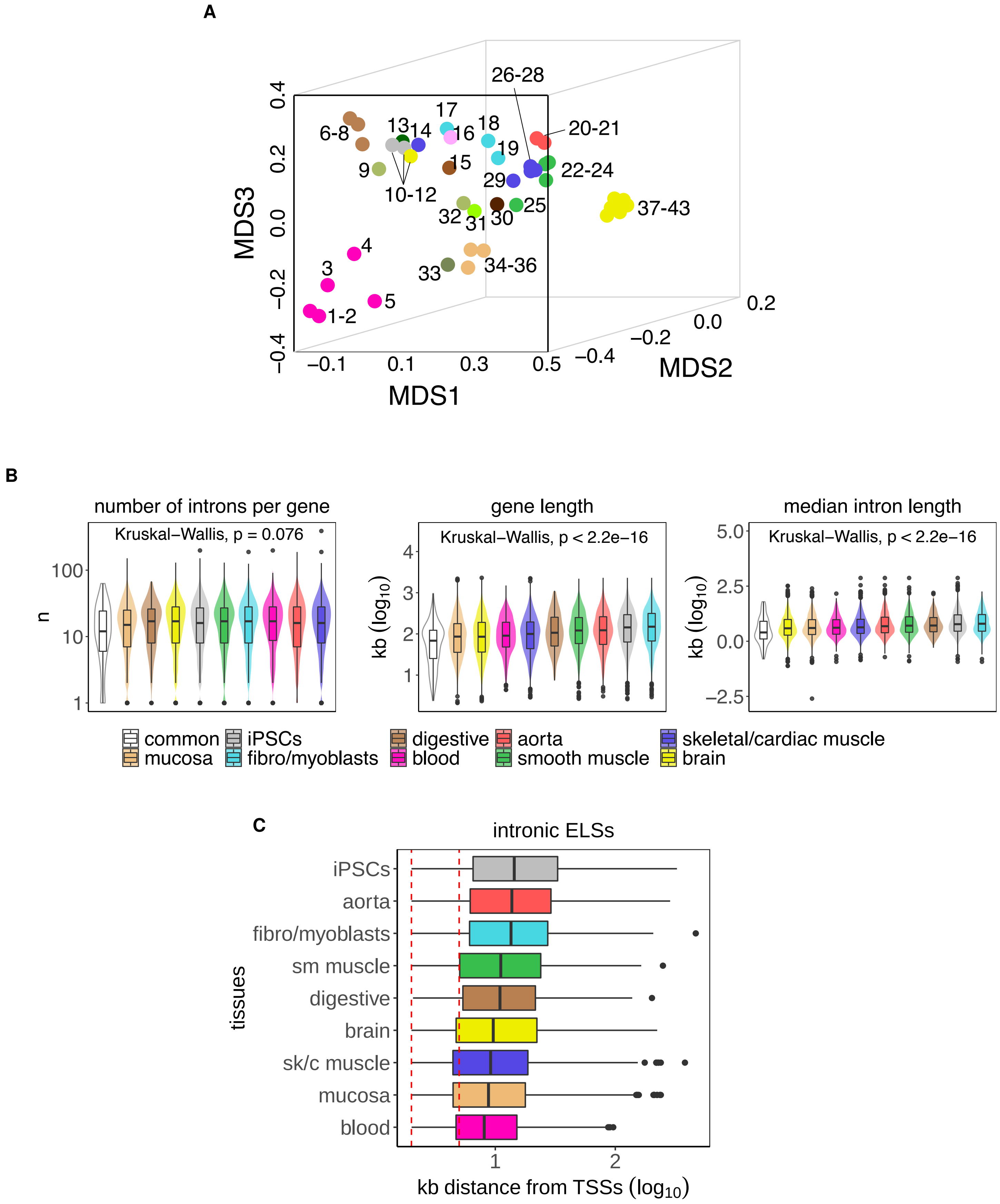
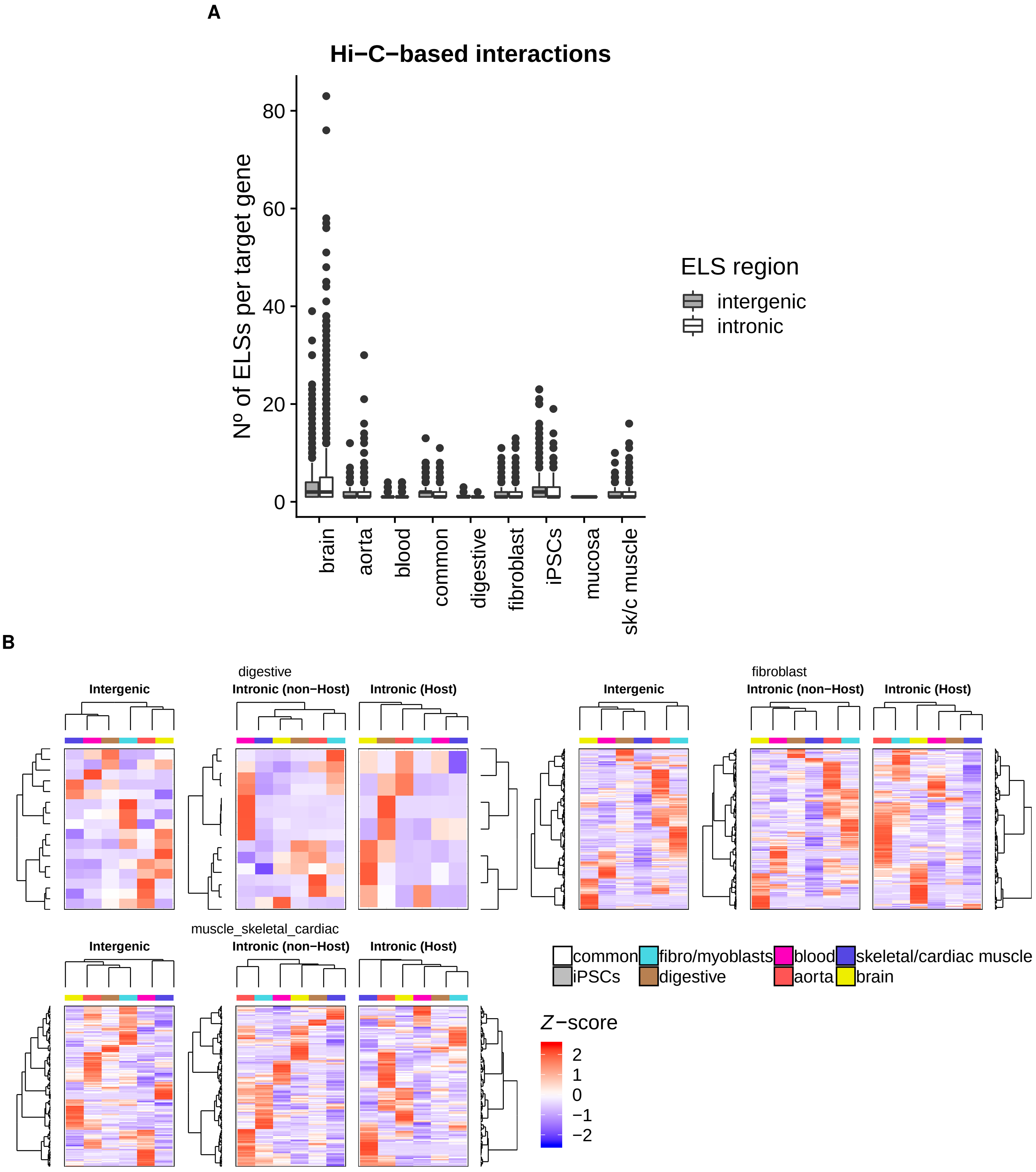
Supplementary Figures and Tables

List of Figures

