Functional autonomy of distant-acting human enhancers.

Axel Visel[1], Jennifer A. Akiyama[1], Malak Shoukry[1], Veena Afzal[1], Edward M. Rubin[1,2], and Len A. Pennacchio[1,2,*]

One sentence summary:

Human tissue-specific enhancers are highly modular and functionally autonomous regulatory units with additive spatiotemporal activities in embryonic development.

[1] Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA.

[2] U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598 USA.

* To whom correspondence should be addressed: Len A. Pennacchio, Genomics Division, One Cyclotron Road, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: LAPennacchio@lbl.gov, Phone: (510) 486-7498, Fax: (510) 486-4229.

Abstract

Many human genes are associated with dispersed arrays of transcriptional enhancers that regulate their expression in time and space. Studies in invertebrate model systems have suggested that these elements function as discrete and independent regulatory units, but the *in vivo* combinatorial properties of vertebrate enhancers remain poorly understood. To explore the modularity and regulatory autonomy of human developmental enhancers, we experimentally concatenated up to four enhancers from different genes and used a transgenic mouse assay to compare the *in vivo* activity of these compound elements with that of the single modules. In all of the six different combinations of elements tested, the reporter gene activity patterns were additive without signs of interference between the individual modules, indicating that regulatory specificity was maintained despite the presence of closely-positioned heterologous enhancers. Even in cases where two elements drove expression in close anatomical proximity, such as within neighboring subregions of the developing limb bud, the compound patterns did not show signs of cross-inhibition between individual elements or novel expression sites. These data indicate that human developmental enhancers are highly modular and functionally autonomous and suggest that genomic enhancer shuffling may have contributed to the evolution of complex gene expression patterns in vertebrates.

Main Text

The regulation of many human genes is controlled by multiple discrete enhancer sequences with different tissue specificities (e.g. *1-4*). Such enhancers activate gene expression independent of their orientation (*5*) and are commonly scattered across large

noncoding intervals (*6, 7*), in some extreme cases functioning >1 Mb from their gene promoter target (*8, 9*). While progress towards genome-wide annotation of developmental enhancers has been made by coupling comparative genomic approaches to experimental studies in mice and fish (*10-12*), the functional and evolutionary significance of their dispersed arrangement remains unclear. Structural modularity of enhancer architecture may facilitate evolutionary fine tuning of distinct aspects of expression patterns (*13, 14*), but observations in invertebrate models (*15, 16*) have also raised the possibility that intergenic translocation of preformed enhancer modules may have contributed to the evolution of complex gene expression patterns in vertebrates. However, it remains unclear if this proposed mechanism of regulatory evolution is feasible since it assumes that enhancers accurately retain their individual activities when placed into a new genomic context containing previously resident heterologous enhancers.

To explore the prevalence of possible positive or negative interactions among human developmental enhancers, we recombined enhancer modules from different, functionally unrelated genes and studied their regulatory *in vivo* properties during embryonic development in transgenic mice. We selected for this purpose six *in vivo* validated enhancers (ref. *10, 11*; E1-E6, Fig. 1a). When individually coupled to a minimal heat shock protein 68 promoter linked to a LacZ reporter gene (ref. *11, 17*; Fig. 1b), each of these elements drove reproducible tissue-specific expression in transgenic mouse embryos. Representative patterns are shown in Fig. 1c and their strong reproducibility in independent transgenic embryos is reported in supplemental table 2. These enhancers were selected for analysis based on their expression patterns which are easily distinguished at the resolution of whole-mount staining, yet also include features that are located in close spatial proximity such as within the limb bud or in neighboring regions of the forebrain, midbrain, and hindbrain. All elements are located on different human chromosomes and thereby are expected to regulate independent gene(s), with the exception of E1 and E2 which are within introns of different genes on chromosome 1 and are located more than 5 Mb apart from each other. All elements are located at least 30 kb away from the closest known promoter, with 4 elements found within genes (E1, E2, E3 and E6) and the remaining 2 elements found between genes (E4 and E5).

