Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene *PRODH*

Maria Suntsova[a,b], Elena V. Gogvadze[a], Sergey Salozhin[c], Nurshat Gaifullin[d,e], Fedor Eroshkin[a], Sergey E. Dmitriev[f], Natalia Martynova[a], Kirill Kulikov[e], Galina Malakhova[a], Gulnur Tukhbatova[c], Alexey P. Bolshakov[c], Dmitry Ghilarov[a], Andrew Garazha[a,b], Alexander Aliper[a,b], Charles R. Cantor[g,1], Yuri Solokhin[e], Sergey Roumiantsev[b], Pavel Balaban[c], Alex Zhavoronkov[b], and Anton Buzdin[a,b,1]

[a]Group for Genomic Regulation of Cell Signaling Systems, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow 117997, Russia; [b]Laboratory of Bioinformatics, D. Rogachyov Federal Research Center of Pediatric Hematology, Oncology and Immunology, Moscow 117198, Russia; [c]Laboratory of Molecular Neurobiology, Institute of Higher Nervous Activity and Neurophysiology, Moscow 117485, Russia; [d]Faculty of Fundamental Medicine and [f]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119192, Russia; [e]Department of Pathology, Russian National Research Medical University named after N. I. Pirogov, Moscow 117997, Russia; and [g]Department of Biomedical Engineering, Boston University, Boston, MA 02215

Contributed by Charles R. Cantor, October 19, 2013 (sent for review August 2, 2013)

Using a systematic, whole-genome analysis of enhancer activity of human-specific endogenous retroviral inserts (hsERVs), we identified an element, hsERV$_{PRODH}$, that acts as a tissue-specific enhancer for the *PRODH* gene, which is required for proper CNS functioning. PRODH is one of the candidate genes for susceptibility to schizophrenia and other neurological disorders. It codes for a proline dehydrogenase enzyme, which catalyses the first step of proline catabolism and most likely is involved in neuromediator synthesis in the CNS. We investigated the mechanisms that regulate hsERV$_{PRODH}$ enhancer activity. We showed that the hsERV$_{PRODH}$ enhancer and the internal CpG island of *PRODH* synergistically activate its promoter. The enhancer activity of hsERV$_{PRODH}$ is regulated by methylation, and in an undermethylated state it can up-regulate PRODH expression in the hippocampus. The mechanism of hsERV$_{PRODH}$ enhancer activity involves the binding of the transcription factor SOX2, whch is preferentially expressed in hippocampus. We propose that the interaction of hsERV$_{PRODH}$ and *PRODH* may have contributed to human CNS evolution.

human-specific endogenous retrovirus | DNA methylation | central nervous system | human speciation | retroelement

Understanding the molecular basis of phenotypic differences between humans and chimpanzees can provide important clues to human-specific behavioral peculiarities and neurological disorders. For this purpose we conducted a genome-wide analysis of human-specific endogenous retroviral (hsERV) inserts that may induce new regulatory pathways by acting as promoters and enhancers (1, 2). HsERVs of the HERV-K(HML-2) group are one of the four families of transposable elements that were able to transpose at the time of the radiation of human lineage from the lineage of its most closely related species, chimpanzee (3). At least 50% of all hsERV elements exhibit promoter activity in human tissues (4). We found only six hsERV inserts in the upstream regions of known human genes, close to transcription start sites. Three of them displayed strong enhancer activity in transient transfection experiments; of these three, only one—near the *PRODH* gene—matched the transcriptional activity pattern of its endogenous genomic copy. This copy of hsERV is a full-length, almost intact betaretrovirus belonging to the HERV-K(HML-2) group. *PRODH* encodes a mitochondrial enzyme proline, dehydrogenase (oxidase), that converts proline to D-1-pyrroline-5-carboxylate (5). PRODH regulates proline catabolism, which is vital for normal CNS functioning. Several *PRODH* mutations are associated with neuropsychiatric disorders, such as schizophrenia (6). Gene knockouts in mice cause severe changes in the executive functioning of the brain (7). Given the potential importance of *PRODH* in brain functioning and disease, we attempted to characterize its newly recognized hsERV$_{PRODH}$ enhancer. We showed that hsERV$_{PRODH}$ enhancer activity is regulated by methylation and that the hsERV$_{PRODH}$ enhancer and *PRODH* internal CpG island act synergistically to activate its promoter. *PRODH* transcription analyses demonstrated the highest expression level in the hippocampus, where hsERV$_{PRODH}$ is hypomethylated. Moreover, the hsERV$_{PRODH}$ enhancer, together with the *PRODH* promoter and CpG island, caused neuron-specific expression. We also found that hsERV$_{PRODH}$ contains two functional sites for binding the transcription factor SOX2 that activates its enhancer activity and is expressed predominantly in hippocampus. Our data shed light on the transcriptional regulation of *PRODH* and identify a human-specific enhancer that is activated in hippocampus.

Results

Enhancer Activity Tests of Individual hsERVs. Bioinformatic screening of all 133 hsERVs previously identified by us and other investigators (8) resulted in the identification of six elements inserted in the close vicinity (<5 kb of the transcription start sites) of known human genes. These hsERVs mapped upstream of *SOCS4*, *PRODH*, *NDUFV1*, *ZFP3*, *KIAA1919*, and *c3orf17* (Fig. S1*A*).

We next attempted to demonstrate the enhancer activity of these elements using transient transfection experiments. The

Significance

We identified a human-specific endogenous retroviral insert (hsERV) that acts as an enhancer for human *PRODH*, hsERV_PRODH. *PRODH* encodes proline dehydrogenase, which is involved in neuromediator synthesis in the CNS. We show that the hsERV_PRODH enhancer acts synergistically with the CpG island of *PRODH* and is regulated by methylation. We detected high *PRODH* expression in the hippocampus, which was correlated with the undermethylated state of this enhancer. *PRODH* regulatory elements provide neuron-specific transcription in hippocampal cells, and the mechanism of hsERV_PRODH enhancer activity involves the binding of transcriptional factor SOX2. Because *PRODH* is associated with several neurological disorders, we hypothesize that the human-specific regulation of *PRODH* by hsERV_PRODH may have played a role in human evolution by upregulating the expression of this important CNS-specific gene.

