mucosa- and aorta-specific ELSs are not represented, since we did not allow outer samples given their small cluster sizes (see Methods). sk/c = skeletal/cardiac; sm = smooth.

Fig. 2. Intronic location of tissue-specific ELSs. **A**: Proportions of common and tissue-specific ELSs, identified in the 33 selected human adult samples, that overlap intronic, exonic and intergenic regions. Error bars represent the 95% confidence interval (sk/c = skeletal/cardiac; sm = smooth). **B**: Proportion of eQTL-ELSs with respect to the total amount of ELSs in each cluster. **C**: Number of intergenic (Ing) and intronic (Intr) cluster-specific ELSs

harboring eQTLs detected in the analysed GTEx tissue samples. Common and iPSCs-specific ELSs were annotated with a composition of tissue-specific significant eQTLs (see Methods). Coloured cells represent the proportion of region-specific eQTL-ELSs over the total amount of eQTL-ELSs per cluster. Significant differences were observed between common and tissue-specific annotated eQTL-ELSs (Chi square test *p*-value ≤ 0.05), showing that common annotated ELSs are highly associated with intergenic regions. **D**: Proportion of Hi-C-ELSs with respect to the total amount of ELSs in each cluster. **E**: Number of intergenic (Ing) and intronic (Intr) cluster-specific ELSs overlapping Hi-C-based detected fragments in the analysed Hi-C tissue samples. Common ELSs were annotated with a composition of tissue-specific significant Hi-C fragments (see Methods). Coloured cells represent the proportion of Hi-C ELSs over the total amount of tissue-specific Hi-C-ELSs per cluster. Significant differences were observed between common and non-common annotated Hi-C-ELS (Chi square test *p*-value ≤ 0.05).

Fig. 3. Intronic enhancers regulate hosting and non-hosting genes. A: Proportions of Hi-C-ELSs that target their host gene. These proportions were calculated over the total amount of intronic Hi-C-ELSs within each cluster. **B**: *Z*-score normalized median gene expression levels, across GTEx tissue categories, of the genes targeted by intergenic and intronic Hi-C-ELSs. Intronic Hi-C-ELSs are distinguished between those targeting their host gene (Host), and those that target a gene outside their hosting region (non-Host). Dendrograms show the hierarchical clustering of target genes (rows) and GTEx tissue categories (columns). **C**: Top three significantly enriched GO terms found in the genes targeted by host and non-host intronic Hi-C-ELSs. *p*-values (FDR corrected) are shown for each enriched term.

Fig. 4. Differential TFs programs activate intronic and intergenic ELSs in a tissue-specific manner. **A**: Barplots reporting the significantly enriched TFBSs in intronic and intergenic tissue-specific ELSs. **B**: *Z*-score normalized median gene expression, across GTEx tissue categories, of the TFs that bind to significantly enriched TFBSs in each group.

Fig. 5. Dynamic localisation of ELSs throughout embryonic development. **A**. Correlation between ELSs' shareness amongst embryonic samples and frequency of their intergenic localization (error bars represent the 95% confidence interval). **B**. MDS representation of embryonic samples defines three main groups of tissues (ESCs in gray; neural progenitors in yellow; more differentiated tissues in green). **C**. ELSs specific to neural progenitors and differentiated tissues are more frequently intronic, while common ELSs are preferentially intergenic. **D**. Dynamics of the localisation of active ELSs during ESC-derived maturation stages (hESC, neural progenitors (NPCs) and neurons). ELSs defined by H3K27ac+/H3K4me3- peaks during ESC-derived neural maturation, that also overlap with ENCODE ELSs, increasingly distribute in intronic regions as maturation advances.

References

Alver, B. H., Kim, K. H., Lu, P., Wang, X., Manchester, H. E., Wang, W., Haswell, J. R., Park, P. J., and Roberts, C. W. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nature Communications*, 8.

Bieberstein, N. I., Oesterreich, F. C., Straube, K., and Neugebauer, K. M. (2012). First exon length controls active chromatin signatures and transcription. *Cell Reports*, 2(1):62–68.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A., et al. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell*, 171(3):557–572.

Chen, C., Lee, G. A., Pourmorady, A., Sock, E., and Donoghue, M. J. (2015). Orchestration of neuronal differentiation and progenitor pool expansion in the developing cortex by SoxC genes. *Journal of Neuroscience*, 35(29):10629–10642.

Chen, C., Yu, W., Tober, J., Blobel, G. A., Speck, N. A., and Correspondence, K. T. (2019). Spatial Genome Re-organization between Fetal and Adult Hematopoietic Stem Cells. *Cell Reports*, 29(12):4200–4211.

Chorev, M. and Carmel, L. (2012). The function of introns. *Frontiers in Genetics*, 3.

Choukrallah, M. A., Song, S., Rolink, A. G., Burger, L., and Matthias, P. (2015). Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. *Nature Communications*, 6.

Church, D. M., Schneider, V. A., Steinberg, K. M., Schatz, M. C., Quinlan, A. R., Chin, C. S., Kitts, P. A., Aken, B., Marth, G. T., Hoffman, M. M., et al. (2015). Extending reference assembly models. *Genome Biology*, 16(1):13.

Cuscó P. and Filion, G. J. (2016). Zerone: A ChIP-seq discretizer for multiple replicates with built-in quality control. *Bioinformatics*, 32(19):2896–2902.

De Laat, W. and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, 502(7472):499–506.

DeMare, L. E., Leng, J., Cotney, J., Reilly, S. K., Yin, J., Sarro, R., and Noonan, J. P. (2013). The genomic landscape of cohesin-Associated chromatin interactions. *Genome Research*, 23(8):1224–1234.

DI Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319.

Eisenberg, E. and Levanon, E. Y. (2013). Human housekeeping genes, revisited. *Trends in Genetics*, 29(10):569–574.

Ernst, J. and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology*, 28(8):817–825.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49.

Gilbert, N., Boyle, S., Sutherland, H., Heras, J. d. L., Allan, J., Jenuwein, T., and Bickmore, W. A. (2003). Formation of facultative heterochromatin in the absence of HP1. *The EMBO Journal*, 22(20):5540–5550.

Gillies, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin Heavy Chain Gene. *Cell*, 33(3):717-728.

Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L. E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6(5):479–491.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3):311–318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, 38(4):576–89.

Heinz, S., Romanoski, C. E., Benner, C., and Glass, C. K. (2015). The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology*, 16(3):144–154.

Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrbücher für wissenschaftliche Botanik*, 69.

Hilbert, L., Sato, Y., Kuznetsova, K., Bianucci, T., Kimura, H., Jülicher, F., Honigmann, A., Zaburdaev, V., and Vastenhouw, N. L. (2021). Transcription organizes euchromatin via microphase separation. *Nature Communications*, 12(1):1360.

Jacob, A. G. and Smith, C. W. (2017). Intron retention as a component of regulated gene expression programs. *Human Genetics*, 136(9):1043–1057.

Jung, I., Schmitt, A., Diao, Y., Lee, A. J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., et al. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics*, 51(10):1442–1449.

Kawase, S., Imai, T., Miyauchi-Hara, C., Yaguchi, K., Nishimoto, Y., Fukami, S. I., Matsuzaki, Y., Miyawaki, A., Itohara, S., and Okano, H. (2011). Identification of a novel intronic enhancer responsible for the transcriptional regulation of musashi1 in neural stem/progenitor cells. *Molecular Brain*, 4(1).

Khandekar, M., Brandt, W., Zhou, Y., Dagenais, S., Glover, T. W., Suzuki, N., Shimizu, R., Yamamoto, M., Lim, K. C., and Engel, J. D. (2007). A Gata2 intronic enhancer confers its pan-endothelia-specific regulation. *Development*, 134(9):1703–1712.

Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. A., Sloane-Stanley, J. A., McGowan, S. J., De Gobbi, M., Hosseini, M., Vernimmen, D., et al. (2012). Intragenic enhancers act as alternative promoters. *Molecular Cell*, 45(4):447–458.

Lancaster, M. A. and Knoblich, J. A. (2014). Generation of cerebral organoids from human pluripotent stem cells. *Nature Protocols*, 9(10):2329–2340.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Current Biology*, 20(17):754–763.

Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1):189–192.

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, 47(W1):W199–W205.

Lis, R., Karrasch, C. C., Poulos, M. G., Kunar, B., Redmond, D., Duran, J. G., Badwe, C. R., Schachterle, W., Ginsberg, M., Xiang, J., et al. (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature*, 545(7655):439–445.

Lu, L., Liu, X., Huang, W. K., Giusti-Rodríguez, P., Cui, J., Zhang, S., Xu, W., Wen, Z., Ma, S., Rosen, J. D., et al. (2020). Robust Hi-C maps of enhancer-promoter interactions reveal the function of non-coding genome in neural development and diseases. *Molecular Cell*, 79(3):521–534.

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: Fast, accurate and versatile alignment by filtration. *Nature Methods*, 9(12):1185–1188.

Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., et al. (2015). The human transcriptome across tissues and individuals. *Science*, 348(6235):660–665.

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., et al. (2015). Mapping long-range

promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, 47(6):598–606.

Ott, C. J., Blackledge, N. P., Kerschner, J. L., Leir, S. H., Crawford, G. E., Cotton, C. U., and Harris, A. (2009). Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proceedings of the National Academy of Sciences of the United States of America*, 106(47):19934–19939.

Pennacchio, L. A., Loots, G. G., Nobrega, M. A., and Ovcharenko, I. (2007). Predicting tissue-specific enhancers in the human genome. *Genome Research*, 17(2):201–211.

Pérez-Lluch, S., Blanco, E., Tilgner, H., Curado, J., Ruiz-Romero, M., Corominas, M., and Guigó, R. (2015). Absence of canonical marks of active chromatin in developmentally regulated genes. *Nature Genetics*, 47(10):1158–1167.

Pervouchine, D. D., Djebali, S., Breschi, A., Davis, C. A., Barja, P. P., Dobin, A., Tanzer, A., Lagarde, J., Zaleski, C., See, L. H., et al. (2015). Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. *Nature Communications*, 6.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*.

Rand, E. and Cedar, H. (2003). Regulation of imprinting: A multi-tiered process. *Journal of Cellular Biochemistry*, 88(2):400–407.

Rose, A. B. (2019). Introns as gene regulators: A brick on the accelerator. *Frontiers in Genetics*, 9.

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L., et al. (2016). A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports*, 17(8):2042–2059.

Shaul, O. (2017). How introns enhance gene expression. *International Journal of Biochemistry and Cell Biology*, 91:145–155.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4):272–286.

The ENCODE Project Consortium (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583(7818):699–710.

The GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204–213.

