

UCSF

UC San Francisco Previously Published Works

Title

Ultraconserved Enhancers Are Required for Normal Development

Permalink

<https://escholarship.org/uc/item/9nr7s9m3>

Journal

Cell, 172(3)

ISSN

0092-8674

Authors

Dickel, Diane E

Ypsilanti, Athena R

Pla, Ramón

et al.

Publication Date

2018

DOI

10.1016/j.cell.2017.12.017

Peer reviewed



Published in final edited form as:

Cell. 2018 January 25; 172(3): 491–499.e15. doi:10.1016/j.cell.2017.12.017.

Ultraconserved Enhancers Are Required for Normal Development

Diane E. Dickel^{1,*}, Athena R. Ypsilanti^{2,#}, Ramón Pla^{2,#}, Yiwen Zhu¹, Iros Barozzi¹, Brandon J. Mannion¹, Yupar S. Khin², Yoko Fukuda-Yuzawa¹, Ingrid Plajzer-Frick¹, Catherine S. Pickle¹, Elizabeth A. Lee¹, Anne N. Harrington¹, Quan T. Pham¹, Tyler H. Garvin¹, Momoe Kato¹, Marco Osterwalder¹, Jennifer A. Akiyama¹, Veena Afzal¹, John L.R. Rubenstein², Len A. Pennacchio^{1,3,4,*}, and Axel Visel^{1,3,5,*}

¹Functional Genomics Department, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

²Department of Psychiatry, Neuroscience Program, UCSF Weill Institute for Neurosciences, and the Nina Ireland Laboratory of Developmental Neurobiology, University of California San Francisco, San Francisco, CA 94158, USA

³U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

⁴Comparative Biochemistry Program, University of California Berkeley, Berkeley, CA 94720, USA

⁵School of Natural Sciences, University of California Merced, Merced, California, 95343, USA

Summary

Non-coding “ultraconserved” regions containing hundreds of consecutive bases of perfect sequence conservation across mammalian genomes can function as distant-acting enhancers. However, initial deletion studies in mice revealed that loss of such extraordinarily constrained sequences had no immediate impact on viability. Here we show that ultraconserved enhancers are required for normal development. Focusing on some of the longest ultraconserved sites genome-wide, located near the essential neuronal transcription factor *Arx*, we used genome editing to create an expanded series of knockout mice lacking individual or combinations of ultraconserved enhancers. Mice with single or pairwise deletions of ultraconserved enhancers were viable and fertile, but in nearly all cases showed neurological or growth abnormalities, including substantial

*To whom correspondence should be addressed: D.E.D., dedickel@lbl.gov; L.A.P., lapennacchio@lbl.gov; A.V., avisel@lbl.gov (lead contact).

#These authors contributed equally.

Author Contributions

D.E.D., A.R.Y., R.P., Y.Z., B.J.M., Y.S.K., I.P.-F., C.S.P., E.A.L., A.N.H., Q.T.P., T.H.G., M.K., M.O., J.A.A., and V.A. performed experiments, including making and/or characterizing all transgenic and knockout mouse lines. D.E.D., I.B., and Y.F.-Y. analyzed RNA-Seq data. D.E.D., A.R.Y., R.P., J.L.R.R., L.A.P., and A.V. planned the study and wrote the manuscript with input from the remaining authors.

Declaration of Interests

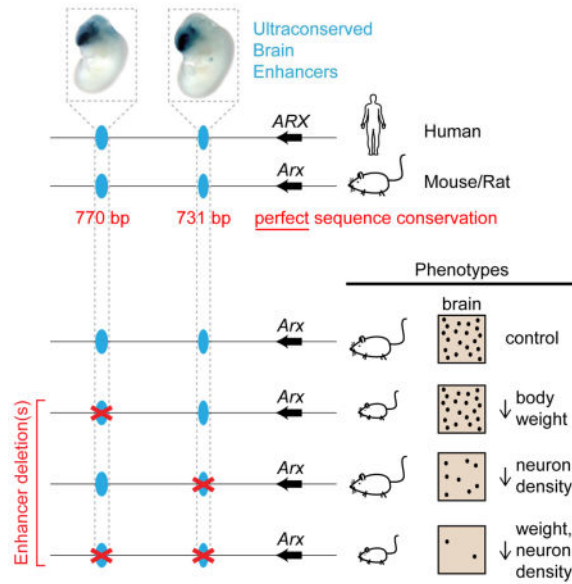
J.L.R.R. is cofounder, stockholder, and currently on the scientific board of *Neurona*, a company studying the potential therapeutic use of interneuron transplantation.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

alterations of neuron populations and structural brain defects. Our results demonstrate the functional importance of ultraconserved enhancers and indicate that remarkably strong sequence conservation likely results from fitness deficits that appear subtle in a laboratory setting.

In brief

In contrast to initial studies suggesting that loss of ultraconserved enhancers had no impact on viability, these sequences are now shown to be required for normal development.



Introduction

One of the most astonishing observations from the sequencing of the human, mouse, and rat genomes was the finding of 481 loci of perfect sequence conservation between these species (minimum length 200 contiguous base pairs), despite the approximately 80 million years since their last common ancestor (Bejerano et al., 2004). Dubbed “ultraconserved” sequences, more than half of these sites display no evidence for protein-coding function. These noncoding elements are enriched near transcription factor genes expressed during embryogenesis (Bejerano et al., 2004; Sandelin et al., 2004), suggesting a possible role in regulating the expression of essential developmental genes. Indeed, we previously tested more than 200 noncoding ultraconserved sites using *in vivo* transgenic mouse reporter assays, which revealed that at least half act as transcriptional enhancers during embryonic development (Pennacchio et al., 2006; Visel et al., 2008).

Given their likely role in regulating the expression of vital genes, along with their extreme sequence conservation, noncoding ultraconserved elements were broadly assumed to be essential for life. However, the targeted deletion of four ultraconserved enhancers resulted in mice that were viable, fertile, and did not have any obviously detrimental phenotypes (Ahituv et al., 2007). This perplexing observation resulted in much speculation as to why these sites retained extreme sequence conservation in the apparent absence of deleterious

effects, including hypotheses such as undetected subtle phenotypes selected against over longer time periods, functional redundancy of enhancers, or even atypical molecular functions (Ahituv et al., 2007; Chiang et al., 2008; McLean and Bejerano, 2008).

In the present study, we generated a series of mouse lines with single or pairwise deletions of ultraconserved enhancers to explore these possible explanations. We focused on a region on the X chromosome that contains the longest ultraconserved sites genome-wide (Bejerano et al., 2004). The region also encompasses the *Aristaless-related homeobox (ARX/Arx)* gene, mutations in which cause a variety of severe neurological phenotypes (Gecz et al., 2006; Kitamura et al., 2002). At least four of the ultraconserved sites within this region show enhancer activity in the developing forebrain that is similar to *Arx* gene expression (Pennacchio et al., 2006; Visel et al., 2013). We observed that nearly all single and pairwise enhancer deletions resulted in non-lethal, but identifiable abnormalities, including reduced growth and impaired brain development. Our results demonstrate that loss of ultraconserved enhancers, while broadly compatible with life, causes phenotypes that appear subtle in the laboratory but likely lead to decreased fitness over evolutionary time.

Results

To examine the biological function of ultraconserved sequences beyond initial deletion studies (Ahituv et al., 2007), we focused on an 800 kb region of the X chromosome in which eight non-exonic ultraconserved sites are found, along with the *Arx* gene (Bejerano et al., 2004; Figure 1). Each of the eight sites contains up to three adjacent ultraconserved elements of 200 bp of perfect conservation, and the locus harbors four of the five longest non-exonic ultraconserved sites in the human genome (Figure 1). We previously showed that five of the eight sites have transcriptional enhancer activity in the embryonic day 12.5 (E12.5) forebrain in transgenic mouse reporter assays, including four with highly reproducible activity in the telencephalon (Figure 1) and one with weakly reproducible activity in the diencephalon (Pennacchio et al., 2006).

Of the four telencephalon enhancers, two drive gene expression in similar regions of the ventral telencephalon, while the other two are active in similar dorsal domains, suggesting that pairs of enhancers at this locus may be active in the same cell types and potentially have redundant functions (Visel et al., 2013; Figure 1). Collectively, these dorsal and ventral telencephalon enhancer activity patterns are strikingly similar to the expression pattern of *Arx* (Visel et al., 2013; Figure 1), suggesting a regulatory interaction between these enhancers and this neighboring gene in the developing forebrain. Because of the high density of unusually long and possibly functionally redundant ultraconserved enhancers, this locus provides a testbed for assessing the role of ultraconserved enhancers in development. In particular, it is an ideal locus to assess possible explanations, such as enhancer redundancy or sub-lethal deleterious phenotypes, that were proposed to explain the initial observations that ultraconserved enhancers are not required for viability.

