ความหลากหลายของลำดับเบสของดีเอนเอและลิงค์เกจดิสอิควิลิเบรียมของยีน GABA Transporter-1 (*SLC6A1*) ใน 5 กลุ่มประชากรเพื่อการประยุกต์ใช้ในงานวิจัยด้านเภสัชพันธุศาสตร์

นางสาว รุ่งนภา หิรัญสถิต

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีวเวชศาสตร์ (สหสาขาวิชา)
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

SEQUENCE VARIATION AND LINKAGE DISEQUILIBRIUM IN THE GABA TRANSPORTER-1 GENE (*SLC6A1*) IN FIVE POPULATIONS: IMPLICATIONS FOR PHARMACOGENETIC RESEARCH

Miss Rungnapa Hirunsatit

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Biomedical Sciences
(Interdisciplinary Program)
Graduate School
Chulalongkorn University
Academic Year 2007

Thesis Title	SEQUENCE VARIATION AND LINKAGE DISEQUILIBRIUM IN THE GABA TRANSPORTER-1 GENE (*SLC6A1*) IN FIVE POPULATIONS: IMPLICATIONS FOR PHARMACOGENETIC RESEARCH
By	Miss Rungnapa Hirunsatit
Field of Study	Biomedical Sciences
Thesis Advisor	Professor Apiwat Mutirangura
Thesis Co-advisor	Professor Joel Gelernter

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

..... Dean of the Graduate School
(Assistant Professor M.R. Kalaya Tingsabadh, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Assistant Professor Tewin Tencomnao, Ph.D.)

..... Thesis Advisor
(Professor Apiwat Mutirangura, M.D., Ph.D.)

..... Thesis Co-advisor
(Professor Joel Gelernter, M.D.)

..... External Member
(Associate Professor Vip Viprakasit, M.D., D. Phil.)

..... Member
(Professor Vorasuk Shotelersuk, M.D.)

..... Member
(Assistant Professor Kanya Suphapeetiporn, M.D., Ph.D.)

..... Member
(Assistant Professor Rasmon Kalayasiri, M.D.)

รุ่งนภา หิรัญสถิต : ความหลากหลายของลำดับเบสของดีเอนเอและลิงค์เกจดิสอิควิลิเบรียมของยีน GABA Transporter-1 *(SLC6A1)* ใน 5 กลุ่มประชากรเพื่อการประยุกต์ใช้ในงานวิจัยด้านเภสัชพันธุศาสตร์ (SEQUENCE VARIATION AND LINKAGE DISEQUILIBRIUM IN THE GABA TRANSPORTER-1 GENE (*SLC6A1*) IN FIVE POPULATIONS: IMPLICATIONS FOR PHARMACOGENETICS RESEARCH) **อ. ที่ปรึกษา: ศ.ดร.อภิวัฒน์ มุทิรางกูร**, 63 หน้า.

โปรตีน GABA transporter-1 (GAT-1) ถูกสร้างขึ้นโดยยีน *SLC6A1* เป็นเป้าหมายใหม่ที่น่าสนใจสำหรับการรักษาโรคทางจิตเวช การศึกษาความแตกต่างทางพันธุกรรมของคนจากกลุ่มประชากรที่แตกต่างกันจะช่วยให้ความเข้าใจในเชิงเภสัชพันธุศาสตร์ได้ดียิ่งขึ้น โดยทำการศึกษารูปแบบความหลากหลายทางพันธุกรรมและลิงค์เกจดีสอีควิลิเบรียมของยีน *SLC6A1* ในประชากร 5 กลุ่ม จาก 3 ทวีปเปรียบเทียบกัน พิจารณาความแตกต่างของลำดับนิวคลีโอไทด์บนจีโนมความยาวรวมทั้งสิ้น 12.4 กิโลเบส โดยซีเควนซ์ส่วนที่เป็นโปรโมเตอร์ เอกซอน และบางส่วนของอินทรอนที่อยู่ติดกับเอกซอน ในกลุ่มตัวอย่างชาวอเมริกันเชื้อสายแอฟริกา ชาวอเมริกันเชื้อสายยุโรป ชาวไทย ชาวมัง ชาวฟินนิช (ฟินแลนด์) จำนวนรวมทั้งสิ้น 40 คน เพื่อหาความแตกต่างของลำดับนิวคลีโอไทด์ในบริเวณที่ต้องการจะศึกษา โดยเฉพาะที่เป็นโพลีมอร์ฟิซึ่มของนิวคลีโอไทด์เดี่ยว หรือสนิพ (single nucleotide polymorphism) ความหลากหลายที่พบรวมทั้งสิ้น 63 ตำแหน่ง และนอกจากนั้นยังทำการศึกษาลิงเกจดิสอีควิลิเบรียม เพื่อหารูปแบบของการเรียงตัวของ สนิพ 16 ตำแหน่งที่ครอบคลุม ยีน *SLC6A1* ในประชากรกลุ่มที่ใหญ่ขึ้นเปรียบเทียบกันทุกกลุ่มประชากร จากการศึกษาพบว่าในกลุ่มชาวอเมริกันเชื้อสายแอฟริกาพบโพลีมอร์ฟิซึ่มที่มีลักษณะจำเพาะ โดยเฉพาะอย่างยิ่งพบ VNTR (variable number tandem repeat) ที่มีการเพิ่มขึ้นของลำดับเบสที่ซ้ำกันจำนวน 21 เบสที่บริเวณที่เป็นโปรโมเตอร์ของยีน ซึ่งไม่ปรากฏว่ามีในประชากรกลุ่มอื่น เมื่อพิจารณาที่ค่าลิงค์เกจดีสอีควิลิเบรียมของยีนนี้พบว่าน้อยมากในทุกกลุ่มประชากร โดยในกลุ่มอเมริกันเชื้อสายแอฟริกามีความหลากหลายทางพันธุกรรมมากที่สุด และไม่พบความแตกต่างของรูปแบบของลิงค์เกจดีสอีควิลิเบรียมในกลุ่มอเมริกันเชื้อสายยุโรปและกลุ่มชาวฟินนิช ส่วนในกลุ่มประชากรมังแม้จะมีค่าของลิงค์เกจดีสอีควิลิเบรียมสูงที่สุด เมื่อพิจารณาจากการเรียงตัวของสนิพในบริเวณดังกล่าว แต่ข้อมูลที่ได้จากวิธีนี้ก็ไม่สามารถที่จะหาสนิพที่เป็นตัวแทนของสนิพตัวอื่นในขอบเขตลิงค์เกจดีสอีควิลิเบรียมเดียวกัน (Tag SNP) ได้ นอกจากที่กล่าวมายังได้ทำการศึกษาการทำงานของชิ้นส่วนดีเอนเอที่มีและไม่มีส่วนของจำนวนซ้ำของ 21 เบส เปรียบเทียบกัน และเปรียบเทียบกับคอนสตรัคที่ไม่มีส่วนโปรโมเตอร์นี้อยู่เลย พบว่าส่วนที่มีการเพิ่มขึ้นของจำนวนซ้ำของ 21 เบสจะมีการทำงานสูงขึ้นอย่างมีนัยสำคัญเมื่อเทียบกับคอนสตรัคที่ไม่มี และเมื่อใช้วิธี Gel Shift Assays สามารถตรวจพบโปรตีนที่มาเกาะส่วนของลำดับเบสดังกล่าว ซึ่งอาจเป็นไปได้ว่าบริเวณดังกล่าวมีส่วนในการควบคุมการทำงานของยีนนี้ ค่าลิงค์เกจดีสอีควิลิเบรียมที่ค่อนข้างต่ำและการที่บริเวณของยีน *SLC6A1* อาจเป็นบริเวณที่เกิดรีคอมไบเนชั่นค่อนข้างสูง ยีนนี้จึงเป็นตัวอย่างหนึ่งของปัญหาที่เกิดกับการศึกษาความสัมพันธ์ทางพันธุศาสตร์ที่อาศัยแฮพโพลไทป์ อย่างไรก็ตามการเกิดจำนวนซ้ำของ 21 เบส อาจมีคุณสมบัติทำให้การทำงานของยีนนี้เพิ่มขึ้นและมีนัยสำคัญต่อการตอบสนองของระบบประสาท GABA ต่อการรักษาด้วยยาในกลุ่มที่มีเป้าหมายที่โปรตีน GABA transporter-1 นี้ที่แตกต่างกันในกลุ่มประชากรที่มีเชื้อสายแอฟริกา

สาขาวิชา ชีววิทยาศาสตร์ **ลายมือชื่อนิสิต** รุ่งนภา หิรัญสถิต

ปีการศึกษา 2550 **ลายมือชื่ออาจารย์ที่ปรึกษา** .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

4789718520: MAJOR BIOMEDICAL SCIENCES

KEY WORD: GABA TRANSPORTER-1 / GAT-1 / *SLC6A1* / LINKAGE DISEQUILIBRIUM / PHARMACOGENETICS

RUNGNAPA HIRUNSATIT: SEQUENCE VARIATION AND LINKAGE DISEQUILIBRIUM IN THE GABA TRANSPORTER-1 GENE (*SLC6A1*) IN FIVE POPULATIONS: IMPLICATIONS FOR PHARMACOGENETICS RESEARCH. THESIS ADVISOR: PROF APIWAT MUTIRANGURA, M.D., Ph.D., 63 pp.

GABA transporter-1 (GAT-1; genetic locus *SLC6A1*) is emerging as a novel target for treatment of neuropsychiatric disorders. To understand how population differences might influence strategies for pharmacogenetic studies, we identified patterns of genetic variation and linkage disequilibrium (LD) in *SLC6A1* in five populations representing three continental groups. We resequenced 12.4 kb of *SLC6A1*, including the promoters, exons and flanking intronic regions in African-American, Thai, Hmong, Finnish, and European-American subjects (total n=40). LD in *SLC6A1* was examined by genotyping 16 SNPs in larger samples. Sixty-three variants were identified through resequencing. Common population-specific variants were found in African-Americans, including a novel 21-bp promoter region variable number tandem repeat (VNTR), but no such variants were found in any of the other populations studied. Low levels of LD and the absence of major LD blocks were characteristic of all five populations. African-Americans had the highest genetic diversity. European-Americans and Finns did not differ in genetic diversity or LD patterns. Although the Hmong had the highest level of LD, our results suggest that a strategy based on the use of tag SNPs would not translate to a major improvement in genotyping efficiency. In addition, we performed reporter assays using *SLC6A1* promoter constructs containing the insertion or non-insertion variants or using a promoterless construct. The insertion variant showed significantly more activity than the non-insertion promoter variant in multiple cell lines. Gel shift assays determined that nuclear proteins interact with the insertion sequence.

Owing to the low level of LD and presence of recombination hotspots, *SLC6A1* may be an example of a problematic gene for association and haplotype tagging-based genetic studies. However, the 21-bp insertion polymorphism leads to a dramatic increase in *SLC6A1* promoter activity based on *in vitro* experiment. This polymorphism may prove useful in predicting clinical response to pharmacological modulators of SLC6A1 as well as GABAergic function in individuals of African descent.

Field of study: Biomedical Sciences

Academic year: 2007

Student's signature: Rungnapa Hirunsatit

Advisor's signature: Apiwat Mutirangura

Co-advisor's signature:

ACKNOWLEDGEMENTS

This project cannot be completed without all voluntary blood donors in Thailand, Finland and African-Americans and European-Americans in USA. Moreover, all supporting grants were from the National Institutes of Health (The Thai-US Drug Dependence Genetic Training Program), the Alcoholic Beverage Medical Research Foundation (ABMRF), the US Department of Veterans Affairs [VA Alcohol Research Center, VA Mental Illness Research, Education and Clinical Center (VA MIRECC), and VA Veterans Research Enhancement Award Program (VA REAP)] and Thailand Research Fund. Technical help of Greg Dalton-Key, AnnMarie Lacobelle Wantroba and Katarzyna Kanarek is also greatly appreciated. Very special thanks to my parents and my husband (Chupong Ittiwut) for all their supporting.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

LIST OF ABBREVIATIONS

GABA	Gamma-aminobutyric acid
GAT	GABA transporter
BGT-1	betaine/GABA transporter-1
SLC6A1	solute carrier family 6 (neurotransmitter transporter, GABA), member 1
LD	linkage disequilibrium
SNPs	single nucleotide polymorphisms
ANOVA	analysis of variance
-MR	recombination model
-X10	run 10 times
π	nucleotide diversity
θ	Watterson's estimator of zeta
EA	European-American
AA	African-American
HWE	Hardy-Weinberg equilibrium
LOD	log of the likelihood odds ratio

CHAPTER I

INTRODUCTION

Background and rationale

Gamma-aminobutyric acid (GABA) is the most ubiquitous inhibitory neurotransmitter in the brain. Abnormal function of the GABA system has been implicated in almost every common neurological and psychiatric disorder. Augmentation of brain GABA function is the presumed therapeutic mechanism of several classes of medications, including the benzodiazepines, gabapentin, and pregabalin, which are used in treatment of many psychiatric and neurological disorders, addictions, and pain (1, 2). Identifying genetic factors responsible for variation in clinical response holds promise as a way to improve the clinical application of these medications; it may be possible to identify those who are most likely to respond to treatment and those who are at risk to develop adverse effects. For these kinds of studies, it will be important to identify polymorphisms in genes encoding the components of the brain GABA system. Although the HapMap project (3) will identify the major basic blocks of linkage disequilibrium in these genes, detailed patterns of genetic variation, including the discovery of novel functional variants, can only be accomplished through resequencing.

A novel class of medications was recently developed that act by blockade of GABA transporters (GAT), thereby inhibiting GABA uptake. Four main subtypes of GABA transporters, the GAT-1, GAT-2, GAT-3 and Betaine/GABA transporter-1 (BGT-1), have been identified through molecular cloning. The first medication in the GABA reuptake inhibitor class 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 GAT-1 sites (4). Tiagabine was developed for treatment of seizure disorders but is being used by many psychiatrists to treat a variety of conditions in which augmentation of brain GABA function is thought to be desirable to alleviate clinical symptoms. Preliminary studies suggest that tiagabine is effective in treatment of anxiety (5), sleep disorders (6), depression (7) and addictions (2, 8, 9). Animal studies suggest that most of the tiagabine's common adverse effects, such as muscle twitching and sedation,

arise from specific inhibition of GAT-1 (10). Tiagabine has a rare, but serious adverse effect - non-convulsive status epilepticus - which has raised concerns about its off-label use in treatment of psychiatric disorders (11). Pharmacogenetic study could plausibly help to identify those subjects who are at greatest risk for developing side-effects to tiagabine and other medications that inhibit GAT-1.

The *SLC6A1* gene, which encodes the GAT-1 protein, is an obvious candidate for pharmacogenetic studies of tiagabine and other GAT-1 inhibitors (12). In anticipation of larger pharmacogenetic studies, we aimed to identify novel genetic variation and examine linkage disequilibrium in the *SLC6A1* gene in five populations representing three major continental groups [European (EA and Finnish), African (AA) and Asian (Thai and Hmong)]. Populations considered isolated, Finnish and Hmong, and two mixed populations, EA and AA, were examined, to understand how population differences should influence planning for pharmacogenetic studies. By comparing isolated and mixed populations, we hoped to shed further light on the purported benefits of isolated populations in mapping complex genetic traits (13, 14). Furthermore, our goal was to identify a set of haplotype tagging markers for studies focusing on response to GAT-1 inhibition.

Research questions

- From the sequencing result of promoters, all coding regions and flanking intronic regions of *SLC6A1* gene, which polymorphism that we found? Is there any population specific variant?
- How about the pattern of LD and tagging SNP of *SLC6A1* in all 5 populations?
- Can we find the population specific SNP or haplotype that possible response to GAT-1 inhibitors?

Objectives

To identify population specific polymorphism of *SLC6A1* gene and investigate potential candidate polymorphism to respond to GAT-1 inhibitors.

Hypothesis

If we find population specific SNP or haplotype consequence to *SLC6A1* gene expression, these polymorphisms are possible to present difference response to GAT-1 inhibitors in all 5 populations.

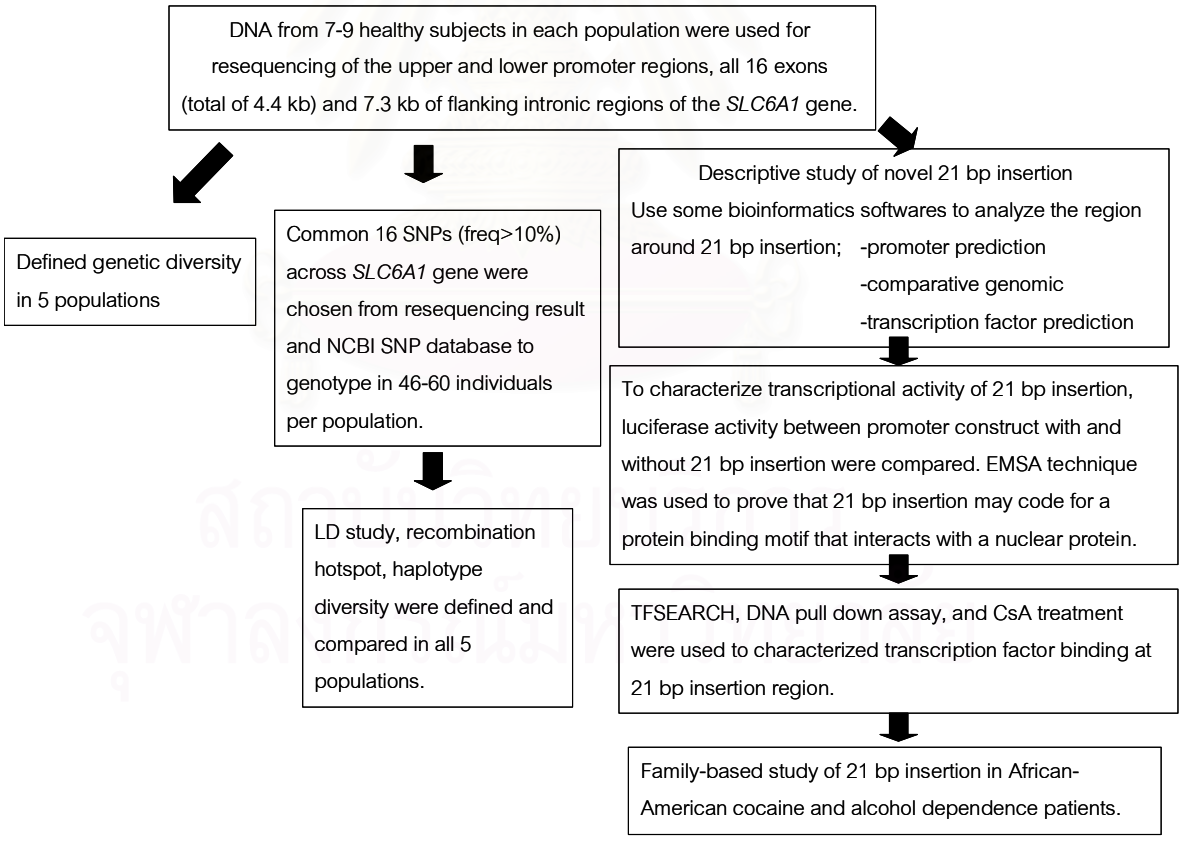
Key words

GABA transporter 1 (GAT-1), *SLC6A1*, linkage disequilibrium, pharmacogenetics

Expected benefit and application

We expect to see functional and population specific polymorphism of *SLC6A1* gene and we can apply this variant in clinical treatment of GAT-1 inhibitor in these 5 populations.

