รูปแบบลิงก์เกจดีสอีควิลิเบรียมระหว่างประชากรของยีน *GABRA2* และ *GABRG1* ที่ตำแหน่งกลุ่มยีน GABA บนโครโมโซม 4 ของมนุษย์

นายชูพงศ์ อิทธิวุฒิ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาชีวเวชศาสตร์ (สหสาขาวิชา)

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

INTER-POPULATION LINKAGE DISEQUILIBRIUM PATTERNS OF *GABRA2* AND *GABRG1* GENES AT THE GABA LOCUS ON HUMAN CHROMOSOME 4

Mr. Chupong Ittiwut

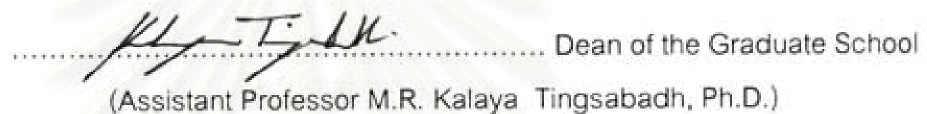
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

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

Copyright of Chulalongkorn University

Thesis Title	INTERPOPULATION LINKAGE DISEQUILIBRIUM PATTERNS OF *GABRA2* AND *GABRG1* GENES AT THE GABA CLUSTER LOCUS ON HUMAN CHROMOSOME 4
By	Mr. Chupong Ittiwut
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.)

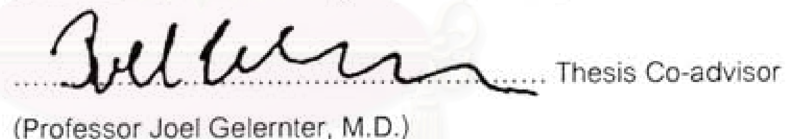
THESIS COMMITTEE

Chairman

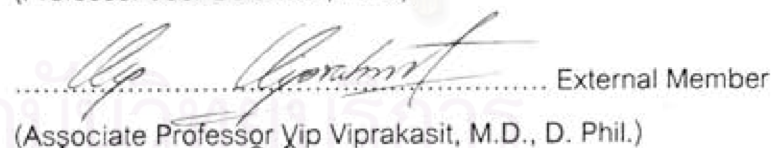
(Assistant Professor Tewin Tencomnao, Ph.D.)

Thesis Advisor

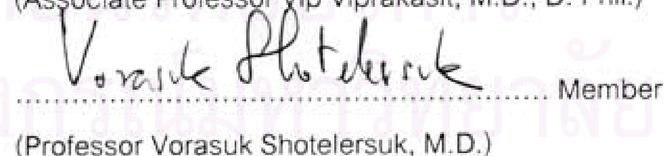
(Professor Apiwat Mutirangura, M.D., Ph.D.)


Thesis Co-advisor

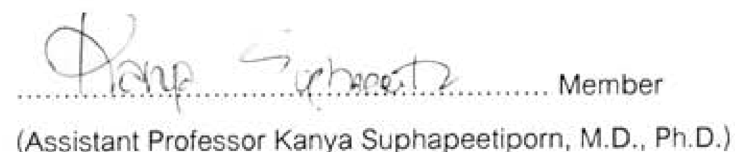
(Professor Joel Gelernter, M.D.)


External Member

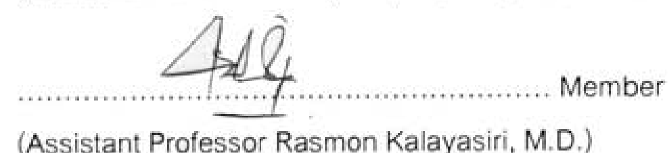
(Associate Professor Vip Viprakasit, M.D., D. Phil.)


Member

(Professor Vorasuk Shotelersuk, M.D.)


Member

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


Member

(Assistant Professor Rasmon Kalayasiri, M.D.)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชูพงศ์ อิทธิวุฒิ : รูปแบบลิงก์เกจดีสอีควิลิเบรียมระหว่างประชากรของยีน *GABRA2* และ *GABRG1* ที่ตำแหน่งกลุ่มยีน GABA บนโครโมโซม 4 ของมนุษย์ (INTER-POPULATION LINKAGE DISEQUILIBRIUM PATTERNS OF *GABRA2* AND *GABRG1* GENES AT THE GABA LOCUS ON HUMAN CHROMOSOME 4) **อ. ที่ปรึกษา: ศ.ดร.อภิวัฒน์ มุทิรางกูร,** 59 **หน้า.**

ยีน *GABRA2* และ *GABRG1* ซึ่งสร้างโปรตีนแอลฟา 2 และแกมม่า 1 ซึ่งเป็นส่วนประกอบของรีเซพเตอร์ $GABA_A$ ยีนทั้งสองนี้อยู่ที่ตำแหน่งแขนสั้นของโครโมโซมที่ 4 ในมนุษย์ (4p) พบความสัมพันธ์ของยีน *GABRA2* กับการติดสุราเรื้อรัง อย่างไรก็ตามยังไม่มีการศึกษาในแง่ของการทำงานของยีนที่สัมพันธ์กับการติดสุราอย่างแน่ชัด เพื่อที่จะเข้าใจถึงความสัมพันธ์ดังกล่าวมากยิ่งขึ้น การศึกษาในครั้งนี้ได้ทำการศึกษารูปแบบของลิงเกจดีสอีควิลิเบรียมและแฮพโพลไทป์ของยีนทั้งสอง ซึ่งอยู่ห่างกันประมาณ 90 กิโลเบส ซึ่งเป็นไปได้ว่ามาร์กเกอร์ที่อยู่บนยีนทั้งสองนี้อาจมีความสัมพันธ์กันในเชิงของการเกิดลิงเกจดีสอีควิลิเบรียม จึงได้ทำการศึกษารูปแบบของแฮพโพลไทป์ของสนิ๊พ (SNP-single nucleotide polymorphism) จำนวน 13 ตำแหน่ง ใน 5 กลุ่มประชากรที่แตกต่างกัน ได้แก่ ประชากรอเมริกันเชื้อสายยุโรป ประชากรอเมริกันเชื้อสายแอฟริกัน ประชากรเชื้อสายจีน (ในที่นี้แบ่งเป็นสองกลุ่มได้แก่ชาวฮั่นที่อยู่ที่แคลิฟอร์เนีย ที่อพยพมาจากไต้หวัน ฮ่องกง และกลุ่มของชาวไทยเชื้อสายจีน) ชาวไทย และชาวไทยเชื้อสายม้ง จากการศึกษาพบว่าในชาวม้งบริเวณของลิงเกจดีสอีควิลิเบรียมระหว่างสองยีนนี้ขยายออกไปถึง 280 กิโลเบส ในขณะที่ในประชากรกลุ่มอื่นๆ ที่บริเวณเดียวกันนี้จะพบบล็อกของบริเวณที่เป็นลิงเกจดีสอีควิลิเบรียมจำนวนอย่างน้อย 2 บล็อก ความแตกต่างที่พบนี้อาจมีนัยสำคัญบางประการที่เกี่ยวข้องกับลักษณะการติดสุราเรื้อรังที่แตกต่างกันไปในแต่ละประชากร ซึ่งเป็นไปได้ว่าไม่เพียงแต่ยีน *GABRA2* เท่านั้นที่มีความสัมพันธ์กับการติดสุราเรื้อรัง แต่อาจรวมถึงยีน *GABRG1* ที่อยู่ใกล้เคียงกันนั้นด้วย เราจึงได้ทำการศึกษาต่อไปในกลุ่มประชากรที่มีตัวอย่างของคนที่เป็นสุราเรื้อรังและกลุ่มควบคุมที่เหมาะสม โดยทำการจีโนไทป์สนิ๊พ 6 สนิ๊พที่เป็นตัวแทนสนิ๊พทั้ง 13 สนิ๊พข้างต้น ซึ่งอยู่ในตำแหน่งที่ครอบคลุมทั้งสองยีน โดยทำการศึกษาในกลุ่มประชากรอเมริกันเชื้อสายแอฟริกัน ที่ติดสุราเรื้อรังจำนวน 276 คนทำการศึกษาเทียบกับกลุ่มควบคุมจำนวน 242 คน จากผลการศึกษาพบว่า พบความสัมพันธ์ของสนิ๊พเดี่ยวบางตัวกับการติดสุราเรื้อรังได้แก่ rs10938426 ซึ่งอยู่ที่อินทรอน 1 ของยีน *GABRG1* โดยมีค่านัยสำคัญทางสถิติเมื่อพิจารณาจีโนไทป์ของกลุ่มผู้ป่วยเทียบกับกลุ่มควบคุมที่ค่า p=0.044 และยังพบความสัมพันธ์อย่างมีนัยสำคัญที่ p= 0.008 และ 0.007 เมื่อพิจารณาจีโนไทป์และอัลลีล ที่สนิ๊พ rs279869 ซึ่งอยู่บนอินทรอน 6 ของยีน *GABRA2* เมื่อทำการศึกษาความสัมพันธ์เชิงแฮพโพลไทป์ของสนิ๊พทั้ง 6 ตำแหน่งด้วยกันพบว่ามีนัยสำคัญทางสถิติที่ค่า p= 0.0027 (ที่ปรับค่านัยสำคัญโดยวิธี Bonferroni) และค่านัยสำคัญทางสถิติเพิ่มขึ้นอย่างมากเมื่อพิจารณาเฉพาะที่ตำแหน่งสนิ๊พทั้งสองตัวแหน่งดังกล่าวในกลุ่มผู้ป่วยเทียบกับกลุ่มควบคุม โดยมีค่า p= 0.00013 การศึกษาครั้งนี้บอกให้เราทราบว่า มีความสัมพันธ์ของสนิ๊พของยีน *GABRA2* และ *GABRG1* โดยเฉพาะในแง่ความสัมพันธ์แบบแฮพโพลไทป์ของแต่ละสนิ๊พจากแต่ละยีน ซึ่งเป็นหลักฐานสำคัญที่อาจบ่งบอกความสัมพันธ์ในแง่ของการทำงานของโปรตีนทั้ง 2 ตัวที่สัมพันธ์กับการเสี่ยงต่อการติดสุราเรื้อรังในกลุ่มประชากรอเมริกันเชื้อสายแอฟริกัน

สาขาวิชา ชีวเวชศาสตร์
ปีการศึกษา 2550

ลายมือชื่อนิสิต...
ลายมือชื่ออาจารย์ที่ปรึกษา...
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม...

4789716220: MAJOR BIOMEDICAL SCIENCES

KEY WORD: LINKAGE DISEQUILIBRIUM; ALCOHOL DEPENDENCE; *GABRA2*; *GABRG1*; HAPLOTYPE

CHUPONG ITTIWUT: INTER-POPULATION LINKAGE DISEQUILIBRIUM (LD) PATTERNS OF *GABRA2* AND *GABRG1* GENES AT THE GABA LOCUS ON HUMAN CHROMOSOME 4. THESIS ADVISOR: PROF. APIWAT MUTIRANGURA, M.D., Ph.D., 59 pp.

GABRA2 and *GABRG1*, which encode the α2 and γ1 subunits, respectively, of the $GABA_A$ receptor, are located in a cluster on chromosome 4p. The *GABRA2* locus has been found to be associated with alcohol dependence (AD) in several studies, but no functional variant that can account for this association has been identified. In order to understand the reported associations, we sought to understand LD patterns and haplotype structure of these genes. With close intergenic distance, ~90 kb, it was anticipated some markers might show intergenic LD. Variation in 13-SNP haplotype block structure was observed in 5 different populations: European American, African American, Chinese [Han and Thai], Thai, and Hmong. In Hmong, a 280 kb region of considerably higher LD spans the intergenic region, whereas in other populations, there were two or more LD blocks cross this region. To understand better the findings of association of AD with markers located at the 3' region of *GABRA2* which may be attributable to the adjacent gene, *GABRG1*, we genotyped Six tag single nucleotide polymorphisms (SNPs) that span *GABRG1* and *GABRA2* in 276 African-Americans (AAs) with AD and in 242 AA controls. Individual single SNP associations were tested by chi-square. Nominally significant allele frequency differences were identified for rs10938426, at *GABRG1* intron 1, with p= 0.044. Significant differences in both genotype and allele frequency (p=0.008 and 0.007 respectively) were observed at rs279869, located at *GABRA2* intron 6. We performed haplotype association analysis by means of PHASE. Six-SNP haplotypes combining SNPs from both gene loci showed differences between controls and AD subjects, p=0.0027, significant after Bonferroni correction. Two-SNP haplotype composed of rs10938426 and rs279869, showed greater significance (p=0.00013). This finding suggests significance of the interrelationship between these two genes and the possibility of risk loci in each of them. A two-SNP haplotype composed of one SNP from each gene suggests possible interaction of these genes and supports the involvement of both in predisposition to AD in AAs.

Field of study: Biomedical Sciences

Academic year: 2007

Student's signature:

Advisor's signature:

Co-advisor's signature:

Acknowledgments

This work was supported in part by grants from the National Institutes of Health (D43 TW06166, R01 AA11330, R01 DA12690, K24 DA15105, K08 AA13732, P50 AA12870, K24 AA13736, R01 AA011321, P50 AA03510, K24 MH64122, P50 DA018197) and the U.S. Department of Veterans Affairs (VA Alcohol Research Center; VA Mental Illness Research, Education and Clinical Center).

I would like to thank many people who have been greatly contributing to the completion of this thesis. The great majority of contributions have been provided by Professor Apiwat Mutirangura and Professor Joel Gelenter who are the thesis advisor and co-advisor respectively. As the principal investigators of the Thai-US Drug Dependence Training Program, they gave me one of the best opportunities in my life to go for doing the 3-year-long research project at Yale University, the place at which I had been warmly welcomed by Professor Joel, all people in Gelernter Lab, and many friends there who share every moments of happiness together.

Many thanks are due highly to all my friends from Mutirangura Lab, Faculty of Medicine, Chulalongkorn University.

This research would not be completed without extraordinary helps from my dear parents, my brother, my grandparents, and all of my family members. They always play their important role in inspiring, sharing, and supporting me all the time even in the hardest time of my life. It is my great honor and pleasure to say many thanks and express my deepest gratitude to my beloved wife, Rungnapa (Moo) Hirunsatit, who stands by my sides everywhere and every time.