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518(7540):556–559.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9):R137.
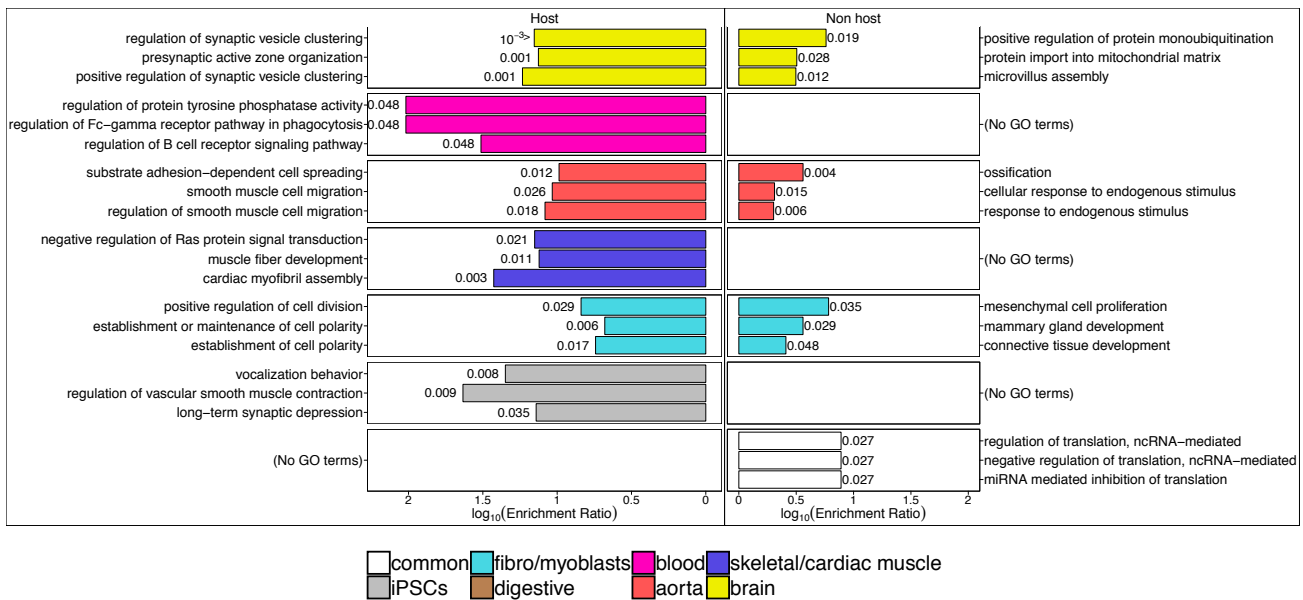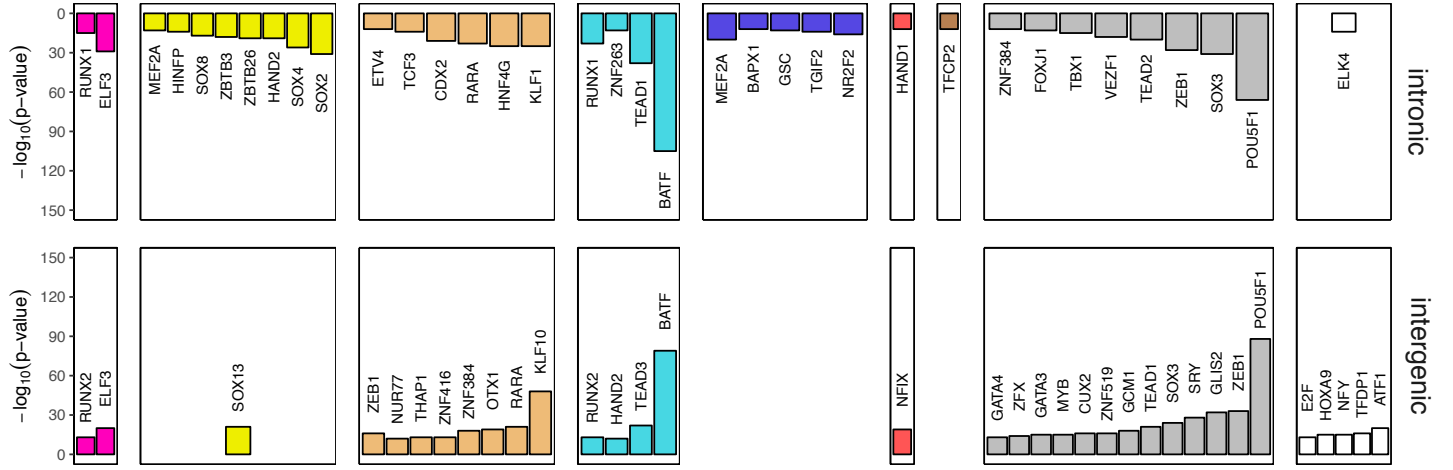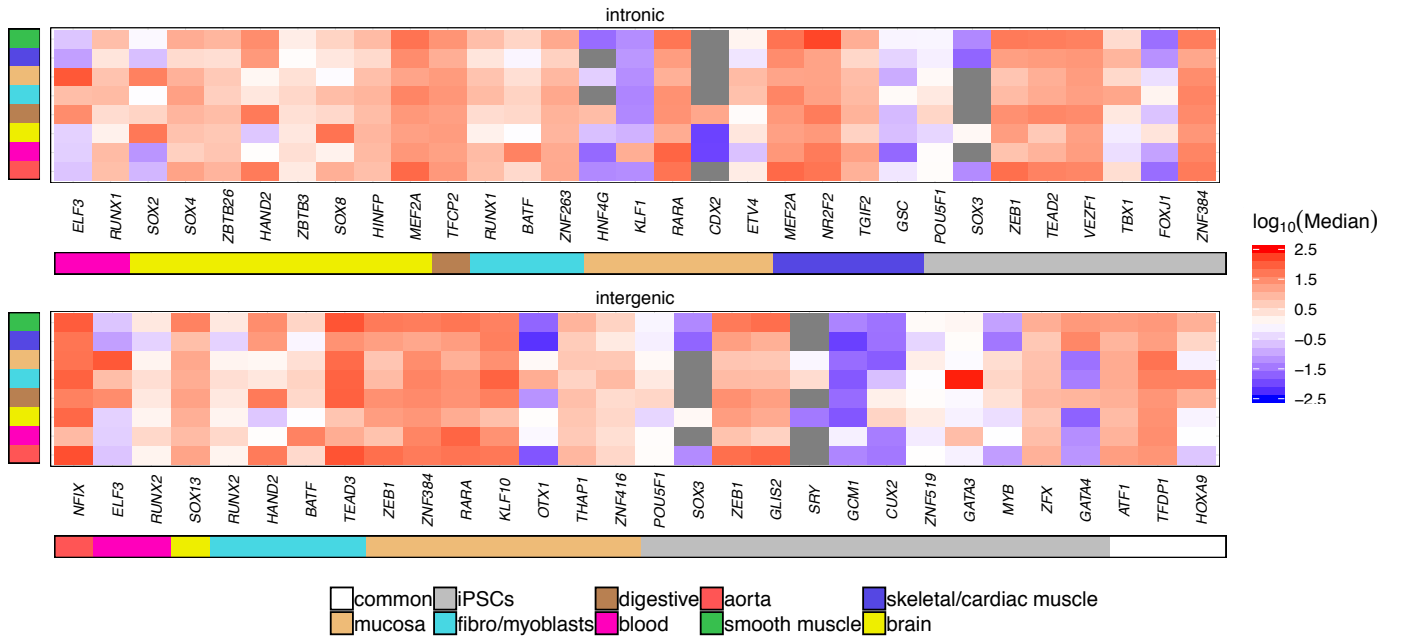
A

B
aorta

Intergenic   Intronic (non–Host) Intronic (Host)

blood

Intergenic   Intronic (non–Host) Intronic (Host)

brain

Intergenic   Intronic (non–Host) Intronic (Host)

Z–score

C

	Host	Non host	
regulation of synaptic vesicle clustering	10^{-3}>	0.019	positive regulation of protein monoubiquitination
presynaptic active zone organization	0.001	0.028	protein import into mitochondrial matrix
positive regulation of synaptic vesicle clustering	0.001	0.012	microvillus assembly
regulation of protein tyrosine phosphatase activity	0.048		
regulation of Fc–gamma receptor pathway in phagocytosis	0.048		(No GO terms)
regulation of B cell receptor signaling pathway	0.048		
substrate adhesion–dependent cell spreading	0.012	0.004	ossification
smooth muscle cell migration	0.026	0.015	cellular response to endogenous stimulus
regulation of smooth muscle cell migration	0.018	0.006	response to endogenous stimulus
negative regulation of Ras protein signal transduction	0.021		
muscle fiber development	0.011		(No GO terms)
cardiac myofibril assembly	0.003		
positive regulation of cell division	0.029	0.035	mesenchymal cell proliferation
establishment or maintenance of cell polarity	0.006	0.029	mammary gland development
establishment of cell polarity	0.017	0.048	connective tissue development
vocalization behavior	0.008		
regulation of vascular smooth muscle contraction	0.009		(No GO terms)
long–term synaptic depression	0.035		
(No GO terms)		0.027	regulation of translation, ncRNA–mediated
		0.027	negative regulation of translation, ncRNA–mediated
		0.027	miRNA mediated inhibition of translation

$\log_{10}$(Enrichment Ratio) $\log_{10}$(Enrichment Ratio)

common   fibro/myoblasts   blood   skeletal/cardiac muscle
iPSCs   digestive   aorta   brain

