Published in final edited form as:

Pharmacogenet Genomics. 2009 January ; 19(1): 53–65. doi:10.1097/FPC.0b013e328318b21a.

Twenty-one-base-pair insertion polymorphism creates an enhancer element and potentiates *SLC6A1* GABA transporter promoter activity

Runghana Hirunsatit^{a,e}, Elizabeth D. George^a, Barbara K. Lipska^f, Hani M. Elwafi^a, Lisa Sander^a, Carolyn M. Yrigollen^b, Joel Gelernter^{a,e,c}, Elena L. Grigorenko^b, Jaakko Lappalainen^{a,e}, Shrikant Mane^d, Angus C. Nairn^a, Joel E. Kleinman^f, and Arthur A. Simen^a

^a Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut

^b Department of Psychology and Child Study Center, Yale University School of Medicine, New Haven, Connecticut

^c Department of Genetics and Neurobiology, Yale University School of Medicine, New Haven, Connecticut

^d Molecular Biophysics and Biochemistry and the Yale Keck Center, Yale University School of Medicine, New Haven, Connecticut

^e VA Connecticut Healthcare System, West Haven

^f National Institutes of Mental Health, Section on Neuropathology, Clinical Brain Disorders Branch, GCAP, IRP, NIMH, NIH, Bethesda, Maryland, USA

Abstract

Objective—Sodium-dependent and chloride-dependent γ -aminobutyric acid (GABA) transporter 1 (SLC6A1) is the target of a number of drugs of clinical importance and is a major determinant of synaptic GABA concentrations. We resequenced the human *SLC6A1* gene previously and discovered a novel 21 bp insertion in the predicted promoter region that creates a second tandem copy of the sequence. Here we sought to determine the functional relevance of this variation.

Methods—We used reporter assays, mobility shift assays, quantitative PCR, and proteomics methods as well as postmortem expression analysis for this work.

Results—Reporter assays showed that the insertion allele significantly increases promoter activity in multiple cell lines. The zinc finger transcription factor ZNF148 was found to significantly transactivate the promoter and increase expression when overexpressed but could not account for the differences in activity between the two alleles of the promoter. Copy number of the insertion

Correspondence to Dr Arthur A. Simen, MD, PhD, Divisions of Molecular Psychiatry, Human Genetics, and Aging Research, Department of Psychiatry, Yale University School of Medicine, Suite 901, New Haven, CT 06511-6624, USA Tel: +203 785 7361; fax: + 203 785 7357; e-mail: Arthur.Simen@yale.edu.

Present address: Runghana Hirunsatit, Center for Excellence in Genetics of Cancer and Human Diseases, Chulalongkorn University, Bangkok, Thailand; Jaakko Lappalainen, AstraZeneca Pharmaceuticals, Wilmington DE 19850, USA

Supplementary data

Supplementary data are available at the *Pharmacogenetics and Genomics* journal online (www.pharmacogeneticsandgenomics.com).

Conflicts of interest: Dr Gelernter has received financial support or compensation from the following: related to consultation for Columbia University, the University of CT Health Center, NIH, and Faegre & Benson; related to grant reviews for the National Institutes of Health; and related to academic lectures and editorial functions in various scientific venues (including for the ACNP). Dr Lappalainen is currently employed by AstraZeneca Pharmaceuticals. His effort on this project was restricted to the time that he was employed by Yale University and the West Haven VA Healthcare System.

sequence was associated with exponentially increasing activity of a downstream promoter, suggesting that the insertion sequence has enhancer activity when present in multiple copies. *SLC6A1* promoter genotype was found to predict *SLC6A1* RNA expression in human postmortem hippocampal samples. These results suggest that the insertion polymorphism leads to increased *SLC6A1* promoter activity because, in part, of creation of an enhancer element when present as multiple copies. Genotyping individuals from Tanzania in this study suggested that the insertion allele has its origin in Africa.

Conclusion—On account of the effect of the insertion on promoter activity, this relatively common polymorphism may prove useful in predicting clinical response to pharmacological modulators of *SLC6A1* as well as GABAergic function in individuals of African descent.

Keywords

African continental ancestry group; enhancer elements; γ -aminobutyric acid plasma membrane transport proteins; transcription factors; zinc fingers

Introduction

Abnormal function of the γ -aminobutyric acid (GABA) system has been implicated in a wide variety of psychiatric and neurological disorders including mood disorders [1,2], seizure disorder [3], drug dependence [4], and alcohol dependence [5]. GABA receptor agonists and antagonists such as benzodiazepines, gabapentin, and pregabalin, are efficacious in the treatment of many of these disorders [6–10]. GABA transporters play an important role in terminating GABA action by removing GABA from the synaptic cleft. Four GABA transporters have been identified: *SLC6A1* (GAT-1), *SLC6A13* (GAT-2), *SLC6A11* (GAT-3), and *SLC6A12* (BGT-1). They can be distinguished by their pharmacological properties and tissue distribution. *SLC6A1* is the most abundantly expressed GABA transporter subtype in the brain and serves to cotransport two sodium ions, one chloride ion, and one GABA molecule across membranes [11].

A study in rat cortex showed that *SLC6A1* expression overlaps with the expression of the 67 kDa isoform of the GABA synthetic enzyme glutamic acid decarboxylase (GAD67). Most cortical *SLC6A1* expression was seen in cortical layer IV, followed by layers II–III. GAT-1 positive axonal terminals were seen to impinge preferentially on the soma and proximal dendrites of pyramidal cells, and expression was also noted in astrocytes, suggesting that *SLC6A1* is involved in the reuptake of GABA from axonal terminals, as well as extrasynaptic uptake by astrocytes. In contrast, *SLC6A11* is mostly localized to astrocytes and *SLC6A13* is predominantly expressed in the leptomeninges, in neurons, and glia [12].

Pharmacological blockade of GABA transporters, especially *SLC6A1*, is an important therapeutic mechanism, which presumably acts to increase extracellular GABA levels. The first GABA reuptake inhibitor to become available for clinical use was tiagabine [(–)-(R)-1-[4, 4-Bis(3-methyl-2-thienyl)3-butenyl] nipecotic acid hydrochloride], which selectively blocks *SLC6A1* sites [13]. Tiagabine was developed for the treatment of seizure disorders but is being used to treat a variety of psychiatric conditions. Preliminary studies suggest that tiagabine is effective in the treatment of anxiety [14], sleep disorders [15], addiction [16], and depression [17]. Mice lacking *SLC6A1* have decreased aggression, anxiety, and depression-like behavior [18,19] as well as altered sensitivity to ethanol [20], further illustrating the potential for *SLC6A1* inhibition to treat a variety of psychiatric disorders.

The *SLC6A1* gene resides on chromosome 3p25-p24, spans 46.5 kb, and includes 16 exons (Fig. 1). This gene encodes a protein of 599 amino acids with a molecular weight of 67 kDa. The March 2006 genome build shows two transcripts for *SLC6A1*, the first transcript contains

all 16 exons and the other is coded by exons 2–16. All 16 exons and the predicted promoter regions of the *SLC6A1* gene were resequenced in our earlier study [21]. No nonsynonymous SNPs were found but we found a 21-bp insertion polymorphism in the predicted promoter region upstream of exon 1 that creates a second tandem copy of the sequence and therefore creates a variable number of tandem repeats (VNTR) polymorphism. We will refer to this sequence that is present in one or two copies as GAT1-21 (GGGTGGGGAGAGGGAGGGAGG).

