

# UCSF

## UC San Francisco Previously Published Works

### Title

A high-resolution enhancer atlas of the developing telencephalon.

### Permalink

<https://escholarship.org/uc/item/51k4p4qg>

### Journal

Cell, 152(4)

### ISSN

0092-8674

### Authors

Visel, Axel  
Taher, Leila  
Girgis, Hani  
et al.

### Publication Date

2013-02-01

### DOI

10.1016/j.cell.2012.12.041

Peer reviewed



Published in final edited form as:

*Cell*. 2013 February 14; 152(4): 895–908. doi:10.1016/j.cell.2012.12.041.

## A High-Resolution Enhancer Atlas of the Developing Telencephalon

**Axel Visel<sup>1,2,\*</sup>, Leila Taher<sup>3,9</sup>, Hani Girgis<sup>3</sup>, Dalit May<sup>1,10</sup>, Olga Golonzhka<sup>4</sup>, Renee Hoch<sup>4</sup>, Gabriel L. McKinsey<sup>4</sup>, Kartik Pattabiraman<sup>4</sup>, Shanni N. Silberberg<sup>4</sup>, Matthew J. Blow<sup>2</sup>, David V. Hansen<sup>5,6,11</sup>, Alex S. Nord<sup>1</sup>, Jennifer A. Akiyama<sup>1</sup>, Amy Holt<sup>1</sup>, Roya Hosseini<sup>1</sup>, Sengthavy Phouanavong<sup>1</sup>, Ingrid Plajzer-Frick<sup>1</sup>, Malak Shoukry<sup>1</sup>, Veena Afzal<sup>1</sup>, Tommy Kaplan<sup>7,8</sup>, Arnold R. Kriegstein<sup>5,6</sup>, Edward M. Rubin<sup>1,2</sup>, Ivan Ovcharenko<sup>3</sup>, Len A. Pennacchio<sup>1,2</sup>, and John L. R. Rubenstein<sup>4</sup>**

<sup>1</sup>Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>2</sup>U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

<sup>3</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

<sup>4</sup>Department of Psychiatry, Rock Hall, University of California at San Francisco, San Francisco, CA 94158-2324, USA

<sup>5</sup>Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA 94143, USA

<sup>6</sup>Department of Neurology, University of California, San Francisco, CA 94143, USA

<sup>7</sup>Department of Molecular and Cell Biology, California Institute of Quantitative Biosciences, University of California, Berkeley, CA 94720, USA

<sup>8</sup>School of Computer Science and Engineering, The Hebrew University, Jerusalem, Israel

### Summary

The mammalian telencephalon plays critical roles in cognition, motor function, and emotion. While many of the genes required for its development have been identified, the distant-acting regulatory sequences orchestrating their *in vivo* expression are mostly unknown. Here we describe a digital atlas of *in vivo* enhancers active in subregions of the developing telencephalon. We identified over 4,600 candidate embryonic forebrain enhancers and studied the *in vivo* activity of 329 of these sequences in transgenic mouse embryos. We generated serial sets of histological brain sections for 145 reproducible forebrain enhancers, resulting in a publicly accessible web-based data collection comprising over 32,000 sections. We also used epigenomic analysis of human and mouse cortex tissue to directly compare the genome-wide enhancer architecture in these species. These data provide a primary resource for investigating gene regulatory mechanisms of telencephalon development and enable studies of the role of distant-acting enhancers in neurodevelopmental disorders.

---

\*Correspondence: [avisel@lbl.gov](mailto:avisel@lbl.gov) (A.V.).

<sup>9</sup>Present address: Institute for Biostatistics and Informatics in Medicine and Ageing Research, University of Rostock, Germany.

<sup>10</sup>Present address: Department of Family Medicine, Clalit Health Services, and The Hebrew University-Hadassah Medical School, Jerusalem, Israel.

<sup>11</sup>Present address: Department of Neuroscience, Genentech, Inc., South San Francisco, CA 94080, USA.

## Introduction

The telencephalon houses the cerebral cortex and basal ganglia, structures that are pivotal for human brain functions (Wilson and Rubenstein, 2000). Impaired telencephalic development and function are associated with major neurological and neuropsychiatric disorders including mental deficiency, cerebral palsy, epilepsy, schizophrenia and autism (Lewis and Sweet, 2009; Walsh et al., 2008a). Significant progress has been made towards defining spatially resolved gene expression patterns in the developing and adult mouse and human brain on a genomic scale (Diez-Roux et al., 2011; Gong et al., 2003; Gray et al., 2004; Lein et al., 2007; Portales-Casamar et al.; Visel et al., 2004; Zeng et al., 2012). In contrast, the distant-acting gene regulatory sequences that are critical for orchestrating the spatial and temporal expression of genes in the developing and adult brain remain poorly defined, despite evidence from large-scale human genetic studies demonstrating the contribution of regulatory sequences to a wide spectrum of human traits and disorders (Durbin et al., 2010; Maurano et al., 2012), and despite anecdotal direct evidence for a critical requirement for enhancers in brain development (Kurokawa et al., 2004; Shim et al., 2012).

Unlike protein-coding genes, enhancers involved in specific biological processes are difficult to identify because they reside in the vast and poorly characterized non-coding portion of the genome, and can be located hundreds of thousands of base pairs away from the promoters of the target genes they regulate (Lettice et al., 2003). The introduction of enhancer prediction methods based on extreme evolutionary conservation (Nobrega et al., 2003; Pennacchio et al., 2006; Visel et al., 2008) and chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Visel et al., 2009a) increased the efficiency of identifying enhancers. Importantly, ChIP-seq experiments performed directly on tissues can provide accurate predictions of the broad general anatomical region in which an enhancer is active (Visel et al., 2009a). Nevertheless, the spatial resolution of these methods is limited and detailed *in vivo* studies are required to precisely define the activity patterns of enhancers at high resolution.

To address the need for an improved understanding of the *cis*-regulatory architecture and gene networks active during telencephalic development, we combined sequence conservation- and ChIP-seq-based enhancer prediction with large-scale histological activity analysis of human telencephalon enhancers in transgenic mice. We demonstrate how the high-resolution neuroanatomical annotation of enhancer activities can be used to develop computational sequence classifiers for enhancers active in different subregions of the telencephalon. We also directly compare the genome-wide enhancer architecture active in the mouse and human cortex using ChIP-seq from these tissues, and provide examples of downstream applications for enhancers identified through this work.

## Results

### Genome-Wide Identification of Candidate Forebrain Enhancers

To generate a genome-wide set of forebrain enhancer candidate sequences, we collected forebrain tissue from embryonic day [e]11.5 mouse embryos and performed tissue-ChIP-seq using an antibody for the enhancer-associated protein p300. Results were analyzed alongside previously described data to increase sampling depth (see Supplemental Methods). Genome-wide enrichment analysis led to the identification of 4,425 non-coding regions genome-wide that are distal from transcription start sites and significantly enriched in p300 binding in the e11.5 forebrain (Suppl. Table S1). Since p300 was previously shown to be associated with active tissue-specific enhancers (Blow et al., 2010; Visel et al., 2009a), these sequences were predicted to be distant-acting forebrain enhancers. As a complementary approach to

identify candidate enhancers, we also used extreme sequence conservation in conjunction with genomic location. Thus, we scrutinized sequences under extreme evolutionary constraint (Siepel et al., 2005; Visel et al., 2008) in the genomic vicinity of 79 genes with a known role in forebrain development or function (Suppl. Table S2), and identified 231 additional candidate forebrain enhancer sequences (Suppl. Table S3). Combined, these two datasets comprised a total of 4,656 noncoding sequence elements that we hypothesized to be enriched in forebrain enhancers.

### Transgenic Validation and Characterization of Enhancers

To validate candidate telencephalon enhancer sequences and define their *in vivo* activities in more detail, we selected 329 elements predicted to be enhancers by conservation and/or ChIP-seq for experimental testing (Suppl. Table S4). Nearly all of these selected elements were located near genes with a known function in the forebrain (Suppl. Table S2). In order to focus on the most conserved core regulatory architecture of mammalian telencephalon development, only ChIP-seq peaks that were detectably conserved between the human and mouse genome were tested. Regardless of the identification method, all tested sequences showed evidence of significant evolutionary constraint (phastCons scores ranging from 415 to 931, median 798; Suppl. Table S4). The selected candidate enhancer sequences were amplified from human genomic DNA, cloned into an enhancer reporter vector (Hsp68-LacZ), and used to generate transgenic mice by pronuclear injection (see **Methods**). Transgenic embryos were stained for reporter gene (LacZ) activity at e11.5 and reporter expression was annotated using established reproducibility criteria (Pennacchio et al., 2006). Only elements that drove expression in the forebrain in at least three embryos, each of them corresponding to an independent transgenic integration event, were considered as reproducible forebrain enhancers. In total, 105 of 329 (32%) candidate sequences tested were reproducible forebrain enhancers at e11.5, of which 36 showed reproducible expression exclusively in the forebrain (Suppl. Table S4). For comparison, in previous transgenic assays of p300 binding sites in two different non-neuronal tissues, limb buds and heart, only 4/155 (2.6%) tested sequences had reproducible forebrain enhancer activity at e11.5 (Blow et al., 2010; Visel et al., 2009a). Enhancer candidate sequences that overlapped p300 ChIP-seq peaks were more enriched in verifiable *in vivo* forebrain enhancers than extremely conserved sequences that showed no evidence of p300 binding (58% compared to 23%, Suppl. Table S4). Selected examples of reproducible forebrain enhancers whose *in vivo* activity was confirmed in transgenic mice are shown in Fig. 1 and whole-mount images for all validated enhancers are accessible online through the Vista Enhancer Browser (Visel et al., 2007).