Methodology framework



CHAPTER II

REVIEWS AND RELATED LITERATURES

Pharmacogenetics

Among various types of human DNA variation including repeat sequence, ins/del polymorphism and copy number variation (CNV), also reveal in human genome. Single nucleotide polymorphisms (SNPs) were applied to variety of studies including pharmacogenetics. Regarding the Human Genome Project (HGP) (15), 2.1 million SNPs were found in every 1250 bp on average through the genome (0.1% of genome) and presented individually difference. It is very useful for human genome scanning for gene mapping and also benefit to the studies of evolution and inter-individual difference in drug response (16).

Different effects of drugs or agents on the treatment of different individuals are of interested for individualized medicine. The main goal of the pharmacogenetics or pharmacogenomics (screening whole genome) studies that try to define how a gene affects drug sensitivity/resistant (17). Several documents reported that genetic factors influence heterogeneity of individually response to treatments and could be inherited (18, 19). Drug-susceptible phenotype can be assumed as a kind of disease phenotype. Several approaches for disease mapping gene were capable to apply to pharmacogenetics study of drug-susceptible phenotype. More than one gene is contributed to complex phenotype. Polymorphism of drug metabolizing enzymes, *CYP450*, and targeted receptor proteins are crucial factors for drug physiological pathway (20). It is very interesting that some of these polymorphisms are specific in some geographic population but not found in others. Thus these specific polymorphisms contributable to drug-susceptibility phenotype are possibly useful only in a specific population. Although the international HapMap project (3) have already identified a major blocks of linkage disequilibrium in these genes, more detailed patterns of global genetic variation, including discovery of novel functional variation, can only be accomplished through resequencing. Expressed phenotype (level of protein and pharmacokinetic of protein) from candidate genotype

(haplotype) must be determined to further benefit to predict human responsiveness to some agents.

Screening polymorphisms of gene of interest is not difficult in post-genomic era because several tools are available. One kind of polymorphism, non-synonymous SNPs, was indicated to be functional SNPs because of changing amino acid. This variation is a priority to determine disease susceptibility and pharmacogenetics. Numerous recently studies revealed synonymous SNPs (silence mutations) can affect protein expression by involving in RNA stability (21, 22). There is requiring more data to support this hypothesis. However, in this study we focus on population specific polymorphisms (SNP and haplotype) and find out the specific LD pattern in each population for further high-density genetic maps to drug-susceptibility phenotype. Moreover, diversity of SNP frequency in different populations is further possible point to predict drug dose requirement in individual ethnicity.

Human genetic diversity

Genetic diversity can be devided into 2 levels; individual and population levels. The variation can present in nucleotide, gene and chromosome of individuals in population (23). This study tries to characterize diversity of *SLC6A1* gene in intra- and inter- populations; African-American, European-American, Finnish, Thai and Hmong. The genetics diversity found in this study can possibly be applied to the treatment of GAT-1 inhibitor in these 5 populations. We hope to see correlation between some genetic variations and different responses to agents in different ethnic groups. However, functional genetic variation in terms of SNPs or haplotype that affect protein level should be first priority to be classified.

In addition, genetic diversity in our populations may reflect human evolution since our sample populations were recruited from 3 major continents. The differences between these populations are not only influenced by genetic factors but also environmental factors. For example, the people in the area of high intensity of UV radiation will have dark skin color (high level of melanin) to protect their skins. This may be the effect of natural selection or adaptive advantage on genetic

variations specific in different geographic area. Recently data presented African population the world oldest population which passed several times of genetic recombination event. Africa has more diversity than non-African populations. In another way, new populations were migrated out of Africa and bring only few alleles of gene to settle the new population. That make non-African populations has less diversity than Africa (24).

From our result we found 21-bp insertion that present in only African-American and other populations in Africa (Tanzania and Zambia). It's possible that this variation happen long time ago in African population before migrated out of Africa. We screen in other European populations; Russia and Finnish. No this variation was presented. We don't know yet what is a main factor to eliminate this variation from non-African populations.

Nucleotide diversity (π) is a statistic for single population sample and represents average of pairwise nucleotide difference in the sample of (25).

$$\pi = (N/N-1) \sum_{ij} p_i p_j \pi_{ij}$$

N = number of sequences in sample

p_i and p_j are frequency of alleles in sample

π_{ij} = the proportion of different site in sequence i and j

Sample calculation for nucleotide diversity (π)

		Frequency
A	AATCGATCGCCCATTCTGTG	0.3
B	AATCTATCGACCATTCTATG	0.5
C	AATCGATCGACCAGTCTATG	0.2

A and B A and C B and C

π = (3/2) {[0.3*0.5*(3/20)] + [0.3*0.2*(3/20)]+[0.5*0.2*(2/20)]}

= (3/2){0.0225+0.009+0.01}

= 0.06225

Ardlie KG, Kruglyak L and Seielstad M (26) reported correlation between LD (D' or r^2) with nucleotide diversity with a bottleneck effect via simulations based on starting at 1,000 generations ago (coalescent theory), effective population size equal 10,000 (N_e) and exponentially expansion until 5 billion currently. They conclude that LD increases but nucleotide decreases with the severity of the bottleneck. Our sequencing result presented nucleotide diversity in exon (π_{EX} = 5.5-7.0 x 10^{-5}) lower than in intron (π_{INT} = 1.1-2.3 x 10^{-4}). However, these values are lower than normal diversity in previous screening in 49 human genes (π_{EX} = 1.1 x 10^{-3} and π_{INT} = 3-4 x 10^{-4}) (25). In this study we used another parameter, Watterson's estimator of theta (θ), to report diversity of *SLC6A1* sequence. This number is the observed number of SNPs adjusted for sample size and new mutation rate expected to occur in each generation following coalescent theory (27).

Linkage disequilibrium (LD) and recombination process

Alleles at two loci (or more) in same LD are found in the same haplotype because two (or more) link loci coexist in the recombination process. This always use in genetic in population and talk about less or more allele (haplotype) frequency in population. Alleles that coexist together when recombination happen that called all these alleles staying in same haplotype and in the same LD block. The length of LD block around genetic markers need to be defined because we try to narrow down the position of causative mutation. For example, a SNP showed association with a disease. We need to define LD block around this SNP to find out where is the real causative mutation. This SNP may be the real one or another polymorphism locating in this LD block. We can use this approach for mapping disease causing gene linked to some genetic markers. In addition, the pattern of LD block can be different in varied population. Knowing about basic phenomenon (LD block and recombination hotspot) around interested region in specific population is the priority knowledge for association mapping (28). The recombination hotspot region is an opposed event with LD. Thus, low LD region (high rate of recombination) is more difficult to mapping gene by using close-linked genetic markers.

Linkage disequilibrium between 2 SNPs (pairwise) can be measured by D' parameter which is a population-based parameter reflecting recombination in all of generations since the rise of the individual polymorphic alleles is under study (26, 29, 30). A simple measure of pairwise LD is presented below.

LD is $D_{ij} = x_{ij}-p_iq_j$ (31)

x_{ij} is the observed of frequency of haplotype A_iB_j,

p_i and q_j are the frequencies of allele A_i and B_j at loci A and B

If the expected frequency of gamete A_iB_j is p_iq_j, it can be implied that no statistic association between alleles adjacent. The normalized measure of D' (32) is widely preferred due to it is statistically corrected, as its range is the same for allele frequencies

D' is defined as $D'ij$; $D'_{ij} = D_{ij}/ D_{max}$,

$D_{max} = \min [p_iq_j, (1- p_i)(1- q_j)]$ when $D_{ij} <0$, or

$D_{max} = \min [p_i (1- q_j), (1- p_j) q_j]$ when $D_{ij} >0$

Another parameter that can use to define pairwise LD is r^2 (33).

$r^2 = D/(p_{Ai}p_{Aj}p_{Bi}p_{Bj})$

GABA transporter 1

Gamma-aminobutyric acid (GABA) is the major inhibitory neurotransmitter in the mammalian brain. GABA released from the presynaptic neurons binds with the $GABA_A$, $GABA_B$ and $GABA_C$ receptors on postsynaptic and presynaptic terminals. GABA is removed from the synaptic cleft by GABA transporters (GAT), which terminates GABA action. Four main subtypes of GABA transporters have been identified through molecular cloning and which can also be distinguished by their pharmacological properties. These isoforms are *GAT-1, GAT-2, GAT-3 and BGT-1*. From in situ hybridization in rat, GABA transporter 1 (GAT-1) mRNA is the most ubiquitously expressed subtype in many areas of the brain; telencephalon in the olfactory bulb, diencephalons in the retina, midbrain, pons, medulla and cerebellum (34). The GAT-1 function is believed to be a key event in terminating synaptic GABA currents. For studies focusing on genetically influenced aberrant function of the

brain gamma-aminobutyric acid (GABA) system and variation in response to its pharmacological modulation, it will be important to identify polymorphisms in the genes that encode the components of the brain GABA system. GABA is by far the most ubiquitous inhibitory neurotransmitter in brain and aberrant function of the GABA system has been implicated in almost every common neurological and psychiatric disorder which further supports the need for detailed cataloging of genetic variation in the GABA system genes.

The *GAT-1* (*SLC6A1*) gene resides on chromosome 3p25-p24 and spans 46.5 kb with 16 exons. This gene encodes a protein of 599 amino acids with a molecular weight of 67 kDa. GAT-1 is a Na^+ dependent transporter and has 12 transmembrane spanning regions with intracellular N- and C- termini (35). The current genome build (36) shows 2 transcripts for *GAT-1 (SLC6A1)*, the first transcript contains all 16 exons and the other is coded by exons 2 to 16. The *SLC6A1* protein was found in several specie and the nucleotide and amino acid homology between human and other specie (chimp, rat, mouse and dog) are presented in table below (37, 38).

Table 1 The table shows percent of nucleotide and amino acid homology with human *SLC6A1*.

Organism	Human similarity	
	Nucleotide similarity (%)	Amino acid similarity (%)
chimpanzee *(Pan troglodytes)*	98.89	99.16
rat *(Rattus norvegicus)*	89.98	97.83
mouse *(Mus musculus)*	90.09	97.83
dog *(Canis familiaris)*	93.32	98.83

CHAPTER III

METERIALS AND METHODS

DNA samples

For re-sequencing of the *SLC6A1* gene, 40 genomic DNA samples were collected from unrelated individuals representing 5 different populations: EA (n=7), AA (n=9), Finnish (n=8), Thai (n=8), and Hmong (n=8). The Finnish subjects were unrelated parents of adolescent subjects who were participating in an epidemiological study focusing on the identification of risk factors for early-onset mental illness and substance dependence in Finland (39). The Thai and Hmong populations were collected in Thailand as part of an ongoing genetic association and population genetic study. The Thais selected for resequencing had grandparents and parents of Thai ancestry (Thai-Thai) or had mixed Thai and Chinese ancestry (Thai-Chinese), by subject report. These samples were obtained from a blood drive in Bangkok, Thailand. The Hmong subjects were recruited in a Hmong village in the northern part of Thailand. The AA and EA samples have been described earlier elsewhere (40). Both EA and AA samples were self-identified and confirmed as such by Bayesian marker clustering (40). All subjects provided informed consent as approved by the appropriate institutional review boards. In addition, 46 EA, 60 AA, 59 Thai, 47 Finnish and 48 Hmong individuals were genotyped for 16 *SLC6A1* SNPs to examine linkage disequilibrium (LD) in this gene. The Thai subjects selected for examination of linkage disequilibrium were Thai-Thai. Recruitment and population characteristics of subjects selected for *SLC6A1* genotyping were identical of subjects selected for resequencing (39, 40). The participants were recruited from non-clinical populations. All studies described in this article were conducted according to the Declaration of Helsinki. The studies were approved by the institutional review boards of Yale University School of Medicine, West Haven VA Hospital, Northern- Ostrobothnia Hospital District (University of Oulu, Finland) and Chulalongkorn University (Bangkok, Thailand). All subjects signed a written informed consent for participation in this study.

For family-based study of novel 21 bp insertion, 319 pedigrees (249 nuclear families, 848 persons) of African-American were recruited from 4 sites; Yale University

School of Medicine (APT Foundation; New Haven, CT), University of Connecticut Health Center (UConn; Farmington, CT), McLean Hospital (Harvard Medical School; Belmont, MA), and Medical University of South Carolina (MUSC; Charleston, SC). These families were recruited on the basis of containing at least one affected sibling pair for cocaine or opioid dependence; additional relatives were recruited where possible (41, 42). Probands with major psychotic illness (schizophrenia or schizoaffective) subjects were excluded. All subjects of self-identified AA were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) (41, 43). Cocaine dependence diagnosis was based on DSM-IV criteria.

Promoter prediction

The ElDorado program of the Genomatix software package was used to predict the location of the *SLC6A1* promoter region (44). TFSERCH was used to find transcription proteins that bind with the specific region of DNA. The sequence of the *SLC6A1* gene submitted to the promoter region analysis was obtained from the National Center for Biotechnology Information (NCBI) (45).

Amplification and sequencing

For sequencing of *SLC6A1*, the upper and lower promoter regions, all 16 *SLC6A1* exons (total of 4.4 kb) and 7.3 kb of flanking intronic regions were amplified. About 70 bp of the predicted 601 bp of the lower promoter region were not included in the sequence analysis. Approximately 12.4 kb of the *SLC6A1* gene was amplified, corresponding to about 25% of the total length of the gene. All primers were designed with the PRIMER3 software (46). Primers were obtained from Invitrogen (Carlsbad, CA). PCR amplification was optimized before sequencing by testing different cycling conditions. Betaine (Sigma Aldrich, St. Louis, MO) at 0.5-1 M final concentrations was added to the reactions, as needed, to enhance specificity and yield of PCR amplification. PCR reactions were carried out in 15 μl volumes containing 20 ng genomic DNA, 200 μM of dNTPs mix (Stratagene, La Jolla, CA), 1 μM of mixed primers forward and reverse, 1X PC2 buffer, 0.75 U of KlenTaq1^{TM} (Ab Peptides, St Louis, MO) and 0.5-1 M betaine when needed. Thermocycling conditions consisted of an initial

denaturation step at 95°C for 5 min, 30 cycles of denaturation step at 95°C for 30 sec, an annealing step at 60-65°C 30 sec, and an extension step at 72°C. The duration of the extension step varied from 30 sec to 2 min depending on the length of the amplicon. After optimization, genomic DNA samples from each population were PCR amplified followed by purification with MinElute PCR purification columns (Qiagen, Valencia, CA) or the reaction mixtures were treated with ExoSAP-IT (USB, Cleveland, OH) to remove excess nucleotides and primers. Purified PCR samples were sequenced in the forward and reverse directions at Yale University W.M Keck Foundation Biotechnology Resource Laboratory. Sequencing reactions were conducted using the BigDye Terminator v3.1 cycle sequencing kit and an ABI 9800 Thermocycler (Applied Biosystem, Foster city, CA). Sequencing reactions were analyzed on an ABI 3730 xl DNA Analyzer (Applied Biosystems, Foster city, CA). Owing to technical problems, approximately 300 bp in exon 7 and intron 8 (2.4% from total 12.4 kb sequenced region) is missing in the sequencing data in Thai and Hmong populations (see additional data file).

Owing to the repeat elements contained in the identified upper promoter sequence and homologous sequences within the *SLC6A1,* the upper promoter region and parts of exon 1 were amplified using nested PCR. In addition, for amplification of a 180 bp fragment located in the junction of the 5' upstream region and exon 1, a region which is very high in CG content, 7-deaza-dGTP (New England BioLabs, Beverly, MA) was added to the reactions.

Genotyping and linkage disequilibrium study

A total of 16 SNPs were chosen for genotyping in population samples to examine haplotype structure of the *SLC6A1* gene. Nine SNPs chosen for genotyping were identified through resequencing: -24321A/C, -1529A/G, 949A/G, 3164C/T, 14351A/G, 16009A/G, 16116C/T, 20172C/T and 20622A/G. The remaining seven SNPs, -29477C/T, -17590C/T, -13071A/G, -9765C/T, 7772A/G, 13269C/T, and 16605C/T, were chosen from the NCBI dbSNP (47) collection. Of the 16 SNPs studied, 14 were available through Applied-Biosystem's Assay-On-Demand service (Applied Biosystems, Foster city, CA). One assay was custom designed and obtained through the ABI's Assay-by-Design service (Applied Biosystems, Foster city, CA). PCR amplification of the 5' nuclease

assays were conducted using 1 ng of DNA, 1X TaqMan universal PCR master mix (Applied Biosystems, Foster city, CA), 0.5X SNP genotyping assay mix (Applied Biosystems, Foster city, CA). PCR conditions were as follows: denaturation step of 95 ^{o}C for 10 min, followed by 50 cycles of 95 ^{o}C for 15 sec and 60 ^{o}C for 1 min. Amplification was performed on PTC-200 cyclers (MJ Research, Hercules, CA) and data were analyzed using the ABI Prism 7900HT Sequence Detector System and software version 2.1 (Applied Biosystem, Foster city, CA). All samples were run in duplicate for quality control purposes. Based on comparison of the duplicate runs, we estimated the genotyping error rate to be less than 0.05%. The -24321A/C SNP was genotyped using 7-deaza-dGTP sequencing because its location inside a GC-rich region made it very difficult to design a 5' nuclease assay for this SNP.