Here we examine the molecular consequences of this VNTR polymorphism in *SLC6A1*. We show that the insertion polymorphism significantly increases promoter activity in part through the creation of a functional enhancer element, and identify the transcription factor ZNF148 as an important positive regulator of transcriptional activity. Expression analysis in postmortem brain samples determined that *SLC6A1* genotype significantly predicts *SLC6A1* expression in hippocampus. We provide evidence that the insertion allele is likely derived from Africa and is unique to individuals in our sample with African ancestry. These results identify a genetic variant that may have important implications for therapeutic response to inhibitors of *SLC6A1* as well as GABAergic function in individuals with African ancestry.

Materials and methods

DNA samples

Human DNA samples were obtained in full compliance with Yale and NIH Human Investigation Committee regulations.

Cell culture

All cell lines were obtained from American Type Culture Collection (ATCC; Manassas, Vermont, USA). Mouse embryonic carcinoma cells (P19) and human embryonic kidney 293 cells (HEK-293) were cultured in Dulbecco's modified Eagle medium (GIBCO invitrogen cell culture, Carlsbad, California, USA). Media were supplemented with 10% fetal bovine serum, 2 U/ml penicillin, 2 µg/ml streptomycin, and 2mmol/l L-glutamine (GIBCO invitrogen cell culture). Human neuroblastoma cells [SK-NBE(2)] were cultured in a 1: 1 mixture of Eagle's minimum essential medium and F-12K media (ATCC) supplemented with 10% fetal bovine serum, 2 U/ml penicillin, and 2 µg/ml streptomycin. All cells were grown in a humidified incubator at 37°C and 5% CO₂.

Electromobility shift assay

Nuclear protein extracts from P19, SK-N-BE(2), and HEK-293 were prepared using the NE-PER Nuclear and Cytoplasmic Extraction kit (PIERCE, Rockford, Illinois, USA) and quantified by BCA protein assay kit (PIERCE). Two sets of double-stranded DNA probes were constructed coding for one copy of GAT1-21. One set of probes was labeled with LI-COR IRDye 700-red, the other with LI-COR IRDye 800-green phosphoramidite (LI-COR Bioscience, Lincoln, Nebraska, USA). The IRDye 800 competitor probe was used in a manner analogous to a nonradioactive competitor probe in radioisotope- based electromobility shift assay (EMSA) assays. For the EMSA binding reactions, a simple competition of the two probes was performed where the ratio of probe to competitor was varied from 1: 1, 1: 2, 1: 5, and 1: 10. Probes were mixed with 1x binding buffer [2.5mmol/l dithiothreitol (DTT)/0.25% Tween-20], 1 µg of Poly (dI-dC) (LI-COR bioscience) and mixed with nuclear lysates. Binding reactions were carried out for 30 min at room temperature with end-over-end mixing and loaded on 4% native acrylamide gels containing 0.38 mol/l of glycine. Gel imaging was carried out using an Odyssey Imaging System (LI-COR Bioscience) at 700 and 800nm wavelengths.

Construction of reporter plasmids

Genomic DNA from samples that were determined to be homozygous for the *SLC6A1* insertion or noninsertion alleles with otherwise identical sequences were identified by DNA sequencing. Genomic DNA was amplified by nested PCR. The first PCR used the primers 5'-CTGGGCTGGAGAGAAGGAATCTTTT-3' and 5'-ATGCAACTCTCGCCTCTGTTCCAG-3' to yield a DNA fragment of 1.52 kb containing the 5' UTR and exon 1 of the *SLC6A1* gene. PCR reactions were carried out in 50 µl volumes containing 50 ng genomic DNA, 10mmol/l of each dNTP (New England BioLabs, Beverly, Massachusetts, USA), 2 µmol/l of primer, 1x reaction buffer, 1 µl of PfuUltra II Fusion HS DNA Polymerase (Stratagene, La Jolla, California, USA) and 10% sulfolane. PCR products were gel purified and amplified in a second PCR reaction using primers 5'-CTGAGTTCCTGGGGACCCAGAGGGAAGG-3' and 5'-CGAGCGGCGCCTCTGCTCCTTCATGTGG-3'. Each 50 µl of reaction contained 1 µl of purified PCR product, 10 mmol/l of each dNTP (New England BioLabs), 2 µmol/l of each primer, 1x reaction buffer, 1 µl of PfuUltra II Fusion HS DNA Polymerase (Stratagene), and 10% dimethyl sulfoxide. PCR products were gel purified and subcloned into the pSTBlue-1 vector (EMD Biosciences, Darmstadt, Germany). Recombinant plasmid clones were sequenced in both directions for verification. Desired clones with and without the 21 bp insertion with otherwise identical sequences and sequences identical to the Genbank reference sequence were selected, digested with *NheI* and *MluI* (New England BioLabs), and subcloned into the pGL3-basic reporter vector (Promega, Madison, Wisconsin, USA) to create pGL3-*SLC6A1*-Ins (two copies of GAT1-21) and pGL3-*SLC6A1*-Del (one copy of GAT1-21).

Transcription factor constructs

The transcription factors SP1, MAZ1, and ZNF148 were amplified from mouse cDNA using Pfu Ultra II polymerase (Stratagene), cloned in to pSTBlue-1 (Novagen), and subjected to DNA sequencing. SP1 primers included 5'-CGAATTCGCCACCATGAGCGACCAAGATCACTCCATGGA-3' and 5'-CAAGCTTTCACCTTGTCGTCATCGTCTTTGTAGTCGAAGCCATTGCCACTGATATT AATGGACT-3'. ZNF148 primers included 5'-CTCTAGAGCCACCATGAACATTGACGACAAACTGGAAGGATTG-3' and 5'-CAAGCTTTCACCTTGTCGTCATCGTCTTTGTAGTCGCCAAAAGTCTGGCCAGTTGT GGC-3'. MAZ1 primers included 5'-CGAATTCGCCACCATGTTCCCGGTGTTTCCTTGACACGC-3' and 5'-CAAGCTTTCACCTTGTCGTCATCGTCTTTGTAGTCCCAGGGTTGGGAGGGAAGTG GC-3'. Inserts were then subcloned into the pcDNA3.1-Zeo vector (Invitrogen) for overexpression studies. Data were analyzed using SYSTAT version 12 by one-way analysis of variance (ANOVA).

Transient transfection and dual-luciferase assay

P19 and HEK-293 cells were transiently transfected with pGL3-*SLC6A1*-Ins, pGL3-*SLC6A1*-Del, or the promoterless pGL3-basic using Fugene 6 transfection reagent (Roche, Basel, Switzerland). SK-N-BE(2) cells were transfected by calcium phosphate. Cells were treated with 125 mmol/l CaCl₂ in 2 × (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)-buffered saline buffer (500mmol/l (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 100 mmol/l Na₂HPO₄, 1 mol/l NaCl, pH=7.10) for 6 h after 15% glycerol shock for 40 min. Transfections were optimized to yield greater than 60% efficiency as determined by lacZ plasmid transfection and X-gal histochemistry. pRL-SV40 plasmid (Promega) coding for Renilla luciferase was included in all transfection reactions to allow normalization of luciferase activity for differences in transfection efficiencies. Forty-eight hours after transfection cells were harvested and luciferase activity was measured by the Dual-Luciferase Reporter Assay

System (Promega) on a DTX 880 Series Multimode Detector (Beckman Coulter, Fullerton, California, USA). Firefly luciferase luminescence values were divided by Renilla luciferase luminescence values from the same transfection to control for differences in transfection efficiency. Of note, one-way ANOVA revealed no significant differences in Renilla luminescence for these experiments. Band intensities were determined using NIH ImageJ software and analyzed using SYSTAT version 12 by one-way ANOVA.