### High-Resolution Analysis of Telencephalon Enhancer Activity Patterns

To define the precise spatial expression patterns of telencephalic enhancers active at e11.5, we performed high-resolution analysis on a set of 145 enhancers (Suppl. Table S5). These sequences were selected from the 105 forebrain enhancers discovered in the present study and from complementary sets of forebrain enhancers identified at whole-mount resolution in previous enhancer screens (Pennacchio et al., 2006; Visel et al., 2008; Visel et al., 2009a). For each enhancer, a full set of contiguous coronal paraffin sections (average: 220 sections) was obtained. Full-resolution digital images of over 32,000 sections are available through the Vista Enhancer Browser (Visel et al., 2007). Selected sections of patterns driven by different enhancers in subregions of the pallium and subpallium are shown in Figures 2 and 3, illustrating the diversity of spatial specificities observed.

In addition to the spatial activity patterns of all 145 enhancers studied at e11.5, we also examined the temporal activities of a subset of these enhancers at later prenatal stages of telencephalon development (Figures 2S and 3F-G). These temporal comparisons showed

that the spatial patterns of enhancer activity were largely constant. In two cases, enhancers active in subregions of the subpallium at e11.5 displayed characteristic features of subpallial cell populations (interneurons) that tangentially migrate to the pallium. At e13.5, these cells had just arrived in the ventrolateral pallium (hs692 and hs799), and by e15.5 they were in the dorsal pallium (hs799, arrowheads in Fig. 3F-G). These results support that enhancers regulate both spatial and temporal aspects of telencephalic gene expression in patterns consistent with the biology of these regions and cell types.

To facilitate analysis by computational methods, we devised a standardized neuroanatomical annotation scheme for the e11.5 stage of telencephalon development (Fig. 2A and 3C, Suppl. Fig. S1, Suppl. Table S5). All telencephalon enhancer activity patterns examined in this study were annotated using this standardized annotation scheme, in some cases complemented by descriptions that further subdivide the standardized domains or are restricted to subsets of cells (Suppl. Table S5). The standardized annotations assigned to each enhancer through this annotation effort enable computational analysis of enhancer activity patterns, as well as a comparison to expression patterns of their presumptive target genes at this stage of development.

### Comparison of Enhancer Activities to Gene Expression Patterns

To test whether the telencephalon enhancers examined at high resolution generally recapitulate the spatial expression patterns of their presumptive target genes, we compared their LacZ reporter activities to RNA *in situ*-hybridization data. For example, the *Arx* gene is expressed both in subpallial and pallial regions, with increasing expression in pallial regions from e11.5 to e13.5 (Fig. 4A). We found that there are at least four distant-acting telencephalic enhancers in this extended locus, two of which drive subpallial and two of which drive pallial expression, indicating that developmental *Arx* regulation is more complex than initially suggested (Colasante et al., 2008). In addition, comparison of other genes with well-established roles in telencephalon development (*Lef1*, *Wnt8b*, *Gsx2*, *Nr2f1*) to nearby enhancers also revealed examples of spatially concordant enhancer activity and RNA expression (Fig. 4B-E). A recurring feature of these comparisons is the restriction of individual enhancer activities to subregions of the respective gene expression patterns, supporting the modular structure of telencephalic enhancer architecture. For instance, hs687 activity in the LGE matches *Gsx2* RNA expression, while the latter is also expressed in the MGE and hs1172 activity in the pallium matches *Nr2f1* RNA expression, while the gene is also expressed in the subpallium.

To assess whether these illustrative examples are representative of a general congruence between enhancer activity patterns and the expression of nearby genes, we performed a quantitative correlation analysis across the available data set (see Supplemental Experimental Procedures for details). Overall, we found a highly significant correlation between the activity patterns of enhancers and telencephalic expression patterns of nearby annotated genes ( $P=0.0003$ , Mann-Whitney test, Fig. 4F). In addition to the high-resolution comparisons of enhancer and gene activity patterns, we also examined whether the genome-wide set of 4,425 forebrain enhancer candidate sequences identified by ChIP-seq from embryonic mouse forebrain tissue is associated with genes with known functions in the telencephalon. Unbiased genome-wide assessment (McLean et al., 2010) showed highly significant enrichment in genes that cause forebrain-related phenotypes when deleted in mouse models (Suppl. Table S6). These observations support on a genomic scale that the large set of forebrain candidate enhancers predicted by ChIP-seq in this study is enriched near genes that are involved in telencephalon development.

## Sequence Analysis of Subregion-Specific Enhancers

A large set of telencephalon enhancers, analyzed at high spatial resolution and annotated to a standardized scheme, offers the possibility to examine sequence features that are associated with *in vivo* activity in different telencephalic subregions. To explore this regulatory code, we trained a Random Forests (RF) classifier (Breiman, 2001; Bureau et al., 2005; Cummings and Segal, 2004; Lunetta et al., 2004) to discriminate between enhancers active in 1. pallium only, 2. pallium and subpallium (compound pattern), or 3. subpallium only, and random genomic sequences (see Fig. 5 and Supplemental Methods). Classification is based on the presence or absence of combinations of sequence motifs matching known transcription factor binding sites (Bryne et al., 2008; Matys et al., 2006). The five most relevant motifs distinguishing the three classes of enhancers and their respective importance are shown in Fig. 5B (for additional motifs, see Suppl. Fig. S2 and Suppl. Table S8). We did not observe any single motif that was sufficient to accurately discriminate between the different classes of enhancers, suggesting that only the combinatorial binding of multiple transcription factors determines the observed spatial regulatory activity. The majority of the most discriminatory motifs (at least 60% of the top 15 motifs characterizing enhancers active in each of the telencephalic subregions considered) correspond to predicted binding sites for homeodomain-containing transcription factors, consistent with the known critical role of these proteins in telencephalon development (Hebert and Fishell, 2008). Suppl. Fig. S3 summarizes the enrichment of the 15 most relevant motifs for enhancer activity in the three different telencephalic subregions considered. Despite possible ambiguities associated with computational transcription factor binding site predictions, the RF classifier accurately predicts approximately 80% of the sequences (see Supplemental Methods, Suppl. Table S9). Sequence motifs with high quantitative importance for discriminating between different classes of telencephalon enhancers are overall more conserved in evolution compared to non-important motifs, supporting their functional relevance (Suppl. Fig. S4).

These computational predictions of relevant sequence motifs provide a starting point for experimental studies aimed at understanding the transcription factor binding site content of telencephalon enhancers in more detail. To illustrate the value of a large set of enhancers with known sequences and activity patterns for studying genetic dependencies in telencephalon development, we tested a subset of subpallial enhancers for their direct regulation by two major subpallial transcription factors, Dlx2 and Ascl1 (see Supplemental Methods). In a cell-based luciferase assay, we observed that Dlx2 and/or Ascl1 significantly increased reporter expression when co-transfected with 13 of 20 tested enhancers (Fig. 5C). Of note, these enhancers are located near several genes with known roles in subpallium development and the results are consistent with previous studies demonstrating that Dlx2 regulates the expression of Arx, Meis2 and Sp8, and that Ascl1 regulates the expression of Sox4 (Castro et al., 2011; Colasante et al., 2008; Long et al., 2009). Considering the expected complexity of the spectrum of transcription factors binding to different subsets of telencephalon enhancers (Fig. 5B, Suppl. Table S8), complementary scalable methods will be required to experimentally validate all binding sites within each of the enhancers identified. Our cell-based studies of a small subset of these sequences highlight, however, that the combined knowledge of the genomic location, the spatial activity, and the upstream transcription factors of discrete, distant-acting regulatory sequences generates hypotheses that are directly testable in genetic *in vivo* systems.