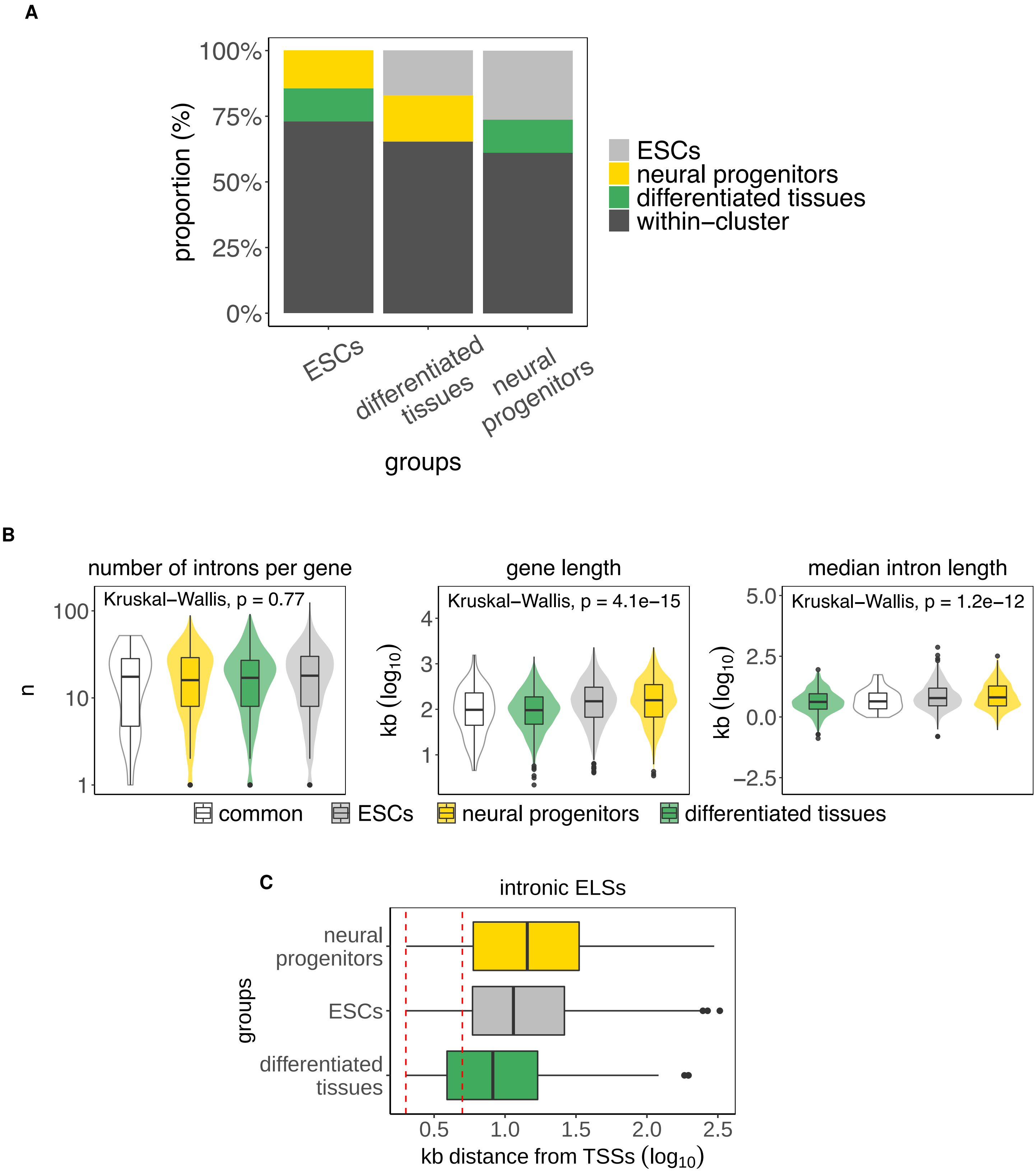
List of Tables



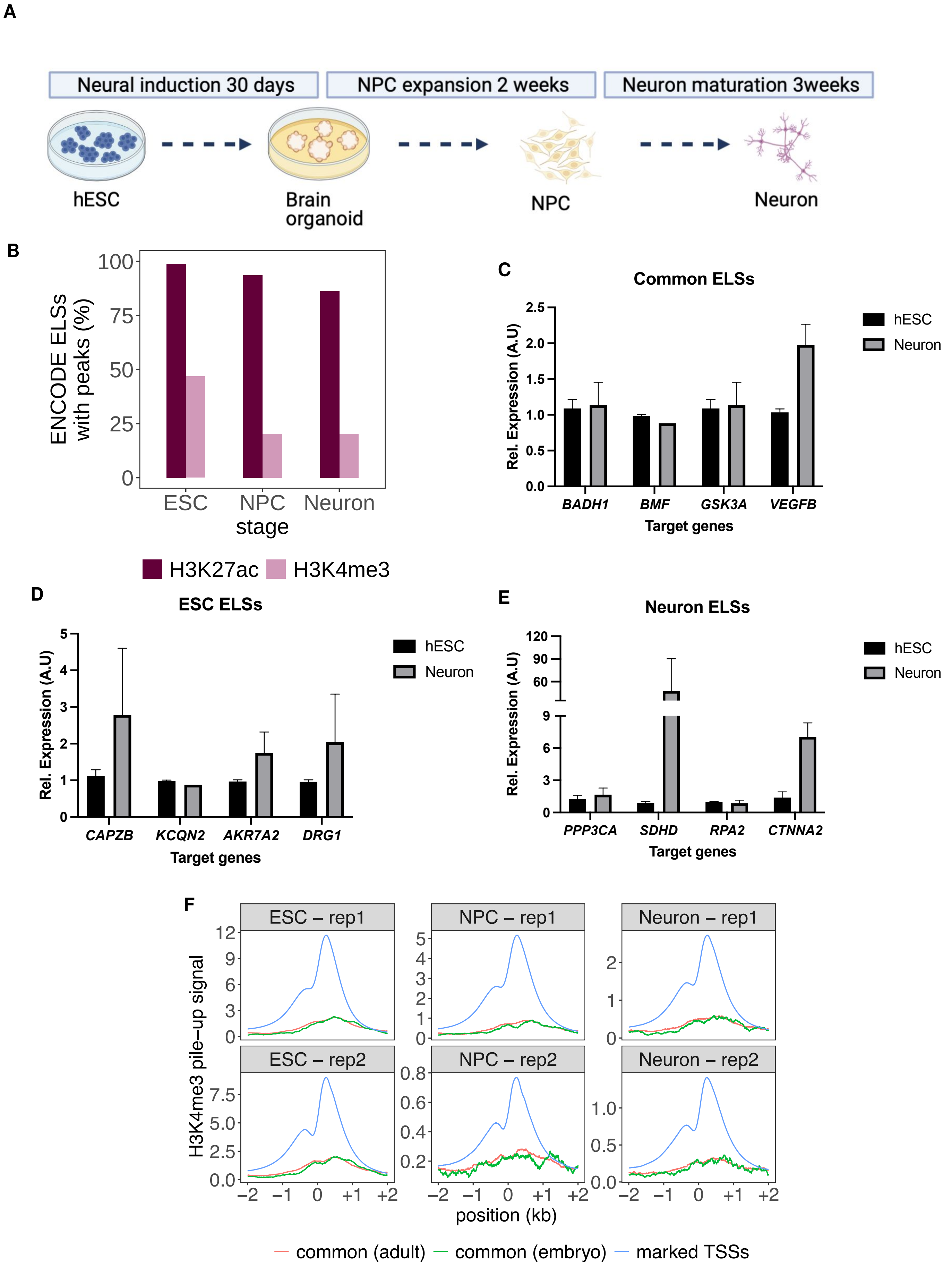
Supplementary Fig. 1. **A**: Multidimensional scaling (MDS) representation of the dissimilarities among the 43 human adult samples based on the pattern of activity of ELS-cCREs. The binary distance between a given pair of samples was computed considering presence/absence vectors of 921,166 distal ELSs (+/- 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.