These enhancers are experimentally defined by their ability to drive reporter gene expression in transgenic mouse embryos, but (as for most human developmental enhancers) it is unknown whether additional *cis*-regulatory cues such as general or tissue-specific repressor/silencer activities are also embedded in these elements that might affect the activity of other regulatory elements in their vicinity. To determine the combinatorial properties of these human enhancers, we generated five constructs containing pair-wise tandem fusions of the heterologous enhancers described above (Fig. 1b). For each construct, we obtained multiple transgenic embryos at e11.5 (representing 8 to 15 independent transgenic integration events, see suppl. table 2). In each of the cases studied we observed reproducible patterns that were a direct superimposition of the two individual patterns (Fig. 1d-h). For instance, as one representative example, a construct containing E2 (forebrain) coupled to E5 (medial-dorsal and lateral cell populations of the midbrain and ventral hindbrain) targets reporter gene expression to the same respective subregions of the fore-, mid- and hindbrain as observed for E2 and E5 alone (Fig. 1c,e). We observed no decrease in the reproducibility of enhancer activities in the tandem fusion constructs in

comparison to each enhancer element alone (suppl. table 2). Conversely, the concatenation of heterologous enhancers did not result in any reproducible staining in additional anatomical structures or domains outside of those observed for the individual enhancer constructs. These data indicate that in all instances tested, the two enhancers retained their specificity independent of each other despite the artificial coupling of enhancer modules that regulate different genes.

To test whether more complex combinations of enhancer modules would result in positive or negative interactions, we concatenated four different enhancers from unrelated genes and tested the activity of this compound construct at embryonic day 11.5. Similar to the additive results observed in the combination of two discrete enhancers, independent transgenic embryos for the compound 4-mer construct had highly reproducible patterns that included all the major features of the individual patterns, while not introducing additional reproducibly stained structures (Fig. 1i, suppl. table 2). Thus, even when multiple distant-acting enhancers were combined in close proximity in a single construct, the individual enhancer units retained their distinct spatial specificities at this time-point.

In addition to spatial properties, we also examined temporal aspects of expression driven by concatenated enhancers. We selected for this purpose construct E3+E4, containing a pair of enhancers that drove expression in the developing limb (Fig. 1f), where morphological changes enable precise developmental stage matching of independently generated transgenic embryos. When tested individually at stages ranging from e10.5 to e12.5, enhancer E3 alone targets expression to the apical ectodermal ridge (AER) and the surface ectoderm of the limb bud (Fig. 2a-d). In contrast, E4 alone does not drive limb staining at e10.5, but targets expression to a sharply restricted central cell population in the limb at e11.5 that continues throughout e12.5 (Fig. 2e-h). In order to compare the developmental progression of expression driven by constructs E3, E4, and E3+E4, we collected multiple transgenic embryos at e10.5, e11.5 and e12.5 for each construct. We found that the compound construct E3+E4 drove AER expression at all time-points examined, whereas the medial expression domain was first observed at e11.5 as with E4 alone, indicating that the developmental onset of expression driven by E4 is not affected by the presence of E3 in its immediate proximity. Taken together, these results indicate that the functional independence of human developmental enhancers and the absence of obvious regulatory interference among them could allow the generation of complex spatiotemporal expression patterns through modular intergenic recombination of enhancers.

The high degree of regulatory autonomy observed in this study suggests that functional independence and spatiotemporal additivity are common features of human distant-acting enhancer modules. Our results indicate that emerging collections of human and other vertebrate enhancers (*10-12*) provide a toolbox enabling the design of regulatory composites driving customized, complex *in vivo* expression patterns in a predictable manner due to the additive nature of the individual components. These observations also have potential evolutionary implications, as it has been proposed that duplication of regulatory elements into new genomic locations may have contributed to the emergence of complex gene expression patterns (*14, 18, 19*). *Bona fide* examples of intergenic enhancer shuffling in the human genome remain to be identified, but recent comparative genomic evidence suggests that exaptation of transposable genome elements occurred on a pervasive scale (*20*). While

some of these mobile elements gave rise to functional enhancers (*21, 22*), it remains uncertain whether such transposon-derived elements typically arose *de novo* or if partially preformed *cis*-regulatory functions were already embedded at the time of their translocation. The remarkable modular additivity of spatial and temporal enhancer activities observed in this study suggests that functional distant-acting enhancers, if translocated into new genomic environments, have the potential to transfer aspects of expression patterns between genes without disrupting the function of pre-existing enhancers, supporting intergenic enhancer shuffling as a possible mechanism of vertebrate genome evolution.