## Human Brain CHIP-seq

Our large-scale transgenic testing and high-resolution analysis of telencephalon enhancers focused on sequences that are highly conserved in evolution, with the goal to characterize the most conserved core regulatory architecture of mammalian telencephalon development. However, epigenomic methods also enable the systematic discovery of poorly conserved

and lineage-specific enhancers (Schmidt et al., 2010). To explore possible differences between human and mouse telencephalon enhancers in more detail, we determined the genome-wide occupancy of the enhancer-associated proteins p300/CBP in human fetal (gestational week 20) cortex (Fig. 6A,B). ChIP-seq analysis identified 2,275 peaks (candidate enhancers) genome-wide that were located at least 2.5kb from the nearest transcript start site. Comparison with transcriptome data from human fetal cortex tissue revealed a 2.7-fold enrichment in candidate enhancers within 2.5-20kb of the transcript start sites of genes highly expressed in fetal human cortex ( $P < 1e-14$ , binomial distribution), with significant enrichment up to 220kb away from promoters ( $P < 0.001$ , binomial distribution, Fig. 6C). In contrast, no enrichment of p300/CBP binding sites was observed near genes highly expressed in other tissues. Similar to candidate enhancers predicted from mouse e11.5 forebrain, unsupervised statistical enrichment analysis of functional gene annotations (McLean et al., 2010) showed significant association with genes implicated in nervous system-related phenotypes (Suppl. Table S6). While many extremely conserved non-coding sequences in the human genome are enhancers active in the developing nervous system (Pennacchio et al., 2006), we observed that a third (36.5%) of ChIP-seq-predicted human brain candidate enhancers are under weak (phastCons  $< 350$ ) or no detectable evolutionary constraint, suggesting that subsets of human brain enhancers may not be functionally conserved in mice.

At gestational week 20, the human cortex is considerably further developed than the mouse pallium at embryonic day 11.5, and instead corresponds broadly to early postnatal stages in mouse (Clancy et al., 2007). To enable a direct experimental comparison between the two species, we performed p300/CBP ChIP-seq on mouse postnatal (P0) cortex tissue. Using identical methods as for human tissue, we identified 1,132 candidate enhancers (distal ChIP-seq peaks). The majority (58%) of human-derived peaks showed significant or suggestive (sub-significant) enrichment in ChIP-seq reads at the orthologous site in the mouse genome (Fig. 6D). The remaining 42% either showed no enrichment in the orthologous mouse region or were not alignable to the mouse genome. While the lower sequencing coverage in the mouse data set may lead to an underestimation of mouse-compared to human-specific peaks (compare Fig. 6D/E), the presence of 307 peaks in non-alignable regions of the human genome (Fig. 6D) supports that a non-negligible proportion of human brain enhancers emerged in evolution after the divergence of primates and rodents from their last common ancestor.

Similar to the large collection of telencephalon enhancers identified and characterized at e11.5, ChIP-seq peaks derived from human fetal cortex are expected to include enhancers with a variety of *in vivo* activity patterns. To illustrate this, we examined the *in vivo* activities of candidate enhancers from human fetal cortex in postnatal transgenic mice. Two examples of such enhancers driving reproducible expression in a minimum of three independent transgenic animals are shown in Fig. 6F-K. Consistent with the ChIP-seq prediction, both enhancers were active in the cortex (red arrows), as well as additional, but distinct and reproducible regions of the telencephalon.

To illustrate the value of the genome-wide sets of human and mouse candidate enhancers for the interpretation of human genetic datasets, we compared the genomic position of these sequences with different catalogs of regions in the human genome implicated in neurodevelopmental, neurological or neuropsychiatric diseases. We intersected the genome-wide sets of candidate enhancers identified in the three different ChIP-seq experiments a) with lead single nucleotide polymorphisms (SNPs) from genome-wide association studies of relevant traits (Hindorff et al., 2009), b) with catalogs of syndromic microdeletions and -duplications (Firth et al., 2009), and c) with a set of autism-associated rare copy number variants (Marshall et al., 2008; Szatmari et al., 2007). Fourteen lead SNPs from genome-

wide association studies, including SNPs associated with attention deficit hyperactivity disorder, bipolar disease and schizophrenia were found to be located within predicted forebrain enhancers. Moreover, 381 enhancers mapped within recurrent microdeletions or duplications associated with neurological phenotypes, and 421 enhancers overlapped copy number variants present in autism cases, but not healthy controls. While further experimental studies will be required to examine possible causal roles of variants affecting enhancer sequences, the genome-wide sets of candidate enhancers identified from human and mouse brain tissue through this study provide a starting point to explore the role of telencephalon enhancers in human diseases.

### Telencephalon Enhancers as Molecular Reagents

The enhancers described in our high-resolution atlas can be used as molecular reagents to drive *in vivo* expression of reporter or effector genes to specific telencephalic subregions of interest, owing to the reproducibility of their activity patterns (Fig. 7A). To illustrate some of the resulting applications, we coupled enhancer hs1006, associated with the *WNT8B* gene, to a minimal Hsp68 promoter, followed by a tamoxifen-inducible Cre recombinase (CreER<sup>T2</sup>), an internal ribosomal entry site, and a green fluorescent protein (GFP) reporter (Fig. 7B). In stable transgenic mouse lines generated with this construct, termed CT2IG-hs1006, GFP expression at e11.5 was indistinguishable from LacZ reporter expression (Fig. 7A/B). GFP expression in these stable lines facilitates a temporally resolved mapping of enhancer activity. A comparison of GFP activity at e12.5, e15.5, and e17.5 with *Wnt8b* RNA expression indicated that enhancer activity spatially coincides with *Wnt8b* gene expression, supporting that this enhancer controls region-specific expression of the gene over an extended period of prenatal telencephalon development.

Since expression of the compound effector/reporter transcript in CT2IG-hs1006 mice faithfully resembled *Wnt8b* expression across multiple stages of development, the chemically inducible CreER<sup>T2</sup> recombinase can be used for spatially and temporally highly restricted genomic recombineering applications, such as neuronal fate mapping studies. To demonstrate this, we crossed CT2IG-hs1006 mice with Rosa26-LacZ mice (Fig. 7B) (Indra et al., 1999). Tamoxifen induction of CreER<sup>T2</sup> in pregnant compound CT2IG-hs1006:Rosa26-LacZ mice at e10.5 leads to recombination only in the small proportion of pallial cells in which the enhancer is active at this time point. LacZ staining at later stages revealed the spatial fate of cells in which the enhancer was active at e10.5. For example, hs1006-driven e10.5 → e12.5 fate mapping marked pallial cell populations with a distribution that is clearly distinct from hs1006 activity at this time point (compare e12.5 patterns in Figures 7C and 7D). These data highlight the utility of these enhancers to precisely drive gene expression in the developing brain and their value as a rich resource for a diversity of uses.

### Discussion

This work provides a comprehensive resource for basic studies of telencephalon enhancers. Our targeted screen identified the genomic location of thousands of candidate enhancers putatively active in the embryonic forebrain. The mapping and annotation of the activity patterns of nearly 150 human telencephalon enhancers at histological resolution in transgenic mice provides insight into the regulatory architecture of individual genes that are required for forebrain development and will facilitate studies of molecular genetic pathways by identifying the genomic regions to which upstream transcription factors bind.

Our analysis revealed several cases of enhancers that drive similar patterns and are associated with the same gene (e.g. Fig. 4A) in a manner reminiscent of the ‘shadow enhancers’ observed in invertebrate models (Frankel et al., 2010; Hong et al., 2008). The

data provided through this work will support the identification of minor spatial activity differences between such enhancers, as well as the functional exploration of their apparent redundancies. It is also remarkable that a large proportion of enhancers examined in this study drove patterns that were at least partially different from all other enhancers examined, highlighting the complexity of the developing forebrain, as well as the regulatory sequence code orchestrating its development.

The motif-based classifiers derived from enhancers active in different subregions of the telencephalon demonstrate the value of systematically annotated enhancer activity data sets for computational studies aimed at deciphering the correlation between the transcription factor binding sites present in an enhancer and its precise spatial activity pattern. Beyond such functional genomic studies, the enhancers identified and characterized in this work provide a comprehensive set of molecular reagents that can be used to target gene expression to defined subregions of the developing brain, or to defined cell states when differentiating stem cells *in vitro*. This will enable tissue-specific homologous recombination and deletion strategies or expression of reporter and selectable genes, as illustrated in Fig. 7.

Finally, results from this study are expected to enable and facilitate the functional genomic exploration of the role of enhancers in human brain disorders. There is accumulating evidence that non-coding sequence variants, as well as copy number variation in coding and non-coding portions of the genome have important impacts on a wide spectrum of disorders including bipolar, schizophrenia, autism, intellectual disability and epilepsy (Cooper et al., 2011; Durbin et al., 2010; International Schizophrenia Consortium, 2008; Malhotra et al., 2011; Sebat et al., 2007; Vacic et al., 2011; Visel et al., 2009b; Walsh et al., 2008b). However, the functional interpretation of non-coding sequence or copy number variants remains a major challenge and few potentially causative connections linking neurological traits to molecular variation in enhancers have been identified (e.g., (Poitras et al., 2010)). Thus, the systematic mapping and high-resolution analysis of telencephalon enhancers through this work is expected to provide functional genomic insights to guide studies that will mechanistically relate individual non-coding sequence and copy number variants to brain disorders.

## Experimental Procedures

### Chromatin immunoprecipitation followed by sequencing (ChIP-seq)

ChIP-seq on forebrain tissue isolated from e11.5 CD-1 strain mouse embryos, using an antibody directed against p300, was performed according to previously described procedures (Visel et al., 2009a). For human tissue ChIP-seq and the matched mouse postnatal cortex data set, an anti-acCBP/p300 pan-specific antibody was used (May et al., 2011).

