

# บทที่ 1

## บทนำ

### 1.1 หลักการ ทฤษฎี เหตุผล และสมมุติฐาน

เนื่องจากในปัจจุบันนี้ความรู้ทางด้านเทคโนโลยีชีวภาพได้พัฒนาไปอย่างรวดเร็ว ทำให้เราได้รับความรู้ต่าง ๆ อย่างมากมาย รวมถึงได้รับข้อมูลทางชีวภาพต่าง ๆ ทั้งของมนุษย์ พืช และสัตว์ ไม่ว่าจะเป็นข้อมูลการแสดงออกของยีน (Gene Expression) หรือข้อมูลลำดับของดีเอ็นเอ (DNA Sequence) ซึ่งมนุษย์ได้พยายามที่จะนำข้อมูลทางชีวภาพต่าง ๆ เหล่านี้มาทำการศึกษาวิเคราะห์ เพื่อให้สามารถเข้าใจกลไกการทำงานต่าง ๆ ของร่างกายซึ่งถูกควบคุม โดยการทำงานของยีนต่าง ๆ ร่วมกัน ซึ่งหากเราสามารถเข้าใจถึงกลไกการทำงานต่าง ๆ ของยีนแล้ว ก็อาจจะทำให้เราเข้าใจถึงพัฒนาการของร่างกายและอาจสามารถทำการป้องกันโรคต่าง ๆ ที่มีสาเหตุมาจากความผิดปกติของยีนได้ หรือเพื่อทำการศึกษาเกี่ยวกับข้อมูลการแสดงออกของยีนพืชเพื่อทำการพัฒนาสายพันธุ์หรือเพื่อป้องกันโรคที่เกิดกับพืชได้ เป็นต้น

สำหรับการวิเคราะห์ข้อมูลการแสดงออกของยีนหรือข้อมูลดีเอ็นเอไมโครอาร์เรย์นั้น ได้มีการศึกษาวิจัยกันในหลายด้าน เช่น การนำข้อมูลการแสดงออกของยีนมาทำการจัดกลุ่ม (Gene Clustering) การทำนายกลุ่มหน้าที่ของยีนจากการแสดงออกของยีน (Gene Classification) และการหาเครือข่ายการควบคุมกันระหว่างยีน (Gene Regulatory Network) ซึ่งเราจะพบว่าการศึกษาเครือข่ายการควบคุมกันระหว่างยีนนี้เป็นสิ่งที่มีสำคัญต่อการศึกษาวิจัยทางด้านชีวภาพ การเข้าใจกลไกการควบคุมกันระหว่างยีนอาจนำไปสู่การเข้าใจกระบวนการต่าง ๆ ภายในเซลล์ที่มีความซับซ้อนได้ โดยลักษณะต่าง ๆ ของสิ่งมีชีวิตที่แสดงออกมานั้นอาจเกิดมาจากการแสดงออกร่วมกันของยีนต่างๆ โดยที่การแสดงออกของยีนหนึ่ง ๆ อาจถูกควบคุมจากยีนอื่นหรือปัจจัยอื่น (Regulator) เช่น โปรตีนทรานสคริปต์ชันแฟกเตอร์ (Transcriptional factor: TF) เป็นต้น เพื่อให้เซลล์ต่าง ๆ สามารถทำงานได้อย่างเป็นปกติ แม้เมื่อมีความผิดปกติเกิดขึ้นที่ยีนใด ๆ ก็จะทำให้สามารถรู้ได้ว่ายีนนี้จะส่งผลกระทบต่อการทำงานของยีนใดบ้างที่ถูกยีนที่ผิดปกตินี้ไปควบคุมการทำงานอยู่ ซึ่งอาจทำให้เราสามารถคิดหาวิธีที่จะทำการป้องกันหรือทำการรักษาได้ หรือในบางครั้งโรคต่าง ๆ ที่เกิดมาจากสาเหตุทางพันธุกรรม การควบคุมกันระหว่างยีนอาจเข้ามามีบทบาทสำคัญในการพิจารณาว่ายีนใดเป็นยีนที่เข้ามาควบคุม (Regulator Gene) ยีนอื่น ๆ จนทำให้เกิดอาการผิดปกตินั้น ๆ ขึ้น ซึ่งเราสามารถยับยั้งหรือป้องกันโรคโดยยับยั้งการทำงานของยีนที่เป็นตัวควบคุมนั้นได้