TABLE OF CONTENTS

Page

Abstract (Thai) IV
Abstract (English) V
Acknowledgement VI
Table of Contents VII
List of Figures IX
List of Tables X
List of Abbreviations XI
CHAPTER I: Introduction 1
Research Questions and Objective 4
Hypothesis, Conceptual Framework, Key Words and
Expected Benefits and Application 5
CHAPTER II: Reviews and Related Literatures
Gene finding for complex diseases 6
Role of haplotypes in candidate gene studies 7
D' as measure of LD 9
HapMap and the Han Chinese population 10
Hmong population 11
Hmong population represents inbred or "isolated" population 13
LD in isolated populations 14
Genetic of $GABA_A$ receptor and AD 15
$GABA_A$ receptor 15
GABA neurotransmitter 16
GABA synthesis and Kreb's (Tricarboxylic) cycle 16
TaqMan method 17
CHAPTER III: Materials and Methods
Populations
-Interpopulation LD block definition 20

-AAs for association study 20

Selection of SNP markers

-Thirteen SNPs for LD block definition 21

-Tagging SNPs 21

Genotyping 22

Statistical analysis

-LD block definition 23

Association between *GABRA2/GABRG1* variations in AA population 26

CHAPTER IV: Results

Allele frequency differences between populations 27

Analysis of LD and Tagging SNPs 27

Haplotype frequencies 30

GABRA2 and *GABRG1* Variation Associated with Alcohol Dependence in African American Population 33

CHAPTER V: Conclusion and Discussion 37

References 47

APPENDICE 56

BIBLIOGRAPHY 59

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF FIGURES

Figure Page

1. Location of SNP markers genotyped across *GABRA2* and *GABRG1* 3

2. The signal and noise effect 14

3. Composition of $GABA_A$ receptor 17

4. TaqMan (5' Nuclease Assay) genotyping method 18

5. Overview of the procedure for using TaqMan SNP Genotyping Assays 19

6. LD structure and location of 6 htSNPs 22

7. Comparisons of LD structure and htSNPs over the *GABRG1/GABRA2* 29

8. LD patterns computed based on genotype data from HapMap 43

LIST OF TABLES

Table | Page

1. Populations included in Phase I and II HapMap 11

2. Location and prevalence of allele frequencies of *GABRG1* and *GABRA2* 28

3. Distribution of common haplotype among population: 13 SNPs (A-M) 30

4. Haplotype frequencies reconstructed by comparing Block Ib (SNP C-D-E-F) in European Americans to other populations 32

5. Haplotype frequencies reconstructed by compared Block II (SNPs G-L) in European American to other populations 32

6. *GABRG1-GABRA2* SNP information in African American 34-35

7. *GABRG1-GABRA2* 6-SNP haplotype association (SNP A to L) 36

8. Two SNP haplotype association (SNP D and J) from table 2 result 36

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF ABBREVIATIONS

GABA	Gamma-aminobutyric acid
GABRA2	Alpha-2 subunit of $GABA_A$ receptor
GABRG1	Gamma-1 subunit of $GABA_A$ receptor
LD	Linkage disequilibrium
SNPs	single nucleotide polymorphisms
AD	Alcohol dependence
EA	European-American
AA	African-American
THT	Thai Thai
THC	Chinese Thai
HM	Hmong
USC	Chinese American
YRI	Yoruba in Ibadan, Nigeria
CHB	Han Chinese in Beijing, China
JPT	Japanese in Tokyo, Japan
CEU	Population from Center d'Etude du Polymophisme, Utah, USA
NT	Nucleotide
HWE	Hardy-Weinberg equilibrium

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER I

INTRODUCTION

γ-aminobutyric acid (GABA) is the major inhibitory neurotransmitter in human brain. It is widely involved in the modulation of neurotransmission, particularly via regulation of neuronal excitability (1, 2) . GABA exerts its effects through interaction with GABA receptors. Functionally, fast synaptic inhibition in the mammalian CNS is mediated largely by activation of the γ-amino butyric acid type A ($GABA_A$) receptors (3), a ligand-gated receptor that is expressed in many regions of mammalian brain. The working receptor is composed of five protein subunits that form a chloride channel that remains closed until it binds its ligand. Each subunit is a long chain polypeptide with four putative α-helical cylinder domains embedded within the cell membrane, with the N-terminal end being extracellular (4, 5). The 19 distinct $GABA_A$ receptor subunits have been classified into α, β, γ, δ, ε, π, and ρ types (6). Two α and two β subunits are typically part of the pentameric assemblies of $GABA_A$ receptors. The particular subunit composition varies widely among brain regions and species. GABA binds specifically to the recognition site in the $GABA_A$ receptor and forms a chloride ion-selective channel that mediates neuronal membrane potential (7-9) . Other pharmacologically important molecules, for example, barbiturates, benzodiazepines, ethanol, and the anesthetic steroids, also interact with these receptors, either directly or indirectly.

Alcohol dependence (MIM 103780), AD, is one of the complex psychiatric diseases in which genetic and environmental factors were found to play a crucial role. Its heritability was estimated in at least 50% for both men and women (10). Genetic susceptibility has been implicated to the involvement of the brain GABAergic system and AD due especially to the evidences that implicated adaptation of $GABA_A$ receptor to ethanol exposure (11-14). For this line of physiological evidences make genes coding for $GABA_A$ receptor compelling candidate genes for AD. According to 19 heterogeneous subunits (6 of α, 3 of β, 3 of γ, δ, ε, π and 3 of ρ class) of $GABA_A$ receptor already classified (15),

Among the 19 distinct subunits of mammalian $GABA_A$ receptors identified to date (6), the γ-aminobutyric acid α-2 receptor (encoded by *GABRA2*) [OMIM 137140] is the

subtype that has received the greatest attention in relation to alcohol dependence risk. Based on its physiological role and its direct interactions with ethanol (16), as well as evidence from genetic linkage studies, it was suggested that the gene encoding $GABA_A$ receptor α-2 subunit should be considered a candidate locus for influencing risk for alcohol dependence (10, 17, 18) . Results from a genomewide linkage scan by the Collaborative Study on the Genetics of Alcoholism (COGA) showed suggestive linkage to alcohol dependence in the region of the GABA receptor gene cluster on chromosome 4p13-12. This region contains the *GABRG1, GABRA2, GABRA4,* and *GABRB1* genes, encoding γ-1, α-2, α-4, and β1 subunits of the $GABA_A$ receptor, respectively. Association studies considering some of these loci indicated that a genetic predisposition to alcohol dependence is related to polymorphic variation at or near *GABRA2* (19-22). Edenberg et al. 2004 (19) reported significant association between SNPs in the *GABRA2* flanking genes and alcohol dependence (AD). They found that the region of strongest association with AD extended from intron 3 to 58 kb beyond the 3′ end of the gene, spanning 164 kb. This study included 31 SNPs within or closely flanking *GABRA2* that were significantly associated with alcohol dependence. They also found that all consecutive three-SNP haplotypes within *GABRA2* 1 of the 5 haplotypes at the 5′ end of the gene, and all 43 of the haplotypes starting within exon 3 and extending to the 3′ end of the gene -- were significantly associated with AD. The association was confirmed by Covault et al. (20), who compared allele frequencies of 10 *GABRA2* SNPs spanning the coding region in European American (EA) controls and subjects with AD. The strongest evidence of association was shown across a region encompassed by seven of the 10 SNPs, from rs279837 in intron 3 through rs567926 in the 3' downstream region. This location provided the best evidence for association in all studies published to date. Subsequent studies have shown association of these *GABRA2* SNPs and alcohol dependence in populations of European ancestry; including Russians, Germans, Finns, and American Indians (21-24). These replicated findings established the importance of *GABRA2* variation – or variation at loci mapping close enough to *GABRA2* to be in linkage disequilibrium (LD) with associated markers -- on the risk for AD.

GABRA2 spans approximately 140 kb. There are 10 transcribed exons present in human *GABRA2* mRNA (NM_000807), with four major isoforms, consisting of

combinations of two alternative 5' and 3' exons caused by alternative splicing and (potentially) alternative promoter use (25). LD data from previous studies have employed sets of markers that extended only through the 3' end of the *GABRA2* gene; these data showed strong pairwise LD between the last two markers at the 3'end of the haplotype block of each study, consistent with the possibility of an extended LD block in the 3' direction (20-23). Across the intergenic region in the 3' direction, the gene encoding the γ-1 subunit (*GABRG1*) of the $GABA_A$ receptor is located. This gene spans approximately 83kb and consists of 9 exons (26). A consistent pattern of LD was previously constructed in EAs; this study (24) also showed an extended (193-kb) LD region, with D' 0.6-1.0. This region of LD spans from rs279867 in *GABRA2* intron 6 across the intergenic region to rs1391168, which located in intron 1 of *GABRG1* (19).

We analyzed 13 SNPs, located either within these genes or in the intergenic region. We aimed to provide sufficient SNP density such that our SNP set would be informative for most other non-genotyped SNPs that map to the region (Figure 1). To evaluate whether the effect of differing genetic background among populations with different ethnicities is reflected in distinct patterns of LD blocks, six populations were included: European American, African American, and Han Chinese American samples from the US, and Hmong, Thai, and Chinese Thai from Thailand.

Figure 1 location of SNP markers genotyped across GABRA2 and GABRG1 in GABA cluster on chromosome 4p13-1. The order of the genes from telomere to centromere on chromosome 4p (distances not to scale)

In this study, we consider the LD relationships within and between *GABRA2* and *GABRG1*, to improve our understanding of the previously reported associations between *GABRA2* and AD. The extent of LD in the populations of African, European, and Chinese

descents, represent by African American, European American, and Chinese American (USC), presented in this study, are comparable to those observed in the YRI, CEU, and CHB populations, in the International HapMap projects. These data are complementary; some of our samples of unrelated subjects are larger than those included in HapMap, and although the samples are of comparable ethnicity they are not identical. Understanding LD patterns and the *GABRA2/GABRG1* haplotype structure by extending it to multiple populations also allows inferences about the evolutionary history of any variant that increases AD risk.

Though the *GABRA2* locus has been found to be associated with alcohol dependence (AD) (19) association of AD with markers located at the 3' region of *GABRA2* has been replicated in several studies, but one recent study (27) suggests the possibility that the signal may be attributable to the adjacent gene, *GABRG1*, located 90kb distant in the 3' direction. The aim of the study was to investigate these two possibly-associated genes in African Americans, a population where decreased linkage disequilibrium compared to European-Americans (28) may permit the source of an association signal (or signals) to be distinguished.

Research Questions

1. How different are the LD patterns of the *GABRA2/GABRG1* region between 6 different populations?
2. Is there any association between the *GABRA2* and *GABRG1* SNPs and alcohol dependence in a selected (with availability of phenotype data and proper number of cases and controls) African American population?

Objectives

1. To investigate intergenic LD patterns different between 6 different populations including European American, African American, Chinese (Han and Thai), Thai, and Hmong.
2. To determine association between *GABRA2* and *GABRG1* variations (SNPs), in terms of genotype, allele, and haplotype, and alcohol dependence in African American population.

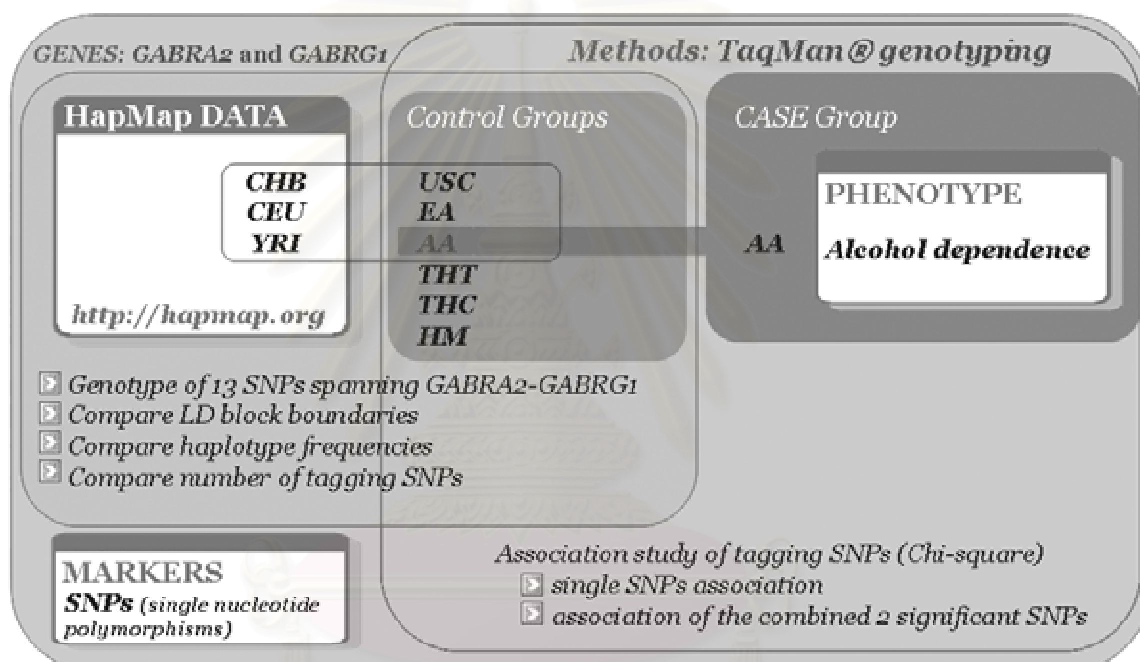
Hypothesis

1. Different population will manifest different LD patterns

2. Closely related populations, for example Chinese Thai and Han Chinese , show same LD pattern where as populations distant in geographic and genetic components such as Thais and African American show significant different LD patterns.

3. *GABRG1* SNPs which locate in the 3' direction of *GABRA2* associate with alcohol dependence in African American.

Conceptual Framework



Key Words

Linkage disequilibrium, LD block, Haplotype block, Association study, alcohol dependence, Haplotype-tag SNP, htSNPs

Expected Benefits and Application

Understanding LD patterns and the GABRA2/GABRG1 haplotype structure by extending it to multiple populations will facilitate inferring the evolutionary history of any variant that increases AD risk. In addition, with the pharmacogenetic approach patient with different *GABRA2/GABRG1*variants or patterns of variant (e.g. genotype, allele, and haplotype) may manifest different responses to drug or medical treatments.

CHAPTER II
LITERATURE REVIEWS

Gene finding for complex diseases

The research trend being of interested by many human geneticists all over the world is still "gene finding" as it used to be over the past few years. Moreover the complete sequencing of the human genome, as said, mostly finished in 2003 highly encourage the progress of this approach to identify disease-associated gene. Whereas gene finding for Mendelian disorders has dramatically been successful in number and function of gene able to be identified, there have been few of genes identified for complex disorders. The hassle, for example, for gene finding in complex diseases is that; in Mendelian disorder, mutations are generally deleterious and segregation of phenotype can be found in families. For complex diseases, besides its controversy for the clear cut effect of genetic component, very few evidences of genetic risk factors have been identified. The famous model of gene finding for complex diseases widely used initially for Alzheimer's disease is apolipoprotein E gene (29, 30) and the gene coding factor V Leiden clotting factor for thrombophilia (31, 32)

This very limited success of gene finding study in any complex diseases may be due to the difficulty to simply characterize how the genetics of complex disease should be compared to the well-defined genetic transmission pattern which was the important key role for the successfulness of gene maping for Mendelian diseases (33). Due to its complicating circumstances such as phenocopies, heterogeneity, clinical expression variability, age at onset, polygenic effect, and the ratio between genetic and environmental effects, the mapping gene approach used in Mendelian disorder is rarely success in complex disease. Focusing on the two main hypotheses of genetics of complex disease, the first one is about multi-equivalent risk model; many rare variants exist in population with varying effect on risks (each variant being a strong risk factor). The next hypothesis is the restricted polymorphism model which describes a relatively small pool of common disease allele exists. This latter hypothesis considers each factor a weak risk factor that seems to equally account for small effect. These risk factors can probably be both genetic and environmental, which are able to increase or decrease

disease risk (34). Definitely, complex disease is defined as disease where susceptibility is controlled multiple genetic and environmental risk factors and where each of these factors has only a modest effect on susceptibility (35). The typical complex diseases include, for instance, asthma, diabetes, and obesity. Also, it is widely used terms "common disease" for complex diseases because its high incidence found in a particular population as being common.

Understanding of LD pattern in human population has hopefully been to date the challenging way to facilitate the discovery of genes that influence human complex traits. The existence of the human genome project following by the HapMap project is interested people and especially many researchers worldwide to the way of characterization of human disease genes transmitted and being unique in any LD pattern in any population. The blossoms of recent improvements of SNP genotyping technology have enabled researchers to extend their study. Regarding association between SNP marker and the diseases, it can be inferred that significantly difference in SNP allele frequency between populations suggests difference in selection of the gene involved (36). Not only to mapping gene by a small scale of candidate gene approach including single SNP or few catalog of SNPs association, but researcher also extend their study to studying of various size of LD beyond the smaller region to broader views of genome easily. Considering LD levels, observed genetic-physical distance relationships can also reflect the LD unique outcome which leads to the more understanding of ancestral recombination, mutation, natural selection and population history (35).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Role of haplotypes in candidate gene studies

Cardon and Abecasis (2002)(35) defined the term "haplotype block" as "a discrete chromosome region of LD and low haplotype diversity". Expectedly any pair of adjacent SNPs exist within haplotype blocks is in strong LD, whereas other pairs will show much weaker association. Such blocks are also hypothesized to be regions of low recombination flanked by the two recombination hot spots. Allelic dependence will be improved in any high LD region. Even a tiny fraction of SNPs markers to be typed in this region can be facilitated to a chance of detecting association without genotyping a

number of markers located within this region. Controversially, within any low LD region, low correlation between markers will be found. The recruitment of more or all marker will be needed to be typed in order to characterize adequately all information given by this region.

It is already clear that understanding LD patterns in common haplotypes will be sufficient for study rare variants or any suspected mutations playing partial or great combination role to a particular complex traits. So there have been number of recent genetic studies focusing on mutations or polymorphisms within challenging low LD regions with the ultimate goal to characterize and map candidate genes for complex diseases. The previous articles by Gray et al (2000) (37) and Johnson et al. (2001) (38) have shed light on a lower density of markers within the non-recombination haplotype block being sufficient for disease-mapping purpose. The sufficient SNP markers selected were called haplotype-tag SNPs (htSNPs) (38). All information present in the LD block could be captured by the htSNPs, therefore raises possibility to reduce cost of genotyping and cut down labor to analyze data from a much higher marker. Moreover, by the time the search of genetic determinants of common disease has been extremely of interest, htSNP could reduce the problem of disease mapping to only two key steps; finding LD block and then, by cohort case-control study, finding any specific haplotype with strongest association to the disease. However, the caution of defining haplotype block remains carefully needed. Marker density is main factors that can affect haplotype block characterization. It should be kept in mind that probably, in some genomic regions or in some populations, the well-defined haplotype block and high-frequency SNP might not exist. This can probably be solved by other appropriate statistics employed to the study. Nevertheless, effect of inappropriate statistic method on unwell-defined LD block remains unclear. Maybe, the reason why a large haplotype block cannot be defined might be due to the existence of recombination hotspots within such region which reduce the larger haplotype to the smaller ones. Any sub-block existing may lower haplotype diversity than that of the larger one (39).

