

**ชื่อเรื่องวิทยานิพนธ์** การตรวจหาชิ้นที่ถูกควบคุมร่วมจากลำดับดีเอ็นเอและระดับการแสดงออกของยีน โดยอาศัยฮิดเดนมาร์คอฟโมเดล

**ผู้เขียน** นาย ภาสกร ใจเจริญวิญญู

**ปริญญา** วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

**อาจารย์ที่ปรึกษาวิทยานิพนธ์** ผู้ช่วยศาสตราจารย์ ดร.จิรยุทธ ไชยจาวรณิช

### บทคัดย่อ

งานวิจัยฉบับนี้ มีวัตถุประสงค์เพื่อนำเสนอกระบวนการตรวจสอบกลุ่มของยีนที่ถูกควบคุมร่วม โดยขั้นตอนแรกจะใช้วิธีการวิเคราะห์ข้อมูลระดับการแสดงออกของยีนด้วยฮิดเดนมาร์คอฟโมเดลซึ่งเป็นส่วนสำคัญหลักของงานวิจัย หลังจากนั้นจึงนำผลการวิเคราะห์ข้อมูลที่ได้ไปทำการวิเคราะห์ลำดับดีเอ็นเอเป็นขั้นตอนสุดท้าย ในขั้นตอนแรกนั้นจะใช้ฮิดเดนมาร์คอฟโมเดลในกระบวนการหาความสัมพันธ์ระหว่างยีน การจัดกลุ่มยีน ซึ่งจะนำไปสู่การแบ่งกลุ่มยีนเพื่อตรวจหา กลุ่มของยีนที่ถูกควบคุมร่วมในที่สุด นอกจากนี้ยังได้มีการนำเสนอเทคนิคต่างๆเพื่อช่วยให้การวิเคราะห์มีประสิทธิภาพสูงขึ้น อันได้แก่ การนำความรู้ทางชีววิทยาไปฝังไว้ในโมเดล การเลือกข้อมูลมาวิเคราะห์โดยใช้รหัสโมเดล และการใช้หลักการของเบย์เซียนอินฟอร์เมชันในการสร้างโมเดล จากผลการทดลองแสดงให้เห็นได้ว่าวิธีการวิเคราะห์ข้อมูลระดับการแสดงออกของยีนนั้น ให้ผลการแบ่งกลุ่มที่ดีและค่อนข้างสอดคล้องกับรายงานทางชีววิทยา เมื่อนำเอากลุ่มของยีนที่ถูกควบคุมร่วมกันดังกล่าวไปทำการวิเคราะห์ในขั้นตอนสุดท้ายโดยใช้ RSATools ก็สามารถตรวจสอบพบ โอลิโกนิวคลีโอไทด์ร่วมในส่วนหน้าของยีนทุกกลุ่ม และคาดว่าจะเป็นการดำเนินงานที่ ยีนควบคุมร่วมมาเกาะ ซึ่งสามารถนำไปใช้เป็นข้อมูลในการวิเคราะห์เชิงลึกทางชีววิทยาได้ในอนาคต

<b>Thesis Title</b>	Detection of Co-regulated Genes from DNA Sequences and Gene Expression Levels using Hidden Markov Model.
<b>Author</b>	Mr. Phasit Charoenkwan
<b>Degree</b>	Master of Science (Computer Science)
<b>Thesis Advisor</b>	Assistant Professor Dr. Jeerayut Chaijaruwanich

### ABSTRACT

The objective of this study was to propose a strategy for detecting co-regulated genes. First, gene expression data analysis using hidden Markov model was implemented as the main part of this research. Then, the obtained results were used in DNA sequences analysis. The hidden Markov model was used in the process of genes correlation prediction and genes classification, leading to genes clustering for the detection of co-regulated genes. Furthermore, various techniques such as embedding the biological knowledge to the model, selecting significantly expressed genes using the null model and using the Bayesian information criterion to construct the model were introduced for better efficiency in the analysis of gene expression data. Moreover, these experimental results provided gene clusters corresponding to existing biological knowledge. Therefore, these clusters of genes presumably co-regulated by same regulators were finally sent to RSATools for upstream DNA sequences analysis. The common oligo-nucleotides were found in all groups of genes and were expected to be the common binding sites for regulator genes, possibly leading to further rigorous biological research.