Human Molecular Genetics, 2005, Vol. 14, No. 10 ***1393–1404***
doi:10.1093/hmg/ddi148
Advance Access published on April 13, 2005

SLC18A2 promoter haplotypes and identification of a novel protective factor against alcoholism

Zhicheng Lin*, Donna Walther, Xiao-Ying Yu, Suxia Li, Tomas Drgon and George R. Uhl

Molecular Neurobiology Branch, IRP, NIDA, NIH, DHHS, Baltimore, MD 21224, USA

Received February 7, 2005; Revised and Accepted April 4, 2005

The vesicular monoamine transporter 2 (VMAT2, *SLC18A2*) takes up cytosolic monoamines into intracellular secretory vesicles, preventing their neurotoxicity in the cytosol and discharging them into extracellular space by exocytosis. It has been shown that one-copy deletion of the VMAT2 gene increases locomotion activity significantly in response to drug treatments and dopamine neuron death rate in response to neurotoxin treatments in knockout mice. Little is known about promoter polymorphisms and their influence on *SLC18A2* promoter activity. We have re-sequenced a 17.4 kb DNA in the *SLC18A2* promoter region for Caucasians and revealed 47 polymorphisms that confer 13 haplotypes. One of the haplotypes reaches a frequency as high as 65%, likely due to positive selection. *In vitro* analysis showed a 20% difference in promoter activity between two frequent haplotypes and identified some of the polymorphisms that influence promoter activity. Four haplotype-defining single nucleotide polymorphisms (hdSNPs) can define the frequent haplotypes and by genotyping these hdSNPs, we find that haplotypes with −14234G and −2504C of *SLC18A2* promoter region represent a protective factor against alcoholism ($P = 0.0038$ by Fisher's exact tests). Therefore, *SLC18A2* promoter haplotypes defined here create a foundation for transcriptional characterization of individuality and for association study on monoamine-related human diseases.

INTRODUCTION

The vesicular monoamine transporter 2 (VMAT2) is an important molecule for the function of monoaminergic neurons that are key participants in locomotion, reward and mnemonic brain systems. Acting to deplete cytosolic monoamines (dopamine, serotonin, norepinephrine and histamine) by uptake into intracellular vesicles and to discharge the monoamines into extracellular space, VMAT2 prevents neurotoxicity of these monoamines in the cytosols and regulates neurotransmission. Levels of VMAT2 expression help determine the efficacy of spatio-temporal buffering of intracellular monoamines and minimize the degree of neurotoxicity.

Expression of VMAT2 displays a distinct tissue pattern. VMAT2-reactive immunostaining is only seen in central, peripheral and enteric neurons (1). In the brain, VMAT2 is expressed only in certain regions such as nigra compacta, hypothalamus, ventral tegmental area (VTA), subventricular zone cells and olfactory bulb (2–4). In particular, VMAT2 is expressed in monoaminergic and histaminergic neurons. Within VTA, VMAT2 expression level is different from subdivision to subdivision (5). The regulatory mechanisms underlying these tissue-specific expression patterns are unknown.

Data from knockout mice indicate that expression of VMAT2 is essential for survival and that different expression levels have behavioral consequences. Homozygous (−/−) knockout mice could survive less than two postnatal weeks. Heterozygous (+/−) knockout mice displayed attenuated rewarding effects of amphetamine by conditioned place preference and increased sensitivity to neurotoxicity of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) when compared with wild-type mice (6). These gene dosage and other pharmacological experiments have indicated clearly that reduced expression of VMAT2 gene could cause increased sensitivity to neurotoxicity (7) and abnormal behaviors, likely due to impaired vesicular regulation of neurotransmitter release into synaptic cleft (8). Deletion of VMAT2 gene reduced the capacity of primary cultures to release dopamine, 5-HT and histamine (9,10). The +/− mice displayed much greater locomotion activity in response to apomorphine, cocaine, amphetamine and ethanol injection (11). Furthermore, +/− male mice consumed more high-concentration (up to 32%) alcohol than did the +/+ mice (12).

Environmental stimuli could regulate expression of the VMAT2 gene. Lu and Wolf (13) reported that VMAT2 mRNA levels in rat midbrain were increased significantly

*To whom correspondence should be addressed at: Molecular Neurobiology Branch, National Institute on Drug Abuse, National Institutes of Health, 5500 Nathan Shock Drive, Baltimore, MD 21224, USA. Tel: +1 4105502843; Fax: +1 4105501535; Email: zlin@intra.nida.nih.gov

3 days after repeated amphetamine administration. In experiments designed to examine stress influence on expression of monoaminergic genes, long-term immobilization increased VMAT2 mRNA levels by 50% in brainstem A1 and A2 neurons of rats (14). Polychlorinated biphenyls are ubiquitous environmental contaminants. These contaminants bioaccumulate in higher tropic levels of the food chain and target the dopaminergic neurons. In mice, acute exposure to aroclor 1260, a polychlorinated biphenyl, decreased VMAT2 expression levels by 59% (15).

There is sparse information about the genetic architecture of the human VMAT2 gene (*SLC18A2*), especially the promoter region. We and others (16,17) have identified the 16 exons of *SLC18A2* that span >35 kb of genomic DNA on human chromosome 10. Twenty-three single nucleotide polymorphisms (SNPs) have been reported for *SLC18A2* in 450 Caucaisians and 304 Japanese, 19 of the SNPs in introns and four in exons (18–20). Two of the exon SNPs, located in exons 3 and 15 and with frequencies of <1%, confer amino acid mutations Thr → Met and Met → Ile that do not affect substrate uptake but do influence sensitivity to the VMAT2 inhibitor reserpine (21). There is no information in the literature about any polymorphisms in the *SLC18A2* promoter region.

We now report results of studies that seek polymorphisms in 17.4 kb upstream of *SLC18A2* exon 2, the haplotypes that these polymorphisms confer and the ways in which these haplotypes correlate with differences in *SLC18A2* expression *in vitro*. We report identification of common haplotypes in *SLC18A2* 5′ flanking regions that are likely to be responsible for significant aspects of *SLC18A2*, including expression and related human diseases.

RESULTS

Discovery of polymorphisms

We sought common variants and haplotypes by re-sequencing the 17 449 nucleotides 5′ to the *SLC18A2* ATG codon in exon 2 in DNA from each of 23 unrelated individuals of self-reported European-American ancestry. These individuals provide 95% power to detect polymorphisms of 6% frequency and 91% power to detect polymorphisms of 5% frequency. For 13 of the individuals, we obtained complete genotype data, providing 95% power to detect polymorphisms of 11% frequency and 74% power to detect polymorphisms of 5% frequency. These sequences appeared to be overlapped by another gene(s), because six different expressed sequence tags (ESTs) of unknown function have homology to this genomic region (Fig. 1). ESTs c–f are probably associated with the presences of repetitive sequences (discussed subsequently).

From sequencing 23 individuals for the 17.4 kb promoter region, we identified 43 polymorphisms, including 37 SNPs, five insertion/deletion polymorphisms (indels) and one simple sequence length polymorphism (SSLP) in the *SLC18A2* promoter (17 449 bp) region upstream of the exon 2 (Fig. 1). The 5′ SSLP located at −5199 bp has at least six alleles and two of them are major ones: allele *a* of 38 bp and allele *b* of 42 bp. This SSLP displayed 0.5428 heterogeneity. The average polymorphism rate is 393 bp per site. Seventeen (40%) of these polymorphisms are novel and 26 (60%) are contained in current NCBI *SLC18A2* database listings (Fig. 1).

The pattern of polymorphism occurrences looks interesting. Forty-two percent of the polymorphisms were rare mutations with a frequency of <11 and 13% were in the 30–40% category, as expected; 16% of the polymorphisms fit into the 11–30% category, less than half of the expected. In contrast, 29% were highly frequent (>40%) polymorphisms, 2.34-fold higher than expected (Table 1).