Genotyping of the length polymorphisms

Amplification of the region containing the 21 bp short/long VNTR and 2 bp GG/-GG insertion/deletion polymorphisms was accomplished using primers 5' AAGGAGAGAGATTGGAGCG 3' and 5' CTTCTTTCCTCTCGCATTC 3' (Invitrogen, Carlsbad, CA). PCR reactions were conducted in 15 μl volumes containing 20 ng genomic DNA, 200 μM of dNTPs mix (Stratagene, La Jolla, CA), 1 μM of mixed reverse and forward primers, 1X PC2 buffer, 0.75 U of KlenTaq1^{TM} (Ab Peptides, St Louis, MO) and 1 M Betaine. The thermocycling conditions consisted of an initial step at 95^{o}C for 5 min, 30 cycles of denaturation at 95^{o}C for 30 sec, annealing 60^{o}C 30 sec, and extension 72^{o}C 30 sec. The lengths of the PCR products corresponding to the long and short alleles are 166 bp and 145 bp. The long and short alleles were separated using 3% metaphore agarose and gel electrophoresis (ISC BioExpress, Kaysville, UT). The GG/-GG insertion/deletion polymorphism was genotyped using direct sequencing of the PCR product as described in above.

Cell culture

All cell lines were obtained from ATCC. 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, CA). Media were supplemented

with 10% fetal bovine serum, 2 U/ml penicillin, 2 μg/ml of streptomycin, and 2 mM L-Glutamine (GIBCO invitrogen cell culture, Carlsbad, CA). Human neuroblastoma cells [SK-N-BE(2)] were cultured in a 1:1 mixture of Eagle's Minimum Essential Medium (EMEM) and F-12K media (ATCC, Manassas VA) supplemented with 10% non-heat-inactivated fetal bovine serum, 2U/ml penicillin, and 2 μg/ml streptomycin. All cells were grown in a humidified incubator at 37°C and 5% CO_2.

Electromobility Shift Assay (EMSA)

Nuclear protein extracts from P19, SK-N-BE(2), and HEK-293 were prepared using the NE-PER Nuclear and Cytoplasmic Extraction kit (Pierce) and quantified by BCA protein assay kit (Pierce). Two sets of double stranded DNA probes were constructed coding for one copy of the VNTR (5'-AGGAGGGGGTGGGGAGAGGGA-3'). 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). The IRDye 800 competitor probe was used in a manner analogous to a non-radioactive competitor probe in radioisotope-based 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 (2.5 mM DTT/0.25% Tween-20, 1 μg of Poly (dI-dC; LI-COR bioscience, Lincoln, Nebraska) and mixed with nuclear lysates. Binding reactions were carried out for 30 minutes at room temperature with end-over-end mixing and loaded on 4% native acrylamide gel containing 0.38 M glycine. Gel imaging was carried out using an Odyssey Imaging System (LI-COR bioscience, Lincoln, Nebraska) at 700 nm and 800 nm wavelengths. For supershift experiments, 1 μl of human SP1 antibody (Upstate, Charlottesville VA) was added in a binding reaction and incubated for 30 minutes at room temperature before running on a 4% polyacrylamide gel.

Construction of reporter plasmids

Genomic DNA from samples that were determined to be homozygous for the *SLC6A1* insertion or non-insertion alleles with otherwise identical sequences were identified by DNA sequencing. Genomic DNA was amplified by nested PCR. The first

PCR utilized 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 (-25543 to -24023) of the *SLC6A1* gene. PCR reactions were carried out in 50 μl volumes containing 50 ng genomic DNA, 10 mM of each dNTP (New England BioLabs, Beverly, MA), 2 uM of primer, 1x reaction buffer, 1 μl of PfuUltra II Fusion HS DNA Polymerase (Stratagene, La Jolla, CA) and 10% sulfolane. Thermocycling conditions consisted of an initial denaturation step at 98°C for 2 min, touch down PCR from 70-50°C and 25 cycles of denaturation at 95°C for 20 sec, annealing step at 58°C 20 sec, and extension at 72°C 90 sec. PCR products were gel purified and amplified in a second PCR reaction using primers 5'-CTGAGTTCCTGGGGACCCCAGAGGGAAGG-3' and 5'-CGAGCGGCGCCTTGCTCC TTCATGTGG-3' that amplified positions -25516 to -24103 of the *SLC6A1* gene. Each 50 μl reaction contained 1 μl of purified PCR product, 10 mM of each dNTP (New England BioLabs, Beverly, MA), 2 μM of each primer, 1x reaction buffer, 1 μl of PfuUltra™ II Fusion HS DNA Polymerase (Stratagene, La Jolla, CA), and 10% DMSO. Thermocycling was conducted as follows: denaturation at 98°C for 2 min, 30 cycles of 95°C for 20 sec, 65°C 20 sec, and 72°C for 90 sec. 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 were selected, digested with *NheI* and *MluI* (New England BioLabs, Beverly, MA), and subcloned into the pGL3-basic reporter vector (Promega, Madison WI) to create pGL3-*SLC6A1*-Ins (21 bp *SLC6A1* insertion) and pGL3-*SLC6A1*-Del (non-insertion).

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 mM $CaCl_2$ in 2x HEPES-buffered saline (HBS) buffer (500 mM Hepes, 100 mM Na_2HPO_4, 1M NaCl pH=7.10) for 6 hours followed by 15% glycerol shock for 40 minutes (48, 49). Transfections were optimized to yield >60%

efficiency as determined by lacZ plasmid transfection and x-gal histochemistry. pRL-SV40 plasmid (Promega, Madison WI) coding for Ranilla 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, Madison WI) on a DTX 880 Series Multimode Detector (Beckman Coulter, Fullerton CA).

DNA affinity pull down assay

DNA templates that contained 0, 1 or 2 copies of the VNTR were designed for pull down assays. The one and 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 (Figure 1) (50) using 4 primers, 5'-TGTAAAACGACGGCCAGTGGCAGACAGGCTGGTGACCCAGG ATGA-3', 5'-TCTCTTCCTCCCTCCCTCGCCTGCCCCGCCGT-3', 5'-GCAGGCGAGG GAGGGAGGAAGAGA-3' and 5'-CTTCTTTCCTCTCGCATTC-3'. PCR products were labeled at the 5' ends by amplification with a dual biotin containing M13 oligonucleotide sequence (51). Biotinylated DNA was incubated with 1x binding buffer (10 mM Tris, 50 mM KCl, 1 mM DTT pH=7.5) and nuclear protein lysates at room temperature for 30 minutes 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 minutes with continuous mixing. Complexes were purified using a micocentrifuge tube magnet (Dynal). Supernatants containing unbound material were removed using aspiration and beads were washed 3 times with buffer containing 20 mM Tris, 100 mM KCl, 2 mM DTT, pH=7.5. Complexes were eluted by adding 6x SDS sample buffer and heating at 95°C for 5 minutes. One-dimensional electrophoresis was carried out using 10% SDS-PAGE gels run at 10 mA/gel at room temperature. Gels were stained by coomassie blue (GelCode Blue Stain reagent, PIERCE, Rockford IL) and destained overnight with ultra pure water, or were stained using silver staining (SilverSNAP Stain kit II, PIERCE, Rockford IL). Bands were cut and were subjected to liquid chromatography followed by mass spectroscopy (Yale/NIDA Proteomic Center,

Yale University). Gels fragments were trypsinization 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: 2 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) and Bioinformatics Harvester (52).

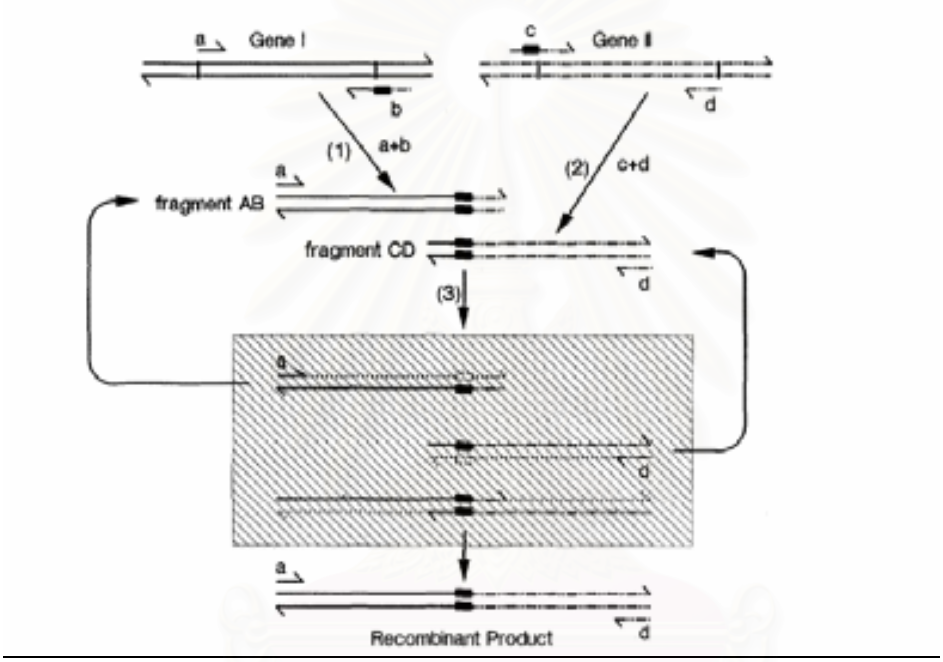


Figure 1 *The illustration of overlap extension PCR technique (50).*

NFAT signaling assays

P19 and HEK-293 cells were transiently transfected with the luciferase constructs. After 48 hours, cells were pretreated with vehicle or 500 nM of cyclosporin A (CsA; Sigma-Aldrich, St Louis MO) for 30 minutes before stimulation with 25 ng/ml of phorbol-12-myristate-13-acetate (PMA; Sigma-Aldrich, St Louis MO) for 3 hours. Luciferase activity was subsequently measured.

Statistical analysis and software

Indices of sequence variation in *SLC6A1* were calculated using a web application SLIDER (53). These indices included the number of polymorphic sites, nucleotide diversity per base pair (π) and the Watterson's estimator of theta (θ).

Nucleotide diversity per base pair (π) describes the mean number of differences per site between two sequences chosen at random from a sample of sequences. The Watterson's estimator of theta (θ) is the observed number of SNPs adjusted for the sample size and new mutation rate expected to occur in each generation (27). In addition, for each subject we calculated the number of heterozygous SNPs observed in the sequence data. The number of heterozygous SNPs was compared between populations using ANOVA followed by post hoc Fisher's Least Significant Difference-test.

PHASE software, which implements a Bayesian algorithm for haplotype reconstruction, was used to estimate haplotype frequencies (54, 55). PHASE's options –X10 and –MR were used to estimate recombination rates across *SLC6A1* (56, 57). The value on the Y-axis of Figure 3 shows changes in recombination parameter (ρ) per base pair of *SLC6A1* exceeding the background recombination rate (56, 57). The average recombination rate was estimated based on 1,000 burn-ins and 1,000 iterations. Recombination frequencies at *SLC6A1* were compared visually between our and HapMap data. No statistical analyses were performed. To evaluate haplotype diversity among populations, we studied how often the most common haplotypes were shared or disjoint. The haplotypes were identified using a sliding window analysis across every three consecutive *SLC6A1* SNPs. The four most common three-SNP haplotypes in each window and in each population were identified. The rationale for choosing the four most common haplotypes for this analysis was that visual inspection of the haplotype frequencies told us that in each window and in each population virtually all variation in haplotype diversity was captured by the four most common haplotypes. The average percent of all haplotypes captured by the top four haplotypes was 96%. We then calculated how many times each of the common three-SNP haplotype was disjoint between the populations. A summary pairwise score derived for the populations is presented in Table 4.

LD patterns in *SLC6A1* were visualized using HAPLOVIEW version 3.2.(58) We used the T*agger* algorithm, implemented in HAPLOVIEW, to search for haplotype tagging SNPs in *SLC6A1* (59). We used the default *Tagger* thresholds r^2>0.8 and LOD score > 3. POWERMARKER (60) was used to calculate allele frequencies and

examination of Hardy-Weinberg equilibrium (HWE). To illustrate differences in the span of LD in the five populations, r^2 was plotted against physical distance. To do this, r^2 was calculated for all SNP pairs. Because these values were not normally distributed, median values are presented. Physical distance (bp) was divided into distance bins to illustrate population differences in LD span across a range of physical distances. Median r^2 in distance bins (0.1-10kb, 10.01-20kb etc) in different populations is presented in Figure 4. No statistical analyses were performed on these data.

Tests of association for 21 bp insertion with cocaine and alcohol dependence in African-American families were performed by means of FBAT (Family-Based Association Tests) (61) analysis.

CHAPTER IV

RESULTS

Algorithmic promoter prediction

The sequence for *SLC6A1*, located in contig NT_022517, which contains the sequence for human chromosome 3, was used as a reference. ElDorado software (44) located two putative *SLC6A1* promoter regions in the contig at positions 10973950-10974553 and 10998399-10998999 (+ strand). The first region is located in an area surrounding the boundary between the 5' region and exon 1 of *SLC6A1* and is estimated to be 604 bp long. Hereafter, we refer to this region as the upper promoter region. The second putative promoter region of 601 bp was located in the area flanking the boundary between intron 2 and exon 3. This region extends 70 bp across the *SLC6A1* start codon in the 3' direction in exon 3. We refer to this region as the lower promoter region.

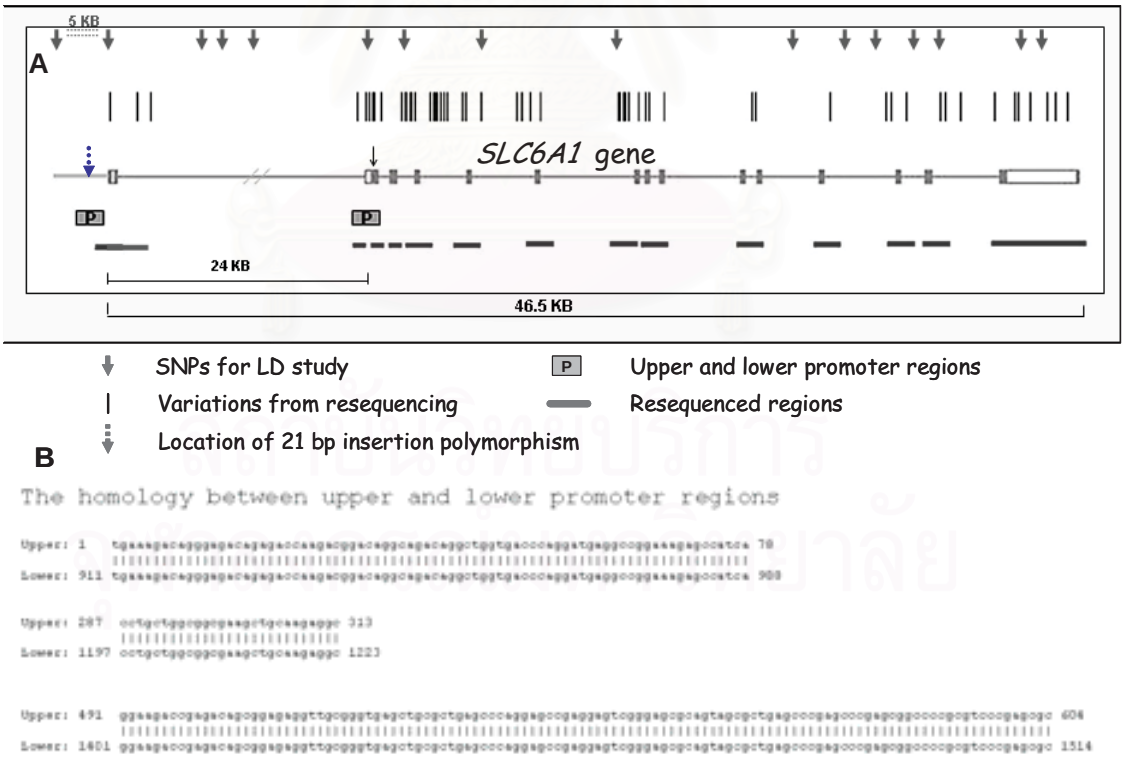


Figure 2 *SLC6A1 gene map, location of SNP markers in LD study, locations of variation from resequencing result, promoter regions and position of 21 bp insertion were presented in A. The homology between upper and lower promoter regions were shown in B.*

Nucleotide diversity in *SLC6A1* in five populations

The *SLC6A1* promoter regions, 16 exons, and the flanking intronic regions were amplified and resequenced in 7-9 samples from each of the five populations. A total of 61 SNPs were found in the sequencing screening sample, including 5 SNPs in the promoter regions, 3 synonymous coding sequence SNPs, and 53 non-coding region SNPs. No non-synonymous SNPs were found. The identified *SLC6A1* variants, including frequency in the screening sample, location, and flanking sequence are provided in the table 2.