Materials and Methods

Enhancer reporter constructs. All enhancer sequences were PCR amplified from human genomic DNA (Clontech) using the primers listed in supplemental table 1. PCR fragments were cloned into the pENTR plasmid (Invitrogen), transferred into an Hsp68-LacZ reporter vector containing a Gateway cassette using LR recombination (Invitrogen; *11, 17*) and sequence validated.

Compound enhancers. To generate compound enhancers, inserts from individual constructs were subcloned by standard molecular cloning techniques. Sequence and orientation of enhancers in the final constructs are indicated in supplemental table 1. Residual multiple cloning site fragments of up to 48bp residing between enhancers are also listed in supplemental table 1.

Transgenic mice. Transgenic mouse embryos were generated by pronuclear injection in accordance with protocols reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee. Zygotes at 0.5 dpc for pronuclear injection were collected from FVB strain donor females (Charles River) and, after injection, transferred into pseudopregnant CD-1 strain recipient females (Charles River). Embryos were collected and stained for LacZ activity as previously described (*6*).

Assessment of reporter gene expression. Only anatomical structures in which reporter gene expression was present in at least three embryos resulting from independent transgene integration events were considered reproducible. Reproducibilities for all patterns observed with individual and compound enhancers are listed in supplemental table 2.

Acknowledgments

The authors wish to thank Shyam Prabhakar, Rotem Sorek, Marcelo Nobrega, James Noonan and Nadav Ahituv for critical comments on the manuscript; Keith Lewis, Amy Holt, Ingrid Plajzer-Frick, Sengthavy Phouanenavong, and Sumita Bhardwaj for technical support. L.A.P./E.M.R. were supported by the Berkeley-PGA, under the Programs for Genomic Applications, funded by National Heart, Lung, & Blood Institute, and L.A.P. by the National Human Genome Research Institute. Research was performed under Department of Energy Contract DE-AC02-05CH11231, University of California, E.O.

Lawrence Berkeley National Laboratory. A.V. was supported by an American Heart Association postdoctoral fellowship.

References

1. W. S. Simonet, N. Bucay, R. E. Pitas, S. J. Lauer, J. M. Taylor, *J Biol Chem* **266**, 8651-4 (May 15, 1991).

2. R. J. Schwartz, E. N. Olson, *Development* **126**, 4187-92 (Oct, 1999).

3. J. B. Burch, *Semin Cell Dev Biol* **16**, 71-81 (Feb, 2005).

4. A. A. Abbasi *et al.*, *PLoS ONE* **2**, e366 (2007).

5. J. Banerji, S. Rusconi, W. Schaffner, *Cell* **27**, 299-308 (Dec, 1981).

6. M. A. Nobrega, I. Ovcharenko, V. Afzal, E. M. Rubin, *Science* **302**, 413 (Oct 17, 2003).

7. E. de la Calle-Mustienes *et al.*, *Genome Res* **15**, 1061-72 (Aug, 2005).

8. L. A. Lettice *et al.*, *Hum Mol Genet* **12**, 1725-35 (Jul 15, 2003).

9. T. Sagai, M. Hosoya, Y. Mizushina, M. Tamura, T. Shiroishi, *Development* **132**, 797-803 (Feb, 2005).

10. A. Visel *et al.*, *Nat Genet* **40**, 158-60 (Feb, 2008).

11. L. A. Pennacchio *et al.*, *Nature* **444**, 499-502 (Nov 23, 2006).

12. A. Woolfe *et al.*, *PLoS Biol* **3**, e7 (Jan, 2005).

13. E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution* (Academic Press, Burlington, MA, ed. 1st, 2006).

14. S. B. Carroll, J. Grenier, S. Weatherbee, *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Wiley-Blackwell, Oxford, ed. 2nd, 2005).