Ultraconserved Forebrain Enhancers Are Active in *Arx*-expressing Cells

First, we sought to verify that these enhancers are active in *Arx*-expressing cells and to determine their cell type specificity, since this information can be used to guide downstream

phenotyping of enhancer deletion mice. We created a series of transgenic reporter mice, each with one of the ultraconserved telencephalon enhancers driving expression of a fluorescent reporter gene in the embryonic forebrain (see **Methods**). We then performed single-cell RNA-Seq using Drop-Seq (Macosko et al., 2015), separately for each enhancer, on E12.5 whole forebrains from these embryos (Figure 2A, Table S1). Across all experiments, we reconstructed gene expression profiles for nearly 5,000 single cells (Figures 2B and S1A, Table S1). We masked the expression of the reporter gene and clustered the cells by gene expression similarity, which identified nine distinct cell types present, including two cell types that had robust expression of *Arx*. The first is a population of radial glia and neural progenitor cells, and the second is a cluster containing immature neurons, including both GABAergic and glutamatergic types (Figure 2B, see Figure S1A–C and Table S2 for relevant marker genes). We next examined which cells displayed expression of the reporter gene, which revealed that all four ultraconserved telencephalon enhancers drove reporter gene expression in at least one of these two *Arx*-expressing cell types (Figures 2C and S1D). Both ventral forebrain enhancers (hs119 and hs121) are active predominantly in the immature neuron cluster, particularly in neurons expressing GABAergic marker genes (Figure S1B,D), consistent with the known activity of hs121 in GABAergic neurons later in embryogenesis (Colasante et al., 2008). The dorsal enhancer hs123 is active predominantly in the radial glia/neural progenitor cells, with additional activity in immature neurons that have a glutamatergic identity (Figure S1C–D). Although few reporter-expressing cells were recovered for the dorsal enhancer hs122, the majority are radial glia/neural progenitor cells (Figure S1D). Together, these results support that the four ultraconserved enhancers collectively activate the expression of *Arx* in at least two different *Arx*-expressing cell populations.

Single and Double Knockout of Ultraconserved Enhancers Results in Viable Mice

To examine the *in vivo* function of these ultraconserved enhancers, we used CRISPR/Cas9 genome editing in fertilized eggs to generate a series of mouse lines, each missing one of the four enhancers (Figures 3A and S2, Table S3). This included one ultraconserved enhancer (hs121, also named UC467) whose deletion was previously reported to produce viable animals (Ahituv et al., 2007). Additionally, to examine the possibility of functional redundancy between ultraconserved enhancers limiting deleterious effects, we generated deletions of pairs of enhancers with similar activity patterns (both dorsal or both ventral) on the same haplotype (Figures 3A and S2, Table S3). In all cases, viable founder mice for each targeted deletion were obtained (Table S3). Taking advantage of the location of *Arx* on the X chromosome, we out-crossed correctly targeted founder mice to wild-type animals to obtain hemizygous-null males for phenotyping. Male *Arx* gene knockout (hemizygous-null) mice die within two days after birth and have severe developmental defects, including smaller brains and testicular and pancreatic anomalies (Collombat et al., 2003; Kitamura et al., 2002). In contrast, males lacking any of the single or pairs of ultraconserved *Arx* forebrain enhancers are viable, and none displayed increased lethality *in utero* or perinatally (Figure 3B and Table S4). Furthermore, hemizygous-null males from all enhancer deletion lines were capable of siring offspring, ruling out severe infertility or testicular defects (Figure 3B and Table S4). Similarly, heterozygous-null females were viable and fertile (Figure 3B and Table S4), and viable homozygous-null females were obtained in all cases tested (Figure

3B). All told, individual loss of any of four ultraconserved forebrain enhancers near *Arx* resulted in mice that are both viable and fertile.

Remarkably, mice missing pairs of potentially redundant ultraconserved enhancers are similarly viable and fertile and have no obvious abnormalities (Figure 3B). These results demonstrate that even deletions of the longest ultraconserved sites in the genome do not have broadly catastrophic consequences to an individual organism's viability or fertility, further refuting the assumption that extreme evolutionary sequence constraint translates to indispensable function. These results are additionally inconsistent with the hypothesis that functional redundancy between ultraconserved sites broadly explains the absence of obviously deleterious phenotypes.

Pairwise Loss of Ultraconserved Enhancers Results in Decreased *Arx* RNA

We next examined whether the expression of *Arx* in the developing forebrain is altered upon loss of individual or pairs of ultraconserved enhancers. RNA-Seq on whole forebrain from male E11.5 embryos hemizygous-null for single or pairs of enhancers showed that both double enhancer deletions result in reduced *Arx* expression ($P = 0.002$ by edgeR, see **Methods**, Figure 4A–B, Table S5). No other gene within 10 Mb of the enhancers had significantly reduced expression, indicating that these ultraconserved enhancers act specifically on *Arx*. In contrast to the double knockouts, none of the single enhancer deletions profiled led to substantial down-regulation of *Arx*, as measured by RNA-Seq (Figure S3).

To further explore how the enhancer deletions affect *Arx* expression in specific subregions of the developing forebrain, and to identify changes that may not be captured by whole tissue mRNA sequencing, we performed *Arx* RNA *in situ* hybridization on embryonic forebrain tissue sections (Figures 4C–D and S4). In all cases where loss of *Arx* expression was observed by RNA-Seq, this deficit was in the region of the brain where the enhancers are active. For example, the combined loss of both enhancers active in the ventral telencephalon (hs119/hs121) results in reduced *Arx* expression in the ventral telencephalon. *In situ* hybridization also revealed a reduction of *Arx* expression in the dorsal forebrain of hs122-null embryos that was not detected by RNA-Seq (Figures 4D and S4). Together, these results confirm that the four ultraconserved enhancers regulate the brain region-specific expression of *Arx*.

Ultraconserved Enhancer Deletions Cause Growth and Brain Defects

To further explore the phenotypic consequences of the loss of ultraconserved enhancers near *Arx*, we performed a detailed characterization of growth and brain morphology for mice missing single or pairwise enhancer combinations. First, we used postnatal weight gain as a broad proxy for general organismal fitness (Crawley, 2007). While the majority of enhancer deletions resulted in no significant changes in growth, the loss of the hs119 ventral forebrain enhancer resulted in a 5–10% reduction in body mass (postnatal weeks 6–16, Figures 5A and S5, $P < 0.05$; two-tailed t-test). This phenotype was exacerbated by combined loss of both ventral forebrain enhancers (hs119/hs121), which resulted in an ~15% reduction in

body weight throughout the entire developmental window examined (Figure 5A, $P < 0.05$ at 17 of 18 postnatal time points measured; two-tailed t -test).

Guided by the severe forebrain phenotypes observed in mice hemizygous-null for the *Arx* gene (Kitamura et al., 2002), we next screened all lines for neuroanatomical changes, beginning with the ventral enhancer deletions (hs119, hs121, hs119/hs121). Deletion of the *Arx* gene results in defects in the ventral telencephalon, in particular a substantial decrease in striatal cholinergic neurons (Colombo et al., 2007; Marsh et al., 2016). While postnatal (P30–P63) mice lacking the ventral forebrain enhancers showed no obvious change in the overall size or structure of the striatum (Figure 5B), immunohistochemical analysis of choline acetyltransferase (ChAT)-expressing neurons revealed a 62% reduction in the density of cholinergic neurons in the striatum for males missing hs121 (Figure 5B–C and Table S6, $P < 0.001$; ANOVA). Males hemizygous null for hs119 and hs121 combined had a near-complete (97%) loss of cholinergic neurons in the striatum (Figure 5B–C and Table S6, $P < 0.001$; ANOVA). We also examined the postnatal densities of several neocortical GABAergic interneuron populations that originate in the ventral telencephalon during embryogenesis, some of which are known to be altered in *Arx* gene knockout mice (Colombo et al., 2007; Kitamura et al., 2002). Loss of hs121 alone, or in combination with hs119 (hs119/hs121 KO), resulted in a 21–24% decrease in the density of parvalbumin (PV)-expressing neocortical interneurons and a moderate (37–52%) increase in vasointestinal peptide (VIP) interneurons (Figure S6, Table S6, $P < 0.05$; ANOVA). In summary, the individual loss of one ultraconserved ventral telencephalon enhancer (hs121) alone causes substantial alterations in neuron populations in the postnatal mouse brain, and this phenotype further increases in severity when both ventral forebrain enhancers of *Arx* are lost. These results demonstrate a role for both ventral forebrain *Arx* enhancers in the development of specific classes of neurons.

Finally, we assessed neuronal phenotypes resulting from the loss of ultraconserved *Arx* enhancers that are active in the dorsal telencephalon. Brain sections from postnatal hemizygous null males and matched littermate controls (aged 13 to 28 weeks) revealed obvious abnormalities to the hippocampus caused by the loss of one of the dorsal enhancers (hs122), consistent with the dysgenesis of the hippocampus observed in *Arx* gene knockouts (Colasante et al., 2015; Kitamura et al., 2002). Specifically, loss of hs122 led to a 26% reduction in the length of the dentate gyrus and a 36% reduction in the area of its granule cell layer ($P < 0.05$; two-tailed t -test), in the absence of substantial changes in the length of other hippocampal structures examined (Figures 6A–B and S7, Table S6). Additionally, the dentate gyrus had a severely disorganized appearance in hs122-null mice (Figures 6A and S7A–B). Loss of hs123 alone did not result in any obvious deficits in the hippocampus, and combined loss of hs122 and hs123 resulted in a phenotype similar to hs122 alone (Figures 6A–B and S7, Table S6). Thus, in combination with the ventrally active enhancers, our results reveal a role for three out of four ultraconserved *Arx* enhancers in the development of the brain.