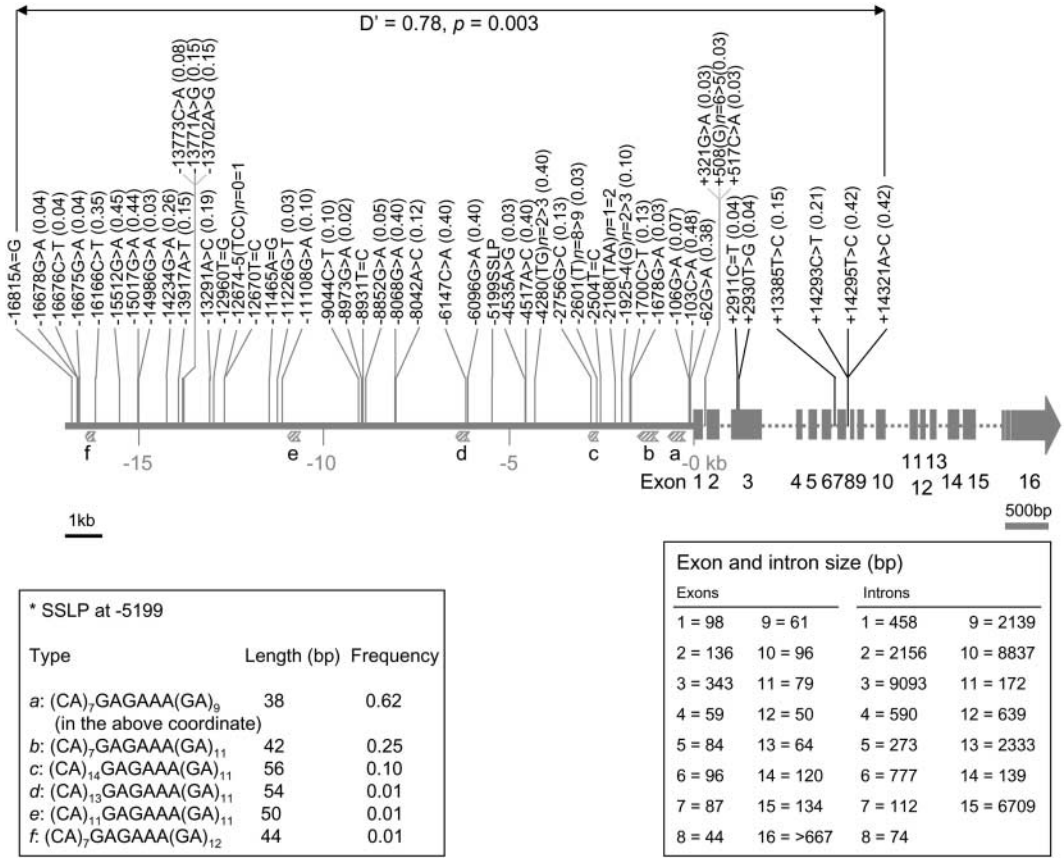
Additional SNPs were found to lie downstream of exon 2, by sequencing the *SLC18A2* gene region for a few of the individuals. They included six SNPs in the downstream region, including +2911C/T and +2930T/G in exon 3, +13385T/C in intron 6 and +14293C/T, +14295T/C and +14321A/C in intron 7. The +2911C/T is a missense mutation, Thr68 → Met, in the hVMAT2 protein. These six SNPs are consistent with what has been reported previously (18).

High nucleotide diversity and positive selection

We assessed nucleotide diversity (θ), heterozygosity (π) and deviations from the expectations based on neutral mutation models (23) for the polymorphisms that lie in the *SLC18A2* promoter region (Table 2). In general, SNPs show higher θ- and π-values than other types of polymorphisms; π-values are higher for the indel than for the SSLP. The π-value is 0.0007926, slightly higher than one (0.000637) observed for the downstream non-coding regions of *SLC18A2*, whereas the θ-values are similar (18). Intron 1 shows values of θ but not π higher than the 5′ region. Tajima *D*-values are significant only for the SSLP, which displays $D = 2.0761$ ($P < 0.05$). However, the 5′ region displays positive Tajima *D*-values, but intron 1 displays negative values for all three types of polymorphisms (Table 2).

High linkage disequilibrium

These common allelic variants provide markers for linkage disequilibrium (LD) analysis. The 17.4 kb promoter region examined in this study displayed high LD ($D' = 0.78$, $P = 0.003$) between the SNP +517C/A in intron 1 and the upstream SNP −16815A/G (Fig. 1). This promoter region displayed evident blocks of restricted diversity. As shown in the GOLD graph, a majority (three-quarters) of the areas reached D'-values >0.8 (Fig. 2). However, some of the alleles rarely transmitted together with others. Among the possible 903 polymorphism pairs that can be constructed from the 43 biallelic polymorphisms assessed in these samples, 66.0% show D'-values of 1 and 17.3% show D'-values between 0.5 and 1. Only a few polymorphism pairs display no or little LD: no LD between −15512G/A and −11226G/A, +321G/A or +517C/A, between −11226G/A and $-2108(TAA)_{1/2}$ or −103C/A; between $-2108(TAA)_{1/2}$ or −103C/A and +321G/A or +517C/A. The polymorphisms −15017G/A, −12960T/G, $-12674(TCC)_{0/1}$, −12670T/C and −11465A/G each have low D'-values (nominally 0.062) with −11226G/A, +321G/A or +517C/A; −8931T/C has the same D'-values with −11226G/A, +321G/A or +517C/A; −16675G/A has a lower D'-value (0.048) with −62G/A. Although these



* SSLP at -5199

Type	Length (bp)	Frequency
a: $(CA)_7GAGAAA(GA)_9$ (in the above coordinate)	38	0.62
b: $(CA)_7GAGAAA(GA)_{11}$	42	0.25
c: $(CA)_{14}GAGAAA(GA)_{11}$	56	0.10
d: $(CA)_{13}GAGAAA(GA)_{11}$	54	0.01
e: $(CA)_{11}GAGAAA(GA)_{11}$	50	0.01
f: $(CA)_7GAGAAA(GA)_{12}$	44	0.01

Exon and intron size (bp)

Exons		Introns	
1 = 98	9 = 61	1 = 458	9 = 2139
2 = 136	10 = 96	2 = 2156	10 = 8837
3 = 343	11 = 79	3 = 9093	11 = 172
4 = 59	12 = 50	4 = 590	12 = 639
5 = 84	13 = 64	5 = 273	13 = 2333
6 = 96	14 = 120	6 = 777	14 = 139
7 = 87	15 = 134	7 = 112	15 = 6709
8 = 44	16 = >667	8 = 74	

Figure 1. Schematic of *SLC18A2* 5′ region, exons and polymorphisms. Horizontal line represents chromosomal DNA from left (5′, centromere) to right (3′, telomere); black, 5′ region. Gray blocks represent exons (as numbered underneath them); gray arrow represents exon 16 whose 3′ end is not yet defined. Introns are indicated as dashed lines and are not to scale. The sizes of 16 exons and 15 introns are listed in the right box. Exon 1 is defined according to the cDNA clone (L23205, from substantia nigra) of Erickson and Eiden (22), although another cDNA clone (BG718762, from testis) starts at −26. Coordinate is −1 for the first base pair 5′ to exon 1 and is +1 for the first base pair of exon 1. Vertical lines represent locations of polymorphisms with gray ones in the 5 promoter region and black ones in the downstream regions; each polymorphism is indicated by its coordinate, its polymorphism (major allele > minor allele or the two allele have the same frequency ' = ') and the frequency of the minor allele in the parenthesis. LD (*D*′ and its associated *P*-value) is indicated below the horizontal double-arrow for a pair of polymorphisms. A SSLP located at −5199 is detailed of its alleles, length and frequency in the left box. Hatched arrow heads, ESTs whose sequence sources are: a, AW292760 (from lung) and BF512349; b, BI911851(from leukocyte); c, BX644039 and 63 other entries (from neuroblastoma cell lines and peripheral tissues); d, T02968 (from fetal brain); e, BF893626 (from duodenal adenocarcinoma cell line); f, AW961582 and 55 other entries (from both brain and peripheral tissues). Gray scale bar, 1 kb for 5′ region; black scale bar, 500 bp for the exons. Twenty-six polymorphisms contained in the NCBI database (Build 34.3) are −16815, rs2619103; −16676, rs2532797; −16166, rs3026047; −15512, rs2619102; −15017, rs2803820; −14234, rs363371; −13917, rs363323; −13771, rs2619099; −13702, rs2619098; −13291, rs2803818; −12960, rs2532798; −11465, rs363324; −9044, rs363329; −8931, rs2532799; −8852, rs12414934; −8068, rs10886050; −8042, rs2619097; −6147, rs10886051; −6096, rs2532801; −4535, rs7099849; −4517, rs4081624; −4280, rs10689256; −2756, rs2532802; −2108, rs3981221; −1700, rs11197931; +517, rs2619094 (the underlined are those used by the International HapMap Project).

D′-values should be viewed with caution as the associated *P*-values are >0.89, they suggest a relatively greater rate of recombination between each pair.

Identification of one major and 12 minor haplotypes

We used the program Haplotyper to estimate the haplotypes in the initial sample of 15 individuals subjected to re-sequencing and with the most complete genotype data (Table 3). Of the 15 individuals, seven were controls and eight were abusers. We obtained complete genotype data for 13 individuals. We missed 35% data for the other two, which represented 4.7% of the overall data and was apparently tolerated by Haplotyper, because the posterior probabilities were >0.92 for inferring the haplotypes of these two individuals. As a result, inference from these 15 individuals' genotype data revealed 13 haplotypes, from *A* to *M*. *A* is the major haplotype with the highest frequency of 43% and *B* had the highest second frequency of 17%. The rest of the haplotypes had low frequency, <7%. We have identified four SNPs, −14234G/A, −11465A/G, −2504T/C and −1700C/T, that appear to differentiate *A* and *B* from one another and each from the rest of the 13 haplotypes (Table 3) (discussed subsequently). We term these haplotype-defining SNPs as hdSNPs. Haplotypes *A, D, F* and *L* carry allele *a* of SSLP; *B, C, E, I* and *J* carry allele *b*; *M* carries allele *d*; *H* and *K* carry allele *f* and *G* carries allele *g* (Table 3 and Fig. 1).