D’ as measure of LD

Theoretically, two loci is belonged to genetic linkage equilibrium when the population is influenced by the condition in which random mating, no genetic mutation, no either selective advantage or disadvantage, and no migration exists. Under this condition, the genetic disequibrum in which at least two linked loci will not be observed. However LD between loci does exist since the model of genetic equilibrium is not fit to all human populations and all assumptions to maintain equilibrium in human gene pool never come true. Linkage disequilibrium (LD) between loci are observed and measured by a population-based D’ parameter. Numerous of LD region found throughout the human genome can be influenced by the age of crossover generations by generations, gene conversion, selective advantages or disadvantages of single or linked alleles, and demographic history, such as founder effect and admixture (40-42).

Strong LD should be virtually found, if there is no recombination hotspot exists within any chromosome region or between the two adjacent markers. In contrast, in the region with recombination events has occurred, no or little LD is measured. Among various statistic method used for LD measures, a simple measure of LD is $D_{ij} = x_{ij}-p_iq_i$ (43), where xij is the observed of frequency of haplotype AiBj, pi and qj are the frequencies of allele Ai and Bj at loci A and B respectively. 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’ (44) is widely preferred due to it is statistically corrected, as its range is the same for allele frequencies D’ is defined as D'_{ij}; $D'_{ij} = D_{ij}/$ Dmax, where $Dmax = min[p_iq_j, (1- p_i)(1- q_j)]$ when $D_{ij} <0$, or $D\ max = min[p_i\ (1- q_j), (1- p_i)\ q_j]$ when $D_{ij} >0$.

Where as another measure of Linkage Disequilibrium is designated as r^2 (45); which $r^2= D/(p_{Ai}p_{Aj}p_{Bi}p_{Bj})$. This parameter has more reliable sample properties than |D'| when low allele frequencies are observed. Though it is controversy between using |D'| and r^2 for LD measure, online article from http://genestat.org (46) evaluated the use of these two standard LD measures as followed;

1. D and |D'| are best if plotting association between disease and markers for the purpose of "Simple fine mapping" (47).

2. r^2 is best for association mapping in sense of most powerful or useful SNPs and/or haplotype because there is a simple linear relationship with sample size (48)
3. r^2 depends on marker allele frequencies and can be difficult to interpret when comparing multiple markers in a region (43)
4. |D'| is directly related to recombination fraction and its generalization to more than two loci is the only measure of LD not sensitive to allele frequencies
5. |D'|=1 if 2 or 3 haplotypes are present and r^2=1 if 2 haplotypes present (for pairwise LD)
6. Intermediate values of |D'| hard to interpret

(**Note**: This conclusion has been adapted and copied from reference (46))

HapMap and the Han Chinese population

The international HapMap project aims to determine the common patterns of DNA sequence variants, their frequencies, and correlations between them in the four large reference populations at a density of 1 SNP every 5kb. This Big Mega Project is also aimed at selection of tagging SNPs (tSNPs) or haplotype tagging SNPs (htSNPs) that capture information of other variations in the genome to minimize the number of SNPs need to genotyped for genome-wide and candidate-gene association in a certain population. Han Chinese in Beijing (CHB), China, were recruited to the HapMap project as the other three populations including a population from Centre d'Etude du Polymorphisme Humain reference individuals from Utah, USA (CEU), Japanese in Tokyo, Japan (JPT), and Yoruba in Ibadan, Nigeria (YRI) (Table 1) (49-52).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 1 Populations Included in Phase I and II HapMap (http://ccr.coriell.org)

Population	Microtiter Plates	Individual DNA Samples	Individual Cell Cultures
Yoruba in Ibadan, Nigeria **[YRI]**	HAPMAPPT03 HAPMAPPT04	180	180
Han Chinese in Beijing, China **[CHB]**	HAPMAPPT02* HAPMAPPT05**	90	90
Japanese in Tokyo, Japan **[JPT]**	HAPMAPPT02* HAPMAPPT05**	91	91
CEPH Collection **[CEU]** [NIGMS Human Genetic Cell Repository]	HAPMAPPT01 HAPMAPPT06	90	90

*The HAPMAPPT02 plate has 45 individuals from the Japanese in Tokyo and 45 individuals from the Han Chinese in Beijing.
**The HAPMAPPT05 plate has 46 individuals from the Japanese in Tokyo and 45 individuals from the Han Chinese in Beijing.

In this study, Chinese samples were ordered from another catalog of Han Chinese from the Coriell Institute for Medical Research different from the HapMap CHB population recruited from Beijing. Our samples are provided in the "Human population collection" of Han People of Los Angeles (California, USA) Panel of 100 (50 males and 50 females) named HD100CHI. Samples are detailed in the appendix back of this dissertation. All information is available at the Coriell Institute official website (53). The ethnicity of our "Han Chinese" was self-identified from individuals with all four grandparents born in Taiwan, China, or Hong Kong. These areas are occupied by the Southern Han Chinese geographically separated from the Northern Han Chinese by their dwellings in the lower or upper Yangtze River respectively (54-57).

Hmong Population

Hmong population is also named ***Miao*** in Chinese, referred to many of Non-Han population, and sometimes referred to ***Meo***. This name is opposed by Hmongs in Lao area as it sounds like a similar-sounding word means "***cat***" in the Lao language. Similarly, in Thailand, Hmong Thais find offensive as they are being looked down when

they are called "***Maew*** or ***แม้ว***". There are two groups of Hmong in Southeast Asia distinguished by differences in language and custom: Hmong ***Der*** (White Hmong) and Mong ***Leng*** (BlueHmong) (58). The word "**Hm**ong" comes from to Hmong Der, where as Mong Leng spells **M**ong without the beginning of H. Demographically pushed and invaded by stronger majority of Han Chinese from Southwest China during the late of 19th and early 20th centuries, the early major wave of Hmong population had been migrated through the North of Vietnam, Laos, Myanmar, and Thailand, the new world believed as rich, ample, and sterile field. The early Hmong settlers in these new lands have been pushed to the higher mountainous part by the early stronger settlers who already occupied the more ample lowland (58, 59). During the 1950s-1980s, Hmong in Indochina were supported by the United States CIA for the secret war to fight against the communist Laos and Vietnam. The Hmong army under the famous Hmong General Vang Pao, the former Royal Lao Army, played an important role in supporting the Americans in Vietnam War. When the United States troops were pulled out, majority of the Hmong population survived from the War had been fled to Thailand. Some of Hmong refugees stayed at the refugee camps in Thailand of which the major camp is at Wat Tham Krabok in Saraburi province, waiting for transfer to the third countries. The first 3,466 Hmong refugees arrived in America in December 1975. Eventually, the other were transferred and later resettled in France, the United States, Australia, and Canada (60).

With approximately 120,000 Hmong individuals populated in Thailand, the population structure is influenced by the unique history of intermarriage and harmonious relations. Regarding harmonious relation which refers to a state of proper equilibrium, Hmongs pay more attention to deal and greet with people they meet or coming to their area in order to exchange names, ages, clans, and generations to establish their relationship to address if they and the visitors belong to the same clan (relatives descent from a common ancestor). So they will be able to establish precise relationship within the clan. If not, they will establish their relationship through the marriage of their kin, beginning with their wives and aunts, in order to express their obligate to help each other. There are 19 Hmong clans in Laos. Clan membership is obtained by birth, marriage for women and adoption (58).

Hmong population represents inbred or “isolated” population

Since the Hmong populations are believed migrated from Southern China with inbred population originating from a limited number of founder, the population structure is seemingly unique according to their small founders, a new age of populations, and high rate of intermarriage (58). In our current study Hmong population is considered “***isolated population***” according to the criteria of the population which is founded by “*a limited number of individuals,* **and** *geographical and or cultural barriers have restricted migration and breeding with other groups”*, as defined by Charkraborty and Deka 2002 (61). Compared to any outbred populations in which the genetics and environmental heterogeneity is high (34), there appear to be two advantages of using isolated population for genetic mapping (62). First, low heterogeneity is exhibited in genetically isolated population as demonstrated by a small numbers of polymorphis SNPs and mutations in specific disease related genes. Second, isolated population is homogeneous that can reduce artefactual association of population stratification in which two or more are admixed (63). Reducing in genetic heterogeneity, as found in isolated population, can increase genotypic relative risk (GRR) because the homogeneous population with decreasing of genetic background noise (64) (Figure 2).

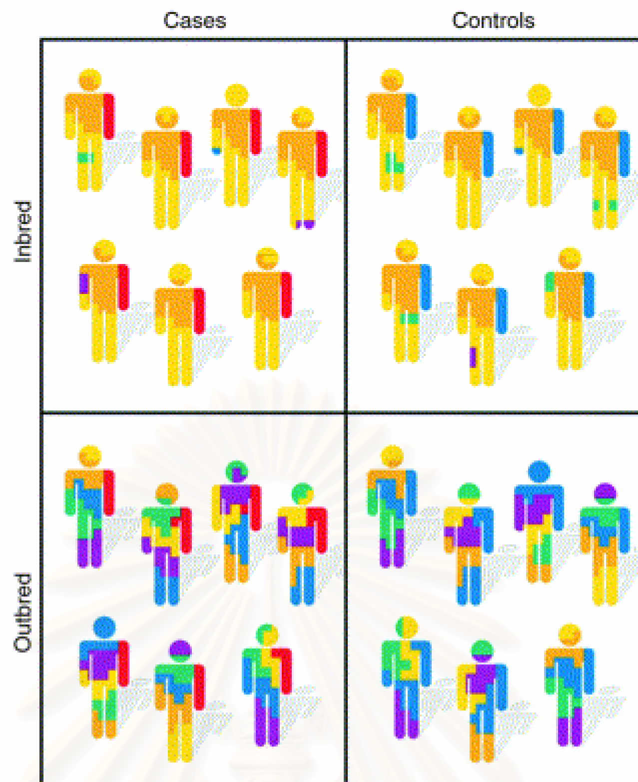


Figure 2 **The signal and noise effect.** Zak NB *et al.* 2001 (62) described how the genetic background noise effect association mapping by comparing inbred-isolated population to the outbred population. In this picture the background noise as obviously seen in outbred population, represented by multicolor in body parts of individuals, are not presented in inbred population. Low heterogeneity in isolated population is shown by fewer color in both case and controls in this population. This makes the disease predisposing signal, represented by red arms in cases of both groups, obviously detected in inbred population. Disease bearing individuals in outbred group, though are shaded by red in their arms, the other colorful noise make the red signal less obvious.

LD in isolated populations

Isolated populations are informative in studies illustrating the effects of population demographics on genetic and genomic structure and can greatly facilitate genetic mapping. To map genes for complex disease in very young isolated population, with theoretically long expected LD, easily enable geneticists to find a locus with a limited number of markers (34).

Genetic of $GABA_A$ receptor and AD

Though AD is influenced by the interplay between genetics and environmental factors, it is very interesting to dissect and find genes contributable to alcohol addict. It is also challenging to know if an individual inherit genetic component predisposing to alcoholism. It is better for providing proactive suggestion for avoiding alcohol use in individual susceptible to AD. Plausibly, the genetic factors, narrowing to specific gene variations, will predispose person with susceptibility to become an alcoholic. However, people who has genetic predisposition will never be dependent if they never expose or drink alcohol.

There were so many lines of evidence reported on the involvement of $GABA_A$ receptors and behavioral effects of alcohol in animal model (65, 66). Finding that GABA is involved in alcohol abuse and dependence supports a current theory that "predisposition to alcoholism might be inherited as part of a general state of brain overactivation" (19). People with susceptibility to AD may inherit a variety of genes that contribute to this state, e.g. normalizing the excitory state by alcohol drinking maybe lead people with a hyperexcited nervous system to use alcohol more frequently in order to normalize brain circuits. That, in turn, would put them at greater risk for developing AD.

$GABA_A$ receptor

Based on its physiological role and its direct interactions with ethanol (16), as well as evidence from genetic linkage studies, it was suggested that the gene encoding the $GABA_A$ receptor α-2 subunit should be considered a candidate locus for influencing risk for AD (10, 17, 18). Results from a genome-wide linkage scan by the Collaborative Study on the Genetics of Alcoholism showed suggestive linkage to AD in the region of the GABA receptor gene cluster on chromosome 4p13–p12. This region contains the *GABRG1, GABRA2, GABRA4,* and *GABRB1* genes, encoding γ-1, α-2, α-4, and β-1 subunits of the $GABA_A$ receptor, respectively. Association studies considering some of these loci indicated that a genetic predisposition to AD is related to polymorphic variation at or near *GABRA2* (19-22, 27). Edenberg et al. (19) reported a significant association between SNPs in the GABRA2 flanking genes and AD (AD). They found that

the region of strongest association with AD extended from intron 3 to 58 kb beyond the 3′ end of the gene, spanning 164 kb. This study included 31 SNPs within or closely flanking *GABRA2* that were significantly associated with AD. They also found that all consecutive 3-SNP haplotypes within *GABRA2—that 1* of the 5 haplotypes at the 5′ end of the gene and all 43 of the haplotypes starting within exon 3 and extending to the 3′ end of the gene—were significantly associated with AD.

GABA neurotransmitter

γ-Aminobutyric acid (GABA) is the major inhibitory neurotransmitter in the human brain highly concentrated in the substantia nigra and globus pallidus nuclei of the basal ganglia, followed by the hypothalamus, the periaqueductal grey matter and the hippocampus. It is the second most common brain neurotransmitter to glutamate which is the most common (excitory) neurotransmitter. GABA is presented in 30-40 percent of all synapses, the regions or gaps between presynaptic and postsynaptic neuron. It is widely involved in the modulation of neurotransmission, particularly via regulation of neuronal excitability (1, 2). GABA exerts its effects through interaction with GABA receptors. Functionally, fast synaptic inhibition in the mammalian CNS is mediated largely by activation of the γ-aminobutyric acid type A ($GABA_A$) (Figure 3) (3, 15).

GABA synthesis and Kreb’s (Tricarboxylic) cycle

The synthesis of GABA is linked to Kreb’s cycle, which breaks down fat, carbohydrate, protein, and Acetyl co-enzyme A and release oxygen and high energy phosphate compound, in animal mitochondria. Alpha-ketoglutarate, an intermediate compound in Kreb’s cycle, is catalyzed and trasaminated to glutamic acid by GABA transaminase. GABA will be synthesized in the next step using glutamic acid as substrate and glutamic decarboxylase (GAD) as catalyst. GABA can then be catabolized to succinic semialdehyde back into Kreb’scyle by catalysis of GABA transaminase enzyme. This reaction is known as GABA shunt.

