

Supplementary Materials.

Len A. Pennacchio, Gabriela G. Loots, Marcelo A. Nobrega and Ivan Ovcharenko

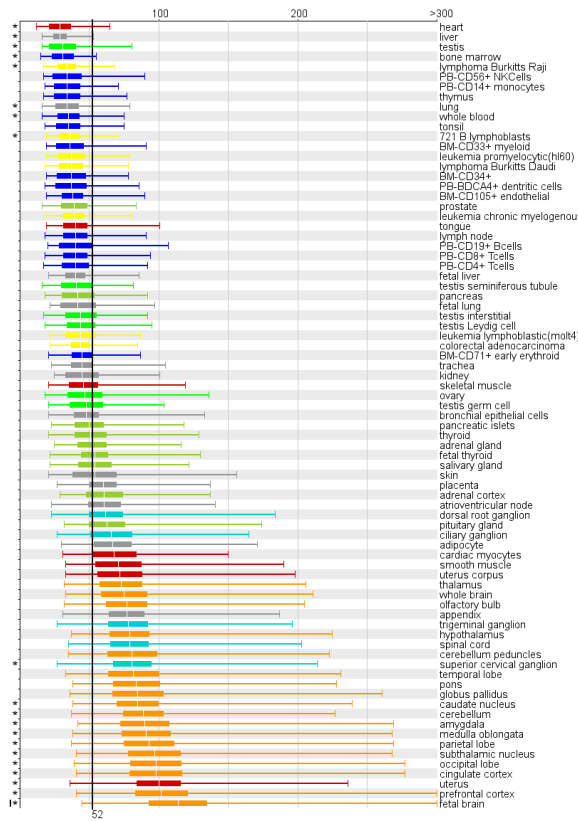
Predicting Tissue-Specific Enhancers in the Human Genome

Supplementary *EI* database.

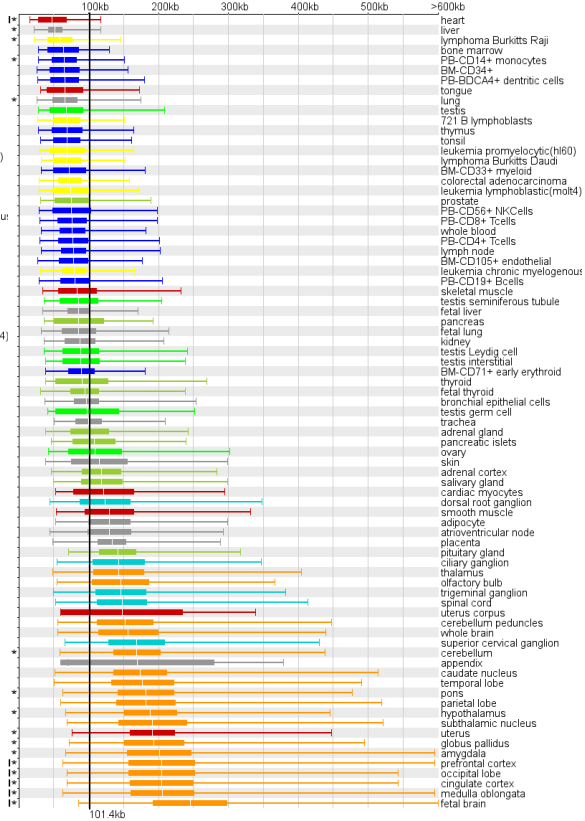
We have compiled the optimization information on tissue-specific scoring of candidate enhancers as well as other underlying datasets at <http://www.dcode.org/EI>. This database contains positional and annotation information for 7,187 human and 5,585 mouse genome tissue-specific candidate enhancers. It also includes tissue-specificity annotation of Gumby (Prabhakar et al. 2006) candidate enhancers and all human-mouse noncoding conserved elements in the human genome.

Figure S1. Analysis of all available human and mouse tissues. Number of noncoding ECRs (A – human, C – mouse) and locus length (B – human, D – mouse) of genes highly expressed in different tissues. Bound horizontal lines represent the inter-quartile range (the distance between the 25th and 75th percentiles) of tissue-specific distributions, solid colored rectangles measure the standard error in median calculations (white lined inside colored rectangles). Statistically significant ($<5\%$) distributions that deviate from the global median (represented by a solid vertical black line) are marked by an asterisk on the left side bar. Tissues where the median value is 2-fold smaller or larger than the median are marked by a vertical line on the left side bar. Panel E depict a correlation in number of noncoding ECRs (ncECR) plotted against the locus length for different tissues.

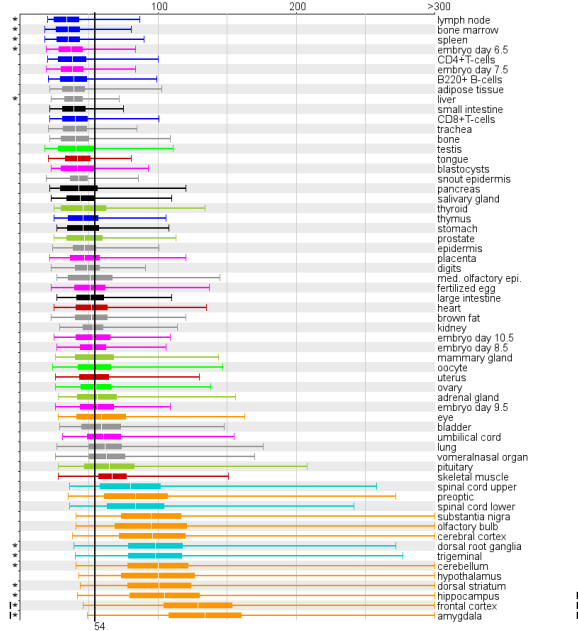
A



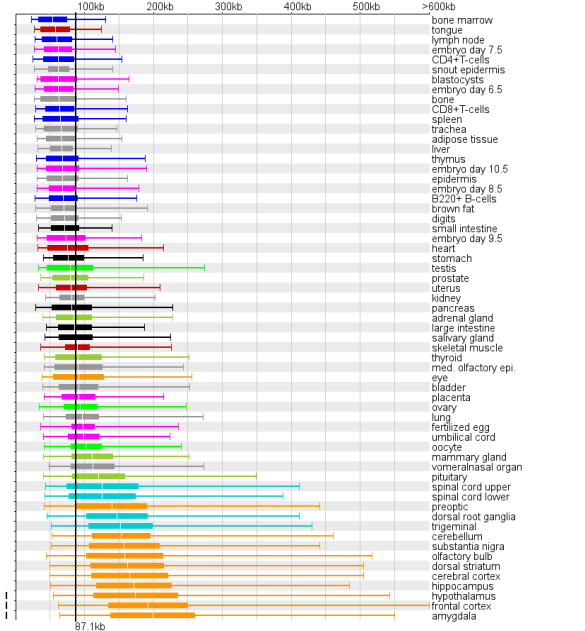
B



C



D



E

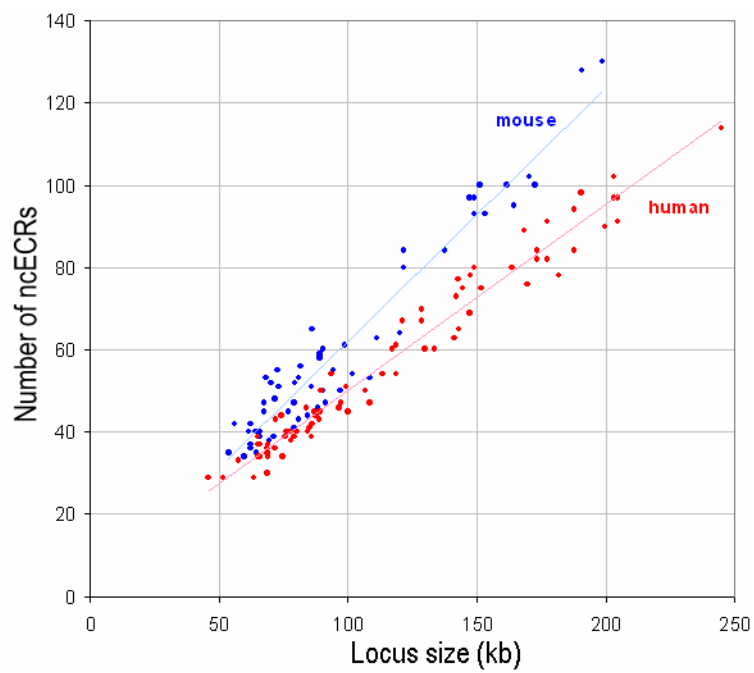
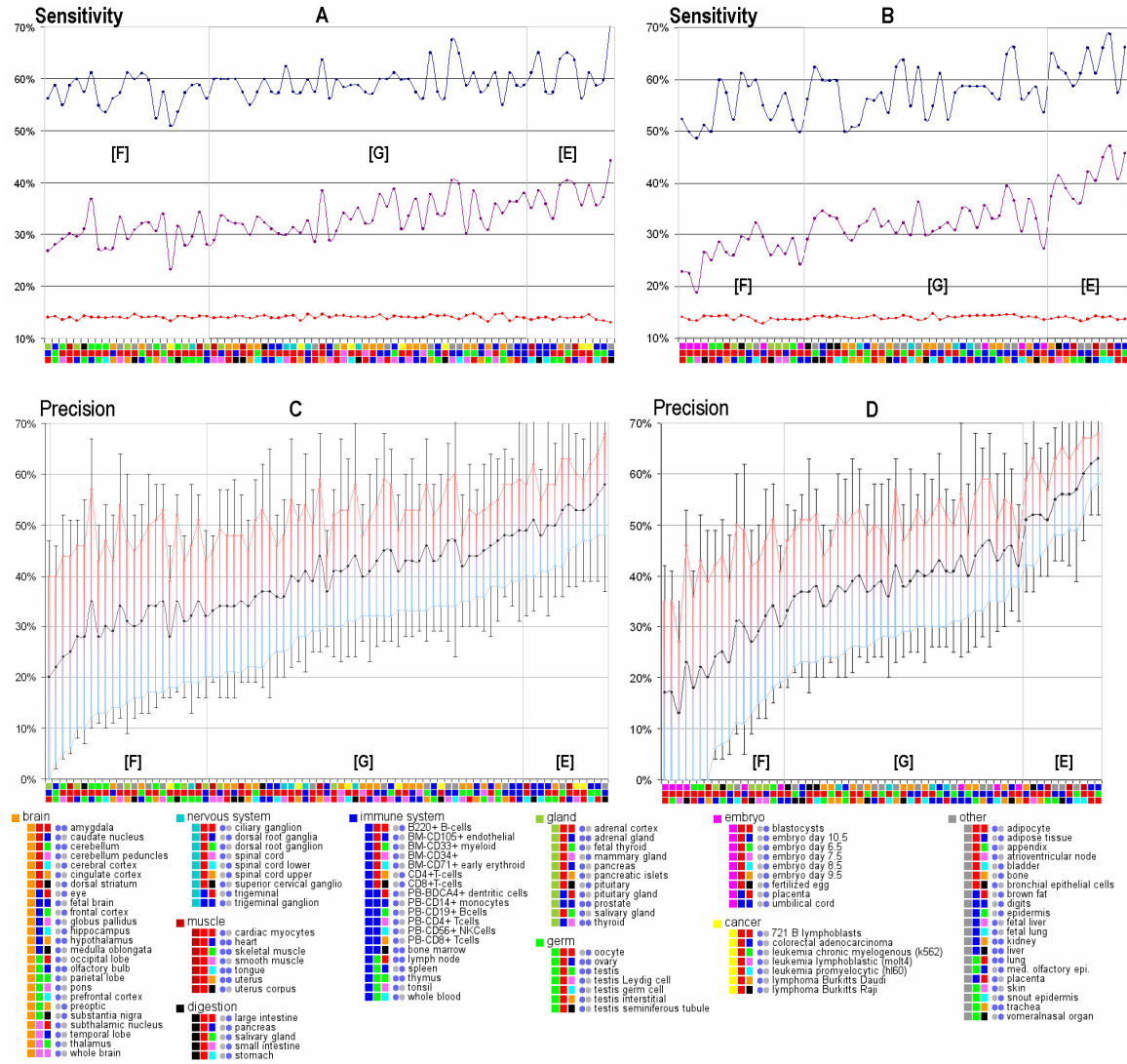