Discussion

Ultraconserved sites have fascinated biologists since their identification, and their perfect extended sequence conservation led to the widespread assumption that these sequences are generally required for organismal viability. However, this is clearly contradicted by deletion studies of individual non-exonic ultraconserved sites, which have uniformly resulted in viable and fertile mice (this study; Ahituv et al., 2007; Nolte et al., 2014). In this study, we sought to explain the puzzling observation that loss of sequences with extreme conservation results in no obvious detrimental effects *in vivo*. By focusing on the numerous and unusually long ultraconserved elements near the essential *Arx* gene, we were able to assess whether functional redundancy between ultraconserved enhancers or the presence of nonlethal, but nonetheless detrimental, phenotypes is most likely to explain this surprising observation. We deleted pairs of potentially redundant enhancers with similar activity profiles, removing up to 2,086 bp (hs119+hs121) of perfectly human-rodent conserved sequence. Surprisingly, even this combinatorial removal of extremely long ultraconserved enhancers yielded viable and fertile mice, undermining the hypothesis that functional redundancy is the primary explanation for the lack of obvious viability deficits. These results further underscore that perfect conservation is not analogous to essentiality.

In the absence of the conclusive identification of phenotypes explained by the loss of tissue-specific enhancers for previous ultraconserved site knockouts, a variety of alternative hypotheses to explain the extreme conservation at these sites were proposed, some of which were not based on the conventional model of negative selection (Chiang et al., 2008; Harmston et al., 2013). In this study, we resolve this debate by identifying important biological functions for the majority of individual ultraconserved sites examined. Namely, individual loss of three of the four ultraconserved enhancers examined resulted in a deficit of growth or neurodevelopment (Figure 7). We posit that negative selection against these detrimental phenotypes is likely at least partly responsible for the conservation observed. For this study, we focused on ultraconserved enhancers that are located on the X chromosome and exclusively examined hemizygous male mice. This raises the question whether the heterozygous loss of ultraconserved enhancers, either for X chromosome loci in females or for autosomal loci generally, will also broadly result in identifiable phenotypes. Based on previous mouse knockouts of less well conserved autosomal enhancers, we speculate that heterozygous loss of some ultraconserved enhancers will be sufficient to cause similar, albeit less severe, developmental phenotypes and deficits in gene expression (Dickel et al., 2016; Johnson et al., 2012). Collectively, these results support that loss of ultraconserved elements leads to fitness defects that are generally subtle in a laboratory setting but likely selected against over evolutionary time.

STAR Methods

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
rabbit anti-parvalbumin	Swant Swiss Abs	Cat# PV 28
rat anti-somatostatin (clone YC7)	Millipore	Cat# MAB354
goat anti-somatostatin (D-20)	Santa Cruz	Cat# sc-7819
rabbit anti-Vasointestinal peptide	Immunostar	Cat# 20077; RRID: AB_572270
goat anti-Choline Acetyltransferase (ChAT)	Millipore	Cat# AB144P
mouse Anti-Calretinin, (clone 6B8.2)	Millipore	Cat# MAB1568
Anti-Digoxigenin-AP, Fab fragments	Roche	Cat# 11093274910
Chemicals, Peptides, and Recombinant Proteins		
DIG RNA labeling mix	Roche	Cat# 11277073910
T3 polymerase	Roche	Cat# RPOLT3-RO
DNaseI	Roche	Cat# 4716728001
Sheep serum	Sigma	Cat# S2263
Blocking reagent	Roche	Cat# 11096176001
BM Purple	Roche	Cat# 11442074001
Accumax Cell Dissociation Solution	Innovative Cell Technologies	Cat# AM105
Maxima H Minus Reverse Transcriptase	Thermo Fisher Scientific	Cat# EP0753
Perfluorooctanol	Sigma	Cat# 370533-25G
Exonuclease I	New England Biolabs	Cat# M0293L
Critical Commercial Assays		
RNA Clean & Concentrator	Zymo Research	Cat# R1015
mMESSAGE mMACHINE T7 kit	Thermo Fisher Scientific	Cat# AM1344
MEGAsortscript T7 kit	Thermo Fisher Scientific	Cat# AM1354
MEGAclear kit	Thermo Fisher Scientific	Cat# AM1908
Barcoded dT Beads for Drop-Seq	(Macosko et al., 2015) ChemGenes	Lot# 011416B
TruSeq Stranded mRNA Sample Prep Kit	Illumina	Cat# RS-122-2101
RNAqueous Total RNA Isolation Kit	Thermo Fisher Scientific	Cat# AM1912
Drop-seq Microfluidic Device	(Macosko et al., 2015) Nanoshift	N/A
KAPA HiFi HotStart ReadyMix	Kapa Biosystems	Cat# KK2602
Nextera XT DNA Library Preparation Kit	Illumina	Cat# FC-131-1096
TURBO DNA-free Kit	Thermo Fisher Scientific	Cat# AM1907
Deposited Data		
RNA-seq data, including Drop-seq for enhancer transgenics, and bulk tissue RNA-seq for enhancer knockouts	This study	GEO: GSE100394
Experimental Models: Organisms/Strains		
Mouse: FVB	Charles River	http://www.criver.com/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: FVB/ <i>Arx</i> - hs119	This paper	N/A
Mouse: FVB/ <i>Arx</i> - hs121	This paper	N/A
Mouse: FVB/ <i>Arx</i> - hs122	This paper	N/A
Mouse: FVB/ <i>Arx</i> - hs123	This paper	N/A
Mouse: FVB/ <i>Arx</i> - hs119 hs121	This paper	N/A
Mouse: FVB/ <i>Arx</i> - hs122 hs123	This paper	N/A
Oligonucleotides		
A full list of primers and oligos is provided in Table S7	IDT	N/A
Recombinant DNA		
Hsp68-mCherry vector	This paper	N/A; available from the authors
Hsp68-GFP vector	This paper	N/A; available from the authors
hs119-hsp68-GFP reporter vector	This paper	N/A; available from the authors
hs121-hsp68-mCherry reporter vector	This paper	N/A; available from the authors
hs122-hsp68-GFP reporter vector	This paper	N/A; available from the authors
hs123-hsp68-mCherry reporter vector	This paper	N/A; available from the authors
pDD921 plasmid encoding human optimized Cas9	(Kvon et al., 2016)	N/A; available from the authors
gRNA cloning vector	(Mali et al., 2013)	Addgene Plasmid #41824
Software and Algorithms		
Drop-seq published pipeline v. 1.11	(Macosko et al., 2015)	http://www.mccarrolllab.com/dropseq
STAR v. 2.4.1d	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
SIMLR	(Wang et al., 2017)	http://bioconductor.org/packages/release/bioc/html/SIMLR.html
DBSCAN	(Ester et al., 1996)	https://cran.r-project.org/web/packages/dbscan/
MAST	(Finak et al., 2015)	https://github.com/RGLab/MAST
ROCR	(Sing et al., 2005)	https://cran.r-project.org/web/packages/ROCR/
cutadapt_v1.1	(Martin, 2011)	https://github.com/marcelm/cutadapt
TopHat v2.0.6	(Trapnell et al., 2009)	https://github.com/infphilo/tophat
HTSeq	(Anders et al., 2015)	https://github.com/simon-anders/htseq
edgeR	(Robinson et al., 2010)	http://bioconductor.org/packages/release/bioc/html/edgeR.html
Cufflinks v2.2.1	(Trapnell et al., 2010)	https://github.com/cole-trapnell-lab/cufflinks
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
SPSS Statistics v15	SPSS Inc.	https://www.ibm.com/products/spss-statistics
Other		
Guide RNA cloning protocol	(Mali et al., 2013)	http://www.addgene.org/static/cms/files/hCRISPR_gRNA_Synthesis.pdf
List of human-mouse-rat ultraconserved sequences	(Bejerano et al., 2004)	https://users.soe.ucsc.edu/~jill/ultra.html
Drop-Seq experimental protocol	(Macosko et al., 2015)	http://www.mccarrolllab.com/dropseq/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by, the Lead Contact Axel Visel (avisel@lbl.gov).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee. All mice used in this study were housed at the Animal Care Facility (ACF) of LBNL. Mice were monitored daily for food and water intake, and animals were inspected weekly by the Chair of the Animal Welfare and Research Committee and the head of the animal facility in consultation with the veterinary staff. The LBNL ACF is accredited by the American Association for the Accreditation of Laboratory Animal Care (AAALAC). All transgenic and knockout mouse models generated as part of this work were made in the *Mus musculus* FVB strain.

Transgenic (enhancer-reporter) embryos used for Drop-seq were developmental stage E12.5. These embryos were not assessed for gender, which is not outwardly obvious at E12.5. Therefore, we expect that a roughly equal number of male and female embryos were assessed. Additional information is provided in the **METHOD DETAIL** section below (see *Generating Transgenic Embryos* and *Single-Cell RNA-Seq*).

Knockout (enhancer deletion) animals phenotyped were the following ages: RNA-seq (E11.5), *in situ* hybridization (E12.5 for dorsal knockouts, E13.5 for ventral knockouts), body weight phenotyping (3–20 weeks postnatal), neurological phenotyping (postnatal ages from P30 to adult, exact ages of animals are given in Table S6). Phenotyping of enhancer deletion mice was restricted to males because the enhancers deleted reside on the X chromosome, which is subject to random inactivation in females. Additional detailed information regarding specific phenotyping analyses, including selection criteria, is provided in the **METHOD DETAIL** section below (beginning at *Generation of Enhancer Knockout Mice*).

