

วิทยานิพนธ์นี้นำเสนอวิธีการจำแนกแบบไม่มีพารามิเตอร์ (Non-parametric Classification Technique) ที่ใช้สำหรับเลือกชุดเครื่องหมายพันธุกรรมแบบสองอัลลีล (Candidate Bi-allelic Genetic Marker) ที่ดีที่สุดที่สามารถอธิบายถึงความเสี่ยงต่อการเป็นโรค (Disease Susceptibility) ได้และสามารถใช้ในการศึกษาอันตรกิริยาระหว่างยีน (Gene-Gene Interaction) หลักการทำงานของวิธีการที่นำเสนอจะทำการแมป (Mapping) ระหว่างแฮปโลไทป์ที่ได้จากการอนุมาน (Inferred Haplotype) กับสถานะของการเป็นโรคหรือไม่เป็นโรค (Case/Control Status) ซึ่งมีการทำงานเป็นวงรอบ ในการหาชุดของ SNP ที่ดีที่สุดจากชุดของ SNP ทั้งหมดที่เป็นไปได้ที่สร้างขึ้นจากวิธีการจัดหมู่ (Combination) แต่ละ SNP ส่วนเกณฑ์ที่ใช้ในการเลือกชุดของ SNP ที่ดีที่สุดนั้นจะใช้ค่าของความถูกต้องการจำแนก (Prediction Accuracy) , ความสามารถในการแพร่ของแฮปโลไทป์จากอันดับต่ำไปสูง (Low-to-High Order Haplotype Propagation Capability) และชุดสั้นป็นน้อยที่สุด ในระหว่างขั้นตอนการหาชุดของ SNP ที่ดีที่สุดนั้น จะมีการสร้างแฮปโลไทป์ที่มีความยาวต่าง ๆ กัน (Variable-Length Haplotypes) ขึ้น ดังนั้นจึงเรียกรวมวิธีได้นำเสนอนี้จากคุณลักษณะดังกล่าวว่า Variable-Length Haplotype Construction for Gene-Gene Interaction (VarHAP) ซึ่ง VarHAP ได้ทำการทดลองเปรียบเทียบกับวิธีอื่น ๆ อีกสองวิธีคือ วิธี Multifactor Dimensionality Reduction (MDR) และ โปรแกรม FAMHAP โดยใช้หลายปัญหาที่มีอันตรกิริยาระหว่างสองยีนเป็นปัญหาในการทดลอง ผลการทดลองนั้นได้แสดงให้เห็นว่า VarHAP เหมาะสมสำหรับทุกสถานการณ์ของปัญหาไม่ว่าจะเป็นในกรณีที่มีความไม่สมดุลในการเชื่อมโยงอย่างอ่อน (Weak Linkage Disequilibrium) หรือมีความไม่สมดุลในการเชื่อมโยงกันอย่างแข็งแรง (Strong Linkage Disequilibrium) ท่ามกลางข้อมูล SNP ที่นำมาศึกษา

## Abstract

200594

This thesis presents a non-parametric classification technique for identifying a candidate bi-allelic genetic marker set that best describes disease susceptibility in gene-gene interaction studies. The developed technique functions by creating a mapping between inferred haplotypes and case/control status. The technique cycles through all possible marker combination models generated from the available marker set where the best interaction model is determined from multiple criteria including prediction accuracy, low-to-high order haplotype propagation capability and model parsimony. Since variable-length haplotypes are created during the best model identification, the developed technique is referred to as a variable-length haplotype construction for gene-gene interaction (VarHAP) technique. VarHAP has been benchmarked against a multifactor dimensionality reduction (MDR) program and a haplotype interaction technique embedded in a FAMHAP program in various two locus interaction problems. The results reveal that VarHAP is suitable for all interaction situations with the presence of weak and strong linkage disequilibrium among genetic markers.