Figure S2. Sensitivity (A,B) and precision (C,D) of the *EI* method in recognizing human (A,C) and mouse (B,D) tissue-specific enhancers. Blue and red curves on sensitivity plots measure the percentage of highly and lowly expressed gene loci with tissue-specific enhancers, respectively; while the red curve estimates *EI* sensitivity for *de novo* enhancer recognition. Lower and upper bound estimates of precision along with their average are given in red, blue, and black on precision plots, respectively. Standard deviation is also depicted for each lower and upper bound estimate. Tissues that are plotted along the horizontal axis of these graphs are color coded; the legend is present at the bottom of the figure. Blue and gray circles indicate the presence or absence of a tissue, respectively in either the human (left) or the mouse (right) data. Tissues are split into three groups based on the lower-bound estimate of the corresponding *EI* method precision in enhancer recognition - F (poor ; less than 20%), G (good; from 20% to 40%), and E (excellent; over 40%).

Figure S2.



Enrichment of individual TFBS in tissue-specific candidate enhancers.

We found that several individual TF motifs were significantly enriched in 43 human and 29 mouse tissues yielding 27K human and 29K mouse TF-tissue associations (Table S1). Many of these associations (including AP-4 and OCT-1 in brain; ETS and ELK1 in T- and B-cells) were consistent with previously reported TF enrichments in loci of tissue-specific gene expression (Blanchette et al. 2006) despite the use of a different approach to identify candidate enhancers utilized by the referenced study. Several associations reported here are novel or add a member to a known cohort of synergistically acting TFs. For example, we observed a strong association among MEF2, SRF, MyoD, Myogenin, and LBP1 (aka UPB1) TFBSs in candidate enhancers corresponding to mouse genes highly expressed in skeletal muscle (Table S1B). While the first four of these TFs have been shown to play key roles during muscle development (Li et al. 2005; Naya and Olson 1999), little is known about LBP1 except that it regulates extraembryonic angiogenesis (Parekh et al. 2004). Consistent with previously reported results (Blanchette et al. 2006), we demonstrate that enrichment of TFs can potentially act as an indicator of TF-tissue association, but the presence or absence of a TFBS does not represent a reliable signature of tissue specificity. For example, while MEF2 is strongly linked to high expression in skeletal muscle it is also linked to high expression profiles in 17 additional tissues. Similarly, while SP1 is the primary candidate for high expression in cardiac myocytes (SP1 is known to be a key regulator of expression in these cells (Azakie et al. 2006)), SP1 putative TFBS are present in 57% of the loci of highly expressed genes and 43% of the loci of lowly expressed genes, in cardiac myocytes. Therefore SP1 putative TFBS fail to

provide an accurate indicator whether a locus is linked to cardiac myocytes or not, based selectively on the presence or absence of SP1 binding sites. We also find that about half of tissues (including liver and kidney) do not display a significant increase in the density of any individual TF. Therefore the cooperative analysis of multiple TFs is required to describe the tissue-specific architecture of gene regulation; supporting the idea that gene regulation is directed by multiple TFs that synergistically bind to arrays of heterogeneous TFBS.

Table S1. List of TFs with enriched densities of putative TFBS in human (A) and mouse (B) loci of highly expressed genes ($p < 0.05$). Up to 10 enriched TFs with the lowest p -values are listed (ordered by an increasing p -value). No TF enrichment was detected for 38 human and 32 mouse tissues.

A)

Human tissue	List of TFs
721 B lymphoblasts	E2F4DP2
PB-CD19+ Bcells	ETS, ELK1
PB-CD4+ Tcells	ETS1, ELK1
PB-CD56+ NKCells	ETS
PB-CD8+ Tcells	ELK1, CETS1P54, ETS1
adipocyte	NF1, CEBPB
amygdala	MEF2, MAZ, MZF1, CREBATF, ZF5, AP2, NRF1, MAZR, E2F1, KROX
appendix	POU3F2, EVI1, PEBP, MEF2, COREBINDINGFACTOR, OSF2, AP1, GATA3, GFII1, CART1
atrioventricular node	BEL1, VMAF
cardiac myocytes	SP1, ZF5, E2F1
caudate nucleus	KROX, MAZ, AP2, HIC1, MEF2, MZF1, ZF5, NGFIC, EGR3, UF1H3BETA
cerebellum	MAZR, AP2, KROX, UF1H3BETA, MAZ, HIC1, TFIII, ZF5, TAL1ALPHA47, EGR
cerebellum	KROX, UF1H3BETA, AP2, MAZ, CACBINDINGPROTEIN, ZF5, SP1, MAZR, GC, EGR
peduncles	
ciliary ganglion	TAL1ALPHA47, IPF1, TAL1BETA47, HNF6, CDP, NKX61, TAL1BETAITF2, HNF3B, HNF3ALPHA, CDXA
cingulate cortex	MAZ, MZF1, CACBINDINGPROTEIN, AP2, NRSF, KROX, CREBATF, MAZR, AP1, MEF2
colorectal	ZF5, SP1
adenocarcinoma	
fetal brain	NRF1, OCT, MZF1, MAZR, MEF2, TAXCREB, OCT1, E2F1, TCF11MAFG, AHR
fetal thyroid	CACBINDINGPROTEIN, AP2GAMMA, SP1
globus pallidus	AP2, MZF1, TCF11MAFG, MEF2, PAX2, MAZ, NRSF, CACBINDINGPROTEIN, CREBATF, AP1
Heart	AP2GAMMA
Hypothalamus	MEF2, MAZ, AP1, ATF4, MZF1, CREBATF, BACH2, ZIC2, VJUN, AP2
leukemia	E2F4DP2, E2F1DP2

lymphoblastic(molt4)	
Lung	CACBINDINGPROTEIN, AP2GAMMA, NFKAPPAB50
lymph node	ETS
medulla oblongata	MEF2, MAZ, P53_DECAMER, MZF1, AP2, TAXCREB, NGFIC, EGR2, PAX3, EGR1
occipital lobe	AP2, MAZ, MEF2, MZF1, CREBATF, AP2ALPHA, MAZR, NGFIC, P300, AP1
olfactory bulb	PAX4, SRF, PAX8, LEF1TCF1, HIC1, NFKB, AP1, CACCCBINDINGFACTOR_Q, PAX3, MAZ
Ovary	RP58, PXR, E4BP4, AFP1
parietal lobe	MEF2, CACBINDINGPROTEIN, MZF1, AP2, MAZ, ZIC3, CREBATF, MAZR, ATF4, PAX3
pituitary gland	CREBATF, CREB, CHOP, VBP, TCF11, AP1, HNF6, CLOX
Pons	MAZ, MZF1, MEF2, ATF4, CACBINDINGPROTEIN, UF1H3BETA, MYOGENIN, AP2, LEF1, LMO2COM
prefrontal cortex	KROX, NGFIC, AP2, MAZ, MZF1, MEF2, CACBINDINGPROTEIN, CREBATF, ZF5, EGR1
skeletal muscle	RSRFC4, MEF2
smooth muscle	SP1, AP2, CETS1P54, ETS2, COREBINDINGFACTOR, ETS
spinal cord	MAZ, SP1, MZF1, MEF2, SREBP, AP2ALPHA, AP2, ZIC2, CACBINDINGPROTEIN, UF1H3BETA
subthalamic nucleus	AP2, MAZ, CACBINDINGPROTEIN, MEF2, MZF1, NRSE, CREBATF, VMAF, MAZR, UF1H3BETA
superior cervical ganglion	TATA, RP58, GATA3, TCF4, HNF1, OCT1, DTYPEPA, CEBP, FXR, LEF1TCF1
temporal lobe	MZF1, MAZ, KROX, AP2, CACBINDINGPROTEIN, MEF2, MAZR, CREBATF, TAXCREB, UF1H3BETA
Thalamus	MAZ, SP1, KROX, CACBINDINGPROTEIN, MZF1, UF1H3BETA, HIC1, ZF5, MAZR, MEF2
trigeminal ganglion	EVII, HNF6, TATA, GATA3, HLF, CDP, LEF1, LEF1TCF1, IRF, IPF1
Uterus	SRF, COREBINDINGFACTOR, FREAC2, FOXJ2, FOXO3, FOXO1, HNF6, STAT, HNF1, MEF2
uterus corpus	COREBINDINGFACTOR, PEBP, HNF3, STAT, GATA, NFAT, GATA1, AP1, NF1, GATA3
whole brain	MZF1, HIC1, KROX, AP2, MAZ, MAZR, CACBINDINGPROTEIN, ZF5, NRF1, SP1

B)