METHOD DETAILS

Ultraconserved Sequence Annotations—The complete list of human-mouse-rat ultraconserved sequences (Bejerano et al., 2004) was downloaded from <https://users.soe.ucsc.edu/~jill/ultra.html>. Genomic coordinates were converted to the hg19 human genome assembly using liftOver (Rosenbloom et al., 2015). Current functional annotations (exonic vs. non-exonic) of all ultraconserved elements were obtained by using the UCSC Genome Browser's Table Browser function (Rosenbloom et al., 2015) to intersect all ultraconserved sites with hg19 gene annotations, including RefSeq (Pruitt et al., 2005), UCSC Genes (Hsu et al., 2006), and GENCODE v24lift37 (Harrow et al., 2012). Intersections were performed on June 26, 2017. Ultraconserved elements were annotated as exonic if they displayed any overlap with an exon of an annotated gene from any of the gene sets considered. Otherwise, ultraconserved elements were annotated as non-exonic.

Generating Transgenic Embryos—Enhancer names (hs numbers) used are the unique identifiers from the VISTA Enhancer Browser (www.enhancer.lbl.gov). All four enhancers

were previously tested in transgenic mouse assays using a minimal promoter and LacZ reporter gene, and the results for whole mount embryos (Pennacchio et al., 2006) and sectioned forebrain tissue (Visel et al., 2013) were reported previously. For Drop-Seq, transgenic enhancer-reporter embryos were generated using the same techniques (Kothary et al., 1989; Pennacchio et al., 2006), with the exception of a fluorescent transgene (mCherry or GFP) in place of LacZ. Briefly, enhancers were PCR amplified (primers in Table S7) and cloned into a fluorescent reporter vector via the Gateway cloning method (Thermo Fisher Scientific). Transgenic mice were generated using *Mus musculus* FVB strain mice, and sample sizes were determined empirically based on our experience performing >2,000 transgenic enhancer assays and on the tissue volume requirements of the single-cell RNA-Seq method. Mouse embryos were excluded from further analysis if they did not express the fluorescent reporter gene or were not at the correct developmental stage. As no direct comparisons were made between embryos with and without the enhancer-reporter transgene, randomization and experimenter blinding were unnecessary and not performed.

Single-Cell RNA-Seq—F0 (hs119, hs121, hs122) or F1 (hs123) generation embryos aged E12.5 were screened for fluorescent reporter transgene expression using an MZ16 Stereomicroscope (Leica Microsystems) outfit with a SOLA SM 365 Light Source (Lumencor). Whole forebrain tissue was dissected from all embryos displaying fluorescent reporter activity in a pattern identical to the previously described enhancer activity pattern. For each enhancer, forebrain tissue from at least two independent founders or founder lines was pooled together. Room temperature Accumax Cell Dissociation Solution (Innovative Cell Technologies) was added to the tissue, and this was briefly pipetted to generate single cell suspensions. Cells were then immediately passed through a 40 micron strainer to remove non-dissociated tissue. Samples were kept on ice throughout the collection and preparation process to prevent RNA degradation. For each experiment, human HEK293T/17 cells (ATCC) were spiked in to a final concentration of 2.5% to serve as an internal quality control for doublet frequency. The final cell concentration was then adjusted to 50 cells/ μ l. Single-cell RNA sequencing was performed using the Drop-Seq method (Macosko et al., 2015) according to protocol version 3.1 from <http://www.mccarrolllab.com/dropseq/>. Briefly, cells and ChemGenes Beads (Lot 011416B) were captured in aqueous droplets containing lysis buffer using a microfluidic device (Nanoshift). Droplets were recovered and broken with the addition of 6X SSC and perfluorooctanol (Sigma) followed by mixing and centrifugation. Beads released from the droplets were recovered, washed, and suspended in reverse transcriptase buffer. Captured mRNAs were reverse transcribed using Maxima H-Reverse Transcriptase (Thermo Fisher Scientific). Beads were then washed, and excess bead primers were removed with Exonuclease I (NEB). Beads were again washed and counted. Template switching PCR was performed using Kapa HiFi Hotstart Readymix (Kapa Biosystems). cDNAs were captured and purified using AMPure XP beads (Beckman Coulter), and library quality was assessed using a BioAnalyzer High Sensitivity Assay (Agilent). Sequencing libraries were generated using the Nextera XT kit (Illumina), purified with AMPure beads, and sequenced paired-end on an Illumina HiSeq2500.

Raw data analysis and digital expression quantification was carried out using the Drop-Seq published pipeline version 1.11 (Macosko et al., 2015). *DetectBeadSynthesisErrors* and

DigitalExpression were run setting *NUM_CORE_BARCODES* to 10,000. The alignment was performed using STAR version 2.4.1d (Dobin et al., 2013). Custom genomes and transcriptomes were generated using iGenome annotations for mm10 and hg19, along with the sequences of the transgenes, mCherry and GFP (SUN1-sfGFP). All the downstream analyses were performed using the statistical computing environment R v.3.3.1 (www.r-project.org). First, for each one of the five libraries, the expression counts for each transcript was retrieved for the top 10,000 STAMPs. By means of the human cells spiked-in, the purity was determined using 90% as cutoff (namely, if 90% or more of the transcripts assigned to a STAMP were from human or mouse, that STAMP was assigned to either a human or a mouse cell, otherwise it was considered as a doublet). Next, starting from the 50 STAMPs showing the highest number of detected transcripts, the doublet-rate was estimated (Macosko et al., 2015). The 51st STAMP was then added and the doublet-rate re-calculated, and so forth up to the 10,000th STAMP. A threshold of 90% doublet-rate was finally applied to the resulting curve, which in turn allowed the identification of a set of high-quality STAMPs. A further threshold on the minimum number of detected transcripts was applied, so that STAMPs showing less than 1,000 detected molecules were discarded. Results from the five libraries were then merged together. Digital expression for each STAMP was normalized to Transcripts Per Million (TPM). The expression of the two transgenes (mCherry, GFP) was excluded from any of the following clustering steps. Genes with detectable expression in less than or equal to 10 STAMPs were also discarded. Dimensionality reduction was performed using the *single-cell interpretation via multikernel learning* approach (SIMLR; Wang et al., 2017). The STAMPs were then clustered in the resulting two-dimensional space using density-based clustering (DBSCAN; Ester et al., 1996). The *dbscan* R package was employed, setting *eps* to 3 and *minPts* to 25. MAST (Finak et al., 2015) was then run to detect differentially expressed genes. The Two-sample Likelihood Ratio Test implemented in the *LRT* function of the *MASTR* package allowed the identification of marker genes for each cluster. Briefly, given a cluster, each STAMP was either flagged as belonging or not belonging to it. Those genes identified as up-regulated in the cluster at a FDR ≤ 0.05 (Benjamini-Hochberg correction; Benjamini and Hochberg, 1995) were classified as markers for the cluster. More stringent sets of marker genes were determined by filtering these lists based on the Area Under the Curve (AUC), which is an estimate on how well a certain gene predicts a cell as belonging to a certain cluster. AUCs were calculated using the *ROCR* R package (Sing et al., 2005). Given a cluster, a list of stringent markers was defined as those genes identified by MAST (see above) also showing an AUC ≥ 0.6 . Enriched pathways and gene ontologies were identified using mouseMine (Motenko et al., 2015). Cell type cluster identities were determined by examining the enriched gene ontology lists and by comparing marker gene sets for each cell type to marker genes used previously for forebrain (such as Chen et al., 2017). The identification of the cell types in which each enhancer is active was robust to the choice of clustering algorithm, as similar analyses using the *k*-means based SC3 method (Kiselev et al., 2017), resulted in identical conclusions.

As a bulk tissue control, whole forebrain from E12.5 littermates not expressing the reporter gene was dissected, pooled, and RNA was extracted using the RNAqueous Total RNA Isolation Kit (Thermo Fisher Scientific). Resulting total RNA was incubated with

ChemGenes Beads to specifically capture mRNA. Reverse transcription, library generation, and sequencing for the bulk control were carried out as described for the Drop-Seq libraries (see above). Pseudobulk profiles for each Drop-Seq library were generated by calculating the sum of the single-cell counts for each gene. The resulting profiles were normalized as TPM (transcripts per million sequenced UMIs) and compared to FPKM values from the bulk tissue control. Gene expression correlation values between pseudobulk and bulk are reported in Table S1.

Generation of Enhancer Knockout Mice—*Arx* enhancer knockout mice were generated using *in vivo* Cas9 editing methods adapted from those previously reported (Wang et al., 2013; Yang et al., 2013) (Figure S2). Briefly, Cas9 was amplified from plasmid pDD921 using primers T7Cas9_F and PolyACas9_R (Kvon et al., 2016; Table S7), and the resulting fragment was *in vitro* transcribed using the mMESSAGE mMACHINE T7 kit (Thermo Fisher Scientific).

Pairs of single guide RNAs (sgRNAs) were designed to target sites flanking each of the enhancers to be deleted (target site sequences listed in Table S7). sgRNAs were designed to have minimal possible off-target positions and were constructed using 60-mer oligos (Integrated DNA Technologies, Table S7) and a gRNA cloning vector (Addgene plasmid 41824; Mali et al., 2013) according to a publicly available protocol (http://www.addgene.org/static/cms/files/hCRISPR_gRNA_Synthesis.pdf). Complete sgRNA sequences were amplified from the cloning vector using primers listed in Table S7, and the resulting DNA fragments were *in vitro* transcribed using the MEGAshortscript T7 kit (Thermo Fisher Scientific). Cas9 mRNA and transcribed sgRNAs were purified using the MEGAclear kit (Thermo Fisher Scientific). Single stranded donor oligos (IDT, Table S7) were used to encourage homology driven repair to generate multiple F0 mice with identical deletion events for all enhancer deletions except hs123.