Author contributions: M.S., E.V.G., S.S., F.E., S.E.D., A.P.B., C.R.C., Y.S., S.R., P.B., A.Z., and A.B. designed research; M.S., E.V.G., N.G., F.E., S.E.D., N.M., K.K., G.M., G.T., A.P.B., D.G., A.G., A.A., Y.S., and A.B. performed research; M.S., E.V.G., S.S., S.E.D., A.G., A.A., C.R.C., S.R., A.Z., and A.B. analyzed data; and M.S., E.V.G., S.S., S.E.D., C.R.C., S.R., A.Z., and A.B. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: ccantor@sequenom.com or bu3din@mail.ru.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1318172110/-/DCSupplemental.

genomic sequence upstream of the transcription start site including the hsERV element (LTR^+ constructs) or excluding the hsERV element (LTR^- constructs) was cloned into a luciferase-reporter construct for each of the six genes (Fig. 1*A*) and was used in a luciferase assay (Fig. 1*B* and Fig. S1*B*). For better accuracy, we used a Dual-Luciferase Reporter Assay (Promega). For the upstream regions of *SOCS4*, *PRODH*, and *KIAA1919*, we detected strong and statistically significant enhancer activity, with a normalized luciferase activity LTR^+/LTR^- ratio of seven- to 10-fold (Fig. 1*B*). For *c3orf17*, there was only a twofold difference in the luciferase activities of the LTR^+ and LTR^- constructs (Fig. S1*B*). The enhancer effects of the hsERVs were restricted to one of the four cell lines tried, Tera-1, suggesting that it is tissue specific.

A number of studies suggest that transient transfection data may differ dramatically from the expression patterns seen for endogenous copies of the same genes when located in their native, genomic environment (9). Thus, examination of endogenous gene copy expression is needed to refine transient transfection data. We used quantitative RT-PCR (qRT-PCR) to evaluate transcriptional activities of all genes in our study in the same four cell lines used for the transient transfection experiments (Fig. 1*C* and Fig. S1*C*). In five of the six genes there was no correlation between the transcriptional activities of the endogenous gene copies and their transfected upstream regulatory regions for either the LTR^+ and LTR^- constructs. However, for *PRODH* (Fig. 1 *B* and *C*) we found a correlation between the promoter activities of the LTR^+ construct and the endogenous copy of this gene, with clearly enhanced activities for both in Tera-1 cells, but there was no correlation between the transcription of the endogenous copy and the LTR^- construct. Thus, only three elements displayed strong enhancer activity (by increasing the expression of neighboring genes from the authentic promoters by seven- to 10-fold) when tested in four human cell lines of the different etiologies. However, we cannot exclude the possibility that the other hsERVs tested may be able to show rather distinct enhancer activities in the other human cell types. Of these three hsERVs, only one element, $hsERV_{PRODH}$, displayed enhancer activity in the luciferase tests that matched the transcriptional pattern of its endogenous copy.

We further attempted to compare the promoter activities of the human $hsERV_{PRODH}$ -lacking (LTR^-) and $hsERV_{PRODH}$ -containing (LTR^+) *PRODH* upstream regions (Fig. 1*D*) with the orthologous copy of the chimpanzee *PRODH* upstream region that lacks this retroviral insertion. For this purpose we amplified the chimpanzee 2.4-kb-long *PRODH* upstream region and cloned it into a luciferase-reporter vector. We found that the activity of the chimpanzee upstream sequence was comparable to that of the human LTR^- construct and was weaker than that of the LTR^+ construct. These data demonstrate that, in some cell types, the human *PRODH* upstream region harboring the $hsERV_{PRODH}$ insert is significantly more transcriptionally active than the orthologous chimpanzee sequence.

The $hsERV_{PRODH}$ enhancer therefore could represent a human-specific regulation of gene expression that arose after the separation of human and chimpanzee ancestors.

$HsERV_{PRODH}$, a Human-Specific Enhancer That Acts Synergistically with a CpG Island to Activate *PRODH* Transcription. Little is known about the regulation of *PRODH* transcription. Sequence analysis indicates that this gene includes an internal CpG island (CpG_{PRODH}) that partly overlaps with its second exon (Fig. 2*A*). To investigate whether the inclusion of CpG_{PRODH} affects the activity of the $hsERV_{PRODH}$ enhancer, we added a 230-bp-long fragment overlapping with the CpG_{PRODH} sequence to the luciferase reporter construct (Fig. 2*B*). This fragment included the whole first *PRODH* noncoding exon, the short first intron (62 bp), and a portion of the second exon up to the translation initiation codon. For these experiments, we used Tera-1 cells. We found that presence of the CpG_{PRODH} sequence strongly reinforced the activity of the $hsERV_{PRODH}$ enhancer (Fig. 2*B*). In contrast, CpG_{PRODH} had no effect on the promoter activity of the *PRODH* upstream sequence that lacked $hsERV_{PRODH}$ (Fig. 2*B*). These data suggest that $hsERV_{PRODH}$ and the CpG_{PRODH} may act synergistically in *PRODH* transcriptional regulation. We also identified a DNase hypersensitivity cluster inside $hsERV_{PRODH}$, which appeared to be even more important than that in the beginning of the *PRODH* gene itself, as reflected by its cluster score values (Fig. 2*A*). This finding clearly indicates that $hsERV_{PRODH}$ and CpG_{PRODH} have functional activity in vivo.