15. C. V. Kirchhamer, L. D. Bogarad, E. H. Davidson, *Proc Natl Acad Sci U S A* **93**, 13849-54 (Nov 26, 1996).

16. S. Gray, P. Szymanski, M. Levine, *Genes Dev* **8**, 1829-38 (Aug 1, 1994).

17. R. Kothary *et al.*, *Development* **105**, 707-14 (Apr, 1989).

18. R. J. Britten, E. H. Davidson, *Q Rev Biol* **46**, 111-38 (Jun, 1971).

19. M. Kermekchiev, M. Pettersson, P. Matthias, W. Schaffner, *Gene Expr* **1**, 71-81 (Apr, 1991).

20. C. B. Lowe, G. Bejerano, D. Haussler, *Proc Natl Acad Sci U S A* **104**, 8005-10 (May 8, 2007).

21. A. M. Santangelo *et al.*, *PLoS Genet* **3**, 1813-26 (Oct, 2007).

22. G. Bejerano *et al.*, *Nature* **441**, 87-90 (May 4, 2006).

23. A. Siepel *et al.*, *Genome Res* **15**, 1034-50 (Aug, 2005).

24. G. Bejerano *et al.*, *Science* **304**, 1321-5 (May 28, 2004).

Figure Legends

Figure 1

Spatial additivity of tissue-specific enhancers fused from different genes. a) genomic environment of six conserved enhancers used in this study. A 50 kb genomic interval bracketing each enhancer is shown, including intron/exon structure of overlapping genes (black) and conservation in 17 vertebrates (color shaded boxes; (*23*)). All enhancers included an ultraconserved core region (suppl. table 1; (*24*)). b) Single and compound enhancer constructs for *in vivo* testing. c) Enhancer activity of single elements at mouse embryonic day 11.5. d)-i) *In vivo* activity of heterologous compound enhancers. d) E1+E2, e) E2+E5, f) E3+E4, g) E1+E5, h) E5+E6, i) E1+E2+E5+E6. Only one representative embryo is shown for each single and compound pattern, see supplemental table 2 for reproducibility across independent transgenic animals.

Figure 2

Temporal and spatial additivity of individual enhancer activities within the developing limb. a)-l) dorsal surface view of forelimb buds of individual embryos transgenic for E3 (a-d), E4 (e-h) and the compound enhancer E3+E4 (i-l). For each construct, embryos collected at e10.5 (a,e,i), e11.5 (b,f,j), and e12.5 (c,d,g,h,k,l) and representative limbs were stage-matched based on morphology. The transgenic status of the embryo shown in e) was confirmed by genotyping.