Cladistic analysis shows that *A* and *B* are located in two distinct clades (Fig. 3). *A* is probably the oldest lineage. *F, L* and *E* were all derived from *A* but failed to expand, likely due to

Table 1. Frequency of polymorphisms

Frequency of polymorphisms	Expected (%)[a]	Observed 5′	Intron 1	Total	OR[b]
<11%	40.6	16 (38%)	3 (100%)	19 (42%)	1
11–30%	34.2	7 (17%)	0	7 (16%)	0.45
30–40%	13.2	6 (14%)	0	6 (13%)	0.95
>40%	12.0	13 (31%)	0	13 (29%)	2.34

Polymorphisms were grouped into 5′ and intron 1. Data from 23 individuals (complete genotyping of 13 and incomplete of 10) are used for *SLC18A2*. Eleven percent is the detectable frequency with a power of 0.95 for a sample size of 26 chromosomes.
[a]Calculated according to the method of Glatt *et al.* (18).
[b]Odds ratio is total versus expected, category <11% is the referent.

Table 2. Nucleotide diversity and average heterozygosity for *SLC18A2* 5′ region

Polymorphism location		Base pair screened	No. of polymorphisms	θ-value	π-value	Tajima's *D*-value
5′	SNP		35	0.0005429	0.0006683	0.6880
	Indel		4	0.0000621	0.0001002	1.4757
	SSLP		1	0.0000155	0.0000366	2.0761*
	Total	16893	40	0.0006205	0.0008051	0.9354
Intron 1	SNP		2	0.0011444	0.0003359	−1.5373
	Indel		1	0.0005722	0.0001679	−1.1741
	Total	458	3	0.0017165	0.0005038	−1.7613
Overall	SNP		37	0.0005557	0.0006558	0.5070
	Indel		5	0.0000751	0.0001014	0.8454
	SSLP		1	0.0000150	0.0000354	2.0761*
	Total	17449	43	0.0006458	0.0007926	0.6831

Nucleotide diversity (θ) and average heterozygosity (π) were obtained from 26 chromosomes. Analyses were grouped based on the locations and forms of polymorphisms. The 98 nucleotides of exon 1 were excluded for the analysis of the 5′ region but included for the overall analysis; 5′SSLP was not included in indel.
*D-values are significant for $P < 0.05$ based on β distribution.

negative selection pressure. The rest of the haplotypes form a distal clade and are all closely related, with *B* as the favored lineage during human history.

Different *in vitro* promoter activity conferred by different haplotypes

We used the luciferase reporter system and human neuroblastoma cell lines to examine promoter activity of different haplotypes and to identify functional polymorphisms. We constructed expression vectors using the *SLC18A2* 5′ haplotypes *A* and *C* across the 6.3 kb region that extends from −5555 to +751 bp. The 6.3 kb DNAs that represented haplotypes *A* and *C* differed from each other by six SNPs and two indels. When compared with *B, C* had an additional polymorphism at −1700 and provided more power for identifying functional polymorphisms. Fragments were inserted into Promega pGL3-enhancer and the activities of luciferase were compared in transiently expressing IMR-32 and SH-SY5Y cells (Fig. 4). Haplotype *C* displayed higher promoter activity than *A* consistently, 30% higher in IMR-32 and 18% higher in SH-SY5Y cells. These *in vitro* functional assays suggest that 5′ haplotypes *A* and *C* confer different promoter activities.

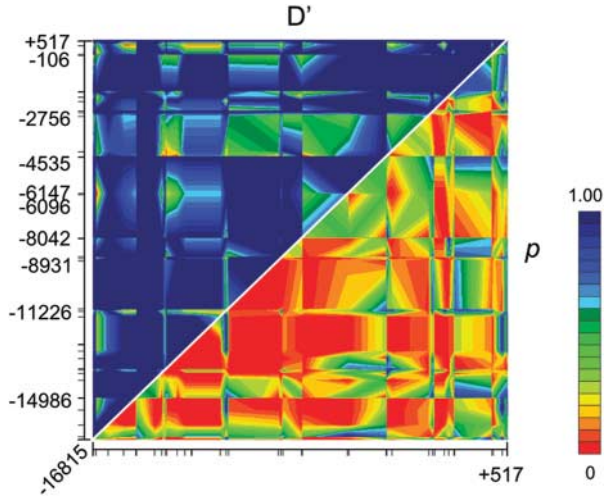


Figure 2. GOLD plot of blocks of restricted haplotype diversity in the *SLC18A2* 5′ flanking region. Coordinates are indicated on the left hand side and at the bottom. Upper left half, D'-values and lower right half for the associated P-values. The color scales are shown at right hand side: dark blue=1 and red = 0.05 or less for either D' or P.

Detailed chimerism analysis identified polymorphisms important for promoter activity. We progressively switched the *A* alleles into *C* alleles on the *A* background, from the 3′ to the 5′ end of the 6.3 kb fragments. Polymorphisms −62G/A and −103A/C decreased promoter activity by one-third of the haplotype *A* activity, as shown by *C*1. SNP at −1700 did not influence promoter activity in these cell lines, as shown by *C*2. Switching allele *T* into *C* at −2504 and insertion of an additional copy of trinucleotide TAA at −2108 restored the promoter activity to haplotype *A* levels, as shown by *C*3. The last three polymorphisms at −4280, −4517 and the SSLP together seemed to contribute an additional 15% of haplotype *A* activity, as shown by comparing *C*3 and *C* (Fig. 4).

Association with alcoholism

We wanted to know whether the major haplotypes were associated with alcoholism. We used the four hdSNPs, −14234G/A, −11465A/G, −2504T/C and −1700C/T, to genotype 333 (Collaborative Study on Genetics of Alcoholism) COGA-unrelated individuals (189 controls and 144 alcoholic patients). Haplotypes were inferred by PL-EM. Again, *A* and *B* were the two most frequent or major haplotypes, with frequency of 65.2 and 13.8%, respectively. This COGA population missed haplotype *I*, but had six additional haplotypes *N, O, P, Q, R* and *S*. GACC (representing haplotypes *E, F* and *L*) appeared to be associated significantly with alcoholism by Fisher's exact tests ($P = 0.0245$, Table 4). Haplotype *R* tended to be over-represented in alcoholics ($P = 0.0804$ by Fisher's exact tests). Overall, the *SLC18A2* promoter region was associated with alcoholism at marginally significant levels ($P = 0.0699$, based upon χ^2-values by case–control

Table 3. Promoter region haplotypes of *SLC18A2*

ID	Haplotype[a] (below coordinates)	Frequenly (%)	SSLP allele[b]
	−16815 −16678 −16676 −16675 −16166 −15512 −15017 −14986 −14234 −13977 −13773 −13371 −13702 −13291 −12960 −12674 −3 −12670 −11465 −11226 −11108 −9041 −8931 −8852 −8068 −8042 −6147 −6096 −4535 −4517 −4208 −7 −2756 −2601 −2504 −2108 −7 −1925 −4 −1700 −1678 −106 −103 −62 −321 −508 −513 −517		−5199
A	AGCGCAGCGACAAAT0TAGGCTGGACGAA2G8T12CGGAGG6C	43	*a*
D	AGCGCGACGACAAAG1CGGGCTGGACGAA2G8T12CGGAGG6C	3.3	*a*
E	AGCACAGCGACAAAT0TAGGCTGAAAAAC3G8C22CGGCAG6C	3.3	*b*
F	AGCACAGCGAAGAAT0TATGCTGGACGAA2C8C12CGGAGA6A	3.3	*a*
G	AGCACGACGACAAAG1CGGGCCGAACGAC3C9C23CGACGG6C	3.3	*g*
H	GGCGCGACGTCGGCG1CGTATCAGCCGAC3C8C22CGGCGA6A	3.3	*f*
I	GGCGCGATGTCAACG1CGGGTCAGCCGGA3G8C22TGGCGG6C	3.3	*b*
J	GGCGCGACGTAGGCG1CGGATCGGCCGAA2G8T23CGGCGG6C	3.3	*b*
K	GGCGCGACGTAGGCG1CGGATCGGCCGAC3C8C23CGACGG6C	3.3	*f*
L	GGCGTGGCGACAAAT0TAGGCCGAAAAAC2G8C22CGGCAG6C	3.3	*a*
B	GGCGTGACAACAAAG1CGGGCCGAAAAAC3G8C22CGGCAG6C	17	*b*
C	GGCGTGACAACAAAG1CGGGCCGAAAAAC3G8C22TGGCAG6C	6.7	*b*
M	GATGCGACGTCGGCG1CGGGCCGGCAAAA2C8C22CAGCGG5C	3.3	*d*

[a]Underline indicates the four haplotype-defining SNPs (hdSNPs).
[b]SSLP allele that each haplotype carries, see Figure 1 for sequences.

analysis; Table 4). Most interestingly, the *G* allele at −14234 together with the *C* allele at −2504 had a 14% frequency and was over-represented in controls when compared with 6.9% in alcoholic patients, representing a significant protective factor against alcoholism ($P = 0.0038$ by Fisher's exact tests, Table 4).