ในการนำข้อมูลการแสดงผลของยีนมาทำการศึกษาวเคราะห์เพื่อทำการหาการควบคุมกันระหว่างยีนโดยทั่วไปนั้นจะอาศัยวิธีการของเครือข่ายเบย์เซียน (Bayesian network) โดยทำการวิเคราะห์จากข้อมูลการแสดงผลของยีนโดยตรง ซึ่งทำให้บางครั้งเราอาจพบว่าข้อมูลที่นำมาศึกษาอาจมีจำนวนน้อยเกินไปอาจไม่เพียงพอที่จะสามารถทำการสรุปผลลัพธ์ได้อย่างถูกต้อง ดังนั้นจึงต้องมีการนำเทคนิควิธีการที่สามารถจำลองข้อมูลขึ้นมาจากข้อมูลที่เราสงสัยใจที่มีอยู่แล้ว (Observation data) ให้มีจำนวนที่มากขึ้นเป็นข้อมูลของประชากร และทำการหาความสัมพันธ์ระหว่างข้อมูลกับปัจจัยหรือที่เรียกว่า พารามิเตอร์ (parameter) ต่าง ๆ ที่ส่งผลต่อการแสดงผลของข้อมูล เพื่อให้สามารถทำการศึกษาค้นคว้าได้ผลลัพธ์ที่มีความถูกต้องหรือน่าเชื่อถือมากยิ่งขึ้น ซึ่งวิธีการหนึ่งที่ถูกนำมาช่วยในการแก้ปัญหานี้ก็คือ เทคนิควิธีการมาร์คอฟเชนมอนติคาร์โล (Markov Chain Monte Carlo)

ดังนั้นในงานวิจัยฉบับนี้จึงได้นำเสนอและประยุกต์ใช้วิธีการเบย์เซียนและมาร์คอฟเชนมอนติคาร์โล มาทำการวิเคราะห์ข้อมูลทางชีววิทยาซึ่งเป็นข้อมูลการแสดงผลของยีนหรือดีเอ็นเอไมโครอาร์เรย์ เพื่อทำการจำลองความสัมพันธ์ระหว่างยีนและปัจจัยที่มีผลกระทบต่อการแสดงผลของยีน จากนั้นจึงทำการสร้างแบบจำลองการควบคุมกันระหว่างยีน โดยอาศัยเทคนิควิธีการมาร์คอฟเชนมอนติคาร์โลเข้ามาช่วยในการสร้างแบบจำลองเบย์เซียน โดยหลักการและทฤษฎีต่างๆที่ใช้ในการทำงานวิจัยชิ้นนี้จะอธิบายไว้ในบทที่ 2 จากนั้นจึงเข้าสู่ส่วนของการทดลองเพื่อทำการสร้างแบบจำลองของข้อมูลซึ่งแบ่งออกเป็นสองการทดลอง คือ บทที่ 3 เป็นการจำลองการศึกษาการส่งผลกระทบของปริมาณความเข้มข้นกลูโคสและความหนาแน่นเซลล์ต่อระดับการแสดงผลของยีนในยีสต์สายพันธุ์ *แซคคาโรไมซีส เซรีวิซิเย* และบทที่ 4 เป็นการอนุมานแบบจำลองการควบคุมกันระหว่างยีนด้วยเครือข่ายเบย์เซียน โดยอาศัยวิธีการมาร์คอฟเชนมอนติคาร์โล และในส่วนสุดท้ายคือบทที่ 5 ซึ่งเป็นบทสรุปภาพรวมของงานวิจัยทั้งหมดรวมทั้งข้อเสนอแนะและแนวทางในการทำงานวิจัยต่อไปในอนาคต

## 1.2 วัตถุประสงค์ของการวิจัย

ในการศึกษาและวิจัยครั้งนี้เพื่อทำการวิเคราะห์ข้อมูลดีเอ็นเอไมโครอาร์เรย์โดยอาศัยวิธีการมาร์คอฟเชนมอนติคาร์โล โดยแบ่งวัตถุประสงค์ของวิทยานิพนธ์ออกเป็น 2 ส่วน ดังนี้คือ

1.2.1 นำเสนอวิธีการในการวิเคราะห์ข้อมูลดีเอ็นเอไมโครอาร์เรย์ด้วยวิธีการเบย์เซียนและมาร์คอฟเชนมอนติคาร์โล

1.2.2 สร้างแบบจำลองการควบคุมกันระหว่างยีน จากข้อมูลดีเอ็นเอไมโครอาร์เรย์ด้วยเทคนิควิธีการมาร์คอฟเชนมอนติคาร์โล

### 1.3 ประโยชน์ที่จะได้รับจากการศึกษาเชิงทฤษฎี และ/หรือ เชิงประยุกต์

สำหรับประโยชน์ที่คาดว่าจะได้รับจากการวิเคราะห์ข้อมูลจีโนมไมโครอาร์เรย์โดยอาศัยวิธีการมาร์คอฟเชนมอนติคาร์โลในงานวิจัยครั้งนี้ มีดังต่อไปนี้คือ