Cas9 mRNA (100 ng/μL final concentration), two sgRNAs (25 ng/μL final concentration each), and donor oligo (200 ng/μL) were mixed in microinjection buffer (10mM Tris, 0.1mM EDTA, pH 7.5), and the resulting mixture was injected into the cytoplasm of E0.5 FVB strain mouse embryos using standard procedures. All single enhancer knockouts were generated using fertilized eggs from wild-type FVB mice. Double enhancer knockouts were generated iteratively, using fertilized eggs harboring one of the deletions and delivering sgRNAs to target the other. Deletion events were identified by genotyping (primers listed in Table S7), and breakpoints were confirmed with Sanger sequencing. Details for founder mouse lines used are listed in Table S3. Founder lines were expanded and maintained by outcrossing mutation carriers with wild-type FVB animals to minimize the likelihood of a line harboring an off-target mutation.

RNA-Seq for Enhancer Knockouts—Individual forebrains were microdissected from E11.5 embryos by an experimenter blinded to the embryos' genotypes. Genotyping and gender determination for embryos were carried out using primers listed in Table S7. Samples were chosen for RNA-Seq such that a wild type littermate was sequenced for each enhancer knockout embryo sequenced. Because the enhancers reside on the X chromosome, only male embryos were used for RNA-Seq. RNA was isolated using the RNAqueous Total RNA

Isolation Kit (Thermo Fisher Scientific). RNA samples were DNase-treated with the TURBO DNA-free Kit (Thermo Fisher Scientific), and RNA quality was then assessed using a 2100 Bioanalyzer (Agilent) with an RNA 6000 Nano Kit. RNA sequencing libraries were made using the TruSeq Stranded mRNA Sample Prep Kit (Illumina). RNA-Seq libraries were subjected to an additional purification step using Agencourt AMPure XP beads (Beckman Coulter) to remove remaining high molecular weight products (Dickel et al., 2016). RNA-Seq library quality and concentration were assessed using a 2100 Bioanalyzer with the High Sensitivity DNA Kit (Agilent) and a Qubit Fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). RNA-Seq libraries were sequenced with single end 50 bp reads on a HiSeq 2500 or 4000 (Illumina) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. To eliminate batch variation effects, all knockout and wild-type embryos for a given line were processed in the same batch, and the resulting barcoded libraries were pooled together for sequencing on the same flow-cell lane. RNA-Seq analysis was carried out as previously described (Dickel et al., 2016). Briefly, filtering and mapping were performed using Illumina-provided software, along with cutadapt_v1.1 (Martin, 2011), TopHat v2.0.6 (Trapnell et al., 2009), and HTSeq (Anders et al., 2015). For each sample, between 35 and 58 million usable reads remained following quality filtering. Differential gene expression analysis was performed with edgeR (Robinson et al., 2010) and excluded genes with very low expression values, defined by Cufflinks v2.2.1 (Trapnell et al., 2010) as having a fragment per kilobase of transcript per million mapped reads (FPKM) less than 1. For differential gene expression analysis, the *P*-values reported in figures are those prior to false discovery rate (FDR) correction, and the *P*-values following genome-wide FDR correction for the double enhancer deletion lines are reported in Table S5.

In situ Hybridization—Isolated embryonic brains (E12.5 or E13.5) used for *in situ* hybridization (ISH) were fixed overnight in 4% paraformaldehyde (PFA), transferred to 15% sucrose overnight followed by 30% sucrose overnight for cryoprotection, and then embedded and frozen in OCT compound for cryosectioning. Cryostat sections were cut coronally at 20 μ m thickness. For *in situ* hybridization a rostro-caudal series of at least ten sections were examined. Section ISHs were performed using previously published digoxigenin-labeled riboprobes as previously described using RNase-free reagents and solutions (Stanco et al., 2014). Briefly, sections were washed in PBS, steamed in 10mM sodium citrate (pH=6), washed in PBS then acetylated for 10 minutes (1.3% Triethanolamine, 0.06% HCl, 0.38% acetic anhydride). The sections were then prehybridized for 30 mins at 65°C (in a solution of 50% formamide, 50ug/mL heparin, 50ug/mL yeast tRNA, 5xSSC and 1%SDS) and then hybridized with DIG-labeled riboprobes overnight at 65°C. Sections were then washed in 5xSSC (pH 4.5) for 5 minutes at room temperature then washed twice in 0.2X SSC at 72°C for 30 minute s. After a 5 minute wash in NTT (0.15M NaCL, 0.1M Tris pH 8, 0.1% Tween), sections were blocked in 5% heat inactivated sheep serum in 2% blocking reagent in NTT for 1 hour at room temperature. Sections were then incubated overnight in anti-Digoxigenin-AP antibody, and following three washes in NTTML (0.15M NaCL, 0.1M Tris pH 9.5, 0.1% Tween, 50mM MgCl₂, 2mM Levamisole), the signal was revealed using BM purple at 37°C. Sections were then fixed in 4% para formaldehyde and coverslipped.

The *Arx* probe was designed against exon 5 and 6 (chrX: 9054137-90543401 in mm9), linearized with PstI, and transcribed with a DIG labeling mix using the T3 polymerase (Roche). The probe was purified using an RNA clean and concentrator kit (Zymo Research).

Body Weight Measurements—Hemizygous-null males used for body weight analyses were selected based on availability. For all lines, controls were the available wild-type male littermates of the hemizygous null males analyzed. Mice were weighed using a standard electronic laboratory scale. Measurements were taken weekly beginning at weaning, with each animal being weighed on a consistent day of the week from week to week, by a handler un-blinded to genotype.

Sample Selection for Neurological Phenotyping—Because the *Arx* locus resides on the X chromosome, phenotyping was performed only on male mice. All animals used for neurological phenotyping ranged from generation F2 to F4. With the exception of two of the hs119 deletion animals and one wild-type control, all animals phenotyped resulted from at least two generations of out crossing to wild-type to minimize possible off-target effects. For the dorsal forebrain enhancer knockouts (hs122, hs123, hs122/hs123), animals were selected using a matched littermate strategy. For each knockout animal phenotyped, a wild-type male littermate was phenotyped as a control. For the ventral forebrain enhancer knockouts (hs119, hs121, hs119/hs121), the four wild-type controls were littermates of at least one of the knockout animals phenotyped. Details for animals phenotyped are provided in Table S6.

Brain Sectioning and Immunohistochemistry—Postnatal animals were anesthetized with intraperitoneal avertin (0.015 ml/g of a 2.5% solution) and perfused transcardially with PBS and then with 4% paraformaldehyde (PFA). Whole brain was isolated and then subjected to 4–5 hours fixation in 4% PFA, cryoprotection in 30% sucrose in PBS, and cut frozen coronally on a freezing sliding microtome at 40 μ m. For immunohistochemistry all primary and secondary antibodies were diluted in PBS containing 10% Normal Serum, 0.25% Triton X-100 and 2% BSA. The following primary antibodies were used: rabbit anti-parvalbumin (1:3000, Swant Swiss Abs), rat anti-somatostatin (1:200, Millipore), goat anti-somatostatin (1:100, Santa Cruz), mouse anti-calretinin (1:1,000, Millipore), rabbit anti-VIP (1:300, Immunostar) and goat anti-ChAT (1:100, Millipore). Sections processed for immunofluorescence were counterstained with Hoechst 33342 (Thermo Fisher Scientific). For indirect immunohistochemistry, sections were incubated with biotinylated secondary antibodies (Jackson), diluted 1:300, and processed by the ABC histochemical method (Vector Laboratories). Black reaction, to enhance contrast, was obtained by mixing 0.05% diaminobenzidine (DAB) plus 2.5% in a volume of 1% Cobalt chloride and 1% nickel ammonium sulfate. The reaction was carried out by adding 0.01% hydrogen peroxide (Adams, 1981).

Image Acquisition and Analysis—Brightfield images were taken using a Coolsnap camera (Photometrics) mounted on a Nikon Eclipse 80i microscope using NIS Elements acquisition software (Nikon). Brightness and contrast were adjusted and images merged using Photoshop or ImageJ software (Schneider et al., 2012).

Cell Counting—For assessing cell density in the postnatal neocortex on brightfield sections, 10x images were taken of the somatosensory cortex, encompassing all neocortical layers, from both hemispheres for each replicate. Images were opened with ImageJ (Schneider et al., 2012) to delimit and measure the region of interest (ROI). All cells in the ROI were counted by an investigator blinded to the animal's genotype. Cell counts were divided by the ROI area to determine cell density. Statistical analyses for cell counting were done with SPSS Statistics v15 software (SPSS Inc.). The statistical significance of multiple comparisons on continuous data was performed using the ANOVA test with a Tukey HSD or Scheffe posthoc to determine the significance between groups after checking that the data fit to a normal distribution (assessed by Shapiro-Wilk normality test) and that the variances were equal (determined by Levene's test).