A

B

C introns | exons | intergenic regions

common | ESCs | neural progenitors | differentiated tissues

D

Dynamics in ELS throughout development

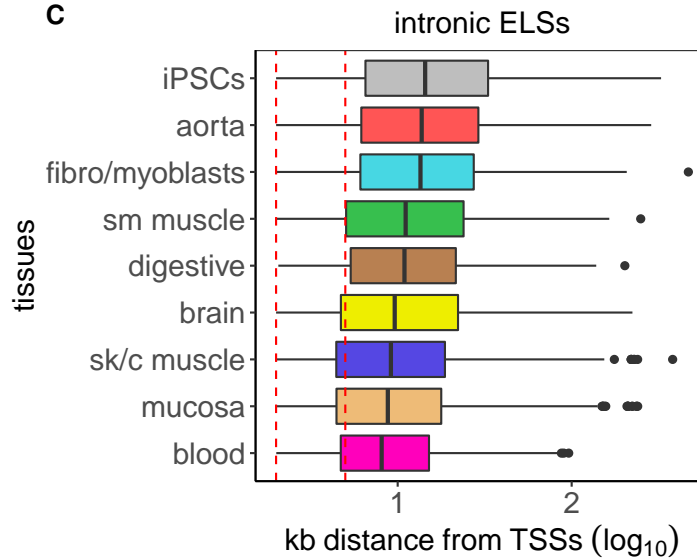# Supplementary Figures and Tables

List of Figures

List of Tables

Supplementary Fig. 1. **A**: Multidimensional scaling (MDS) representation of the dissimilarities among the 43 human adult samples based on the pattern of activity of ELS-cCREs. The binary distance between a given pair of samples was computed considering presence/absence vectors of 921,166 distal ELSs (+/- 2 kb from any annotated TSS). The correspondence between samples and numbers is reported in Supplementary Table 1 in Supplementary_File.pdf. **B**: Features of genes hosting intronic ELSs in each cluster of adult samples: (1) number of introns per hosting gene, (2) length of hosting gene, (3) median intron length per hosting gene. **C**: Distributions of distances of tissue-specific intronic ELSs from annotated TSSs. The minimum distance from either the start or the end of every ELS was considered. Vertical dashed red lines correspond to 2 and 5 kb. sk/c = skeletal/cardiac; sm = smooth.
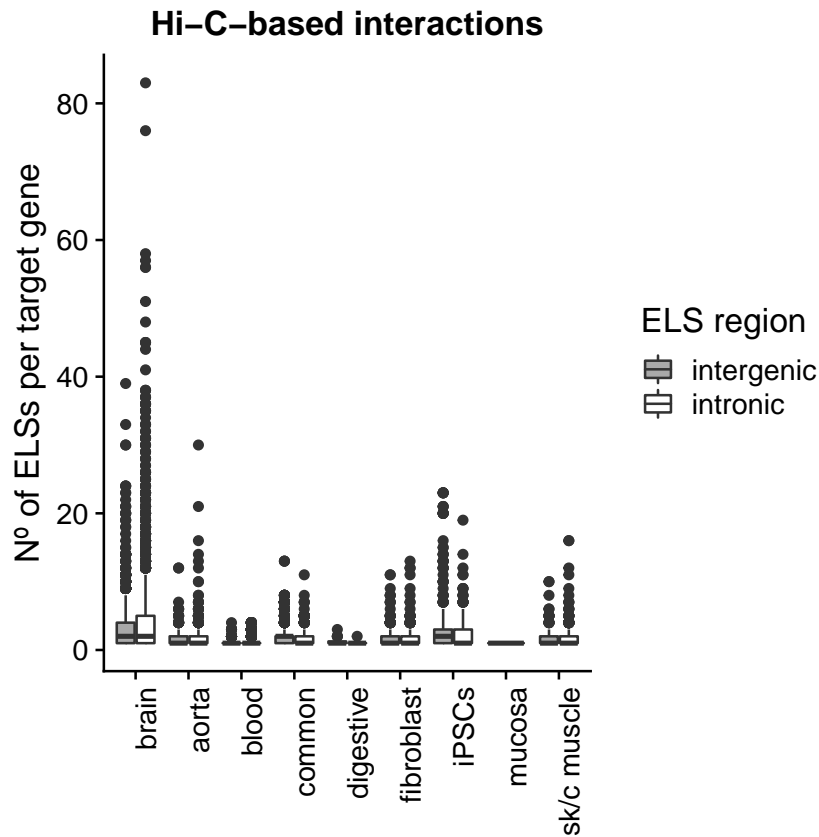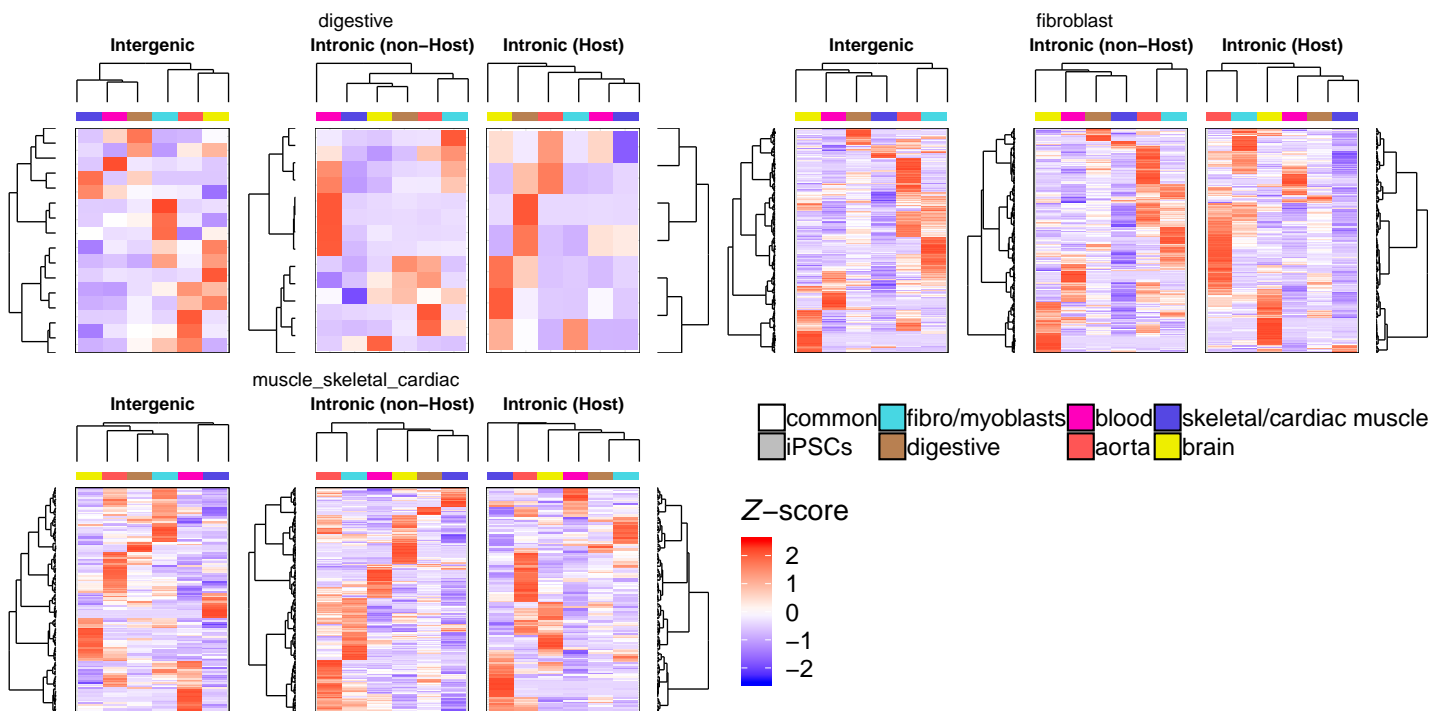
Supplementary Fig. 2. **A**: Distributions of the number of common and tissue-specific ELSs regulating a single target gene according to the Hi-C-based data. sk/c = skeletal/cardiac. **B**: *Z*-score normalized median gene expression across GTEx tissue categories of the genes targeted by intergenic and intronic Hi-C-ELSs in digestive, fibroblast and skeletal/cardiac muscle tissues. Intronic Hi-C-ELSs are distinguished between those targeting their host gene (Host), and those that target a gene that is not their host gene (non-Host). Dendrograms show the hierarchical clustering of target genes (rows) and GTEx tissue categories (columns).
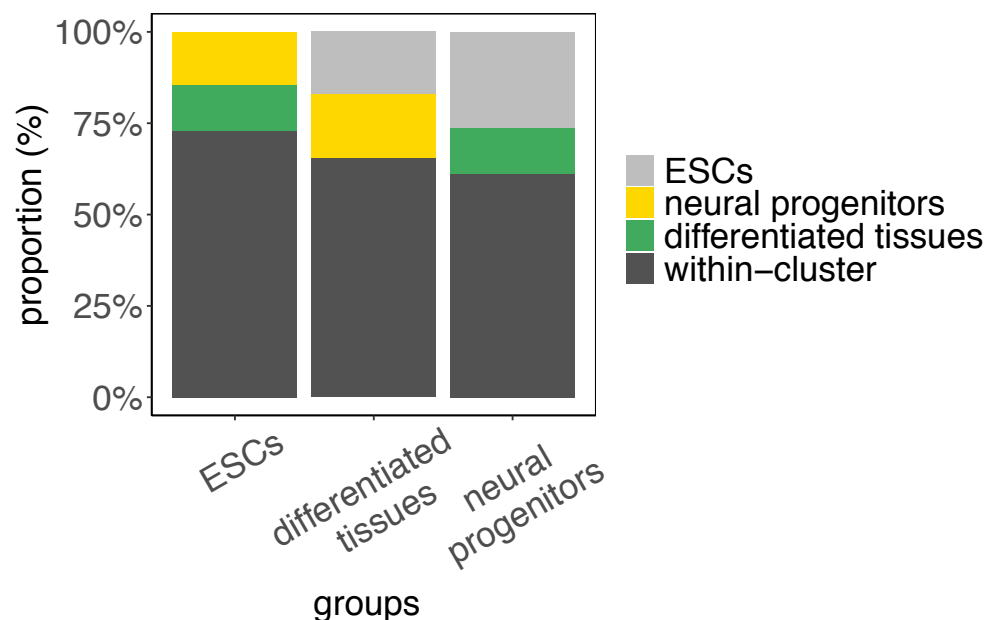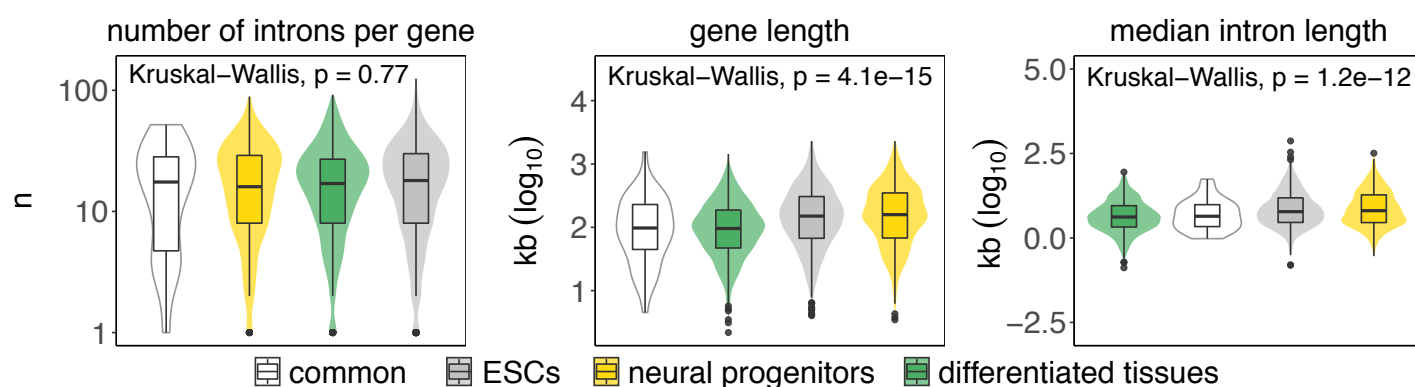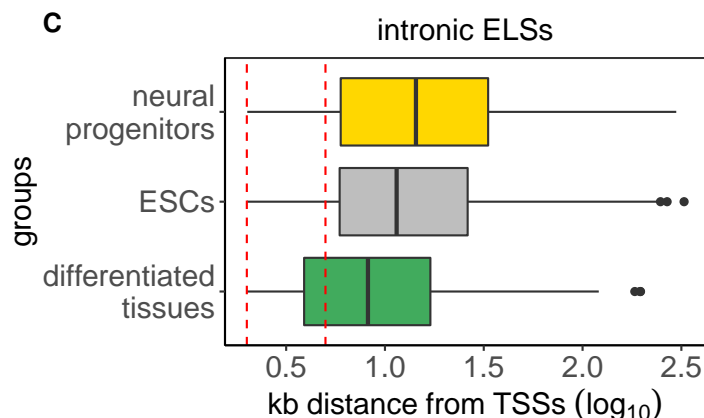
**Supplementary Fig. 3**. **A**: Group-specific ELSs in embryonic samples (analogous to Fig. 1D). The barplot represents the type of outer samples observed within sets of ESCs-, differentiated tissues- and neural progenitors-specific ELSs. **B**: Features of genes hosting either common or specific intronic ELSs identified in embryonic samples (analogous to Supplementary Fig. 1B in Supplementary_File.pdf): (1) number of introns per hosting gene, (2) length of hosting gene, (3) median intron length per hosting gene. **C**: Distributions of distances of group-specific intronic ELSs from annotated TSSs (analogous to Supplementary Fig. 1C in Supplementary_File.pdf). The minimum distance from either the start or the end of every ELS was considered. Vertical dashed red lines correspond to 2 and 5 kb.
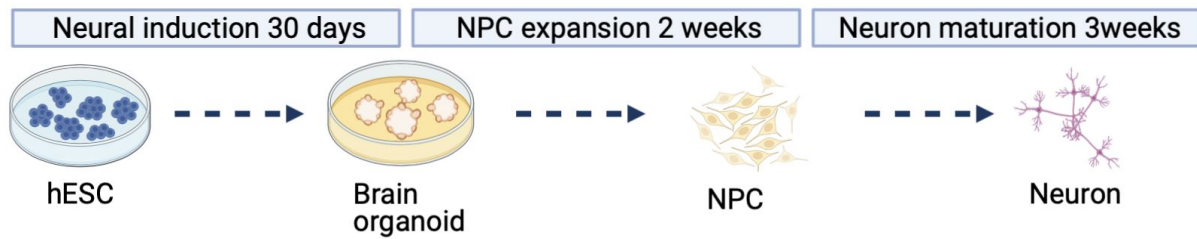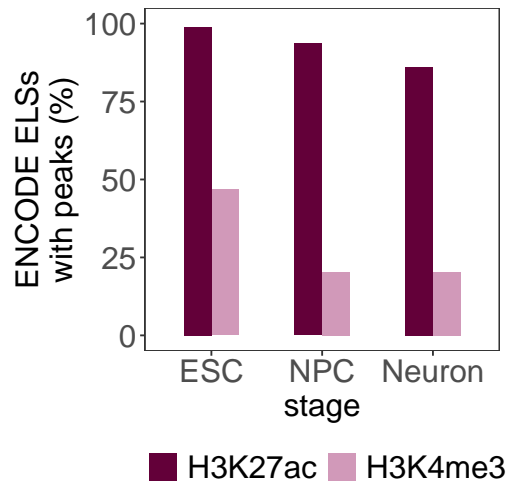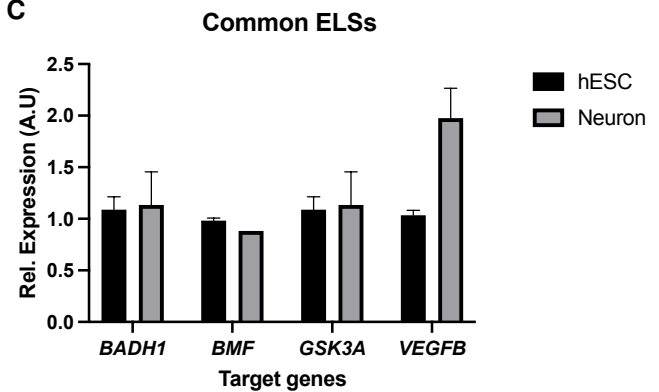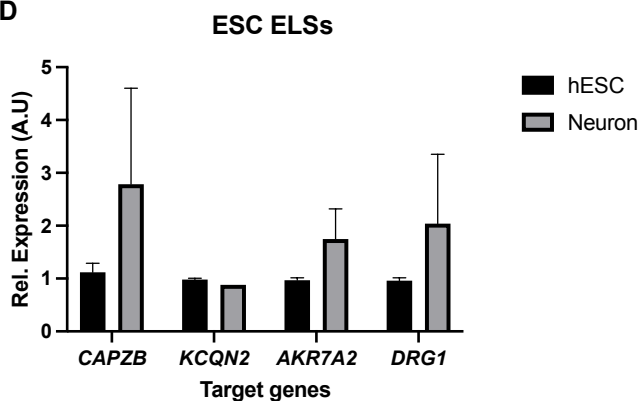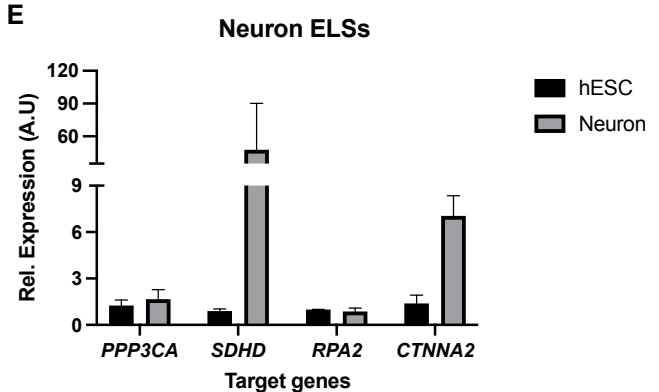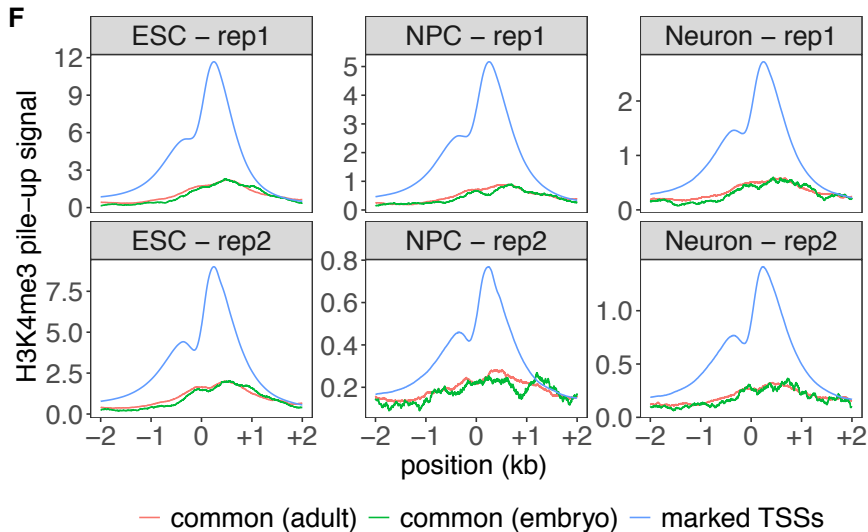