The Methylation Status of $hsERV_{PRODH}$ Regulates *PRODH* Transcription. DNA methylation is thought to be involved in silencing HERV-K (HML 2) transcriptional regulatory elements (10, 11). To test this possibility, we transfected Tera-1 cells with in vitro-methylated LTR^-, LTR^+, or control constructs (the control included an SV40 promoter upstream of the luciferase reporter). The degree of methylation varied depending on the length of treatment with CpG-DNA methylase (0–30 min) (Fig. S2*A*). Results show that methylation dramatically reduces both $hsERV_{PRODH}$ and SV40 reporter transcription activities. This result is consistent with the idea that DNA methylation is vital for the regulation of HERV-K (HML 2) transcriptional activities.

These transient transfection results showed that $hsERV_{PRODH}$ hypermethylation may inhibit enhancer activity, and hypomethylation may increase enhancer activity. To uncover the methylation status of $hsERV_{PRODH}$ in the four human cell lines used thus far in this study, we performed bisulfite sequencing covering the whole length of the $hsERV_{PRODH}$ LTR. We found that the $hsERV_{PRODH}$ was extensively methylated in cell lines HepG2, A549, and NGP127 but was hypomethylated in Tera-1 cells (Fig. S2*B*). These low levels of methylation and increased expression of *PRODH* in Tera-1 cells

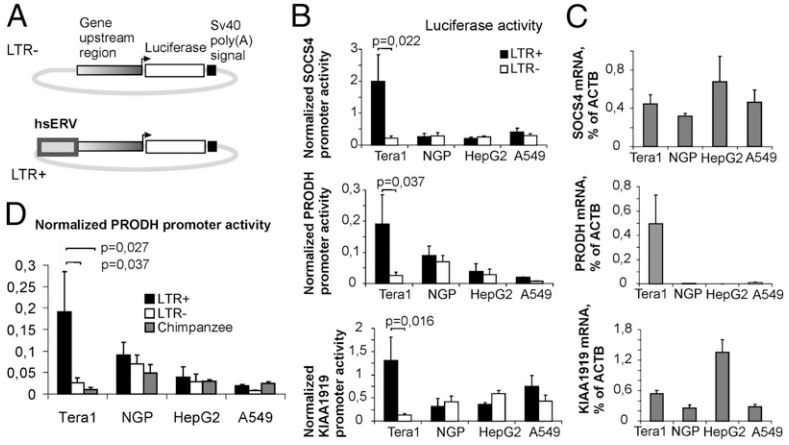

Fig. 1. Comparison of luciferase reporter assay data with transcriptional activities of the endogenous gene copies. (*A*) Scheme of luciferase reporter constructs. (*B*) Promoter activities of LTR^+ and LTR^- constructs in Tera-1, NGP127, HepG2, and A549 cell lines (normalized to the SV40 promoter activity). (*C*) mRNA levels of *SOCS4*, *PRODH*, and *KIAA1919* genes in cell lines measured by qRT-PCR relative to the endogenous β-actin (*ACTB*) gene expression. (*D*) Promoter activities of human and chimpanzee *PRODH* upstream regions. LTR^-, human *PRODH* promoter region; LTR^+, human *PRODH* promoter region including hsERV LTR; Chimpanzee, orthologous chimpanzee PRODH promoter (lacking hsERV). Data show means ± SD of three independent experiments.

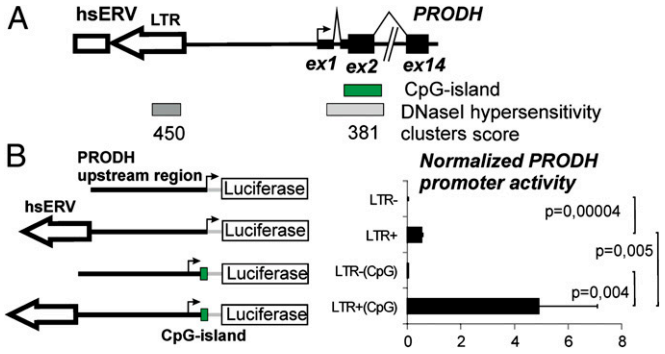

Fig. 2. Effect of the PRODH CpG island on hsERV$_{PRODH}$-enhancer activity. (*A*) The *PRODH* gene upstream region. Black bars indicate *PRODH* exons; black arrow, *PRODH* transcription start site; open arrow, hsERV LTR; open bar, hsERV internal region; green bar, CpG island; gray bar, DNase I hypersensitive clusters. Numbers under gray bars indicate cluster scores. Data were taken from the University of California Santa Cruz Genome Browser, http://genome.ucsc.edu. (*B*) (*Left*) Schematic representation of luciferase reporter constructs. (*Right*) Relative *PRODH* promoter activity, normalized to SV40 promoter activity. Data show means ± SD of three independent experiments.

suggest that the hsERV$_{PRODH}$ may be an active, tissue-specific enhancer in Tera-1 cells. A similar analysis of methylation of CpG$_{PRODH}$ revealed that, unlike hsERV$_{PRODH}$, CpG$_{PRODH}$ was hypomethylated in all of the cells tested (Fig. S2*B*).

PRODH encodes an isoform of proline dehydrogenase that, according to previous reports (7, 12) (human http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc002zoj.4&hgg_prot=E7EQL6&hgg_chrom=chr22&hgg_start=18900286&hgg_end=18923806&hgg_type=knownGene&db=hg19&hgsid=294069969), is predominantly transcribed in the heart, lung, liver, and CNS. We confirmed these reports with qRT-PCR data on samples of human testicular, bladder, heart, lung, and mixed brain tissues (Fig. S2*C*). The highest *PRODH* transcription was seen in brain samples, congruent with a putative role for *PRODH* in CNS functioning.

We examined hsERV$_{PRODH}$ and CpG$_{PRODH}$ methylation in the same brain samples using bisulfite sequencing. As in our previous data, CpG$_{PRODH}$ was equally undermethylated in all tissues (Fig. S2*D*). In contrast, hsERV$_{PRODH}$ was heavily methylated in most molecules; however, in several molecules hsERV$_{PRODH}$ was almost completely free of methylation (arrows in Fig. S2*D*). Methylated alleles may represent cells in which the hsERV$_{PRODH}$-enhancer element is suppressed, whereas nonmethylated alleles may correspond to cells with active enhancer.