Figure 1

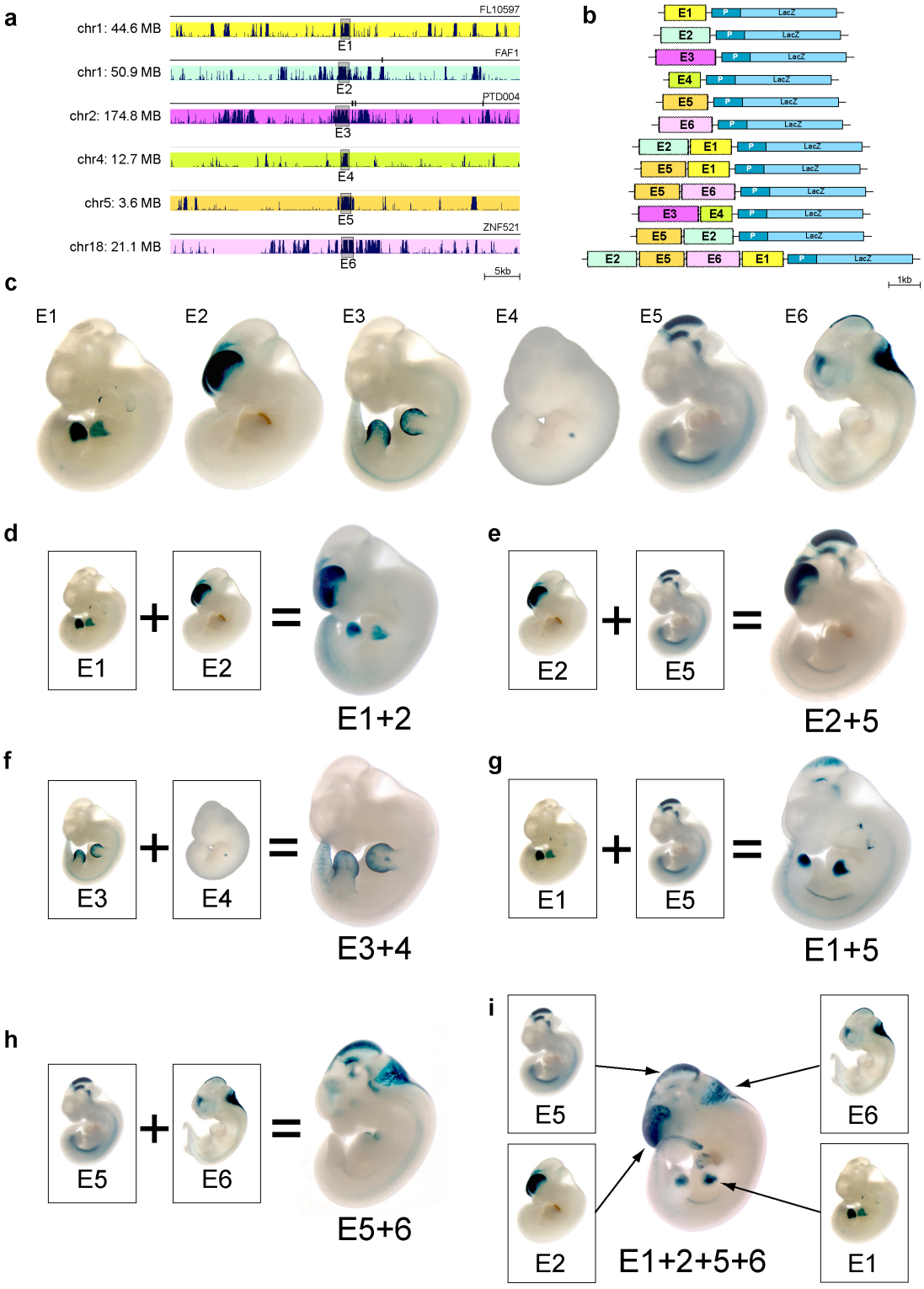
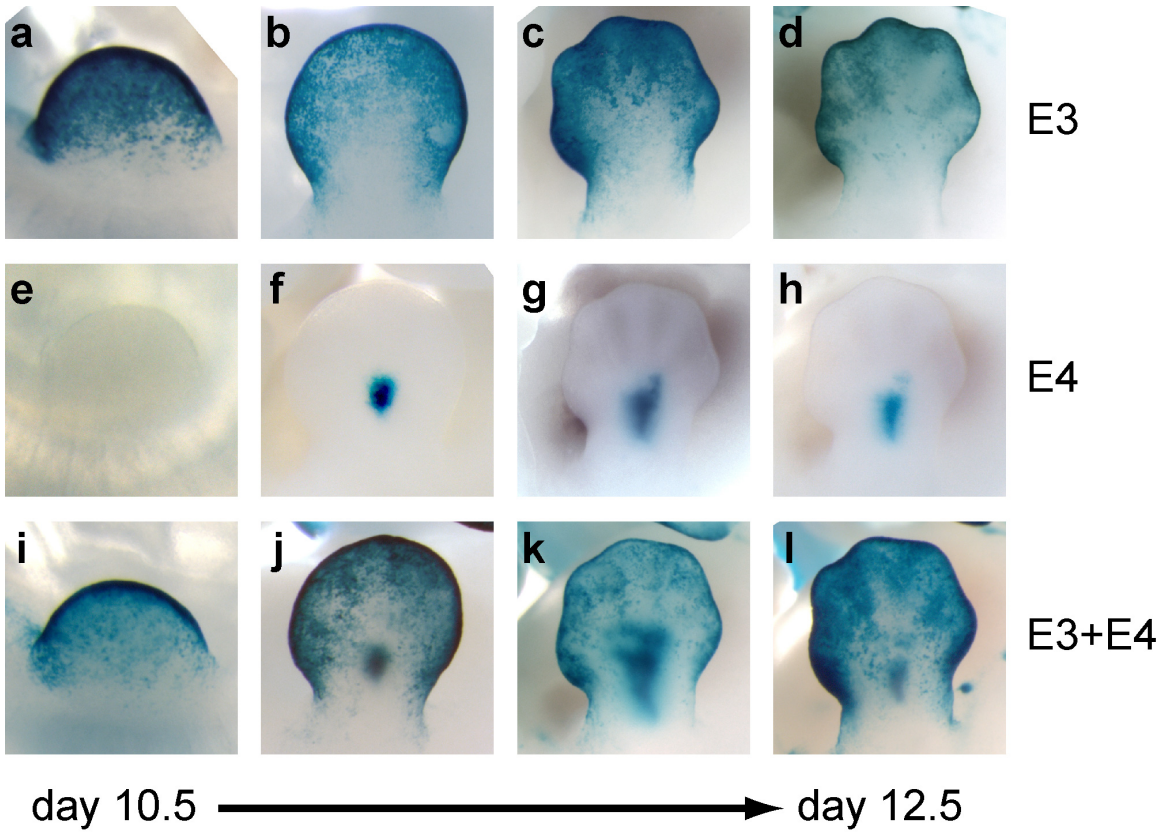