GOLD plotting reveals that the alcoholic patients display LD patterns different from the controls in the proximate promoter region (from −1.7 to −6.2 kb, Fig. 5). The D'-values range from 0.595 (−11 465 versus −1700) to 0.938 (−11 465 versus −2504) for controls but from 0.769 (−14 234 versus −2504) to 0.916 (−14 234 versus −1700) for the alcoholic patients. Overall, the alcoholic patients display stronger LD than the controls in the proximate promoter region.

The SSLP marker was not associated with drug abuse vulnerability. We also looked at an NIDA sample as a population of illegal drug abuse, using the SSLP marker. We did transmission-based disequilibrium tests for 107 families, of which at least one member was an abuser. The results indicated that there was no significant transmission disequilibrium or maternal transmission disequilibrium between the unaffected and the affected (abusers, data not shown).

DISCUSSION

Our data on population genetics, functional genomics and association study of *SLC18A2* promoter region provide a novel insight into human genetics of *SLC18A2*, an unexplored territory. *SLC18A2* promoter has one major haplotype that represents up to two-thirds of chromosomal copies for the promoter region in the populations, based upon the two mixed Caucasian populations that we examined here; polymorphisms within the 5 kb proximal region contribute to S*LC18A2* promoter activity, based upon *in vitro* assays; S*LC18A2* promoter region is associated marginally with alcoholism, based upon our case–control analysis.

The International HapMap Project (http://www.hapmap.org/index.html.en) uses eight SNPs in the 17.4 kb *SLC18A2* promoter region. Two of the SNPs, rs2619103 and rs2532798, are −16815A/G and −12960T/G, respectively, as revealed in this study (Fig. 1). In the HapMap Caucasian samples, these two SNPs form two haplotypes 'AT' and 'GG' that have frequencies of 71 and 29%, respectively, when compared with 50 and 40% in our modest sample (Table 3). These two SNPs have heterogeneities of ∼0.4 (0.5 in our sample); the six other SNPs each have heterogeneities of <0.07 in the Caucasian samples but 0.22–0.5 in Chinese, Japanese and Africans. Therefore, our haplotype information is much more comprehensive and representative of Caucasian samples.

Evidence is provided here that the promoter region has blocks of restricted diversity that lead to the formation of a major haplotype and a few minor ones. First, GOLD display of LD shows that a majority of the areas have high D'-values and less than a quarter of the areas have relatively low D'-values, consistent with the observations from COGA samples (Fig. 5). In particular, intron 1, SNPs at −4517, −11 108 and −11 226, are the areas that display more diversity and low LD (Fig. 2). Such diversity is probably generated by the presence of genomic DNA repeats (Fig. 6). The 17.4 kb promoter has 40.5% sequences with high homology to various chromosomal repeats and another 59.5% as unique regions. SNPs at −1678, −1700, −4517, −6096, −6147 and −11 108 are located in those repeats and happen to be at the low D' areas. Overall, three-quarters of the polymorphisms that generate low D' are located in the repetitive sequences (Figs 2 and 6) (see Results). Secondly, Tajima's *D*-values are positive for all the polymorphisms in the 5′ region but negative for all polymorphisms in intron 1, which again indicates that intron 1 and the 5′ region represent two blocks of restricted diversity (Table 2). Sixty percent of the polymorphisms that we report here have been deposited in the NCBI databases. However, we are able to define the common haplotypes for

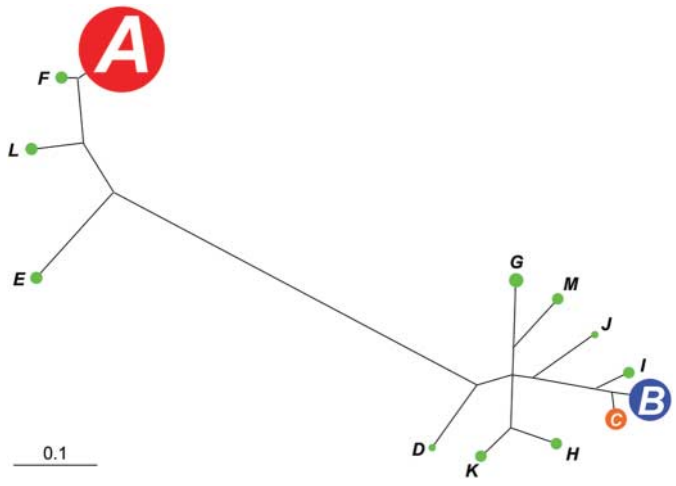


Figure 3. Cladogram of *SLC18A2* promoter haplotypes (*A*–*M*). The relatedness of these haplotypes was obtained using 17.4 kb genomic DNA sequences that represent the corresponding haplotypes. The size of the circle represents the frequency of the haplotype: the first three haplotypes *A, B* and *C* are in red, blue and orange, respectively, and the rest in green. Bootstrap values from 2000 simulations: 1.0 for all nodes except 0.60 for the *D* node. Scale bar: substitution per site.

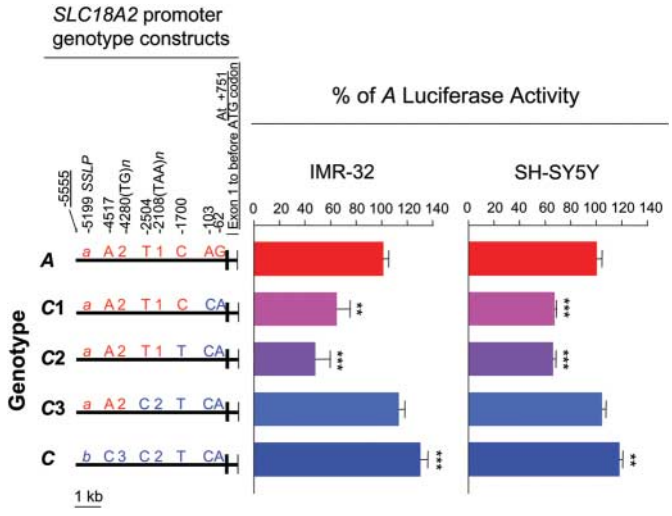


Figure 4. Luciferase activities from *SLC18A2* 5′ haplotype constructs. Left panel: genotypic description of a 6.3 kb fragment (horizontal lines) of the *SLC18A2* promoter region that corresponds to the region from +751 to −5555 in Figure 1 and contains eight polymorphisms comparing haplotypes *A* and *C*. Five expression plasmids used here are pGL3e-hVMAT2-6.1A (*A*), pGL3e-hVMAT2-6.1C1 (*C*1), pGL3e-hVMAT2-6.1C2 (*C*2), pGL3e-hVMAT2-6.1C3 (*C*3) and pGL3e-hVMAT2-6.1C (*C*). The vertical black bars represent exon 1 and vertical lines represent 5′ part of exon 2. Scale bar, 1 kb. These expression plasmids are *SLC18A2* promoter (6.3 kb)-Luc hybrids (see Materials and Methods) for luciferase activity assays in either IMR-32 (middle panel) or SH-SY5Y cells (right panel). Alleles of haplotype *A* are indicated in red and those of haplotype *C* in blue; chimeras are *C*1, *C*2 and *C*3. The positions (coordinates) of these polymorphisms are indicated above the *A* construct. Luciferase activity is expressed in percentage of what (red) haplotype *A* construct displayed and indicated in a blue bar for haplotype *C* construct or intermediate colors for the chimeric constructs. Analysis of variance (ANOVA) and Student's *t*-test: $^{**}P < 0.01$ and $^{***}P < 0.001$, $n = 10$–14.

the 17.4 kb promoter region in a systematic way. The haplotypes that we have defined here are the first for this important gene, to the best of our knowledge.

Genetic analysis of nucleotide diversity indicates strong positive selection of *SLC18A2* promoter haplotypes. Cladistic analysis of the haplotypes reveals a large distance between the two most frequent haplotypes *A* and *B*. Evidently, *A* has been heavily favored during human biological history (Fig. 3). Definition of the *SLC18A2* 5′ sequence variants also allows estimations of the deviations from expectations based on neutral mutation models. Although our modest sample size could underestimate some of the Tajima *D*-values, it is interesting that the significant *D*-value (2.08) for the 5′ SSLP appears to indicate positive selection pressure (18). Such observations, coupled with the retention of haplotype *A* in the current populations, might even be consistent with an unbalanced selection of the major 5′ haplotype. This unbalanced selection would provide one contributor to the restricted haplotype diversity in this region that we identify in the current population samples.