Overall, our data suggest that the activity of the hsERV$_{PRODH}$ enhancer is regulated differentially through its methylation and that the expression of *PRODH* is regulated by mechanism(s) other than promoter and CpG$_{PRODH}$ methylation. Moreover, the unmethylated state of the CpG$_{PRODH}$ that we observed in all tissues under investigation could not by itself be sufficient for providing high transcriptional activity for that gene.

hsERV$_{PRODH}$–CpG$_{PRODH}$ Interplay in Human Brain. Because *PRODH* is brain specific, and because the brain contains a fraction of cells with unmethylated hsERV$_{PRODH}$, we next examined *PRODH* expression and CpG$_{PRODH}$ in different brain tissue types using qRT-PCR. We performed qRT-PCR on 20 brain tissue types taken from a single donor (Fig. 3*A*) and found that *PRODH* is highly expressed in the hippocampus, parietal lobe, and occipital lobe (Fig. 3*B*). We also detected differences between the left and

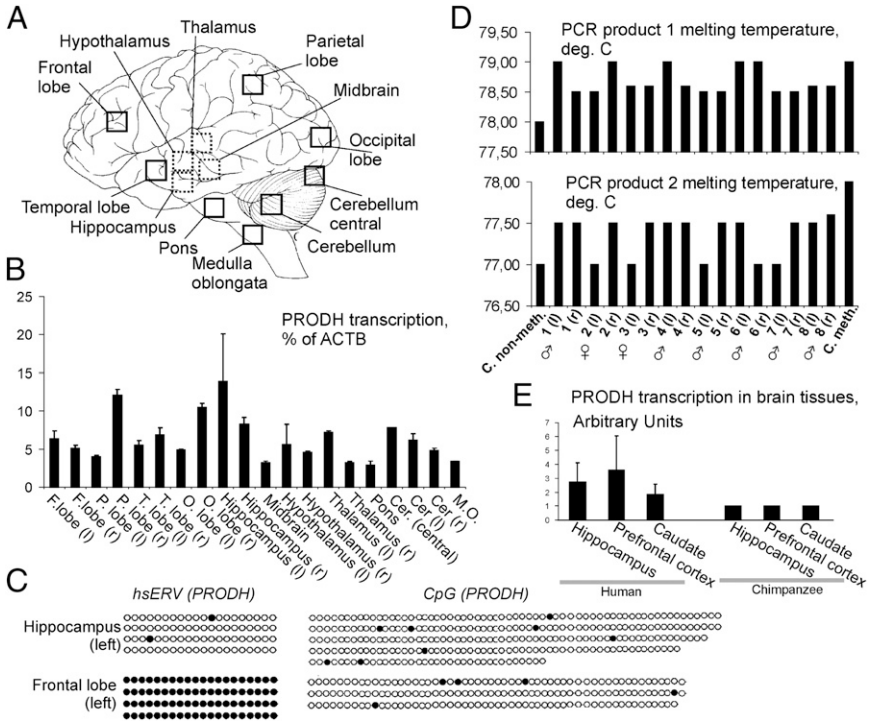

Fig. 3. Functional characterization of the *PRODH* locus in the brain tissues. (*A*) Schematic view of the brain sections investigated. (*B*) Expression of *PRODH* in human brain tissues measured using qRT-PCR relative to *ACTB*. Data show means ± SD of three independent experiments. (*C*) Representative methylation patterns of the hsERV$_{PRODH}$ (*Left*) and CpG$_{PRODH}$ (*Right*) in the left hippocampus and left hemisphere of the frontal lobe. Black circle, methylated CG dinucleotide; white circle, unmethylated CG dinucleotide. (*D*) High-resolution melting profiling of bisulfite-treated DNA from the left (l) and right (r) hemisphere of human hippocampi. Data are shown for two PCR-amplified bisulfite-treated CG-rich fragments of the hsERV$_{PRODH}$. C.meth, methylated sequence control; C non-meth, unmethylated sequence control. (*E*) Transcriptional activity of *PRODH* in human and chimpanzee brain tissues. The data were extracted from the NCBI GEO database, and the fold-change differences in gene-expression levels in individual human tissue samples and in the average chimpanzee tissue were calculated. Arbitrary units represent the fold-change difference between the *PRODH* expression in human samples and the chimpanzee median *PRODH* expression.

right hemispheres of the parietal lobe, occipital lobe, and thalamus (Fig. 3*B*), but further studies using a greater sampling are required to validate the significance of this effect.

We next used a bisulfite sequencing assay to assess the methylation status of hsERV$_{PRODH}$ and CpG$_{PRODH}$ in these tissues. CpG$_{PRODH}$ was hypomethylated in all tissues. In contrast, hsERV$_{PRODH}$ was strongly methylated in all tissues except the left hemisphere of the hippocampus (Fig. 3*C*). This region also contained the highest transcriptional activity of *PRODH* among the investigated tissues (Fig. 3*B*). Subsequent analyses using a methylation-sensitive high-resolution melting assay on eight additional, independent left and right hippocampus samples obtained from the healthy human donors revealed that undermethylation of the hsERV$_{PRODH}$ in the hippocampus is generally characteristic of the human population (Fig. 3*D*).

Finally, we compared *PRODH* gene transcription in human and chimpanzee using available microarray gene-expression data extracted from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo/). We interrogated 18 samples of human brain tissue and 18 samples of chimpanzee brain tissue, representing hippocampus, prefrontal cortex, and caudate (Dataset S1). These samples included seven human and six chimpanzee hippocampal samples. In all the brain sections tested we observed clear up-regulation of PRODH in human as compared with the chimpanzee samples (Fig. 3*E*).