Overexpression assays

Human SK-N-BE(2) cells (ATCC) were plated into 24-well polystyrene plates (BD Falcon) coated with polyethyleneimine (800 kD; Sigma) at a density of 2×10^5 cells per well. The next day cells were transfected with 200 ng per well of cDNA coding for ZNF148 in the CMV-driven expression vector pCMV-MCS (Stratagene), or with pCMV-MCS as a control. After 48 h cells were harvested and total RNA was prepared using the RNeasy kit (Qiagen). For each well, 0.5 μ g of RNA was reverse transcribed using the Superscript II kit and dT primer (Invitrogen) and subsequently treated with 0.5 μ l of RNase H (Invitrogen) for 30 min. cDNA was subjected to real-time quantitative PCR (qPCR). TATA-binding protein (TBP) was utilized for the purpose of normalization and was detected with the primers TBP 5'-TCAGCAGTCAACGTCCCAGCAGGCA-3' and TBPR 5'-GCTGCGGTACAATCCCAGAACTCTCCGA-3'. Human *SLC6A1* was detected with the primers S1F 5'-TGCCGTGGAAACAGTGCGACAACCC-3' and S1R 5'-TGCATGTTGCGCTCCCAGAACTCCACC-3' by cycling for 10 min at 95°C followed by 40 cycles of 95°C for 20 s, 72°C for 20 s, and 80°C for 10 s. Fluorescence was acquired at the 80°C step. Amplification efficiencies were determined for each primer set using the dilution series method. Data were analyzed by calculating ratios of *SLC6A1* expression relative to TBP expression as $[(\text{Efficiency } SLC6A1)^{-CT(SLC6A1)} / (\text{Efficiency TBP})^{-CT(TBP)}]$ where the efficiencies (ranging from 1 to 2) were calculated by the dilution series method and CT(i) refers to the observed threshold crossing for gene i. Data were analyzed using SYSTAT version 12 using one-way ANOVA.

DNA affinity pull-down assay

DNA templates that contained zero, one, or two copies of GAT1-21 were designed for pull-down assays. The noninsertion (one copy) and insertion (two copy) templates were generated by PCR from pGL3-*SLC6A1*-Del and pGL3-*SLC6A1*-Ins, respectively. The zero-copy template was generated by deletion mutagenesis using overlap extension PCR [22] using four primers, 5'-TGTAACGACGGCCAGTGGCAGACAGGCTGGTGACCCAGGATGA-3', 5'-TCTCTTCCCTCCCTCGCCTGCCCCGCCGT-3', 5'-GCAGGCGAGGGAGGGAGGAAGAGA-3', and 5'-CTTCTTTCCTCTCGCATTC-3'. PCR products were labeled at the 5' ends by amplification with a dual biotin containing M13 oligonucleotide sequence [23]. Biotinylated DNA was incubated with 1x-binding buffer (10mmol/l Tris, 50 mmol/l KCl, 1 mmol/l DTT pH=7.5) and nuclear protein lysates at room temperature for 30min with end-over-end mixing. DNA-protein complexes were mixed with Dynabeads MyOne Streptavidin beads (DynaL Biotech ASA, Smestad Oslo, Norway) and incubated for 30 min with continuous mixing. Complexes were purified using a microcentrifuge tube magnet (DynaL). Supernatants containing unbound material were removed using aspiration and beads were washed three times with buffer containing 20mmol/l Tris, 100 mmol/l KCl, 2mmol/l DTT, pH=7.5. Complexes were eluted by adding 6 $\times$ SDS sample buffer and heating at 95°C for 5 min. One-dimensional electrophoresis was carried out using 10% SDS-polyacrylamide gel electrophoresis gels run at 10 mA/gel at room temperature. Gels were stained by coomassie blue (GelCode Blue Stain reagent, PIERCE) and destained overnight with ultra pure water, or were stained using silver staining (SilverSNAP Stain kit II, PIERCE). Bands were cut and were subjected to liquid chromatography followed by mass spectroscopy (Yale/NIDA Proteomic Center, Yale University, New Haven, Connecticut,

USA). Gel fragments were trypsinized and proteins were analyzed on a Waters Q-ToF ABI mass spectrometer. All MS/MS spectra were searched using the automated Mascot algorithm against the NCBI nr database. Criteria for positive protein identification were: two or more MS/MS spectra matched the same protein entry in the database and the matched peptides were derived from the type of enzymatic digestion performed on the protein. We report only proteins that fulfilled these criteria and where at least 10 peptides matched the protein. Proteins were annotated using STRING (Search Tool for the Retrieval of Interacting Proteins; <http://string.embl.de/>) and Bioinformatics Harvester (<http://harvester.embl.de/>). Liquid chromatography/tandem mass spectrometry was conducted by the Yale Proteomics Center.

Enhancer constructs

The GAT1-21 sequence (GGGTGGGGAGAGGGAGGGAGG) was cloned upstream of the SV40 promoter in the pGL3-Promoter vector (Promega) in the forward direction in one, two, or four tandem copies. As a control, two tandem copies of a randomly generated sequence (TTATGACGTTATTCTACTTTG) with no homology to any REFSEQ promoter sequence was inserted upstream of the same promoter. Constructs were verified by DNA sequencing. For these experiments we noted that the resulting constructs led to increased Renilla activity. This is a well-known difficulty with the dual luciferase method, and is thought to be because of the ability of very strong promoters to affect other promoters in *trans*. To prevent any spurious results from regulation of Renilla expression, we normalized our data for these experiments within but not between constructs, with mean Renilla values normalized to control (pGL3 promoter) values. To verify that this procedure did not bias our results, we compared our results to non-normalized values (i.e. absolute firefly luciferase luminescence values) and observed precisely the same pattern that we report here. Two independent sets of experiments were conducted. Data were analyzed using SYSTAT version 12 using one-way or two-way ANOVA with Tukey honestly significant difference (HSD) post-hoc comparisons.

Human postmortem expression analysis

Postmortem brains were collected at the Clinical Brain Disorders Branch, National Institute of Mental Health with informed consent from the legal next of kin under National Institute of Mental Health protocol 90-M-0142 and processed as described previously [24]. In brief, hippocampus and dorsolateral prefrontal cortex (DLPFC) were dissected. Diagnoses were determined by independent reviews of clinical records and family interviews by two psychiatrists using Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria. Toxicological analysis was conducted on every patient. Normal controls had no known history of psychiatric symptoms or substance abuse, and negative toxicological results. Positive toxicology was not an exclusion criterion for cases with schizophrenia.

For this study, postmortem brain tissue exclusively from African-American individuals between the ages of 18 and 74 years of age (mean=43.21 years) was used because the *SLC6A1* insertion polymorphism is found only in individuals of African descent. The sample included 43 males and 30 females and 69 individuals were successfully genotyped. Of these 73 patients, 49 were healthy controls and 24 had a diagnosis of schizophrenia (14 males and 10 females). RNA from hippocampus and DLPFC was reverse transcribed and subjected to real-time qPCR.