The *in vitro* expression data not only suggest differences in promoter activity between haplotypes, but also identify candidate polymorphisms that likely influence the promoter activity. Polymorphisms at −62, −103 and −4517 are all located in consensus sequences for transcription factor (TF) binding. For example, allele *A* at −103 of haplotype *A* is in a Sp1-like activating site (GGGAGG), a MAP kinase (ERK1)-related Sp1 site and an activating *cis*-acting element (24–26). Consistent with previous observations, mutation of this site reduces promoter activity (Fig. 4). The SNPs at −2108 and −2504 are obviously functional as well, but there are no known TF consensus binding sites for these SNPs. We did not include *B* in this study. *B* has all the polymorphisms as does *C*, except that *B* has *C*, instead of *T* at −1700, which means *A* and *B* would have seven polymorphisms on the 6.3 kb fragment. As −1700C/T does not seem functional, we expect that *B* would have the same promoter activity as would *C* in Figure 4. These *in vitro* data may not necessarily reflect the haplotype activity *in vivo* due to other unknown activities and known mechanisms such as *SLC18A2* DNA methylation (Takahashi and Uhl, unpublished data). However, they help to identify individual *cis*-acting elements. Further mutagenesis study and one-component hybridization analyses should help to identify the corresponding TFs.

Significant association between *SLC18A2* promoter and alcoholism that our data suggest here is consistent with several lines of evidence pointing to an association between *SLC18A2* and different diseases including addiction. It has been found that VMAT2 reduced expression levels in cocaine abusers, based on both [(3)H]dihydrotetrabenazine binding and western blots of postmortem brain tissues (27). In a binding study with (+)[(11)C]dihydrotetrabenazine, VMAT2 protein levels were elevated in untreated major depression patients (28) and in the ventral brainstem region of bipolar disorder patients, which was correlated differentially between men and women (29). In Flinders sensitive line rats, a genetic animal model for clinical depression in humans, VMAT2 mRNA levels were elevated in the Dorsal raphe when compared with control Sprague–Dawley rats

Table 4. Haplotype association with alcoholism

Defined haplotype[a]	SNP				(%)[b]	COGA unrelated individuals[c]		Fisher's exact test	
	−14234	−11465	−2504	−1700		CTL	Alcoholic	*P*-value[d]	
A	G	A	T	C	(65.2)	241 (63.8)	193 (67.0)	0.4120	[0.9]
B	A	G	C	C	(13.8)	48 (12.7)	44 (15.3)	0.3653	[0.9]
C	A	G	C	T	(4.2)	16 (4.2)	12 (4.2)	1.0	[1.0]
D, J	G	G	T	C	(1.1)	4 (1.1)	3 (1.0)	1.0	[1.0]
E, F, L	G	A	C	C	(4.5)	23 (6.1)	7 (2.4)	**0.0245**	[2.6]
G, H, K, M	G	G	C	C	(6.0)	26 (6.9)	12 (4.2)	0.1768	[1.7]
N	A	A	T	C	(2.9)	11 (2.9)	8 (2.8)	1.0	[1.1]
O	G	A	C	T	(0.8)	4 (1.1)	1 (0.3)	0.3959	[3.1]
P	A	A	C	C	(0.8)	2 (0.5)	3 (1.0)	0.6569	[0.5]
Q	G	A	T	T	(0.5)	3 (0.8)	0 (0)	0.2626	–
R	A	G	T	T	(0.5)	0 (0)	3 (1.0)	0.0804	–
S	A	G	T	C	(0.3)	0 (0)	2 (0.7)	0.1866	–
					(100)	378 (100)	288 (100)		
Globtal test[e]					$\chi^2 = 18.54$, df = 11, *P*-value = 0.0699				
E-O[f]	G	×	C	×	(11.0)	53 (14.0)	20 (6.9)	**0.0038**	[2.2]

[a]As defined in Tables 3 and 5; haplotype *I* (GGCT) is not present in this COGA sample and not listed in this table; *N*, *O*, *P*, *Q*, *R* and *S* are new haplotypes in this sample.
[b]Number in parentheses, haplotype frequency as %.
[c]Haplotype counts; CTL, control; number in parentheses, haplotype frequency as %.
[d]Bold, statistically significant; number in parentheses, odds ratio; –, inaccurate numbers not shown.
[e]Case-control analysis by CLUMP.
[f]Haplotypes with *G* at −14234 and *C* at −2504.

(30). More interestingly, a recent genome-wide scan study searching for vulnerability loci for nicotine dependence has suggested *SLC18A2* to be one of the candidate genes for nicotine dependence (31). These data are consistent with the observations that the heterozygous VMAT2 knockout mice gained significant sensitivity to drugs such as cocaine, ethanol and amphetamine (6,11), suggesting that alterations in *SLC18A2* expression may underlie different human diseases. In this study, we did not observe any positive association between *SLC18A2* markers and abuse of illegal drugs. Among the 15 individuals used for polymorphism discovery and haplotype inference, there were no differences in haplotype distribution between the seven controls and the eight abusers. TDT analysis of 107 families using the SSLP marker failed to show any significant association. In the future, large samples are required to be genotyped for the four hdSNPs in order to validate the current findings and to investigate possible association between *SLC18A2* and vulnerability to abuse of illegal drugs.

The fact that *G* at −14234 together with *C* at −2504 is over-represented significantly in controls has two important implications. The first important implication is that, in haplotype association studies, individual haplotypes need to be investigated, but combination of various haplotypes needs to be considered as well. This is because that many of the polymorphisms are not functionally important and the involvement of these non-functional haplotypes in defining haplotypes could 'dilute' functionally distinct haplotypes. The second important implication is that alleles at −14 234 and −2504 or other alleles that are closely linked to these two locations are likely functional and these alleles function together. 'TCCCG**G**CGCT' (allele *G* at −14 234 in bold) is a consensus sequence for repressor GCF that is expressed in many tissues including brain, but 'TCCCG**A**CGCT' (allele *A* at −14 234 in bold) is not (32). At this point, we do not know any TF that binds to the −2504 forward locus. *In vitro* data show that the promoter activity can be doubled by switching −2504T and $-2108(TAA)_1$ into −2504C and $-2108(TAA)_2$, indicating that the −2504C site could be a TF binding site and this locus merits further investigation (Fig. 4).

These current *SLC18A2* genetic data thus add to our understanding of *SLC18A2* structure and function by revealing polar diversity and functional significance of polymorphism in *SLC18A2* promoter sequences. Our documentation of haplotype-dependent differences in *SLC18A2 in vitro* promoter activity is essential for understanding the significance of results from VMAT2 knockouts. As demonstrated, the haplotypes defined here provide a platform where transcriptional characterization of individual differences at this locus and further association studies should be pursued to understand their roles in related human diseases.

MATERIALS AND METHODS

Subjects

A total of 608 individual genomic DNAs were used in this study. Re-sequencing studies used DNAs from 23 unrelated European-American participants. Of these participants, four control individuals reported no lifetime histories of any significant use of any addictive substance, whereas eight polysubstance abusers reported substantial use and dependence on at least one illegal substance. Eleven other European-American participants were provided by Dr Gabbay at the Uniformed Service University of the Health Sciences. TDTs used 252

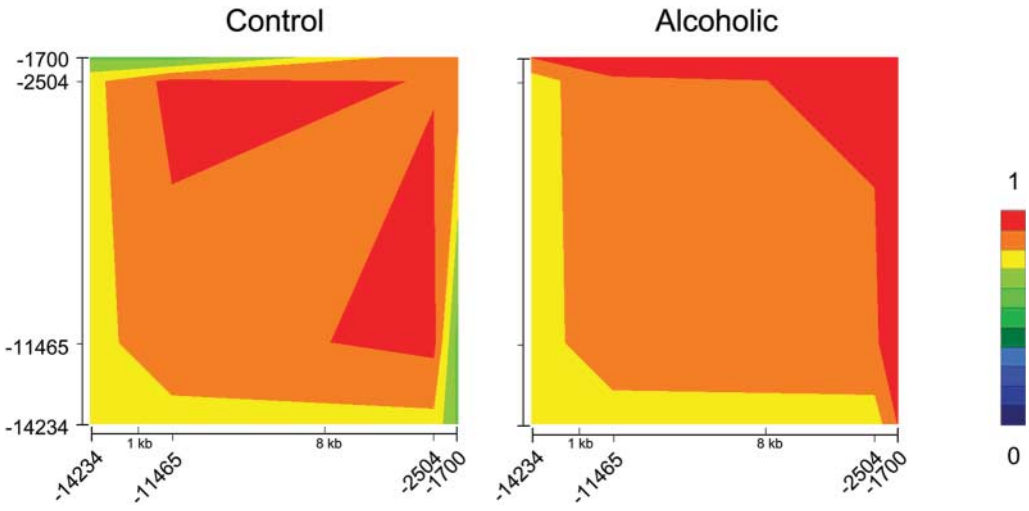


Figure 5. GOLD plot of LD based on the genotype data with the four hdSNPs for both control individuals (left) and alcoholic patients (right). Coordinates are indicated on the left hand side and at the bottom. The associated *P*-values are all <0.0001. The color scale bar for both panels is shown at further right hand side: dark blue=0 and red = 1.

individuals from 107 families, of which at least one member was an abuser. Seventy-seven percent were Caucasians. Eight of 107 families had information from both parents and 37 of them from only single parents. Seventy-three were unaffected, 179 were abusers and all recruited at NIDA-IRP (National Institute on Drug Abuse-Intramural Research Program) in Baltimore, MD, USA. Association studies used DNAs from 333 (189 control and 144 alcoholic) unrelated individuals selected from COGA pedigrees. From each of the COGA pedigrees, we selected the grand parents and the offspring's spouses that came from outside the pedigree as unrelated individuals. Of the COGA samples, 92.5% (including 92% of alcoholics and 93% of controls) were self-reported European-Americans and the rest were African-Americans. The 189 control individuals were all pure unaffected based upon DSM-IIIR, Feighner and JCD-10. All of 144 alcoholic individuals met at least two of the DSM-IIIR, Feighner and JCD-10 criteria for alcohol dependence. Genotyping of this COGA sample with other unlinked markers has shown negative association and suggested that this sample has little ethnic stratification. This study was approved by NIDA Institutional Review Board.