Cell-Type Specificity of *PRODH* Transcriptional Activity. We next examined whether the transcriptional activity of *PRODH* is restricted to specific hippocampal cell types. We transformed cultured rat hippocampal cells with a lentiviral construct (Fig. 4*A*) containing GFP under the control of a constitutive CMV promoter and a reporter gene for the red fluorescent protein DsRed under the control of the *PRODH* promoter. We detected three clearly distinct morphological cell types in hippocampal tissue: neurons, astrocytes, and microglial cells. At least 50 fluorescent cells were observed for each cell type. In all cases, DsRed fluorescence was restricted to neurons, but GFP fluorescence was shown in all three cell types (Fig. 4 *B* and *C*), suggesting that *PRODH* expression in the hippocampus is specific to neurons.

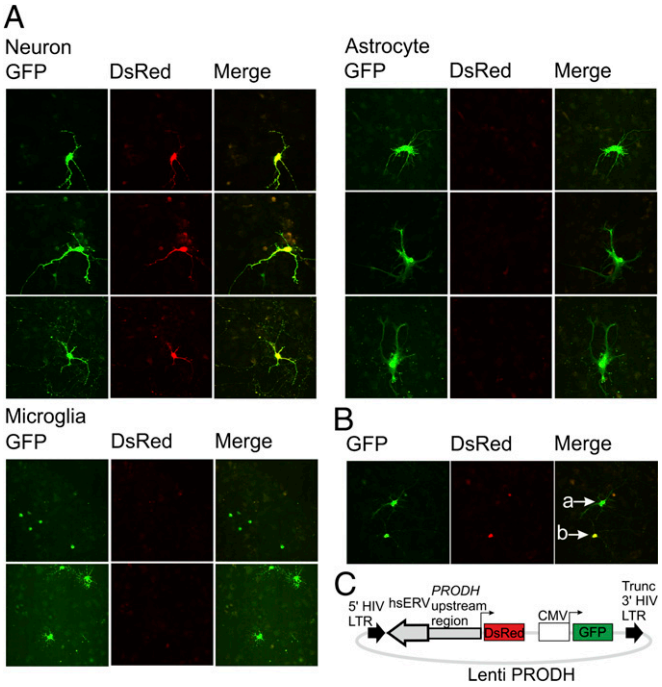

Fig. 4. Transformation of rat hippocampal cells with a lentiviral construct carrying the upstream human *PRODH* region. (*A*) Schematized structure of the lentiviral construct. Red fluorescent protein (DsRed) was placed under the transcriptional control of the PRODH promoter; the GFP gene was placed under the control of the constitutive CMV promoter (+ control). (*B*) *PRODH* promoter activity in different hippocampal cell types. (*C*) Comparison of DsRed vs. GFP fluorescence in astrocytes (a) and neurons (b).

Mechanism of hsERV$_{PRODH}$ Enhancer Activity. To identify which transcription factors are responsible for hsERV$_{PRODH}$ enhancer activity, we bioinformatically screened the hsERV$_{PRODH}$ LTR for transcription factor-binding sites (TFBSs) using the MatInspector (13) and Alibaba2.1 (www.gene-regulation.com/pub/programs/alibaba2/index.html) programs. We identified ~200 putative TFBSs. To identify TFBSs that may contribute functionally to hsERV$_{PRODH}$ enhancer activity, we compared transcriptomes of Tera-1 (hsERV$_{PRODH}$ enhancer–positive) cells with those of the cells that do not display hsERV$_{PRODH}$ enhancer activity, i.e., A549, NGP127, HepG2, and NT2/D1 (hsERV$_{PRODH}$-enhancer–negative) (Fig. S3) and with the sample of human left hemisphere hippocampal tissue. We identified a set of transcription factor genes that were up-regulated in enhancer-positive cells (Datasets S2 and S3). Only two genes, *SOX2* and *NFKB1*, exhibited TFBSs that had been predicted previously using bioinformatics. Subsequent qRT-PCR analyses showed that both genes were up-regulated simultaneously in Tera-1 cells and in the hippocampus and were down-regulated in all four enhancer-negative cell types (Datasets S2 and S3). A subsequent dendrogram analysis of gene-expression profiles showed that Tera-1 and hippocampus samples cluster together (Fig. S4).

To determine if *SOX2* and *NFKB1* functionally affect hsERV$_{PRODH}$ activity, we overexpressed these genes in Tera-1 cells and in the enhancer-negative cell line NT2/D1. Overexpression of *NFKB1* had no effect on enhancer activity of hsERV$_{PRODH}$ in either cell line (Fig. S5*A*), but overexpression of *SOX2* in NT2/D1 cells resulted in a strong enhancer effect for hsERV$_{PRODH}$ (Fig. 5*A*, *Lower*). This effect also was seen in Tera-1 cells, despite their higher background levels of *SOX2* (Fig. 5*A*, *Upper*).

We next conducted an EMSA with radiolabeled double-stranded oligonucleotides corresponding to predicted hsERV$_{PRODH}$ TFBSs for *NFKB1* and *SOX2* gene products (Fig. S5*B*). We found no bands for predicted hsERV$_{PRODH}$-binding sites for *NFKB1* (Fig. S5 *B* and *D*), suggesting that the NFKB1 is not involved in mediating the enhancer activity of hsERV$_{PRODH}$. In a similar experiment with SOX2, two binding sites within the hsERV$_{PRODH}$ sequence showed a strong positive signal (Fig. S5 *B* and *C*). Experiments with the competitor oligonucleotides confirmed that SOX2–TFBS binding was sequence specific for both the investigated recognition sites (Fig. S5*B*). Both these sites contained the predicted recognition motif AACAAAG. Subsequent experiments comparing TFBS binding of nuclear extracts from cell types confirmed binding in enhancer-positive cell types (Tera-1 and SOX2-overexpressed NTD/D1) and no binding in enhancer-negative cells (Fig. 5*B*). Our assays also found that four-nucleotide mutations in the SOX2 recognition motif disabled binding of SOX2 to the enhancer (Fig. 5*C*).