We quantified mRNA expression levels of *SLC6A1* transcripts by real-time qPCR, using an ABI Prism 7900 sequence detection system (Applied Biosystems). The assay Hs01104469_m1, which contained the primers and a probe, was purchased from Applied Biosystems. Each 10 µl of reaction contained 500 nl of the assay, 5 µl of Taqman Universal PCR Mastermix (Applied Biosystems) containing Hot Goldstar DNA Polymerase, dNTPs with dUTP, uracil-*N*-glycosylase, passive reference, and 100 ng of cDNA template. PCR cycle

parameters were 50°C for 2min, 95°C for 10 min, 40 cycles of 95°C for 15s, and 60°C for 1min. PCR data were acquired from the Sequence Detector Software (SDS version 2.0, Applied Biosystems) and quantified by a standard curve method with serial dilutions of pooled cDNA derived from commercially available human brain reference RNA. In each experiment, the R^2 value of the curve was more than 0.99, and controls comprising no-template cDNA resulted in no detectable signal. All samples were measured in a single plate for each transcript, and cycle number at threshold values were in the linear range of the standard curve. All measurements were performed in triplicate, and the gene expression level was calculated as an average of the triplicates. Expression was normalized to the geometric mean of β -actin, β -2- microglobulin, and β -glucuronidase. *SLC6A1* genotypes were determined as detailed above except that PCR products were sized on an Agilent Bioanalyzer and PCR products were also subjected to direct sequencing to confirm the genotypes. Genotypes were scored as 0 (only noninsertion alleles present) or 1 (at least one insertion allele present). This coding scheme was used in part because only one patient was homozygous for the insertion allele in our sample. Data were analyzed using a multiple regression model with SAS PROC GLM version 9.1.3. This model included diagnosis, age at death, *SLC6A1* genotype, sex, and their interactions, as well as an index of RNA quality (Agilent RIN index). Interaction effects were interpreted by calculating a categorical age at death variable as a median split of the age at death variable. Means and standard errors for graphical analysis were calculated with SAS GLM using model-based least squares means and standard errors.

Results

SLC6A1 promoter activity

To evaluate the functional significance of the copy number of GAT1-21 on *SLC6A1* promoter activity, promoter variants were used to drive luciferase expression in reporter assays in three cell lines (Fig. 2). A two-way ANOVA revealed a main effect of construct [promoterless control vector vs. one-copy (noninsertion) promoter containing vector vs. two-copy (insertion) promoter containing vector; $F(2,34)=41.99$, $P<0.001$]. No significant effect of cell line [$F(2,34)=2.33$, $P>0.05$] was present and there was no cell line $\times$ construct interaction effect [$F(4,34)=0.79$, $P>0.05$]. Post-hoc analyses revealed that the two-copy (insertion) and one-copy (noninsertion) variants showed significantly more activity than the promoterless control (pGL3-basic) vector ($P<0.05$ and $P<0.001$ by Tukey HSD test, respectively) and the two-copy (insertion) variant showed significantly higher activity than the one-copy (noninsertion) variant ($P<0.001$ by Tukey HSD test). On average, the two-copy (insertion) promoter generated about twofold more luciferase expression than the one-copy (noninsertion) promoter across the four replications tested. These data suggest the 1.4-kb region upstream from exon 1 has promoter activity and that the 21-bp insertion polymorphism that creates two copies of the GAT1-21 sequence significantly increases the activity of the *SLC6A1* promoter.

Electromobility shift assay study

Because the 21-bp insertion polymorphism increased the activity of the *SLC6A1* promoter, we hypothesized that the insertion sequence may code for a protein binding motif that interacts with a nuclear protein, such as a transcription factor, to enhance transcription, and that two copies of GAT1-21 may bind such a protein more effectively than a single copy of GAT1-21. We generated a double stranded DNA probe coding for GAT1-21 labeled with one infrared fluorophore (IRDye 700) and interacted this probe with nuclear proteins from HEK-293 and SKN- BE(2) cells in gel shift assays. To establish specificity we used a competitor probe with the same sequence labeled with an infrared dye with distinct spectral properties (IRDye 800) and we used the IRDye 800- labeled probe in a manner analogous to a 'cold' competitor in a radioactive EMSA assay. As shown in Fig. 3a, nuclear proteins from both cell lines tested led to retardation of the migration of the probe, and this interaction could be reduced by competition

with a competitor double stranded oligonucleotide. These results suggest the presence of nuclear proteins that specifically bind to GAT1-21.

Pull-down and transactivation studies

We used two approaches to identify proteins that interact with the insertion sequence region that may regulate promoter activity. First, we attempted to identify proteins that directly interact with the promoter using pull-down assays. Biotinylated double-stranded DNA probes coding for the *SLC6A1* promoter with zero, one, or two copies of GAT1-21 were interacted with nuclear lysates from human neuroblastoma cells [SK-N-BE(2)], captured with streptavidin magnetic beads, separated by SDS–polyacrylamide gel electrophoresis, and coomassie blue or silver stained. As a control, a streptavidin-only condition was included. Protein bands that were specific to the presence of DNA probe were identified, excised, and proteins identified by liquid chromatography/tandem mass spectrometry. No protein bands were identified that were specifically associated with the two-copy (insertion) version. The proteins that did bind to the promoter probes were therefore analyzed together. Pathway analysis and annotation using STRING (<http://string.embl.de/>) revealed that many of the proteins identified undergo known protein–protein interactions (Supplementary Table 1). These data define a protein complex that interacts with the *SLC6A1* promoter that may shed light on the transcriptional efficacy of the promoter. These DNA–protein interactions, however, cannot account for the differences in promoter activity between the insertion and noninsertion variants.

The direct biochemical approach that we used is likely to be limited in sensitivity for low-abundance proteins such as transcription factors. We therefore used a bioinformatics approach to predict transcription factors that can bind within the vicinity of the polymorphism using the TESS program (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>). This analysis suggested the presence of potential MAZ1, ZNF148 (ZBP-89), and SP1 transcription factor binding sites (Fig. 3b). The mouse, human, and monkey *SLC6A1* genes are very homologous. In all these species the gene has 16 exons and a large intron (> 20 kb) is found between exons 1 and 2, but this intron is not present in the current genome build for rat. A single copy of a sequence homologous to GAT1-21 is present in the human, monkey, mouse, and rat reference sequences. The first SP1 and MAZ sites are conserved in all five species and the second SP1 and ZNF148 sites are conserved only in primates.

We cloned these transcription factors, created overexpression constructs, and tested each construct for its ability to transactivate the *SLC6A1* promoter. Our goal with these studies was to determine whether the copy number of GAT1-21 influences sensitivity to the effects of these transcription factors, and we therefore normalized the data to the luminescence resulting from the same construct without cotransfected transcription factor. The luminescence values for both constructs therefore begin at 100% of control, although as we have seen the two-copy (insertion) variant has more activity than the single-copy (noninsertion) variant. As shown in Fig. 3c, these transcription factors differed significantly in their ability to modulate *SLC6A1* function [$F(3,8)=1091.37$, $P<0.001$ for overall effect of transcription factor]. ZNF148 dramatically increased the activity of the *SLC6A1* promoter ($P<0.001$ by Tukey HSD Post-Hoc test) relative to transfection with a control vector (pBluescript, Stratagene) whereas MAZ1 and SP1 decreased promoter activity by a small but significant amount (both $P<0.001$ by Tukey HSD Post-Hoc tests).