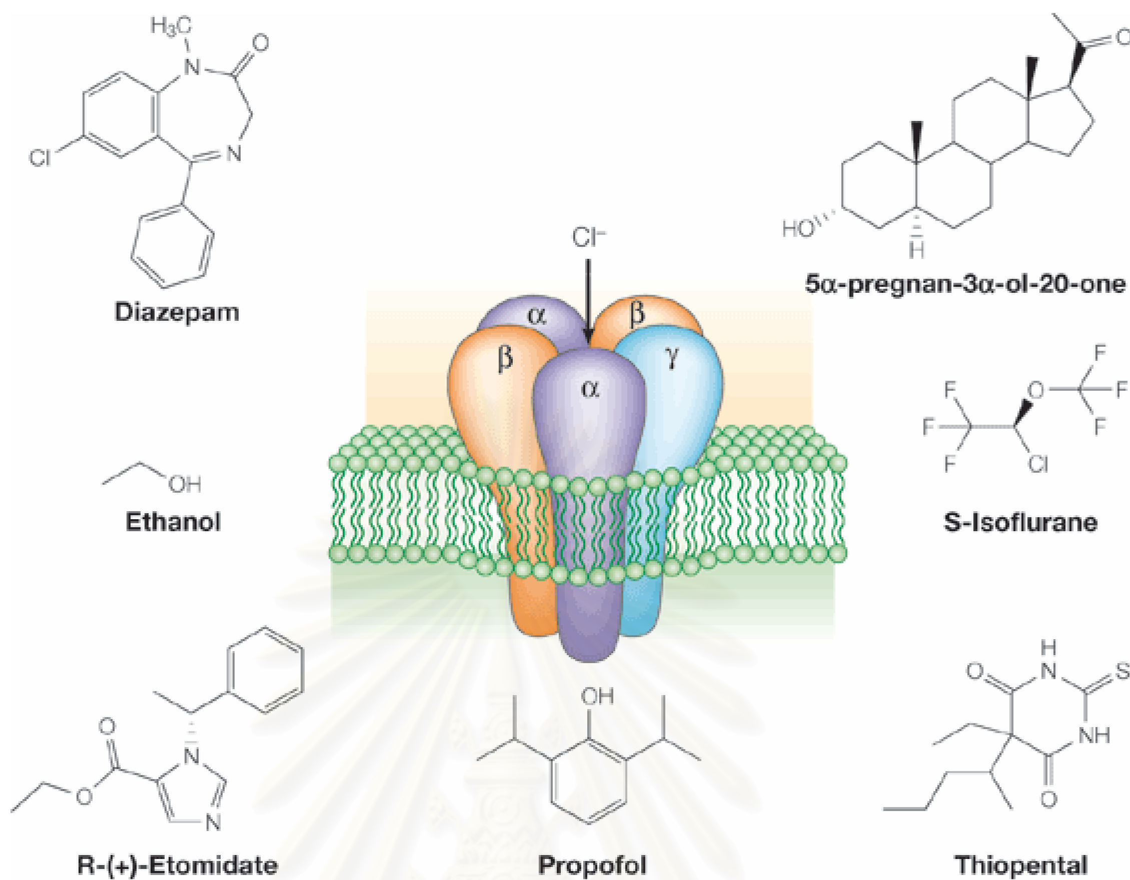


Figure 3 Composition of the $GABA_A$ receptor and its typical allosteric modulators (15)

Taqman Method

The SNP Taqman® genotyping method (67) (Figure 4) is a name given to the method applicable to high-throughput genotyping for any biallelic SNP marker. Each SNP genotyping requires 2 assays that consist of 2 primers for amplifying the sequence of interest and two Taqman MBG probes specific to each allele mixed into a single tube. The assay is commercially available for ordering for both "*on demand*" (ready to order) and "*by design*" (made when order) basis.

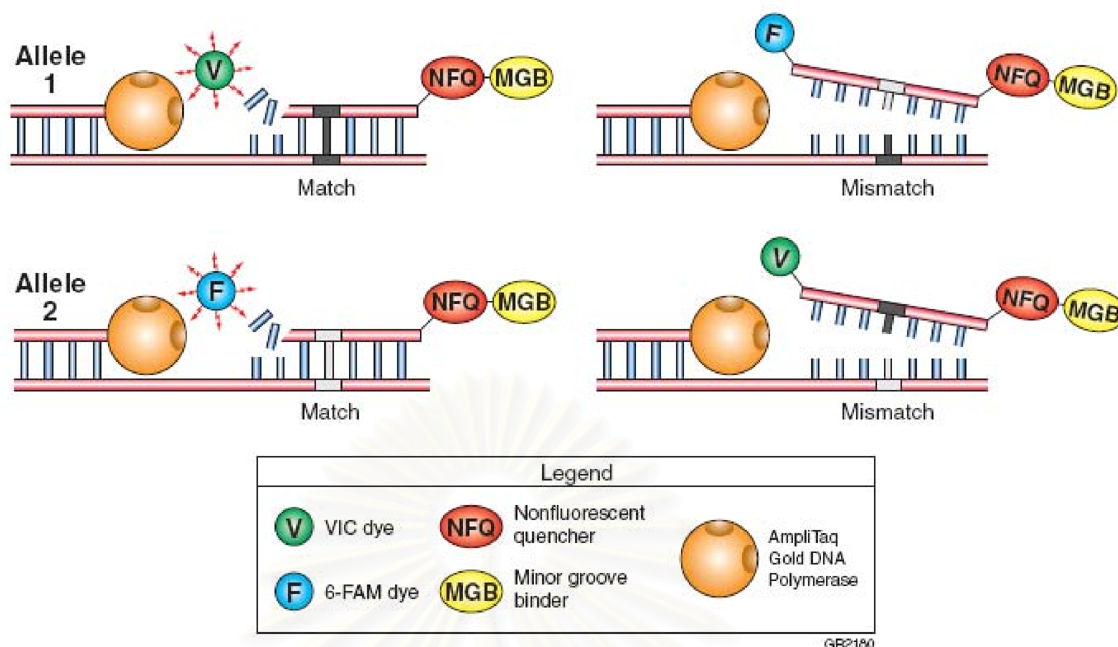


Figure 4 Taqman ® (5´ Nuclease Assay) genotyping method (67).

Two reporter dyes are synthesized and attached to each end of Taqman® MGB (Minor Groove Binding) probe. One allele is named VIC that referred to VIC dye attached to 5' end of the probe. The other allele can be detected by the FAM probe which has FAM dye on its 5' end. These probes are modified to increase the melting temperature (Tm) without increasing probe length. The greater differences in Tm values between matched and mismatched probes produce more accurate allelic discrimination. Non fluorescent quencher (NFQ) are modified at 3' end of the probes which allows the detection system to measure reporter dye contributions more accurately. The applied Biosystem platform for Taqman® genotyping is shown in Figure 5.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

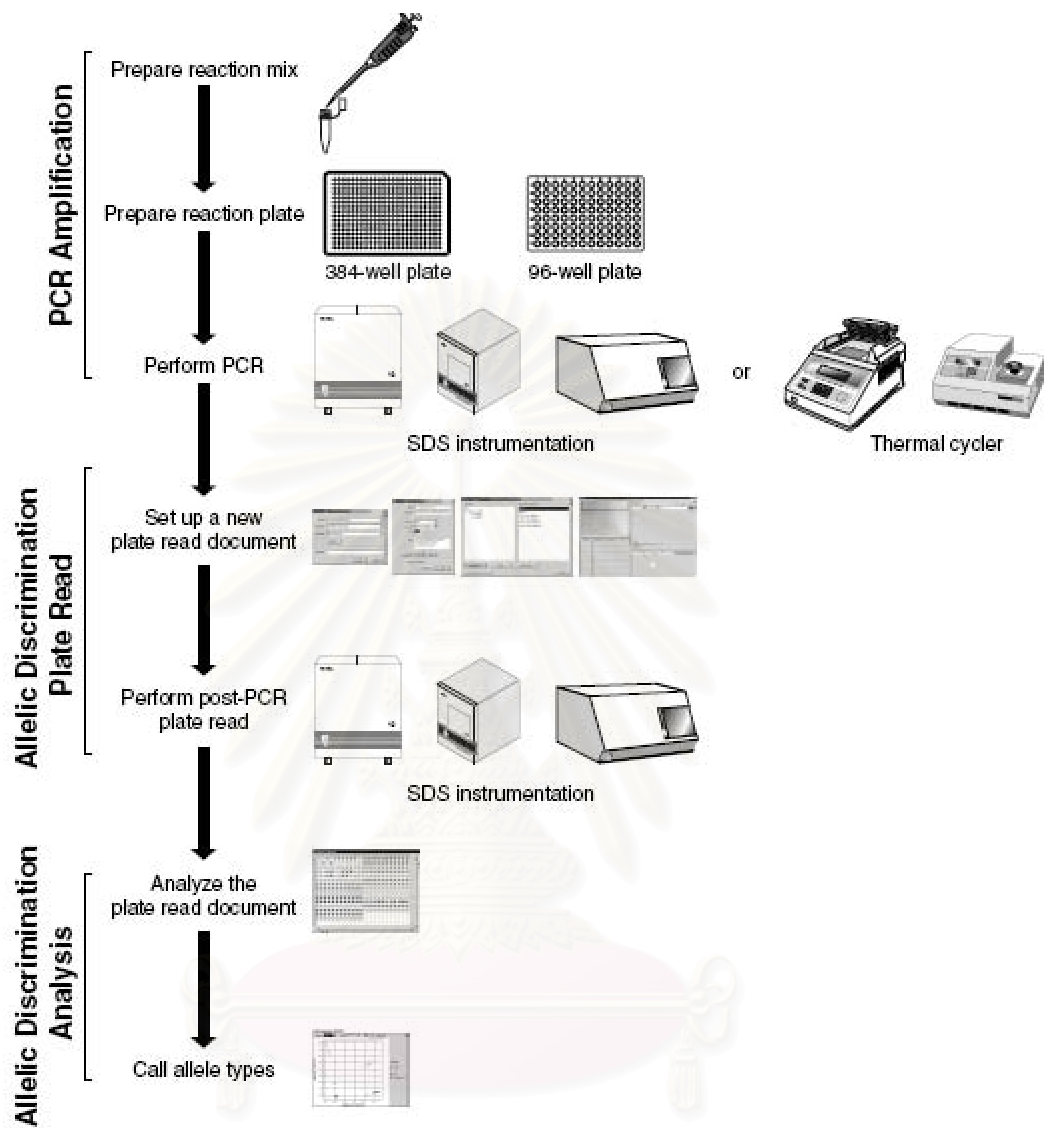


Figure 5 Overview of the procedure for using TaqMan SNP Genotyping Assays (67).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER III
MATERIALS AND METHODS

Populations

Interpopulation LD block definition

A total of 450 unrelated subjects were included in this study, from Thailand and The United States. The study sample of Thai populations was recruited as adult blood donors aged 18 years and older from the Thai Red Cross. Only individuals who reported that all four grandparents were of either Thai (THT) or Chinese Thai(THC) ethnicity were included, resulting in panels of 56 THTs and 38 THCs. Forty-eight subjects of Hmong (HM) descent were recruited from a hill tribe population near Chiang Mai, Thailand. Related subjects were identified and excluded based on a ML-Relate analysis (68), as described elsewhere (69). Samples from the US were from three different sources. Chinese (USC) DNA samples were obtained from a Han Chinese sample panel (Catalog # HD100CHI) available from Coriell Cell Repositories, Camden, NJ. This sample is comprised of 50 males and 50 females, all of whom reported that all four of their grandparents were born in Taiwan, China or Hong Kong. The 160 European American (EA) and 48 African American (AA) subjects were recruited at the University of Connecticut Health Center or at the VA Connecticut Healthcare System, West Haven Campus. All subjects enrolled in the study provided informed consent as approved by the institutional review board at the appropriate institution.

AAs for association study

Additional 276 cases diagnosed with AD and 242 healthy control subjects of African descent in the US who were screened to exclude psychiatric involvement were recruited from 3 addiction clinic sites including the University of Connecticut Health Center, the VA Connecticut Healthcare System, West Haven Campus, McLean Hospital, and Medical University of South Carolina. All subjects enrolled in the study provided informed consents as approved by the institutional review board at the appropriate institution. Additional affected or unaffected parents and siblings were excluded from the study. Ethnicity of all subjects was identified according to self identification and systematically confirmed by Bayesian model-based clustering method using genetic

marker information (69). All subjects were also interviewed for AD and other psychiatric disorders by trained interviewers using the Structured Clinical Interview for DSM-IV disorders (SCID-I) criteria for psychiatric diagnosis. Informed consents were obtained from all participants prior to enrollment as approved by the institutional review board at each clinical site. DNA samples were mostly obtained from immortalized cell line. Only a minority of samples were extracted directly from peripheral blood or saliva. All samples recruited to this study were checked for consistency with the set of samples previously reported by Covault *et al* (27). Any sample that was already used was excluded from current study.

Selection of SNP markers

Thirteen SNPs for LD block definition

We selected 13 SNPs with minor allele frequency > 0.15, based on their use in previous studies, and/or map position. These SNPs span the 312.6 kb bp region including *GABRA2* and *GABRG1*, and were genotyped in the six populations described above. Six SNPs that map to *GABRA2* (rs567926, rs534459, rs529826, rs279869, rs279858, and rs279837) are a subset of the 10 SNPs reported previously by Covault et al.(20) and are designated here as SNPs G to L, respectively. The other seven SNPs include three intergenic SNPs and four that map to *GABRG1*, selected from the NCBI database based on LD and intermarker distance. No SNPs resulting in amino acid change are known to map in this region; only one known nonsynonymous SNP (rs279858) in the *GABRA2* coding region was genotyped (Table 2). All of these SNPs were polymorphic in all populations.

Tagging SNPs

Six haplotype tagging (ht) with minor allele frequency (MAF) of 15% or higher, including rs1497571, rs10938426, rs10033451, rs567926, rs279869, and rs279837, that span a 312.6 kb region including *GABRG1* and *GABRA2*, were identified by using the htSNP approach in *HAPLOVIEW* software version 3.32 (available at http://www.broad.mit.edu/mpg/haploview) (70), from set of 13 SNPs that span the 312.6-

kb region including *GABRA2* and *GABRG1* previously genotyped in 48 AA individuals (Figure 6) (28).

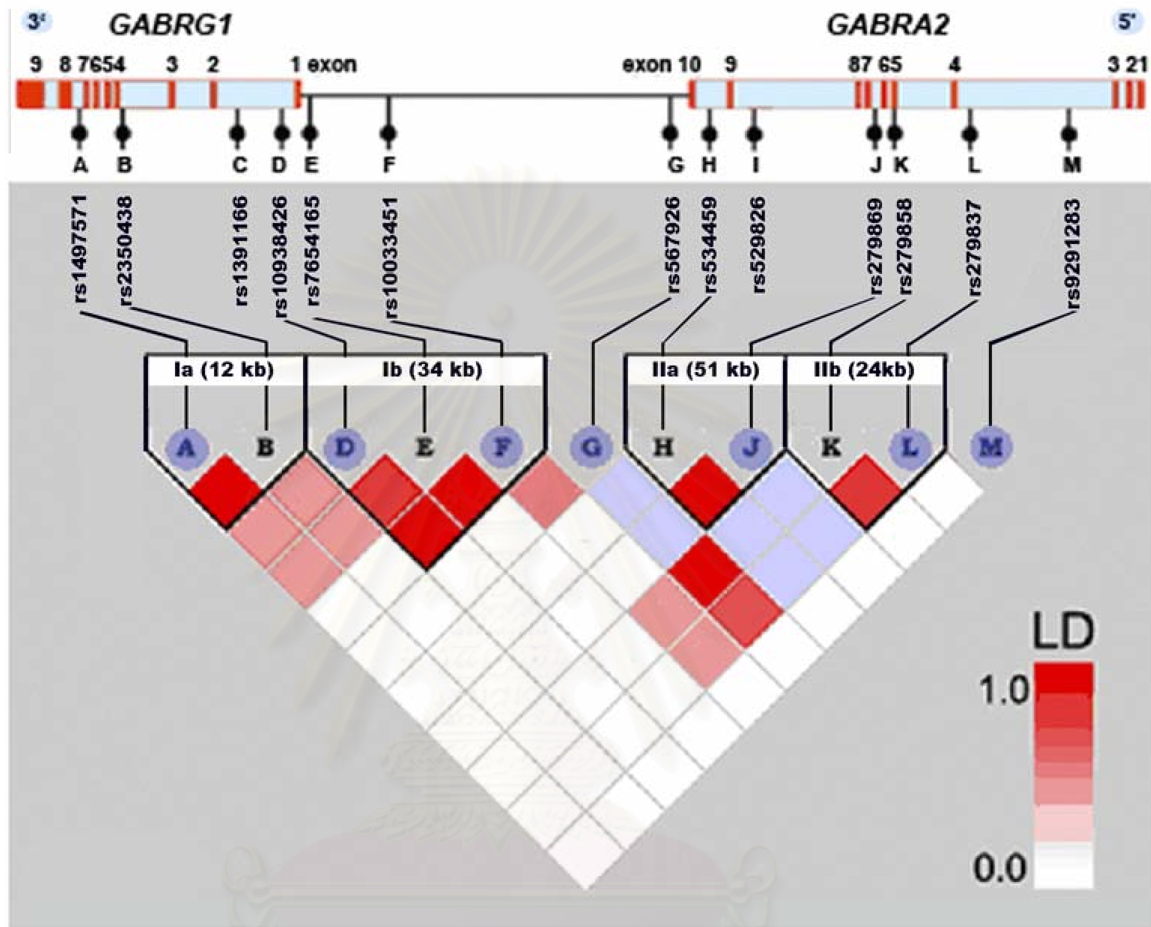


Figure 6 (28) LD structure and location of 6 htSNPs over the GABRG1/GABRA2 region on chromosome 4p13-12 in African American population (N=48). Blue marks represent htSNPs. Markers with>25% missing genotypes are excluded as they are not informative for haplotype block definition by HAPLOVIEW

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Genotyping

The TaqMan method, a fluorogenic assay based on 5'-nuclease activity (71), was employed for genotyping. All 13 SNPs were identified in all subjects by using specific assays synthesized and designed by Applied Biosystems (Foster City, CA). All of these TaqMan probe primers are available as ABI “Assays-on-Demand.” Each PCR reaction was performed with a reaction volume of 2μl including 1X concentration of

Taqman 2X Universal PCR Master Mix (Applied Biosystems), 1X concentration of 20X assay mix, distilled water, 1X concentration of 100X BSA, and 2 ng DNA. PCR amplification was accomplished using an ABI 9700 thermocycler at 95°C for 10 min, followed by 15 s at 92°C, then 60 s at 60°C for 40 cycles, before detection in the ABI PRISM 7900HT Sequence Detection System (SDS) and analysis using software available from the Applied Biosystems.