Hippocampal Measurements—Brain sectioning and image acquisition were performed as described above. For this analysis, sections were stained with DAPI. Measures of dentate gyrus length were obtained using ImageJ by tracing the total length of the superior and inferior blades of the gyrus on serial sections of right and left hippocampi following the appearance of the superior limb of the dentate gyrus. The total length of the *Cornu Ammonis* (CA) was measured similarly. Surface area measures of the dentate gyrus were performed for the whole dentate gyrus and restricted to the granule cell layer. Measures in figures are presented normalized to the length or area of age-matched wild-type littermates, with raw measurements provided in Table S6. Measurements were performed blinded to genotype.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details regarding specific quantification and statistical analyses are provided throughout the main text and figures and in the **METHOD DETAILS** section above. Unless otherwise indicated, statistical analyses and plot generation were performed using R (www.r-project.org).

DATA AND SOFTWARE AVAILABILITY

All RNA-Seq data are available through GEO under identification code GSE100394. No software was generated for this project.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by U.S. National Institutes of Health grants to L.A.P. and A.V. (R24HL123879, R01HG003988, UM1HG009421, UM1HL098166) and to J.L.R.R. (NINDS R01NS34661 and NIMH R37MH049428). The single cell RNA sequencing work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory (to D.E.D.). A.R.Y. was supported by a grant from Fondation Fyssen. M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. Research was conducted at the E.O. Lawrence Berkeley National Laboratory and performed under U.S. Department of Energy Contract DE-AC02-05CH11231, University of California. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, which is supported by National Institutes of Health Grants S10RR029668 and S10RR027303. We would like to thank qb3 MacroLab for plasmid pML6B and J. Doudna (University of California Berkeley) for plasmid pMJ920 and for advice on Cas9/CRISPR-mediated genome editing.

References

- Adams JC. Heavy metal intensification of DAB-based HRP reaction product. *J Histochem Cytochem.* 1981; 29:775. [PubMed: 7252134]
- Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 2007; 5:e234. [PubMed: 17803355]
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015; 31:166–169. [PubMed: 25260700]
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science.* 2004; 304:1321–1325. [PubMed: 15131266]
- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological).* 1995; 57:289–300.
- Chen YJJ, Friedman BA, Ha C, Durinck S, Liu J, Rubenstein JL, Seshagiri S, Modrusan Z. Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types. *Sci Rep.* 2017; 7:45656. [PubMed: 28361918]
- Chiang CWK, Derti A, Schwartz D, Chou MF, Hirschhorn JN, Wu CT. Ultraconserved elements: analyses of dosage sensitivity, motifs and boundaries. *Genetics.* 2008; 180:2277–2293. [PubMed: 18957701]
- Colasante G, Collombat P, Raimondi V, Bonanomi D, Ferrai C, Maira M, Yoshikawa K, Mansouri A, Valtorta F, Rubenstein JLR, 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]
- Colasante G, Simonet JC, Calogero R, Crispi S, Sessa A, Cho G, Golden JA, Broccoli V. ARX regulates cortical intermediate progenitor cell expansion and upper layer neuron formation through repression of Cdkn1c. *Cereb Cortex.* 2015; 25:322–335. [PubMed: 23968833]
- Collombat P, Mansouri A, Hecksher-Sorensen J, Serup P, Krull J, Gradwohl G, Gruss P. Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev.* 2003; 17:2591–2603. [PubMed: 14561778]
- Colombo E, Collombat P, Colasante G, Bianchi M, Long J, Mansouri A, Rubenstein JLR, Broccoli V. Inactivation of Arx, the murine ortholog of the X-linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation. *J Neurosci.* 2007; 27:4786–4798. [PubMed: 17460091]
- Crawley, JN. *What's Wrong With My Mouse?: Behavioral Phenotyping of Transgenic and Knockout Mice.* Wiley; 2007.
- del Río JA, de Lecea L, Ferrer I, Soriano E. The development of parvalbumin-immunoreactivity in the neocortex of the mouse. *Brain Res Dev Brain Res.* 1994; 81:247–259. [PubMed: 7813046]
- Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, May D, Spurrell CH, Plajzer-Frick I, Pickle CS, et al. Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nat Commun.* 2016; 7:12923. [PubMed: 27703156]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29:15–21. [PubMed: 23104886]
- Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd.* 1996; 96:226–231.
- Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 2015; 16:278. [PubMed: 26653891]
- Gez J, Cloosterman D, Partington M. ARX: a gene for all seasons. *Curr Opin Genet Dev.* 2006; 16:308–316. [PubMed: 16650978]
- Gould E, Woolf NJ, Butcher LL. Postnatal development of cholinergic neurons in the rat: I. Forebrain. *Brain Res Bull.* 1991; 27:767–789. [PubMed: 1664779]
- Harmston N, Baresic A, Lenhard B. The mystery of extreme non-coding conservation. *Philos Trans R Soc Lond, B, Biol Sci.* 2013; 368:20130021. [PubMed: 24218634]

- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012; 22:1760–1774. [PubMed: 22955987]
- Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC Known Genes. *Bioinformatics.* 2006; 22:1036–1046. [PubMed: 16500937]
- Johnson KD, Hsu AP, Ryu M-J, Wang J, Gao X, Boyer ME, Liu Y, Lee Y, Calvo KR, Keles S, et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J Clin Invest.* 2012; 122:3692–3704. [PubMed: 22996659]
- Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat Meth.* 2017; 14:483–486.
- Kitamura K, Yanazawa M, Sugiyama N, Miura H, Iizuka-Kogo A, Kusaka M, Omichi K, Suzuki R, Kato-Fukui Y, Kamiirisa K, et al. Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nat Genet.* 2002; 32:359–369. [PubMed: 12379852]
- Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development.* 1989; 105:707–714. [PubMed: 2557196]
- Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissières V, Pickle CS, Plajzer-Frick I, Lee EA, et al. Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell.* 2016; 167:633–642. e11. [PubMed: 27768887]
- Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.* 2015; 161:1202–1214. [PubMed: 26000488]
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, Dicarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* 2013; 339:823–826. [PubMed: 23287722]
- Marsh ED, Nasrallah MP, Walsh C, Murray KA, Nicole Sunnen C, McCoy A, Golden JA. Developmental interneuron subtype deficits after targeted loss of Arx. *BMC Neurosci.* 2016; 17:35. [PubMed: 27287386]
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal.* 2011; 17:10–12.
- McLean C, Bejerano G. Dispensability of mammalian DNA. *Genome Res.* 2008; 18:1743–1751. [PubMed: 18832441]
- Motenko H, Neuhauser SB, O’Keefe M, Richardson JE. MouseMine: a new data warehouse for MGI. *Mamm Genome.* 2015; 26:325–330. [PubMed: 26092688]
- Nolte MJ, Wang Y, Deng JM, Swinton PG, Wei C, Guindani M, Schwartz RJ, Behringer RR. Functional analysis of limb transcriptional enhancers in the mouse. *Evol Dev.* 2014; 16:207–223. [PubMed: 24920384]
- 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]
- Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2005; 33:D501–D504. [PubMed: 15608248]
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140. [PubMed: 19910308]
- Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 2015; 43:D670–D681. [PubMed: 25428374]
- Sandelin A, Bailey P, Bruce S, Engström PG, Klos JM, Wasserman WW, Ericson J, Lenhard B. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics.* 2004; 5:99. [PubMed: 15613238]
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth.* 2012; 9:671–675.

- Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005; 21:3940–3941. [PubMed: 16096348]
- Stanco A, Pla R, Vogt D, Chen Y, Mandal S, Walker J, Hunt RF, Lindtner S, Erdman CA, Pieper AA, et al. NPAS1 represses the generation of specific subtypes of cortical interneurons. *Neuron*. 2014; 84:940–953. [PubMed: 25467980]
- Taniguchi H, He M, Wu P, Kim S, Paik R, Sugino K, Kvitsiani D, Kvitsani D, Fu Y, Lu J, et al. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron*. 2011; 71:995–1013. [PubMed: 21943598]
- Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105–1111. [PubMed: 19289445]
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010; 28:511–515.
- 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, Taher L, Girgis H, May D, Golonzhka O, Hoch RV, McKinsey GL, Pattabiraman K, Silberberg SN, Blow MJ, et al. A high-resolution enhancer atlas of the developing telencephalon. *Cell*. 2013; 152:895–908. [PubMed: 23375746]
- Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglu S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat Meth*. 2017; 14:414–416.
- Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013; 153:910–918. [PubMed: 23643243]
- Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013; 154:1370–1379. [PubMed: 23992847]

Highlights

- The *Arx* locus has a high density of very long ultraconserved forebrain enhancers
- Single or pairwise deletion of these enhancers results in viable mice in all cases
- In most cases, individual enhancer loss results in developmental anomalies
- Selection against these fitness deficits likely contributes to ultraconservation