We next sought to determine whether ZNF148 regulates cellular *SLC6A1* expression (Fig. 3d). We transfected human SK-N-BE(2) neuroblastoma cells with ZNF148 in a CMV-driven expression vector or the expression vector alone as a control ($N=6$ wells per condition). After 48 h RNA was purified and *SLC6A1* RNA expression was determined by real-time qPCR and was normalized to the housekeeping gene TBP. ZNF148 overexpression did not significantly regulate the housekeeping gene TBP ($P=0.599$ by t -test) but did significantly upregulate

SLC6A1 expression (mean SLC6A1 expression relative to TBP 0.853 ± 0.064 for ZNF148 expressing cells, 0.595 ± 0.047 for controls; $P=0.009$ by *t*-test). These results suggest that ZNF148 regulates endogenous SLC6A1 expression in a manner that is similar to its effects on our SLC6A1 promoter reporter construct.

ZNF148 is a transcription factor that can inhibit or enhance promoter activity depending on the promoter involved. ZNF148 is thought to complex with SP1 to inhibit the activity of certain promoters such as that of vimentin [25,26] and inhibits SP1 activation of the ornithine decarboxylase promoter [27]. ZNF148, however, enhances expression of STAT1 [28] and activates transcription of intestinal alkaline phosphatase [29]. Of most importance here, these transcription factors modulated activity of the noninsertion and insertion promoter variants to an equal extent. Taken together with the pulldown results, we find no evidence that differential protein interactions determine the differences in promoter activity between the insertion and noninsertion *SLC6A1* variants although we cannot exclude this possibility.

Intrinsic enhancer function

We next sought to determine whether copy number of GAT1-21 affects promoter activity by acting as an intrapromoter enhancer element. We inserted zero, one, two, or four copies of the sequence in the forward direction in front of an SV40 promoter, which in turn drives expression of luciferase, and as an additional control we inserted two copies of a random sequence in front of the same promoter (Fig. 4a). Relative to the control (SV40 promoter alone), the two-copy and the four-copy constructs showed significantly more promoter activity than the SV40 promoter alone ($P<0.05$ by Tukey HSD test), but the single-copy and random-sequence constructs did not (Fig. 4b). The four-copy construct showed significantly more activity than all of the other constructs ($P<0.05$ by Tukey HSD test). The two-copy construct showed a small (but statistically significant) amount of additional activity than the single-copy construct ($P<0.05$ by Tukey HSD test). The construct containing two copies of a random sequence differed significantly from the two and four-copy GAT1-21 sequence constructs but not from the single-copy and control constructs. Overall, these data suggest that two copies of GAT1-21 may increase promoter activity in part through the creation of a functional enhancer element.

To determine whether the insertion region is present elsewhere in the human genome we performed a search using BLAT (<http://genome.ucsc.edu>) for the insertion sequence. As shown in Table 1 a total of five genomic loci in addition to SLC6A1 show at least 95% identity with GAT1-21. Whether this sequence regulates gene expression in these other genomic contexts will require further research to determine.

Postmortem expression analysis

To determine whether the GAT1-21 polymorphism can influence SLC6A1 expression in human brain we conducted postmortem expression analysis and genotyping of a cohort of 43 African-American males and 30 African-American females. Of these 73 patients, 49 were healthy controls and 24 had a diagnosis of schizophrenia (14 males and 10 females) and genotypes were successfully obtained from 69 patients. RNA from hippocampus and prefrontal cortex was analyzed by real-time qPCR and DNA samples were genotyped for GAT1-21. Genotype frequencies were 72.46% noninsertion/noninsertion, 26.09% noninsertion/insertion, and 1.45% insertion/insertion, implying an allele frequency of 14.49% for the insertion allele, similar to what we have reported elsewhere (see below). These genotype frequencies closely match Hardy-Weinberg expectations (Pearson's $\chi^2=0.19$, $P=0.6625$).

A multiple regression model including diagnosis, age at death, *SLC6A1* genotype, sex, and their interactions, as well as an index of RNA quality (Agilent RIN index) was fit to the data. For the DLPFC samples, only RNA integrity reached statistical significance [$F(1,42)=6.31$,

$P=0.016$] and the DLPFC samples were therefore dropped from further analysis. In contrast, the RNA integrity term was not significant for the hippocampal samples [$F(1,45)=2.10$, $P=0.1547$]. Significant main effects of age [$F(1,45)=15.36$, $P=0.0003$] and genotype [$F(1,45)=7.17$, $P=0.0103$] were seen, as well as age $\times$ genotype [$F(1,46)=7.09$, $P=0.0107$], genotype $\times$ sex [$F(1,46)=7.75$, $P=0.0078$], and age $\times$ genotype $\times$ sex [$F(1,46)=9.37$, $P=0.0037$] interaction effects. Diagnosis and interaction effects with diagnosis were nonsignificant (all $P>0.05$).

We sought to interpret the interaction effects by dichotomizing age into values below ('young') and above ('older') the median age of our sample (44 years) and we calculated means and standard errors within the context of our regression model using least squares means and standard errors. As shown in Fig. 5, older patients showed lower SLC6A1 expression than younger patients. Younger male patients expressed more SLC6A1 than younger female patients, but this trend was weaker and in the opposite direction for older patients. Finally, patients with the insertion allele generally showed higher SLC6A1 expression than patients without this allele, consistent with the results of our promoter studies. A weak trend in the opposite direction was noted for older female patients. A simple correlation analysis (Fig. 5b) also revealed that SLC6A1 expression markedly declines with age ($r = -0.5187$, $P<0.0001$). It is therefore possible that the observed age $\times$ genotype interaction effect is due in part to an age-related decline in SLC6A1 expression because of aging effects with a concomitant decline in influence of genetic factors. These data also suggest that sex affects SLC6A1 expression. SLC6A1 is apparently under the influence of a variety of genetic and nongenetic factors.

Population differences

We previously genotyped the *SLC6A1* VNTR polymorphism (Fig. 1) in 46 European–American, 60 African–American, 59 Thai, 47 Finnish, and 46 Hmong individuals [21] and found this polymorphism only in the African–American population with an allele frequency of 39%. These data suggest but do not establish that the insertion allele has its origins in Africa. To provide additional evidence for this hypothesis we genotyped this polymorphism in 69 Tanzanian individuals in this study. In the Tanzanian cohort, we obtained genotype frequencies of 2.9, 30.4, and 66.7% for the 2/2, 1/2, and 1/1 copy genotypes, implying an allele frequency of 18.1%. These genotype frequencies do not deviate significantly from Hardy–Weinberg equilibrium expectations ($\chi^2=1.44$, $df=2$, $P>0.05$). We have not observed the two-copy allele in hundreds of additional genotypes that we have recently conducted with individuals without African ancestry. The presence of the insertion allele in individuals of African descent but not other populations is consistent with this allele having arisen in Africa but not having been carried with the major migrations from Africa that founded the rest of the world's populations.

Discussion

In experiments involving mouse embryonic carcinoma cells (P19), human neuroblastoma cells [SK-N-BE(2)], and human embryonic kidney (HEK-293) cells, luciferase reporter assays showed that the 1.4-kb fragment upstream from exon 1 is in fact a functional promoter for *SLC6A1*. EMSA studies showed that nuclear proteins from mouse as well as human cell lines interact with GAT1-21. It is interesting to note that *SLC6A1* has two promoter elements, one upstream of exon 1 and the other upstream of exon 2. An increase in the efficacy of the most 5'-promoter element that we studied here because of two copies of GAT1-21 may serve to bias the ratio of upstream to downstream promoter transcripts generated. Exon array analysis of human brain tissue suggests that the predicted first and second SLC6A1 transcripts are in fact expressed, and publically available transcriptome sequencing data support the existence of SLC6A1 transcripts originating from the first and second exons of the gene as shown in Fig.