Mouse tissue	List of TFs
B220+ B-cells	ETS, PU1, ETS1, CETS1P54
CD4+T-cells	PU1, ETS
amygdala	YY1, AP2, OCT1, RFX1, HNF6, ZIC2, PAX8, MZF1, MAZ, E2F
cerebellum	AP2, SP1, E2F, GC, MTF1, MZF1, MAZ, YY1, CHCH, ZIC3
cerebral cortex	AP2, MAZ, KROX, ZF5, MZF1, NRSE, E2F, E2F1, VDR, EGR
dorsal root ganglia	SP1, SRF, YY1, HNF6, MMEF2, OCT1, HOXA3, GC, NRF1, NF1
dorsal striatum	AP2, UF1H3BETA, CHCH, MAZ, RFX1, KROX, AP2ALPHA, MZF1, MTF1, SPZ1
embryo day 10.5	E2F1, E2F1DP1RB, E2F4DP2, E2F4DP1, E2F1DP2, E2F, E2F1DP1, CACBINDINGPROTEIN, MAZ, UF1H3BETA
embryo day 8.5	MAZR, MAZ, E2F1, E2F1DP2, E2F, E2F4DP1, E2F1DP1RB, UF1H3BETA, CACBINDINGPROTEIN, ARP1
embryo day 9.5	E2F4DP2, E2F1DP2, E2F1, E2F, E2F4DP1, CACBINDINGPROTEIN, MAZ, E2F1DP1RB, ZF5, E2F1DP1
epidermis	MYOGENIN
eye	MTATA, CRX, OCT
frontal cortex	YY1, AP2, CREB, HNF6, MAZ, PAX8, NGFIC, CREBP1, ZIC2, RFX1
heart	RSRFC4, MEF2
hippocampus	AP2, MAZ, UF1H3BETA, CACBINDINGPROTEIN, E2F, KROX, SP1, VDR, MZF1, ZF5
hypothalamus	YY1, HNF6, EVII, ZIC2, SRF, RFX1, AP2, OCT1, POU3F2, MAZ
lymph node	PU1
olfactory bulb	MAZ, SP1, E2F1, AP2, ZF5, RFX1, CHCH, E2F, KROX, NGFIC
pituitary	YY1, GLI, AHRARNT, BARBIE, NRSE
preoptic	AP2, MAZ, YY1, GC, SP1, MZF1, HIC1, MAZR, CHCH, ZIC2
skeletal muscle	RSRFC4, MEF2, MYOGENIN, HMEF2, LBP1, MMEF2, MYOD, SRF
small intestine	DR1
snout epidermis	MYOGENIN
spinal cord lower	MAZ, AP2, SP1, GC, YY1, MAZR, GLI, ZIC1, ZIC2, TFIII
spinal cord upper	MAZ, YY1, ZIC2, GLI, AP2, SRF, MAZR, TFIII, ZIC3, ZIC1
substantia nigra	AP2, MAZ, YY1, GLI, MTF1, CREB, HNF6, ZIC2, SP1, MZF1
tongue	MYOGENIN
trigeminal	PBX1, CBF, SRF, NF1, CDPCR1, YY1, HNF6, OCT1, VMYB, TEF
umbilical cord	SP1, AP1, CACBINDINGPROTEIN

Table S2. Top 10 TFs with the highest tissue-specificity importance, in descending order.

A – human, B – mouse.

A)	
Tissue	List of TFs
721 B lymphoblasts	NRF2, ETS2, STAT, ELK1, CETS168, NFY, YY1, STAT6, CETS1P54, IRF
BM-CD105+ endothelial	E2F1DP1RB, E2F4DP1, NFY, E2F4DP2, CAAT, PR, ZTA, NRSF, AR, HES1
BM-CD33+ myeloid	TEL2, PEA3, ICSBP, CETS1P54, ETS2, ETS, NKX25, ISRE, NFKAPPAB50, CEBPB
BM-CD34+	E2F, E2F4DP2, STAF, NFMUE1, ELF1, CETS1P54, ELK1, NFY, CEBP, MEIS1
BM-CD71+ early erythroid	E2F1DP1, NFY, E2F1DP1RB, E2F4DP2, GATA, NKX61, LXR, E2F4DP1, NF1, NRSF
PB-BDCA4+ dendritic cells	PEA3, ISRE, CREB, CETS1P54, ELF1, COREBINDINGFACTOR, ELK1, CETS168, DEAF1, IRF2
PB-CD14+ monocytes	PU1, PEA3, ELK1, ETS, CETS1P54, TEL2, ETS2, ISRE, NERF, NFMUE1
PB-CD19+ Bcells	CETS1P54, TEL2, ETS, PU1, YY1, ELK1, PEA3, NRF2, ELF1, OCT
PB-CD4+ Tcells	ELK1, ETS1, PEA3, ICSBP, NRF2, NERF, TAXCREB, CBF, USF, CETS168
PB-CD56+ NKCells	PEA3, ETS, CETS1P54, PU1, NERF, STAT1, E2F1DP1, USF, PEBP, STAT5A
PB-CD8+ Tcells	CETS1P54, ELK1, PEA3, PU1, NERF, NRF2, ETS1, ICSBP, NKX25, ETS
adipocyte	IK1, SRF, CEBPB, NKX25, TAL1ALPHA47, TAXCREB, CETS1P54, RP58, E2F1DP1RB, USF
adrenal cortex	GATA, DEAF1, HSF2, TFIIA, STAT6, SOX9_B1, CP2, CREBATF, EGR, MEF3
adrenal gland	TAXCREB, PPARG, DEAF1, LMO2COM, GATA1, ZIC3, PXR, IK1, GATA3, HAND1E47
amygdale	PAX8, PBX1, GRE, PADS, RFX1, NFMUE1, TAL1BETAITF2, VJUN, ZIC3, MEF2
appendix	CART1, NFKB, S8, PBX1, CLOX, TCF4, POU3F2, IRF2, GATA3, RP58
atrioventricular node	RFX1, RP58, TAL1BETAE47, BEL1, TCF4, PBX1, MYOGENIN, VBP, NRSE, PXR
bone marrow	PPAR_DR1, ETS, DR1, TAL1BETAE47, HES1, ELK1, LMO2COM, PAX5, EBOX, TAXCREB
bronchial epithelial cells	BACH2, HIC1, TCF4, P53, STAT6, SREBP1, SRF, T3R, NFKAPPAB65, SOX
cardiac myocytes	MYOD, STAT5B, ETS1, TAL1, SRF, NFKAPPAB65, E2F1DP1RB, MYOGNF1, BACH2, CETS1P54
caudate nucleus	TCF11, NGFIC, HEB, MEF2, OCT, TCF11MAFG, BEL1, EBOX, ZIC3, PBX1
cerebellum	TAL1BETAITF2, TAL1ALPHA47, ZIC3, TAL1, FXR_IR1, TTF1, ER, SEF1, PR, NRSF
cerebellum peduncles	TAL1ALPHA47, TAL1BETAITF2, NRSF, ZIC2, P53_DECAMER, PADS, TAL1, VJUN, MYOD, SMAD
ciliary ganglion	TAL1ALPHA47, TAL1BETAITF2, HP1SITEFACTOR, NFAT, NKX3A, PR, CDP, RORA1, HSF1, ISRE
cingulate cortex	TAL1BETAITF2, BRN2, ZIC3, LHX3, VJUN, TAXCREB, P53_DECAMER, FREAC3, PBX1, NRSF
colorectal adenocarcinoma	ZIC3, E2F4DP2, VDR, BACH2, CREL, DR1, PAX, E2F1DP1RB, E2A, ZIC2
dorsal root ganglion	TBX5, CDP, VMAF, SOX9_B1, PBX1, NERF, OCT1, LEF1, DR4, VBP
fetal brain	CLOX, TCF11, OCT, SOX, OCT1, TAL1BETAITF2, PBX1, CDPCR3, TAACC, SMAD
fetal liver	COUP, HNF1, PPARG, TAL1, OSF2, HNF4, MYOGENIN, ZIC2, HNF4ALPHA, E2F1DP1RB
fetal lung	NFKB, STAT, TATA, TFIIA, FXR_IR1, NKX25, PPARG, EGR, NFKAPPAB65, IK1
fetal thyroid	SRF, TTF1, GATA2, TAL1, NF1, USF, MSX1, PAX, EBOX, MTF1
globus pallidus	BRN2, TAL1BETAITF2, NFKB, TCF11MAFG, MYOD, NRSF, LHX3, CEBPA, P53_DECAMER, TAL1BETAE47
heart	MYOGENIN, LMO2COM, ETS1, LFA1, TBX5, MEF3, CETS1P54, TAL1, NERF, MEF2
hypothalamus	ZIC2, NFMUE1, TAL1BETAITF2, HSF1, SMAD4, PR, P53_DECAMER, NKX25, STAF, GLI
kidney	HNF1, CP2, LMO2COM, PPARG, PAX, ETS2, ER, HNF4_DR1, ZIC2, HNF4ALPHA
leukemia chronic myelogenous (k562)	NFY, CAAT, SF1, GCNF, TCF4, STAT, ELK1, VDR, HES1, PEA3
leukemia lymphoblastic (molt4)	STAT, CETS1P54, E2F, VMYB, CBF, E2F1, XBP1, NFY, YY1, FAC1
leukemia promyelocytic (hl60)	CETS1P54, VJUN, XPF1, IRF, TAL1BETAE47, E2F1, NFY, CREB, HSF, NKX22
liver	HNF4, HNF1, TAL1, DR1, MYOGENIN, STAT, ER, MTF1, USF, COUP_DR1
lung	NFKB, CACBINDINGPROTEIN, NFKAPPAB50, TBX5, ETS, TTF1, CETS1P54, ZIC3, LXR_DR4, ZIC1
lymph node	NERF, PU1, ETS, NFKB, CETS168, NFKAPPAB, CBF, LMO2COM, HSF, CETS1P54
lymphoma Burkitts Daudi	E2F1, NRSF, VMYB, CETS1P54, NKX22, ETS2, NFKAPPAB, NRF2, E2F1DP1RB, PAX6
lymphoma Burkitts Raji	NFKAPPAB, NFMUE1, NRF2, IRF, TAL1BETAE47, PAX6, XPF1, ZIC3, NERF, PU1
medulla oblongata	TAL1BETAITF2, P53_DECAMER, NKX61, CDP, CEBPA, POU6F1, CRX, PBX, TAL1ALPHA47, OCT
occipital lobe	TAL1BETAITF2, TCF11MAFG, HEN1, P53, NFKB, P300, VMYB, FREAC3, USF, NGFIC
olfactory bulb	IPF1, STAT, TAL1ALPHA47, SRF, NFKB, PR, FXR, RP58, PBX1, HNF3ALPHA
ovary	RP58, TAL1ALPHA47, PXR, POU6F1, S8, E4BP4, LMO2COM, GATA6, GF11B, PBX1
pancreas	AREB6, DELTAEF1, HNF4_DR1, USF, CP2, HNF4, SF1, HEN1, OCT, NF1
pancreatic islets	NKX25, ICSBP, VJUN, ARP1, HNF1, ER, STAT, TAL1, OCT, LMO2COM
parietal lobe	ZIC3, TAL1BETAITF2, BRN2, OCT, VJUN, TAL1ALPHA47, VMYB, CDP, MEF2, SOX9_B1
pituitary gland	LHX3, PBX1, CHOP, ZIC2, TCF11, NKX61, CREBATF, OCT, DEAF1, VMAF
placenta	NCX, TTF1, LMO2COM, ATF6, NFMUE1, COREBINDINGFACTOR, HNF4_DR1, GLI, POU3F2, ZIC2
pons	P53_DECAMER, PBX, LEF1, TAL1ALPHA47, CDPCR1, LEF1TCF1, TAL1BETAITF2, EBOX, CEBPA, MZF1
prefrontal cortex	TAL1BETAITF2, HOX13, TCF11, BRN2, GZF1, MEF2, SOX9_B1, NKX25, CRX, VMYB