Supplementary Fig. 4. **A**: Scheme depicting the differentiation protocol of ESCs into NPCs and neurons. **B**: Overlap between ENCODE embryonic common ELSs and ChIP-seq peaks of H3K27ac and H3K4me3 observed in ESCs, NPCs and neurons. **C-E**: Gene expression analysis in hESC and ESC-derived neurons of genes targeted by ENCODE ELSs experimentally validated by common (**C**), ESC-specific (**D**) and neuron-specific (**E**) ChIP-seq peaks (see also Supplementary Table 14). Relative quantification was performed against hESC gene expression values and reference gene was *ACTB*. This analysis was performed in triplicates. **F**. Pile-up H3K4me3 signal for ENCODE common embryonic and adult ELSs in ESCs, NPCs and Neurons detected in each of the ChIP-seq replicates. The signal at marked TSSs is comparatively higher than at ELSs, suggesting low promoter activity in the selected ELSs.
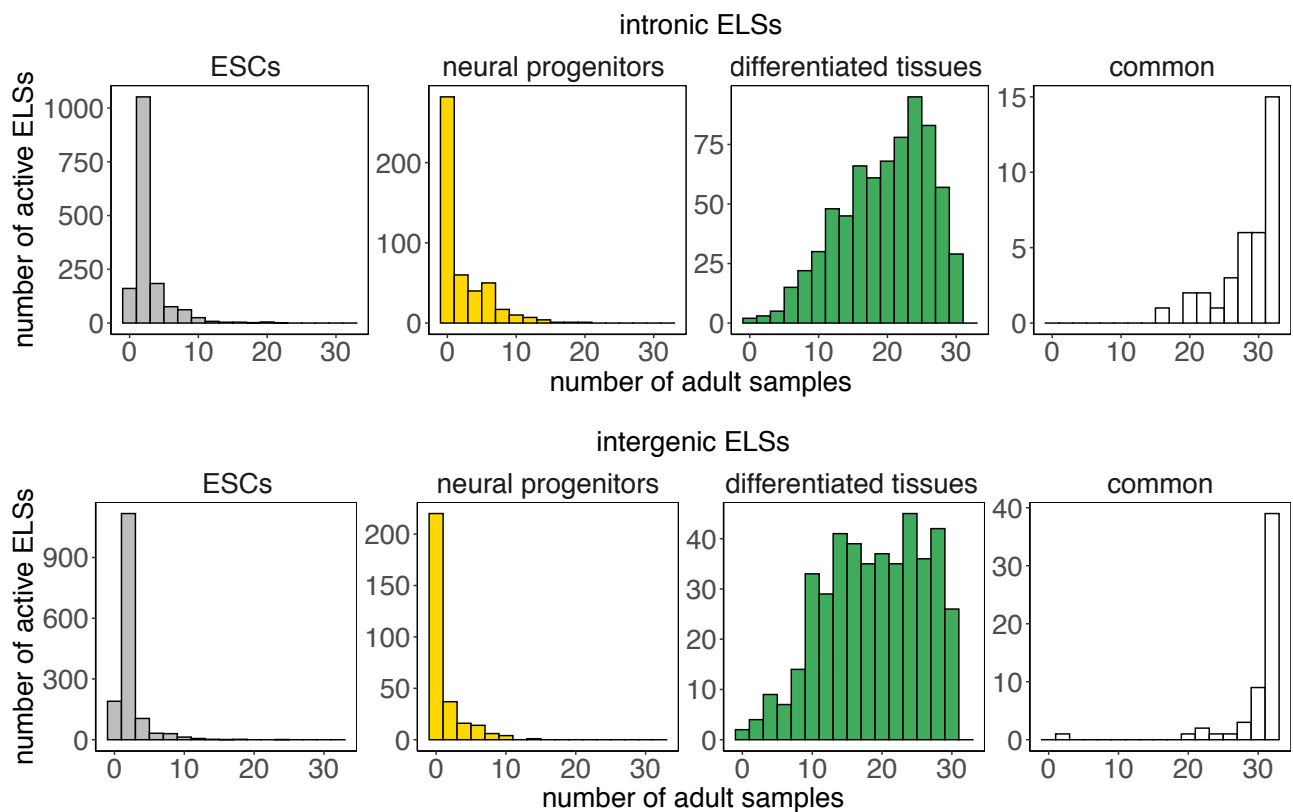
Supplementary Fig. 5. Overlap of ELSs between embryonic and adult tissues. ESCs- and neural progenitors-specific ELSs are not active in any or very few adult samples (n = 33), independently of their genomic location (intronic *vs.* intergenic). Instead, most of the embryonic common ELSs, especially those intergenic, are also active in adult tissues. Intermediate distributions are observed for ELSs specific to differentiated embryonic tissues.
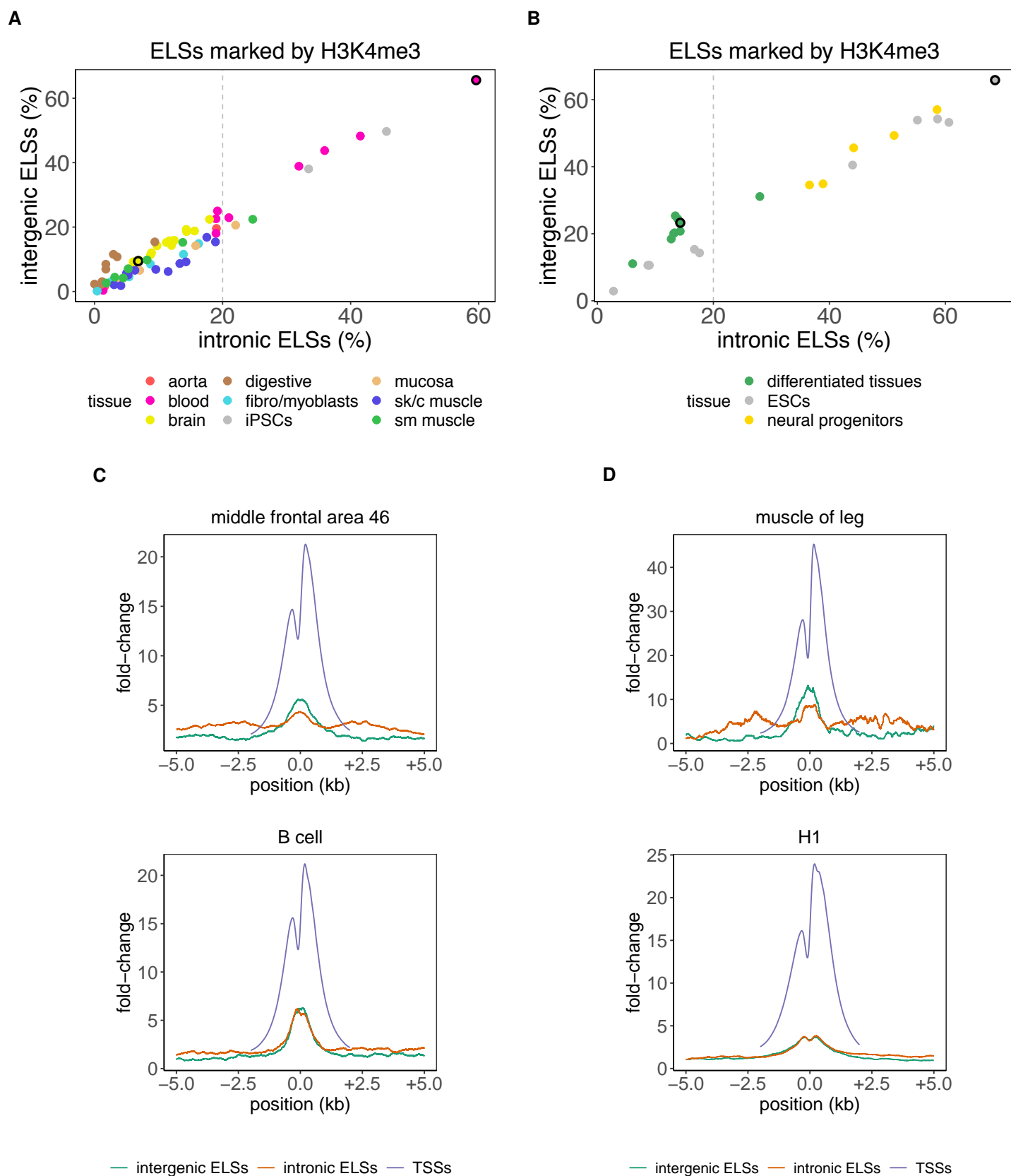
Supplementary Fig. 6. **A, B**: Proportions of intergenic and intronic tissue-specific ELSs marked (i.e. showing presence of called peaks) by H3K4me3 signal across adult (**A**) and embryonic (**B**) samples (sk/c = skeletal/cardiac; sm = smooth). **C, D**: Aggregated H3K4me3 fold-change signal in marked TSSs, as well as marked intronic and intergenic ELSs. Two representative samples were selected for adult (**C**) and embryonic (**D**) tissues. The samples used are indicated by bolded circles in **A** (brain: middle frontal area 46; blood: B cell) and **B** (differentiated tissues: muscle of leg; ESCs: H1). The signal at marked TSSs was computed over +/- 2 kb from the TSS. The signal at marked ELSs, instead, was computed over a +/- 5 kb region from the center of the ELS.
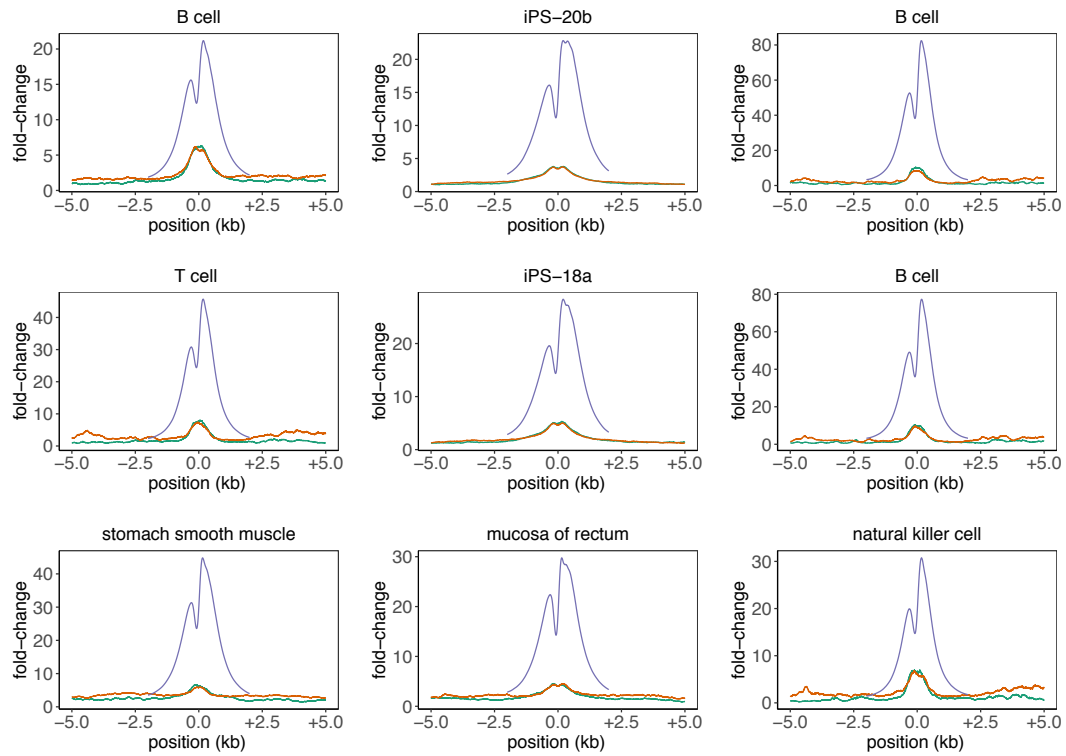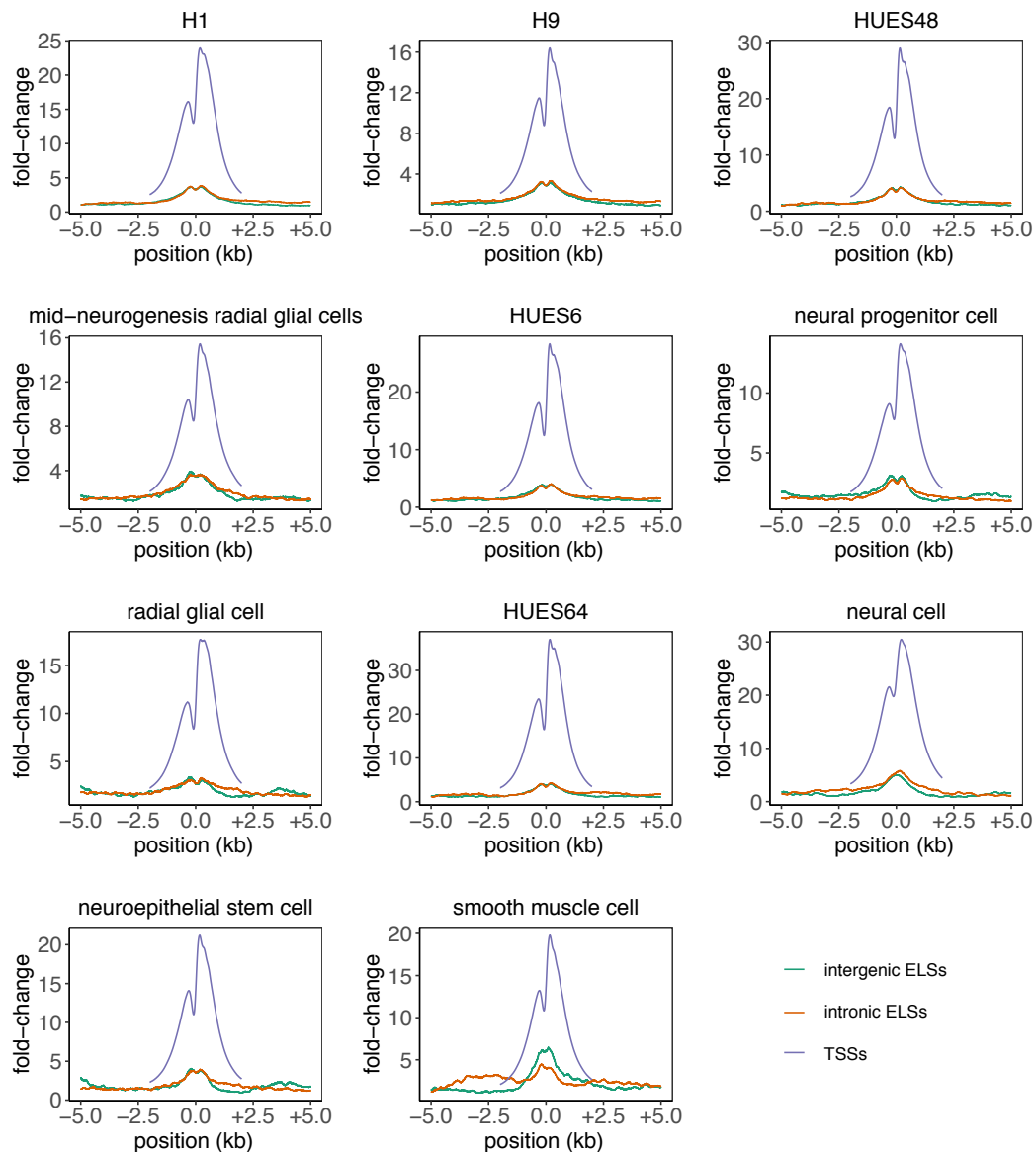
Supplementary Fig. 7. **A, B**: Analogous representations to Supplementary Figs. 6C-D for adult (**A**) and embryonic (**B**) samples reporting $\geq 20\%$ of ELSs marked by H3K4me3 in Supplementary Figs. 6A-B.