All genotypes were replicated twice and compared to each other for quality control. Data from genotyped plates with fewer than 3 out of 384 (0.78 %) mismatches (genotyping error) were employed in the analysis (with exclusion of any mismatched genotypes). Genotyping of plates with a higher error rate was repeated; if there were 3 or few discordances between two 384-well plates, then the concordant genotypes from these two plates were used and the (3 or fewer) discordances, discarded.

Statistical analysis

LD block definition

Allele frequencies for each SNP marker were calculated by using POWERMARKER(72). The Hardy-Weinberg equilibrium (HWE) exact test was applied. The extent of LD was estimated using the confidence-interval method in HAPLOVIEW software version 3.32 (available at http://www.broad.mit.edu/mpg/haploview)(70), based upon pairwise |D'| calculation between markers. LD measures and haplotype block structure were obtained using the program Haploview. Minimum percentage of genotype was set to exclude markers in which less than 75% of individuals were genotyped. Most markers exceeded this ratio, except rs1391166, rs10938426 and rs 7654165 in the THC population, and markers rs1391166 and rs529826 in the AA population. Genotype data for SNPs in the same region ranging from chromosome 4 nucleotides 45,874,178 – 46,226,602, which span these two genes and cover all of the 13 SNPs, was downloaded from the HapMap project website (http://www.hapmap.org) and were analyzed by HAPLOVIEW to compare LD block structures between our populations and the three HapMap populations of comparable ethnicities (EA comparable to CEU; AA comparable to YRI; and Asian populations: USC, Thais (THT, THC) and HM, comparable to CHB from HapMap). Comparisons were made in two

scales. The 11 SNPs presented in our study that were also genotyped in the HapMap population were selected for HAPLOVIEW analysis, but there were no available genotypes for SNPs B and I (rs2350438 and rs529826, respectively) in the HapMap dataset. LD patterns, including all informative SNPs (with minor allele frequency >0.15), were selected from HapMap for analysis. Cutoff rate for minor allele frequency (MAF) at the level of 0.15 in this study is based on data presented in HapMap. This was reported previously by Bonnen et al. (2006) (73) in a whole-genome study in 5 populations including the four HapMap populations (JPT, CHB, CEU, and YRI). The allele frequency distribution of SNPs in these populations was similar to that in others for markers with a minor allele frequency > 15%. This minimum MAF threshold represented > 75% of the 110,000 genotyped SNPs in HapMap populations. We also analyzed subsets of our dataset defined as haplotype tagging SNPs (htSNPs) selected from LD blocks using the Tagger function within Haploview. Analysis was performed using both "pairwise" and "aggressive tagger" Haploview functions. This function is an extension algorithm developed by Carlson et al (74) and based on de Bakker's tagger (75).

Haplotype blocks definition were considered a standard output of Haploview. One of the major utilities of haplotype studies such as this is to try to define basic haplotype blocks of this region in specific populations. No haplotype data are presently available for Thai and Hmong populations. We can use HapMap data as preliminary data but they are not specific to the populations that we have. These data will be useful for marker selection in future studies using Thai and Hmong populations and for an improved understanding of interlocus LD. That is, we can use haplotype block data to define "tagging" SNPs in each block for further association studies in these populations.

We used the "Confidence Interval" option, based on the algorithms published by Gabriel et al., Science 2002 (76), available as the default parameter in HAPLOVIEW to identify LD blocks. This option provides a method to define LD blocks using 95% interval bounds on D' value. A block is "created" if 95% of informative comparisons are in strong LD. The other two block definitions in HAPLOVIEW are: (a) "Four Gamete Rule" option described in Wang et al., Am J Hum Genet., 2002 (77), and (b) "Solid Spine of LD" option. The Four Gamete Rule option creates blocks by computing four possible two-marker haplotypes generated by each pair of markers. If all four are observed with

at least a frequency of 0.01, recombination is assumed to have occurred. The Solid Spine option creates blocks by inferring the first and the last markers which are in strong LD with all other intermediate markers (whereas intermediate marker themselves are not necessary in LD with each others). Using different HAPLOVIEW block definitions may (and often do) give different patterns of LD blocks. Thus, interpretation of LD patterns is dependent on and should refer to the options based on different algorithms mentioned in the instructions.

The minimum 75% call rate threshold was set for the HAPLOVIEW "minGeno" in this study. The threshold, as indicated in the HAPLOVIEW instructions, can range between 0 and 1, and has a default value of 0.5. The 75% minGeno value used in this study is acceptable according to the HAPLOVIEW instructions. For all 13 loci genotyped in 6 populations (i.e. 78 locus-population pairs), 71 of 78 loci (91%) showed a greater than 75 percent genotyping rate. The other seven loci with lower completion rates (67%-73%) in some populations (rs1391166 and rs rs529826 in AAs, rs10938426, rs7654165, rs10033451 and rs279837 in THCs, and rs7654165 in THT) were excluded from HAPLOVIEW analysis. These loci were not shown in the LD plot in Figure 7 (in Chapter 4: Results).

Previously, systematic error and a significant effect of missing data on haplotype analyses was shown when missing data for any locus is larger than 10% (78). However different software was used for that study (JLIN for LD plot and SIMHAP for haplotype analysis). Potential thresholds used by one program may not be applicable to another. In addition, most of the markers in our study map to high-LD blocks where they share information with multiple other markers, We believe that default software thresholds used in this study are acceptable.

To examine haplotype frequencies of the htSNPs, PHASE 2.0.2 software (79), based on Bayesian statistical methods, was employed to reconstruct and estimate haplotype frequencies.

Association between *GABRA2/GABRG1* variations in AA population

For individual SNP association, genotype frequency and allele frequency of each SNPs were determined in both case and two groups. Chi-square test corresponding to p-value <0.05 in any single SNP comparison was employed to determine goodness of fit between observed and expected genotype or allele frequencies comparing subjects with AD and controls. Linkage disequilibrium blocks which refer to nonrandom association of alleles containing group of 6 tagging SNPs and block of the two nominally significant single SNPs (rs10938426 stratified by rs279869), were constructed. Haplotype frequency estimation of the commonly observed haplotype patterns in case and control groups was inferred using the computer PHASE program (79) (version 2.02 available at http://www.stat.washington.edu/stephens/phase.html). Chi-square test was also used to test for association between Multi-marker haplotype and AD phenotype. BonFerroni correction for multiple significance tests was implicated by dividing the alpha level at 0.05 with the more number of common haplotypes presented in case and control group. The Bonferroni-corrected level of significance required for these comparisons was observed when a nominal p-value was less than 0.0125 (0.05/4) and 0.005 (0.05/10) for 2-SNP and 6-SNP haplotype association respectively.

CHAPTER IV
RESULTS

Allele frequency differences between populations

Allele frequencies in each population are presented in Table 2. The genotype distributions of all SNPs were consistent with Hardy-Weinberg equilibrium expectations, in all populations.

Analysis of LD and Tagging SNPs

LD and haplotype block structure results are summarized in Figure 7. Blocks were defined by LD analysis function and the confidence-interval method (74-76) selected within the parameter list. In the HM population, high LD (D' > 0.85) extended from rs1497571 in intron 7 of *GABRG1* to rs279837 in intron 3 of *GABRA2*, a 280 kb segment, whereas in other populations, there were two or more LD blocks across this region. The LD block boundary observed in other populations corresponds to lower D' between rs10033451 and rs567926, which are located 89 kb apart in an intergenic region. Based on the SNP tagging function in HAPLOVIEW, we determined that the largest number of htSNPs required was seven, in the AA population, whereas the smallest number required was three, in the HM population (Figure 7).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 2 Location and prevalence of allele frequencies of GABRG1 and GABRA polymorphisms

Marker information					Populations					
					THC	HM	THT	AA	USC	EA
	Marker	Position	gene	NT	N=38	N=48	N=59	N=48	N=100	N=160
A	rs1497571	45,900,178	*GABRG1* intron 7	[C/T]	0.66/0.34	0.73/0.27	0.62/0.38	0.38/0.62	0.57/0.43	0.56/0.44
B	rs2350438	45,912,197	*GABRG1* intron 3	[C/T]	0.66/0.34	0.72/0.28	0.62/0.38	0.37/0.63	0.57/0.43	0.56/0.44
C	rs1391166	45,946,795	*GABRG1* intron 1	[A/T]	0.39/0.61	0.28/0.72	0.35/0.65	0.77/0.23	0.45/0.55	0.50/0.50
D	rs10938426	45,959,163	*GABRG1* intron 1	[A/G]	0.66/0.34	0.72/0.28	0.59/0.41	0.30/0.70	0.56/0.44	0.51/0.49
E	rs7654165	45,967,995	Intergenic	[C/T]	0.62/0.38	0.73/0.27	0.64/0.36	0.30/0.70	0.55/0.45	0.53/0.47
F	rs10033451	45,993,260	Intergenic	[C/T]	0.28/0.72	0.24/0.76	0.34/0.66	0.29/0.71	0.38/0.62	0.44/0.56
G	rs567926	46,082,697	Intergenic	[A/G]	0.46/0.54	0.43/0.57	0.49/0.51	0.81/0.19	0.54/0.46	0.61/0.39
H	rs534459	46,097,733	*GABRA2* intron 9	[A/G]	0.50/0.50	0.55/0.45	0.48/0.52	0.71/0.29	0.46/0.54	0.39/0.61
I	rs529826	46,112,580	*GABRA2* intron 8	[C/T]	0.48/0.52	0.55/0.45	0.47/0.53	0.70/0.30	0.42/0.58	0.39/0.61
J	rs279869	46,148,923	*GABRA2* intron 6	[G/T]	0.50/0.50	0.43/0.57	0.48/0.52	0.28/0.72	0.50/0.50	0.60/0.40
K	rs279858	46,155,521	*GABRA2* exon 5 (silent)	[C/T]	0.48/0.52	0.57/0.43	0.52/0.48	0.19/0.81	0.53/0.47	0.39/0.61
L	rs279837	46,180,251	*GABRA2* intron 3	[A/G]	0.48/0.52	0.36/0.64	0.45/0.55	0.78/0.22	0.44/0.56	0.61/0.39
M	rs9291283	46,212,761	*GABRA2* intron 3	[A/G]	0.13/0.87	0.18/0.82	0.17/0.83	0.33/0.67	0.04/0.96	0.26/0.74

Position is given according to the numbering in Genbank database entry NT_006238 (Homo sapiens chromosome 4 genomic contig, reference assembly). NT represents nucleotide change from allele 1, Genbank NT_006238 strand, to allele 2. Both are precent in the parenthesis [allele1/allele2].Abbreviations: THC=Chinese Thai, HM= Hmong, THT=Thai, AA=African American, USC=Chinese American, EA=European American

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

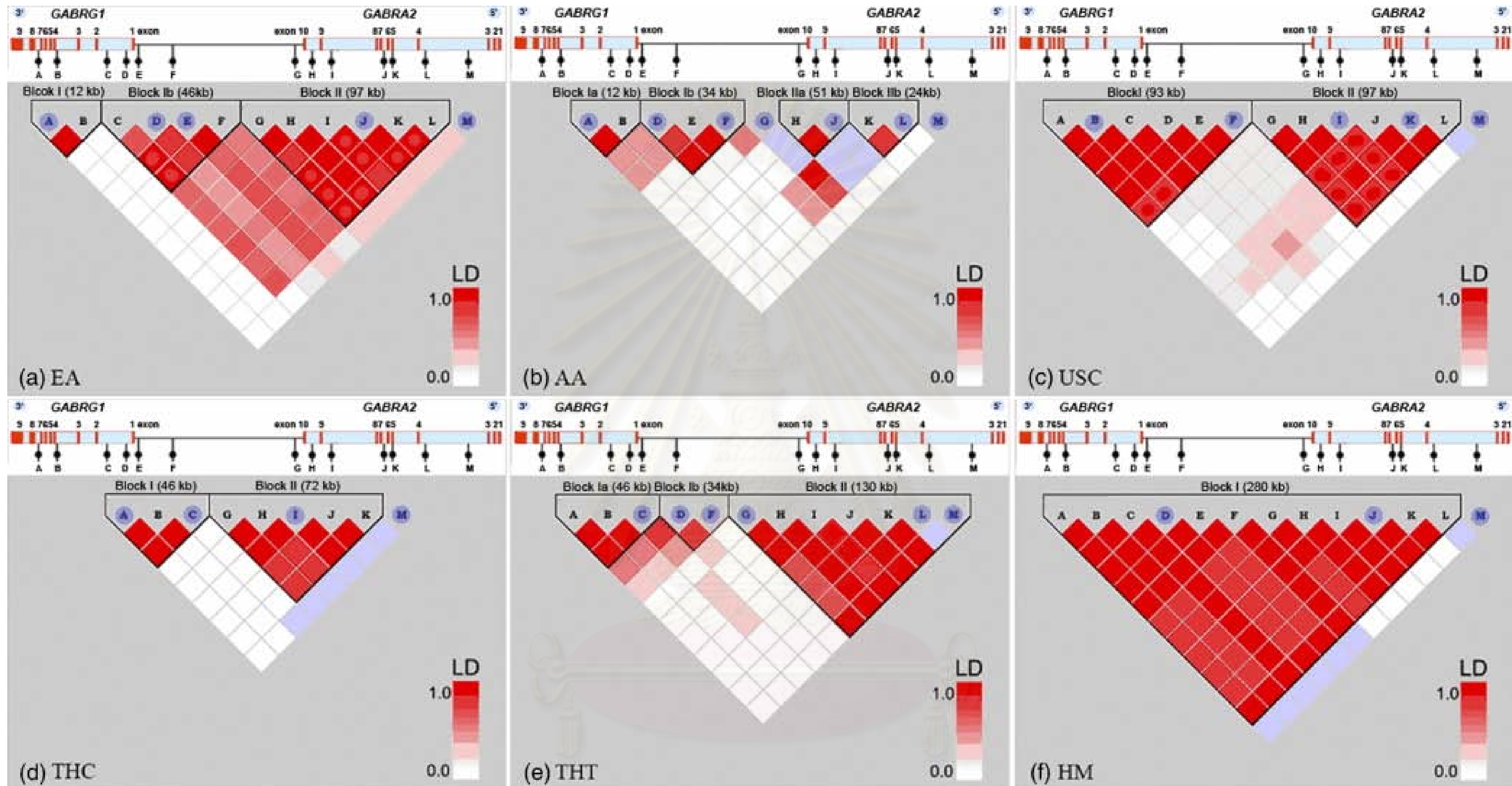


Figure 7 Comparisons of LD structure and htSNPs over the GABRG1/GABRA2 region on chromosome 4p13–p12 across five different ancestral populations. Blue marks represent htSNPs shown in each population. Some markers with > 25% missing genotypes are excluded as they are not informative for haplotype block definition. (a) EA, European American; (b) AA, African American; (c) USC, Chinese American; (d) THC, Chinese Thai; (e) THT, Thai; (f) HM, Hmong.

Haplotype frequencies

Haplotype frequencies (for those >0.05) are summarized in Table 3. The number of common haplotypes in populations varied between 4 (HM) and 8 (AA). Haplotype diversity also varied among populations. Four common haplotypes in the HM population accounted for 88% of the total information of all haplotypes. For the other populations, the set of population-specific common haplotypes represented 62-77% of the total information.