prostate	LXR_DR4, NFMUE1, MYOGENIN, TATA, TBX5, TAL1, HNF4_DR1, SRF, NMYC, PPAR_DR1
salivary gland	IRF2, ICSBP, OCT, IRF1, S8, TATA, TBX5, IRF, XBP1, GATA2
skeletal muscle	RSRFC4, SRF, MYOGENIN, TATA, DR1, LFA1, GATA1, MEF3, PPARG, NFKB
skin	RP58, E2A, HSF1, GATA3, TEF, TCF11, NF1, PIT1, MAF, BACH2
smooth muscle	GFI1, MYOGNF1, STAT5B, TAL1ALPHA47, BACH2, NFKAPPAB65, PAX8, MSX1, ZTA, ETS2
spinal cord	MEIS1, NFKB, TEF, CDPCR3, HSF1, ZIC1, SOX5, TAL1BETAITF2, GCM, TCF11MAFG
subthalamic nucleus	TAL1BETAITF2, TAL1BETAE47, NRSF, TCF11, P53_DECAMER, CEBPA, SF1, PBX1, VMAF, NKX61
superior cervical ganglion	TATA, TCF4, S8, NFE2, DTYPEPA, RP58, VMAF, TAL1ALPHA47, PAX6, FXR
temporal lobe	LHX3, TAL1BETAITF2, BRN2, PBX, P53_DECAMER, OCT, SOX9_B1, NRSF, PBX1, TCF11
testis	SRF, VMYB, LBP1, POU3F2, GFI1B, CREB, RFX1, E12, RFX, RP58
testis Leydig cell	MEIS1AHOXA9, YY1, LBP1, GATA2, PBX1, RFX1, MEIS1BHOXA9, CREB, RFX, RP58
testis germ cell	GFI1B, FXR_IR1, POU3F2, GATA2, MEIS1AHOXA9, MEIS1, PADS, YY1, ERR1, XBP1
testis interstitial	MEIS1BHOXA9, NFKAPPAB50, GATA2, PADS, TITF1, MEIS1AHOXA9, YY1, LBP1, POU3F2, ERR1
testis seminiferous tubule	GFI1B, POU3F2, TTF1, NFKAPPAB50, OCT1, TITF1, ZIC2, GFI1, MEIS1, MEIS1AHOXA9
thalamus	NRSF, TAL1ALPHA47, TAL1BETAITF2, TAL1, NFMUE1, FOXM1, SMAD4, HEB, EBOX, BEL1
thymus	PEA3, RREB1, PU1, ETS, NERF, CP2, ETS1, NFKAPPAB, TAL1, USF
thyroid	TTF1, MTF1, AREB6, PAX2, USF, XBP1, NFKB, LEF1TCF1, EVI1, E2
tongue	SRF, NF1, BACH2, MEF3, CP2, SREBP1, MYOGENIN, TATA, TBX5, E2A
tonsil	ETS, PU1, CREL, ISRE, NFKAPPAB, ICSBP, MAF, ER, STAT5A, NFKB
trachea	TATA, DELTAEF1, STAT, TBX5, FREAC4, IK1, NFKAPPAB50, NFAT, ZIC3, TEL2
trigeminal ganglion	GATA3, CDPCR3HD, IRF, COMP1, EVI1, IRF7, S8, VMAF, TCF4, NKX61
uterus	HNF6, TST1, FOXO3, SRF, POU3F2, HELIOSA, FOXJ2, HFH8, TAL1ALPHA47, STAT4
uterus corpus	IPF1, HP1SITEFACTOR, NF1, GATA1, STAT4, NKX61, RP58, TAL1ALPHA47, NFKB, GFI1B
whole blood	PU1, PEA3, TEL2, TCF4, ETS, STAT5A, NFKB, COREBINDINGFACTOR, IRF, NRSF
whole brain	TAL1BETAITF2, ZIC3, LHX3, VMAF, MZF1, OCT, MSX1, TAL1, NGFIC, ZIC2

B)

Tissue	List of TFs
B220+ B-cells	ETS1, PU1, NFKB, ETS, ELK1, NRF2, NFKAPPAB50, IRF2, ICSBP, ISRE
CD4+T-cells	PU1, ETS, ETS1, NRF2, NFKB, ICSBP, CETS1P54, PBX, HAND1E47, OSF2
CD8+T-cells	PU1, ETS, E2A, CETS1P54, ETS1, TBX5, NFKB, TAL1, CETS168, PEA3
adipose tissue	ISRE, IRF2, ELK1, COREBINDINGFACTOR, STAT, IK3, ETS1, PU1, PAX9, STAT5B
adrenal gland	GCNF, CEBPDDELTA, TCF4, ETS, TGIF, MYCMAX, LEF1TCF1, CHOP, PBX1, CEBP
amygdale	PAX8, HNF6, NRSF, OCT1, P53_DECAMER, YY1, TEF, RFX1, PBX, TAL1ALPHA47
bladder	SRF, PAX8, GLI, NFAT, ETS, ZIC1, DEC, VDR, IK1, SREBP1
blastocysts	NRSF, CMYB, E2F, TST1, ZTA, GATA3, FOXO4, HNF3B, YY1, ROAZ
bone	ETS, ELK1, MEF2, PU1, ELF1, SRF, TGIF, ETS1, ZID, MYCMAX
bone marrow	OSF2, PU1, PAX6, MYOGENIN, PPARG, ELK1, PEBP, TEL2, ETS1, RORA1
brown fat	COUP_DR1, PPARA, MYOGENIN, HNF4ALPHA, IK2, SRF, SREBP1, DEC, ERR1, RSRFC4
cerebellum	P53_DECAMER, YY1, ZID, NRSF, MEIS1BHOXA9, FOXJ2, OCT1, LEF1TCF1, NRSE, PAX8
cerebral cortex	CDPCR1, NRSF, PAX8, P53_DECAMER, NRSE, HNF6, TEF, GCNF, OCT1, TCF11
digits	TATA, BEL1, TCF4, NFAT, NCX, VDR, GLI, SREBP1, NF1, SP1
dorsal root ganglia	YY1, HNF6, PBX1, CDPCR1, ZID, CDXA, TEF, SMAD4, ERR1, PR
dorsal striatum	RFX1, SMAD4, NRSF, PAX9, TAL1ALPHA47, HNF6, PAX8, RORA2, NRSE, CDPCR1
embryo day 10.5	E2F4DP2, E2F1DP1RB, E2F1DP2, MYCMAX, OLF1, TAL1BETAE47, RREB1, UF1H3BETA, R, CDPCR1
embryo day 6.5	TCF4, ZIC3, ZID, HNF4, HNF4_DR1, NRSF, NFMUE1, CACCCBINDINGFACTOR_Q, TFIIA, COUP
embryo day 7.5	TCF4, ER, LMO2COM, TAL1ALPHA47, YY1, TBX5, PAX5, PAX3, STAT, ZID
embryo day 8.5	TCF4, ZID, HNF1, HOXA3, GATA, GZF1, ZIC3, POU6F1, TITF1, IK1
embryo day 9.5	E2F4DP2, LEF1, RREB1, E2F, UF1H3BETA, TCF4, SOX5, MYCMAX, IK1, ZID
epidermis	RSRFC4, TAL1, MYOGENIN, TCF4, SRF, LBP1, HEB, GLI, SREBP, LXR_DR4
eye	CRX, FOXP3, GATA1, OCT, PR, SRF, FXR_IR1, ATF1, HNF4ALPHA, NFKAPPAB65
fertilized egg	POU6F1, TCF4, TAL1BETAITF2, HTF, STAT, YY1, LEF1TCF1, TBP, TAL1, SF1
frontal cortex	P53_DECAMER, HNF6, PAX8, TAL1ALPHA47, HMGYI, OCT1, MEIS1BHOXA9, RORA2, EN1, NRSF
heart	RSRFC4, ERR1, SRF, MEF2, VJUN, MYOGENIN, LBP1, PAX6, CEBPA, PADS
hippocampus	PAX9, SMAD3, NRSE, P53_DECAMER, CDPCR1, NRSF, POU3F2, TAL1, HNF6, HEN1
hypothalamus	PAX8, YY1, PITX2, PAX3, P53_DECAMER, TEF, POU3F2, PBX1, NRSF, TAL1
kidney	HNF4, HNF1, HNF4_DR1, HNF4ALPHA, STAT, DR1, PPAR_DR1, PAX, CBF, HSF1
large intestine	HNF4, HNF4ALPHA, HNF4_DR1, STAT, STAT5A, AREB6, COREBINDINGFACTOR, TCF4, HSF1, ER
liver	HNF4, HNF1, HNF4_DR1, SREBP1, OLF1, PPAR_DR1, COUP, STAT, PAX2, FXR_IR1
lung	TITF1, NERF, ETS, FREAC7, STAT, SOX9_B1, GCNF, MEF3, HFH4, LHX3
lymph node	ETS, ETS1, PU1, CREL, ETS2, NFKAPPAB65, IRF2, NFKAPPAB50, NFKB, ICSBP
mammary gland	MYOGNF1, STAT, NRSF, HNF4_DR1, GCNF, PPARG, ERR1, TCF4, NRSE, NCX
med	TAL1ALPHA47, MEIS1, CDPCR3HD, SRF, IK3, CIZ, DELTAEF1, PPARA, CRX, TCF11
olfactory bulb	CHOP, PAX8, TAL1ALPHA47, SMAD3, PBX, P53_DECAMER, OCT1, NRSF, YY1, NERF
oocyte	YY1, TST1, LPOLYA, LEF1TCF1, LEF1, TAL1ALPHA47, NKX62, ZTA, SOX, HOX13
ovary	HNF4ALPHA, STAT, ZID, MYB, TCF4, SRF, PAX3, PEBP, IK2, LBP1