	Biosample Term Name	Biosample Type	Samples' Cluster	ENCODE File ID
1	natural killer cell	primary cell	blood	ENCFF529UWB
2	T cell	primary cell	blood	ENCFF098NHL
3	B cell	primary cell	blood	ENCFF379TAE
4	CD14-positive monocyte	primary cell	blood	ENCFF967MJU
5	peripheral blood mononuclear cell	primary cell	blood	ENCFF509DPX
6	pancreas	tissue	digestive	ENCFF681HOL
7	body of pancreas	tissue	digestive	ENCFF768JUC
8	stomach	tissue	digestive	ENCFF992HIZ
9	right lobe of liver	tissue	-	ENCFF476MEG
10	iPS-18a	cell line	iPSCs	ENCFF920QRH
11	iPS-20b	cell line	iPSCs	ENCFF231KWX
12	bipolar neuron	in vitro differentiated cells	-	ENCFF045GKW
13	thyroid gland	tissue	-	ENCFF296SZK
14	gastrocnemius medialis	tissue	-	ENCFF322RAX
15	endocrine pancreas	tissue	-	ENCFF055CJM
16	ovary	tissue	-	ENCFF586NXH
17	myotube	in vitro differentiated cells	fibro/myoblasts	ENCFF120MMC
18	skeletal muscle myoblast	primary cell	fibro/myoblasts	ENCFF037UZZ
19	fibroblast of lung	primary cell	fibro/myoblasts	ENCFF495RTY
20	aorta	tissue	aorta	ENCFF178GDW
21	thoracic aorta	tissue	aorta	ENCFF257XAQ
22	stomach smooth muscle	tissue	sm muscle	ENCFF726JTT
23	rectal smooth muscle tissue	tissue	sm muscle	ENCFF093MDL
24	vagina	tissue	sm muscle	ENCFF904XYE
25	muscle layer of duodenum	tissue	sm muscle	ENCFF862BGI
26	gastrocnemius medialis	tissue	sk/c muscle	ENCFF863OGG
27	right cardiac atrium	tissue	sk/c muscle	ENCFF278RUJ
28	skeletal muscle tissue	tissue	sk/c muscle	ENCFF311MNY
29	subcutaneous abdominal adipose tissue	tissue	sk/c muscle	ENCFF725QLM
30	esophagus	tissue	-	ENCFF442HYL
31	lung	tissue	-	ENCFF598QTT
32	liver	tissue	-	ENCFF645PQQ
33	spleen	tissue	-	ENCFF821ESA
34	mucosa of rectum	tissue	mucosa	ENCFF759YFL
35	mucosa of rectum	tissue	mucosa	ENCFF403IPC
36	colonic mucosa	tissue	mucosa	ENCFF867TJN
37	middle frontal area 46	tissue	brain	ENCFF070EXF
38	caudate nucleus	tissue	brain	ENCFF508GKP
39	angular gyrus	tissue	brain	ENCFF942KAC
40	layer of hippocampus	tissue	brain	ENCFF159NZA
41	substantia nigra	tissue	brain	ENCFF233VRB
42	temporal lobe	tissue	brain	ENCFF810IQU
43	cingulate gyrus	tissue	brain	ENCFF494WCN

Supplementary Table 1. ENCODE catalogues of cell type-specific candidate cis-Regulatory Elements (cCREs) for 43 human adult samples. The accession number (ENCODE File ID) allows to uniquely identify the catalogue on the ENCODE portal (`https://www.encodeproject.org/`). The color palette was inspired by the Genotype Tissue Expression (GTEx) Project.

Samples	Tissue-specific ELSs
mucosa	6,205
blood	750
iPSCs	10,966
fibro/myoblasts	2,207
digestive	302
aorta	6,231
smooth muscle	2,825
skeletal/cardiac muscle	5,467
brain	13,054

Samples	Common ELSs
all	555

Supplementary Table 2. [upper panel] Number of ELSs that are specific to each of the 9 clusters of 33 selected human adult samples. Tissue-specific ELSs are those active in 100% (iPSCs, fibro/myoblasts, digestive, mucosa and aorta) or $\geq$ 80% (all other clusters) of the samples within a cluster. In addition, they are active in 0 (iPSCs, fibro/myoblasts, digestive, mucosa and aorta) or at most 1 (all other clusters) outer sample (i.e. a sample that does not belong to the considered cluster). [lower panel] Number of ELSs active in $\geq$ 95% (i.e. n = 31) of the 33 selected human adult samples (common ELSs).