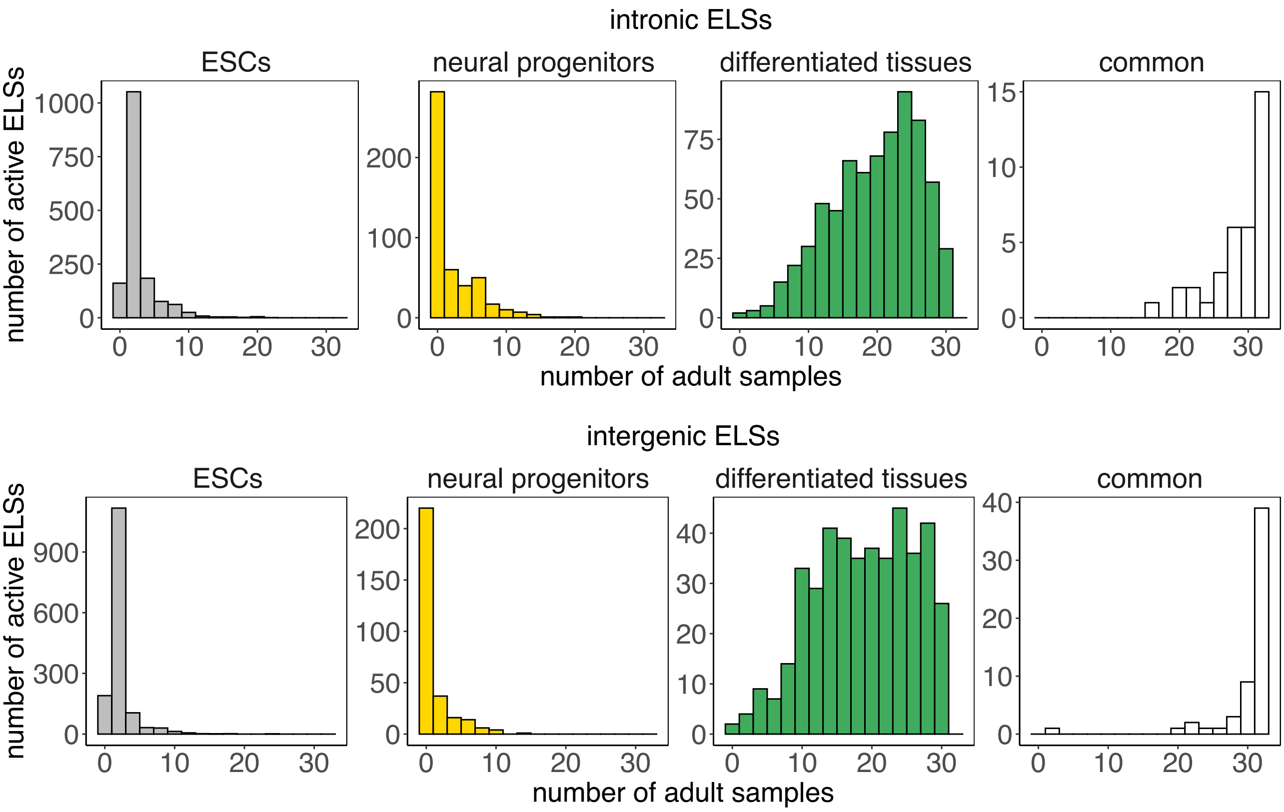
Supplementary Fig. 2. **A**: Distributions of the number of common and tissue-specific ELSs regulating a single target gene according to the Hi-C-based data. sk/c = skeletal/cardiac. **B**: *Z*-score normalized median gene expression across GTEx tissue categories of the genes targeted by intergenic and 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).



