

**EVALUATION OF FACTORS AFFECTING GENETIC  
ASSOCIATION STUDY USING SNP-BASED ANALYSIS IN THAI  
POPULATION**



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY  
(CLINICAL PATHOLOGY)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY  
2008**

**COPYRIGHT OF MAHIDOL UNIVERSITY**

การประเมินผลของปัจจัยที่ส่งผลกระทบต่อการศึกษาความสัมพันธ์ทางพันธุกรรมโดยใช้การวิเคราะห์ด้วยความหลากหลายของลำดับเบสเดี่ยวในประชากรไทย

(EVALUATION OF FACTORS AFFECTING GENETIC ASSOCIATION STUDY USING SNP-BASED ANALYSIS IN THAI POPULATION)

สมหญิง พรหมโส 4737430 RACP/D

ปร.ค. (พยาธิวิทยาคลินิก)

คณะกรรมการควบคุมวิทยานิพนธ์ : วสันต์ จันทราทิตย์, Ph.D., ชัยชัย สุระ, M.D., วัลยา จงเจริญประเสริฐ, M.D., จินตนา โตเจริญธนาผล, Ph.D.

บทคัดย่อ

การศึกษานี้ได้ทำการพิจารณาถึงปัจจัยที่ส่งผลกระทบต่อการศึกษาความสัมพันธ์ทางพันธุกรรม โดยเฉพาะอย่างยิ่งในประชากรไทย ซึ่งการประเมินถึงปัจจัยเหล่านี้ไม่ว่าจะเป็น วิธีการจีโนไทป์ ความแตกต่างของเชื้อชาติ ความเป็นชั้นภูมิของประชากร และการเรียงลำดับความสำคัญของสโนิปส์ คณะผู้วิจัยได้ทำการทดลองในกลุ่มตัวอย่างสองกลุ่ม คือ กลุ่ม 32 คนไทย ซึ่งทำการถอดรหัสด้วยวิธีการ resequencing ของ 13 ยีนที่สัมพันธ์กับโรคหลอดเลือดหัวใจ และ กลุ่ม 280 คนไทย ที่เก็บตัวอย่างจากประชากรสี่ภาคของประเทศไทย คือ ภาคเหนือ อีสาน กลาง และได้ ทำการจีโนไทป์ ใน 1,536 ของลำดับเบสเดี่ยวใน 188 ยีนที่สัมพันธ์กับยา

ผลการทดลองพบว่าประมาณ 30% ของลำดับเบสเดี่ยวทั้งหมด 194 ตำแหน่งที่พบด้วยวิธีการ resequencing เป็นลำดับเบสเดี่ยวที่เพิ่งค้นพบใหม่ในการศึกษานี้ ลำดับเบสเดี่ยวที่เพิ่งค้นพบใหม่นี้ส่วนใหญ่อยู่นอกบริเวณขอบเขตแฮปโลไทป์ที่กำหนดโดยการคำนวณของลำดับเบสเดี่ยวที่เคยพบก่อนหน้านี้ หรือที่แสดงผลในฐานข้อมูลสโนิปส์ เพียง 5.77% ของลำดับเบสเดี่ยวที่เพิ่งค้นพบใหม่นี้อยู่ในบริเวณดังกล่าว การเลือกตัวแทนของลำดับเบสเดี่ยวที่ใช้ในการจีโนไทป์แสดงให้เห็นว่า 8 ใน 13 ยีนจะได้ประโยชน์มากขึ้นเมื่อทำการเลือกจากการใช้ข้อมูลของลำดับเบสเดี่ยวที่เพิ่งค้นพบเข้ามาวิเคราะห์ด้วย โดยเฉพาะอย่างยิ่งเมื่อมีการค้นพบเพียงหนึ่งตำแหน่งของลำดับเบสเดี่ยวที่เพิ่งค้นพบใหม่ การเลือกตัวแทนลำดับเบสเดี่ยวก็มีการเปลี่ยนแปลง นอกจากนี้ผลการทดลองพบว่าไม่มีมาร์คเกอร์ใดให้ผลแตกต่างอย่างมีนัยสำคัญระหว่างประชากรสี่ภาคของประเทศไทย ในขณะที่ 3 ลำดับเบสเดี่ยว ได้แก่ rs4147930 rs2285526 และ rs4148230 ถูกพบด้วยค่าความถี่มากกว่า 0.10 และพบเพียงในประชากรไทยเมื่อเทียบกับประชากรของ Hapmap ค่าสถิติ F แสดงให้เห็นหลักฐานบางอย่างถึงการมี population substructure อย่างมีนัยสำคัญของประชากรสี่ภาค จุดที่แตกต่างมากที่สุด คือ บริเวณระหว่างภาคใต้-ภาคเหนือ ( $F_{ST}: 6.09 \times 10^{-3}$ ) และ ภาคใต้-ภาคอีสาน ( $F_{ST}: 8.51 \times 10^{-3}$ ) ความแตกต่างนี้ถูกสนับสนุนด้วยค่าการวิเคราะห์ด้วย genomic control และ structured association ค่าแลมดาอยู่ระหว่าง 0.950 – 1.658 ภาคกลางเป็นภาคที่มีการกระจายข้อมูลและรวมเข้าไปในกลุ่มประชากรของภาคที่เหลือ