Genomic location	Tissue cluster	FDR	Odds ratio	Confidence interval
intronic	mucosa	2.0E-07	1.62	1.35-1.94
	iPSCs	7.6E-14	1.96	1.64-2.35
	fibro/myoblasts	3.2E-13	2.05	1.68-2.49
	digestive	4.7E-07	2.11	1.58-2.84
	blood	1.7E-14	2.44	1.93-3.07
	aorta	4.7E-29	2.76	2.30-3.32
	sm muscle	3.7E-37	3.38	2.79-4.11
	sk/c muscle	1.5E-49	3.89	3.23-4.69
	brain	1.8E-66	4.66	3.90-5.58
exonic	mucosa	9.9E-06	0.30	0.19-0.51
	iPSCs	2.0E-06	0.28	0.18-0.46
	fibro/myoblasts	4.0E-08	0.15	0.07-0.30
	digestive	2.7E-03	0.15	0.02-0.63
	blood	1.8E-02	0.44	0.21-0.90
	aorta	2.1E-04	0.38	0.24-0.63
	sm muscle	3.5E-04	0.37	0.21-0.64
	sk/c muscle	2.5E-04	0.39	0.24-0.65
	brain	3.0E-02	0.61	0.40-0.99
intergenic	mucosa	8.3E-05	0.70	0.58-0.84
	iPSCs	9.0E-10	0.58	0.48-0.69
	fibro/myoblasts	6.2E-09	0.57	0.47-0.69
	digestive	4.3E-05	0.55	0.41-0.73
	blood	6.6E-12	0.45	0.36-0.57
	aorta	6.0E-24	0.40	0.33-0.48
	sm muscle	1.1E-31	0.33	0.27-0.40
	sk/c muscle	2.7E-43	0.28	0.23-0.34
	brain	1.6E-63	0.22	0.19-0.26

Supplementary Table 3. For each cluster of samples we assessed, with Fisher's exact test, significant differences in the proportions of common *vs.* tissue-specific ELSs that overlap intronic, exonic and intergenic regions. *p*-value (FDR-corrected), odds ratio and confidence interval are reported for each test. sk/c = skeletal/cardiac; sm = smooth.

Group	Genes ∩ ELSs			
	Introns	**Exons**	**Both**	**Total**
mucosa	1,245 (82.56%)	51 (3.38%)	212 (14.06%)	1,508
blood	335 (85.24%)	14 (3.56%)	44 (11.20%)	393
iPSCs	1,910 (84.03%)	59 (2.60%)	304 (13.37%)	2,273
fibro/myoblasts	749 (86.89%)	15 (1.74%)	98 (11.37%)	862
digestive	129 (90.21%)	3 (2.10%)	11 (7.69%)	143
aorta	1,058 (79.31%)	47 (3.52%)	229 (17.17%)	1,334
smooth muscle	656 (81.59%)	29 (3.61%)	119 (14.80%)	804
skeletal/cardiac muscle	1,298 (80.82%)	49 (3.05%)	259 (16.13%)	1,606
brain	1,523 (64.51%)	145 (6.14%)	693 (29.35%)	2,361
common	144 (83.24%)	14 (8.09%)	15 (8.67%)	173

Supplementary Table 4. Number of genes whose introns and/or exons intersect tissue-specific and common ELSs identified in adult samples.

Tissue	GO term	Description
aorta	GO:0031589	cell-substrate adhesion
	GO:0043062	extracellular structure organization
	GO:2000147	positive regulation of cell motility
	GO:0043087	regulation of GTPase activity
	GO:0061564	axon development
blood	GO:0042110	T cell activation
	GO:0051056	regulation of small GTPase mediated signal transduction
	GO:0002764	immune response-regulating signaling pathway
	GO:0002521	leukocyte differentiation
	GO:0050900	leukocyte migration
brain	GO:0061564	axon development
	GO:0050808	synapse organization
	GO:0022604	regulation of cell morphogenesis
	GO:0099177	regulation of trans-synaptic signaling
	GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
skeletal/cardiac muscle	GO:0003012	muscle system process
	GO:0042692	muscle cell differentiation
	GO:0007517	muscle organ development
	GO:0051056	regulation of small GTPase mediated signal transduction
	GO:0034330	cell junction organization
smooth muscle	GO:0043062	extracellular structure organization
	GO:0003012	muscle system process
	GO:0019932	second-messenger-mediated signaling
	GO:0003013	circulatory system process
	GO:0099177	regulation of trans-synaptic signaling
mucosa	GO:0051056	regulation of small GTPase mediated signal transduction
	GO:0038127	ERBB signaling pathway
	GO:0034330	cell junction organization
	GO:0043087	regulation of GTPase activity
	GO:0032970	regulation of actin filament-based process
digestive	-	-
fibro/myoblasts	GO:0043087	regulation of GTPase activity
	GO:0010975	regulation of neuron projection development
	GO:0051056	regulation of small GTPase mediated signal transduction
	GO:0090130	tissue migration
	GO:2000147	positive regulation of cell motility
iPSCs	GO:0010975	regulation of neuron projection development
	GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
	GO:0022604	regulation of cell morphogenesis
	GO:0061564	axon development
	GO:0050808	synapse organization
common	GO:0034330	cell junction organization
	GO:1903706	regulation of hemopoiesis
	GO:1901652	response to peptide
	GO:0002521	leukocyte differentiation
	GO:0035264	multicellular organism growth

Supplementary Table 5. Significantly enriched GO terms (Biological Process) associated with genes hosting intronic ELSs identified in adult samples. Only the top five enriched terms are shown for each group.

Tissue	Hosting	GO term	Description
aorta	intronic	-	-
	intergenic	MF:0004499	N,N-dimethylaniline monooxygenase activity
		MF:0004024	alcohol dehydrogenase activity, zinc-dependent
		MF:0004022	alcohol dehydrogenase (NAD) activity
blood	intronic	-	-
	intergenic	BP:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II
		BP:0060333	interferon-gamma-mediated signaling pathway
		BP:0050852	T cell receptor signaling pathway
brain	intronic	BP:0000226	microtubule cytoskeleton organization
		BP:0030030	cell projection organization
		BP:0120036	plasma membrane bounded cell projection organization
	intergenic	CC:0033267	axon part
		CC:0005815	microtubule organizing center
		CC:0015630	microtubule cytoskeleton
fibro/myoblasts	intronic	CC:0015629	actin cytoskeleton
	intergenic	-	-
digestive	intronic	-	-
	intergenic	MF:0035591	signaling adaptor activity
mucosa	intronic	BP:0044281	small molecule metabolic process
		MF:0016289	CoA hydrolase activity
		MF:0008395	steroid hydroxylase activity
	intergenic	MF:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
skeletal/cardiac muscle	intronic	BP:0006085	acetyl-CoA biosynthetic process
		BP:0006520	cellular amino acid metabolic process
		BP:0019752	carboxylic acid metabolic process
	intergenic	CC:0071556	integral component of lumenal side endoplasmic reticulum membrane
		MF:0042605	peptide antigen binding
		CC:0098553	lumenal side of endoplasmic reticulum membrane
smooth muscle	intronic	MF:0016408	C-acyltransferase activity
		MF:0019842	vitamin binding
		MF:0050662	coenzyme binding
	intergenic	-	-
iPSCs	intronic	-	-
	intergenic	-	-
common	intronic	-	-
	intergenic	-	-

Supplementary Table 6. Significantly enriched GO terms associated with the intergenic and intronic eQTL-ELSs' target genes. Only the three top enriched Biological Process (BP) terms are shown for each analysis, when no BP terms are found Molecular Function (MF) and Cellular Component (CC) terms are shown instead.

Tissue	Hosting	GO term	Description
aorta	intronic	BP:0014910	regulation of smooth muscle cell migration
		BP:0048660	regulation of smooth muscle cell proliferation
		BP:0003205	cardiac chamber development
	intergenic	MF:0004722	protein serine/threonine phosphatase activity
		CC:0031012	extracellular matrix
blood	intronic	BP:0150079	negative regulation of neuroinflammatory response
		BP:0042093	T-helper cell differentiation
		BP:0002294	CD4-positive, alpha-beta T cell differentiation in immune response
	intergenic	BP:0048006	antigen processing and presentation, endogenous lipid antigen via MHC class Ib
		BP:0061737	leukotriene signaling pathway
		BP:0048007	antigen processing and presentation, exogenous lipid antigen via MHC class Ib
brain	intronic	BP:1990709	presynaptic active zone organization
		BP:0098698	postsynaptic specialization assembly
		BP:0099068	postsynapse assembly
	intergenic	BP:0042426	choline catabolic process
		BP:0055070	copper ion homeostasis
		BP:1902003	regulation of amyloid-beta formation
fibro/myoblasts	intronic	BP:0072273	metanephric nephron morphogenesis
		BP:0061383	trabecula morphogenesis
		BP:0030010	establishment of cell polarity
	intergenic	BP:0007442	hindgut morphogenesis
		BP:1902260	negative regulation of delayed rectifier potassium channel activity
		BP:0001946	lymphangiogenesis
iPSCs	intronic	BP:0045636	positive regulation of melanocyte differentiation
		BP:0061550	cranial ganglion development
		BP:0045986	negative regulation of smooth muscle contraction
	intergenic	BP:0021825	substrate-dependent cerebral cortex tangential migration
		BP:0043383	negative T cell selection
		BP:0030318	melanocyte differentiation
skeletal/cardiac muscle	intronic	CC:0071005	U2-type precatalytic spliceosome
		CC:0031674	I band
		CC:0030018	Z disc
	intergenic	-	-
smooth muscle	intronic	-	-
	intergenic	-	-
digestive	intronic	-	-
	intergenic	-	-
mucosa	intronic	-	-
	intergenic	-	-
common	intronic	-	-
	intergenic	BP:0007156	homophilic cell adhesion via plasma membrane adhesion molecules
		BP:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
		BP:0034728	nucleosome organization

Supplementary Table 7. Significantly enriched GO terms associated with the intergenic and intronic Hi-C-ELSs' target genes. Only the three top enriched Biological Process (BP) terms are shown for each analysis, when no BP terms are found Molecular Function (MF) and Cellular Component (CC) terms are shown instead.

Tissue	Hosting	GO term	Description
brain	host	GO:2000809	positive regulation of synaptic vesicle clustering
		GO:2000807	regulation of synaptic vesicle clustering
		GO:1990709	presynaptic active zone organization
		GO:0048790	maintenance of presynaptic active zone structure
		GO:1905274	regulation of modification of postsynaptic actin cytoskeleton
	non-host	GO:1902527	positive regulation of protein monoubiquitination
		GO:0030150	protein import into mitochondrial matrix
		GO:0030033	microvillus assembly
		GO:1901976	regulation of cell cycle checkpoint
		GO:0044743	protein transmembrane import into intracellular organelle
skeletal/cardiac muscle	host	GO:0055003	cardiac myofibril assembly
		GO:0046580	negative regulation of Ras protein signal transduction
		GO:0048747	muscle fiber development
		GO:0051058	negative regulation of small GTPase mediated signal transduction
		GO:0055013	cardiac muscle cell development
	non-host	-	
blood	host	GO:1905449	regulation of Fc-gamma receptor signaling pathway in phagocytosis
		GO:1903613	regulation of protein tyrosine phosphatase activity
		GO:0050855	regulation of B cell receptor signaling pathway
		GO:0045589	regulation of regulatory T cell differentiation
		GO:0050853	B cell receptor signaling pathway
	non-host	-	
aorta	host	GO:0014910	regulation of smooth muscle cell migration
		GO:0014909	smooth muscle cell migration
		GO:0034446	substrate adhesion-dependent cell spreading
		GO:0014812	muscle cell migration
		GO:0007179	transforming growth factor beta receptor signaling pathway
	non-host	GO:0001503	ossification
		GO:0071495	cellular response to endogenous stimulus
		GO:0009719	response to endogenous stimulus
fibro/myoblasts	host	GO:0051781	positive regulation of cell division
		GO:0030010	establishment of cell polarity
		GO:0007163	establishment or maintenance of cell polarity
		GO:0046578	regulation of Ras protein signal transduction
		GO:0010631	epithelial cell migration
	non-host	GO:0010463	mesenchymal cell proliferation
		GO:0030879	mammary gland development
		GO:0061448	connective tissue development
		GO:0007178	transmembrane receptor protein serine/threonine kinase signaling
		GO:0008284	positive regulation of cell proliferation
iPSCs	host	GO:0003056	regulation of vascular smooth muscle contraction
		GO:0071625	vocalization behavior
		GO:0060292	long-term synaptic depression
		GO:0097106	postsynaptic density organization
		GO:0009187	cyclic nucleotide metabolic process
	non-host	-	-
common	host	-	
	non-host	GO:0035278	miRNA mediated inhibition of translation
		GO:0040033	negative regulation of translation, ncRNA-mediated
		GO:0045974	regulation of translation, ncRNA-mediated
		GO:0034644	cellular response to UV
		GO:0045047	protein targeting to ER

Supplementary Table 8. Significantly enriched GO terms associated with the target genes of Hi-C-ELSs' that are host and non-host of these ELSs. Only the five top enriched Biological Process (BP) terms are shown.