Polymerase chain amplifications of genomic sequences and polymorphism discovery

Pairs of 20–33 base polymerase chain reaction (PCR) primers allowed amplification of 0.5–3.2 kb DNA fragments from VMAT2 genomic sequences from individual genomic DNA samples using 95°C for 3 min, 45 cycles of 94°C for 0.5 min, 55°C for 1 min, 72°C for 3 min, followed by 72°C for 7 min. Annealing temperatures and extension times varied slightly according to the product sizes and primer T_m-values. Each 50 μl reaction contained 10 ng of genomic DNA, 1 μM of each oligonucleotide primer, 1.5 mM $MgCl_2$, 1× PCR buffer, 5% dimethylsulfoxide (DMSO) and 2 U of *Taq* I Gold polymerase (PerkinElmer Inc., Wellesley, MA, USA). GC-rich sequences around exon 1 were amplified using *Tsg* polymerase (Lamda Biotech Inc., St Louis, MO, USA) or Herculase (Stratagene, La Jolla, CA, USA). Each PCR product underwent DNA sequencing for polymorphism discovery. PCR products containing SSLP or indel (insertion/deletion) variants were cloned into pCR4-TOPO cloning vector (Invitrogen, Carlsbad, CA, USA), and DNAs purified (Qiagen, Valencia, CA, USA) from five to eight transformants for each genomic DNA were sequenced. A total of ~300 kb genomic DNA was re-sequenced to discover polymorphisms.

Power calculation

The probability of detecting a variant with allele frequency p in a sample of n individuals was calculated as $1-(1-p)^{2n}$. The allele frequency of a variant that can be detected with a probability of P in a sample of n individuals was calculated as $1-\sqrt[2n]{1-P}$.

Genetic data analysis

GCG (Genetics computer group, University of Wisconsin, Madison, WI, USA) software was used for DNA sequence analyses, Haplotyper (http://www.people.fas.harvard.edu/~junliu/Haplo/docMain. htm) (33) and PL-EM (http://www.people.fas.harvard.edu/~junliu/plem/) (34) programs for haplotype inference, HaploXT (http://archimedes.well.ox.ac.uk/pise/haploxt-simple.html) and GOLD (http://www.sph.umich.edu/csg/abecasis/GOLD/) (35) for graphical display of LD, 2LD (http://www.iop.kcl.ac.uk/IoP/Departments/PsychMed/GEpiBSt/software.shtml) (36) for LD analyses, Repeatmasker Web Server (http://www.repeatmasker.org/) for identification of repetitive sequences, Tfsitescan (http://www.ifti.org, updated on January 31, 2005) for identification of mammalian TF binding consensus sites, ClustalX (http://www.bioinformatics.ubc.ca/resources/tools/index.php?name = clustalx) (37), TreeView

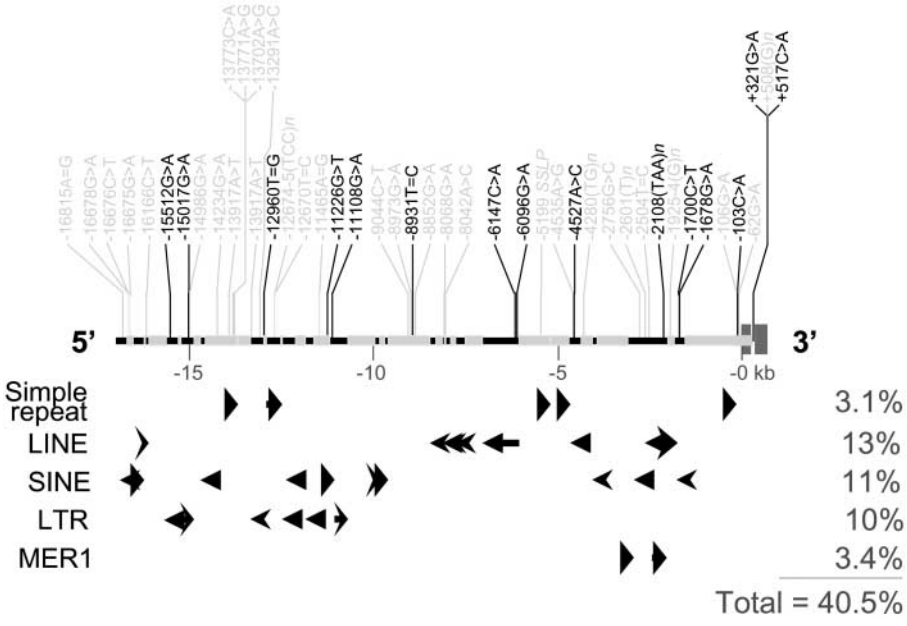


Figure 6. DNA sequence repeats (black arrows and arrow heads) and locations of low-LD polymorphisms (black vertical lines and labels) in the 17.4 kb *SLC18A2* promoter region (the thick horizontal line, 5′ to 3′ ends). Types of repeats identified here include simple repeats, LINE (long interspersed nuclear element), SINE (short interspersed nuclear element), LTR (long terminal repeat) and MER1 (human medium reiterated frequency repeat_type 1). Details of the presence of each repeat are LINE: three LINE1s, four LINE2; SINE: six ALUs and four MIRs; LTR: three MaLRs, one ERVL, and one ERV_classI. Repetitive sequences are in pieces of black horizontal lines and unique sequences in gray. The sequence length that each type of repeats represent is indicated as a percentage on the right hand side. Gray vertical boxes represent exons 1 and 2.

(http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) (38) for cladistic analysis and CLUMP (http://www.hgmp.mrc.ac.uk/About/Courses/2002/comp.linkage/clump.html) (39) for case–control analysis of promoter haplotypes. Fisher's exact tests and odds ratio calculations were all performed by InStat (PrismGraph Software Inc., San Diego, CA, USA). Linkage-based association analysis was performed using TDT (version 3) program in statistical analysis for genetic epidemiology (SAGE) (40).

Estimate of nucleotide diversity and statistic D

θ was estimated as the number of segregating sites, S, divided by a_1, where $a_1 = \sum_{i=1}^{25} 1/i = 3.8160$, divided by the number of nucleotides sequenced (23). π was estimated by $k = \sum_{j=1}^{S} 2p_j(1 - p_j)$, divided by $1 - (1/26) = 0.9615$, divided by the number of nucleotides sequenced, where p_j is the observed frequency of the jth SNP (23). Statistic D was calculated according to Tajima (23).

DNA manipulation: haplotype and reporter gene constructions

A bacterial artificial chromosome (BAC) clone (ID no. CIT987SK-1143A11; Genbank accession no. AC005880; size: 146 388 bp) containing the entire *SLC18A2* gene and having a genotype (haplotype) *A* in the *SLC18A2* promoter region was purchased from Research Genetics (now integrated into Invitrogen). The BAC DNA was isolated and purified according to the protocol provided by the company. A 5.4 kb *Bam* HI/*Not* I DNA fragment was freed from this clone and cloned into a pBluescript SK phagemid vector (Stratagene). A 741 bp fragment, corresponding to the immediate upstream region from the *Not* I site to the base pair prior to the ATG methionine codon, was amplified from an individual's genomic DNA whose genotype of the *SLC18A2* promoter region was homozygote *A*, the same as the BAC clone. The 5′ primer used was 5′-CGCTGGGTCCGGGGAAGAGG-3′ and 3′ primer 5′-GGGGGGCGGCCGCGGATCCGGCTCC GGGGCTCCGCTGCG-3′, where a *Bam* HI site (underlined) was designed for shuttling. This PCR product was digested with *Not* I and ligated into the *Not* I site which was located at the 3′ end of the 5.4 kb fragment, yielding a 6131 bp fragment that represented the 6131 bp region upstream of the ATG codon or contained the 5560 bp region upstream of the exon 1 of *SLC18A2* gene. The resulted plasmid was termed pBSv2p6kbA. Another haplotype, *C*, had eight polymorphisms when compared with *A*. To get the genotype *C* of an equivalent clone, we amplified four fragments from a homozygote *C* individual's genomic DNA progressively from the 3′ end to the 5′ end. These four fragments were (i) the aforementioned 741 bp fragment containing SNPs −62G/A and −103A/C and restriction sites *Not* I/*Bam* HI for shuttling (same primers); (ii) a 1534 bp fragment containing SNP −1700C/T and restriction sites *Swa* I/*Sfi* I (5′ primer 5′-CATGGCAGTTTTG TATTCTT-3′ and 3′ primer 5′-TATGGTGGGGAGTTGGGG AGGTGAGGGT-3′); (iii) a 1404 bp fragment containing an indel $-2108(TAA)_{1/2}$, SNP −2504T/C and restriction sites *Hind* III/*Swa* I (5′ primer 5′-GCTCCTGCATGGCCTCCCC-3′ and 3′ primer 5′-AACACTGGAAGCCACACCAAACTCC-3′) and (iv) a 2176 bp fragment containing another indel $-4280(TG)_{2/3}$, SNP −4517A/C, SSLP −5199(38 bp or