In the screening sample, 33 of the 61 SNPs discovered (54%) were present in more than one population. More SNPs unique to a single population or continental group were found in the AAs than in all of the other populations combined (counting Finns and EAs, and Thai and Hmong, together). Sixteen population-specific SNPs were observed in the AAs, one in the EAs, 6 in the Finnish, 4 in the Thai, and 2 in the Hmong (see table 2); i.e., a total of 13 in the populations other than AAs. Three common population-specific SNPs were found in AA population (-24794A/G, -24126G/T and 17885 A/C, minor allele frequencies = 0.39, 0.33 and 0.22) (These estimates of population allele frequencies have to be interpreted with caution since the data were derived from sequencing only 14-18 chromosomes per population). Analysis of nucleotide diversity based on the sequencing data indicated that AAs had the greatest number of polymorphic sites (n=41), highest nucleotide diversity per bp (2.77×10^{-4}) and highest Watterson's estimator of theta (θ) (2.55×10^{-4}). Inspection of nucleotide diversity in the Hmong, Finnish, EA, and Thai populations revealed no marked differences among these four populations (Table 3). As expected, nucleotide diversity was lower in exons as compared to intronic regions (π_{EX} = 5.5-7.0 x 10^{-5}, π_{INT} = 1.1-2.3 x 10^{-4}). Interestingly, nucleotide diversity was not higher in the AAs when exonic sequence only was considered ($\pi_{AA} = 5.9 \times 10^{-5}$, $\pi_{EA} = 6.4 \times 10^{-5}$, $\pi_{Finn} = 7.1 \times 10^{-5}$, $\pi_{Thai} = 5.6 \times 10^{-5}$, $\pi_{Hmong} = 5.6 \times 10^{-5}$). The average number of heterozygous SNPs per

Table 2 List of SNPs discovered in *SLC6A1* by resequencing 40 individuals from Thai (n=8), Hmong (n=8), European-American (n=7), African-American (n=9) and Finnish (n=8) populations.

Name	rs	position	location	alleles	flanking sequences	EA	AA	Finn	Thai	Hmong
-29477	rs1710879	11004420	5'upstream	T/C	TCATA**T/C**GTAGTT	0.71/0.29 (0.197-0.383)	0.99/0.01 (-0.008-0.028)	0.75/0.25 (0.162-0.338)	0.76/0.24 (0.163-0.317)	0.78/0.22 (0.137-0.303)
-25070	rs2675156	11008827	5'upstream	A/C	ACCCC**A/C**ACCCA	0.64/0.36 (0.109-0.611)	0.28/0.72 (0.073-0.487)	0.63/0.37 (0.133-0.607)	0.63/0.37 (0.133-0.607)	0.75/0.25 (0.038-0.462)
-24993	rs2697151	11008904	5'upstream	A/G	GCAGT**A/G**CTGCA	0.36/0.64 (0.109-0.611)	0/1	0.19/0.81 (-0.002-0.382)	0.13/0.87 (-0.035-0.295)	0/1
-24924	rs2675157	11008973	5'upstream	A/G	GACCA**A/G**GACGG	0.64/0.36 (0.109-0.611)	0.61/0.39 (0.165-0.615)	0.63/0.37 (0.133-0.607)	0.94/0.06 (-0.056-0.176)	0.81/0.19 (-0.002-0.382)
-24794	rs41469948	11009103	5'upstream	A/G	AGGGA**G/A**GGAGG	0/1	0.39/0.61 (0.165-0.615)	0/1	0/1	0/1
-24788	rs41342644	11009109	5'upstream	ins21 bp	**[GGGTGGGGGAGAG GGAGGGAGG]**	0/1	0.39/0.61 (0.165-0.615)	0/1	0/1	0/1
-24780	rs41539860	11009117	5'upstream	GG/-GG	AGAGG**[GG]**AGAAA	0.70/0.30 (0.06-0.54)	0.77/0.23 [b] (0.036-0.424)	0.67/0.33 (0.100-0.560)	0.78/0.22 (0.017-0.423)	0.84/0.16 (-0.02-0.34)
-24534	rs41338544	11009363	5'upstream	C/T	GCGGG**C/T**CCTGC	1/0	1/0	0.94/0.06 (-0.056-0.176)	1/0	1/0
-24343	rs41429044	11009554	Exon1 (5'UTR)	C/T	GCGCA**T/C**CGGAG	0/1	0/1	0.06/0.94 (-0.056-0.176)	0/1	0/1
-24321	rs34189945	11009576	Exon1 (5'UTR)	A/C	CCCGG**A/C**GCAGC	0.64/0.36 (0.109-0.611)	0.33/0.67 (0.113-0.547)	0.75/0.25 (0.038-0.462)	0.69/0.31 (0.083-0.537)	0.75/0.25 (0.038-0.462)
-24321	rs34189945	11009576	Exon1 (5'UTR)	A/C	CCCGG**A/C**GCAGC	0.66/0.34 (0.243-0.437)	0.59/0.41 (0.322-0.498)	0.68/0.32 (0.226-0.414)	0.64/0.36 (0.273-0.447)	0.66/0.34 (0.245-0.435)
-24271	rs41486951	11009627	Intron1	A/G	AGGGC**G/A**GCCGG	0/1	0/1	0/1	0.06/0.94 (-0.056-0.176)	0/1
-24126	rs41506145	11009771	Intron1	G/T	CCACA**T/G**GAAGG	0/1	0.33/0.67 (0.113-0.547)	0/1	0/1	0/1
-17590	rs1568074	11016307	Intron1	T/C	GAGGC**T/C**TGGGC	0.51/0.49	0.22/0.78	0.59/0.41	0.5/0.5	0.51/0.49

						(0.388-0.592)	(0.146-0.294)	(0.311-0.509)	(0.410-0.590)	(0.390-0.590)
-13071	rs1710892	11020826	Intron1	G/A	TCTGC**G/A**GTTTT	0.53/0.47 (0.368-0.572)	0.31/0.69 (0.227-0.393)	0.63/0.37 (0.272-0.468)	0.5/0.5 (0.410-0.590)	0.5/0.5 (0.400-0.600)
-9765	rs1710891	11024132	Intron1	C/T	CAGGA**C/T**TGGAT	0.55/0.45 (0.348-0.552)	0.5/0.5 (0.411-0.589)	0.6/0.4 (0.301-0.499)	0.53/0.47 (0.380-0.560)	0.63/0.37 (0.237-0.467)
-1569	rs41422847	11032328	Intron1	C/T	CATCC**C/T**TCCTC	1/0	1/0	1/0	1/0	0.94/0.06 (-0.056-0.176)
-1529	rs6342	11032368	Exon2 (5'UTR)	A/G	ACCCA**G/A**GGTGG	0.43/0.57 (0.171-0.689)	0.11/0.89 (-0.035-0.255)	0.44/0.56 (0.197-0.683)	0.44/0.56 (0.197-0.683)	0.06/0.94 (-0.056-0.176)
-1529	rs6342	11032368	Exon2 (5'UTR)	A/G	ACCCA**G/A**GGTGG	0.48/0.52 (0.378-0.582)	0.22/0.78 (0.146-0.294)	0.67/0.33 (0.235-0.425)	0.34/0.66 (0.255-0.425)	0.14/0.86 (0.071-0.209)
-1213	rs41386254	11032684	Intron2	A/G	GCCCA**A/G**TTTCC	1/0	0.94/0.06 (-0.05-0.17)	1/0	1/0	1/0
-396	rs11923810	11033501	Intron2	T/C	TTAAG**T/C**ACTTA	1/0	0.78/0.22 (0.029-0.411)	1/0	0.87/0.13 (-0.035-0.295)	1/0
-234	rs41388950	11033663	Intron2	T/C	AAGGG**C/T**GGGTG	0/1	0.22/0.78 (0.029-0.411)	0/1	0/1	0.44/0.56 (0.197-0.683)
-218	rs11919775	11033679	Intron2	G/C	AAGAG**G/C**GCTTA	1/0	0.78/0.22 (0.029-0.411)	0.94/0.06 (-0.056-0.176)	0.94/0.06 (-0.056-0.176)	1/0
-17	rs41362845	11033880	Exon3 (5'UTR)	C/T	CCCTG**C/T**GTCCA	1/0	1/0	0.94/0.06 (-0.056-0.176)	1/0	1/0
949	rs2928078	11034846	Intron4	G/A	CAGAA**G/A**CTAGG	0.43/0.57 (0.171-0.689)	0.33/0.67 (0.113-0.547)	0.5/0.5 (0.255-0.745)	0.56/0.44 (0.197-0.683)	0.06/0.94 (-0.056-0.176)
949	rs2928078	11034846	Intron4	G/A	CAGAA**G/A**CTAGG	0.47/0.53 (0.368-0.572)	0.25/0.75 (0.173-0.327)	0.31/0.69 (0.217-0.403)	0.62/0.38 (0.292-0.468)	0.76/0.24 (0.155-0.325)
1063	rs7643585	11034960	Intron4	C/T	TGCCT**C/T**GTGCT	1/0	0.83/0.17 (-0.004-0.344)	1/0	1/0	0.56/0.44 (0.197-0.683)
1064	rs41391147	11034961	Intron4	G/C	GCCTC**G/C**TGCTC	1/0	1/0	0.94/0.06 (-0.056-0.176)	1/0	1/0
1076	rs41444444	11034973	Intron4	C/T	GTGCC**T/C**GACAG	0/1	0/1	0/1	0.06/0.94 (-0.056-0.176)	0/1
1098	rs1728803	11034995	Intron4	A/G	GATAC**A/G**TGGTG	0.5/0.5 (0.238-0.762)	0.83/0.17 (-0.004-0.344)	0.44/0.56 (0.197-0.683)	0.5/0.5 (0.255-0.745)	0.81/0.19 (-0.002-0.382)

1114	rs41381549	11035011	Intron4	A/C	TGCTC**A/C**CTGAC	1/0	1/0	1/0	0.94/0.06 (-0.056-0.176)	1/0
1532	rs3817585	11035429	Intron5	G/C	ACCCA**G/C**AGCCC	1/0	1/0	1/0	0.87/0.13 (-0.035-0.295)	0.94/0.06 (-0.056-0.176)
1606	rs41524646	11035503	Intron5	G/A	GGTTT**G/A**TCTTT	1/0	0.89/0.11 (-0.035-0.255)	1/0	1/0	1/0
1724	rs2933307	11035621	Intron5	G/C	AAGTG**C/G**GTATT	0.43/0.57 (0.171-0.689)	0.67/0.33 (0.113-0.547)	0.50/0.50 (0.255-0.745)	0.62/0.38 (0.142-0.618)	0.06/0.94 (-0.056-0.176)
1773	rs9827626	11035670	Intron5	C/T	AAATC**C/T**TGACT	1/0	0.78/0.22 (0.029-0.411)	1/0	0.87/0.13 (-0.035-0.295)	1/0
1775	rs2928077	11035672	Intron5	G/C	ATCCT**G/C**ACTCT	0.57/0.43 (0.171-0.689)	0.61/0.39 (0.165-0.615)	0.5/0.5 (0.255-0.745)	0.56/0.44 (0.197-0.683)	0.94/0.06 (-0.056-0.176)
1897	rs41370144	11035794	Intron5	C/T	TTAAA**C/T**ATAGA	1/0	1/0	1/0	1/0	0.87/0.13 (-0.035-0.295)
2008	rs1710888	11035905	Intron5	C/T	ATTCA**C/T**AGTAG	1/0	1/0	1/0	0.87/0.13 (-0.035-0.295)	1/0
2884	rs41497953	11036781	Intron5	C/T	TGCTG**C/T**GAAGG	1/0	0.89/0.11 (-0.035-0.255)	1/0	1/0	1/0
2916	rs41478447	11036813	Intron5	C/T	GACCA**T/C**GAGGG	0/1	0.06/0.94 (-0.05-0.17)	0/1	0/1	0/1
3164	rs3774070	11037061	Intron6	C/T	GAGCC**C/T**GGCAG	1/0	0.61/0.39 (0.165-0.615)	0.87/0.13 (-0.035-0.295)	0.87/0.13 (-0.035-0.295)	0.94/0.06 (-0.056-0.176)
3164	rs3774070	11037061	Intron6	C/T	GAGCC**C/T**GGCAG	0.94/0.06 (0.011-0.109)	0.5/0.5 (0.411-0.589)	0.93/0.07 (0.018-0.122)	0.89/0.11 (0.054-0.166)	0.92/0.08 (0.026-0.134)
4647	rs11712912	11038544	Intron6	A/G	GCACC**G/A**GCAAA	0/1	0.06/0.94 (-0.05-0.17)	0/1	0.13/0.87 (-0.035-0.295)	0.06/0.94 (-0.056-0.176)
4766	rs3856786	11038663	Intron6	G/A	GCAGA**G/A**GAGTG	1/0	0.94/0.06 (-0.05-0.17)	1/0	0.87/0.13 (-0.035-0.295)	1/0
4965	rs3856787	11038862	Intron6	G/A	AATGG**G/A**CCTCA	1/0	0.94/0.06 (-0.05-0.17)	1/0	1/0	1/0
5193	rs6344	11039090	Exon7	G/T	ATCAC**G/T**CTGGC	1/0	1/0	0.94/0.06 (-0.056-0.176)	[a]	[a]
7772	rs10510403	11041669	Intron7	A/G	GCATT**A/G**AAGTA	0.82/0.18	0.89/0.11	0.86/0.14	0.74/0.26	0.81/0.19

						(0.101-0.259)	(0.054-0.166)	(0.070-0.210)	(0.181-0.339)	(0.112-0.268)
7942	rs41323844	11041840	Intron7	A/C	TGTGC**A/C**TATGT	1/0	0.89/0.11 (-0.035-0.255)	1/0	0.87/0.13 (-0.035-0.295)	1/0
7944	rs41480248	11041842	Intron7	A/C	TGCAT**A/C**TGTAA	1/0	0.89/0.11 (-0.035-0.255)	1/0	1/0	1/0
7978	rs9822125	11041875	Intron7	A/T	ATGCA**A/T**TACTT	1/0	0.50/0.50 (0.269-0.731)	1/0	0.94/0.06 (-0.056-0.176)	1/0
8159	rs41335049	11042056	Intron7	C/T	AGATA**C/T**GGATG	1/0	0.83/0.17 (-0.004-0.344)	1/0	1/0	1/0
8229	rs41537851	11042126	Intron7	C/T	ACTTT**C/T**CTCCC	1/0	0.94/0.06 (-0.05-0.17)	1/0	1/0	1/0
8302	rs33948309	11042199	Exon8	G/A	GTGAC**G/A**CTGCC	1/0	0.89/0.11 (-0.035-0.255)	1/0	1/0	1/0
8443	rs17532365	11042340	Intron8	G/A	TTTCT**G/A**ACCCTC	1/0	1/0	0.87/0.13 (-0.035-0.295)	1/0	1/0
8830	rs41426745	11042727	Intron9	G/C	TTCCT**G/C**TTGTC	1/0	0.83/0.17 (-0.004-0.344)	1/0	1/0	1/0
9029	rs35972647	11042926	Exon10	C/T	GACTC**C/T**ATCAT	1/0	0.83/0.17 (-0.004-0.344)	1/0	1/0	1/0
11967	rs6770472	11045863	Intron11	G/T	TTGGG**G/T**CTGGG	1/0	0.94/0.06 (-0.05-0.17)	1/0	1/0	1/0
11973	rs36034065	11045869	Intron11	G/A	CTGGG**G/A**CTGCT	1/0	0.89/0.11 (-0.035-0.255)	1/0	1/0	1/0
13269	rs11925331	11047166	Intron12	C/T	CAGCA**C/T**AGAGA	0.96/0.04 (0-0.080)	0.66/0.34 (0.255-0.425)	0.99/0.01 (-0.010-0.030)	0.68/0.32 (0.236-0.404)	0.65/0.35 (0.255-0.445)
14351	rs2272403	11048248	Intron13	G/A	ACACA**G/A**TCTAG	0.93/0.07 (-0.064-0.204)	0.44/0.56 (0.211-0.669)	0.81/0.19 (-0.002-0.382)	0.5/0.5 (0.255-0.745)	0.69/0.31 (0.083-0.537)
14351	rs2272403	11048248	Intron13	G/A	ACACA**G/A**TCTAG	0.88/0.12 (0.054-0.186)	0.38/0.62 (0.293-0.467)	0.89/0.11 (0.047-0.173)	0.62/0.38 (0.292-0.468)	0.64/0.36 (0.264-0.456)
16009	rs41436650	11049906	Intron13	A/G	CAGGC**A/G**TGGGC	1/0	0.89/0.11 (-0.035-0.255)	0.94/0.06 (-0.056-0.176)	1/0	0.87/0.13 (-0.035-0.295)
16009	rs41436650	11049906	Intron13	A/G	CAGGC**A/G**TGGGC	0.97/0.03 (-0.005-0.065)	0.89/0.11 (0.054-0.166)	0.98/0.02 (-0.008-0.048)	0.94/0.06 (0.017-0.103)	0.80/0.20 (0.120-0.280)