pancreas	RP58, STAT, RORA1, CDPCR1, HEN1, SREBP1, MAZR, PPARA, TATA, CDPCR3HD
pituitary	NRSE, PBX1, ZIC2, YY1, TAL1BETAITF2, NRSF, HSF1, SOX5, IK1, IK2
placenta	GLI, ETS, ZIC2, USF, TAL1BETAITF2, CETS1P54, STAT, SRF, HNF4_DR1, TAL1ALPHA47
preoptic	YY1, HNF6, OCT1, HSF, ZID, HSF1, TAL1, NRSE, RFX, PBX1
prostate	MYOGENIN, GR, TCF11, SF1, SREBP1, ROAZ, NERF, LBP1, PTF1BETA, CRX
salivary gland	GCNF, DELTAEF1, CACCCBINDINGFACTOR_Q, TCF11, TGIF, STAT, SF1, LPOLYA, TCF4, SREBP1
skeletal muscle	RSRFC4, MEF2, MEF3, SRF, PPARA, HAND1E47, ERR1, HMEF2, LBP1, MYOGENIN
small intestine	HNF4, DR1, PPAR_DR1, COUP_DR1, COUP, ER, IRF1, RORA1, MYOGNF1, CREL
snout epidermis	TATA, MYOGENIN, TAL1, MYOD, TTF1, MINI20, SREBP1, TBX5, NRSE, E2A
spinal cord lower	PBX1, YY1, PAX8, VMAF, NRSE, CBF, PBX, HNF6, DEAF1, NRSF
spinal cord upper	ZIC2, GCNF, PAX8, VMAF, CART1, ZIC1, HNF6, PBX1, HTF, YY1
spleen	SMAD, PU1, GATA1, RP58, TCF4, LMO2COM, VJUN, AREB6, ETS, PEA3
stomach	DR1, PAX5, T3R, PPARG, DEC, TCF4, ZTA, XPF1, E47, LMO2COM
substantia nigra	PAX8, YY1, P53_DECAMER, TEF, PBX1, DEAF1, LDSPOLYA, NRSE, ZIC2, NRSF
testis	RFX1, ATF3, ER, GATA1, ZID, GATA6, ZIC3, T3R, VDR, ROAZ
thymus	OSF2, PEA3, E2F4DP1, TEL2, XFD3, ETS, COREBINDINGFACTOR, AML, GABP, E2F4DP2
thyroid	RP58, HSF1, GC, MAZR, HNF1, CACCCBINDINGFACTOR_Q, UF1H3BETA, VMAF, TAL1BETAITF2, TAL1ALPHA47
tongue	MYOGENIN, TATA, TBX5, SRF, SREBP1, VMAF, TAL1, E47, OLF1, MINI19
trachea	MEF3, ETS1, NCX, VMAF, CEBP, SRF, ETS, SREBP1, PU1, ISRE
trigeminal	PBX1, TEF, CDPCR1, EN1, GCNF, HNF6, HAND1E47, VJUN, ZID, HSF1
umbilical cord	NERF, SRF, PBX, LFA1, NFKAPPAB65, DEC, VDR, CACBINDINGPROTEIN, NFAT, ZIC2
uterus	SRF, FOXO4, LBP1, IK2, SMAD4, RORA1, ETS2, TFIIA, RP58, ETS
vomerolnasal organ	IK2, PTF1BETA, HSF1, ETS, SRF, RFX1, NFKAPPAB65, LMO2COM, PAX5, TFE

Characterizing predicted associations of TFs with tissue-specificities. In order to reliably characterize the ratio of true positive predictions in the set of predicted TF-tissue associations we manually curated 158 predicted TF-tissue expression associations that correspond to the group of associations with the minimal TF importance threshold of 0.25. Manual curation was performed by seeking published evidence or TF expression in matching tissues (*in situ*, rtPCR, Western or Northern data) (Table S3). Of the 63 predicted TF-tissue associations that are not supported by an increase in gene expression in GNF Atlas2 data, we found 37 TFs (59%) to have at least one line of direct evidence that supports the positive predicted TF-tissue association, indirect evidence supported 10 other TF-tissue associations (16%), while no evidence either supporting or rejecting the predicted association was found for 16 of them (25%; that is 10% of 158 total).

Table S3. Correlation between transcription factor (TF) tissue specificity-prediction and *in vivo* TF expression/function

Transcription Factor	Tissue	Imp	Microarray Expression	rtPCR/in situ Expression	Literature Support	Other important Notes	
BRN2	parietal lobe	0.40	+		McEvilly RJ et al, 2002	cartilage	
	cingulate cortex	0.37	+				
	globus pallidus	0.30	+				
CART1	appendix	0.43	+		Eyquem et al. 2004 (2 papers)		
CETS1P54	PB-CD14+ monocytes	0.33	+				
	PB-CD56+ NKCells	0.27	+				
	PB-CD8+ Tcells	0.63	+				
	PB-CD19+ Bcells	0.62	+				
CLOX	fetal brain	0.86	+		Baroukh N et al. 2005	eye--rods	
COUP	fetal liver	0.46	+				
CRX	medulla oblongata	0.26	+		Li FX et al, 2003 (2 papers)		
E2F	BM-CD34+	0.31	+				
	BM-CD71+ early erythroid	0.30	+				
	BM-CD34+	0.28	+		Cesari F et al, 2004		
ELK1	PB-CD14+ monocytes	0.40	+				
ETS1	tonsil	0.71	+				
	lymph node	0.30	+				
	PB-CD14+ monocytes	0.34	+				

	PB-CD56+ NKCells	0.29	+	Grund EM, et al 2005	
	PB-CD19+ Bcells	0.40	+		
	PB-CD4+ Tcells	0.34	+	Wang J et al. 2006	
	heart	0.26	+	Lie-Venema H, et al 2003; Takahashi T et al, 2001.	
ETS2	PB-CD14+ monocytes	0.26	+		
EVI1	trigeminal ganglion	0.26	+		
FOXJ2	uterus	0.42	+		
FOXO3	uterus	0.50	+		
GATA	adrenal cortex	0.33	+	Jimenez P, et al, 2003	gata6
GATA3	trigeminal ganglion	0.41	+		
HEB	caudate nucleus	0.26	+		
HFH8	uterus	0.40	+		
HNF1	kidney	0.40	+	Senkel S, et al, 2005; Cheret C, et al, 2002	
HNF4	liver	1.03	+	Cheng W, et al, 2006	
ICSBP (IRF8)	salivary gland	0.59	+		
IRF7	trigeminal ganglion	0.25	+	Oussman SS et al., 2005	
IPF1	uterus corpus	0.35	+		
LHX3	cingulate cortex	0.25	+		
	pituitary gland	0.52	+	Cushman LJ et al, 2002	
LMO2COM	heart	0.28	+		
	kidney	0.27	+		
MYOGENIN	heart	0.39	+	Moss JB et al, 1996	
	skeletal muscle	0.29	+	Kanai-Azuma M, et al, 1999	
NERF	lymph node	0.39	+		
NF1 (WSS; NFNS; VRNF; DKFZp686J1293)					
	tongue	0.27	+	Gutmann DH et al, 1995	
	uterus corpus	0.26	+		
NFE2	superior cervical ganglion	0.28	+		cns early embryo
NFKB	lymph node	0.29	+		
	lung	0.38	+		
NFY	BM-CD71+ early erythroid	0.29	+	zhu j et al, 2003	haemopoetic expression
NGFIC	caudate nucleus	0.26	+		neural expression
NKX25	pancreatic islets	0.27	+	Schisler jc et al, 2005; Wang j et al, 2004;	nkx2.2 and nkx6.1 involved in pancreas
function					
NRF2	721 B lymphoblasts	0.40	+	Sander M et al, 2000	nkx2.2 and 2.5 matrices highly similar
NRSF	thalamus	0.26	+	Li J et al, 2004	
	subthalamus nucleus	0.25	+		
OCT (pou2f1)	fetal brain	0.37	+	Sun YM et al, 2005	neuronal development
	parietal lobe	0.36	+	Schonemann MD et al, 1998	neural development
	salivary gland	0.42	+		
P53 DECAMER	medulla oblongata	0.43	+	Fulci G et al. 1998	ubiquitously expressed, brain tumors
PEA3 (ERM)	thymus	0.30	+	Moss AC et al, 2006	cancer
SRF	uterus	0.48	+		
TAL1	fetal liver	0.27	+	Pulford K et al, 1995"	
TAL1ALPHA47	cerebellum peduncles	0.75	+		
	thalamus	0.25	+		
	ciliary ganglion	0.52	+		
	subthalamus nucleus	0.38	+		

POU6F1	medulla oblongata	0.27	+		Cui and Bulleit, 1998; Andersen B et al, 1993	cns expression
PPARG	fetal liver	0.29	+		Memon RA et al, 2000	
PU1	thymus	0.29	+		Nutt SL et al, 2005; Anderson KL et al, 2001	
	tonsil	0.38	+			
	lymph node	0.31	+			
	whole blood	0.60	+			
	PB-CD14+ monocytes	0.80	+			
	PB-CD8+ Tcells	0.26	+			
	PB-CD19+ Bcells	0.35	+			
RREB1	thymus	0.29	+			
RSRFC4 (mef2a)	skeletal muscle	0.30	+		Buchberger A and Arnold HH, 1999	
S8	superior cervical ganglion	0.41	+			
SOX	fetal brain	0.35	+			nervous system development
TAL1BETAITF2	fetal brain	0.26	+			
	temporal lobe	0.40	+			
	parietal lobe	0.44	+			
	occipital lobe	0.31	+			
	prefrontal cortex	0.38	+			
	cingulate cortex	0.58	+			
	cerebellum peduncles	0.39	+			
	subthalamic nucleus	0.72	+			
	globus pallidus	0.30	+			
	medulla oblongata	0.51	+			
TBX5	dorsal root ganglion	0.38	+	+		dorsal root ganglia E13.5
TCF11	prefrontal cortex	0.30	+			
	caudate nucleus	0.46	+			
	occipital lobe	0.28	+			
TCF4	superior cervical ganglion	0.47	+			
TEL2	whole blood	0.28	+			
ZIC2	cerebellum peduncles	0.25	+			
ZIC3 (HTX; HTX1; ZNF203)	cerebellum	0.40	+	+	Aruga J, et al, 2004	early neuronal marker;
E2A (tcf3, itf1)	skin	0.28	-		Merrill BJ et al, 2001; DasGupta R and Fuchs E, 1999	
CDP (cutl1, cux1, phox2)	medulla oblongata	0.28	-	+	Ferrere A et al, 2006; Nieto M et al. 2004	neuronal expression ubiquitous
(neonate)	trigeminal ganglion	0.39	-	+	Ferrere A et al, 2006; Nieto M et al. 2004	neuronal expression ubiquitous
(neonate)						
CEBPA	medulla oblongata	0.28	-			general neural expression (neonate)
BRN2 (PouF3, Oct7 and Otf7)	temporal lobe	0.39	-	+	Huang YT et al, 2005	
ELK1	PB-CD4+ Tcells	0.44	-	+		
ELK1	PB-CD8+ Tcells	0.31	-	+		
GRE (Nr3c1)	amygdala	0.25	-	+	adult brain rtPCR/western; E13.5 in situ	
HNF1	fetal liver	0.45	-			
HOX1.3 (hoxa5, HOX1; HOX1C)	prefrontal cortex	0.32	-	+	Jaksimovic M et al. 2005; Tani M et al, 1989	CNS expression E17.5
LHX3	temporal lobe	0.78	-	+	Bhangoo AP et al, 2006	
MEIS1BHOXA9	testis interstitial	0.25	-	-	Fujimoto S et al, 1998 and Croon E et al, 1998	MEIS1-HOXA9 complex is in testis
NKX61	medulla oblongata	0.32	-	+	Pattyn A et al, 2003; Nelson SB et al, 2005	motor neuron function nkx6 and nkx2