1 (E. George, unpublished observations). Ascertaining definitively whether such a shift occurs and the functional implications of such a shift will require further research.

We initially sought to identify protein factors that may be involved in determining the increased efficacy of the insertion variant promoter relative to the noninsertion variant. Bioinformatics analysis predicted that MAZ1, ZNF148, and SP1 interact with the insertion region, but only ZNF148 was capable of activating the *SLC6A1* promoter. MAZ1 and SP1 were found to have small but statistically significant inhibitory effects on promoter function. The one (noninsertion) and two-copy (insertion) promoters showed similar degrees of enhancement by ZNF148 and similar degrees of inhibition by MAZ1 and SP1. Similarly, pull-down assays failed to identify any differential protein interactions with the noninsertion and insertion variants. This led us to suspect that differential transcription factor affinities were not responsible for the differences in promoter activity between the insertion and noninsertion promoter variants although we cannot rule out this possibility. We therefore explored the possibility that the insertion sequence acts as an intrapromoter enhancer element. When put in front of an SV40 promoter driving luciferase expression, increasing GAT1-21 copy number exponentially increased luciferase expression. The enhancement seen in these experiments was, however, less than the two-fold difference between the one and two-copy GAT1-21 promoters that we observed. A number of possible reasons for these differences are present. First, some degree of enhancer–promoter interaction is likely to occur, such that the insertion sequence enhances particular promoters to a greater or lesser extent. Second, it is possible that additional sequences, perhaps flanking the insertion region, are required for full enhancer efficacy.

Our human postmortem brain expression study was consistent with our reporter assays with regard to the notion that the insertion allele increases expression of *SLC6A1*. The postmortem data also revealed that *SLC6A1* expression is subject to a number of influences other than genotype, including sex and age, which seem to interact with genotype to determine expression levels. Our data revealed a marked decline in *SLC6A1* expression with advancing age. Younger patients showed larger sex and genotype differences than older patients as indicated by a significant sex $\times$ age $\times$ genotype interaction effect. These complex interaction effects can perhaps be understood in light of a model where *SLC6A1* expression is increasingly under the influence of factors related to brain aging rather than genetic factors as individual ages. We observed significant association between GAT1-21 copy number and *SLC6A1* mRNA expression in hippocampus but not DLPFC. We hypothesize that a greater proportion of the variance in the expression of this gene is subject to non-genetic influence in DLPFC than in hippocampus, but this will require further research to establish.

The Recent African Origin or Out-of-Africa model holds that all non-African populations descended from an anatomically modern *Homo sapiens* ancestor that evolved in Africa and then left Africa to inhabit the rest of the world. Although only a small proportion of the genetic variation present in Africa left with these migrations, it is not uncommon to find genetic variations that are unique to individuals of African descent [30]. Recent data from mitochondrial genome sequencing shows that Tanzanians have very high genetic diversity, greater than other African populations, suggesting the possibility that Tanzanians represent the modern descendants of this original source population [31]. Our data are consistent with a model where the 21-bp insertion variant arose before the beginning of the migrations from Africa. The most likely explanation for the lack of this allele in non-African populations is that the allele never left Africa during migrations from that continent. Other models are, however, also possible including genetic drift and natural selection.

These findings have important implications for future pharmacogenetic studies, as this genetic variation may affect clinical response to drugs that bind this gene's protein product in

populations with African ancestry and may be a useful predictor of dosage requirements. Tiagabine is an effective treatment in many psychiatric disorders. Tiagabine has a rare but potentially serious adverse side effect, nonconvulsive status epilepticus, which has raised concerns about its off-label use in the treatment of psychiatric disorders [32]. Our results suggest that patients with the insertion variant may require a higher dose of tiagabine than patients without the insertion variant to achieve the same level of clinical response for a given plasma level. Knowledge of a patient's *SLC6A1* genotype may prove clinically useful in predicting an appropriate dosing regimen for patients with African ancestry and may predict GABAergic function in this population. This may have important implications for risk for conditions such as seizure disorders, drug, and alcohol addiction, as well as mood disorders. Additional research will be required to fully appreciate the clinical significance of the *SLC6A1* VNTR polymorphism. Our data also point to the role of aging and to some extent sex in determining *SLC6A1* expression. Older patients seem to express less *SLC6A1* than younger patients, although this will require confirmation on a protein level. This finding is consistent with a recent postmortem study showing that aging is associated with a loss of [³H]-tiagabine binding in frontal cortex [33]. Lower levels of *SLC6A1* in older patients may imply that lower dosages of drugs such as tiagabine are required to block a certain fraction of transporters. In contrast, lower expression in older patients may suggest that blocking *SLC6A1* will impact GABA levels less than in younger patients because of the lower expression level, which would imply that tiagabine may show lower maximal efficacy in older patients. It is possible that decreased *SLC6A1* expression as consequence of aging is associated with a compensatory increase in the expression of other GABA transporters, but this will require additional research to determine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

R.H. was supported by NIH grant D43 TW06166 and a VA REAP Award. A.S. was supported by NIH grant 1KL2RR024138, NCR grant RR19895, a NARSAD Young Investigator Award, and a VA REAP Award. A.S. and A.N. were supported by NIDA grant DA018343. E.D.G. was supported by NIH grant T32-MH14276. E.L.G. and C.Y. were supported by grant TW006764 (PI: Grigorenko) from the Fogarty Program, NIH and HD048830 (PI: Pugh) from NICHD. J.G. was supported in part by a VA REAP Award. The authors would like to acknowledge the support of the Yale/NIH Neuroscience Microarray Center supported by NIH grant U24 NS051869. The authors thank Kathy Stone and Terence Wu from the Yale/NIDA Proteomics Center as well as Maya M. Davis. We thank Dr Thomas M. Hyde for his involvement in creating the brain collection and his neuroanatomical expertise, and Ms Samhita Kumar and Mr Ross Buerlein for their technical support.

References

1. Sanacora G, Mason GF, Rothman DL, Behar KL, Hyder F, Petroff OA, et al. Reduced cortical gamma-aminobutyric acid levels in depressed patients determined by proton magnetic resonance spectroscopy. *Arch Gen Psychiatry* 1999;56:1043–1047. [PubMed: 10565505]
2. Sanacora G, Mason GF, Rothman DL, Krystal JH. Increased occipital cortex GABA concentrations in depressed patients after therapy with selective serotonin reuptake inhibitors. *Am J Psychiatry* 2002;159:663–665. [PubMed: 11925309]
3. Petroff OA, Rothman DL, Behar KL, Mattson RH. Low brain GABA level is associated with poor seizure control. *Ann Neurol* 1996;40:908–911. [PubMed: 9007096]
4. Ke Y, Streeter CC, Nassar LE, Sarid-Segal O, Hennen J, Yurgelun-Todd DA, et al. Frontal lobe GABA levels in cocaine dependence: a two-dimensional, J-resolved magnetic resonance spectroscopy study. *Psychiatry Res* 2004;130:283–293. [PubMed: 15135161]