16116	rs2675163	11050013	Intron13	C/T	GATGT**C/T**GAGTG	0.29/0.71 (0.052-0.528)	0.11/0.89 (-0.035-0.255)	0.19/0.81 (-0.002-0.382)	0.19/0.81 (-0.002-0.382)	0.69/0.31 (0.083-0.537)
16116	rs2675163	11050013	Intron13	C/T	GATGT**C/T**GAGTG	0.25/0.75 (0.162-0.338)	0.14/0.86 (0.078-0.202)	0.20/0.80 (0.119-0.281)	0.27/0.73 (0.190-0.350)	0.51/0.49 (0.390-0.590)
16605	rs2246543	11050502	Intron14	T/C	GAAAC**T/C**TCTAG	0.64/0.36 (0.109-0.611)	0.94/0.06 (-0.05-0.17)	0.81/0.19 (-0.002-0.382)	0.25/0.75 (0.038-0.462)	0/1
16605	rs2246543	11050502	Intron14	T/C	GAAAC**T/C**TCTAG	0.53/0.47 (0.368-0.572)	0.86/0.14 (0.078-0.202)	0.43/0.57 (0.330-0.530)	0.25/0.75 (0.172-0.328)	0.17/0.83 (0.095-0.245)
17502	rs35957531	11051399	Intron15	C/G	CCCCA**C/G**CCTTC	0.93/0.07 (-0.064-0.204)	1/0	0.75/0.25 (0.038-0.462)	1/0	1/0
17885	rs41517144	11051810	Intron15	A/C	ATGAC**A/C**AGAGA	1/0	0.78/0.22 (0.029-0.411)	1/0	1/0	1/0
17913	rs2697138	11051906	Intron15	G/T	ATGGG**G/T**TGTCA	0.93/0.07 (-0.064-0.204)	1/0	0.75/0.25 (0.038-0.462)	0.81/0.19 (-0.002-0.382)	0.62/0.38 (0.142-0.618)
19475	rs2697134	11053372	Intron15	G/A	GTGGG**G/A**CAGTG	0.93/0.07 (-0.064-0.204)	1/0	0.75/0.25 (0.038-0.462)	0.69/0.31 (0.083-0.537)	0.06/0.94 (-0.056-0.176)
20172	rs2675165	11054069	Exon16 (3'UTR)	C/T	CCCTC**C/T**GAACG	0.93/0.07 (-0.064-0.204)	1/0	0.75/0.25 (0.038-0.462)	0.81/0.19 (-0.002-0.382)	0.56/0.44 (0.197-0.683)
20172	rs2675165	11054069	Exon16 (3'UTR)	C/T	CCCTC**C/T**GAACG	0.83/0.17 (0.093-0.247)	0.97/0.03 (-0.001-0.061)	0.84/0.16 (0.086-0.234)	0.81/0.19 (0.119-0.261)	0.80/0.20 (0.120-0.280)
20177	rs41510044	11054074	Exon16 (3'UTR)	A/G	CGAAC**G/A**CTGCT	0.5/0.5 (0.238-0.762)	0.39/0.61 (0.165-0.615)	0.13/0.87 (-0.035-0.295)	0/1	0/1
20358	rs2675166	11054255	Exon16 (3'UTR)	C/T	TATTC**C/T**CAGGG	0.93/0.07 (-0.064-0.204)	1/0	0.75/0.25 (0.038-0.462)	0.81/0.19 (-0.002-0.382)	0.56/0.44 (0.197-0.683)
20622	rs2944367	11054519	Exon16 (3'UTR)	A/G	GCCCA**A/G**TTTCA	0.71/0.29 (0.052-0.528)	0.61/0.39 (0.165-0.615)	0.75/0.25 (0.038-0.462)	0.75/0.25 (0.038-0.462)	0.62/0.38 (0.142-0.618)
20622	rs2944367	11054519	Exon16 (3'UTR)	A/G	GCCCA**A/G**TTTCA	0.74/0.26 (0.170-0.350)	0.84/0.16 (0.094-0.226)	0.79/0.21 (0.128-0.292)	0.84/0.16 (0.094-0.226)	0.81/0.19 (0.112-0.268)
21271	rs1062246	11055168	Exon16 (3'UTR)	G/A	CTCAC**A/G**TGTGG	0.36/0.64 (0.109-0.611)	0.56/0.44 (0.211-0.669)	0.5/0.5 (0.255-0.745)	0.44/0.56 (0.197-0.683)	0.62/0.38 (0.142-0.618)
21726	rs41276505	11055624	Exon16 (3'UTR)	C/T	TTGCC**C/T**GGGGG	0.93/0.07 (-0.064-0.204)	1/0	1/0	1/0	1/0

SNPs genotyped in the population samples are also provided. A total of 61 SNPs and two length polymorphisms were discovered. Name refers to the position of the polymorphism in relation to the ATG of the *SLC6A1* gene. Column “rs” gives reference SNP number is it was available. Location refers to the location of the polymorphism in the gene. In columns “EA,” “AA,” “Finn,” “Thai,” “Hmong,” allele frequencies of the SNP, estimated based on sequencing of 14-18 chromosomes in each population, is given. “a” no sequence data available for this segment. “b” = frequency was calculated based on the frequency of –GG homozygote and assuming Hardy-Weinberg Equilibrium. “( )”= 95% Confidence Interval for of the minor allele frequency. Highlighted rows = Frequency data obtained by genotyping the SNP in larger samples of Thai (n=59), Hmong (n=48), European-American (n=46), African-American (n=60) and Finnish subjects (n=47). For example, SNP rs34189945 was detected both by sequencing and genotyping and thus listed twice in this table.

person in the approximately 12.4 kb of the *SLC6A1* sequence was the highest in AAs (12.22) and was lowest in the Hmong population (3.75). In Finns, the mean number of heterozygous SNPs was 8.63, in EAs it was 6.57, and in Thai it was 8.38 (Table 3). The number of heterozygous SNPs differed significantly by population (ANOVA $p < 0.0001$). *Post hoc* comparisons showed that the Hmong had fewer heterozygous SNPs than the other populations (each comparison $p<0.009$), with the exception of EAs ($p=0.11$). AAs had more heterozygous SNPs than each of the other populations (each comparison $p<0.032$)

Table 3 Indices of *SLC6A1* nucleotide diversity in five populations.

Population	Number of polymorphic site	Nucleotide diversity per bp (π) (10^{-4})	Watterson's estimator of theta (θ) (10^{-4})	Mean number of heterozygous SNPs per person
European-American	21	1.72	1.41	6.57
African-American	41	2.77	2.55	12.22
Finnish	29	2.11	1.87	8.63
Thai	31	2.14	2.00	8.38
Hmong	24	1.62	1.55	3.75

Linkage disequilibrium in the *SLC6A1* gene

For analysis of linkage disequilibrium in *SLC6A1*, a total of 16 SNPs were genotyped in the AA, Thai, Hmong, EA, and Finnish samples. The SNPs were selected with the goal of encompassing the *SLC6A1* gene with SNPs with allele frequencies >10% in most populations studied in order to allow comparison of LD patterns between populations. Twelve of the 16 SNPs studied met this criterion. SNP rs1710879 was virtually monomorphic in AA population. .The allele frequencies of all 16 SNPs were in Hardy-Weinberg equilibrium (HWE). The LD structure of the *SLC6A1* gene, as detected by this set of SNPs, was evaluated by calculating D' and r^2 using the HAPLOVIEW program (58). The LD structure of the *SLC6A1* gene is presented in Figure 2. Fragmentation of LD into several poorly-defined blocks was noted in all 5 populations studied (Figure 3). There were two short blocks of LD (D' = 0.8-1) observed in all five populations. The first block is located between the markers -29477 (Marker 1, Figure 3) and -24321 (Marker 2, Figure 3). The second LD block,

which is located between markers -17590 (Marker 3, Figure 3) and -9765 (Marker 5, Figure 3), was also found in all populations studied, although the level of LD was lower in AAs. The third LD block was observed in the EA and Finnish populations between markers -1529 (Marker 6, Figure 3) and 3164 (Marker 8, Figure 3).

In accordance with low levels of LD, *Tagger* identified few haplotype tagging SNPs (59). A SNP tagging approach would have allowed omission of three SNPs each in the EA, Finnish, and Thai populations and omission of two SNPs in the Hmong population. None of the 16 SNPs examined in the AA population was identified as a haplotype tagging SNP. We estimated the span of LD in different populations using r^2/distance as an index. The index of LD span was about twofold higher in the Hmong population than in any of the other populations. The differences were greatest in the distance bins <10 kb and 30-40 kb, where the median r^2/distance value was two-to-three fold higher in the Hmong than in the other populations (Figure 4).

Haplotype diversity in *SLC6A1*

To further elucidate the *SLC6A1* haplotype structure, we used PHASE (54, 55) to estimate haplotype frequencies in the five populations. Consistent with the observation of low levels of LD, no common *SLC6A1* haplotypes spanning the entire gene were identified in any of the populations. Haplotype dispersion varied depending on the number of SNPs included in the haplotype. When all 16 SNPs were included in the analysis, no common haplotypes were observed. The most common haplotype had a frequency of 0.048 in the EA, 0.039 in the AA, 0.076 in the Finnish, 0.049 in the Thai and 0.097 in the Hmong populations. Common haplotypes were observed when the analysis was restricted to narrower segments of the gene, i.e., when there were computationally fewer possibilities. We noted that there were differences between populations in the composition and number of common *SLC6A1* haplotypes. To examine this further, we reconstructed each consecutive three-SNP haplotype in each population using a sliding window analysis across

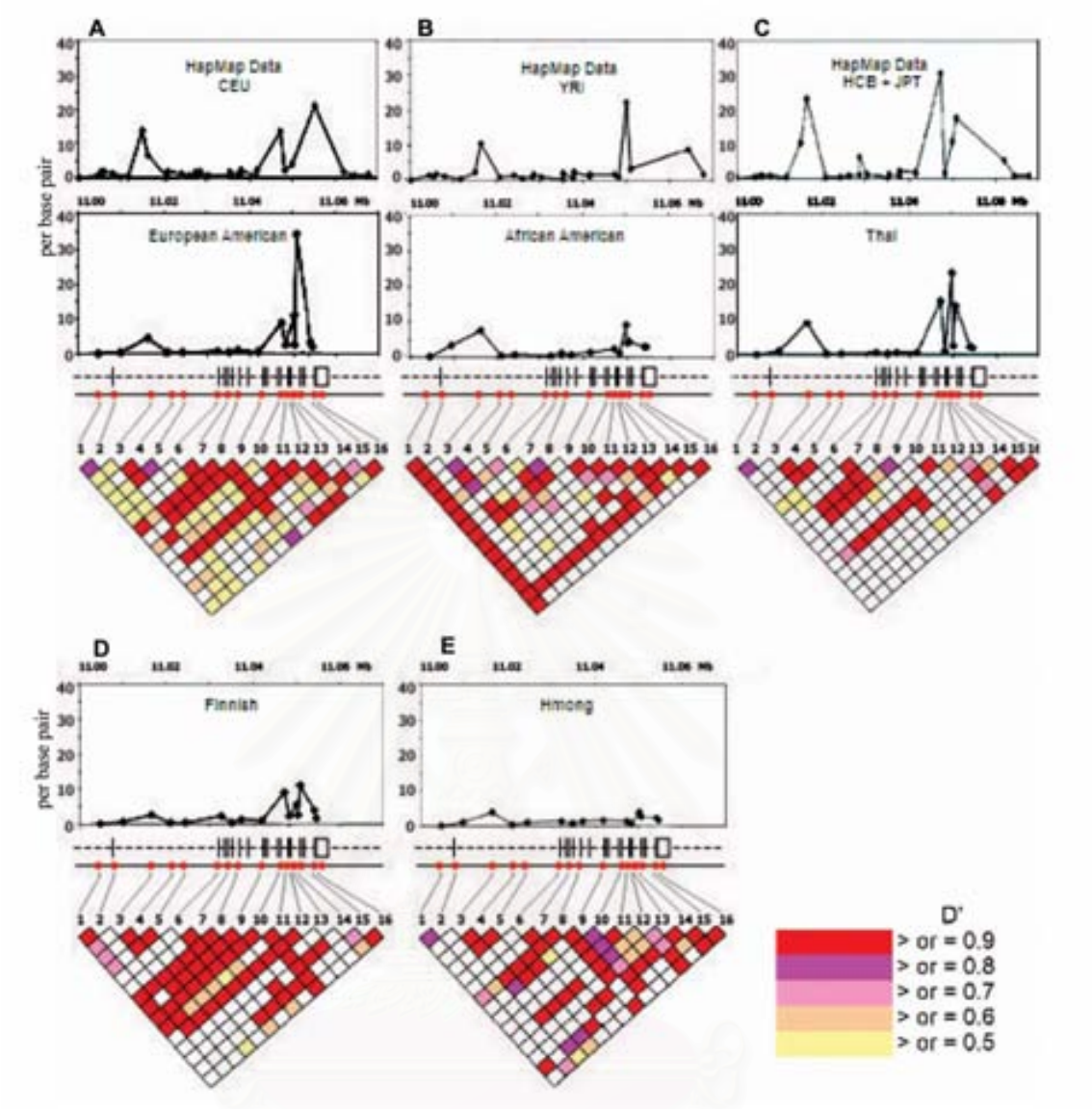


Figure 3 *Illustration of the SLC6A1 LD structure and recombination hotspots in the 5 populations. Upper graphs illustrate elevations from the background recombination rates across the SLC6A1 gene. Y axis represent recombination rate and X axis represents physical distance between the markers. In Figure A, recombination rates for HapMap CEPH Western Europeans (CEU) and European-Americans of the present study are presented. In Figure B, recombination rates for HapMap Yoruban (YRI) and African-Americans of the present study are presented. In Figure C, recombination rates for HapMap combined Han Chinese and Japanese populations (HCB+JPT) and the Thais of the present study are presented. In Figures D and E, recombination rates for the Finns and Hmongs of the present study are presented. In the middle, the exon-intron structure of SLC6A1 and*

location of the markers is presented. LD (D') between the SNPs in SLC6A1 is illustrated in the lower triangular graphs.

the panel of 16 SNPs. The four most common three-SNP haplotypes in each window and in each population were identified. We then calculated how many times each of the common three-SNP haplotypes in each window was disjoint (i.e., not shared) between the populations. A summary pairwise score was calculated for each of the populations, which is presented in Table 4. For example, of all "top-four" three-SNP *SLC6A1* haplotypes, 26.8% were disjoint between AAs and EAs.

Table 4. A summary pairwise percentage score representing the degree to which the four most common haplotypes were disjoint between the five populations.

E	100				
A	26.8	100			
F	16.1	26.8	100		
T	12.7	23.6	16.3	100	
H	26.8	23.2	30.3	17.9	100
	E	**A**	**F**	**T**	**H**

E = European-American, A = African-American, F = Finnish, T = Thai, H = Hmong. For example, of all consecutive "top-four" 3-SNP SLC6A1 haplotypes, 26.8 % were disjoint between African-Americans and European-Americans.

Recombination hotspots in *SLC6A1*

We considered recombination hotspots as an explanation for low level LD in *SLC6A1*. Recombination rates were calculated for the genotype data using the –MR and –X10 options of the PHASE program (56, 57) and genotype data for all 16 *SLC6A1* SNPs. Average recombination rates in the five populations are shown in Figure 3. These data show two areas in *SLC6A1* with elevated recombination rates. The first area is demarcated by markers -29477 (Marker 1) and -13071 (Marker 4) and the second one by markers 7772 (Marker 9) and 20172 (Marker 15), hereafter referred to as the upper and lower hotspots, respectively. There were population differences in the amplitude of the hotspots (Figure 3). In the upper hotspot, which is

located in the areas of intron 1 and exon 1, an increase in the recombination rate was observed in all five populations. The Hmong and AA populations showed lower rate and width of the lower hotspot, located in the area between exon 8 and 16, than did the EAs, Finns or Thai (Figure 3). To further elucidate these findings, we analyzed *SLC6A1* genotypes available through HapMap using PHASE, as described above. Genotyping data for the combined group of Japanese and Han Chinese, Yoruba, and Western Europeans were analyzed and compared to our results. We compared genotype data from our 16 SNPs and genotype data from 27-31 SNPs in HapMap. Two sets of SNPs, our and the HapMap SNPs, are not exactly the same but they are all located within the *SLC6A1* gene (one dot on X-axis in Figure 3 represents one SNP). It appears that genotype data from the two samples (the present study and HapMap) are consistent, both showing evidence for two areas of increase in recombination rate within *SLC6A1*. These results support our findings (Figure 3) regarding two recombination hotspots in *SLC6A1*.

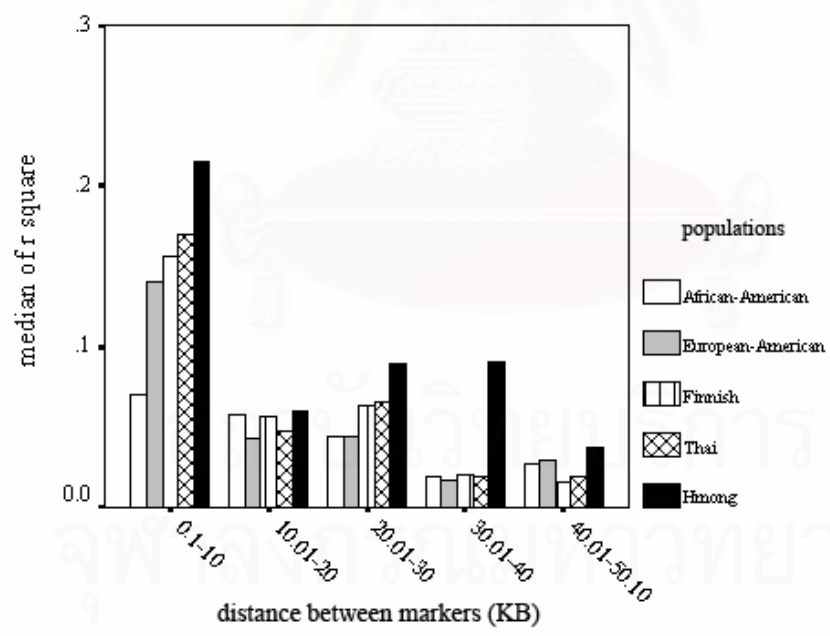


Figure 4 *This graph illustrates LD span (expressed as median r^2) in different populations. Median r^2 calculated for SNP pairs in different distance bins (0.1-10kb, 10.01-20kb etc) is presented.*

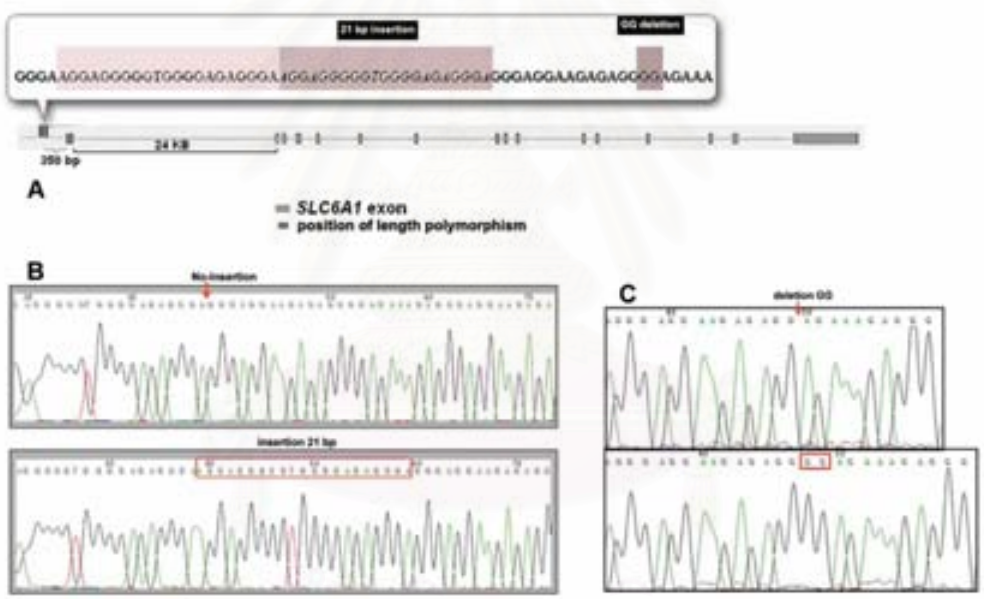
Novel length polymorphisms in the promoter region

Two novel length polymorphisms were identified in the upper promoter of *SLC6A1*. The largest, and perhaps the more interesting polymorphism, is a 21-bp VNTR polymorphism, which is present either as a single element or as two tandem

repeats, hereafter referred to as the *SLC6A1* short and *SLC6A1* long allele, respectively.