NRSF	cerebellum peduncles	0.39	-	+	Zuccato C et al, 2003	control of neuronal genes	
OCT1	fetal brain	0.32	-				
PAX8	amygdala	0.34	-		Stoykova A and Gruss P, 1994	pax expression in fetal brain	
PBX	temporal lobe	0.30	-	+	Ferretti F et al, 2005; Cooper KL et al, 2003	CNS expression/brain	
PEA3	PB-CD4+ Tcells	0.31	-	+	Charrin S and Alcover A, 2006		
	PB-CD8+ Tcells	0.28	-	+			
PPAR DR1 cells	bone marrow	0.29	-	+	Gimble JM et al, 1996	adipogenesis of bone marrow stromal	
RP58	superior cervical ganglion	0.27	-	+	Kang JS et al, 2001	neuronal gene	
SRF	adipocyte	0.27	-	+	Ding W et al, 2001	adipocyte differentiation	
	skeletal muscle	0.30	-	+	Iver D et al, 2006		
	tongue	0.29	-	+	Cen B et al, 2004; Miano JM 2003	SRF tf involved in muscle	
TAL1	cerebellum	0.26	-	+	Bradley CK et al. 2006	critical for neural development	
TAL1BETAITF2	whole brain	0.33	-	+	Jin H et al, 2006	expression targeted in brain	
	cerebellum	0.96	-	+	Jin H et al, 2006	expression targeted in brain	
	ciliary ganglion	0.26	-	+	Bradley CK et al. 2006	neuronal development	
TATA (tbp)	superior cervical ganglion	0.93	-				
	trachea	0.29	-				
TCF11 (nfe2l1)	fetal brain	0.82	-	+			
TEL2	BM-CD33+ myeloid	0.28	-	+	Carella C et al, 2006	haemophoesis	
	PB-CD14+ monocytes	0.31	-	+	Cardone M et al, 2005	b-cells	
	PB-CD19+ Bcells	0.47	-	+	Cardone M et al, 2005	b-cells	
YY1	PB-CD19+ Bcells	0.27	-	+	Zabel MD et al, 2002	activation in b-cells	
ZIC3	parietal lobe	0.44	-	+		expressed in adult mouse brain	
	cingulate cortex	0.26	-	+			
HNF6	uterus	1.26	-	+	(HNF1)	Bingham C et al, 2002	expressed in uterus/mutants have
uterine defects							
ICSBP (IRF8)	pancreatic islets	0.27	-	+	(IRF1)	Baker MS et al, 2003; Nakazawa T et al, 2001	IRF1 involved in pancreatic function
IRF	trigeminal ganglion	0.27	-	+	(IRF7&9)	Ousman SS et al, 2005	upregulated in brain post infection
NFAT	ciliary ganglion	0.26	-	+	(NFAT4)	Groth RD and Mermelstein PG 2003; Zhang YJ et al, 2005	neuronal function
PEA3	BM-CD33+ myeloid	0.26	-	+	(ETS)	Wasylyk B et al, 1993	T-cell development
	whole blood	0.51	-		indirect		
	PB-BDCA4+ dentritic cells	0.28	-		indirect		
	PB-CD14+ monocytes	0.61	-		indirect		
	PB-CD56+ NKCells	0.69	-		indirect		
TAL1ALPHAE47	cerebellum	0.52	-	+			adult brain expression
BACH2	bronchial epithelial cells	0.48	-		no evidence		
CP2	kidney	0.30	-		no evidence		
GFI1B	testis seminiferous tubule	0.30	-		no evidence		
HELIOS/Zfpn1a2	uterus	0.43	-		no evidence		
IRF1	salivary gland	0.30	-		no evidence		
IRF2	salivary gland	0.66	-		no evidence		
NCX/tlx2/Enx/HOX11L1)	placenta	0.43	-		no evidence		
NFKB	appendix	0.26	-		no evidence		
POU3F2 (BRN2; OCT7; OTF7; POUF3)	uterus	0.44	-		no evidence		
RP58	skin	0.42	-		no evidence		
	ovary	0.83	-		no evidence		

SRF	fetal thyroid	0.32	-	no evidence
TAL1ALPHA47	uterus	0.32	-	no evidence
	ovary	0.31	-	no evidence
TST1	uterus	1.12	-	no evidence
TTF1	placenta	0.30	-	no evidence

95 TF have an increased expression in predicted tissues (using UCSC mapping of GNF Expression Atlas 2 data)

63 TF that do not display an increase in expression:

37 TF not expressed in predicted tissues have positive direct evidence of expression or function associated with the predicted tissue specificity

10 TF not expressed in predicted tissues have indirect evidence for positive correlation by linking one of the family members to the predicted tissues specificity

16 TF not expressed in the predicted tissues have no additional evidence

Reconstructing tissue-specific gene regulatory networks.

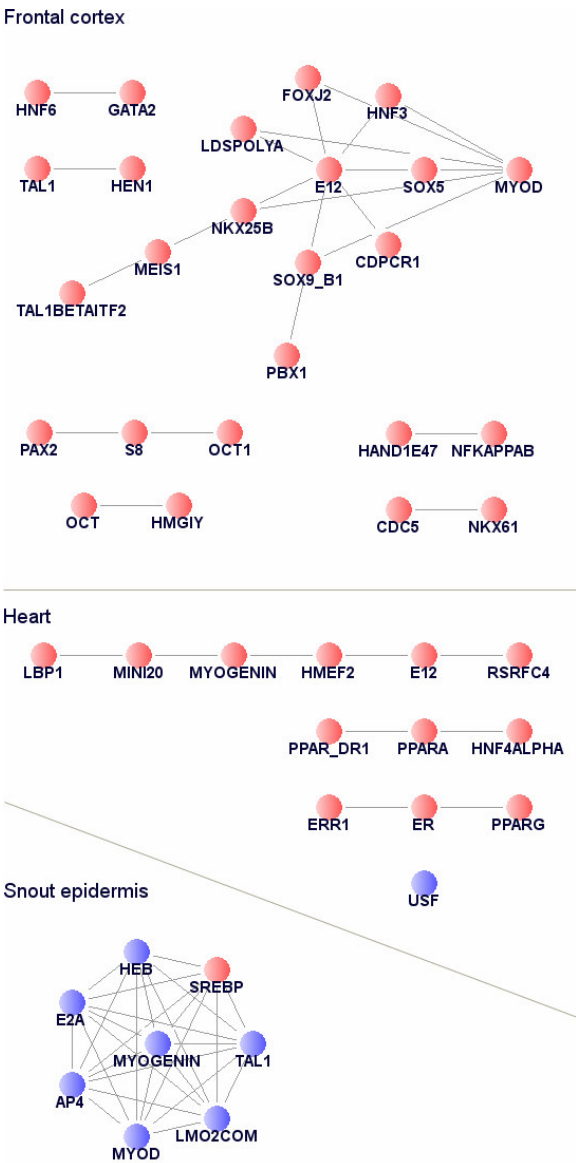
Previous functional dissection of known enhancers has provided evidence of their complex composition of binding sites for multiple TFs. We thus analyzed the distribution of different TFBS in tissue-specific candidate enhancers to determine pairs of TFs with binding sites occurring in close proximity to each other more frequently than expected by chance (see Materials and Methods for details). TF pairs identified in this analysis represented candidate interacting proteins co-functioning in tissue-specific gene expression.

In total, 1,235 human and 713 mouse TF interacting pairs were identified in different tissues; pairs of the same TF binding motif representing 70 and 129 of these pairs, respectively. In several instances we observed a TF “A” independently pairing with two or more different interaction partners (TFs “B” and “C”, for example). In those instances we were able to use this information to build a TF tissue-specific interaction network (B-A-C in the latter example) by joining interacting TF pairs together (Figure S3). For example, we were able to respectively interconnect 12, 6, and 8 different TFs that uniquely describe the topology of the regulatory network in the frontal cortex, heart, and snout epidermis in the mouse (Figure S3). The network between six different heart-specific TFs was linear with 4 TFs serving as the network’s ‘bridge’ without any single instance of intersecting interactions among non-partners, suggesting that heart regulation is consists of multiple narrow regulatory pathways (components) poorly interconnected with each other. The main component of the frontal cortex network consisted of a complex interactome structure based on E12 (aka Tcf21) and MyoD TFs serving as the nuclei of the structure. Interestingly, while independently both E12 and MyoD have been shown to play key regulatory roles in brain development (Gerhart et al. 2001; Heng and Tan 2002), their known cooperative function was previously limited to the muscle program and in our analysis their interaction is

always mediated through a third partner (Etzioni et al. 2005). In contrast, the mouse snout epidermis network of interacting TFs has a very compact structure, where the majority of TFs interact with each other.

It is possible that separate TF interaction components in a tissue-specific regulatory network represent different regulatory pathways therefore the number of components will depend on the complexity of the tissue-specific regulation. The 1-component snout epidermis network, the 4-component heart network, and the 6-component frontal cortex network (Figure S3) are supportive of this hypothesis – where the CNS has the most complex regulatory structure. Furthermore, if we consider 4 individual components of the heart network, we can immediately observe a similarity in TF function from each component. Such, it is possible to link the Myogenin-containing component to muscle development, the PPAR- α component to lipid metabolism, ERR and PPARG component to glucose and lipid metabolism and hormone response, while the solitary USF (aka USF1) component can be associated with hyperlipidemia, cholesterol and triglyceride levels (Coon et al. 2005). Using this approach, we found that one-third of the top 10 TFs with high importance values were linked to at least one interacting partner. This modest observation may be a result of our strict statistical threshold employed in the analysis that potentially limits the extraction to only the major TF interacting partners or the tendency of TFs to couple with several different partners.

Figure S3. Networks of interacting TFs in mouse frontal cortex, heart, and snout epidermis. Gray lines depict pairs of interacting TFs. Blue colored TFs depict TFs that pair with each other in the interactome network.



Analysis of DACH *in vivo* enhancers. To estimate the probability of observing at least 3 out of 5 elements being assigned to brain (and only brain) tissues we generated a non-overlapping dataset of 100,000 elements each centered at a random human ncECR. Each element length was randomly selected from the list of lengths of 5 DACH elements. Organ(s) specificity was scored similarly to the element's tissue specificity. Up to 3 organs with a score of at least half of the maximum element tissue-specificity scores were assigned to an element. We observed 7.6K of these elements to be assigned by the *EI* genome wide annotation specifically to the brain. Using the hypergeometric distribution we estimated the *p*-value to be 0.004.

Table S4. *EI* annotation of 7 DACH1 *in vivo* enhancers. Element names correspond to the original publication [Nobrega, M.A., Ovcharenko, I., Afzal, V. and Rubin, E.M. (2003) Scanning human gene deserts for long-range enhancers. Science, 302, 413]. None of these regulatory elements were detected in the dataset of *EI* tissue-specific candidate enhancers because DACH1 TF, is not classified as an highly expressed gene in any of these tissues, but we were able to overlap these elements with *EI* genome wide predictions.