Tissue	Region	Transcription factors
brain	intronic	MEF2A, HINFP, SOX8, ZBTB3, ZBTB26, HAND2, SOX4, SOX2
	intergenic	SOX13
blood	intronic	ELF3, RUNX1
	intergenic	ELF3, RUNX2
skeletal/cardiac muscle	intronic	MEF2A, GSC, TGIF2, NR2F2
	intergenic	-
smooth muscle	intronic	-
	intergenic	-
fibro/myoblasts	intronic	RUNX1, ZNF263, BATF
	intergenic	RUNX2, HAND2, TEAD3, BATF
iPSCs	intronic	POU5F1, SOX3, ZEB1, TEAD2, VEZF1, TBX1, FOXJ1, ZNF384
	intergenic	POU5F1, ZEB1, GLIS2, SRY, SOX3, GCM1, ZNF519, CUX2, MYB, GATA3, ZFX, GATA4
mucosa	intronic	ETV4, CDX2, RARA, HNF4G, KIF1
	intergenic	ZEB1, THAP1, ZNF416, ZNF384, OTX1, RARA, KLF10
digestive	intronic	TFCP2
	intergenic	-
aorta	intronic	-
	intergenic	NFIX
common	intronic	ELK4(ETS)
	intergenic	HOXA9, TFDP1, ATF1

Supplementary Table 9. Transcription factors corresponding to the significantly enriched transcription factor binding sites (TFBSs) reported by HOMER in each group of ELSs and genomic location.

	Biosample Term Name	Biosample Type	Samples' Group	ENCODE File ID
1	HUES6	cell line	stem cells (ESCs)	ENCFF205SDB
2	HUES64	cell line	stem cells (ESCs)	ENCFF180QLH
3	HUES48	cell line	stem cells (ESCs)	ENCFF086FKD
4	mesendoderm	in vitro differentiated cells	-	ENCFF620BVM
5	H9	cell line	stem cells (ESCs)	ENCFF021HBJ
6	H9	cell line	stem cells (ESCs)	ENCFF505OUS
7	H1	cell line	stem cells (ESCs)	ENCFF051OUV
8	mesodermal cell	in vitro differentiated cells	-	ENCFF250CGY
9	endodermal cell	in vitro differentiated cells	-	ENCFF138DOQ
10	neuroepithelial stem cell	in vitro differentiated cells	neural progenitors	ENCFF138OGZ
11	ectodermal cell	in vitro differentiated cells	-	ENCFF332EYK
12	radial glial cell	in vitro differentiated cells	neural progenitors	ENCFF593TNG
13	neural progenitor cell	in vitro differentiated cells	neural progenitors	ENCFF112ZGF
14	mid-neurogenesis radial glial cells	in vitro differentiated cells	neural progenitors	ENCFF376XBS
15	neural stem progenitor cell	in vitro differentiated cells	neural progenitors	ENCFF455CQW
16	neural cell	in vitro differentiated cells	neural progenitors	ENCFF477EUQ
17	smooth muscle cell	in vitro differentiated cells	differentiated tissues	ENCFF281QON
18	thymus	tissue	differentiated tissues	ENCFF059PHA
19	adrenal gland	tissue	differentiated tissues	ENCFF840ANN
20	IMR-90	cell line	-	ENCFF469PXS
21	fibroblast of lung	primary cell	differentiated tissues	ENCFF292NZP
22	muscle of trunk	tissue	differentiated tissues	ENCFF800YES
23	muscle of leg	tissue	differentiated tissues	ENCFF941JIE
24	stomach	tissue	differentiated tissues	ENCFF198WHL
25	hepatocyte	in vitro differentiated cells	differentiated tissues	ENCFF093BQM
26	large intestine	tissue	differentiated tissues	ENCFF903RGX
27	small intestine	tissue	differentiated tissues	ENCFF543DVJ

Supplementary Table 10. ENCODE catalogues of cell type-specific candidate cis-Regulatory Elements (cCREs) for 27 human embryonic samples. The accession number (ENCODE File ID) allows to uniquely identify the catalogue on the ENCODE portal (`https://www.encodeproject.org/`).

Samples	Group-specific ELSs
ESCs	3,112
neural progenitors	784
differentiated tissues	1,166

Samples	Common ELSs
all	94

Supplementary Table 11. [upper panel] Number of ELSs that are specific to each of the 3 groups of 22 selected human embryonic samples. Group-specific ELSs are active in $\geq$ 80% of the samples within a group, and in at most 1 outer sample (i.e. a sample that does not belong to the considered group). [lower panel] Number of ELSs active in 100% of the 22 selected human embryonic samples (common ELSs).

Genomic location	Samples' Group	FDR	Odds ratio	Confidence interval
intronic	ESCs	3.1E-02	1.67	1.08-2.62
	neural progenitors	1.3E-04	2.45	1.55-3.92
	differentiated tissues	9.9E-05	2.48	1.58-3.94
exonic	ESCs	7.6E-01	0.87	0.14-36.11
	neural progenitors	1.0E+00	1.57	0.23-67.33
	differentiated tissues	8.1E-01	2.04	0.33-84.53
intergenic	ESCs	3.1E-02	0.60	0.39-0.93
	neural progenitors	9.9E-05	0.40	0.25-0.63
	differentiated tissues	9.9E-05	0.39	0.24-0.60

Supplementary Table 12. For each group of samples we assessed, with Fisher's exact test, significant differences in the proportions of common *vs.* group-specific ELSs that overlap intronic, exonic and intergenic regions. *p*-value (FDR-corrected), odds ratio and confidence interval are reported for each test.

Group	Genes ∩ ELSs			
	Introns	Exons	Both	Total
ESCs	907 (89.27%)	21 (2.07%)	88 (8.66%)	1,016
neural progenitors	359 (87.56%)	13 (3.17%)	38 (9.27%)	410
differentiated tissues	492 (86.16%)	24 (4.20%)	55 (9.63%)	571
common	33 (82.50%)	1 (2.50%)	6 (15.00%)	40

Supplementary Table 13. Number of genes whose introns and/or exons intersect group-specific and common ELSs identified in embryonic samples.

Gene	Coordinates	ELS ID	Tissue	Location	Hosting	Peak ChIP-seq K27ac	Primer
PPP3CA	4:102023034-102023302	EH37E0737564	brain	intronic	host	neuron 1	F:GCCAACACTCGCTACCTCTT R:AAGGCCCACAAATACAGCAC
CAPZB	1:19670858-19672569	EH37E0073593	brain	intronic	host	ESC 1	F:CCTGGTCCCCAGTCTATGTG R:ACCACCTTGTCTCTGGCAAT
SDHD	11:113019264-113019880	EH37E0240118	brain	intronic	non-host	neuron 2	F:CAGAATGGTGTGGAGTGCAG R:AGTGGAGAGATGCAGCCTTG
AKR7A2	1:19670858-19672569	EH37E0073593	brain	intronic	non-host	ESC 1	F:GCCGAGATCTGTACCCTCTG R:GAAGAGCTCCGTTTCCACCT
RPA2	1:28307519-28308214	EH37E0078769	brain	intronic	non-host	neuron 1/2	F:CCTTCTCAAGCCGAAAAGAA R:TCATCAACCAAAGTGGCAGA
DRG1	22:31735086-31735685	EH37E0629324	brain	intronic	non-host	ESC 1	F:TTACTCCAAAGGGTGGTGGT R:CAAATCCAATTCGAGCATCA
CSPG5	3:47577542-47578617	EH37E0656905	brain	intergenic	non-host	ESC 1	F:CCACTGCTGCTGTTTCTGG R:CTGCCCTTCACCAGCTCTT
CTNNA2	2:80527433-80528886	EH37E0528734	brain	intronic	host	neuron 1	F:CAGAAAGGCTGTGCTGATGA R:CTTGTCCTGCTACGCACATC
KCNQ2	20:62086208-62086923	EH37E0609018	brain	intronic	host	ESC 1/2, neuron 1	F:CACAGGCAGAAGCACTTTGA R:GAGAGGTTGGTGGCGTAGAA
ACTB	7:5733031-5733564	EH37E0886351	common	intronic	non-host	all	F:ATTGGCAATGAGCGGTTC R:TGAAGGTAGTTTCGTGGATGC
BAHD1	15:40390946-40391339	EH37E0363650	common	intronic	non-host	all	F:GATGATGAGCCTCCTGTGGT R:GCGATGCAAACACTTCATTC
BMF	15:40390946-40391339	EH37E0363650	common	intronic	host	all	F:CAGTGCATTGCAGACCAGTT R:AAGGTTGTGCAGGAAGAGGA
GSK3A	19:40939210-40940400	EH37E0490611	common	intergenic	non-host	all	F:CTCATTTGGGGTCGTGTACC R:GATCTGCAGCTCTCGGTTCT
VEGFB	11:62320405-62321311	EH37E0221959	common	intronic	non-host	all	F:CTGGCCACCAGAGGAAAGT R:CATGAGCTCCACAGTCAAGG

Supplementary Table 14. Selection of brain-specific and common ENCODE ELSs overlapping with hESC-derived neural maturation ChIP-seq. Target genes were identified by Hi-C interaction, and only genes regulated by one ELS in our ENCODE analysis were selected. Specifically, we report: the ELSs coordinates and ID; the type of ELS (brain-specific or common, based on the classification derived from ENCODE adult samples); the ELS genomic location (intronic *vs.* intergenic); the nature of the targeted genes (host or non-host); the presence of peaks in the neural maturation ChIP-seq experiments; the primers used for gene expression analysis.

Group	GO term	Description
neural progenitors	BP: 0060291	long-term synaptic potentiation
	BP: 0050770	regulation of axonogenesis
	BP: 0097061	dendritic spine organization
	CC: 0008328	ionotropic glutamate receptor complex
	CC: 0098878	neurotransmitter receptor complex
	CC: 0014069	postsynaptic density
	MF: 0004970	ionotropic glutamate receptor activity
	MF: 0005089	rho guanyl-nucleotide exchange factor activity
	MF: 0008013	beta-catenin binding
differentiated tissues	BP: 1900020	positive regulation of protein kinase C activity
	BP: 1900040	regulation of interleukin-2 secretion
	BP: 0060766	negative regulation of androgen receptor signaling pathway
	CC: 0098651	basement membrane collagen trimer
	CC: 0098644	complex of collagen trimmers
	CC: 0005583	fibrillar collagen trimer
	MF: 0044548	S100 protein binding
	MF: 0035252	UDP-xylosyltransferase activity
	MF: 0030020	extracellular matrix structural constituent conferring tensile strength
ESCs	BP: 0042908	xenobiotic transport
	BP: 0045986	negative regulation of smooth muscle contraction
	BP: 0098698	postsynaptic specialization assembly
	CC: 0099092	postsynaptic density, intracellular component
	CC: 0031304	intrinsic component of mitochondrial inner membrane
	CC: 0008328	ionotropic glutamate receptor complex
	MF: 0008146	sulfotransferase activity
	MF: 0005547	phosphatidylinositol-3,4,5-triphosphate binding
	MF: 0070300	phosphatidic acid binding
common	CC: 0071565	nBAF complex
	CC: 0016514	SWI/SNF complex
	CC: 0070603	NI/SNF superfamily-type complex

Supplementary Table 15. Significantly enriched GO terms associated with the genes harboring intronic ELSs identified in embryonic samples. Only the top three enriched terms are shown in each analysis (BP: Biological Process; CC: Cellular Component; MF: Molecular Function).