A novel two-base-pair deletion (GG/-GG) at positions (-24780 to -24781) was present in all 5 populations studied. This polymorphism is located only 8 bp 3' to the last base pair of the short/long polymorphism. The -GG allele occurs only in the context of the short allele.

The *SLC6A1* short/long and GG/-GG polymorphisms were genotyped in 46 EA, 60 AA, 59 Thai, 47 Finnish and 48 Hmong individuals. The allele frequency of the long allele was 0.39 in the AA sample but it was not found in other populations. The GG/-GG polymorphism occurred in all 5 populations. The -GG allele frequency was 0.30 in EA, 0.23 in AA, 0.33 in Finnish, 0.22 in Thai and 0.16 in Hmong.



E

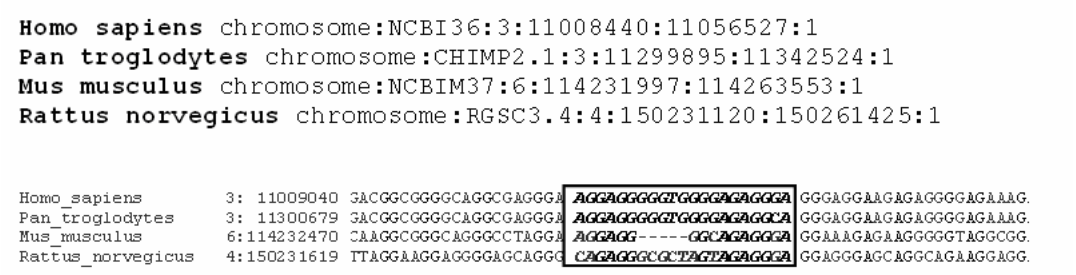


Figure 5 Illustration of the 21 bp insertion/deletion polymorphism (SLC6A1 long and short alleles) and the insertion/deletion GG allele in the SLC6A1 gene. Picture (A) shows position of 21 bp insertion and GG deletion. Picture (B and D) shows the sequence for the SLC6A1 short and long alleles including probe sequence for EMSA study. Sequence for the GG insertion/deletion polymorphism is presented in picture (C) and homology of 21bp insertion with other specie is presented in (E).

Population differences

We previously genotyped the *SLC6A1* VNTR polymorphism (Figure 5) in 46 European-American, 60 African-American, 59 Thai, 47 Finnish and 46 Hmong Chinese individuals (62) 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 verify this hypothesis we genotyped this polymorphism in Tanzanian individuals. In the Tanzanian cohort, we obtained genotype frequencies of 2.9, 30.4, and 66.7% for the Insertion/Insertion, Insertion/Non-insertion, and Non-insertion/Non-insertion genotypes, implying an allele frequency of 18.1% (N=69 individuals genotyped). These genotype frequencies do not deviate significantly from Hardy-Weinberg equilibrium (Chi-square=1.44, Df=2, p>.05). We have not observed the insertion allele in hundreds of additional genotypes 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. We hypothesized that the insertion allele may affect promoter function

and may be an important genetic determinant of *SLC6A1* promoter function in individuals with African ancestry.

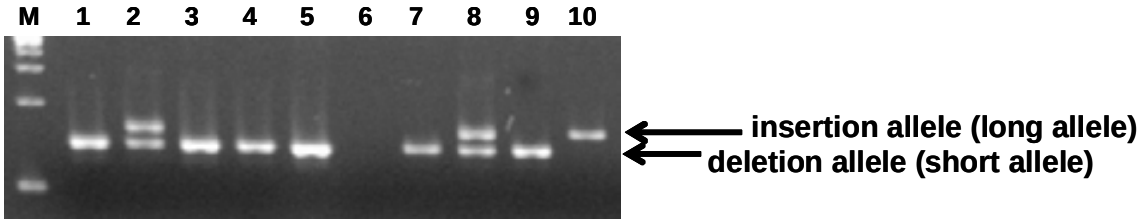
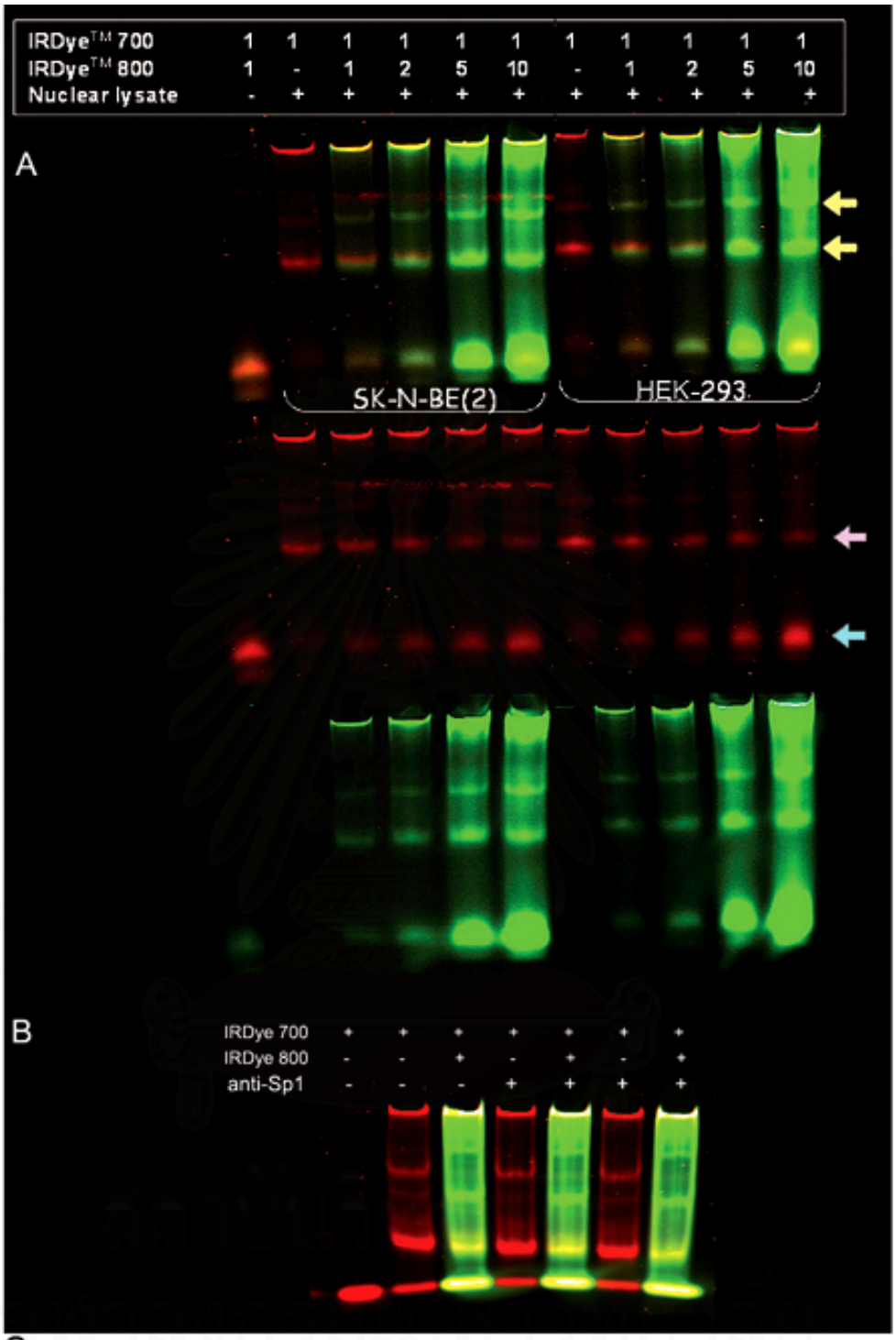


Figure 6 *PCR and gel electrophoresis of 21 bp insertion/deletion SLC6A1 promoter region. (lane M = 100 bp ladder, lane 1,3-5, 7, 9 = homozygous deletion, lane 2 and 8 = heterozygous, lane 10 = homozygous insertion, and lane 6 = negative control.)*

EMSA study

Because the 21-bp insertion polymorphism increased the activity of the *SLC6A1* promoter, we hypothesized that the VNTR may code for a protein binding motif that interacts with a nuclear protein, such as a transcription factor, to enhance transcription, and that the insertion variant may bind such a protein more effectively than the non-insertion variant. We generated a double stranded DNA probe coding for the insertion sequence labeled with one infrared fluorophore (IRDye 700) and interacted this probe with nuclear proteins from HEK-293 and SK-N-BE(2) cells in gel shift assays. To establish specificity we utilized 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 figure 7A, 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 the insertion sequence.

We used a bioinformatics approach to predict transcription factors that can bind within the vicinity of the VNTR using the TFSEARCH program (63). This analysis suggested the presence of overlapping Myeloid Zinc Finger 1 (MZF1) and SP1 binding sites in this region (Figure 7C). MZF1 is a transcription factor involved in



C

```
1  AAGGAGAGAG ATTGGAGCGG GACGGCGGGG CAGGCGAGGG AAGGAGGGGG  entry          score
                                                        --  M00083 MZF1   98.3
          --------->                                        M00077 GATA-3 88.1
                                                      ----  M00008 Sp1    86.3

                              21 bp insertion
51 TGGGGAGAGG GAAGGAGGGG GTGGGGAGAG GGAGGGAGGA AGAGAGGGGA  entry          score
   ----->      **********                                   M00083 MZF1   98.3
               1 copy of VNTR
               --------->                                   M00083 MZF1   98.3
                                                 ------->  M00083 MZF1   94.8
   ----->                                                   M00008 Sp1    86.3
                         ---------->                        M00008 Sp1    86.3
   ------------->                                           M00084 MZF1   85.5
```

Figure 7 *EMSA was used to determine whether nuclear proteins specifically interact with the VNTR. 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 the VNTR labeled with IRdye-700 was used as a probe and a single copy of the VNTR 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.*

hemopoietic development and is also expressed in parathyroid tissue, making this a poor candidate for neuronal regulation. In contrast, SP1 is ubiquitously expressed and is involved in transcriptional regulation in the brain (64-68). We therefore attempted to supershift the VNTR-protein bands with anti-human SP1 but were unable to do so (Figure 7B) suggesting that SP1 is not involved in binding the VNTR.

SLC6A1 Promoter Activity

To determine the functional significance of the insertion polymorphism on *SLC6A1* promoter activity, promoter variants were used to drive luciferase expression in reporter assays in three cells lines (figure 8). As seen in Figure 2, the insertion and deletion variants showed significantly more activity than the promoterless control (pGL3-basic) construct. The pGL3-GAT1-Ins (insertion) construct showed significantly higher activity than the pGL3-GAT1-Del (noninsertion) variant in all 3 cell lines tested. (Figure 2) More than 30-fold more luciferase activity was noted in HEK-293 and SK-N-BE(2) cells from the insertion variant relative to the pGL3 control vector, and 13-fold more activity than control was noted in P19 cells. When the activity of pGL3-GAT1-Ins was compared to pGL3-GAT1-Del (noninsertion), P19 and HEK-293 cells ($p=0.001$) as well as SK-N-BE(2) cells ($p=0.029$) showed significantly more activity for the insertion than the deletion

variant. These data suggest the 1.4 kb region upstream from exon 1 has promoter activity and that the 21-bp insertion polymorphism significantly increases the activity of this *SLC6A1* promoter element.

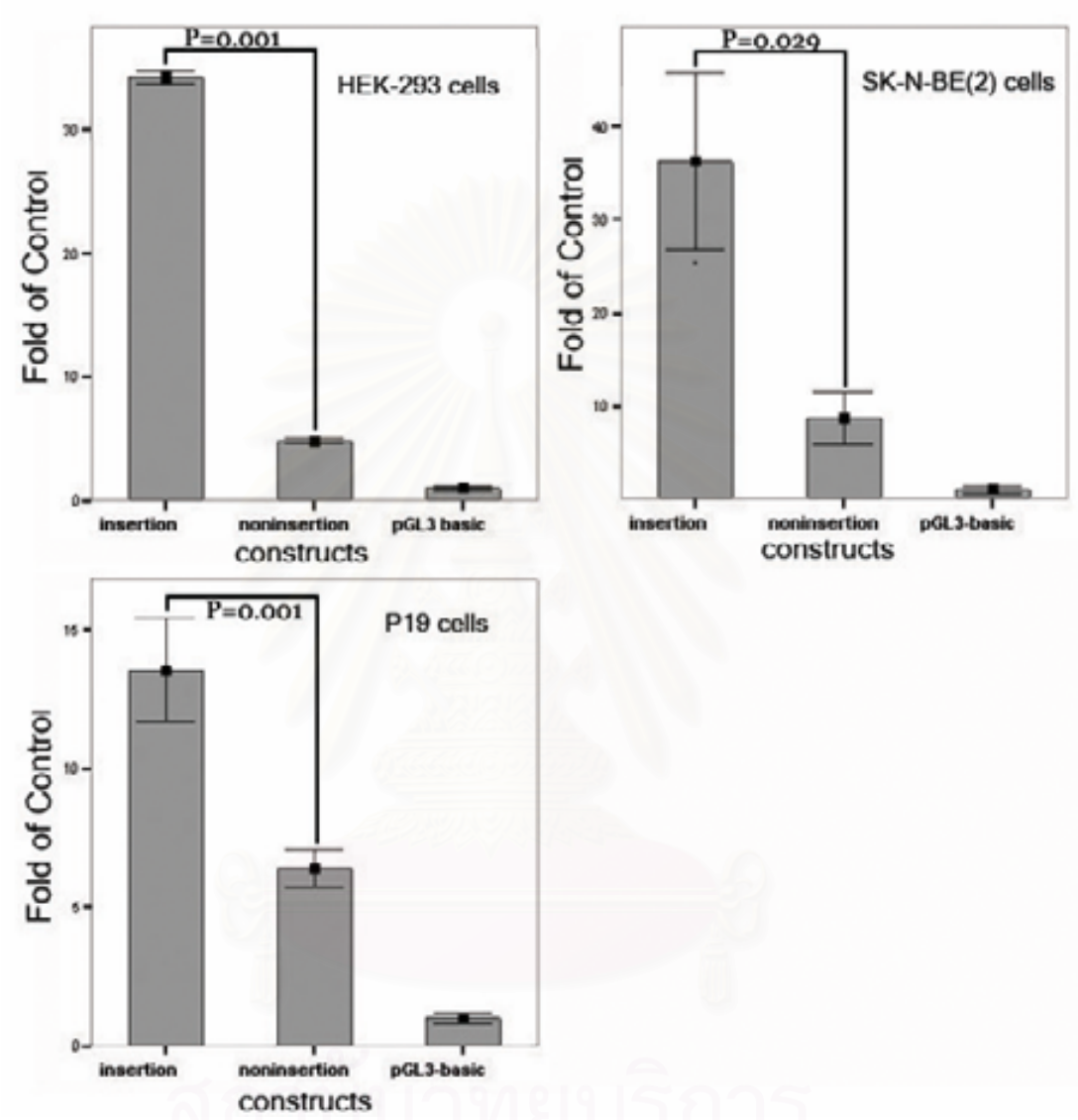


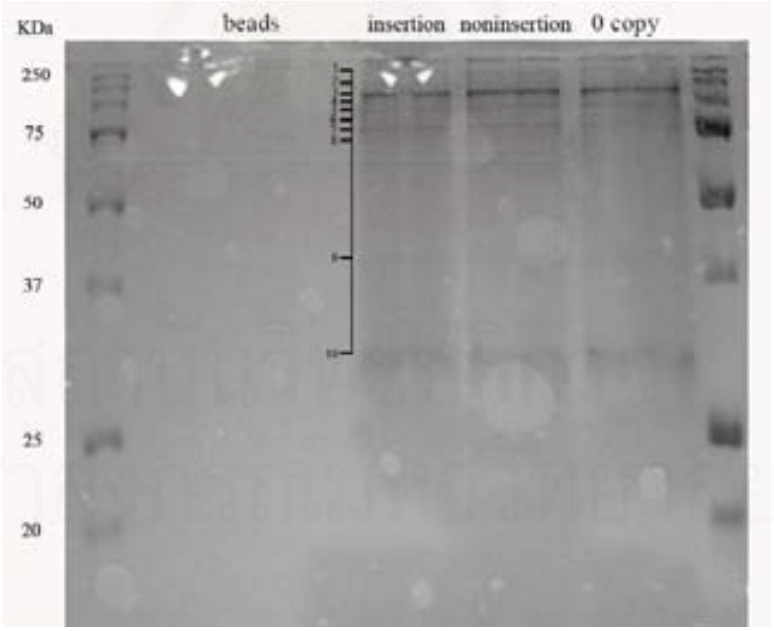
Figure 8 *Luciferase activity in HEK-293, SK-N-BE(2) and P19 cells. The x axis represented type of DNA construct and Y represented activity in fold of control (pGL3-basic) scale.*

Pull-down and transactivation studies

We used two approaches to identify proteins that interact with the VNTR 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 0, 1, or 2 copies of the VNTR were interacted with nuclear lysates from human neuroblastoma cells (SK-N-BE(2)), captured with streptavidin magnetic beads, separated by SDS-PAGE, 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 (LC-MS/MS). Unfortunately no protein bands were identified that were specifically associated with the insertion variant. The proteins that did bind to the promoter probes were therefore analyzed together. Pathway analysis and annotation using STRING (69) revealed that many of the proteins identified undergo known protein-protein interactions (Figure 9B, Table 5). These data define a protein complex that interacts with the *SLC6A1* promoter that may shed light on the transcriptional efficacy of the promoter. However, these DNA-protein interactions cannot account for the differences in promoter activity between the insertion and non-insertion variants.