### Transgenic mouse assays

Enhancer candidate regions were analyzed in transgenic mouse embryos as previously described (Kothary et al., 1988; Pennacchio et al., 2006). Paraffin sections were prepared according to standard protocols. Serial sets of sections were digitally photographed and uploaded to the Vista Enhancer Browser (<http://enhancer.lbl.gov>).

### GFP reporter assays and cell fate mapping

A previously described Cre-ERT<sup>2</sup> construct (Feil et al., 1997) was modified to allow Cre recombinase expression to be driven by the hs1006 enhancer (Fig. 7B). For fate mapping, CT2IG-hs1006 mice were crossed with Rosa26-LacZ reporter mice (Soriano, 1999).

## Luciferase assays

Dlx2 and Ascl1 were selected for luciferase reporter assays due to their well-established roles in subpallial development and because they are representatives of two major groups of transcription factors found among the top motifs of the subpallium classifier (see Supplemental Experimental Procedures). P19 cells were grown by previously described methods (Farah et al., 2000).

## Accession Numbers, Data and Reagent Availability

Images of whole-mount-stained embryos and full sets of e11.5 coronal brain sections are available through the Vista Enhancer Browser, <http://enhancer.lbl.gov>. All enhancer reporter vectors described in this study are freely available from the authors. In addition, archived surplus transgenic embryos for many constructs can be made available upon request for complementary studies. The genome-wide set of ChIP-seq peaks derived from mouse e11.5 forebrain is provided in Suppl. Table S1. Raw data and additional ChIP-seq data sets from postnatal mouse and fetal human cortex are available from GEO under accession number GSE42881.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank Julian Golder and Noah Efron for help with digital image acquisition and data processing; Bing Ren and Zirong Li for help with chromatin immunoprecipitation from embryonic mouse tissue; Inna Dubchak, Simon Minovitsky and Alexandre Poliakov for web site support; staff at San Francisco General Hospital Women's Options Center for their consideration in allowing us to access donated fetal tissue. A.V. and L.A.P. were supported by NINDS grant R01NS062859A and by NHGRI grant R01HG003988. J.L.R.R. was supported by the Nina Ireland, Weston Havens Foundation, NINDS grant R01NS34661, NIMH grant R01MH081880, and NIMH grant R37MH049428. J.L.R.R. and A.R.K. were supported by CIRM RB2-1602. G.M. and S.N.S. were supported by T32 GM007449, K.P. was supported by T32 GM07618, R.H. by F32 MH081431 and O.G. by NARSAD. A.R.K. was supported by NINDS grant R01NS075998. I.O. was supported by the Intramural Research Program of the NIH, National Library of Medicine. Research was conducted at the E.O. Lawrence Berkeley National Laboratory and performed under Department of Energy Contract DE-AC02-05CH11231, University of California.

## References

- Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet.* 2010; 42:806–810. [PubMed: 20729851]
- Breiman L. Random Forests. *Machine Learning.* 2001; 45:5–32.
- Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* 2008; 36:D102–106. [PubMed: 18006571]
- Bureau A, Dupuis J, Falls K, Lunetta KL, Hayward B, Keith TP, Van Eerdewegh P. Identifying SNPs predictive of phenotype using random forests. *Genet Epidemiol.* 2005; 28:171–182. [PubMed: 15593090]
- Castro DS, Martynoga B, Parras C, Ramesh V, Pacary E, Johnston C, Drechsel D, Lebel-Potter M, Garcia LG, Hunt C, et al. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* 2011; 25:930–945. [PubMed: 21536733]
- Clancy B, Finlay BL, Darlington RB, Anand KJ. Extrapolating brain development from experimental species to humans. *Neurotoxicology.* 2007; 28:931–937. [PubMed: 17368774]
- Colasante G, Collombat P, Raimondi V, Bonanomi D, Ferrai C, Maira M, Yoshikawa K, Mansouri A, Valtorta F, Rubenstein JL, et al. Arx is a direct target of Dlx2 and thereby contributes to the

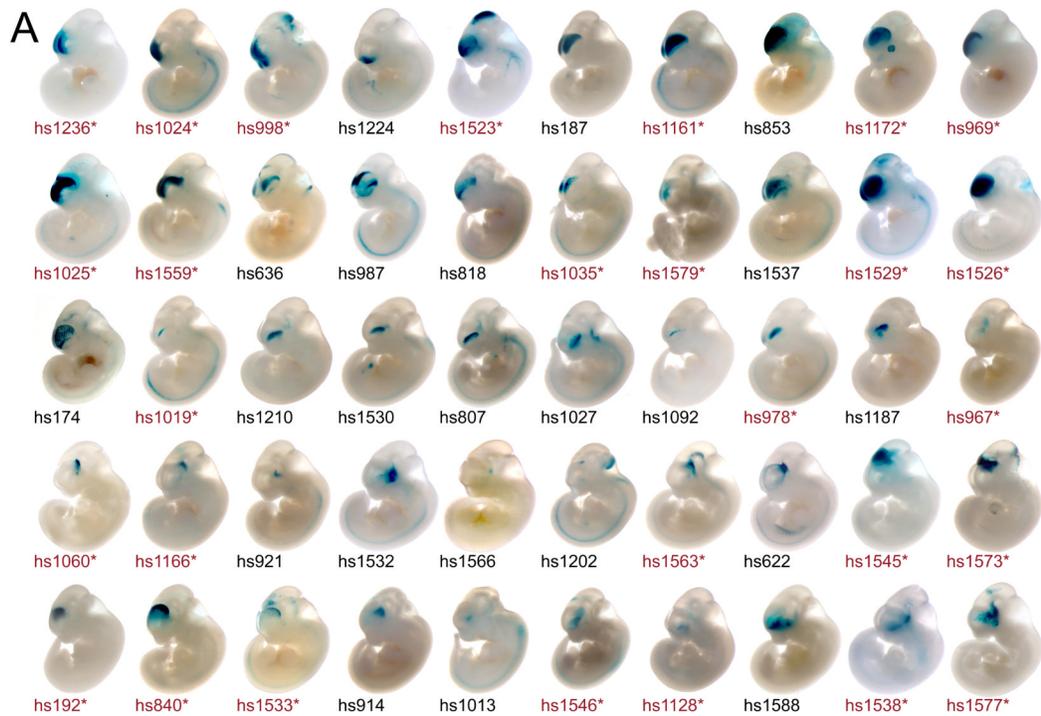
- tangential migration of GABAergic interneurons. *J Neurosci*. 2008; 28:10674–10686. [PubMed: 18923043]
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011; 43:838–846. [PubMed: 21841781]
- Cummings MP, Segal MR. Few amino acid positions in rpoB are associated with most of the rifampin resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics*. 2004; 5:137. [PubMed: 15453919]
- Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol*. 2011; 9:e1000582. [PubMed: 21267068]
- Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
- Farah MH, Olson JM, Sucic HB, Hume RI, Tapscott SJ, Turner DL. Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. *Development*. 2000; 127:693–702. [PubMed: 10648228]
- Feil R, Wagner J, Metzger D, Chambon P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun*. 1997; 237:752–757. [PubMed: 9299439]
- Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009; 84:524–533. [PubMed: 19344873]
- Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*. 2010; 466:490–493. [PubMed: 20512118]
- Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, et al. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*. 2003; 425:917–925. [PubMed: 14586460]
- Gray PA, Fu H, Luo P, Zhao Q, Yu J, Ferrari A, Tenzen T, Yuk DI, Tsung EF, Cai Z, et al. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science*. 2004; 306:2255–2257. [PubMed: 15618518]
- Hebert JM, Fishell G. The genetics of early telencephalon patterning: some assembly required. *Nat Rev Neurosci*. 2008; 9:678–685. [PubMed: 19143049]
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106:9362–9367. [PubMed: 19474294]
- Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. *Science*. 2008; 321:1314. [PubMed: 18772429]
- Indra AK, Warot X, Brocard J, Bornert JM, Xiao JH, Chambon P, Metzger D. Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Res*. 1999; 27:4324–4327. [PubMed: 10536138]
- International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008; 455:237–241. [PubMed: 18668038]
- Kothary R, Clapoff S, Brown A, Campbell R, Peterson A, Rossant J. A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature*. 1988; 335:435–437. [PubMed: 3138544]
- Kurokawa D, Takasaki N, Kiyonari H, Nakayama R, Kimura-Yoshida C, Matsuo I, Aizawa S. Regulation of Otx2 expression and its functions in mouse epiblast and anterior neuroectoderm. *Development*. 2004; 131:3307–3317. [PubMed: 15201223]
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. 2007; 445:168–176. [PubMed: 17151600]

- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet.* 2003; 12:1725–1735. [PubMed: 12837695]
- Lewis DA, Sweet RA. Schizophrenia from a neural circuitry perspective: advancing toward rational pharmacological therapies. *J Clin Invest.* 2009; 119:706–716. [PubMed: 19339762]
- Long JE, Swan C, Liang WS, Cobos I, Potter GB, Rubenstein JL. Dlx1&2 and Mash1 transcription factors control striatal patterning and differentiation through parallel and overlapping pathways. *J Comp Neurol.* 2009; 512:556–572. [PubMed: 19030180]
- Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet.* 2004; 5:32. [PubMed: 15588316]
- Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, Cichon S, Corvin A, Gary S, Gershon ES, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron.* 2011; 72:951–963. [PubMed: 22196331]
- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet.* 2008; 82:477–488. [PubMed: 18252227]
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 2006; 34:D108–110. [PubMed: 16381825]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–1195. [PubMed: 22955828]
- May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, et al. Large-scale discovery of enhancers from human heart tissue. *Nat Genet.* 2011; 44:89–93. [PubMed: 22138689]
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010; 28:495–501. [PubMed: 20436461]
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. *Science.* 2003; 302:413. [PubMed: 14563999]
- Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature.* 2006; 444:499–502. [PubMed: 17086198]
- Poitras L, Yu M, Lesage-Pelletier C, Macdonald RB, Gagne JP, Hatch G, Kelly I, Hamilton SP, Rubenstein JL, Poirier GG, et al. An SNP in an ultraconserved regulatory element affects Dlx5/Dlx6 regulation in the forebrain. *Development.* 2010; 137:3089–3097. [PubMed: 20702565]
- Portales-Casamar E, Swanson DJ, Liu L, de Leeuw CN, Banks KG, Ho Sui SJ, Fulton DL, Ali J, Amirabbasi M, Arenillas DJ, et al. A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proc Natl Acad Sci U S A.* 2010; 107:16589–16594. [PubMed: 20807748]
- Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science.* 2010; 328:1036–1040. [PubMed: 20378774]
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al. Strong association of de novo copy number mutations with autism. *Science.* 2007; 316:445–449. [PubMed: 17363630]
- Shim S, Kwan KY, Li M, Lefebvre V, Sestan N. Cis-regulatory control of corticospinal system development and evolution. *Nature.* 2012; 486:74–79. [PubMed: 22678282]
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005; 15:1034–1050. [PubMed: 16024819]
- Soriano P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet.* 1999; 21:70–71. [PubMed: 9916792]

- Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet.* 2007; 39:319–328. [PubMed: 17322880]
- Vacic V, McCarthy S, Malhotra D, Murray F, Chou HH, Peoples A, Makarov V, Yoon S, Bhandari A, Corominas R, et al. Duplications of the neuropeptide receptor gene *VIPR2* confer significant risk for schizophrenia. *Nature.* 2011; 471:499–503. [PubMed: 21346763]
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009a; 457:854–858. [PubMed: 19212405]
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* 2007; 35:D88–92. [PubMed: 17130149]
- Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet.* 2008; 40:158–160. [PubMed: 18176564]
- Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. *Nature.* 2009b; 461:199–205. [PubMed: 19741700]
- Visel A, Thaller C, Eichele G. [GenePaint.org](http://GenePaint.org): an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* 2004; 32:D552–556. [GenePaint.org](http://GenePaint.org) [PubMed: 14681479]
- Walsh CA, Morrow EM, Rubenstein JL. Autism and brain development. *Cell.* 2008a; 135:396–400. [PubMed: 18984148]
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science.* 2008b; 320:539–543. [PubMed: 18369103]
- Wilson SW, Rubenstein JL. Induction and dorsoventral patterning of the telencephalon. *Neuron.* 2000; 28:641–651. [PubMed: 11163256]
- Zeng H, Shen EH, Hohmann JG, Oh SW, Bernard A, Royall JJ, Glattfelder KJ, Sunkin SM, Morris JA, Guillozet-Bongaarts AL, et al. Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell.* 2012; 149:483–496. [PubMed: 22500809]

### Highlights

- Genome-wide screen for distant-acting enhancers active in the developing forebrain
- High-resolution mapping of *in vivo* enhancer activities in transgenic mice
- Development of computational sequence classifiers for telencephalon subregions
- Comparison of enhancer architecture in developing human and mouse cerebral cortex

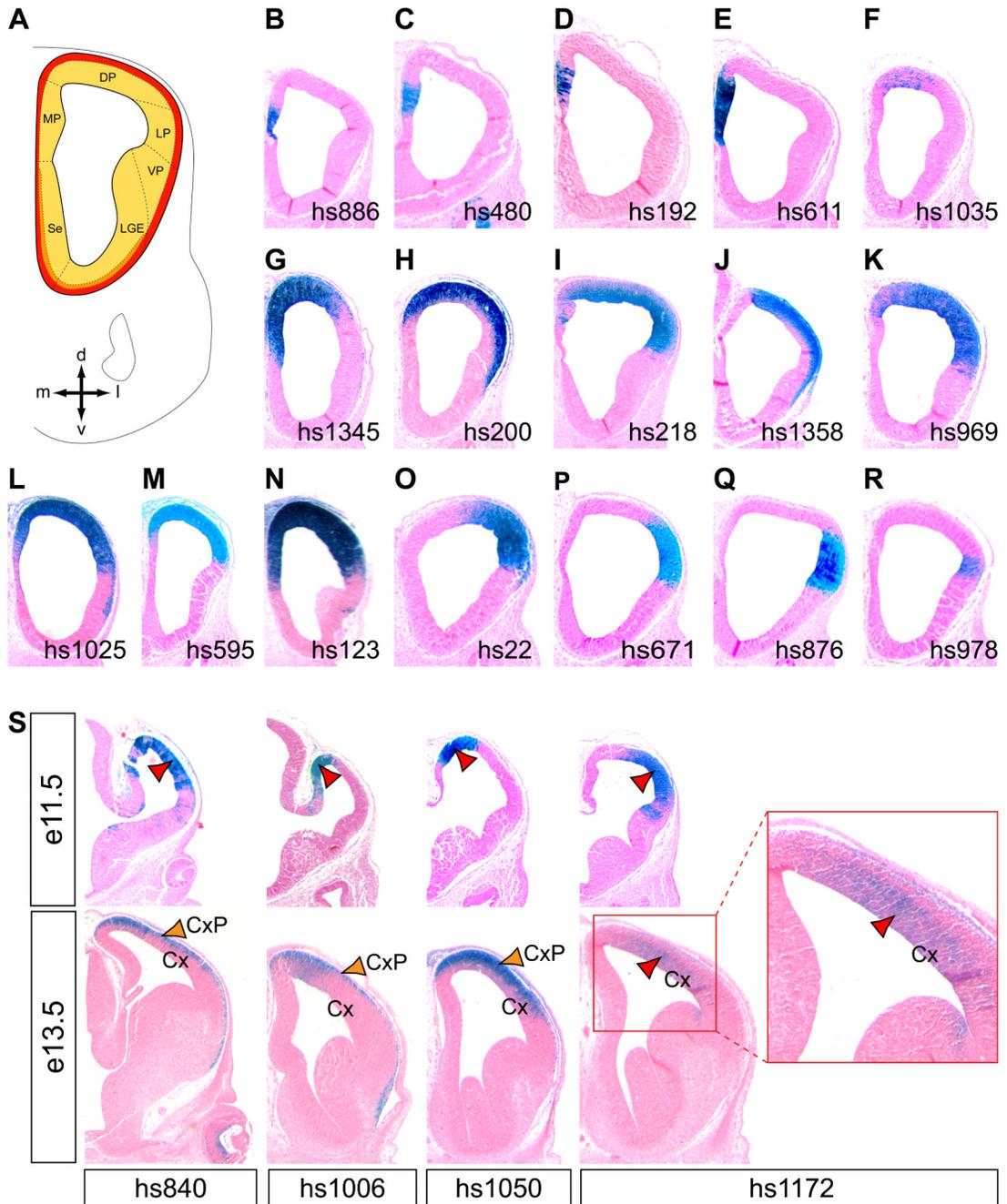


**B** \* enhancers near genes implicated in telencephalon development or function:

enhancer element(s)	gene symbol	gene name
hs967, hs998	ASCL1	achaete-scute complex homolog 1
hs1025	EMX1	empty spiracles homeobox 1
hs1236	EMX2	empty spiracles homeobox 2
hs840	EPHA5	Eph receptor A5
hs1523	FOXP1	forkhead box G1
hs1563	HES1	hairy and enhancer of split 1
hs1533	ID4	inhibitor of DNA binding 4
hs1545	LEF1	lymphoid enhancer binding factor 1
hs1161	LMO4	LIM domain only 4
hs1024, hs1060, hs1172, hs1577	NR2F1	nuclear receptor subfamily 2, group F, member 1 (COUP-TF1)
hs1579	OTX2	orthodenticle homeobox 2
hs1546	POU3F1	POU domain, class 3, transcription factor 1 (OCT-6)
hs978, hs1035, hs1128	POU3F2	POU class 3 homeobox 2 (BRN-2)
hs969, hs1526, hs1529	POU3F3	POU class 3 homeobox 3 (BRN-1)
hs192	SOX2	SRY-box containing gene 2
hs1019	SP4	trans-acting transcription factor 4
hs1166, hs1538	NKX2-1	NK2 homeobox 1 (TITF1)
hs1559	WNT7B	wingless-related MMTV integration site 7B
hs1573	ZIC1	zinc finger protein of the cerebellum 1