Table 3 Distribution of common haplotype among population: 13 SNPs (A-M)

Haplotype; SNPs A-M	THC	HM	THT	AA	USC	EA
C-C-T-A-C-T-A-G-T-G-T-A-G	0.27	0.35	0.25	0.09	0.3	0.2
T-T-T-A-C-T-A-G-T-G-T-A-G						0.16
C-C-A-G-T-C-A-G-T-G-T-A-G						0.08
C-C-A-G-T-C-G-A-C-T-C-G-A						0.08
T-T-A-G-T-C-G-A-C-T-C-G-G	0.07	0.24	0.12		0.21	0.08
C-C-A-G-T-C-G-A-C-T-C-G-G						0.07
T-T-A-G-T-C-G-A-C-T-C-G-A			0.05	0.06		0.05
C-C-T-A-C-T-A-A-C-T-T-A-A				0.06		
C-C-T-A-C-T-A-A-C-T-T-A-G				0.05		
C-C-T-A-C-T-A-G-T-G-T-G-G					0.05	
C-C-T-A-C-T-G-A-C-T-C-G-A		0.17	0.07			
C-C-T-A-C-T-G-A-C-T-C-G-G	0.14	0.12	0.11		0.14	
C-C-T-A-C-T-G-G-T-G-T-A-G	0.06					
T-T-A-G-T-C-A-G-T-G-T-A-G	0.06		0.06	0.05	0.07	
T-T-A-G-T-T-A-A-C-T-T-A-G				0.15		
T-T-A-G-T-T-A-G-T-G-T-A-G				0.06		
T-T-A-G-T-T-G-A-C-T-C-G-G	0.08					
Total	0.68	0.88	0.66	0.52	0.77	0.72

In most of the populations studied, there are two major haplotype blocks. Block I includes SNPs A-F, and spans a region of 93 kb, and block II includes SNPs G-L and extends 98 kb. The LD break (between these blocks) observed in most populations corresponds to lower D' between SNP F and G, which are located 89 kb apart in an intergenic region. The populations that showed different LD structure were the HM, where there was a single block spanning both of these blocks (i.e., SNPs A through L); and the AA population, where we observed four main blocks. In the AAs, block I was divided into Ia (comprised of SNPs A and B) and Ib (SNPs D-F), and block II was divided into blocks IIa (SNPs H and J) and IIb (SNPs K and L), as shown in Figure 7. The EAs and THT also showed a small difference in block definition in which smaller blocks were defined rather than the two main haplotype blocks. Three main haplotype blocks were observed in these populations. The definition of blocks in EAs is as seen in AAs where LD breaks between SNP B and C and creates sub-block Ia and Ib; whereas block I in THT were divided into block Ia and Ib by LD breaks between SNP C and D (Figure 7a and 7e).

Since it was previously reported that in EAs the LD block spanning from SNPs G-L contains haplotypes that are associated with alcohol dependence (20), the two major haplotype blocks (Block Ib and Block II) examined in this population across *GABRA2* and *GABRG1* in this study were used as a standard to compare haplotype frequencies among populations. Three haplotype blocks were reconstructed in the EA population. We analyzed frequencies of the 4-SNP haplotypes (block Ib) and the 6-SNP haplotype (block II) compared to other populations. Block I, with SNPs C-F, covers part of the 5' region, a 46-kb segment of the *GABRG1* gene; and Block II, comprised of SNPs G-L, which spans the 5' flanking region through exon 3 of the *GABRA2* gene. Block II corresponds to that identified in a previous study [15] starting at rs567926 (SNP G) and extending to rs279837 (SNP L), covering a 97 kb region.

Two major block I haplotypes, T-A-C-T and A-G-T-C, were observed in all populations except AA (Table 4). These two common haplotypes accounted for 78% (THC), 97% (HM), 79% (THT), 93% (USC) and 86% (EA) of those observed. In AA, however, the most common haplotypes were T-A-C-T and A-G-T-T (Table 4). For block II (six-locus LD

analysis for SNPs G-L), haplotypes A-G-T-G-T-A and G-A-C-T-C-G represented the common haplotype in other populations with 83-96% of chromosomes, there was only 45% (26% and 19% for A-G-T-G-T-A and G-A-C-T-C-G respectively) in African Americans. The haplotype A-A-C-T-T-A was uniquely found in African American with 48% of total chromosomes (Table 5).

Table 4 Haplotype frequencies reconstructed by comparing Block Ib (SNP C-D-E-F) in European Americans to other populations

Haplotype SNP C-D-E-F	THC	HM	THT	AA	USC	EA
T-A-C-T	0.56	0.73	0.53	0.26	0.55	0.46
A-G-T-C	0.22	0.24	0.26	0.29	0.38	0.40
T-G-C-T						0.05
A-G-T-T	0.16	0.03	0.07	0.40	0.07	
T-G-T-C			0.06			
Total	0.94	1.00	0.92	0.95	1.00	0.91

Table 5 Haplotype frequencies reconstructed by compared Block II (SNPs G-L) in European American to other populations

Haplotype SNP G-H-I-J-K-L	THC	HM	THT	AA	USC	EA
A-G-T-G-T-A	0.40	0.35	0.40	0.26	0.40	0.62
G-A-C-T-C-G	0.43	0.55	0.44	0.19	0.44	0.34
A-G-T-G-T-G	0.02	0.05	0.01	0.02	0.06	0.01
G-G-T-G-T-A	0.08	0.01	0.03		0.01	
A-G-T-T-C-G	0.02	0.02	0.06		0.07	
A-A-C-T-T-A				0.48		
Total	0.95	0.99	0.95	0.95	0.98	0.97

GABRG1 and _GABRA2_ Variation Associated with Alcohol Dependence in African American Population

Nominally significant allele frequency differences were identified for rs10938426, at *GABRG1* intron 1, with *p*= 0.044; and significant differences in both genotype and allele frequency (*p*=0.008 and 0.007 respectively) were observed at rs279869, located at *GABRA2* intron 6 (Table 6). We performed haplotype association analysis by means of PHASE. Common haplotypes inferred from all 6 tagging SNPs and 2 SNPs significantly associated to AD with each responding frequencies were shown in Table 7 and 8 respectively. Distribution of 6-SNP haplotype was observed with 10 patterns with frequency 4% or higher in each that accounted for 85% of the chromosomes among control samples and 80% of AD subjects. Six-SNP haplotypes combining SNPs from both gene loci, *GABRG1* and *GABRA2*, showed differences between controls and AD subjects, *p*=0.0027 (Table 7), significant after Bonferroni correction. Interestingly, the common two-SNP haplotype composed of SNP D and J (rs10938426 (A→G) and rs279869 (G→T)), which all 4 possible haplotypes accounted for all chromosomes in both groups, showed greater statistical significance (*p*=0.00013) for A-T haplotype (Table 8). This highly significant 2-SNP A-T haplotype is consistent with the significant (*p=0.0027)* 6-SNP CATATA haplotype which contains A allele for SNP D and T allele for SNP J. Association analysis of haplotypes defined within each gene showed no other association between any other *GABRG1* or *GABRA2* haplotype and AD risk.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 6 *GABRG1-GABRA2* SNP information, genotype and allele frequencies for African American subjects (*P* < 0.05 bolded text), Allele nucleotide designation refers to NCBI sequence NT_006238

SNP information		Frequencies		Controls vs AD
		Controls	AD	
		(N=242)	(N=276)	Chisq *p*-value
A	CC	0.145	0.142	0.499
hCV3030378	CT	0.483	0.484	
rs149571	TT	0.372	0.375	
C/T	C	0.386	0.384	0.705
GABRG1	T	0.614	0.616	
intron 7				
D	AA	0.079	0.035	0.068
hCV1445604	AG	0.408	0.383	
rs10938426	GG	0.513	0.582	
A/G	A	0.283	0.227	**0.044**
GABRG1	G	0.717	0.773	
Intron 1				
F	TT	0.406	0.392	0.289
hCV11763588	TC	0.452	0.505	
rs10033451	CC	0.142	0.103	
T/C	T	0.632	0.645	0.668
GABRG1	C	0.368	0.355	
26kb 5'				

(continue to next page)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 6 (continued)

SNP information		Frequencies		Controls vs AD
		Controls	AD	
		(N=242)	(N=276)	Chisq *p*-value
G	GG	0.064	0.055	0.592
hCV7537087	GA	0.415	0.378	
rs567926	AA	0.521	0.567	
G/A	G	0.271	0.244	0.332
GABRA2	A	0.729	0.756	
9kb 3'				
J	GG	0.095	0.117	**0.008**
hCV8262927	GT	0.377	0.500	
rs279869	TT	0.528	0.383	
G/T	G	0.284	0.367	**0.007**
GABRA2	T	0.716	0.633	
intron 6				
L	AA	0.533	0.548	0.853
hCV8263070	AG	0.396	0.393	
rs279837	GG	0.071	0.059	
A/G	A	0.731	0.744	0.632
GABRA2	G	0.269	0.256	
intron 3				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 7 *GABRG1-GABRA2* 6-SNP haplotype association (SNP A to L)

haplotype		Controls		AD			
ADFGJL	ADFGJL	allele (N)	E(freq)	allele (N)	E(freq)	chisq	*p*-value
221221	TGTATA	99	0.19	120	0.19	0.1788	0.6724
222122	TGCGTG	76	0.14	82	0.10	0.2082	0.6482
111211	CATAGA	65	0.12	78	0.10	0.0675	0.7950
121221	CGTATA	43	0.08	61	0.09	1.2270	0.2680
222211	TGCAGA	51	0.10	38	0.07	4.6594	**0.0309**
221211	TGTAGA	15	0.04	32	0.06	4.2054	**0.0403**
222221	TGCATA	8	0.02	22	0.05	4.8785	**0.0272**
111221	CATATA	44	0.09	25	0.05	8.9847	***0.0027**
122122	CGCGTG	22	0.05	21	0.05	0.4023	0.5259
221122	TGTGTG	9	0.02	19	0.04	2.3762	0.1232
		432	0.85	498	0.80		

P < 0.05 bolded text

Global p-value = 0.0257 (df=9, chi-square=25.39)

*Significance at threshold 0.005 based on Bonferroni correction corresponding to 0.05/10

Table 8 Two-SNP haplotype association (SNP D and J) from table 2 result

haplotype		Controls		AD		Chisq	*p*-value
		allele (N)	E(freq)	allele (N)	E(freq)		
11	AG	72	0.14	91	0.14	0.215	0.64258
12	AT	58	0.14	31	0.09	14.591	***0.00013**
21	GG	59	0.15	90	0.18	2.713	0.09952
22	GT	307	0.58	374	0.60	0.428	0.51307
		496	1.00	586	1.00		

P < 0.05 bolded text

*Bonferroni correction *p*-value threshold = 0.0125

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER V
DISCUSSION

With methodological developments in high-throughput genotyping, and the development of large-scale genotyping projects such as the international HapMap project, huge quantities of data are being generated making it possible to measure correlations between SNP genotypes and creating better marker maps for association studies. Improved understanding of underlying LD-based mapping approaches has been helpful in elucidating the relationship between common genetic variation and heritable risk for common diseases (49, 80, 81). Polygenic and oligogenic effects are considered to play important roles in influencing complex traits and behaviors. Further, it is likely that the genetic architecture of common, complex traits will differ, at least to some extent, among major population groups. To understand the genetic basis of these traits, we need to understand linkage disequilibrium profiles and haplotype diversity in genomic regions of interest in multiple populations.

In the present study, we focused on two adjacent $GABA_A$ receptor subunit genes, one of which has been associated to risk for alcohol dependence in several previous studies. We did not consider estimating LD in the other two GABA receptor subunit genes (*GABRA4* and *GABRB1*) in the cluster. In stead, we based selection of the two genes studied on the following considerations:

1) Several articles previously reported associations between alcohol dependence and GABRA2, but not the other genes in the cluster;

2) Physical distance between *GABRA2* and *GABRG1* is less than that between *GABRA2* and others – the approximate distance is 0.15 Mb (between *GABRA2* and *GABRG1*), compared to 0.5 and 0.7 MB, the distances from *GABRA2* through *GABRA4* and *GABRB1* respectively. Less distance between the two genes, considered with the report of strong D’ at the last SNP reported by Covault et al. 2004 (20) and Edenberg et al. 2004 (19) at the 3’ end of *GABRA2*, creates a situation where it is plausible to infer that this same linkage disequilibrium might lead to association due (at least in part) to the *GABRG1* contribution.

To define extensively linkage disequilibrium in the chromosomal region, more SNPs covering the *GABRG1* gene located telomeric to *GABRA2* on chromosome 4p were studied. This made it possible to investigate whether these two genes are located within the same LD block, and, more broadly, to ascertain the possibility that the effect observed at *GABRA2* could actually be mediated through a variant mapped to *GABRG1*. We conclude that this is in fact the case, and associations observed with *GABRA2* might be attributable to functional genetic variation at the *GABRG1* locus, or that or there may be disease-related variants at both loci; this may facilitate our understanding of reported associations between *GABRA2* polymorphisms and AD, and permit a more informed search for the functional variant or variants underlying this association – a search that must now extend into the intergenic region and the *GABRG1* locus. It is not known how well genetic information from the 4 HapMap populations (CEU, YRI, JPN, and CHB) represents that of other populations around the world; it is thus useful to answer this question directly, especially for loci of great interest

Allele frequencies for the 13 SNPs of *GABRA2* and *GABRG1* that were genotyped in this study in six different populations from three continents provide a measure of allele distribution and different LD patterns among populations. These results should allow investigators studying populations similar to those characterized herein, or admixed population derivatives, to explore association between these genes and substance dependence and other phenotypes.

During genotyping process, not all samples, in this unrealated case-control study, could be genotyped. These samples were then be characterized as “missing data” or mismatched incase genotyping results were not consistent when replicated. The effect of missing data, genotyping errors, and the use of unrelated samples on haplotype inference can be discussed as followed.

(a) Effect of missing data on haplotype frequency

We estimated haplotype frequency via PHASE (version 2.0.2), based on Stephens-Smith-Donnelly (SSD) algorithm (79, 82) in terms of all 13-SNP and shortened 4- and 6-SNP haplotypes covering each gene based on the principle haplotype pattern found in the EA population.

Effect of missing data on LD and haplotype was reported previously by Hinrichs and Suarez (2005); "the ability to detect linkage disequilibrium (LD) was also substantially reduced by missing data; this in turn could affect tagging SNPs chosen to generate haplotypes" (83). The PHASE program can impute all missing alleles for haplotype scoring. The "–MR" option, suggested to be the most accurate method (set as default in PHASE program), showed similar error rate (0.16 and 0.18) between datasets of no missing data and 5% missing data run in sample size of 50 individuals (PHASE 2.0.2 instructions).

There were 5889 genotypes in this study (453 samples from all populations, 13 SNPs). With 320 genotypes missing, the completion rate is ~0.95 (5.43 % missing data). This is considered acceptable for PHASE.

(b) Effect of genotyping error on haplotype frequency

The overall error rate was < 0.01 (1%). Genotyping error rate was directly determined by regenotyping 100% of the samples. All samples used in this study were replicated in all markers and compared to each replicate for quality control. Genotyped plates with less than 3 samples from overall 384 samples in 384-well plate (0.78%) of mismatch genotype (genotyping error) were then employed in the analysis; discordant genotypes were discarded. With the 95% completion rate, error rate lower than 0.01, and Hardy-Weinberg disequilibrium test that was applied to all loci in our samples, population haplotype frequency estimates are unlikely to be affected substantially.

(c) Effect of using unrelated samples on haplotype frequency

PHASE has been widely used for haplotype frequency estimation, and is considered to employ one of the best reconstruction algorithms (82).

According to the presence of allele frequencies on Table 2, almost all SNPs showed different linkage phase between African American population and the others (i.e., different

major alleles). An explanation for this exceptional phenomenon can be explained. It is well known that in general, non-African populations contain a subset of the genetic variation contained in populations of African ancestry and that allele frequencies can differ greatly between populations. Non-African populations may have drastically different allele frequencies from populations of African origin or ancestry simply due to a bottleneck effect on modern human populations migrating out of Africa during human settlement of Eurasia. Other evolutionary forces such as natural selection may be responsible for the observed allele frequency differences but require specific conditions to operate and can not be inferred from the limited data here. In genomic regions with high LD, such as those studied here, large differences in population allele frequencies should correspond to large differences in population haplotype frequencies, which are also seen in our data.