- 1.3.1 เป็นการประยุกต์ใช้หลักการของมาร์คอฟเชนมอนติคาร์โล เพื่อทำการวิเคราะห์ข้อมูลจีโนมไมโครอาร์เรย์ของยีน
- 1.3.2 ได้เข้าใจในพฤติกรรมการแสดงออกของยีนมากขึ้น
- 1.3.2 ได้แบบจำลองการควบคุมกันของยีนที่มีลักษณะเป็นแบบเครือข่าย
- 1.3.3 ได้รับความรู้เกี่ยวกับวิธีการในการสร้างข้อมูลขึ้นมาจากข้อมูลที่เราสนใจ แต่ข้อมูลนั้นมีอยู่เป็นจำนวนมากต่อการนำไปทำการวิเคราะห์ซึ่งอาจทำให้ได้ผลลัพธ์ที่ไม่ถูกต้อง

### 1.4 ขอบเขตการวิจัย

สร้างแบบจำลองการควบคุมกันระหว่างยีนจากข้อมูลจีโนมไมโครอาร์เรย์ ด้วยวิธีการมาร์คอฟเชนมอนติคาร์โล โดยข้อมูลที่น่ามาทำการวิเคราะห์นั้นเป็นข้อมูลจีโนมไมโครอาร์เรย์ของยีสต์สายพันธุ์ *Saccharomyces Cerevisiae* ซึ่งเป็นข้อมูลจากฐานข้อมูลเอสเอ็มดี (SMD: Stanford Microarray Database) (available at <http://genome-www5.stanford.edu>)

### 1.5 วิธีการวิจัย

ในการศึกษาวิจัยครั้งนี้ได้มีการกำหนดแผนงานวิจัย โดยเริ่มต้นจากการศึกษาข้อมูลทางชีววิทยาก่อน จากนั้นจึงศึกษาทฤษฎีต่างๆ ที่เกี่ยวข้องและเขียน โปรแกรมเพื่อสร้างแบบจำลองการควบคุมกันระหว่างยีน ดังแผนงานวิจัยที่ได้กำหนดไว้ดังต่อไปนี้คือ

- 1.5.1 ศึกษาข้อมูลจีโนมไมโครอาร์เรย์และทฤษฎีต่างๆ ที่เกี่ยวข้อง
- 1.5.2 หาความสัมพันธ์ระหว่างยีนและปัจจัยที่ส่งผลต่อการแสดงออกของยีนในรูปแบบจำลองเชิงกราฟ ด้วยวิธีการมาร์คอฟเชนมอนติคาร์โล
- 1.5.3 ทำการสร้างแบบจำลองของการควบคุมกันระหว่างยีน ด้วยวิธีการมาร์คอฟเชนมอนติคาร์โล
- 1.5.4 ทดสอบ ปรับปรุง และแก้ไขแบบจำลองการควบคุมกันระหว่างยีน โดยอ้างอิงกับความรู้ทางด้านชีววิทยา
- 1.5.6 สรุปผลการวิจัยพร้อมทั้งจัดทำเป็นรายงาน

## 1.6 สถานที่ที่ใช้ในการดำเนินการวิจัยและรวบรวมข้อมูล

สำหรับสถานที่และอุปกรณ์ที่ใช้ในการดำเนินการวิจัยและเก็บรวบรวมข้อมูล รวมทั้งอุปกรณ์และซอฟต์แวร์ (Software) ต่างๆ ที่ใช้ในการวิจัยในครั้งนี้ มีดังนี้คือ

1.6.1 สถานที่และอุปกรณ์ต่างๆ เพื่อใช้ในการดำเนินการวิจัยและรวบรวมข้อมูล รวมทั้งอุปกรณ์ที่ใช้ในงานวิจัย เพื่อให้งานวิจัยในครั้งนี้สำเร็จลงได้ มีดังต่อไปนี้คือ

- (1) ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่
- (2) สำนักหอสมุด มหาวิทยาลัยเชียงใหม่
- (3) ห้องสมุดคณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

1.6.2 อุปกรณ์ที่ใช้ในการวิจัย

- (1) เครื่องคอมพิวเตอร์ส่วนบุคคลจำนวน 1 เครื่อง
- (2) ซอฟต์แวร์ระบบปฏิบัติการ WindowsXP professional
- (3) โปรแกรมภาษา R 2.3.0 เพื่อใช้พัฒนาโปรแกรมที่จะใช้สร้างแบบจำลองการ

ควบคุมกันระหว่างยีน

(4) โปรแกรม WinBUGS 1.4 เพื่อใช้วิเคราะห์ความสัมพันธ์ระหว่างยีนและปัจจัยที่ส่งผลต่อการแสดงออกของยีน