**Figure 1. Expression of a subset of forebrain enhancers identified by conservation or p300 binding at whole-mount resolution**

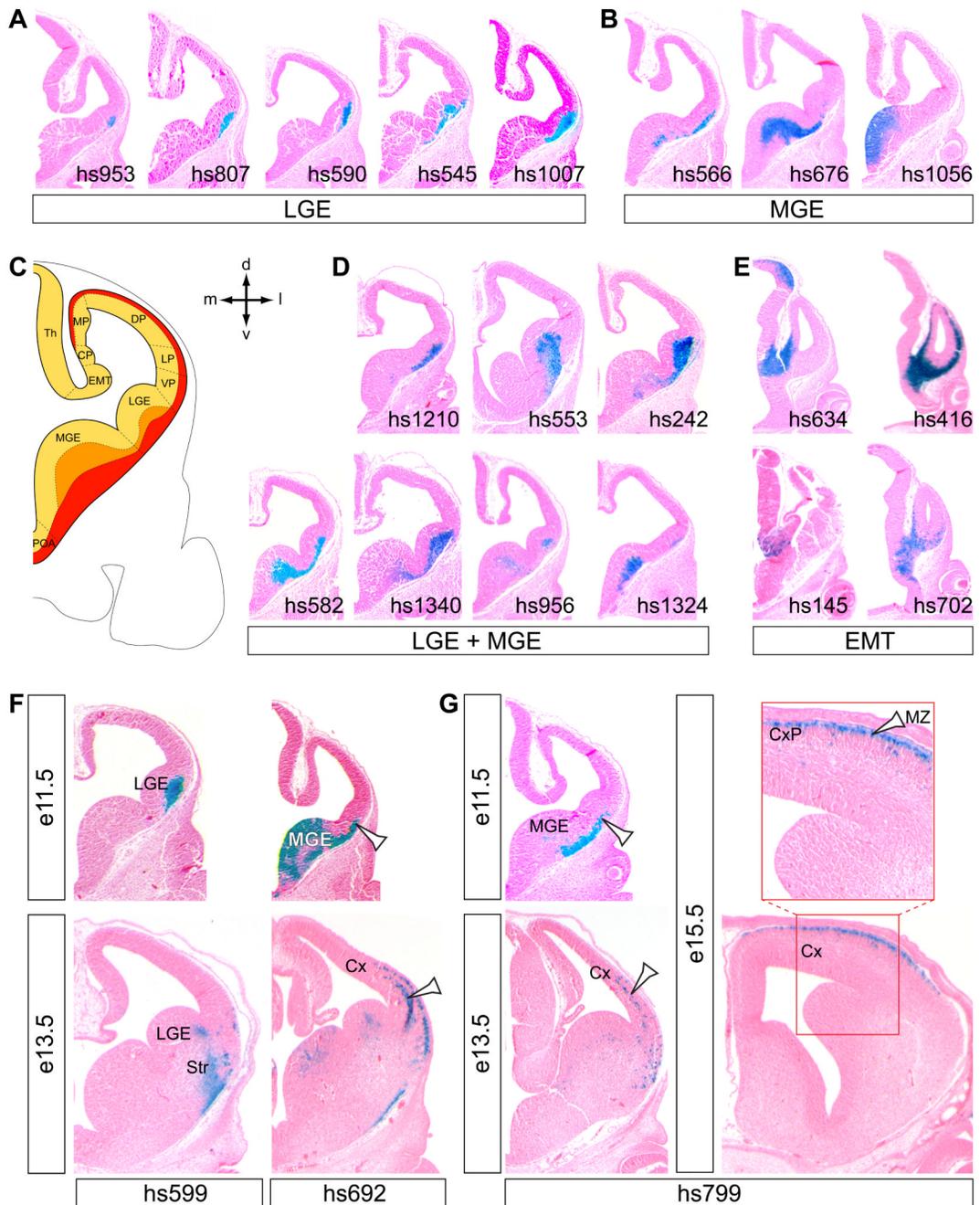
A) A selection of 50 reproducible forebrain enhancers at e11.5 identified in this study. In each case, only one of several (minimum: 3) embryos with the same pattern is shown. Additional embryos obtained with each enhancer construct can be viewed at <http://enhancer.lbl.gov>. Enhancer elements are sorted by broad similarities of patterns as evident at whole-mount resolution. B) Examples of genes implicated in forebrain development that were screened for enhancers in the present study and for which enhancers are shown in A). A full list of all 329 constructs tested in this study, including annotations of enhancer activity patterns and information about neighboring genes are provided in Suppl. Table S4.



**Figure 2. Subset of forebrain enhancers with activity in different dorsoventral subregions of the developing mouse pallium (cortex)**

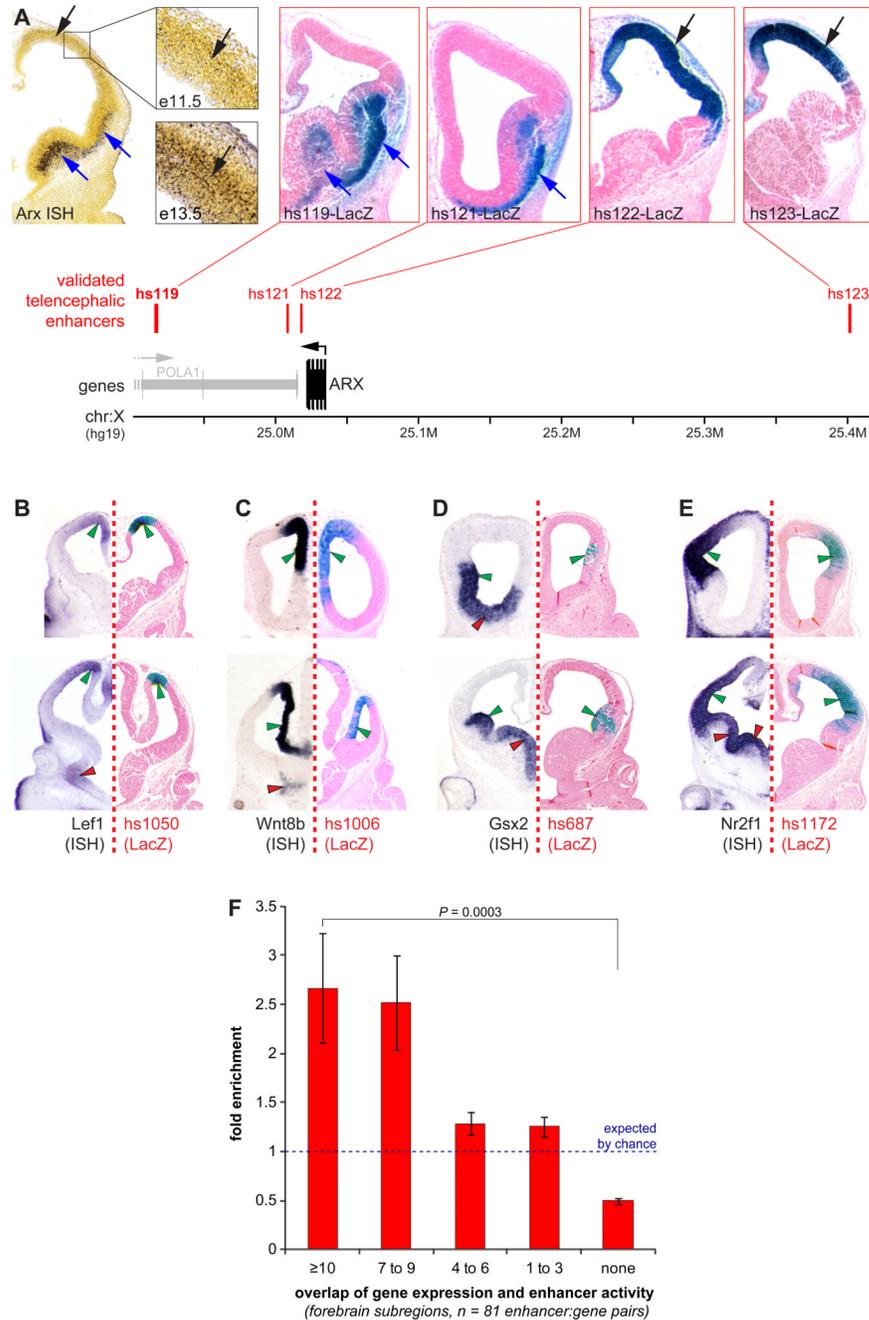
A) Overview of annotated structures in the approximate coronal sectioning plane shown in B)-R). B)-R) Selected enhancers that reproducibly label subregions of the developing pallium. Enhancers are arranged by their spatial specificities in the medial, dorsal, lateral, and ventral pallium. Detailed annotations of all patterns, as well as additional enhancers that drive expression in these subregions are provided in Suppl. Table S5. Full serial sets of sections for each enhancer can be viewed at <http://enhancer.lbl.gov>, using the enhancer IDs indicated in the figure panels. S) Comparison of enhancer activities between e11.5 and e13.5. Red arrowheads indicate activity in neuronal precursor/differentiation zones, orange

arrowheads indicate immature neurons in the cortical plate. Cx, cortex; CxP, cortical plate; DP, dorsal pallium; LGE, lateral ganglionic eminence; LP, lateral pallium; MP, medial pallium; VP, ventral pallium; Se, septum.