5. Krystal JH, Staley J, Mason G, Petrakis IL, Kaufman J, Harris RA, et al. Gamma-aminobutyric acid type A receptors and alcoholism: intoxication, dependence, vulnerability, and treatment. *Arch Gen Psychiatry* 2006;63:957–968. [PubMed: 16952998]
6. Twyman RE, Macdonald RL. Kinetic properties of the glycine receptor main- and sub-conductance states of mouse spinal cord neurones in culture. *J Physiol* 1991;435:303–331. [PubMed: 1722820]
7. Goodchild CS. GABA receptors and benzodiazepines. *Br J Anaesth* 1993;71:127–133. [PubMed: 8393687]
8. LaRoche SM, Helmers SL. The new antiepileptic drugs: clinical applications. *J AM Med Assoc* 2004;291:615–620.
9. Sofuoglu M, Kosten TR. Novel approaches to the treatment of cocaine addiction. *CNS Drugs* 2005;19:13–25. [PubMed: 15651902]
10. Enna SJ, McCarson KE. The role of GABA in the mediation and perception of pain. *Adv Pharmacol* 2006;54:1–27. [PubMed: 17175808]
11. Mager S, Naeve J, Quick M, Labarca C, Davidson N, Lester HA. Steady states, charge movements, and rates for a cloned GABA transporter expressed in *Xenopus* oocytes. *Neuron* 1993;10:177–188. [PubMed: 7679914]
12. Casini G, Rickman DW, Brecha NC. Expression of the gamma-aminobutyric acid (GABA) plasma membrane transporter-1 in monkey and human retina. *Invest Ophthalmol Vis Sci* 2006;47:1682–1690. [PubMed: 16565409]
13. Fink-Jensen A, Suzdak PD, Swedberg MD, Judge ME, Hansen L, Nielsen PG. The gamma-aminobutyric acid (GABA) uptake inhibitor, tiagabine, increases extracellular brain levels of GABA in awake rats. *Eur J Pharmacol* 1992;220:197–201. [PubMed: 1425991]
14. Rosenthal M. Tiagabine for the treatment of generalized anxiety disorder: a randomized, open-label, clinical trial with paroxetine as a positive control. *J Clin Psychiatry* 2003;64:1245–1249. [PubMed: 14658975]
15. Mathias S, Wetter TC, Steiger A, Lancel M. The GABA uptake inhibitor tiagabine promotes slow wave sleep in normal elderly subjects. *Neurobiol Aging* 2001;22:247–253. [PubMed: 11182474]
16. Sofuoglu M, Mouratidis M, Yoo S, Culligan K, Kosten T. Effects of tiagabine in combination with intravenous nicotine in overnight abstinent smokers. *Psychopharmacology (Berl)* 2005;181:504–510. [PubMed: 15983800]
17. Carpenter LL, Schecter JM, Tyrka AR, Mello AF, Mello MF, Haggarty R, et al. Open-label tiagabine monotherapy for major depressive disorder with anxiety. *J Clin Psychiatry* 2006;67:66–71. [PubMed: 16426090]
18. Liu GX, Cai GQ, Cai YQ, Sheng ZJ, Jiang J, Mei Z, et al. Reduced anxiety and depression-like behaviors in mice lacking GABA transporter subtype 1. *Neuropsychopharmacology* 2007;32:1531–1539. [PubMed: 17164814]
19. Liu GX, Liu S, Cai GQ, Sheng ZJ, Cai YQ, Jiang J, et al. Reduced aggression in mice lacking GABA transporter subtype 1. *J Neurosci Res* 2007;85:649–655. [PubMed: 17183588]
20. Cai YQ, Cai GQ, Liu GX, Cai Q, Shi JH, Shi J, et al. Mice with genetically altered GABA transporter subtype I (GAT1) expression show altered behavioral responses to ethanol. *J Neurosci Res* 2006;84:255–267. [PubMed: 16683252]
21. Hirunsatit R, Ilomaki R, Malison R, Rasanen P, Ilomaki E, Kranzler HR, et al. Sequence variation and linkage disequilibrium in the GABA transporter-1 gene (SLC6A1) in five populations: implications for pharmacogenetic research. *BMC Genet* 2007;8:71. [PubMed: 17941974]
22. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 1989;77:61–68. [PubMed: 2744488]
23. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 2000;18:233–234. [PubMed: 10657137]
24. Lipska BK, Deep-Soboslay A, Weickert CS, Hyde TM, Martin CE, Herman MM, et al. Critical factors in gene expression in postmortem human brain: focus on studies in schizophrenia. *Biol Psychiatry* 2006;60:650–658. [PubMed: 16997002]
25. Wiczorek E, Lin Z, Perkins EB, Law DJ, Merchant JL, Zehner ZE. The zinc finger repressor, ZBP-89, binds to the silencer element of the human vimentin gene and complexes with the transcriptional activator, Sp1. *J Biol Chem* 2000;275:12879–12888. [PubMed: 10777586]

26. Zhang X, Diab IH, Zehner ZE. ZBP-89 represses vimentin gene transcription by interacting with the transcriptional activator, Sp1. *Nucleic Acids Res* 2003;31:2900–2914. [PubMed: 12771217]
27. Law GL, Itoh H, Law DJ, Mize GJ, Merchant JL, Morris DR. Transcription factor ZBP-89 regulates the activity of the ornithine decarboxylase promoter. *J Biol Chem* 1998;273:19955–19964. [PubMed: 9685330]
28. Bai L, Merchant JL. Transcription factor ZBP-89 is required for STAT1 constitutive expression. *Nucleic Acids Res* 2003;31:7264–7270. [PubMed: 14654702]
29. Malo MS, Mozumder M, Zhang XB, Biswas S, Chen A, Bai LC, et al. Intestinal alkaline phosphatase gene expression is activated by ZBP-89. *Am J Physiol Gastrointest Liver Physiol* 2006;290:G737–G746. [PubMed: 16384873]
30. Tishkoff SA, Kidd KK. Implications of biogeography of human populations for ‘race’ and medicine. *Nat Genet* 2004;36:S21–S27. [PubMed: 15507999]
31. Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 2007;24:757–768. [PubMed: 17194802]
32. Flowers CM, Racoosin JA, Kortepeter C. Seizure activity and off-label use of tiagabine. *N Engl J Med* 2006;354:773–774. [PubMed: 16481654]
33. Sundman-Eriksson I, Allard P. Age-correlated decline in [3H] tiagabine binding to GAT-1 in human frontal cortex. *Aging Clin Exp Res* 2006;18:257–260. [PubMed: 16804373]

Fig. 1.

Diagram of the *SLC6A1* gene structure. Diagram of the human *SLC6A1* gene showing the location of GAT1-21 that is present in one or two copies that is responsible for the variable number of tandem repeats (VNTR) (hatched) 350 bp 5'upstream of exon 1 and the positions of all 16 exons (solid). Alternative exon usage generates transcripts that include exon 1 through 16 or exon 2 through 16. The starting positions of the two major starting transcripts, denoted T1 and T2, are shown. The expression of exons 1 and 2 was verified using publicly available whole genome exon expression data

(http://www.affymetrix.com/support/technical/sample_data/exon_array_data.affx). In addition, we verified that transcripts originating from exon 1 (T1) and from exon 2 (T2) are in fact expressed using publicly available transcriptome sequencing data (<http://dbtss.hgc.jp/index.html>). These data also support the existence of a transcript originating from within the first intron (not shown in the figure).

Fig. 2.