Table 5. PCR primers and extension probes for genotyping the four SNPs that define the most frequent haplotypes

SNP	Oligonucleotide and amplicon size	T_m and length
−14234G/A	5′ primer 5′-TCAACGCTATGCTGACCAAGA	58°C, 20mer
	3′ primer 3′-AGCTTCAGGGTTTCCTCCACTA	58°C, 22mer
	Amplicon = 150 bp	
	Probe 5′-AGCTGTGAGGCTCCCGGC	62°C, 18mer
−11465A/G	5′ primer 5′-GGAAGTGAAAAATATTCACCCTGGTA	60°C, 26mer
	3′ primer 3′-GATGGACCAGTTCTGAATGGAAGT	59°C, 24mer
	Amplicon = 119 bp	
	Probe 5′-CTGTCAACACCATGGTGGGTT	64°C, 20mer
−2504T/C[a]	5′ primer 5′-CCCCAGCCTTCCAAGTAACTG	50°C, 21mer
	3′ primer 5′-GGACATAAAGATGGCAACAACAGA	59°C, 24mer
	Amplicon = 404 bp	
	Sequencing primer 3SNPf2: 5′-TTTTGTTTTGGATTCAGGGG	
−1700C/T	5′ primer 5′-CCATAGTGCCTGGCACGTAGTA	59°C, 22mer
	3′ primer 5′-GCATAGGCAGCTCTTAGCACAGT	59°C, 23mer
	Amplicon = 114 bp	
	Probe 5′-ATGAGGGGCCGGCACTTA	62°C, 18mer

[a]This SNP did not work on Sequenom. We obtained the genotyping information by sequencing the PCR product using the sequencing primer 3SNPf2, which is 59 bp upstream of the SNP.

a)/(42 bp or *b*) and restriction sites *Bam* HI/*Kpn* I (5′ primer 5′-GGTGCCCATTTCTTGGTTTCTTTTCAC-3′ and 3′ primer 5′-GCACATTTTCTTTCCGCTTCCTCTTC-3′). Progressive shuttling of these four PCR fragments into pBSv2p6kbA produced plasmids pBSv2p6kbAC1, pBSv2p6kbAC2, pBSv2p6kbAC3 and pBSv2p6kbC. pBSv2p6kbAC1 had *C* alleles at −62 and −103; pBSv2p6kbAC had *C* alleles at all eight sites. Cloning the 6131 bp *Bam* HI-fragment in each of the five pBluescript plasmids into *Bgl* II site of pGL3-enhancer (Promega, Madison, WI, USA) generated five expression plasmids pGL3e-hVMAT2-6.1A (A), pGL3e-hVMAT2-6.1C1 (*C*1), pGL3e-hVMAT2-6.1C2 (*C*2), pGL3e-hVMAT2-6.1C3 (*C*3) and pGL3e-hVMAT2-6.1C (*C*).

Transfection and luciferase activity measurements

Purified plasmid DNAs were introduced into SH-SY5Y and IMR-32 cells using Superfect (Qiagen). SH-SY5Y and IMR-32 cells were cultured in DMEM (Dulbecco's modified eagle medium) containing 15% fetal bovine serum (FBS, Life-science Technology, now integrated into Invitrogen) and Vita-Cell minimal medium containing 10% FBS (ATCC, Manassas, VA, USA), respectively, in 5% CO_2 and 100% humidity at 37°C. Cells were split at 1:2 into six-well plates with 30–40% confluence. On the second day, wells were replaced with 1 ml of fresh medium, and cells were transfected 2 h later. Transfection medium was prepared using 2 μg of plasmid DNA mixed at molar ratio 50:1 with control plasmid pTK-RL DNA (Promega) in 100 μl, mixed briefly with 12 μl of Superfect, vortexed for 10 s, incubated at 22°C for 20 min, followed by mixed with 0.6 ml of the normal growth medium. Transfections were begun by replacing media in each well with this mixture, gently covering all the cells, and the transfection mixture/cells were incubated for another 2 days. On the third day, 2 ml of fresh medium was added into each well. On the fourth day, cells in each were washed with 1 ml of phosphate-buffered saline and harvested for luciferase activity measurements using a Dual Luciferase Reporter System Kit (Promega), according to manufacturer's instructions.

Genotyping assays

The four SNPs were genotyped on Sequenom for −14234G/A, −11465A/G and −1700C/T or by sequencing of PCR products for −2504T/C, using oligonucleotides listed in Table 5.

Sequenom Mass Array SNP genotyping was performed according to the manufacturer' instructions (Sequenom, Inc., San Diego, CA, USA). A total of 2.5 ng of COGA genomic DNA was used as a template and 114–150 bp amplicons containing SNPs were obtained using appropriate primers. For PCR, 2.5 ng of genomic DNA was included as templates in 10 μl reaction solution, which contained 0.125 mM dNTPs each, 0.625 μM primers each, 1.25 mM $MgCl_2$, 2% DMSO, 0.05 U *AmpliTaq* GOLD DNA polymerase and 1x reaction buffer (Applied Biosystems, Inc., Foster City, CA, USA). PCR used cycles of 95°C for 3 min, 45 cycles of 94°C for 30 s, 58–61°C for 75 s and 72°C for 25 s, followed by 72°C for 8 min. The amplicons provided templates for primer extension reactions using appropriate oligonucleotide probes. Primer extension products were assessed by MALDI-TOF (Sequenom). For −2504T/C sequencing, primers in the PCR product were digested by ExoSAP-IT (USB, Cleveland, OH, USA). Each 10 μl DNA sequencing reaction contained 7 μl ExoSAP-IT elute, 0.75 μM 3SNPf2, 1× buffer and 1 μl BigDye Terminator v1.1 (ABI). Sequencing cycles used were 96°C for 30 s, 25 cycles of 96°C for 15 s, 50°C for 5 s and 60°C for 4 min, followed by 60°C for 2 min. After addition of 10 μl ddH_2O into the reaction to reach 20 μl, sequencing primers were cleaned up using the DTR V3 Short Plate Kit (catalog no. 89939, Edge Biosystems, MD, USA). The sequencing reactions were denatured at 95°C for 2 min before capillary electrophoresis on 3100 Genetic Analyzer (ABI).

SLC18A2 SSLP at −5 kb was genotyped by PCR amplification using 6-FAM labeled 5′ primer 5′-GCAATGACCATCC TTATCTATGC-3′ and 3′ primer 5′-CGTGGCTGGAGAGTAG GAAG-3′, *AmpliTaq* Gold *Taq* polymerase, an initial 95°C denaturation for 10 min followed by 35 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 45 s and a final extension at 72°C for 10 min. Alleles, 160–177 bp, were resolved using 10% denaturing polyacrylamide gels, an ABI 373 sequencer and Gene Scan software.

ACKNOWLEDGEMENTS

We acknowledge support from NIDA-IRP and especially acknowledge coworkers Frances H. Gabbay for providing human genomic DNAs, Xuguang Zhu, Masatoshi Hazama, Seth Axelard, and Judy Hess for assistance during this work; Mary Pfeiffer for critical reading of this manuscript.

Conflict of Interest statement: None declared.