We next produced SOX2 protein in vitro to determine if TFBS recognition is mediated by SOX2 binding. SOX2-coding mRNA was synthesized and used for translation with the cytoplasmic extract of rabbit reticulocytes. *Renilla* luciferase was used as a negative control. Radiolabeled probes contained either wild-type or mutant SOX2 TFBSs (Fig. 5*D*). We detected binding with both wild-type TFBSs and with the positive control sequence, whereas in the SOX2 translation mix there was no binding with either mutant TFBS (Fig. 5*D*). These data provide strong evidence that the enhancer activity of hsERV$_{PRODH}$ is caused by SOX2 binding. Interestingly, our bioinformatic analysis revealed that both these SOX2 TFBSs are highly conserved among the LTRs of the endogenous retroviral family HERV-K (HML-2) and occur in 84% and 51% of them (i.e., in 557 and 366 genomic copies), respectively (Fig. S6). Thus, HERV-K

GENETICS

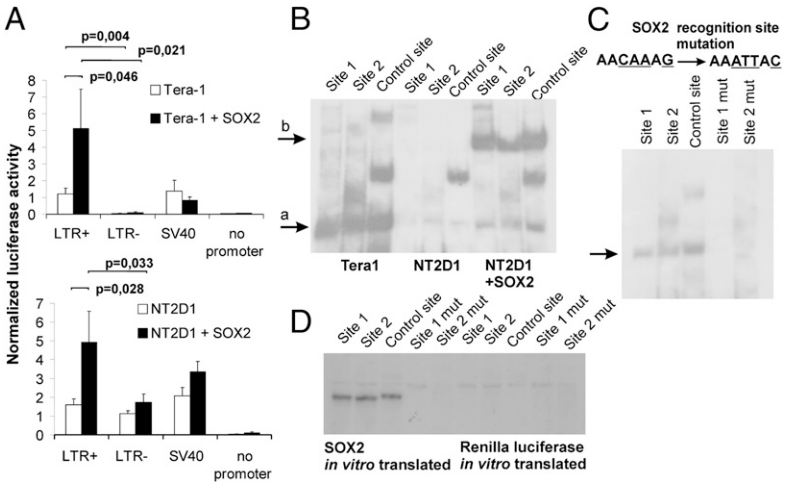

Fig. 5. Regulation of hsERV$_{PRODH}$ enhancer activity by the SOX2 transcription factor. (*A*) Effect of *SOX2* overexpression on hsERV$_{PRODH}$ enhancer activity in Tera-1 (*Upper*) and NT2/D1 (*Lower*) cells. Data show means ± SD of three independent experiments. (*B*) EMSA for SOX2-binding sites with nuclear extracts of Tera-1 cells, NT2/D1 cells, and NT2/D1 cells overexpressing *SOX2*. (*C*) EMSAwith mutated SOX2-binding sites (mutated nucleotide positions are underlined). (*D*) EMSA for the SOX2-binding sites with in vitro-produced SOX2 protein. *Renilla* Luciferase protein was used as a negative control. Sites 1 and 2 SOX2, recognition sites 1 and 2 within hsERV$_{PRODH}$; sites 1 and 2 mut, respective mutated sites; control site, control recognition sequence.

(HML-2)–mediated SOX2 binding in vivo may not be limited to the regulation of the gene *PRODH* transcription but also may be used elsewhere in the human genome.

Discussion

The importance of *PRODH* for CNS functioning is clear, but its precise role is not. A deletion in the genomic locus 22q11, which contains *PRODH*, been implicated in DiGeorge syndrome, a condition associated with several neurological disorders, including schizophrenia (14). PRODH was identified as a schizophrenia-associated gene in some studies (15, 16) but not in another (17). PRODH is a mitochondrial enzyme catalyzing the first step of proline catabolism, and PRODH malfunctions cause hyperprolinemia type I. This condition is known to cause neurological abnormalities (18, 19). *PRODH* also may play a role in the synthesis of neuromediators. Proline can be converted to glutamate in two intermediate steps (20). Glutamate, in turn, may act as a neurotransmitter itself or may be further converted to GABA (21). Because proline may be used for glutamate synthesis in neurons (22), *PRODH* malfunction may cause a neuromediator misbalance (7).

Here, we show that *PRODH* is positively regulated by an hsERV insert. We hypothesize that, at some point in human evolution, the hsERV$_{PRODH}$ insertion may have influenced *PRODH* transcriptional activity significantly and increased its expression in the CNS. We show that the hsERV$_{PRODH}$, together with the *PRODH* promoter, drives neuron-specific expression in cultured hippocampal cells. Remarkably, hsERV$_{PRODH}$ was hypomethylated in human hippocampal samples, where *PRODH* showed the highest transcriptional activity. We also uncovered a mechanism for the regulation of hsERV$_{PRODH}$ involving the binding of SOX2. In the brain, SOX2 is expressed preferentially in the hippocampus, which is known to be one of the brain structures most affected in schizophrenia (23). Our findings suggest a potential role of hsERV$_{PRODH}$ in CNS functioning. Discovery of this CNS-specific PRODH enhancer encourages further studies to discover the molecular features affecting its functional activity and its connection with neurological disorders. Our results can stimulate further studies of how methylation of PRODH regulatory regions interplays with human brain functioning in vivo. hsERV$_{PRODH}$ itself is a repetitive element and is flanked at both ends by other low-diverged genomic repeats (http://genome-euro.ucsc.edu/cgi-bin/hgTracks?position=chr22:18920392-18928317&hgsid=193710100&rmsk=full), making it a difficult target for precise DNA manipulation. However, further development of transgenesis technologies probably will make it possible in the future to explore in full the regulatory potential of hsERV$_{PRODH}$ and of other human-specific inserts of transposable elements.