Reporter assays. Luciferase activity for the insertion and noninsertion promoter variants relative to a promoterless luciferase vector were determined in mouse embryonic carcinoma cells (P19), human embryonic kidney cells (HEK-293), and human neuroblastoma cells [SK-N-BE(2)]. Luciferase reporter constructs were cotransfected with a Renilla luciferase control vector and luciferase activity was expressed as a fraction of Renilla luciferase activity. Four replicates were performed in each cell line. Data were then expressed as activity relative to the promoterless control vector within each cell line. Data were analyzed by two-way analysis of variance with Tukey honestly significant difference post-hoc tests. No significant effect of cell line [$F(2,34)=2.33$, $P>0.05$] was observed and there was no cell line $\times$ construct interaction effect [$F(4,34)=0.79$, $P>0.05$]. Post-hoc analyses revealed that the one and two-copy variants showed significantly more activity than the promoterless control (pGL3-basic) construct ($P<0.05$ and $P<0.001$, respectively) and the two-copy variant showed significantly higher activity than the one-copy variant ($P<0.001$). Asterisks (* $P<0.05$, *** $P<0.001$) pertain to these posthoc tests of copy number. INS sequence, insertion sequence.

Fig. 3.

Gel shift and transactivation experiments. (a) Electromobility shift assay (EMSA) was used to determine whether nuclear proteins specifically interact with GATI-21. Experiments were carried out in human neuroblastoma cells [SK-N-BE(2)] as well as human embryonic kidney cells (HEK-293). A single copy of GATI-21 labeled with IRdye-700 was used as a probe and a single copy of GATI-21 labeled with IRdye-800 was used as a competitor. In the absence of nuclear lysate the probe showed no retardation (lane 1). In the absence of competitor probe, migration through the gel was retarded (lanes 2 and 7) for both cell lines. As the concentration of competitor probe was increased from a 1: 1 to a 1: 10 molar ratio (lanes 3 through 6 and lanes 8 through 11) increasing amounts of probe was displaced from nuclear proteins. The

experiment was repeated a total of three times with identical results. The bottom panel shows densitometry values from the EMSA gel. (b) Alignment of the region proximal to the human VNTR (boxed) in human, rhesus, chimp, rat, and mouse, annotated with the results of a bioinformatics search using the TESS program. Two predicted SP1-binding regions exist, designated here as SP1-1 and SP1-2 (bold). The SP1-2 site overlaps a predicted ZNF148 site (underlined). A predicted MAZ1 site was also detected (lowercase). Note that the SP1-1 and ZNF148 sites are conserved only in primates, but the predicted SP1-1 and MAZ1 sites are more highly conserved. (c) On the basis of the results of the bioinformatics analysis shown in Fig. 3b, we cloned ZNF148, SP1, and MAZ1 from mouse cDNA and overexpressed these proteins in a CMV-driven expression vector along with the noninsertion or insertion form of the *SLC6A1* promoter luciferase reporter in HEK-293 cells. Three replicates were performed. Cells were cotransfected with a Renilla luciferase control vector and luciferase activity was expressed as a fraction of Renilla luciferase activity. Data were then expressed as activity relative to cotransfection with a control DNA (pBluescript vector). Note that normalization was performed separately for the one and two-copy reporter constructs to highlight changes from baseline activity. Both therefore show 100% of control activity in the case of the pBluescript control although the two-copy variant is more active than the one-copy variant (see above). Data were analyzed by analysis of variance followed by post-hoc tests using Tukey honestly significant difference (HSD) tests. The transcription factors differed significantly in their ability to modulate *SLC6A1* function [$F(3,8)=1091.37$, $P<.001$]. ZNF148 increased the activity of the *SLC6A1* promoter ($P<0.001$) whereas MAZ1 and SP1 decreased promoter activity (both $P<0.001$). *** $P<0.001$. (d) To confirm that ZNF148 regulates endogenous *SLC6A1* expression, SK-N-BE(2) cells were seeded at 2×10^5 cells per well in 24 well plates and transfected with 200 ng of ZNF148 in a CMV-driven expression plasmid (pCMV-MCS) or 200 ng pCMV-MCS alone as a control ($N=6$ wells per condition). Cells were harvested 48 h after transfection and RNA was subjected to real time quantitative PCR. ZNF148 overexpression did not significantly regulate the housekeeping gene TATA-binding protein (TPB) ($P=0.599$ by t -test) but did significantly upregulate *SLC6A1* expression (mean *SLC6A1* expression relative to TBP 0.853 ± 0.064 for ZNF148 expressing cells, 0.595 ± 0.047 for controls; $P=0.009$ by t -test). ** $P<0.01$.

Fig. 4.

Enhancer studies. (a) Diagram illustrating the five reporter constructs tested to determine whether the copy number of GAT1-21 determines intrinsic enhancer function. A construct containing an SV40 promoter in front of a firefly luciferase gene was used as a control. One (F), two (FF), or four copies (FFFF) of the GAT1-21 sequence were inserted in the forward (5' to 3') orientation in front of the SV40 promoter. Another construct with two copies of a random sequence (XX) was used as an additional control. (b) Luciferase reporter constructs were cotransfected with a Renilla luciferase control vector. Three replicates were performed in the first experiment and an additional four replicates were performed in the second experiment. Since one-way ANOVA (analysis of variance) showed significant differences in Renilla activity, we sought to prevent any spurious results from regulation of Renilla expression and normalized our data for these experiments within but not between constructs, with mean

Renilla values normalized to control (pGL3 promoter) values. Data were then expressed as activity relative to the control reporter vector. Data were analyzed by ANOVA and post-hoc tests using Tukey HSD tests. Relative to the control (SV40 promoter alone), the two-copy (FF) and the four-copy (FFFF) constructs showed significantly more promoter activity than the SV40 promoter alone, but the single-copy (F) and random-sequence constructs (XX) did not. The four-copy construct (FFFF) showed significantly more activity than all of the other constructs. The two-copy construct (FF) showed a small but statistically significant ($P < 0.05$ by Tukey HSD test) amount of additional activity than the single-copy construct (F). The construct containing two copies of a random sequence (XX) differed significantly from the two-copy (FF) and four-copy (FFFF) constructs but not from the single-copy (F) and control constructs. Overall, these data suggest that the insertion variant may increase promoter activity in part through the creation of a functional enhancer element. * $P < 0.05$ by Tukey HSD test. VNTR, variable number of tandem repeats.

Fig. 5.

Postmortem studies. To determine whether the insertion genotype can influence SLC6A1 expression *in vivo*, we carried out real-time expression analysis and SLC6A1 genotyping on a cohort of African– American patients. (a) Multiple regression modeling revealed significant main effects of age, and genotype, as well as age $\times$ genotype, genotype $\times$ sex, and age $\times$ genotype $\times$ sex interaction effects. To facilitate the interpretation of these results we dichotomized age into values below (‘young’) and above (‘older’) the median age of our sample (44 years) and we calculated means and standard errors within the context of our regression model using least squares means and standard errors. Systematic effects of age, sex, and genotype are evident. (b) SLC6A1 RNA expression was plotted against patient age. The solid

line indicates a linear regression fit. SLC6A1 expression declines significantly with increasing age (Pearson's $r = -0.5187$, $P < 0.0001$).

Table 1

Genomic positions and gene annotations of sequences with at least 95% identity with the insertion sequence 'GGGTGGGGAGAGGGAGGGAGG'

Chromosome	Starting position	Ending position	Number matched	Genes involved
3	11009089	11009109	21	<i>SLC6A1</i> , first intron
22	22509215	22509235	21	<i>DERL3</i> , 5th intron
X	152979186	152979206	21	<i>MeCP2</i> , 2nd intron
2	44996458	44996477	20	No annotated genes in region
11	76507873	76507892	20	<i>CAPN5</i> , 9th intron
15	58076463	58076482	20	Geneless region, 5 kb upstream of FoxB1

Positions are relative to the March 2006 genome build.