A)



B)

Figure 9 *All pulled down proteins were subjected to SDS-PAGE and silver stained (A). "beads" represented no DNA probe in binding reaction, only beads mix with nuclear lysate from SK-N-BE(2) cells. "insertion and noninsertion" represented binding reaction that using biotinylated DNA probe containing 21-bp insertion and noninsertion, respectively. "0 copy" represented no VNTR copy in sequence of biotinylated DNA probe. Interaction of pulled down proteins based on STRING and Harvester seach tool was shown in (B). From MS/MS protein identification, all proteins in (B) reported at least 10 peptides matched the protein.*

Table 5. Non-interacted proteins from pull down experiment and peptides matched score.

Name	Peptides matched score
Heterogeneous nuclear ribonucleoprotein R (HNRPR)	0.027
Antigen identified by monoclonal antibody Ki-67 (MKI67)	0.017
Myosin, heavy polypeptide 10, non-muscle (MYH10)	0.015
Heterochromatin protein 1, binding protein 3 (HP1BP3)	0.015
KIAA0179	0.015
DEAD (Asp-Glu-Ala-Asp) box polypeptide 18 (DDX18)	0.014
Synaptotagmin binding, cytoplasmic RNA interacting protein (SYNCRIP)	0.013
N-acetyltransferase 10 (NAT10)	0.012
Myosin, heavy polypeptide 9, non-muscle (MYH9)	0.009
Programmed cell death 11 (PDCD11)	0.009
DEAD (Asp-Glu-Ala-Asp) box polypeptide 5 (DDX5)	0.009
Heterogeneous nuclear ribonucleoprotein A/B (HNRPAB)	0.007

NFAT signaling inhibition

Based on the proteomics results, NF90 (also known as ILF3) was pulled down by all 3 DNA probes. NF90 is one subunit of the Nuclear Factor of Activated T Cells (NFAT) DNA protein dimer, suggesting that NFAT may regulate expression of *SLC6A1* gene. We studied *SLC6A1* promoter activity under conditions of variable NFAT activity. NFAT is strongly regulated by calcineurin, and Cyclosporine A (CsA) (70) (Figure 10), a potent inhibitor of calcineurin, would be expected to modulate *SLC6A1* activity if the *SLC6A1* promoter is regulated by NFAT. HEK-293 cells were treated with CsA for 30 minutes before stimulation with the phorbol-12-myristalate-13-acetate (PMA) for 3 hours. Luciferase activity in the insertion variant expressing

cells was unchanged by PMA or CsA treatment (Figure 11). These data suggest that although NFAT may interact with the *SLC6A1* promoter, it does not appear to modulate its activity.

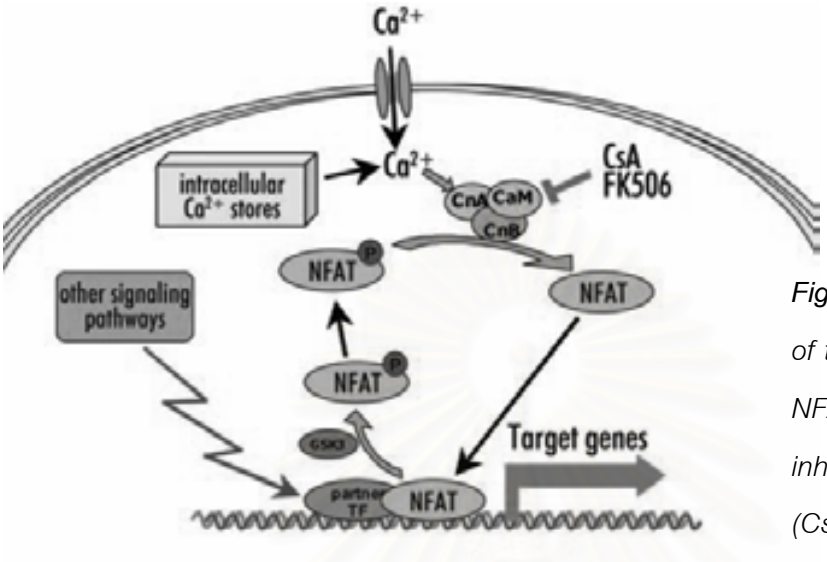


Figure 10 *The illustration of the gene regulation by NFAT protein and its inhibitor, Cyclosporin A (CsA).*

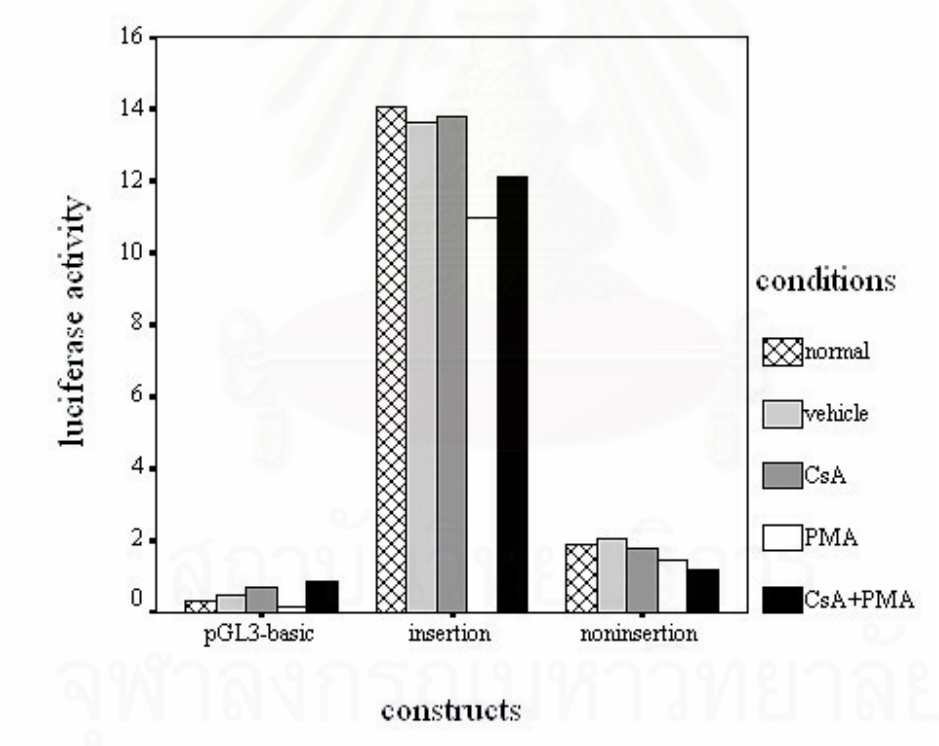


Figure 11 *GAT-1 activity under conditions of variation of NFAT activity. Different conditions represented by different colors of bar.*

Family-based study of novel 21 bp insertion

We analyzed this polymorphism in 848 persons in 249 African-American families by FBAT program (61). Two genetic models were tested (additive, and the computationally identical dominant and recessive). The frequency of this polymorphism in this studied population was 16%. No significant association result between this novel length polymorphism with cocaine dependence (also alcohol dependence). The p-value was 0.352 (Z=0.931) and 0.270 (Z=1.104) in cocaine dependence and alcohol dependence, respectively. Thus, this 21 bp insertion is not associated with cocaine and alcohol dependence in African-American families.

CHAPTER V
CONCLUSION AND DISCUSSION

Conclusion

1. We totally found 63 polymorphisms on *SLC6A1* gene by resequencing.

2. The *SLC6A1* sequence diversity was found highest in African-American population.

3. *SLC6A1* is complicated target for pharmacogenetic studies because its low levels of LD and presence of recombination hotspots would make it difficult to establish an association between genetic markers and response to GAT-1 modulation.

4. Our study suggests that focusing on isolated populations, such as Finns or Hmong, would not provide major benefits to genotyping efficacy of *SLC6A1.*

5. An interesting 21-bp VNTR polymorphisms (*SLC6A1* short and *SLC6A1* long) was discovered in the promoter sequence of the *SLC6A1* gene, especially common in African Americans. This new polymorphism is suggestively a novel candidate for future studies focusing on genetic differences in GAT-1 expression, and hence, response to medications that inhibit GAT-1 function.

6. Some proteins interact with 21-bp insertion polymorphism and lead to a dramatic increase in *SLC6A1* promoter activity.

7. We fail to find the protein (transcription factor) that controls and enhance *SLC6A1* promoter activity.

8. There was no association between 21-bp insertion with cocaine and alcohol dependence.

9. The 21-bp insertion in the promoter region may prove to be useful in predicting clinical response to pharmacological modulators of SLC6A1 as well as GABAergic function especially in individuals of African descent.

Discussion

The goal of the present study was to comprehensively analyze sequence variation and linkage disequilibrium in the *SLC6A1* gene in anticipation of larger

pharmacogenetic studies of tiagabine and other GAT-1 inhibitors. Resequencing 12.4 kb of the *SLC6A1* gene, including all 16 exons and the two putative promoter regions, revealed numerous novel genetic variants. Perhaps the most interesting polymorphism identified was a 21 bp VNTR polymorphism located in the upper promoter sequence of the *SLC6A1* gene (Figure 5). We have termed the alleles as the "*SLC6A1* short," which has one copy of the allele and "*SLC6A1* long," in which the allele is duplicated (Figure 5). Interestingly, the long allele was common in AAs (39%), while in the other populations, it was absent. A likely explanation for the lack of this allele in non-African populations is genetic drift. However, other explanations, such as natural selections, are also possible. Functional studies focusing on understanding whether the short and long allele lead to differential expression of the GAT-1 protein are clearly warranted; these studies may also help in elucidating whether the allele frequency discrepancy between African-Americans and other populations is due to genetic drift or selection. Previously, a comparable promoter region VNTR polymorphism was described in the serotonin transporter gene, which is known to influence the expression of the gene (71). Several studies have shown that this polymorphism partially accounts for differences in the therapeutic response to serotonin selective reuptake inhibitors (SSRIs) (72-75), susceptibility to depression (76, 77) and alcohol dependence (78). The *SLC6A1* short/long is a candidate polymorphism for moderating the response to tiagabine and susceptibility to neuropsychiatric disorders in which GABA dysfunction may play a role, albeit only in AA populations (and any other populations where this variant is present).

Genetic diversity in *SLC6A1* revealed through *SLC6A1* resequencing showed interesting results. Although a limited number of chromosomes were examined, certain trends were found. First, no non-synonymous SNPs were discovered in the 80 chromosomes sequenced, suggesting that the coding sequence of *SLC6A1* has been conserved against common amino-acid altering substitutions through active background selection. Consistent with these results, the nucleotide diversity was much lower in the *SLC6A1* exons compared to the intronic regions examined (79-81). Comparison of the sequence data between populations revealed higher nucleotide diversity in AAs than in other populations (Table 3). In

accordance with these data, the only population in which we found common population specific SNPs, were AAs. In exons, however, nucleotide diversity was no higher in the AA population than in the other populations. In other populations, the SNPs observed in only one population were rarer. Although Finns and Hmong are considered to be isolated populations, nucleotide diversity in these two populations was not different from that observed in EAs or Thais. Overall, no major differences in nucleotide diversity were observed among the Hmong, Thai, EA, and Finnish populations. Together, these findings most likely reflect the older age of the African population relative to the other populations, which had allowed more intronic variation to accumulate in this population, founder effects in non-African populations, and selection pressure conserving the *SLC6A1* exonic sequence. The extent of conservation of GAT-1 amino acid sequence suggests an important role of this protein in normal brain function. The Hmong had a significantly lower number of heterozygous SNPs in comparison to the other populations. One explanation for this finding is that the Hmong subjects may have been distantly related. We postulate that differences in the degree and age of population bottlenecks between the Hmong and the other populations are less likely explanations for the lower heterozygosity observed in the Hmong; we feel that this explanation is less likely because all non-African populations had a low observed frequency of population-specific SNPs and because Hmong nucleotide diversity was not significantly lower than that of the other non-African populations (82). A caveat of this study is that the sample size per ethnicity was small. Consequently, rare non-synonymous SNPs, specific to a population, would have not been identified. It may be useful to resequence larger samples to identify these kinds of variations, at least in the primary target populations of clinical trials. Laboratory methods established for the present study should facilitate such analyses. Another limitation of the study is that the resequencing effort focused on exons. If deep intronic variation in *SLC6A1* contributes to functional variation at the protein level, those variants would have been missed in the present study. A low level of LD in *SLC6A1* was observed in all five populations (Figure 3). Consistent with these results, in the EA, Thai, Hmong, and Finnish populations, only two or three haplotype tagging SNPs in the areas of

preserved LD were identified. In the AAs, no haplotype tagging SNPs were found in the 16 SNPs genotyped. These results suggest that very dense SNP panels would be required to capture common variation in this gene. Using r^2/distance as the index, a longer LD span was observed in the Hmong population than in the other populations (Figure 4). However, higher LD probably would not translate to significant practical improvement in genotyping efficiency overall, because there were no major differences in the number of haplotype blocks in Hmong than in the Finnish, Thai and EA populations. Considering the low level of LD, *SLC6A1* may pose special challenges for association studies both in isolated and mixed populations. The common *SLC6A1* 3-SNP haplotypes were largely the same in the five populations (Table 4), but were not completely overlapping. These results suggest a certain degree of, but not absolute, portability of SNP genotyping sets between the populations. It would interesting to study larger EA, AA, Hmong, Thai and Finnish population samples to further refine the structure and frequencies of the common overlapping and population-specific haplotypes (83). In addition, it will be interesting to study whether haplotype and SNP profile characteristics, such as absence of common non-synonymous substitutions extends to patient populations suffering from various neuropsychiatric disorders, which were not studied here. The present study primarily focused on non-clinical samples and therefore no data were available to assesses whether disease associated variants are present in human populations.

Intrigued by these findings, we examined whether recombination hotspots could explain low levels of LD in *SLC6A1*. Two hotspots were identified using PHASE; the first is located in the areas of exon 1 and intron 1. The second hotspot is located in the area demarcated by exons 8 and 16. As expected, within the hotspots, D' fell off rapidly. For example, in the Finns, in the area of the distal hotspot, D' was only 0.184 between markers 12 and 13, which are spaced 107 bp apart and D' was 0.044 between 13 and 14, which are spaced 489 bp apart. Areas of high recombination, such as seen in *SLC6A1*, potentially limit large-scale association studies, as it would be exceptionally difficult to find risk alleles relying on linkage disequilibrium if the alleles were located inside a recombination hotspot.

We try to pull our data together and try to explain descriptive characteristic of *SLC6A1* gene. Some part of data conflict with others. For example, this gene presented high recombination and low level of LD. But showed lower nucleotide diversity than previous data in human genome (25). In addition, the simulation in coalescent model from Ardlie KG *et al* study (26) reported that nucleotide diversity will be increase when D' and r^2 decrease. One possibility is that *SLC6A1* gene needs to be conserved in nucleotide and amino acid levels conducting to several processes of repair via recombination mechanism (84). Then, it can possible make this gene has high recombination rate, low LD and low diversity as presenting in our result.

The Recent African Origin (RAO) 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. Because 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 (85). Recent data from mitochondrial genome sequencing suggests that Tanzanians have very high genetic diversity, greater than other African populations, suggesting the possibility that Tanzanians represent the modern descents of this original source population (86). Our data are consistent with a model where the 21-bp insertion variant arose prior to 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. However, other models are also possible including genetic drift and natural selection.

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 the 21-bp insertion polymorphism. 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 due to the insertion variant may serve to bias the ratio of upstream to downstream promoter transcripts generated. Whether such a shift occurs and the functional implications of such a shift will require further research.

The insertion polymorphism consists of duplication of a VNTR (5'-AGGAGGGGGTGGGGAGAGGGA-3'). This sequence contains an (A/U)GGG repeat motif. These motifs have been associated with enhancement of splicing (87). It is possible that the insertion may increase the efficacy of intron removal and increase the availability of the mature RNA. However, this mechanism cannot account for the increased activity of the promoter region in isolation in our assays. The insertion region is particularly rich in guanosine nucleotides. Guanosine rich regions of gene promoters are associated with particularly high affinity for transcription factors (88). Consecutive guanosine nucleotides are also known to affect secondary structure. These findings have important implications for future phamacogenetic studies, as this genetic variation is likely to affect clinical response in the African population 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 -non-convulsive status epilepticus - which has raised concerns about its off-label use in the treatment of psychiatric disorders (11). Our results suggest that patients with the 21-bp insertion may require a higher dose of tiagabine than patients without the insertion variant to achieve the same level of clinical response. 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.

REFERENCES

[1] LaRoche SM, Helmers SL. The new antiepileptic drugs: clinical applications. Jama. 2004 Feb 4;291(5):615-20.

[2] Sofuoglu M, Kosten TR. Novel approaches to the treatment of cocaine addiction. CNS Drugs. 2005;19(1):13-25.

[3] Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. Genome Res. 2005 Nov;15(11):1592-3.

[4] 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 Sep 22;220(2-3):197-201.

[5] 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 Oct;64(10):1245-9.

[6] Mathias S, Wetter TC, Steiger A, Lancel M. The GABA uptake inhibitor tiagabine promotes slow wave sleep in normal elderly subjects. Neurobiol Aging. 2001 Mar-Apr;22(2):247-53.

[7] 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 Jan;67(1):66-71.

[8] Gonzalez G, Sevarino K, Sofuoglu M, Poling J, Oliveto A, Gonsai K, et al. Tiagabine increases cocaine-free urines in cocaine-dependent methadone-treated patients: results of a randomized pilot study. Addiction. 2003 Nov;98(11):1625-32.

[9] 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 Sep;181(3):504-10.

[10] Chiu CS, Brickley S, Jensen K, Southwell A, McKinney S, Cull-Candy S, et al. GABA transporter deficiency causes tremor, ataxia, nervousness, and increased GABA-

induced tonic conductance in cerebellum. J Neurosci. 2005 Mar 23;25(12):3234-45.

[11] Flowers CM, Racoosin JA, Kortepeter C. Seizure activity and off-label use of tiagabine. N Engl J Med. 2006 Feb 16;354(7):773-4.

[12] Kurland L, Lind L, Melhus H. Using genotyping to predict responses to anti-hypertensive treatment. Trends Pharmacol Sci. 2005 Sep;26(9):443-7.