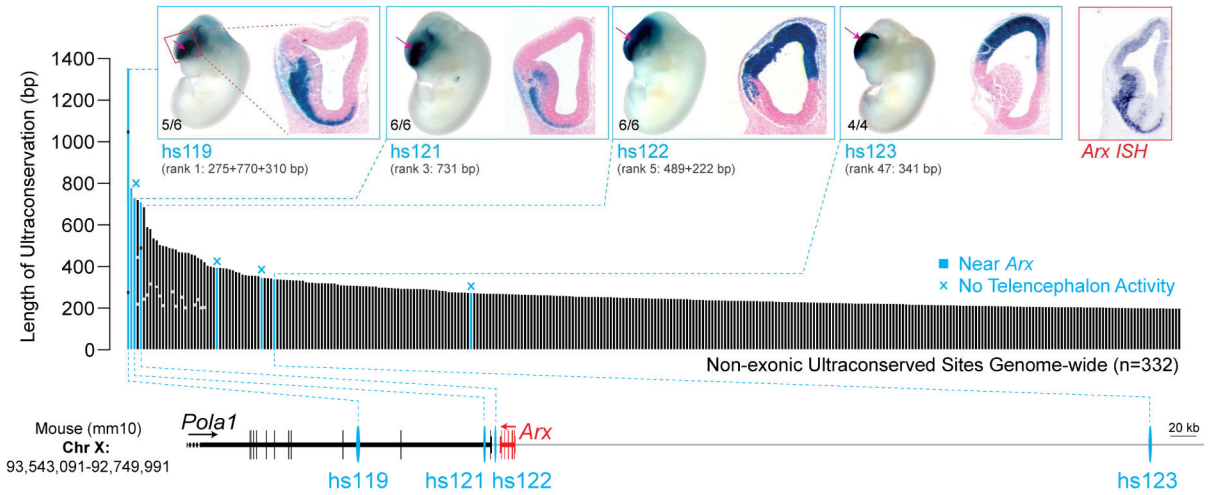


Figure 1. Selection of ultraconserved sites examined

Bar graph shows the length of ultraconservation for all non-exonic human-rodent ultraconserved sites. Regions containing multiple sequences of perfect human-mouse-rat conservation 200bp with less than 1 kb of intervening sequence were combined into a single ultraconserved site, with black or light gray tick marks indicating the length of the individual constitutive ultraconserved elements. **Above:** representative images showing the activity pattern (blue staining, pink arrows) of each telencephalon enhancer near *Arx* in an embryonic day 12.5 (E12.5) transgenic mouse embryo (left; Pennacchio et al., 2006) and specifically in the telencephalon in coronal forebrain sections (right; Visel et al., 2013). VISTA enhancer identifiers are indicated in blue text (hs numbers), and transgenic assay reproducibility is indicated in the bottom left corner of each panel. *Arx* is expressed in both the dorsal and ventral telencephalon at E12.5 (far right). *ISH: in situ* hybridization. **Below:** An overview of the *Arx* genomic locus, located on the X chromosome, showing the position of telencephalon enhancers selected for further study (blue ovals).



Figure 2. Ultraconserved enhancers near *Arx* are active in *Arx*-expressing cells

A) To identify the specific cell type each enhancer is active in, we generated E12.5 transgenic embryos containing each one of the enhancers driving a fluorescent reporter gene and performed single-cell RNA-Seq by Drop-Seq on isolated forebrain tissue. **B)** Cumulative results of single-cell RNA-Seq performed on E12.5 forebrain from all transgenic enhancer-reporter experiments. Each point indicates a unique single cell profiled (4,723 total), and cells were clustered by similarity of RNA expression. Red indicates cells where at least one *Arx* transcript was captured, and these points are scaled by decile of *Arx* expression, as shown in the legend. **C)** Same plot as **B**, color-coded to indicate cells where one of the ultraconserved telencephalic enhancers drove reporter gene expression. Points indicating mCherry expression are scaled by expression level decile (larger points indicate higher expression). See also Figure S1 and Tables S1–S2.

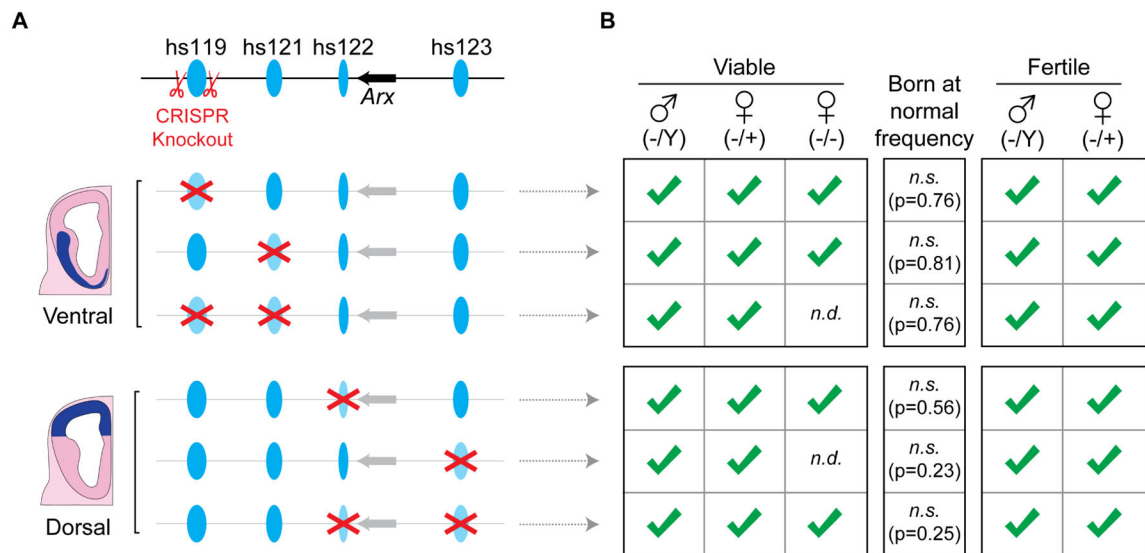


Figure 3. Single and double knockout of ultraconserved enhancers results in viable mice
A) Schematic diagrams of each enhancer knockout line generated using CRISPR/Cas9 editing in fertilized mouse eggs. Ultraconserved forebrain enhancers are shown as blue ovals, with red crosses indicating enhancer deletion alleles. Schematic locus representation is not to scale. **B)** In all cases, loss of individual or pairs of ultraconserved enhancers near *Arx* resulted in viable heterozygous female and hemizygous-null male mice (Figure S2) that did not show any obvious embryonic lethality or loss of fertility (Table S4). *n.d.*: not determined, *n.s.*: not significant. See also Figure S2 and Tables S3–S4.