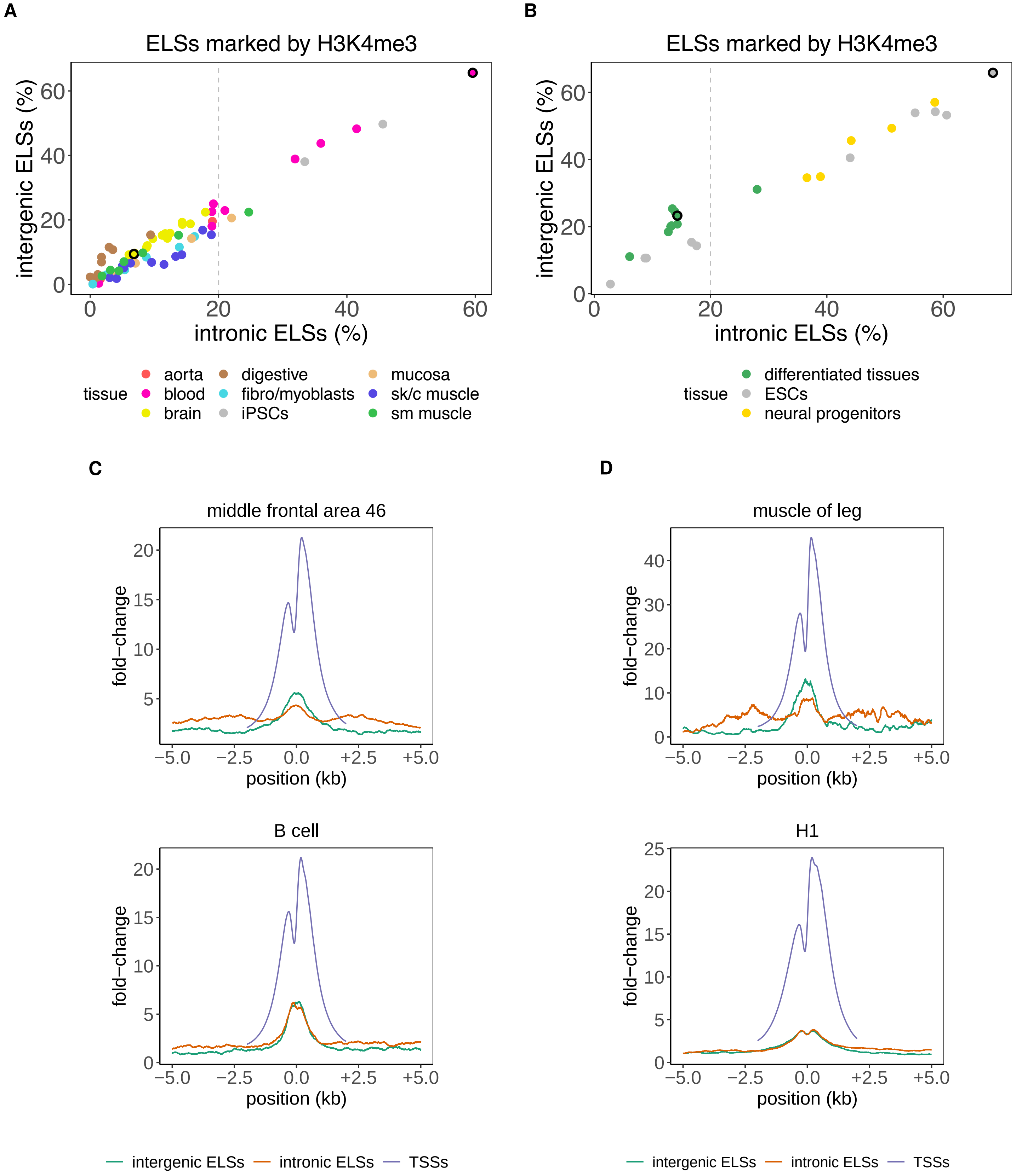
Supplementary Fig. 3. **A**: Group-specific ELSs in embryonic samples (analogous to Fig. 1D). The barplot represents the type of outer samples observed within sets of ESCs-, differentiated tissues- and neural progenitors-specific ELSs. **B**: Features of genes hosting either common or specific intronic 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.