REFERENCES

1. Erickson, J.D., Schafer, M.K., Bonner, T.I., Eiden, L.E. and Weihe, E. (1996) Distinct pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter. *Proc. Natl Acad. Sci. USA*, **93**, 5166–5171.
2. Gonzalez, A.M., Walther, D., Pazos, A. and Uhl, G.R. (1994) Synaptic vesicular monoamine transporter expression: distribution and pharmacologic profile. *Brain Res. Mol. Brain Res.*, **22**, 219–226.
3. Xu, W. and Emson, P.C. (1997) Neuronal stem cells express vesicular monoamine transporter 2 immunoreactivity in the adult rat. *Neuroscience*, **76**, 7–10.
4. Harrington, K.A., Augood, S.J., Kingsbury, A.E., Foster, O.J. and Emson, P.C. (1996) Dopamine transporter (Dat) and synaptic vesicle amine transporter (VMAT2) gene expression in the substantia nigra of control and Parkinson's disease. *Brain Res. Mol. Brain Res.*, **36**, 157–162.
5. Pickel, V.M., Chan, J. and Nirenberg, M.J. (2002) Region-specific targeting of dopamine D2-receptors and somatodendritic vesicular monoamine transporter 2 (VMAT2) within ventral tegmental area subdivisions. *Synapse*, **45**, 113–124.
6. Takahashi, N., Miner, L.L., Sora, I., Ujike, H., Revay, R.S., Kostic, V., Jackson-Lewis, V., Przedborski, S. and Uhl, G.R. (1997) VMAT2 knockout mice: heterozygotes display reduced amphetamine- conditioned reward, enhanced amphetamine locomotion, and enhanced MPTP toxicity. *Proc. Natl Acad. Sci. USA*, **94**, 9938–9943.
7. Johnson-Davis, K.L., Truong, J.G., Fleckenstein, A.E. and Wilkins, D.G. (2004) Alterations in vesicular dopamine uptake contribute to tolerance to the neurotoxic effects of methamphetamine. *J. Pharmacol. Exp. Ther.*, **309**, 578–586.
8. Eiden, L.E., Schafer, M.K., Weihe, E. and Schutz, B. (2004) The vesicular amine transporter family (*SLC18*): amine/proton antiporters required for vesicular accumulation and regulated exocytotic secretion of monoamines and acetylcholine. *Pflugers Arch.*, **447**, 636–640.
9. Fon, E.A., Pothos, E.N., Sun, B.C., Killeen, N., Sulzer, D. and Edwards, R.H. (1997) Vesicular transport regulates monoamine storage and release but is not essential for amphetamine action. *Neuron*, **19**, 1271–1183.
10. Travis, E.R., Wang, Y.M., Michael, D.J., Caron, M.G. and Wightman, R.M. (2000) Differential quantal release of histamine and 5-hydroxytryptamine from mast cells of vesicular monoamine transporter 2 knockout mice. *Proc. Natl Acad. Sci. USA*, **97**, 162–167.
11. Wang, Y.M., Gainetdinov, R.R., Fumagalli, F., Xu, F., Jones, S.R., Bock, C.B., Miller, G.W., Wightman, R.M. and Caron, M.G. (1997) Knockout of the vesicular monoamine transporter 2 gene results in neonatal death and supersensitivity to cocaine and amphetamine. *Neuron*, **19**, 1285–1296.
12. Hall, F.S., Sora, I. and Uhl, G.R. (2003) Sex-dependent modulation of ethanol consumption in vesicular monoamine transporter 2 (VMAT2) and dopamine transporter (DAT) knockout mice. *Neuropsychopharmacology*, **28**, 620–628.
13. Lu, W. and Wolf, M.E. (1997) Expression of dopamine transporter and vesicular monoamine transporter 2 mRNAs in rat midbrain after repeated amphetamine administration. *Brain Res. Mol. Brain Res.* **49**, 137–148.
14. Rusnak, M., Kvetnansky, R., Jelokova, J. and Palkovits, M. (2001) Effect of novel stressors on gene expression of tyrosine hydroxylase and monoamine transporters in brainstem noradrenergic neurons of long-term repeatedly immobilized rats. *Brain Res.*, **899**, 20–35.
15. Richardson, J.R. and Miller, G.W. (2004) Acute exposure to aroclor 1016 or 1260 differentially affects dopamine transporter and vesicular monoamine transporter 2 levels. *Toxicol. Lett.* **148**, 29–40.
16. Surratt, C.K., Persico, A.M., Yang, X.D., Edgar, S.R., Bird, G.S., Hawkins, A.L., Griffin, C.A., Li, X., Jabs, E.W. and Uhl, G.R. (1993) A human synaptic vesicle monoamine transporter cDNA predicts posttranslational modifications, reveals chromosome 10 gene localization and identifies TaqI RFLPs. *FEBS Lett.*, **318**, 325–330.
17. Xu, W., Liu, L., Mooslehner, K. and Emson, P.C. (1997) Structural organization of the human vesicular monoamine transporter type-2 gene and promoter analysis using the jelly fish green fluorescent protein as a reporter. *Brain Res. Mol. Brain Res.*, **45**, 41–49.
18. Glatt, C.E., DeYoung, J.A., Delgado, S., Service, S.K., Giacomini, K.M., Edwards, R.H., Risch, N. and Freimer, N.B. (2001) Screening a large reference sample to identify very low frequency sequence variants: comparisons between two genes. *Nat. Genet.*, **27**, 435–438.
19. Kunugi, H., Ishida, S., Akahane, A. and Nanko, S. (2001) Exon/intron boundaries, novel polymorphisms, and association analysis with schizophrenia of the human synaptic vesicle monoamine transporter (SVMT) gene. *Mol. Psychiatry*, **6**, 456–460.
20. Iwasa, H., Kurabayashi, M., Nagai, R., Nakamura, Y. and Tanaka, T. (2001) Multiple single-nucleotide polymorphisms (SNPs) in the Japanese population in six candidate genes for long QT syndrome. *J. Hum. Genet.*, **46**, 158–162.
21. Burman, J., Tran, C.H., Glatt, C., Freimer, N.B. and Edwards, R.H. (2004) The effect of rare human sequence variants on the function of vesicular monoamine transporter 2. *Pharmacogenetics*, **14**, 587–594.
22. Erickson, J.D. and Eiden, L.E. (1993) Functional identification and molecular cloning of a human brain vesicle monoamine transporter. *J. Neurochem.*, **61**, 2314–2317.
23. Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
24. Ikeda, K., Nagano, K. and Kawakami, K. (1993) Anomalous interaction of Sp1 and specific binding of an E-box-binding protein with the regulatory elements of the Na,K-ATPase alpha 2 subunit gene promoter. *Eur. J. Biochem.*, **218**, 195–204.
25. Pages, G., Stanley, E.R., Le Gall, M., Brunet, A. and Pouyssegur, J. (1995) The mouse p44 mitogen-activated protein kinase (extracellular signal-regulated kinase 1) gene Genomic organization and structure of the 5′ -flanking regulatory region. *J. Biol. Chem.*, **270**, 26986–26992.
26. Hu, H.M., Arcinas, M. and Boxer, L.M. (2002) A Myc-associated zinc finger protein-related factor binding site is required for the deregulation of c-myc expression by the immunoglobulin heavy chain gene enhancers in Burkitt's lymphoma. *J. Biol. Chem.*, **277**, 9819–9824.
27. Little, K.Y., Krolewski, D.M., Zhang, L. and Cassin, B.J. (2003) Loss of striatal vesicular monoamine transporter protein (VMAT2) in human cocaine users. *Am. J. Psychiatry*, **160**, 47–55.
28. Zucker, M., Aviv, A., Shelef, A., Weizman, A. and Rehavi, M. (2002) Elevated platelet vesicular monoamine transporter density in untreated patients diagnosed with major depression. *Psychiatry Res.*, **112**, 251–256.
29. Zubieta, J.K., Huguelet, P., Ohl, L.E., Koeppe, R.A., Kilbourn, M.R., Carr, J.M., Giordani, B.J. and Frey, K.A. (2000) High vesicular monoamine transporter binding in asymptomatic bipolar I disorder: sex differences and cognitive correlates. *Am. J. Psychiatry*, **157**, 1619–1628.
30. Schwartz, K., Yadid, G., Weizman, A. and Rehavi, M. (2003) Decreased limbic vesicular monoamine transporter 2 in a genetic rat model of depression. *Brain Res.*, **965**, 174–179.
31. Sullivan, P.F., Neale, B.M., van den Oord, E., Miles, M.F., Neale,, M.C., Bulik, C.M., Joyce, P.R., Straub, R.E. and Kendler, K.S. (2004) Candidate genes for nicotine dependence via linkage, epistasis, and bioinformatics. *Am. J. Med. Genet.*, **126B**, 23–36.
32. Kageyama, R.and Pastan, I. (1989) Molecular cloning and characterization of a human DNA binding factor that represses transcription. *Cell*, **59**, 815–825.

33. Niu, T., Qin, Z.S., Xu, X. and Liu, J.S. (2002) Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, **70**, 157–169.
34. Qin, Z.S., Niu, T. and Liu, J.S. (2002) Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, **71**, 1242–1247.
35. Abecasis, G.R. and Cookson, W.O. (2000) GOLD—graphical overview of linkage disequilibrium. *Bioinformatics*, **16**, 182–183.
36. Zhao, J.H. (2002) 2LD—two-locus linkage disequilibrium (LD) calculator, http://www.iop.kcl.ac.uk/IoP/Departments/PsychMed/GEpiBSt/software.shtml.
37. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
38. Page, R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.*, **12**, 357–358.
39. Sham, P.C. and Curtis, D. (1995) Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann. Hum. Genet.*, **59**, 97–105.
40. SAGE (2003) Statistical analysis for genetic epidemiology. Statistical Solutions, Cork, Ireland.