a
chr1: 44.6 MB
E1
FL10597
chr1: 50.9 MB
E2
FAF1
chr2: 174.8 MB
E3
PTD004
chr4: 12.7 MB
E4
chr5: 3.6 MB
E5
chr18: 21.1 MB
E6
ZNF521
5kb
b
E1
E2
E3
E4
E5
E6
E2 E1
E5 E1
E5 E6
E3 E4
E5 E2
E2 E5 E6 E1
P
LacZ
1kb
c
E1
E2
E3
E4
E5
E6
d
E1
E2
E1+2
e
E2
E5
E2+5
f
E3
E4
E3+4
g
E1
E5
E1+5
h
E5
E6
E5+6
i
E5
E6
E2
E1
E1+2+5+6

Figure 2



Supplemental table 1

Constructs	
construct name	*order and orientation of DNA fragments*
E1	E1(f)-Hsp68-LacZ
E2	E2(f)-Hsp68-LacZ
E3	E3(f)-Hsp68-LacZ
E4	E4(f)-Hsp68-LacZ
E5	E5(r)-Hsp68-LacZ
E6	E6(f)-Hsp68-LacZ
E1+E2	E2(f)-MF(f)-E1(f)-Hsp68-LacZ
E3+E4	E3(f)-MF(f)-E4(f)-Hsp68-LacZ
E1+E5	E5(r)-MF(f)-E1(f)-Hsp68-LacZ
E2+E5	E5(r)-MF(f)-E2(f)-Hsp68-LacZ
E5+E6	E5(r)-MF(f)-E6(f)-Hsp68-LacZ
E1+E2+E5+E6	E2(f)-aagggtgggcgcgcc-E5(r)-MF(f)-E6(f)-ggcgcgccgacccagctttcttgtacaaagtggtgatgccc-E1(f)-Hsp68-LacZ
(f) = orientation of fragment as indicated below or in human genome (hg18) (r) = reverse orientation of fragment as indicated below or in human genome (hg18)	