**Figure 3. Subset of forebrain enhancers with activity in different subregions of the mouse subpallium (basal ganglia) and eminentia thalami (telencephalic-diencephalic connection)**  
 A), B), D), E) Selected enhancers that target LacZ expression A) predominantly or exclusively to subregions of the LGE, B) predominantly the MGE, D) both the LGE and MGE, and E) the EMT. C) Schematic overview of structures in the approximate sectioning plane shown in A), B), D) and E). Depending on the rostrocaudal extent of staining, for some enhancers more rostral or caudal planes were chosen to illustrate salient features of the respective patterns. F-G) Comparison of enhancer activities between e11.5, e13.5, and e15.5. White arrowheads indicate cell populations whose location is consistent with migration from the MGE, through the LGE, to the cortex. CP, choroid plexus; Cx, cortex;

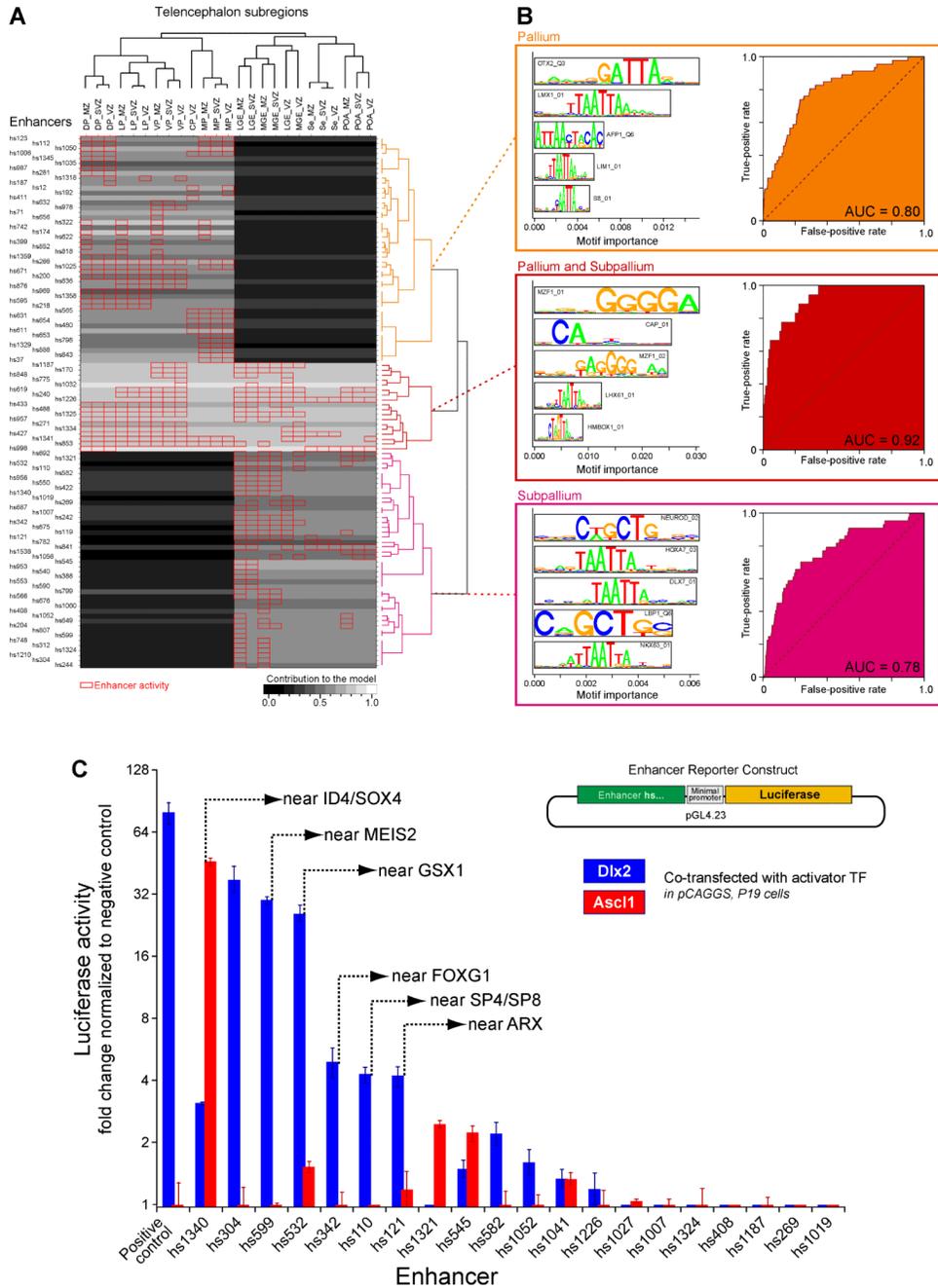
CxP, cortical plate; EMT, eminentia thalami; DP, dorsal pallium; LGE, lateral ganglionic eminence; LP, lateral pallium; MGE, medial ganglionic eminence; MP, medial pallium; MZ, marginal zone; POA, preoptic area; Str, striatum; VP, ventral pallium; Th, thalamus.



**Figure 4. Correlation of spatial enhancer activity patterns with RNA expression patterns of nearby genes**

A-E) Examples of individual enhancers recapitulating aspects of the gene expression patterns. A) The *Arx* gene is expressed in subpallial (blue arrows) and pallial (black arrows) regions. Pallial expression increases from e11.5 to e13.5 (insets). At least four enhancers in the extended locus drive subpallial (hs119, hs121) or pallial expression (hs122, hs123) at e11.5. B-E) Additional examples of overlap in enhancer activity with expression of nearby genes in rostral (top) and more caudal (bottom) areas of the telencephalon at e11.5. In all four cases, there was spatial overlap in activity (green arrowheads), as well as gene expression in additional regions that did not show enhancer activity (red arrowheads). F) To

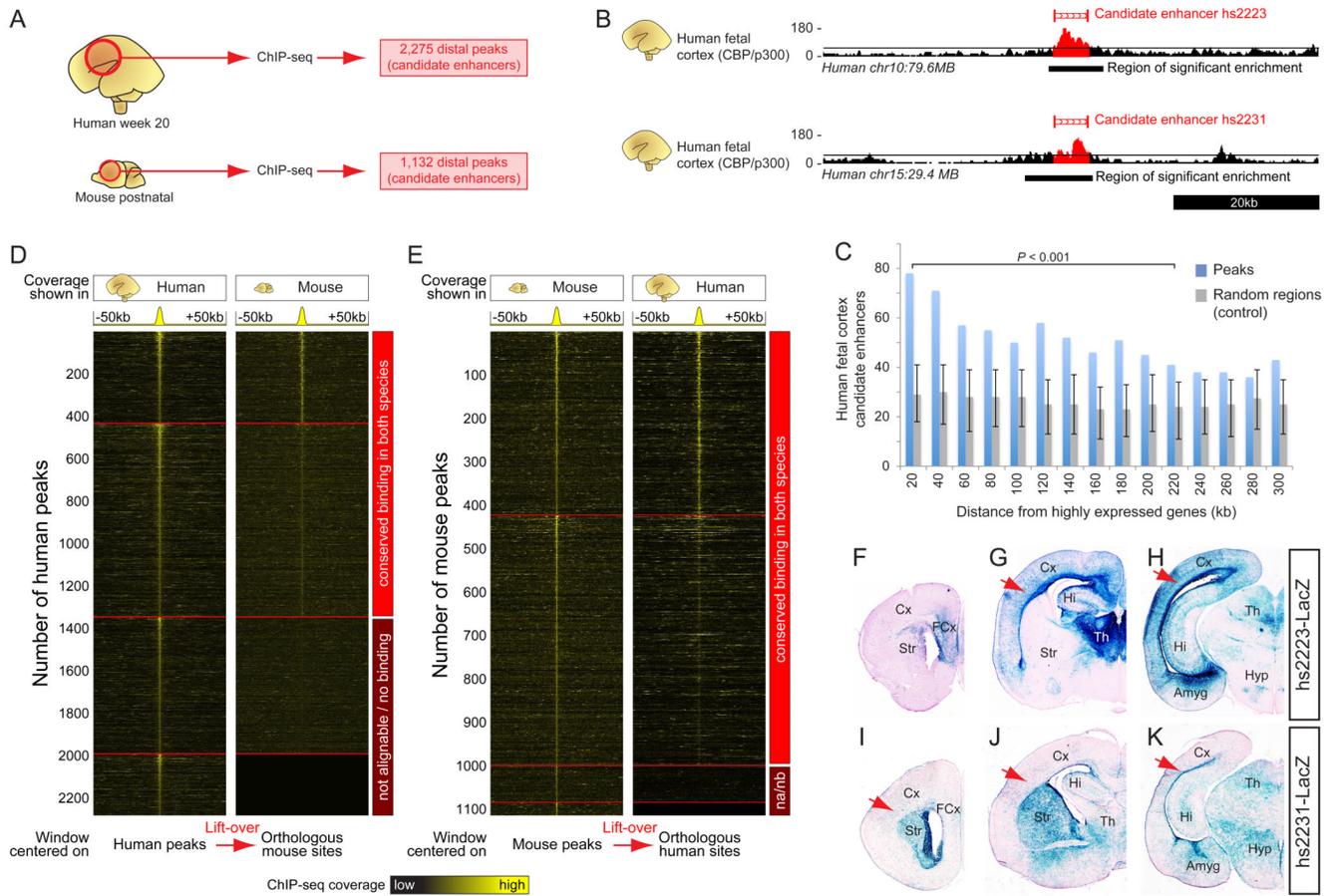
assess overall correlations, the annotated activity patterns of telencephalic enhancers were compared to RNA expression patterns of nearby genes. Compared to randomly assigned enhancer:gene pairs, there is a highly significant enrichment of cases in which concordant enhancer activity and gene expression is observed in one or multiple telencephalic subregions ( $P = 0.0003$ , Mann-Whitney test). *Arx* RNA *in situ* hybridization images in A): Allen Developing Mouse Brain Atlas (<http://developingmouse.brain-map.org>), reproduced with permission from Allen Institute for Brain Science.