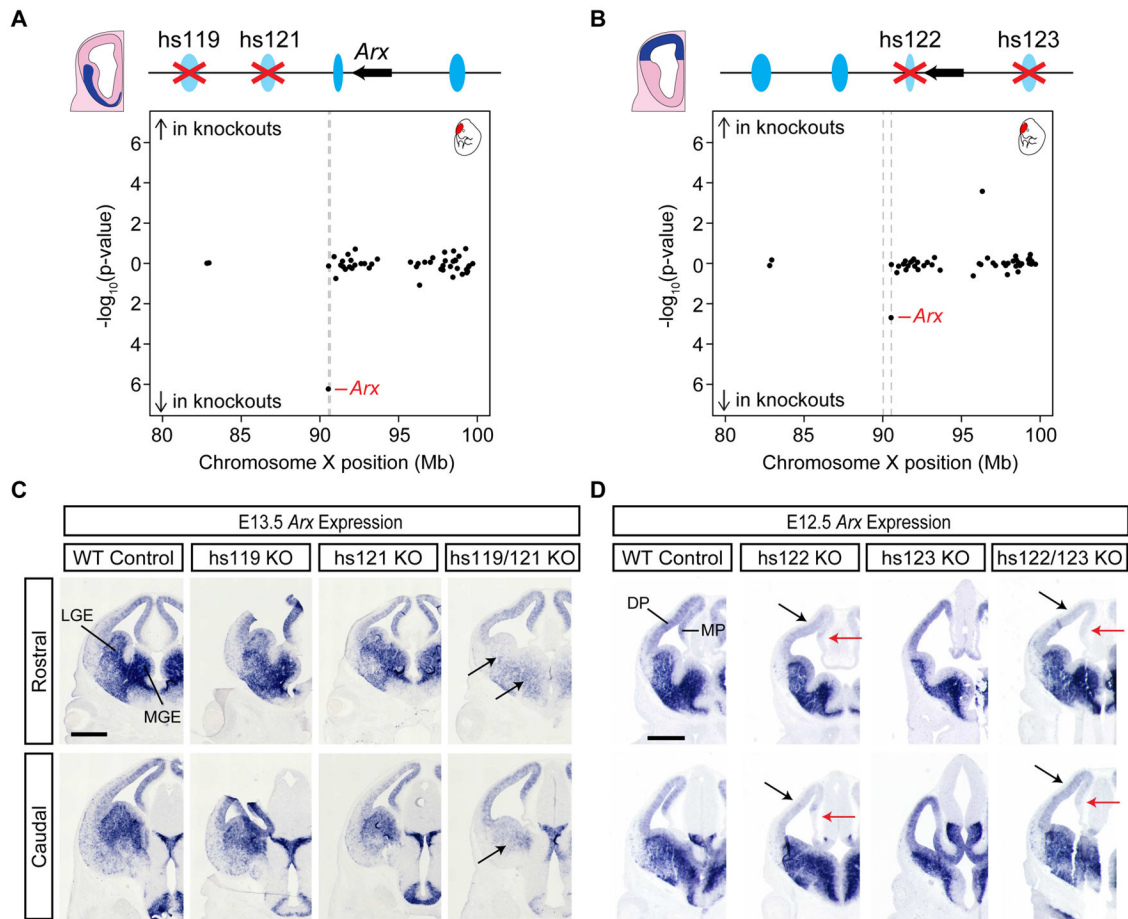


Figure 4. *Arx* expression is diminished in enhancer deletion mice

A,B RNA-Seq results for double knockouts of hs119/hs121 (**A**) and hs122/hs123 (**B**), showing all genes within a 20 Mb window around *Arx*. Whole forebrain from E11.5 embryos was profiled for each pair of deletions (n=2 wild-type, n=2 hemizygous-null males for each panel). Each point represents an individual gene, with its position on the X chromosome (x-axis) and significance of deviation from wild-type (y-axis) indicated. Genes with decreased expression in knockout embryos are plotted with $-\log_{10}(P)$ scores below 0. Dashed gray lines mark the positions of the enhancers. **C** *Arx* expression (blue staining) profiled by *in situ* hybridization in coronal sections of E13.5 forebrain from embryos missing the enhancers active in the ventral forebrain. Embryos missing both enhancers in combination (hs119/hs121 KO) have substantial deficits of *Arx* in the ventral forebrain (black arrows). **D** *Arx* expression in E12.5 forebrain from embryos missing enhancers active in the dorsal forebrain. Embryos missing both enhancers in combination have decreased *Arx* expression in the dorsal pallium (black arrows) and in the medial pallium (red arrows), with hs122-null embryos having similar changes. Scale bars: 500 μm ; LGE: lateral ganglionic eminence, MGE: medial ganglionic eminence; DP: dorsal pallium, MP: medial pallium. See also Figures S3–S4, Table S5.

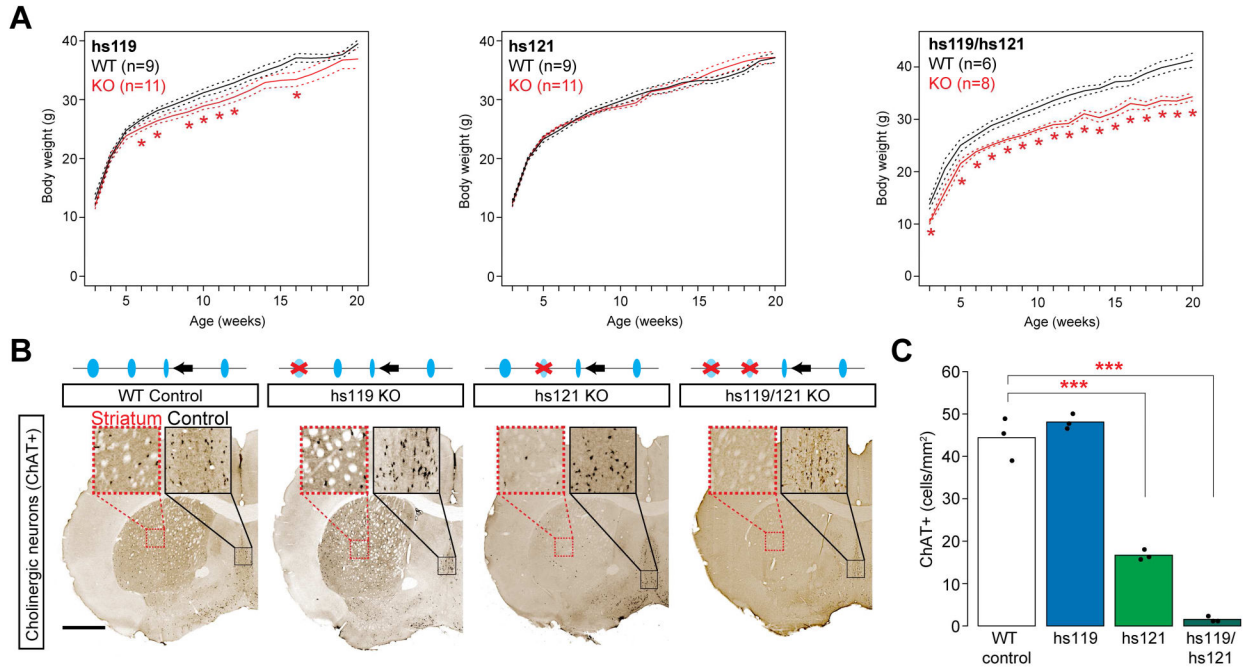


Figure 5. Loss of ultraconserved ventral forebrain enhancers results in growth and brain abnormalities

A) Growth curves for males wild-type (WT, black) or hemizygous null (KO, red) for the ventral forebrain enhancers. Solid lines show mean weight, with dashed lines indicating standard error of the mean (SEM). Sample sizes are shown on each plot. *, $P < 0.05$; two-tailed t-test. **B)** Representative coronal cross-sections of postnatal mouse brains stained for choline acetyltransferase (ChAT, black dots) from a wild-type control and males hemizygous-null (KO) for the enhancers active in the ventral forebrain. The loss of hs121 alone or hs119/hs121 combined results in decreased density of cholinergic neurons in the striatum (red box insets), whereas ChAT+ populations in other structures are unaffected (black box insets). Scale bar: 1 mm. **C)** Quantification of striatal cholinergic neuron density for ventral enhancer knockouts. Bars indicate group means, while individual points represent biological replicates. ***, $P < 0.001$; ANOVA. See also Figures S5–S6 and Table S6

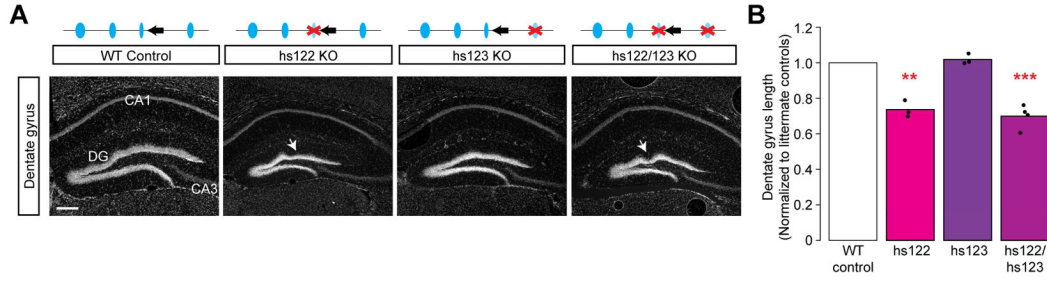


Figure 6. Loss of ultraconserved dorsal forebrain enhancers results in abnormal brain development

A) Representative rostral coronal cross-sections of postnatal hippocampus from a wild-type control, along with males hemizygous null for the enhancers active in the dorsal forebrain. Individual loss of hs122 or combined loss of hs122/hs123 results in a smaller dentate gyrus (DG) with disorganized appearance (white arrowheads). CA1/3: hippocampal *Cornu Ammonis* 1 and 3. Scale bar: 250µm. **B)** Dentate gyrus length for dorsal forebrain enhancer knockouts normalized to wild-type littermate controls. Bars indicate group means, with individual points showing biological replicates. **, $P < 0.01$; ***, $P < 0.001$; two-tailed t -test. See also Figure S7 and Table S6.

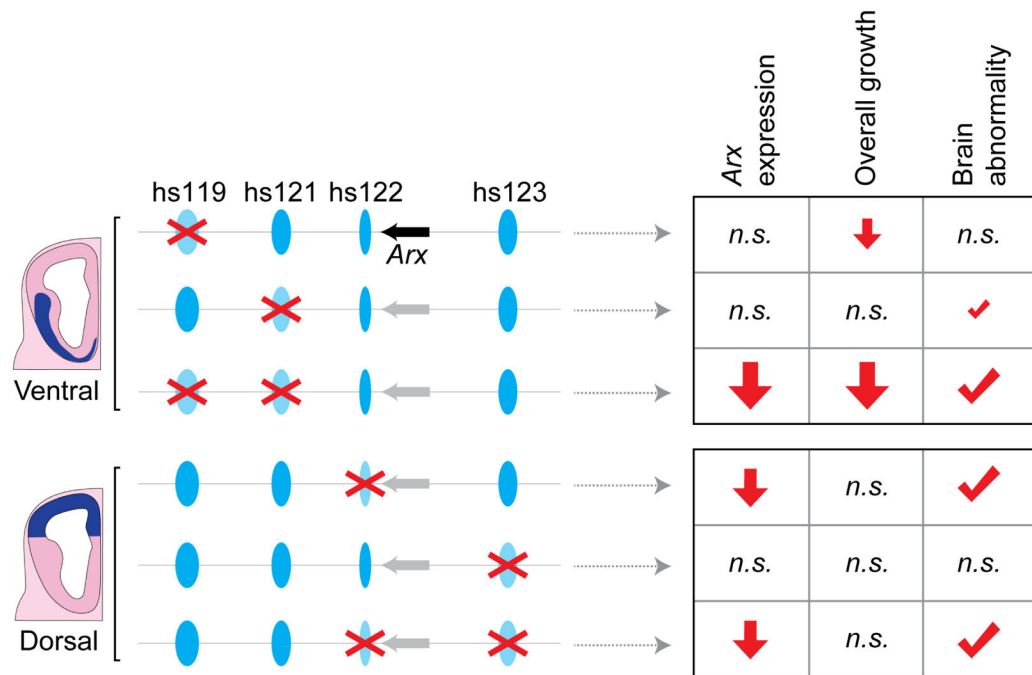


Figure 7. Summary of phenotypes resulting from loss of ultraconserved *Arx* enhancers
 n.s.: no significant changes observed.