Experiment ID	bigBed File ID	bigWig File ID	Biosample Term Name	Samples Cluster
ENCSR960EVO	ENCFF150IXD	ENCFF106UPY	aorta	aorta
ENCSR957BPJ	ENCFF778YRL	ENCFF786OKT	aorta	aorta
ENCSR984KWT	ENCFF096QKQ	ENCFF600PFS	thoracic aorta	aorta
ENCSR930HLX	ENCFF496WSG	ENCFF712GED	thoracic aorta	aorta
ENCSR939UQD	ENCFF006PUX	ENCFF526XGT	B cell	blood
ENCSR000DQR	ENCFF358FDS	ENCFF119SDL	B cell	blood
ENCSR878JSF	ENCFF911TMM	ENCFF804THE	B cell	blood
ENCSR000DQP	ENCFF946HAD	ENCFF831LDI	B cell	blood
ENCSR796FCS	ENCFF317WLK	ENCFF231OTU	CD14-positive monocyte	blood
ENCSR395YXN	ENCFF516NBV	ENCFF692YGS	T cell	blood
ENCSR570AUC	ENCFF876VUG	ENCFF428SHV	natural killer cell	blood
ENCSR206JRX	ENCFF215JSM	ENCFF796FFT	peripheral blood mononuclear cell	blood
ENCSR443SLY	ENCFF274NFA	ENCFF573QMJ	peripheral blood mononuclear cell	blood
ENCSR368YPC	ENCFF541SEP	ENCFF303YKC	peripheral blood mononuclear cell	blood
ENCSR275EAG	ENCFF641SRH	ENCFF482WUA	peripheral blood mononuclear cell	blood
ENCSR535XRY	ENCFF434HTE	ENCFF624ECX	angular gyrus	brain
ENCSR057RET	ENCFF471CNA	ENCFF586RDL	angular gyrus	brain
ENCSR486QMV	ENCFF413UOM	ENCFF433GYN	caudate nucleus	brain
ENCSR840KVX	ENCFF922OIZ	ENCFF284JOB	caudate nucleus	brain
ENCSR032BMQ	ENCFF835HDX	ENCFF893IET	cingulate gyrus	brain
ENCSR693GVU	ENCFF883GXX	ENCFF478BHA	cingulate gyrus	brain
ENCSR383AEO	ENCFF082JYH	ENCFF080XAQ	layer of hippocampus	brain
ENCSR956CFX	ENCFF572YMX	ENCFF266RWF	layer of hippocampus	brain
ENCSR418JIS	ENCFF764ZYN	ENCFF226DRF	layer of hippocampus	brain
ENCSR157EML	ENCFF008YZE	ENCFF438HBY	middle frontal area 46	brain
ENCSR401VZL	ENCFF539MAM	ENCFF476LCL	middle frontal area 46	brain
ENCSR551QXE	ENCFF230LRO	ENCFF531LTO	substantia nigra	brain
ENCSR883QMZ	ENCFF547SJD	ENCFF277PPF	substantia nigra	brain
ENCSR717AJD	ENCFF405ITQ	ENCFF625DED	temporal lobe	brain
ENCSR477BHF	ENCFF939UVF	ENCFF565UAK	temporal lobe	brain
ENCSR876DCP	ENCFF102LZH	ENCFF527JEF	body of pancreas	digestive
ENCSR588PZN	ENCFF447CFW	ENCFF867LQS	body of pancreas	digestive
ENCSR554RQQ	ENCFF485WGE	ENCFF982WVN	body of pancreas	digestive
ENCSR747VED	ENCFF703WUC	ENCFF623TEM	pancreas	digestive
ENCSR315LPR	ENCFF718GRW	ENCFF779EMW	pancreas	digestive
ENCSR063HOI	ENCFF205VUE	ENCFF752TMT	stomach	digestive
ENCSR489ZLL	ENCFF210MKU	ENCFF244PVV	stomach	digestive
ENCSR129NCV	ENCFF248SFF	ENCFF706PVF	stomach	digestive
ENCSR843UEZ	ENCFF350QEB	ENCFF230GOG	stomach	digestive
ENCSR492BHN	ENCFF892PIW	ENCFF647UUQ	stomach	digestive
ENCSR000DWZ	ENCFF219NNT	ENCFF672JJF	fibroblast of lung	fibro/myoblasts
ENCSR000AMW	ENCFF307BRU	ENCFF648WWQ	fibroblast of lung	fibro/myoblasts
ENCSR915QOL	ENCFF498GDR	ENCFF837XME	fibroblast of lung	fibro/myoblasts
ENCSR000ANZ	ENCFF104CCE	ENCFF731WEZ	myotube	fibro myoblasts
ENCSR000ANK	ENCFF286OEL	ENCFF020MPQ	skeletal muscle myoblast	fibro/myoblasts
ENCSR596NOF	ENCFF881WJB	ENCFF211KTI	skeletal muscle myoblast	fibro/myoblasts
ENCSR505JQC	ENCFF779COY	ENCFF637LKZ	iPS-18a	iPSCs
ENCSR989RAL	ENCFF206LCO	ENCFF500GFH	iPS-20b	iPSCs
ENCSR322FGP	ENCFF817MAX	ENCFF956GBB	colonic mucosa	mucosa
ENCSR577DVK	ENCFF874PAW	ENCFF202MXN	colonic mucosa	mucosa
ENCSR276BXF	ENCFF059ORI	ENCFF983WKA	mucosa of rectum	mucosa
ENCSR146DAL	ENCFF911MLR	ENCFF174ZGI	mucosa of rectum	mucosa

Experiment ID	bigBed File ID	bigWig File ID	Biosample Term Name	Samples Cluster
ENCSR098OLN	ENCFF036BUW	ENCFF203DXZ	gastrocnemius medialis	skeletal/cardiac muscle
ENCSR206STN	ENCFF182VGO	ENCFF074GLQ	gastrocnemius medialis	skeletal/cardiac muscle
ENCSR785DJD	ENCFF801JCZ	ENCFF995MET	gastrocnemius medialis	skeletal/cardiac muscle
ENCSR972ETR	ENCFF828XLK	ENCFF596MDR	gastrocnemius medialis	skeletal/cardiac muscle
ENCSR548LZS	ENCFF321AWE	ENCFF358SLL	right cardiac atrium	skeletal/cardiac muscle
ENCSR767NIF	ENCFF159MIF	ENCFF364HCQ	skeletal muscle tissue	skeletal/cardiac muscle
ENCSR346KKE	ENCFF705TXS	ENCFF534NZT	skeletal muscle tissue	skeletal/cardiac muscle
ENCSR703CYD	ENCFF040JUU	ENCFF149LVI	subcutaneous abdominal adipose tissue	skeletal/cardiac muscle
ENCSR294PQF	ENCFF100ODV	ENCFF245YFA	subcutaneous abdominal adipose tissue	skeletal/cardiac muscle
ENCSR605NNZ	ENCFF140MBX	ENCFF076YGE	subcutaneous abdominal adipose tissue	skeletal/cardiac muscle
ENCSR501FTL	ENCFF237JIN	ENCFF487IMQ	subcutaneous abdominal adipose tissue	skeletal/cardiac muscle
ENCSR923IIU	ENCFF578DGE	ENCFF711BWR	subcutaneous abdominal adipose tissue	skeletal/cardiac muscle
ENCSR075PTL	ENCFF744GGI	ENCFF679MIG	muscle layer of duodenum	smooth muscle
ENCSR264APD	ENCFF997ISM	ENCFF180UOM	muscle layer of duodenum	smooth muscle
ENCSR953GFW	ENCFF455LOH	ENCFF225DFI	rectal muscle smooth tissue	smooth muscle
ENCSR168PQI	ENCFF127NZS	ENCFF734LFM	stomach muscle smooth	smooth muscle
ENCSR532FEO	ENCFF931EAQ	ENCFF143PWQ	stomach muscle smooth	smooth muscle
ENCSR647HAQ	ENCFF550ZZM	ENCFF674GXQ	vagina	smooth muscle
ENCSR258UUX	ENCFF707ODD	ENCFF071GTI	vagina	smooth muscle

Supplementary Table 16. List of H3K4me3 ChIP-seq experiments used to perform the analyses in Supplementary Figs. 6A and 6C (see also "re: enrichment of putative alternative promoters in intronic and intergenic tissue-specific ELSs"). The accession numbers (Experiment ID, bigBed file ID, bigWig File ID) allow to uniquely identify the experiment and corresponding files on the ENCODE portal (`https://www.encodeproject.org/`).

Experiment ID	bigBed File ID	bigWig File ID	Biosample Term Name	Samples Cluster
ENCSR715KGX	ENCFF297HKN	ENCFF493TPV	adrenal gland	differentiated tissues
ENCSR442ZOI	ENCFF766JNH	ENCFF634XSS	hepatocyte	differentiated tissues
ENCSR413QXO	ENCFF855JOO	ENCFF191CFV	large intestine	differentiated tissues
ENCSR128QKM	ENCFF783UIQ	ENCFF849QUA	muscle of leg	differentiated tissues
ENCSR714SGY	ENCFF752MYO	ENCFF703OHL	muscle of trunk	differentiated tissues
ENCSR237QFJ	ENCFF749ZSB	ENCFF768LAW	small intestine	differentiated tissues
ENCSR515PKY	ENCFF728OWX	ENCFF394FMW	smooth muscle cell	differentiated tissues
ENCSR202RXT	ENCFF162TRZ	ENCFF690GLN	stomach	differentiated tissues
ENCSR308ZMD	ENCFF017BXO	ENCFF384FQP	thymus	differentiated tissues
ENCSR922CAT	ENCFF980JXF	ENCFF075WFF	mid-neurogenesis radial glial cells	neural progenitors
ENCSR608VNA	ENCFF123NTR	ENCFF379BNK	neural cell	neural progenitors
ENCSR661MUS	ENCFF963JVR	ENCFF644BAH	neural progenitor cell	neural progenitors
ENCSR662PLB	ENCFF655SBQ	ENCFF915BXI	neuroepithelial stem cell	neural progenitors
ENCSR433PUR	ENCFF791AVT	ENCFF038HSI	radial glial cell	neural progenitors
ENCSR003SSR	ENCFF063RLE	ENCFF168UGH	H1	ESCs
ENCSR000AMG	ENCFF127WKA	ENCFF044CKA	H1	ESCs
ENCSR443YAS	ENCFF451DZQ	ENCFF065VIF	H1	ESCs
ENCSR814XPE	ENCFF798FMO	ENCFF347APS	H1	ESCs
ENCSR019SQX	ENCFF908LKM	ENCFF742QHK	H1	ESCs
ENCSR716ZJH	ENCFF044DDA	ENCFF796WXE	H9	ESCs
ENCSR043VGU	ENCFF636FLM	ENCFF494BBO	H9	ESCs
ENCSR153SGD	ENCFF904TVW	ENCFF930LWY	HUES48	ESCs
ENCSR176ABZ	ENCFF126FDP	ENCFF112VVW	HUES6	ESCs
ENCSR894OYM	ENCFF498ZKF	ENCFF854OTV	HUES64	ESCs

Supplementary Table 17. List of H3K4me3 ChIP-seq experiments used to perform the analyses in Supplementary Figs. 6B and 6D (see also "re: enrichment of putative alternative promoters in intronic and intergenic tissue-specific ELSs"). The accession numbers (Experiment ID, bigBed file ID, bigWig File ID) allow to uniquely identify the experiment and corresponding files on the ENCODE portal (`https://www.encodeproject.org/`).