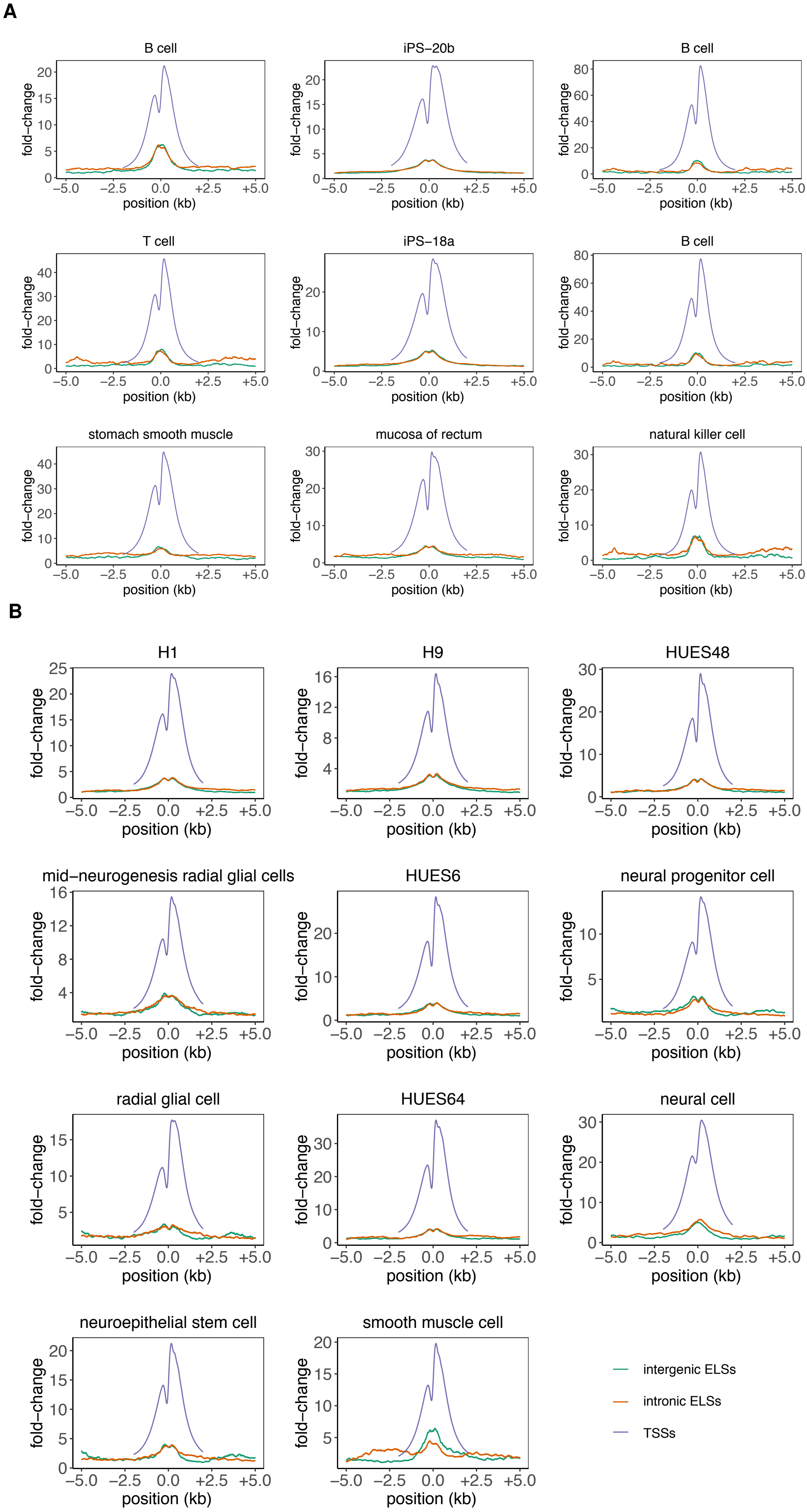
Supplementary Fig. 4. **A**: Scheme depicting the differentiation protocol of ESCs into NPCs and neurons. **B**: Overlap between ENCODE embryonic common ELSs and ChIP-seq peaks of H3K27ac and H3K4me3 observed in ESCs, NPCs and neurons. **C-E**: Gene expression analysis in hESC and ESC-derived neurons 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

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/`).