Methods

Additional methods and detailed protocols are available in *SI Methods*.

Cell Lines. The cell lines Tera-1 (testicular embryonal germ cell tumor), NT2/D1 (partly differentiated testicular germ cells with CNS-precursor cell characteristics), NGP127 (neuroblastoma), HepG2 (hepatocarcinoma), and A549 (lung carcinoma) were used in this study. Cells were grown on DMEM/F12 medium containing 10% (vol/vol) fetal calf serum (Invitrogen) at 37 °C and 5% CO_2.

Tissue Samples. Tissue samples were taken from human heart, lung, bladder, testicles, brain cortex, hippocampus, midbrain, pons, and cerebellum (left, central, right) and from the left and right frontal lobe, parietal lobe, temporal lobe, occipital lobe, hippocampus, hypothalamus, thalamus, and medulla oblongata,. All samples were taken within 24 h after death from adult (19- to 63- y-old) donors killed in road accidents. Informed consent was obtained from donor representatives, and the experiments were approved by the Institutional Review Board of the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry.

Primary Neuronal Cell Culture. Primary neuronal culture was prepared from newborn Wistar rat pup hippocampi.

DNA and RNA Extraction, cDNA Synthesis, PCR Amplification, and Cloning. Genomic DNA was isolated using the Wizard Genomic DNA Purification Kit (Promega), RNA was extracted using the SV Total RNA Isolation System (Promega), and first-strand cDNA synthesis was conducted using the Mint reverse transcription kit (Evrogen) according to the manufacturers' recommendations. PCR amplification of upstream gene regions and bisulfite-converted DNA was conducted with the Encyclo PCR Kit (Evrogen). For qRT-PCR, we used the qPCRmix-HS SYBR (Evrogen). qRT-PCR was performed using a MxPro3000 thermocycler (Stratagene). Primer sequences are listed in Table S1. For amplification of promoter regions we used genomic DNA isolated from human placenta and chimpanzee mixed brain tissues (~40 ng per reaction). The pGEM-T Easy vector (Promega) was used for cloning PCR products, followed by subcloning into a pGL3-basic vector (Promega) and pDsRed-Express-N1 vector (Clontech).

Lentiviral Vector Construction. A 42-bp linker containing Bsp120I and MluI restriction sites was cloned by AvaI and HpaI into a lentiviral vector p156RRLsinPPTCMVGFPPRE (a gift from Alon Chen, the Weizmann Institute of Science, Rehovot, Israel). An expression cassette containing DsRed fluorescent protein under the control of the *PRODH* upstream regulatory

sequence was subcloned into the resulting vector using NotI and MluI restriction sites.

Cell Transfections and Luciferase Assays. Cell transfections were performed in 24-well plates using Unifectin-56 transfection reagent according to the manufacturer's recommendations; 0.5μg DNA was used per well. pGL3-based reporter constructs carrying the upstream region of the tested gene and the firefly luciferase gene were mixed in a ratio of 10/1 with pRL-TK vector used as an internal control. Firefly luciferase values were normalized to those of *Renilla* luciferase. For the overexpression of human transcriptional factors SOX2 and NFKB1, we used pMXs-hSOX2 (Addgene; vector was kindly provided by Maria Lagarkova, Vavilov Institute of General Genetics, Moscow) and phNFKB1 (vector was kindly provided by Alexander Belyavsky, Engelhardt Institute of Molecular Biology, Moscow). Here, *SOX2* and *NFKB1* genes are under the control of the standard CMV promoter. Luciferase activity was measured in quadruplicate 48 h after transfection using the Dual Luciferase Reporter Assay (Promega) and a GENios Pro luminometer (Tecan).

Microscopy. Mounting was conducted with ProLong Gold antifade reagent (Life Technologies). Images were taken on an LSM5Life Confocal microscope and were pseudocolored using ImageJ software (24).

Microarray Hybridization. Microarray hybridization was performed with the HumanHT-12 v4 Expression BeadChip Kit (Illumina) according to the manufacturer's protocol. Microarray hybridization data are available on GEO (NCBI), accession no. GSE43773.

In Vitro DNA Methylation. pGL3-based reporter constructs were methylated in vitro using M.Sss I CpG-methyltransferase (SibEnzyme) according to the manufacturer's protocol.

Bisulfite Sequencing. Bisulfite conversion of genomic DNA was performed using the EpiTect Bisulfite Kit (Qiagen) followed by amplification with the EpiTect Whole Bisulfitome Kit (Qiagen). DNA then was amplified by nested-PCR (primers are given in Table S1).

In Vitro Translation. To obtain mRNA coding for SOX2, PCR products were synthesized using the pMXs-hSOX2 plasmid as the template, with the forward primer SOX2T7, which contained an introduced T7 priming site. The reverse primer, T50, contained a 50-nt-long 3′ poly(T) repeat as described in ref. 24. The RiboMAX kit (Promega) was used for in vitro transcription, which was followed by LiCl precipitation and an m7G capping step (ScriptCap; Epicentre Biotechnologies). This step was followed by a second LiCl precipitation. A capped and polyadenylated mRNA encoding *Renilla* luciferase was obtained in a similar way (25). For in vitro translation, nuclease-treated rabbit reticulocyte lysate (RRL; Promega) was used, and the manufacturer's protocol was followed (26).

Nuclear Extract Preparation. Nuclear extracts were performed as described in ref. 27 and were aliquoted and stored at −70 °C.

EMSA. Annealed oligonucleotides were radiolabeled with an α32P-dATP using Klenow polymerase and then were purified in a polyacrylamide gel and stored at −20 °C. Nuclear extract (5 μg of total protein per reaction) was preincubated with 750 ng of noneukaryotic DNA. Labeled oligonucleotides (30,000 cpm) were added to nuclear extract and incubated for 30 min at 20 °C. After incubation, samples were loaded onto a polyacrylamide gel for electrophoresis. Gels were dried and exposed to X-ray film for 48–120 h.