Element	Match	<i>In vivo</i> annotation	Human <i>EI</i> tissue annotation	Corresponding organs
Dc2	+	Hindbrain	medulla oblongata:16.4 cingulate cortex:8.2	brain
Dc3	+	forebrain, hindbrain, spinal cord, retina	fetal brain:19.7	brain
Dc4	not profiled in EI	Retina	fetal brain:17.8 medulla oblongata:16.0 temporal lobe:9.6	brain
Dc6	-	midbrain, retina, dorsal root ganglia	None	none
Dc7	not profiled in EI	Limbs	None	none
Dc8	+	Forebrain, neural tube	pons:10.3 medulla oblongata:8.8	brain
Dc9	-	hindbrain, neural tube, genital eminence	None	none

Analysis of a dataset of 71 *in vivo* enhancers. Similarly to the DACH test, we generated a non-overlapping dataset of 100K elements each centered at a random human noncoding ECR. Each element length was randomly selected from the list of lengths of 71 *in vivo* brain/CNS enhancers. We observed 11.3K of these elements to be assigned by the genome wide prediction dataset specifically to the brain and/or nervous system. Using the hypergeometric distribution we estimated the *p*-value of correctly assigning brain/CNS tissue-specificity to 20 out of 71 *in vivo* elements to be 0.0001.

Next, we found that there are 1.6K elements that are annotated only as brain and contain a fetal brain annotation as one of the annotations in the tissue list. This allowed us to use hypergeometric distribution to calculate the *p*-value of finding 11 such elements in the list of 71 just by chance to be equal to 2.6e-8.

Table S5. *EI* annotation of 71 elements acting as brain and/or CNS *in vivo* enhancers in E11.5 mice (<http://enhancer.lbl.gov> enhancer database; retrieved on 6/8/06). Four types of annotation classifications corresponded to either a match (+), mixed match to brain/CNS and other tissues (+-), mismatch (-), and an unannotated element (0). Organ(s) classification and scores are separated by a colon in the *EI* annotation records.

Type	Location	<i>In vivo</i> annotation	Genome wide <i>EI</i> annotation
+	chr10:131581076-131582838	hindbrain	Brain:0.813
+	chr10:77660002-77661475	forebrain	Brain:3.955
+	chr13:71198850-71200935	hindbrain	Brain:16.413
+	chr13:71323788-71326336	forebrain, hindbrain, neural tube	Brain:19.685
+	chr15:55212536-55213250	hindbrain, midbrain, neural tube	Brain:18.492
+	chr16:49613812-49615093	hindbrain, midbrain, neural tube	Brain:7.412
+	chr16:49788622-49789863	midbrain, neural tube	Brain:10.095
+	chr16:50134226-50135602	neural tube	nerve:0.870
+	chr16:53208099-53209383	forebrain	Brain:19.835
+	chr16:53780686-53781784	midbrain, neural tube	Brain:5.419

+	chr16:70812067-70813326	forebrain	Brain:3.988
+	chr16:78930094-78931256	hindbrain, midbrain, neural tube	Brain:9.381, nerve:8.562
+	chr2:173939878-173941440	forebrain, midbrain, neural tube	nerve:2.522, brain:1.739
+	chr2:176765577-176766917	hindbrain	Brain:10.589
+	chr5:87275698-87277401	forebrain, hindbrain, midbrain, neural tube	Brain:10.709, nerve:6.407
+	chr5:91072644-91074655	hindbrain	Brain:4.406
+	chr6:41631202-41631655	forebrain, midbrain	Brain:3.777
+	chr7:1038395-1039559	forebrain, midbrain	Brain:15.459
+	chr9:126277954-126280293	midbrain	Brain:5.807, nerve:5.421
+	chr9:8085553-8086166	forebrain, midbrain	Brain:7.473
+-	chr1:50746567-50748310	midbrain	Brain:3.766, immune:3.294
+-	chr1:50877216-50878807	forebrain	Other:2.726, brain:1.801
+-	chr2:45018578-45019304	midbrain	Gland:0.964, nerve:0.928
+-	chr5:158273540-158275189	neural tube	muscle:7.229, other:4.647, brain:3.640
+-	chr7:96278233-96278954	midbrain	Brain:1.337, gland:0.699
-	chr10:126895312-126895993	hindbrain	muscle:4.726
-	chr11:31641933-31643460	midbrain	Germ:5.918, other:5.278
-	chr16:47651967-47652743	neural tube	Germ:3.744
-	chr16:50117697-50118426	midbrain, neural tube	muscle:12.062
-	chr16:50228682-50229540	forebrain	Other:3.444
-	chr16:71296069-71297650	hindbrain, midbrain	muscle:2.371
-	chr16:77068109-77069445	forebrain, hindbrain	muscle:15.119
-	chr4:105703179-105704499	midbrain	cancer:0.947
-	chr5:87727910-87729021	forebrain, hindbrain, midbrain, neural tube	Gland:4.429, germ:2.274
-	chr5:90964368-90964982	forebrain, hindbrain, midbrain	Germ:2.259
-	chr6:98826480-98827231	neural tube	muscle:1.730
-	chr7:21584610-21585242	neural tube	muscle:4.134, other:3.657
-	chr7:26501937-26503042	hindbrain	immune:5.678, gland:3.473
-	chr8:106402706-106404035	midbrain	Other:3.394, gland:2.747, nerve:2.491
-	chr9:79100176-79101778	forebrain	muscle:12.309
0	chr10:77396867-77397624	hindbrain, neural tube	-
0	chr1:10666336-10667384	neural tube	-
0	chr1:209986359-209987919	forebrain	-
0	chr1:38229287-38230699	neural tube	-
0	chr13:94416465-94417820	hindbrain, midbrain, neural tube	-
0	chr13:94416517-94417851	neural tube	-
0	chr1:44384513-44385222	midbrain	-
0	chr15:35028097-35029790	midbrain	-
0	chr1:63176303-63177738	neural tube	-
0	chr16:47470317-47471645	hindbrain	-
0	chr16:53133598-53134853	midbrain	-
0	chr16:53981917-53983239	hindbrain, neural tube	-
0	chr16:84987588-84988227	hindbrain, neural tube	-
0	chr1:87533642-87535103	forebrain	-
0	chr2:164275651-164277265	forebrain	-
0	chr2:174814244-174815870	forebrain	-
0	chr2:236853364-236853848	neural tube	-
0	chr3:138530242-138531969	midbrain	-
0	chr3:182256341-182258504	forebrain	-

0	chr3:62325766-62326758	forebrain	-
0	chr3:62334906-62335609	forebrain, midbrain	-
0	chr3:71373108-71375274	forebrain	-
0	chr5:170560595-170562623	hindbrain	-
0	chr5:3564978-3566399	midbrain	-
0	chr5:93252741-93254078	forebrain, midbrain	-
0	chr6:98222806-98223664	hindbrain, midbrain	-
0	chr7:20776520-20777990	forebrain	-
0	chr7:41964968-41966878	forebrain	-
0	chr9:16700536-16701184	hindbrain	-
0	chr9:78701301-78702668	forebrain, hindbrain, midbrain	-
0	chr9:80956677-80957932	forebrain	-

Table S6. Summary of tissue-specificity annotation of previously predicted candidate regulatory elements in the human genome (Prabhakar et al. 2006).

Tissue	# annotations	Tissue with the largest overlap in annotations; number of overlapping elements	
721 B lymphoblasts	6540	PB-CD19+ Bcells	770
BM-CD105+ endothelial	7749	BM-CD71+ early erythroid	1417
BM-CD33+ myeloid	5267	PB-CD14+ monocytes	626
BM-CD34+	5364	BM-CD105+ endothelial	604
BM-CD71+ early erythroid	6169	BM-CD105+ endothelial	1417
PB-BDCA4+ dendritic cells	3727	PB-CD19+ Bcells	328
PB-CD14+ monocytes	7245	PB-CD19+ Bcells	2036
PB-CD19+ Bcells	9727	PB-CD14+ monocytes	2036
PB-CD4+ Tcells	7262	PB-CD8+ Tcells	1733
PB-CD56+ NKCells	4031	PB-CD14+ monocytes	833
PB-CD8+ Tcells	9039	PB-CD4+ Tcells	1733
adipocyte	8464	uterus	597
adrenal cortex	7706	trigeminal ganglion	830
adrenal gland	4237	adrenal cortex	462
amygdala	10538	medulla oblongata	1138
appendix	17151	fetal brain	2252
atrioventricular node	11673	ovary	1261
bone marrow	7377	heart	744
bronchial epithelial cells	6965	smooth muscle	459
cardiac myocytes	6449	smooth muscle	1082
caudate nucleus	8707	fetal brain	2192
cerebellum	11837	subthalamic nucleus	2020
cerebellum peduncles	6002	cerebellum	1337
ciliary ganglion	16285	uterus	2227
cingulate cortex	11170	temporal lobe	2200
colorectal adenocarcinoma	6312	bone marrow	405
dorsal root ganglion	10258	atrioventricular node	1016
fetal brain	24348	uterus	2768
fetal liver	10093	kidney	1243
fetal lung	8638	uterus	614
fetal thyroid	6555	placenta	461
globus pallidus	7846	temporal lobe	1500
heart	11788	skeletal muscle	1589
hypothalamus	6944	occipital lobe	877
kidney	9808	fetal liver	1243
leukemia chronic myelogenous (k562)	4908	appendix	274
leukemia lymphoblastic (molt4)	5802	721 B lymphoblasts	475
leukemia promyelocytic (hl60)	3589	lymphoma Burkitts Daudi	408
liver	8753	fetal liver	1208
lung	6465	heart	639
lymph node	4590	PB-CD19+ Bcells	395
lymphoma Burkitts Daudi	5470	721 B lymphoblasts	628
lymphoma Burkitts Raji	4269	721 B lymphoblasts	316
medulla oblongata	21902	temporal lobe	2428
occipital lobe	10493	medulla oblongata	1352
olfactory bulb	9919	pons	657
ovary	18231	superior cervical ganglion	2107
pancreas	7529	pancreatic islets	765
pancreatic islets	10086	salivary gland	989
parietal lobe	9865	medulla oblongata	2070
pituitary gland	11562	temporal lobe	2233
placenta	8196	Uterus	564

pons	8457	medulla oblongata	1508
prefrontal cortex	14736	fetal brain	1862
prostate	5034	heart	478
salivary gland	15796	trigeminal ganglion	1262
skeletal muscle	10329	heart	1589
skin	11576	trigeminal ganglion	1197
smooth muscle	9220	cardiac myocytes	1082
spinal cord	6071	fetal brain	551
subthalamic nucleus	10927	cerebellum	2020
superior cervical ganglion	22992	trigeminal ganglion	3993
temporal lobe	16534	medulla oblongata	2428
testis	7002	testis Leydig cell	642
testis Leydig cell	10400	testis interstitial	2413
testis germ cell	11522	testis interstitial	1732
testis interstitial	12013	testis Leydig cell	2413
testis seminiferous tubule	7726	testis interstitial	1453
thalamus	8023	cerebellum	815
thymus	3427	whole blood	238
thyroid	7028	uterus	466
tongue	8625	skeletal muscle	1006
tonsil	5842	PB-CD14+ monocytes	530
trachea	6899	uterus	662
trigeminal ganglion	26143	superior cervical ganglion	3993
uterus	30121	fetal brain	2768
uterus corpus	12843	uterus	1287
whole blood	5348	PB-CD14+ monocytes	1162
whole brain	7206	temporal lobe	1101