DNA frag-ments	**ultra-conserved region included** (*24*)	**VISTA enhancer browser ID** (*11*)	**PCR primers**	**coordinates (human genome, hg18)**
E1	uc.19	#280	fw: 5'-cacctcagcaagggctcgtaaag-3' rv: 5'-ggcagcagtttcaaggttct-3'	chr1:44762411-44763736
E2	uc.25	#200	fw: 5'-cacctggttggctcaaataaatgg-3' rv: 5'-ggcagtgaattcaagccttt-3'	chr1:50937783-50939374
E3	uc.104	#243	fw: 5'-caccgtggccaaagtcacaaacct-3' rv: 5'-actgtgtgtggggaaagacc-3'	chr2:174693790-174695710
E4	uc.140	#259	fw: 5'-caccttcatccccaggcttaat-3' rv: 5'-ccaatagctccttgcctctc-3'	chr4:12618416-12619425
E5	uc.150	#261	fw: 5'-caccgagaggaatgcccctctctt-3' rv: 5'-ccacctctttgctcctgaag-3'	chr5:3564978-3566399
E6	uc.425	#369	fw: 5'-caccgcctggaaggaaacaagatg-3' rv: 5'-gcagacattggctctcctct-3'	chr18:21118751-21120455
MF	multiple cloning site fragment, 5'-aagggtgggcgcgccgacccagctttcttgtacaaagtggtgatgccc-3'			
Hsp68	Hsp68 minimal promoter as previously described (*11, 17*)			
LacZ	beta-galactosidase reporter gene as previously described (*11, 17*)			

Suppl. Table 1: Sequences and constructs used in this study.

Supplemental table 2

				Reproducibility by Single Structures												
Construct	Total e11.5 (LacZ-expressing transgenics)	Reproducible	Ectopic	forebrain - cortex	forebrain - medial	midbrain - dorsal	midbrain - lateral	hindbrain - roof	hindbrain - ventral	limb - distal triangular	limb - ridge	limb - medial	heart	abdominal stripe	tail
E1	4	4	0	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	4/4	n.r.	n.r.	n.r.	n.r.	2/4
E2	11	10	1	10/10	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
E3	10	9	1	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	9/9	n.r.	n.r.	n.r.	n.r.
E4	15	11	4	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	11/11	n.r.	n.r.	n.r.
E5	8	8	0	n.r.	n.r.	7/8	8/8	n.r.	7/8	n.r.	n.r.	n.r.	n.r.	8/8	n.r.
E6	9	9	0	n.r.	8/9	8/9	n.r.	9/9	n.r.	n.r.	n.r.	n.r.	5/9	n.r.	n.r.
E1+E2	11	9	2	9/9	n.r.	n.r.	n.r.	n.r.	n.r.	8/9	n.r.	n.r.	n.r.	n.r.	6/9
E1+E5	11	11	0	n.r.	n.r.	10/11	11/11	n.r.	11/11	10/11	n.r.	n.r.	n.r.	9/11	6/11
E5+E6	11	11	0	n.r.	10/11	11/11	11/11	11/11	9/11	n.r.	n.r.	n.r.	10/11	11/11	n.r.
E3+E4	15	14	1	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	14/14	9/14*	n.r.	n.r.	n.r.
E2+E5	8	7	1	7/7	n.d.**	7/7	7/7	n.r.	7/7	n.r.	n.r.	n.r.	n.r.	7/7	n.r.
E1+E2+E5+E6	10	9	1	8/9	n.d.**	8/9	8/9	8/9	7/9	8/9	n.r.	n.r.	9/9	8/9	6/9

Suppl. Table 2: Reproducibility of staining in individual structures observed with single and tandem enhancer constructs at embryonic day 11.5. Embryos listed as "reproducible" had staining in at least one structure that was reproducible for this construct. Embryos that did not have staining in any reproducible structure are listed as "ectopic". (*) five transgenic embryos in which the medial pattern was absent were collected at e11.5, but were e11.0 or e11.25 as determined by limb morphology. (**) not determined, since in most cases obscured by strong cortex staining. n.r., no reproducible staining observed.