Statistics and Dataset Analysis. For statistical analysis we used STATISTICA software (www.statsoft.com/) and a *t* test. To compare gene expression in human and in chimpanzee, we chose the GSE33010 dataset from the GEO database (www.ncbi.nlm.nih.gov/geo/). We processed the samples analyzed on the platform Affymetrix Human Genome U133 Plus 2.0 Array. Raw microarray data from 36 samples were normalized with the GCRMA algorithm (28) and were summarized using redefined probe set definitions. For pairwise comparison of brain gene-expression profiles, we defined six groups of samples: *Homo sapiens* hippocampus (seven samples), *Pan troglodytes* hippocampus (six samples), *Homo sapiens* prefrontal cortex (four samples), *Pan troglodytes* prefrontal cortex (six samples), *Homo sapiens* caudate (seven samples), and *Pan troglodytes* caudate (six samples). Samples from *Pan troglodytes* were used as a reference.

ACKNOWLEDGMENTS. We thank Dr. S. B. Akopov (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry) for providing cell lines, Drs. S. A. Lukyanov and D. A. Shagin (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry) for assistance with qRT-PCR, and the UMA Foundation for support in manuscript preparation. This work was supported by Russian Foundation for Basic Research Grants 12-04-33094 (to A.B.), 12-04-31028 (to M.S.), 12-04-33196 (to S.E.D.), and 11-04-00682 (to N.M.) and by the Presidium of the Russian Academy of Sciences Program "Dynamics and Conservation of Genomes."

1. Bannert N, Kurth R (2004) Retroelements and the human genome: New perspectives on an old relation. *Proc Natl Acad Sci USA* 101(Suppl 2):14572–14579.
2. Gogvadze E, Stukacheva E, Buzdin A, Sverdlov E (2009) Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions. *J Virol* 83(12): 6098–6105.
3. Mills RE, et al. (2006) Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* 78(4):671–679.
4. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006) At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. *J Virol* 80(21): 10752–10762.
5. Kowaloff EM, Phang JM, Granger AS, Downing SJ (1977) Regulation of proline oxidase activity by lactate. *Proc Natl Acad Sci USA* 74(12):5368–5371.
6. Kempf L, et al. (2008) Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function. *PLoS Genet* 4(11):e1000252.
7. Gogos JA, et al. (1999) The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. *Nat Genet* 21(4):434–439.
8. Buzdin A, et al. (2003) Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: Three master genes were active simultaneously during branching of hominoid lineages. *Genomics* 81(2):149–156.
9. Sestáková B, Vachtenheim J (2006) Distinct co-regulation of endogenous versus transfected MITF-dependent tyrosinase promoter. *Folia Biol (Praha)* 52(5): 161–166.
10. Lavie L, Kitova M, Maldener E, Meese E, Mayer J (2005) CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K (HML-2). *J Virol* 79(2):876–883.
11. Stengel S, Fiebig U, Kurth R, Denner J (2010) Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. *Genes Chromosomes Cancer* 49(5):401–411.
12. Campbell HD, Webb GC, Young IG (1997) A human homologue of the Drosophila melanogaster sluggish-A (proline oxidase) gene maps to 22q11.2, and is a candidate gene for type-I hyperprolinaemia. *Hum Genet* 101(1):69–74.
13. Cartharius K, et al. (2005) MatInspector and beyond: Promoter analysis based on transcription factor binding sites. *Bioinformatics* 21(13):2933–2942.
14. Bassett AS, et al. (2005) Clinical features of 78 adults with 22q11 Deletion Syndrome. *Am J Med Genet A* 138(4):307–313.
15. Liu H, et al. (2002) Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. *Proc Natl Acad Sci USA* 99(6):3717–3722.
16. Bender HU, et al. (2005) Functional consequences of PRODH missense mutations. *Am J Hum Genet* 76(3):409–420.
17. Ripke S, et al.; Multicenter Genetic Studies of Schizophrenia Consortium; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2; Management Committee; Data and Analysis Group; DNA, Genotyping, Data QC and Informatics Group; Publications Committee (2013) Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 45(10):1150–1159.
18. Bavaresco CS, Streck EL, Netto CA, Wyse AT (2005) Chronic hyperprolinemia provokes a memory deficit in the Morris water maze task. *Metab Brain Dis* 20(1):73–80.
19. Clelland CL, et al. (2011) Evidence for association of hyperprolinemia with schizophrenia and a measure of clinical outcome. *Schizophr Res* 131(1-3):139–145.
20. Hu CA, Lin WW, Valle D (1996) Cloning, characterization, and expression of cDNAs encoding human delta 1-pyrroline-5-carboxylate dehydrogenase. *J Biol Chem* 271(16):9795–9800.
21. Johnson JL, Roberts E (1984) Proline, glutamate and glutamine metabolism in mouse brain synaptosomes. *Brain Res* 323(2):247–256.
22. Renick SE, et al. (1999) The mammalian brain high-affinity L-proline transporter is enriched preferentially in synaptic vesicles in a subpopulation of excitatory nerve terminals in rat forebrain. *J Neurosci* 19(1):21–33.
23. McCarley RW, et al. (1999) MRI anatomy of schizophrenia. *Biol Psychiatry* 45(9): 1099–1119.
24. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
25. Dmitriev SE, et al. (2007) Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5′ untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol Cell Biol* 27(13):4685–4697.
26. Dmitriev SE, Terenin IM, Rubtsova MP, Shatskiĭ IN (2003) [Minor secondary-structure variation in the 5′-untranslated region of the beta-globin mRNA changes the concentration requirements for eIF2]. *Mol Biol (Mosk)* 37(3):494–503.
27. Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11(5):1475–1489.
28. Wu Z, Irizarry R, Gentleman R, Murillo FM, Spencer F (2004) A model based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc* 99:909–917.