**Figure 5. Relating sequence motif content to high-resolution activity annotations**

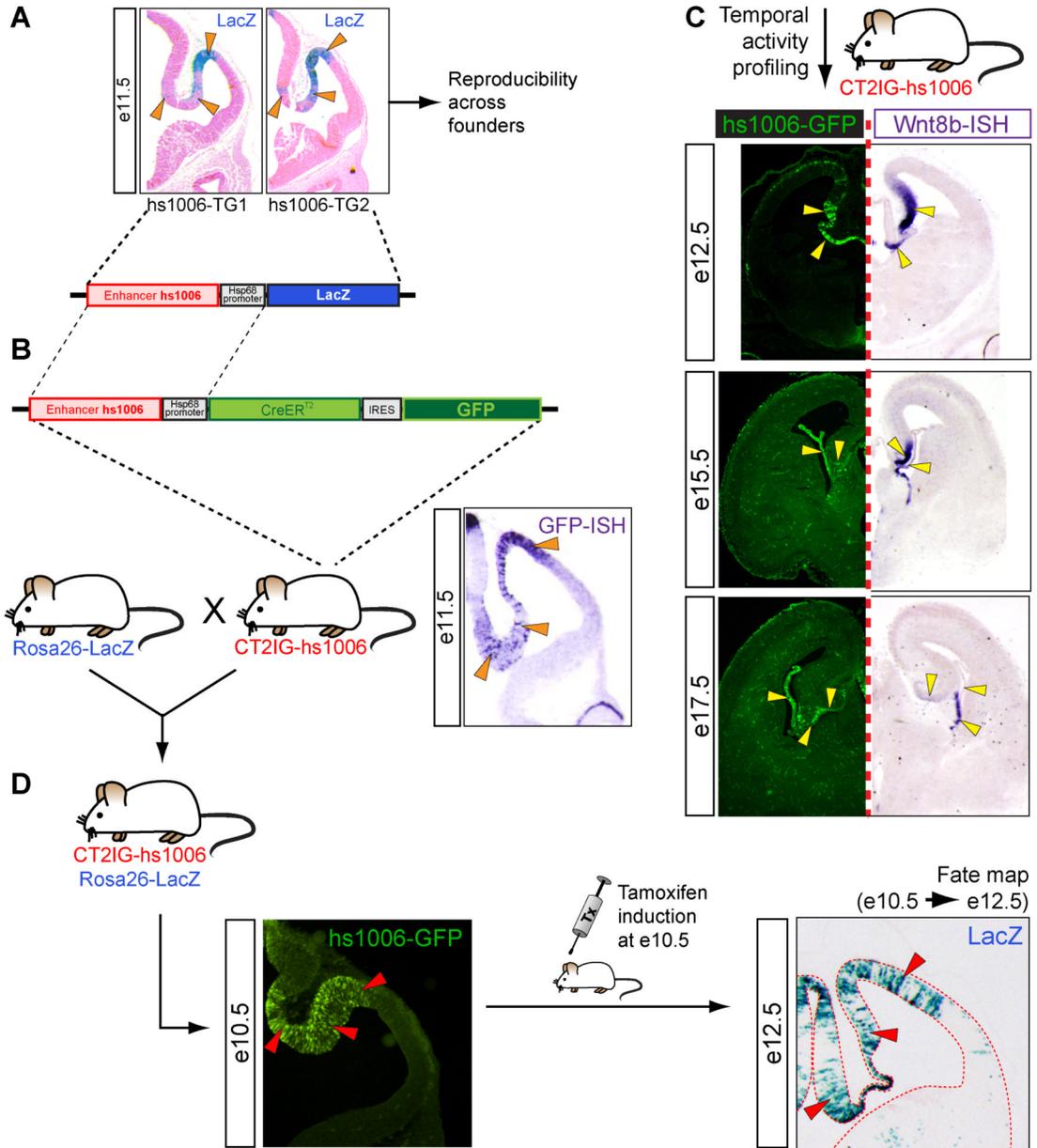
A) Red squares indicate enhancers (rows) active in different telencephalic subregions (columns). Unsupervised clustering (Jaccard's coefficient, average linkage) of telencephalic subregions by similarity of enhancer activity profiles (top dendrogram) largely follows known developmental, functional and topological relations of telencephalic subregions. Clustering (Euclidean distances, Ward's method) of enhancers by similarity of observed activity in telencephalic subregions suggests functional subgroups (right dendrogram). Shades of gray indicate the proportion of decision trees assigning each enhancer to the pallium or subpallium class (for pallium and subpallium enhancers) or to the compound pallium/subpallium class (for compound enhancers). B) The Random Forest (RF) classifier

distinguishes enhancers that are active in pallium only (top), in both pallium and subpallium (center), and in subpallium only (bottom). Left: Top 5 sequence motifs characterizing each class of enhancers and their relative contribution to the classification. Additional motifs are shown in Suppl. Fig. S2. Right: Receiver-operating characteristic (ROC) curves of predictive performances. The area under the curve (AUC) measures the ability of the classifier to limit incorrect predictions while maintaining accuracy in true predictions. For example, the “pallium and subpallium” classifier correctly identifies ~70% of enhancers in this cluster at a false positive rate of 10%. C) Luciferase co-transfection assays of 20 subpallial enhancers with either the transcription factors Dlx2 or Ascl1 in P19 cells.



**Figure 6. Genome-wide experimental comparison of enhancers active during human and mouse cortex development**

A) ChIP-seq analysis was performed on human gestational week 20 and mouse postnatal day 0 cortex tissue, using an antibody directed against the enhancer-associated p300/CBP proteins. B) Two representative peaks (candidate enhancers) identified from the human fetal dataset. C) Predicted human fetal cortex enhancers are significantly enriched in the larger vicinity (up to 220kb away) of genes highly expressed in the human fetal cortex. D) The majority of candidate enhancers identified from human fetal cortex show evidence of p300/CBP binding at orthologous sites in the mouse genome (top two sectors of heat map). However, a substantial proportion of human peaks either shows no evidence of p300/CBP binding at orthologous sites in the mouse genome (third sector), or falls into regions of the human genome that have no known orthologous sequence in the mouse (fourth sector). E) A substantially larger proportion of mouse P0 cortex candidate enhancers was found to be bound by p300/CBP at orthologous sites in the human genome. F-K) Transgenic activity analysis of the two candidate enhancers shown in B) in transgenic mice at postnatal day 1. Each pattern was reproducible in a minimum of three F<sub>0</sub> animals, three sectioning planes from one representative brain per enhancer are shown. Red arrows indicate expression in the cortex.



**Figure 7. Using telencephalon enhancers as tissue-specific reagents**

A) Approach used for generation of the large-scale high-resolution atlas at e11.5. B) Enhancers can be used as drivers of other reporter and effector genes, such as GFP or tamoxifen-inducible Cre recombinase. GFP reporter expression at e11.5 recapitulates the annotated LacZ expression pattern (orange arrowheads). Schematic components of constructs in A) and B) are shown not to scale. C) Stable transgenic lines facilitate temporal profiling of enhancer activity and comparisons with corresponding gene expression patterns. D) Tamoxifen induction at e10.5, followed by LacZ staining at a later timepoint (shown: e12.5) can be used for developmental fate mapping of neuronal cell populations.