[13] Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, et al. The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet. 2000 Jul;25(3):320-3.

[14] Shifman S, Darvasi A. The value of isolated populations. Nat Genet. 2001 Aug;28(4):309-10.

[15] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science (New York, NY. 2001 Feb 16;291(5507):1304-51.

[16] Shastry BS. SNPs in disease gene mapping, medicinal drug development and evolution. Journal of human genetics. 2007;52(11):871-80.

[17] Machalek AZ. Pharmacogenetics Research Leads the Way to Individualized Therapy. http://wwwnigmsnihgov/News/Briefs/PharmacogeneticsResearchhtm. 2006.

[18] Pare CMB, Rees L, Sainsbury MJ. Differentiation of two genetically specific type of depression by the response to anti-depressants. The Lancet. 1962;2:1340-3.

[19] Franchini L, Serretti A, Gasperini M, Smeraldi E. Familial concordance of fluvoxamine response as a tool for differentiating mood disorder pedigrees. Journal of Psychiatry Research. 1998;32:255-9.

[20] Weber WW. Effect of pharmacogenetics on medicine. Environ Mol Mutagen. 2001;37(3):179-84.

[21] Diatchenko L, Slade GD, Nackley AG, Bhalang K, Sigurdsson A, Belfer I, et al. Genetic basis for individual variations in pain perception and the development of a chronic pain condition. Human molecular genetics. 2005 Jan 1;14(1):135-43.

[22] Capon F, Allen MH, Ameen M, Burden AD, Tillman D, Barker JN, et al. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates

association with psoriasis across diverse ethnic groups. Human molecular genetics. 2004 Oct 15;13(20):2361-8.

[23] Ian Harrison ML, Eleanor Sterling. Genetic diversity. Connexions module. 2004.

[24] Wikipedia. Human genetic diversity.

.

[25] Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 1979 Oct;76(10):5269-73.

[26] Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet. 2002 Apr;3(4):299-309.

[27] Bamshad M, Wooding SP. Signatures of natural selection in the human genome. Nat Rev Genet. 2003 Feb;4(2):99-111.

[28] Hey J. What's so hot about recombination hotspots? PLoS Biol. 2004 Jun;2(6):e190.

[29] Laan M, Paabo S. Demographic history and linkage disequilibrium in human populations. Nat Genet. 1997 Dec;17(4):435-8.

[30] Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, et al. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. Hum Mol Genet. 2000 Dec 12;9(20):2947-57.

[31] Hedrick PW. Gametic disequilibrium measures: proceed with caution. Genetics. 1987 Oct;117(2):331-41.

[32] Lewontin RC. The Interaction Of Selection And Linkage. Ii. Optimum Models. Genetics. 1964 Oct;50:757-82.

[33] Mueller JC. Linkage disequilibrium for different scales and applications. Brief Bioinform. 2004 Dec;5(4):355-64.

[34] Conti F, Minelli A, Melone M. GABA transporters in the mammalian cerebral cortex: localization, development and pathological implications. Brain Res Brain Res Rev. 2004 Jul;45(3):196-212.

[35] Fan HP, Fan FJ, Bao L, Pei G. SNAP-25/syntaxin 1A complex functionally modulates neurotransmitter gamma-aminobutyric acid reuptake. J Biol Chem. 2006 Sep

22;281(38):28174-84.

[36] http://www.ensembl.org.

[37] .http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene.

[38] http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC6A1&search=SLC6A1.

[39] Makikyro TH, Hakko HH, Timonen MJ, Lappalainen JA, Ilomaki RS, Marttunen MJ, et al. Smoking and suicidality among adolescent psychiatric patients. J Adolesc Health. 2004 Mar;34(3):250-3.

[40] Yang BZ, Zhao H, Kranzler HR, Gelernter J. Practical population group assignment with selected informative markers: characteristics and properties of Bayesian clustering via STRUCTURE. Genet Epidemiol. 2005 May;28(4):302-12.

[41] Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. Am J Med Genet B Neuropsychiatr Genet. 2005 Jul 5;136(1):45-52.

[42] Gelernter J, Panhuysen C, Wilcox M, Hesselbrock V, Rounsaville B, Poling J, et al. Genomewide linkage scan for opioid dependence and related traits. Am J Hum Genet. 2006 May;78(5):759-69.

[43] Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, et al. Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). Drug Alcohol Depend. 2005 Dec 12;80(3):303-12.

[44] The Genomatix software package [**Error! Hyperlink reference not valid.**.

[45] The National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov/].

[46] The PRIMER3 software [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi].

[47] The NCBI dbSNP [http://www.ncbi.nlm.gov/SNP].

[48] Cold, Spring, Harbor. Calcium phosphate mediated transfection of eukaryotic cells. Nature Methods; 2005. p. 319-20.

[49] Grosjean F, Bertschinger M, Hacker DL, Wurm FM. Multiple glycerol shocks increase the calcium phosphate transfection of non-synchronized CHO cells. Biotechnol Lett.

2006 Nov;28(22):1827-33.

[50] 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 Apr 15;77(1):61-8.

[51] Schuelke M. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. 2000 Feb;18(2):233-4.

[52] http://harvester.embl.de/.

[53] Web application SLIDER [http://genapps.uchicago.edu/index.html].

[54] Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001 Apr;68(4):978-89.

[55] Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 2003 Nov;73(5):1162-9.

[56] Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics. 2003 Dec;165(4):2213-33.

[57] Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet. 2004 Jul;36(7):700-6.

[58] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5.

[59] de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005 Nov;37(11):1217-23.

[60] The POWERMARKER software [http://statgen.ncsu.edu/powermarker/].

[61] Horvath S, Wei E, Xu X, Palmer LJ, Baur M. Family-based association test method: age of onset traits and covariates. Genet Epidemiol. 2001;21 Suppl 1:S403-8.

[62] 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.

[63] http://www.cbrc.jp/research/db/TFSEARCH.html.

[64] Gui J, Song Y, Han NL, Zhou SF, Sheu FS. Involvement of the GC-rich sequence and specific proteins (Sp1/Sp3) in the basal transcription activity of neurogranin gene. Biochem Biophys Res Commun. 2006 Jun 23;345(1):124-32.

[65] Rossner S, Sastre M, Bourne K, Lichtenthaler SF. Transcriptional and translational regulation of BACE1 expression--implications for Alzheimer's disease. Prog Neurobiol. 2006 Jun;79(2):95-111.

[66] Lee J, Kosaras B, Aleyasin H, Han JA, Park DS, Ratan RR, et al. Role of cyclooxygenase-2 induction by transcription factor Sp1 and Sp3 in neuronal oxidative and DNA damage response. Faseb J. 2006 Nov;20(13):2375-7.

[67] Holloway K, Sade H, Romero IA, Male D. Action of transcription factors in the control of transferrin receptor expression in human brain endothelium. J Mol Biol. 2007 Feb 2;365(5):1271-84.

[68] Ramos A, Ho WC, Forte S, Dickson K, Boutilier J, Favell K, et al. Hypo-osmolar stress induces p75NTR expression by activating Sp1-dependent transcription. J Neurosci. 2007 Feb 7;27(6):1498-506.

[69] http://string.embl.de/.

[70] Buchholz M, Ellenrieder V. An emerging role for Ca2+/calcineurin/NFAT signaling in cancerogenesis. Cell Cycle. 2007 Jan 1;6(1):16-9.

[71] Heils A, Teufel A, Petri S, Stober G, Riederer P, Bengel D, et al. Allelic variation of human serotonin transporter gene expression. J Neurochem. 1996 Jun;66(6):2621-4.

[72] Smeraldi E, Zanardi R, Benedetti F, Di Bella D, Perez J, Catalano M. Polymorphism within the promoter of the serotonin transporter gene and antidepressant efficacy of fluvoxamine. Mol Psychiatry. 1998 Nov;3(6):508-11.

[73] Zanardi R, Benedetti F, Di Bella D, Catalano M, Smeraldi E. Efficacy of paroxetine in depression is influenced by a functional polymorphism within the promoter of the serotonin transporter gene. J Clin Psychopharmacol. 2000 Feb;20(1):105-7.

[74] Pollock BG, Ferrell RE, Mulsant BH, Mazumdar S, Miller M, Sweet RA, et al. Allelic variation in the serotonin transporter promoter affects onset of paroxetine treatment

response in late-life depression. Neuropsychopharmacology. 2000 Nov;23(5):587-90.

[75] Yu YW, Tsai SJ, Chen TJ, Lin CH, Hong CJ. Association study of the serotonin transporter promoter polymorphism and symptomatology and antidepressant response in major depressive disorders. Mol Psychiatry. 2002;7(10):1115-9.

[76] Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science (New York, NY. 2003 Jul 18;301(5631):386-9.

[77] Kaufman J, Yang BZ, Douglas-Palumberi H, Houshyar S, Lipschitz D, Krystal JH, et al. Social supports and serotonin transporter gene moderate depression in maltreated children. Proc Natl Acad Sci U S A. 2004 Dec 7;101(49):17316-21.

[78] Feinn R, Nellissery M, Kranzler HR. Meta-analysis of the association of a functional serotonin transporter promoter polymorphism with alcohol dependence. Am J Med Genet B Neuropsychiatr Genet. 2005 Feb 5;133(1):79-84.

[79] Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, et al. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet. 1998 Jul;19(3):233-40.

[80] Solus JF, Arietta BJ, Harris JR, Sexton DP, Steward JQ, McMunn C, et al. Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population. Pharmacogenomics. 2004 Oct;5(7):895-931.

[81] Badagnani I, Chan W, Castro RA, Brett CM, Huang CC, Stryke D, et al. Functional analysis of genetic variants in the human concentrative nucleoside transporter 3 (CNT3; SLC28A3). Pharmacogenomics J. 2005;5(3):157-65.

[82] Nielsen R. Molecular signatures of natural selection. Annu Rev Genet. 2005;39:197-218.

[83] Cheong HS, Lee SO, Choi CB, Sung YK, Shin HD, Bae SC. MERTK polymorphisms associated with risk of haematological disorders among Korean SLE patients. Rheumatology (Oxford). 2007 Feb;46(2):209-14.

[84] Cummings WJ, Zolan ME. Functions of DNA repair genes during meiosis. Curr Top Dev Biol. 1998;37:117-40.

[85] Tishkoff SA, Kidd KK. Implications of biogeography of human populations for 'race' and medicine. Nat Genet. 2004 Nov;36(11 Suppl):S21-7.

[86] Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. Mol Biol Evol. 2007 Mar;24(3):757-68.

[87] Sirand-Pugnet P, Durosay P, Brody E, Marie J. An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken beta-tropomyosin pre-mRNA. Nucleic Acids Res. 1995 Sep 11;23(17):3501-7.

[88] Yafe A, Etzioni S, Weisman-Shomer P, Fry M. Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. Nucleic Acids Res. 2005;33(9):2887-900.

APPENDICE

Appendix

1. CaPi transfection protocol use for SK-N-BE(2) cells.

This method applied from Nature methods April 2005.

Materials

2.5 M $CaCl_2$

Combine : 18.375 g $CaCl_2$ (MW 147 g/mol)

Water to 50 ml and sterile filter

1M NaCl (MW 58.4)

2.92 g in 50 ml water

100 mM Na_2HPO_4 (MW 141.96)

0.71 g in 50 ml water

500 mM HEPES (MW 238.3)

5.96 g in 50 ml water

2x HBS	vol (ml)	final conc.
1M NaCl	14	280 mM
100 mM Na_2HPO_4	0.75	1.5 mM
500 mM HEPES	5	50 mM
Water	30.25	

pH to 7.10 with NaOH

Sterile filter and store at 4 ^{0}C

0.1 xTE and sterile filter

15% glycerol in 1x HBS

7.5 ml glycerol in 42.5 ml 1x HBS Sterile filter

This scale is enough for 5 wells of 12-well plate.

1. Plate cells at 1-4 x 10^5 cells/well (12-well plate) overnight.

2. Change media 1 hour before transfection.

3. Combine 6.25 ug DNA with 25 ul of 2.5 M $CaCl_2$ and adjust final volume to 250 ul by 0.1 x TE.

4. Mix 1 vol of CaPi-DNA (3) with 1 vol of 2x HBS and stand at RT 1 minute (total volume 500 ul).

5. Drop (4) 100 ul to each well and put cell back in incubator for 6 hours.

6. Remove media and wash with PBS 1 time.

7. Add 1.5 ml of 15% glycerol in each well and incubate for 40 minutes in incubator.

8. Wash with PBS and change new media.

2. Reagents for NFAT inhibition

Cyclosporin A (sigma C3662-5mg)

MW = 1202.61

Stock 500 uM or 0.5 mM (1000x)

CsA 1.2 mg (0.0012 g) add ethanol to 2 ml

Filter and keep in 4°C

Working 500 nM

Mix 1 ul of 500 uM (stock) in 1 ml of media (for 1 well of 12-well plate)

PMA (Phorbol 12-myristate 13-acetate) sigma P8139-5 mg

MW = 616.83

Stock 25 ug/ml (1000x)

PMA 0.25 mg (0.00025 g) add ethanol 10 ml

Filter and keep in -20°C

Working 25 ng/ml

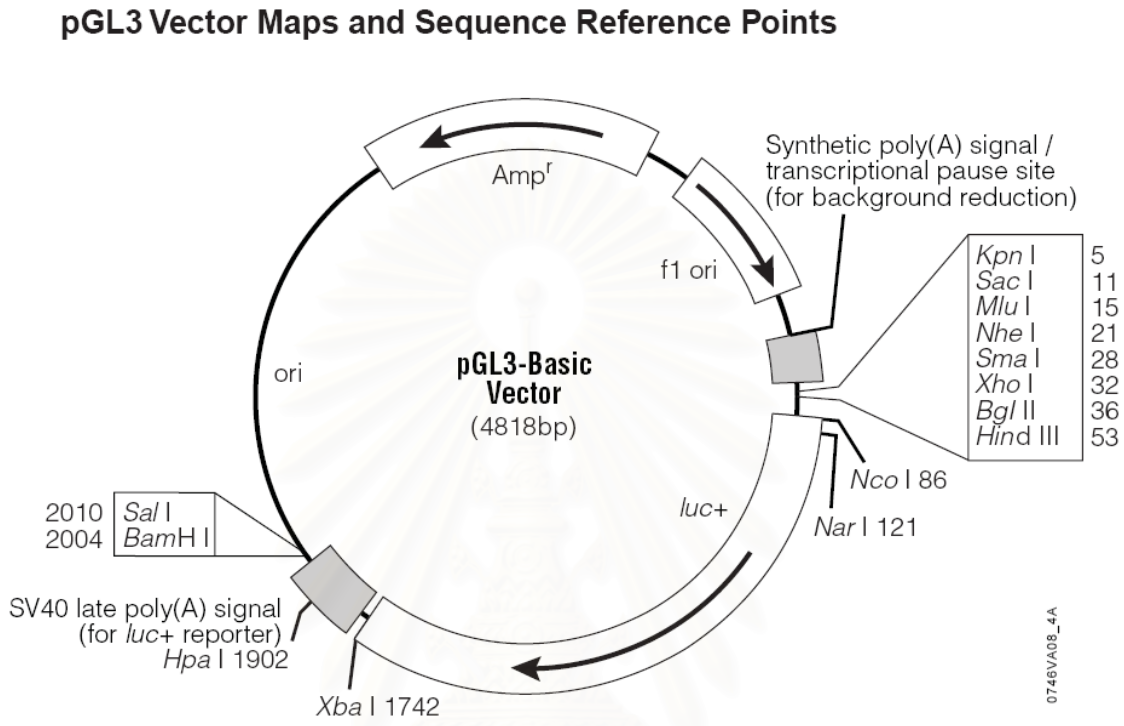
Mix 1 ul of 25 ug/ml (stock) in 1 ml of media (for 1 well of 12-well plate)

Vehicle

Mix 1 ul of ethanol in 1 ml of media (for 1 well of 12-well plate)

3. pGL3 Basic Vector Maps and Sequence Reference Points

(copy from technical manual No.033, Promega)



pGL3-Basic Vector circle map. Additional description: luc+, cDNA encoding the modified firefly luciferase; Ampr, gene conferring ampicillin resistance in E. coli; f1 ori, origin of replication derived from filamentous phage; ori, origin of replication in E. coli. Arrows within luc+ and the Ampr gene indicate the direction of transcription; the arrow in the f1 ori indicates the direction of ssDNA strand synthesis.

pGL3-Basic Vector Sequence Reference Points:

SV40 Promoter	(none)
SV40 Enhancer	(none)
Multiple cloning region	1–58
Luciferase gene (luc+)	88–1740
GLprimer2 binding site	89–111

SV40 late poly(A) signal	1772–1993
RVprimer4 binding site	2080–2061
ColE1-derived plasmid replication origin	2318
Beta-lactamase gene (Ampr)	3080–3940
f1 origin	4072–4527
Synthetic poly(A) signal	4658–4811
RVprimer3 binding site	4760–4779

BIBLIOGRAPHY

My life was mostly related to genetic research around 12 years. In 1995, I obtained a unique opportunity in getting the Development and Promotion of Science and Mathematics Scholarship, in Genetics program, from Chulalongkorn University. Four years later, I graduated from Chulalongkorn University, Bangkok, Thailand with a Bachelor of Science Degree majoring in Genetics, first class honor. Later, from 1999 to 2001, I enrolled and consequently graduated the Master degree in Medical Science Program from the Faculty of Medicine, Chulalongkorn University. After graduation, I have worked as a lecturer in the Department of Biology, Faculty of Science, Mahasarakham University under the condition of the University Development Commission Fund. I occupied the time by teaching undergraduate courses such as General Genetics, Human Genetics, Biochemical Genetics and Biochemistry. Recently, I am a Ph.D. candidate in Biomedical Sciences Program at Chulalongkorn University. I spent 3 years as a postgraduate fellow at Yale University and developed my thesis project there. I attended the Yale-Chula Drug Dependence Genetic Training Program by the suggestion from my mentor (Prof. Apiwat Mutirangura). Attending to the program will provide me an opportunity to gain valuable research experience and as a result deepening my researcher career. I hope I will be graduated soon and I strongly intend to devote all of my ability to be useful for other people.