This present study provides the first comprehensive analysis of patterns of LD spanning these two important genes. In general, LD block size might vary between different populations in which the degree of admixture and differences in ancestries are found. Thus, from generation to generation, differences in LD between markers from the initial populations due to LD decay can be observed by changes in LD block and haplotype structure and long stretches of LD that can reflect haplotype diversity (38, 84). That is clearly the case for these loci in the populations studied. We have systematically analyzed SNPs covering the range of *GABRA2* and *GABRG1* in the GABA receptor cluster on human chromosome 4p, estimating haplotypes in a total of 453 unrelated healthy individuals. The more complex LD structure found in the African American population is consistent with what is known about the population's history, i.e., that it is older than Asian or European populations, which show less complex LD patterns. This is consistent with the observation of Gabriel et al (76) that haplotype blocks found in African populations showed a set of shorter genomic LD blocks than samples from Europe and Asia. While haplotype blocks defined in African Americans indicated the presence of 4 main LD regions spanning all 13 of the SNPs examined, we found that Chinese populations both in Thailand and in the US, and the Thai population were characterized by a two-block structure in this genomic region. In contrast, the Hmong population, which is an isolated minority Hill Tribe dwelling in the

northern part of Thailand, was found to have a unique long-ranged haplotype block structure with a single block encompassing the two blocks observed in the others. This interesting finding in the Hmong population may be attributed to the occurrence of recent selection that can be reflected in long-range haplotypes (85). Alternatively, the Hmong population may be relatively new or have undergone a recent bottleneck (69). A small number of htSNPs are generally sufficient to capture most of haplotype structure in high LD regions (38, 86); consistent with this expectation, only three htSNPs were required to capture the haplotype block diversity in the Hmong population, whereas four htSNPs were necessary for Chinese Thais (THC), five for European American and Chinese American (USC), and six for Thais (THT). In the AA population, there were four haplotype blocks in the same region; seven SNPs were sufficient to capture most of the genetic information of the total 13 SNPs. As can be seen in Figure 7, rs9291283 (SNP M) showed very low LD with the other SNPs examined, falling outside the limits of the haplotype block. Thus, it was tagged as a htSNPs for all groups.

LD analysis based on HapMap data with a denser SNP map (Figure 8) shows that LD definitions in our sample are consistent with those based on well-characterized HapMap populations. The LD regions observed in this study; European American, African American, Chinese American (USC), are similar to those observed in CEU, YRI, and CHB from the HapMap project, respectively (Figure 8); whereas the LD pattern in our Thai populations, compared to HapMap CHB population, is slightly different in Thais (THT) and Chinese Thais (THC) and very different in the Hmong population. This suggests a degree of genetic differentiation among these populations. The Thai population may have unique genetic characteristics, especially the minority Hmong population (recruited in Thailand but present in other Asian countries, including China and Laos, as well).

Additionally, regarding the HapMap Beijing Chinese SNP data (genotyped from 90 individuals as available at the Coriell's catalogs HAPMAPPT02 and HAPMAPPT05), these samples were recently proved to be applicable to the Southern population of Han Chinese in Shanghai, recruited at the city located over 1,000 kilometers south away from Beijing (87). This is at least proved and confirmed that The HapMap CHB population is a good reference

for other Han Chinese population for the future study of complex diseases among Southern Chinese population. In this study, to investigate homogeneity of these two Han Chinese populations, we compared the HD100CHI population with the HapMap CHB population by observation of population structure in terms of linkage disequilibrium patterns based on SNP-based genotyping data. The result showed that these populations were similar in their LD structures at least in the region of *GABRA2* and *GABRG1* genes in the chromosome 4p13-12 region.

Because each population has its own evolutionary history, in which distinct allele frequencies, LD patterns, and haplotype structures can develop, a number of studies show that htSNPs must be defined within specific populations to identify optimal sets of marker for association studies (88-90). Previously, *GABRA2* haplotype structure was reported only in EA, AA, and Russian populations (20, 21, 24); the present study documents LD patterns and haplotypes in three different Asian populations. As shown in European Americans by Covault et al.(20), seven markers spanning the *GABRA2* gene defined an LD block associated with AD. Our study, which used six of those seven markers (we omitted rs279844), also shows a haplotype block defined by SNPs G (rs567926) to L (rs279837) that confirms the same region of strong LD in European Americans.

Haplotype frequencies in both LD blocks found in African Americans were uniquely different from those observed in other populations. In the Hmong population, we observed a small number of 13-SNP haplotype patterns indicative of low diversity, commonly found in an isolated population.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

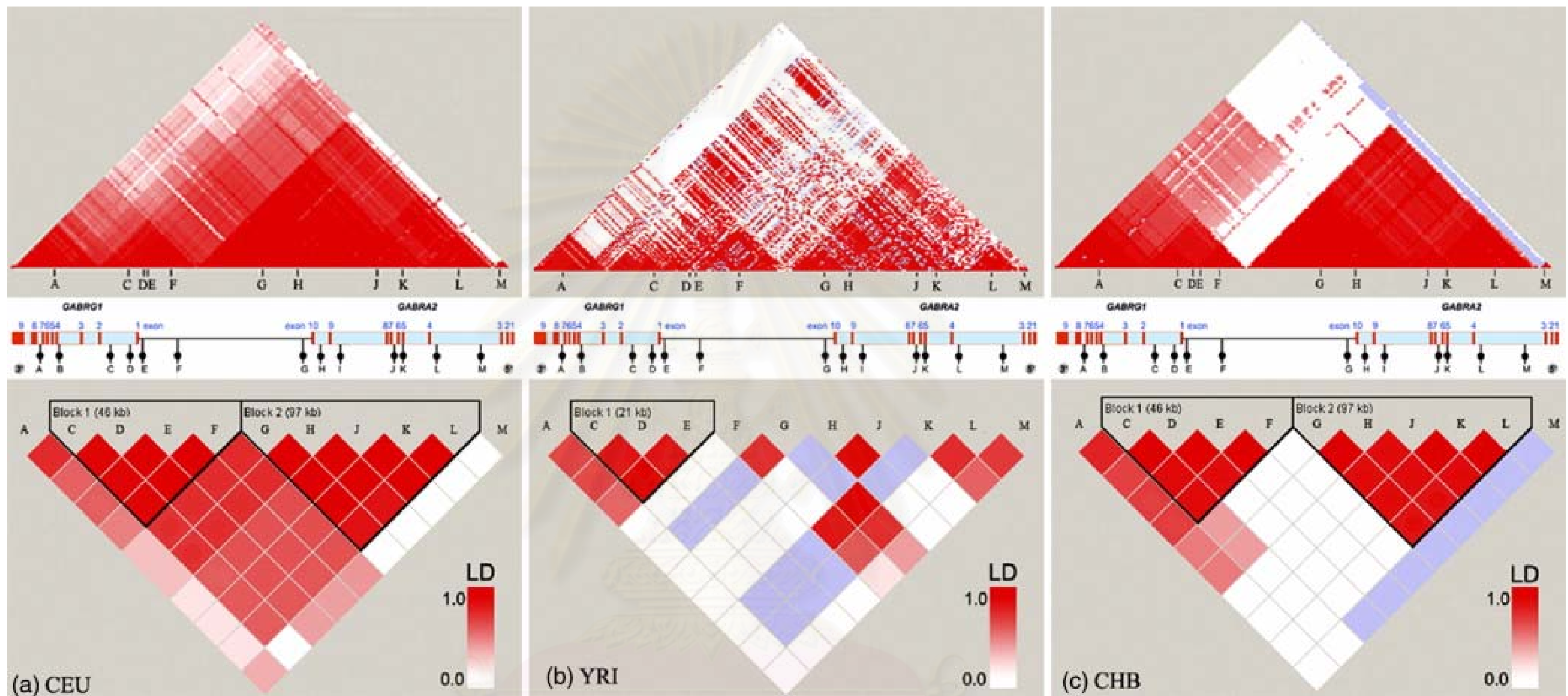


Figure 8 LD patterns computed based on genotype data from the HapMap CEU, YRI, and CHB populations for 11 SNPs used in our study. Note that the 2 SNPs unavailable in HapMap, rs2350438 and rs529826, are not included (28).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

In the analysis of the 4-SNP (C-F) and 6-SNP (G-L) haplotypes, the two haplotype blocks observed in EAs differed from those in AAs. Specifically, the 6-SNP haplotype defined by SNPs G-L identified two complementary common haplotypes, A-G-T-G-T-A and G-A-C-T-C-G, as reported by Covault et al. (20), which together accounted for 92.8% of chromosomes (in the control sample), represented 96% of chromosomes in the European American, 90% in the Hmong, 84% in the Thai (THT) and Chinese American (USC), 83% in the Chinese Thai (THC), and interestingly, only 45% in the African American populations. The African American population showed a specific haplotype, A-A-C-T-T-A, which was not observed in the other populations. In the case of both the 4-SNP and 6-SNP haplotypes, only one recombination event is required to explain the presence of the third unique but common haplotype in the AA sample. The lack of these two common haplotypes in non-African populations may be explained by one of two most likely scenarios: (1) The recombination events took place in Africa after humans migrated out of Africa and subsequently rose to high frequency either from genetic drift or positive selection. (2) The recombination events predate migrations out of Africa, but due to a bottleneck effect, were not represented in founding populations in either Europe or Asia, or were present at such low frequencies that they were lost through genetic drift.

Only small differences were observed between haplotype frequencies in European American and the Asian populations. Further, no significant differences were found in haplotype frequencies between any two Asian groups except in comparisons involving the Hmong.

In conclusion, the first part of this study presents patterns of specific htSNPs and LD block structure in six different populations: European American, African American, Chinese American and three from Asia (i.e., Thai, Chinese Thai and Hmong). LD extended from most of the *GABRA2* gene through the *GABRG1* locus in the same $GABA_A$ cluster region on chromosome 4p, suggesting the possibility of association (and interaction) of both of these two genes with alcohol dependence. Differences in genetic architecture observed in these populations may help to define the physical and genetics regions of *GABRG1* and *GABRA2* that contains an as yet unidentified alcohol dependence-related functional change

To investigate association between *GABRA2/GABRG1* variations and AD in a population available, we selected 6 tagging SNP from total 13 SNPs previously described in 48 AA healthy control subjects with African descent (28). Current larger study populations with 276 ADs and 242 control subjects enable LD studies done to assess association between SNPs markers of *GABRA2/GABRG1* genes and AD. Definition of 4 samller LD blocks with D' value over 0.8 was observed in the *GABRG1/GABRA2* region (28). Regarding the unique LD blocks across these two genes observed in AAs that differ from other non-African reported previously by our group, the LD structure of these genes is consistent with many reports on other genomic regions of populations with African descent. Breakdown of LD owing to presence of smaller LD blocks in African populations were affected by existence of historical differences in population size and structure, long history and the age of the African population, and frequent recombination that decays population LD (91).

Nominal statistic significance for individual SNPs were observed for only 2 from 6 SNPs, rs10938426, at *GABRG1* intron 1, and rs279869, located at *GABRA2* intron 6. To investigate interactions of multiple SNPs within a haplotype, rather than any single SNP, estimation of haplotype patterns was performed to observe haplotype pattern combining all 6 tagging SNPs and haplotype that comprised only the two nominally significant SNPs. The most common 6-SNP haplotype associated with AD (p=0.0027), C-A-T-A-T-A (for rs1497571, rs10938426, rs10033451, rs567926, rs279869, and rs279837), that harbors nucleotide A and T for rs10938426 and rs279869 respectively, was found to account for 10% (44/432) of AD chromosomes and 5% (25/498) of chromosome of control subjects (table 7). The more statistical significance (p=0.00013) was shown in the common 2-SNP haplotype with the same allele of rs10938426 and rs279869 as presented in the significant C-A-T-A-T-A haplotype. Rs10938426 which is located in intron 1 of the *GABRG1* gene is ~100 kb apart from the 3' direction from exon 10, the last exon, of the *GABRA2* gene. The distance between rs10938426 and rs279869, in intron 6 of the *GABRA2* gene, is approximately 190-kb distant (referred to the Genbank NT_006238 strand). In order to further narrow down to other functional or SNPs in the same LD block with greater statistic

significance, a panel of denser SNPs can be mapped within the 34-kb Ib and 51-kb IIa LD region (Figure 6) to find combination of the other possible SNPs that strongly link to the significant SNP markers, rs10938426 and rs279869.

In conclusion, this finding suggests significance of the interrelationship between the *GABRG1 and GABRA2 genes* and the possibility of risk loci for AD in each of them. A two-SNP haplotype composed of one SNP from each gene suggests possible interaction of these genes and supports the involvement of both in predisposition to AD in AAs.

This finding suggests significance of the interrelationship between these two genes and the possibility of risk loci in each of them. A two-SNP haplotype composed of one SNP from each gene suggests possible interaction of these genes and supports the involvement of both in predisposition to AD in AAs.

REFERENCES

(1) McCormick DA, Bal T. Sleep and arousal: thalamocortical mechanisms. Annual review of neuroscience. 1997;20:185-215.

(2) Whittington MA, Traub RD, Jefferys JG. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. Nature. 1995 Feb 16;373(6515):612-5.

(3) Tobler I, Kopp C, Deboer T, Rudolph U. Diazepam-induced changes in sleep: role of the alpha 1 GABA(A) receptor subtype. Proceedings of the National Academy of Sciences of the United States of America. 2001 May 22;98(11):6464-9.

(4) Schofield PR, Darlison MG, Fujita N, Burt DR, Stephenson FA, Rodriguez H, et al. Sequence and functional expression of the GABA A receptor shows a ligand-gated receptor super-family. Nature. 1987 Jul 16-22;328(6127):221-7.

(5) Olsen RW, Tobin AJ. Molecular biology of GABAA receptors. Faseb J. 1990 Mar;4(5):1469-80.

(6) Barnard EA, Skolnick P, Olsen RW, Mohler H, Sieghart W, Biggio G, et al. International Union of Pharmacology. XV. Subtypes of gamma-aminobutyric acidA receptors: classification on the basis of subunit structure and receptor function. Pharmacological reviews. 1998 Jun;50(2):291-313.

(7) Twyman RE, MacDonald RL. Antiepileptic drug regulation of GABAA receptor channels. 4 ed. New York: Wiley-Liss; 1991.

(8) DeLorey TM, Olsen RW. Gamma-aminobutyric acidA receptor structure and function. The Journal of biological chemistry. 1992 Aug 25;267(24):16747-50.

(9) MacDonald RL, Twyman RE. Kinetic properties and regulation of GABAA receptor channels. Ion channels. 1992;3:315-43.

(10) Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, et al. Genome-wide search for genes affecting the risk for alcohol dependence. American journal of medical genetics. 1998 May 8;81(3):207-15.

(11) Harris RA, Valenzuela CF, Brozowski S, Chuang L, Hadingham K, Whiting PJ. Adaptation of gamma-aminobutyric acid type A receptors to alcohol exposure: studies

with stably transfected cells. The Journal of pharmacology and experimental therapeutics. 1998 Jan;284(1):180-8.

(12) Harris RA, Proctor WR, McQuilkin SJ, Klein RL, Mascia MP, Whatley V, et al. Ethanol increases GABAA responses in cells stably transfected with receptor subunits. Alcoholism, clinical and experimental research. 1995 Feb;19(1):226-32.

(13) Harris RA, Mihic SJ, Brozowski S, Hadingham K, Whiting PJ. Ethanol, flunitrazepam, and pentobarbital modulation of GABAA receptors expressed in mammalian cells and Xenopus oocytes. Alcoholism, clinical and experimental research. 1997 May;21(3):444-51.

(14) Iwata N, Virkkunen M, Goldman D. Identification of a naturally occurring Pro385-Ser385 substitution in the GABA(A) receptor alpha6 subunit gene in alcoholics and healthy volunteers. Molecular psychiatry. 2000 May;5(3):316-9.

(15) Belelli D, Lambert JJ. Neurosteroids: endogenous regulators of the GABA(A) receptor. Nature reviews. 2005 Jul;6(7):565-75.

(16) Davies M. The role of GABAA receptors in mediating the effects of alcohol in the central nervous system. J Psychiatry Neurosci. 2003 Jul;28(4):263-74.

(17) Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, et al. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. American journal of medical genetics. 1998 May 8;81(3):216-21.

(18) Zinn-Justin A, Abel L. Genome search for alcohol dependence using the weighted pairwise correlation linkage method: interesting findings on chromosome 4. Genetic epidemiology. 1999;17 Suppl 1:S421-6.

(19) Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, et al. Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. American journal of human genetics. 2004 Apr;74(4):705-14.

(20) Covault J, Gelernter J, Hesselbrock V, Nellissery M, Kranzler HR. Allelic and haplotypic association of GABRA2 with alcohol dependence. Am J Med Genet B Neuropsychiatr Genet. 2004 Aug 15;129(1):104-9.

(21) Lappalainen J, Krupitsky E, Remizov M, Pchelina S, Taraskina A, Zvartau E, et al. Association between alcoholism and gamma-amino butyric acid alpha2 receptor subtype in a Russian population. Alcoholism, clinical and experimental research. 2005 Apr;29(4):493-8.