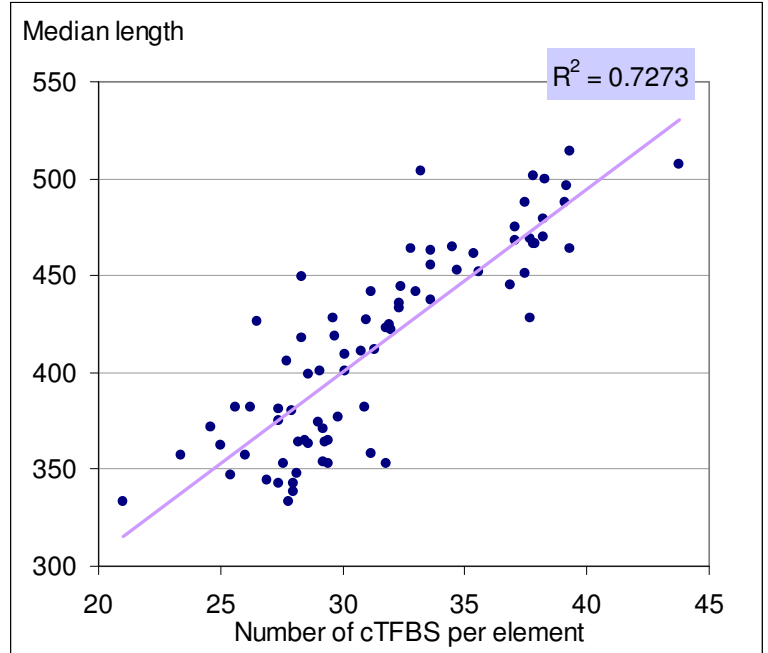
Table S7. Summary of tissue-specificity annotation of mouse-human conserved elements in the mouse genome.

Tissue	# annotations	Tissue with the largest overlap in annotations; number of overlapping elements	
B220+ B-cells	20349	lymph node	6775
CD4+ T-cells	17286	CD8+ T-cells	4655
CD8+ T-cells	18653	CD4+ T-cells	4655
adipose tissue	19800	B220+ B-cells	2759
adrenal gland	26299	Liver	1683
amygdala	42247	frontal cortex	10979
bladder	16081	Digits	1438
blastocysts	16927	Oocyte	1406
bone	14252	skeletal muscle	1445
bone marrow	14510	Thymus	2057
brown fat	23097	Heart	3002
cerebellum	19706	frontal cortex	3859
cerebral cortex	22998	Amygdala	4919
digits	16891	snout epidermis	2848
dorsal root ganglia	17871	Trigeminal	3322
dorsal striatum	19375	frontal cortex	3015
embryo day 10	18514	embryo day 9	4406
embryo day 6	11780	Oocyte	1422
embryo day 7	12219	Oocyte	1000
embryo day 8	16444	embryo day 9	2531
embryo day 9	17754	embryo day 10	4406
epidermis	23832	snout epidermis	4587
eye	38953	Prostate	2074
fertilized egg	35041	Oocyte	4664

frontal cortex	44923	Amygdala	10979
heart	32664	skeletal muscle	6912
hippocampus	19828	cerebral cortex	3171
hypothalamus	26174	Amygdala	5022
kidney	34315	Liver	11087
large intestine	13558	small intestine	2465
liver	37526	Kidney	11087
lung	29171	skeletal muscle	1388
lymph node	24736	B220+ B-cells	6775
mammary gland	22631	Liver	2848
med	22080	salivary gland	2014
olfactory bulb	24969	Amygdala	3778
oocyte	28656	fertilized egg	4664
ovary	16968	vomerlnasal organ	1122
pancreas	24384	Kidney	2827
pituitary	20535	vomerlnasal organ	1890
placenta	17789	fertilized egg	1299
preoptic	12598	Amygdala	1645
prostate	21114	Eye	2074
salivary gland	21968	Med	2014
skeletal muscle	23246	Heart	6912
small intestine	26006	Liver	5336
snout epidermis	22435	Epidermis	4587
spinal cord lower	15162	spinal cord upper	2502
spinal cord upper	23088	Amygdala	3348
spleen	22749	Pancreas	2301
stomach	22246	fertilized egg	1535
substantia nigra	10630	Trigeminal	1697
testis	24234	Oocyte	1774
thymus	26399	bone marrow	2057
thyroid	17964	Kidney	2286
tongue	18014	snout epidermis	4117
trachea	19327	adipose tissue	1956
trigeminal	33260	cerebral cortex	3431
umbilical cord	14070	Bladder	1157
uterus	17724	vomerlnasal organ	1849
vomerlnasal organ	22222	Pituitary	1890

Table S8. Profiling of the number of TFBS per element and the median length of an element in the tissue group for 60k candidate regulatory elements binned by tissue. Tissue assignment is made through linking genes from highly expressed tissue-specific groups to candidate regulatory elements in their loci. The data is sorted by the number of TFBS per element. The figure insert depicts the correlation of the number of TFBS per element with the median element length in a tissue group.

Tissue	# of TFBS per element	Median element length
liver	21.0	333.0
bone marrow	23.4	357.0
fetal liver	24.6	372.0
tonsil	25.0	362.0
heart	25.4	347.0
pancreas	25.6	382.0
lymph node	26.0	357.0
fetal lung	26.2	382.0
trachea	26.5	426.0
lymphoma Burkitts Raji	26.9	344.0
PB-BDCA4+ dendritic cells	27.4	343.0
kidney	27.4	381.0
prostate	27.4	375.0
PB-CD14+ monocytes	27.6	353.0
tongue	27.7	406.0
leukemia promyelocytic (hl60)	27.8	333.0
whole blood	27.9	380.0
721 B lymphoblasts	28.0	343.0
lymphoma Burkitts Daudi	28.0	338.0
PB-CD56+ NKCells	28.1	348.0
PB-CD8+ Tcells	28.2	364.0
pancreatic islets	28.3	418.0
salivary gland	28.3	449.0
PB-CD4+ Tcells	28.5	365.0
bronchial epithelial cells	28.6	399.0
thymus	28.6	363.0
testis	29.0	374.0
testis seminiferous tubule	29.1	401.0
BM-CD33+ myeloid	29.2	354.0
colorectal adenocarcinoma	29.2	371.0
BM-CD105+ endothelial	29.3	364.0
BM-CD34+	29.4	353.0
lung	29.4	365.0
placenta	29.6	428.0
skeletal muscle	29.7	419.0
BM-CD71+ early erythroid	29.8	377.0
testis Leydig cell	30.1	401.0
thyroid	30.1	409.0
fetal thyroid	30.8	411.0
leukemia chronic myelogenous (k562)	30.9	382.0
testis germ cell	31.0	427.0
PB-CD19+ Bcells	31.2	358.0
atrioventricular node	31.2	442.0
adrenal gland	31.3	412.0
leukemia lymphoblastic (molt4)	31.8	353.0



testis interstitial	31.8	423.0
adrenal cortex	31.9	425.0
adipocyte	32.0	422.0
cardiac myocytes	32.3	436.0
skin	32.3	433.0
smooth muscle	32.4	444.0
ovary	32.8	464.0
olfactory bulb	33.0	442.0
appendix	33.2	504.0
ciliary ganglion	33.6	463.0
dorsal root ganglion	33.6	437.0
pituitary gland	33.6	455.0
hypothalamus	34.5	465.0
trigeminal ganglion	34.7	453.0
thalamus	35.4	461.0
spinal cord	35.6	452.0
whole brain	36.9	445.0
caudate nucleus	37.1	468.0
globus pallidus	37.1	475.0
cerebellum peduncles	37.5	451.0
uterus corpus	37.5	488.0
pons	37.7	469.0
temporal lobe	37.7	428.0
amygdala	37.8	466.0
uterus	37.8	501.0
cerebellum	37.9	466.0
cingulate cortex	38.2	470.0
parietal lobe	38.2	479.0
occipital lobe	38.3	500.0
medulla oblongata	39.1	488.0
superior cervical ganglion	39.2	496.0
prefrontal cortex	39.3	464.0
subthalamic nucleus	39.3	514.0
fetal brain	43.8	507.0

REFERENCES:

- Azackie, A., J.R. Fineman, and Y. He. 2006. Sp3 Inhibits Sp1 Mediated Activation of the Cardiac Troponin T Promoter and is Downregulated during Pathologic Cardiac Hypertrophy in vivo. *Am J Physiol Heart Circ Physiol*.
- Blanchette, M., A.R. Bataille, X. Chen, C. Poitras, J. Laganier, C. Lefebvre, G. Deblois, V. Giguere, V. Ferretti, D. Bergeron, B. Coulombe, and F. Robert. 2006. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res* **16**: 656-668.
- Coon, H., Y. Xin, P.N. Hopkins, R.M. Cawthon, S.J. Hasstedt, and S.C. Hunt. 2005. Upstream stimulatory factor 1 associated with familial combined hyperlipidemia, LDL cholesterol, and triglycerides. *Hum Genet* **117**: 444-451.
- Etzioni, S., A. Yafe, S. Khateb, P. Weisman-Shomer, E. Bengal, and M. Fry. 2005. Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. *J Biol Chem* **280**: 26805-26812.
- Gerhart, J., B. Bast, C. Neely, S. Iem, P. Amegbe, R. Niewenhuis, S. Miklasz, P.F. Cheng, and M. George-Weinstein. 2001. MyoD-positive myoblasts are present in mature fetal organs lacking skeletal muscle. *J Cell Biol* **155**: 381-392.
- Heng, J.I. and S.S. Tan. 2002. Cloning and characterization of GRIPE, a novel interacting partner of the transcription factor E12 in developing mouse forebrain. *J Biol Chem* **277**: 43152-43159.
- Li, S., M.P. Czubryt, J. McAnally, R. Bassel-Duby, J.A. Richardson, F.F. Wiebel, A. Nordheim, and E.N. Olson. 2005. Requirement for serum response factor for skeletal muscle growth and maturation revealed by tissue-specific gene deletion in mice. *Proc Natl Acad Sci U S A* **102**: 1082-1087.
- Naya, F.J. and E. Olson. 1999. MEF2: a transcriptional target for signaling pathways controlling skeletal muscle growth and differentiation. *Curr Opin Cell Biol* **11**: 683-688.
- Parekh, V., A. McEwen, V. Barbour, Y. Takahashi, J.E. Rehg, S.M. Jane, and J.M. Cunningham. 2004. Defective extraembryonic angiogenesis in mice lacking LBP-1a, a member of the grainyhead family of transcription factors. *Mol Cell Biol* **24**: 7113-7129.
- Prabhakar, S., F. Poulin, M. Shoukry, V. Afzal, E.M. Rubin, O. Couronne, and L.A. Pennacchio. 2006. Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res*.