โดยสรุปการค้นพบนี้ไม่เพียงแต่แสดงให้เห็นถึงความสำคัญของความสัมพันธ์ทางพันธุกรรมในระหว่างกลุ่มประชากรทางพันธุศาสตร์ และ ยังคงแสดงให้เห็นว่าวิธีการจีโนไทป์ลำดับเบสเดี่ยวด้วยวิธี resequencing มีความสำคัญในการให้ข้อมูลเพิ่มสมบูรณ์ขึ้นสำหรับการศึกษาความสัมพันธ์ทางพันธุกรรมต่อไปในอนาคต โดยเฉพาะอย่างยิ่งในกลุ่มประชากรไทย ข้อมูลความถี่ของลำดับเบสเดี่ยวที่พบในการศึกษานี้ เปิดให้เข้าถึงสืบค้นได้ที่ ฐานข้อมูลไทยสโนิปส์ (<http://thaisnp.biotech.or.th/>) และ ฐานข้อมูลโครงการเภสัชพันธุศาสตร์ (<http://www.pharmagtc.net/>)

**EVALUATION OF FACTORS AFFECTING GENETIC ASSOCIATION STUDY  
USING SNP-BASED ANALYSIS IN THAI POPULATION****SOMYING PROMSO 4737430 RACP/D****Ph.D. (CLINICAL PATHOLOGY)****THESIS ADVISORS: WASUN CHANTRATITA, Ph.D., THUNYACHAI SURA,  
M.D., WALLAYA JONGJAROENPRASERT, M.D., CHINTANA  
TOCHAROENTANAPHOL, Ph.D.****ABSTRACT**

Several factors affecting the study of genetic association, particularly in the Thai population were surveyed in this study. To evaluate these factors (i.e. genotyping technology, ethnic difference, population stratification, and Single Nucleotide Polymorphism (SNP) prioritization), two sample groups were examined: (i) 32 healthy Thai individuals were resequenced in exon flanking regions of 13 genes related to atherosclerosis, and (ii) 280 unrelated healthy Thai samples were recruited from each of four geographical Thai populations, North, Northeast, Central, and South, and were genotyped in 1,536 known polymorphic sites of 188 drug related genes. Around 30% of a total of 194 identified polymorphisms were characterized as novel polymorphisms. Most of the novel SNPs found were located outside haplotype blocks generated by known SNPs in the Thai population. Only 5.77% of the novel SNPs lay in these defined haplotype blocks. The selection of haplotype tagging SNPs shows that 8 out of 13 genes benefit from the ethnic specific genotype information. When at least one novel SNP is present, the chosen tagging SNPs are altered. None of the markers showed significant differences in any of the four subpopulations of Thailand, while three SNPs, rs4147930, rs2285526, and rs4148230, were found in the Thai population with a MAF > 0.10, but not in the other Hapmap populations. F-statistics showed some evidence for significant population substructure in the four geographical areas. The biggest difference was between the South and North ( $F_{ST}: 6.09 \times 10^{-3}$ ) as well as South and Northeast ( $F_{ST}: 8.51 \times 10^{-3}$ ) regions. This difference was also supported by genomic control and structured association analysis. The Lambda inflation factors were between 0.950 – 1.658 among the Thai pairwise population. The genetic structure of the Central Thai population had many features in common with the other three populations. These findings provide not only important insights into the genetic relationships between geographical groups, but also demonstrate that the SNP re-sequencing approach provides more complete information for further genetic association studies, particularly in the Thai population. The marker information and allele frequencies of this study are available at the ThaiSNP database (<http://thaisnp.biotec.or.th/>) and the ThaiPGx (<http://www.pharmagtc.net/>).

**KEY WORDS: SNP DISCOVERY, POPULATION STRATIFICATION, THAI,  
PHARMACOGENETICS, GENOMIC CONTROL****214 P.**