(22) Fehr C, Sander T, Tadic A, Lenzen KP, Anghelescu I, Klawe C, et al. Confirmation of association of the GABRA2 gene with alcohol dependence by subtype-specific analysis. Psychiatric genetics. 2006 Feb;16(1):9-17.

(23) Enoch MA, Schwartz L, Albaugh B, Virkkunen M, Goldman D. Dimensional anxiety mediates linkage of GABRA2 haplotypes with alcoholism. Am J Med Genet B Neuropsychiatr Genet. 2006 Sep 5;141(6):599-607.

(24) Drgon T, D'Addario C, Uhl GR. Linkage disequilibrium, haplotype and association studies of a chromosome 4 GABA receptor gene cluster: candidate gene variants for addictions. Am J Med Genet B Neuropsychiatr Genet. 2006 Dec 5;141(8):854-60.

(25) Tian H, Chen HJ, Cross TH, Edenberg HJ. Alternative splicing and promoter use in the human GABRA2 gene. Brain research. 2005 Jun 13;137(1-2):174-83.

(26) Wilcox AS, Warrington JA, Gardiner K, Berger R, Whiting P, Altherr MR, et al. Human chromosomal localization of genes encoding the gamma 1 and gamma 2 subunits of the gamma-aminobutyric acid receptor indicates that members of this gene family are often clustered in the genome. Proceedings of the National Academy of Sciences of the United States of America. 1992 Jul 1;89(13):5857-61.

(27) Covault J, Gelernter J, Jensen K, Anton R, Kranzler HR. Markers in the 5'-Region of GABRG1 Associate to Alcohol Dependence and are in Linkage Disequilibrium with Markers in the Adjacent GABRA2 Gene. Neuropsychopharmacology. 2008 Mar;33(4):837-48.

(28) Ittiwut C, Listman J, Mutirangura A, Malison R, Covault J, Kranzler HR, et al. Interpopulation linkage disequilibrium patterns of GABRA2 and GABRG1 genes at the GABA cluster locus on human chromosome 4. Genomics. 2008 Jan;91(1):61-9.

(29) Pericak-Vance MA, Bebout JL, Gaskell PC, Jr., Yamaoka LH, Hung WY, Alberts MJ, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. American journal of human genetics. 1991 Jun;48(6):1034-50.

(30) Strittmatter WJ, Weisgraber KH, Huang DY, Dong LM, Salvesen GS, Pericak-Vance M, et al. Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease. Proceedings of the National Academy of Sciences of the United States of America. 1993 Sep 1;90(17):8098-102.

(31) Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature. 1994 May 5;369(6475):64-7.

(32) Majerus PW. Human genetics. Bad blood by mutation. Nature. 1994 May 5;369(6475):14-5.

(33) Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? Current opinion in biotechnology. 1998 Dec;9(6):578-94.

(34) Heutink P, Oostra BA. Gene finding in genetically isolated populations. Human molecular genetics. 2002 Oct 1;11(20):2507-15.

(35) Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. Trends Genet. 2003 Mar;19(3):135-40.

(36) Tahira T, Baba S, Higasa K, Kukita Y, Suzuki Y, Sugano S, et al. dbQSNP: a database of SNPs in human promoter regions with allele frequency information determined by single-strand conformation polymorphism-based methods. Human mutation. 2005 Aug;26(2):69-77.

(37) Gray IC, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. Human molecular genetics. 2000 Oct;9(16):2403-8.

(38) Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, et al. Haplotype tagging for the identification of common disease genes. Nature genetics. 2001 Oct;29(2):233-7.

(39) Kauppi L, Jeffreys AJ, Keeney S. Where the crossovers are: recombination distributions in mammals. Nat Rev Genet. 2004 Jun;5(6):413-24.

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

(41) Laan M, Paabo S. Demographic history and linkage disequilibrium in human populations. Nature genetics. 1997 Dec;17(4):435-8.

(42) 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. Human molecular genetics. 2000 Dec 12;9(20):2947-57.

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

(44) Lewontin RC. The Interaction of Selection and Linkage. Ii. Optimum Models. Genetics. 1964 Oct;50:757-82.

(45) Mueller JC. Linkage disequilibrium for different scales and applications. Briefings in bioinformatics. 2004 Dec;5(4):355-64.

(46) PHOEBE GB. Measures of LD. 2007 September 05, 2007 [cited 2008 April 8, 2008]; Available from: http://www.genestat.org/index.php?n=GeneStat.MeasuresOfLD

(47) Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 1995 Sep 20;29(2):311-22.

(48) Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 2003 Aug;4(8):587-97.

(49) HapMap. The International HapMap Project. Nature. 2003 Dec 18;426(6968):789-96.

(50) The International HapMap Project. Nature. 2003 Dec 18;426 (6968):789-96.

(51) The International HapMap Project. Integrating ethics and science in the International HapMap Project. Nat Rev Genet. 2004 Jun;5(6):467-75.

(52) HapMap. A haplotype map of the human genome. Nature. 2005 Oct 27;437 (7063):1299-320.

(53) The Coriell Institute. HAN PEOPLE OF LOS ANGELES PANEL OF 100 HUMAN VARIATION PANEL 2008 [cited 2008 April 8, 2008]; Available from: http://ccr.coriell.org/Sections/Search/Panel_Detail.aspx?Ref=HD100CHI&PgId=166

(54) Xiao CJ, Cavalli-Sforza LL, Minch E, Du RF. [Geographic distribution maps of human genes in China]. Yi chuan xue bao = Acta genetica Sinica. 2000;27(1):1-6.

(55) Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, et al. Genetic relationship of populations in China. Proceedings of the National Academy of Sciences of the United States of America. 1998 Sep 29;95(20):11763-8.

(56) Wen B, Li H, Lu D, Song X, Zhang F, He Y, et al. Genetic evidence supports demic diffusion of Han culture. Nature. 2004 Sep 16;431(7006):302-5.

(57) Zhao TM, Lee TD. Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Human genetics. 1989 Sep;83(2):101-10.

(58) Duffy J, Harmon R, Ranard D, Thao B, Yang K, Herr P. The Hmong
An Introduction to their History and Culture. Culture Profile. No. 18 June, 2004.

(59) Michaud J. From Southwest China into Upper Indochina: an overview of Hmong (Miao) migrations. Asia Pacific Viewpoint. 1997;38(2):119-30.

(60) Jerald Council, Loehrke J, Service GN. The Hmong Migration. [cited 2008 April 8, 2008]; Available from:
http://www.wisinfo.com/legacy/thailand/images/thailandmap.PDF

(61) Charkraborty R, Deka R, editors. Isolated Populations. . West Sussex: John Wiley and sons, Ltd.; 2002.

(62) Zak NB, Shifman S, Shalom A, Darvasi A. Population-based gene discovery in the post-genomic era. Drug discovery today. 2001 Nov 1;6(21):1111-5.

(63) Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. American journal of human genetics. 2000 Jul;67(1):170-81.

(64) Khoury M, Beaty T, Cohen B, editors. Fundamentals of Genetic Epidemiology. New York, USA Oxford University Press; 1993

(65) Chandra D, Werner DF, Liang J, Suryanarayanan A, Harrison NL, Spigelman I, et al. Normal acute behavioral responses to moderate/high dose ethanol in GABAA receptor alpha 4 subunit knockout mice. Alcoholism, clinical and experimental research. 2008 Jan;32(1):10-8.

(66) Liang J, Suryanarayanan A, Chandra D, Homanics GE, Olsen RW, Spigelman I. Functional consequences of GABAA receptor alpha 4 subunit deletion on synaptic and extrasynaptic currents in mouse dentate granule cells. Alcoholism, clinical and experimental research. 2008 Jan;32(1):19-26.

(67) Taqman. [cited 2008 April 8, 2008]; Available from:
http://www.rockefeller.edu/genomics/pdf/Applied_Biosystems_Taqman_SNP_Genotyping.pdf

(68) Kalinowski ST, Wagner AP, Taper ML. ML-Relate: Software for estimating relatedness and relationship from multilocus genotypes. Mol Ecol Notes 2006;6:576-9.

(69) Listman JB, Malison RT, Sughondhabirom A, Yang BZ, Raaum RL, Thavichachart N, et al. Demographic changes and marker properties affect detection of human population differentiation. BMC Genet. 2007;8:21.

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

(71) Shi MM, Myrand SP, Bleavins MR, de la Iglesia FA. High throughput genotyping for the detection of a single nucleotide polymorphism in NAD(P)H quinone oxidoreductase (DT diaphorase) using TaqMan probes. Mol Pathol. 1999 Oct;52(5):295-9.

(72) Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics (Oxford, England). 2005 May 1;21(9):2128-9.

(73) Bonnen PE, Pe'er I, Plenge RM, Salit J, Lowe JK, Shapero MH, et al. Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. Nature genetics. 2006 Feb;38(2):214-7.

(74) Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. American journal of human genetics. 2004 Jan;74(1):106-20.

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

(76) Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002 Jun 21;296(5576):2225-9.

(77) Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. American journal of human genetics. 2002 Nov;71(5):1227-34.

(78) McCaskie PA, Carter KW, McCaskie SR, Palmer LJ. The effect of missing data on linkage disequilibrium mapping and haplotype association analysis in the GAW14 simulated datasets. BMC genetics. 2005 Dec 30;6 Suppl 1:S151.

(79) Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. American journal of human genetics. 2003 Nov;73(5):1162-9.

(80) Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? Human molecular genetics. 2002 Oct 1;11(20):2417-23.

(81) Klein AP, Tsai YY, Duggal P, Gillanders EM, Barnhart M, Mathias RA, et al. Investigation of altering single-nucleotide polymorphism density on the power to detect trait loci and frequency of false positive in nonparametric linkage analyses of qualitative traits. BMC Genet. 2005 Dec 30;6 Suppl 1:S20.

(82) Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. American journal of human genetics. 2001 Apr;68(4):978-89.

(83) Hinrichs AL, Suarez BK. Genotyping errors, pedigree errors, and missing data. Genetic epidemiology. 2005;29 Suppl 1:S120-4.

(84) Goldstein DB. Islands of linkage disequilibrium. Nature genetics. 2001 Oct;29(2):109-11.

(85) Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002 Oct 24;419(6909):832-7.

(86) Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science. 2001 Nov 23;294(5547):1719-23.

(87) Hu C, Jia W, Zhang W, Wang C, Zhang R, Wang J, et al. An evaluation of the performance of HapMap SNP data in a Shanghai Chinese population: analyses of allele frequency, linkage disequilibrium pattern and tagging SNPs transferability on chromosome 1q21-q25. BMC Genet. 2008 Feb 27;9(1):19.

(88) Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. Human molecular genetics. 2003 Apr 1;12(7):771-6.

(89) Beaty TH, Fallin MD, Hetmanski JB, McIntosh I, Chong SS, Ingersoll R, et al. Haplotype diversity in 11 candidate genes across four populations. Genetics. 2005 Sep;171(1):259-67.

(90) Sawyer SL, Mukherjee N, Pakstis AJ, Feuk L, Kidd JR, Brookes AJ, et al. Linkage disequilibrium patterns vary substantially among populations. Eur J Hum Genet. 2005 May;13(5):677-86.

(91) Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet. 2002 Aug;3(8):611-21.

APPENDICES

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

APPENDICES

Description	HUMAN VARIATION PANEL - HAN PEOPLE OF LOS ANGELES PANEL OF 100
Sex	Males: 50 Females: 50
Aliquot Size	10 μg each
Brief Description	The Han People of Los Angeles Panel of 100 is a selection of 50 male and 50 female samples. Each sample is from an individual of Han ethnicity unrelated to all others in the Panel and has all four grandparents born in Taiwan, China, or Hong Kong.
Price	Commercial Pricing: $3,800.00 Academic and not-for-profit pricing: $3,800.00 Submitters: $3,800.00

List of Han Chinese individuals

Catalog	ID Sex	Age at Sampling
GM17839	Female	64 YR
GM17840	Female	64 YR
GM17775	Female	63 YR
GM17801	Female	60 YR
GM17817	Female	58 YR
GM17841	Female	57 YR
GM17831	Female	56 YR
GM17847	Female	55 YR
GM17844	Female	54 YR
GM17744	Female	51 YR
GM17779	Female	51 YR
GM17795	Female	51 YR
GM17798	Female	51 YR
GM17808	Female	48 YR
GM17754	Female	47 YR
GM17783	Female	47 YR
GM17800	Female	47 YR

Catalog	ID Sex	Age at Sampling
GM17816	Female	47 YR
GM17838	Female	46 YR
GM17814	Female	45 YR
GM17785	Female	44 YR
GM17799	Female	41 YR
GM17793	Female	40 YR
GM17782	Female	37 YR
GM17792	Female	35 YR
GM17843	Female	30 YR
GM17802	Female	28 YR
GM17815	Female	28 YR
GM17811	Female	27 YR
GM17740	Female	23 YR
GM17780	Female	23 YR
GM17738	Female	22 YR
GM17739	Female	22 YR
GM17756	Female	21 YR

Catalog	ID Sex	Age at Sampling
GM17768	Female	21 YR
GM17803	Female	21 YR
GM17833	Female	21 YR
GM17734	Female	20 YR
GM17747	Female	20 YR
GM17752	Female	20 YR
GM17758	Female	20 YR
GM17766	Female	20 YR
GM17733	Female	19 YR
GM17735	Female	19 YR
GM17745	Female	19 YR
GM17746	Female	19 YR
GM17797	Female	19 YR
GM17741	Female	18 YR
GM17757	Female	18 YR
GM17773	Female	18 YR
GM17824	Male	64 YR
GM17837	Male	62 YR
GM17827	Male	61 YR
GM17796	Male	59 YR
GM17761	Male	58 YR
GM17818	Male	58 YR
GM17743	Male	57 YR
GM17776	Male	57 YR
GM17810	Male	56 YR
GM17789	Male	55 YR
GM17854	Male	53 YR
GM17809	Male	51 YR
GM17812	Male	50 YR
GM17794	Male	49 YR
GM17813	Male	46 YR
GM17826	Male	46 YR
GM17853	Male	46 YR

Catalog	ID Sex	Age at Sampling
GM17804	Male	45 YR
GM17805	Male	45 YR
GM17825	Male	45 YR
GM17771	Male	42 YR
GM17791	Male	42 YR
GM17787	Male	41 YR
GM17828	Male	41 YR
GM17790	Male	35 YR
GM17753	Male	34 YR
GM17823	Male	33 YR
GM17842	Male	29 YR
GM17852	Male	25 YR
GM17742	Male	24 YR
GM17850	Male	24 YR
GM17769	Male	23 YR
GM17855	Male	23 YR
GM17856	Male	23 YR
GM17762	Male	22 YR
GM17770	Male	22 YR
GM17806	Male	22 YR
GM17737	Male	21 YR
GM17835	Male	21 YR
GM17846	Male	21 YR
GM17755	Male	20 YR
GM17764	Male	20 YR
GM17765	Male	20 YR
GM17767	Male	20 YR
GM17807	Male	20 YR
GM17749	Male	19 YR
GM17759	Male	19 YR
GM17774	Male	19 YR
GM17845	Male	19 YR
GM17736	Male	18 YR

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

BIBLIOGRAPHY

Chupong Ittiwut is now Ph.D. student in the interdisciplinary program in Biomedical Sciences, Chulalongkorn University. He was born in Chiangrai, Thailand on January 11, 1978. After his elementary and high school educations at Bansonkong (Chiangrai Charoon Rat) school and Samakkhi Wittayakhom school, he left for Chulalongkorn University in Bangkok, where he received his B.Sc. in Genetics and M.Sc. in Medical Science (Molecular Biology and Genetics) in 1998 and 2002 respectively.

While studying Master's (in 1998), he received scholarship from the Ministry of University Affairs to hold academic position in Department of Biology, Faculty of Science, Mahasarakham University, Mahasarakham, after graduation. He had been there 2 years and 4 months as a lecturer and university staff before going back to Chulalongkorn University again for his Ph.D. in June 2004. Supervised by Professor Apiwat Mutirangura, M.D., Ph.D., Chupong had a great opportunity to go abroad for doing the 3-year research project in genetics of addiction at Department of Psychiatry, Yale University, Connecticut, USA, under supervision of Professor Joel Gelernter and supported by the Thai-US Drug Dependence Training Grant.

After finishing doctoral qualifying examination (Monday, February 18, 2008) and thesis proposal defense (Friday, February 29th), he finished his thesis defense on Tuesday April 29th and will be graduated by the end of May of the same year.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย