

บทที่ 3 ผลดำเนินงานวิจัย

3.1 การเก็บและรวบรวมข้อมูล (Data collection)

การทดลองนี้จะใช้ข้อมูล 2 ชุด ดังตารางที่ 3.1 ประกอบด้วย 1) ข้อมูลการแสดงออกของยีนในสภาวะ diauxic shift ของ *Saccharomyces cerevisiae* [15, 16] จากฐานข้อมูล Stanford Microarray (<http://cmgm.stanford.edu/pbrown/explore/array.txt>) เป็นข้อมูลทดสอบสำหรับการวิเคราะห์เริ่มต้นและการพัฒนาโปรแกรมสร้างเครือข่ายการแสดงออกของยีนแบบมีเงื่อนไข (test data) และ 2) ข้อมูลการแสดงออกภายใต้สภาวะ diurnal cycle ของ *Arabidopsis* [14] จากฐานข้อมูล NASC's International Affymetrix Service (NASCArrays) (<http://affymetrix.arabidopsis.info/>) ใช้เป็น case study สำหรับสร้างและศึกษาเครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสังเคราะห์แป้ง

ตารางที่ 3.1 ข้อมูลการแสดงออกของยีนที่ใช้ในงานวิจัยนี้

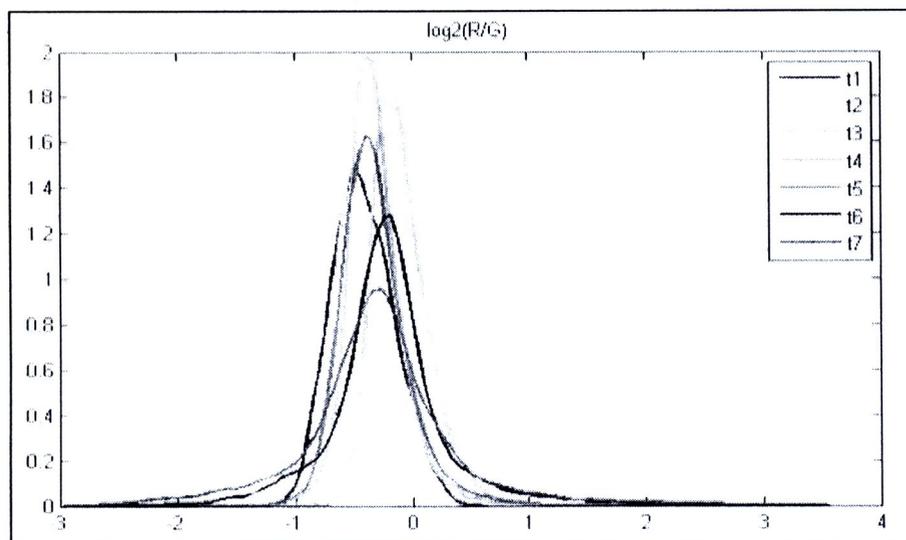
| ข้อมูล | สิ่งมีชีวิต | เทคโนโลยีไมโครอะเรย์ | จำนวนยีนในชุดข้อมูล | Database |
|------------|----------------------|----------------------|---------------------|---|
| Test data | <i>S. cerevisiae</i> | cDNA microarray | 6,153 ยีน | Stanford Microarray Database http://cmgm.stanford.edu/pbrown/explore/array.txt |
| Case study | <i>A. thalian</i> | Affymetrix GeneChips | 22,810 ยีน | NASCArrays http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid=60 |

ข้อมูลการแสดงออกของยีนของ test data เป็นข้อมูลการศึกษาการแสดงออกของยีน 6,153 ยีนภายใต้สภาวะ diauxic shift คือ การเปลี่ยนกระบวนการหมักเป็นการหายใจแบบใช้ออกซิเจนใน 7 ช่วงเวลา (time points; T) คือ ณ เวลาเริ่มต้น (0 ชั่วโมง; T_1) 9.5 ชั่วโมง (T_2), 12 ชั่วโมง (T_3), 13 ชั่วโมง (T_4), 15 ชั่วโมง (T_5), 17 ชั่วโมง (T_6) และ 18.5 ชั่วโมง (T_7) ซึ่งเป็นช่วงเวลาศึกษาการแสดงออกของยีนเมื่อได้รับปริมาณกลูโคสปริมาณ คือ $T_1 = 19$, $T_2 = 18.7$, $T_3 = 17.6$, $T_4 = 14$, $T_5 = 7.5$, $T_6 = 0.2$, และ $T_7 = 0$ g/l Glucose

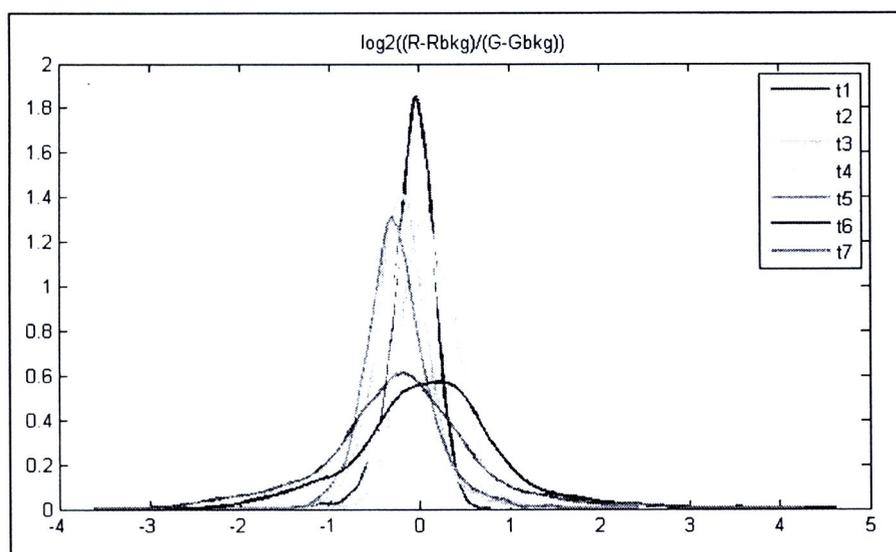
3.2 การเตรียมข้อมูล (Data preprocessing)

ก่อนที่จะนำข้อมูลการแสดงออกของยีนไปใช้นั้น ต้องมีการเตรียมข้อมูลให้เหมาะสม รวมทั้งมีการปรับค่าการแสดงออกของยีนให้อยู่ในมาตรฐานเดียวกัน เรียกว่า normalization เพื่อสามารถนำเอาการแสดงออกของยีนแต่ละยีนมาเปรียบเทียบกันได้ โดยเตรียมข้อมูล galactose pathway ให้อยู่ในรูปแบบ \log_2 ratio และพิจารณาการกระจายตัวของข้อมูลการแสดงออกของยีนในรูปแบบ \log_2 ratio (รูปที่ 3.1) ว่ามีการกระจายตัวของข้อมูลเป็นแบบปกติ (normal distribution) ซึ่งจะต้องมีค่าเฉลี่ยอยู่ที่ 0 แต่จากข้อมูลพบว่าข้อมูลมีการกระจายตัวออกห่าง 0 จึงต้องมีการปรับค่าของข้อมูลให้การกระจายตัวของข้อมูลเข้าใกล้ 0 คือ โดยการหักลบความผิดพลาดเนื่องมาจาก dye bias ด้วยการหักลบค่า background intensity (background correction) ของการแสดงออกของยีนออก (รูปที่ 3.2) อย่างไรก็ตาม จะเห็น

ว่า แม้มีการปรับค่า background ออกแล้ว ยังมีบางช่วงเวลาที่มีการกระจายตัวแตกต่างออกไป ดังนั้น จึงมีการปรับข้อมูล (normalization) ด้วยวิธี lowess เพื่อปรับให้ข้อมูลอยู่มาตรฐานเดียวกัน คือ เข้าใกล้ 0 ให้สามารถเปรียบเทียบข้อมูลการแสดงผลออกของยีนระหว่างช่วงเวลาได้ (รูปที่ 3.3-3.5)



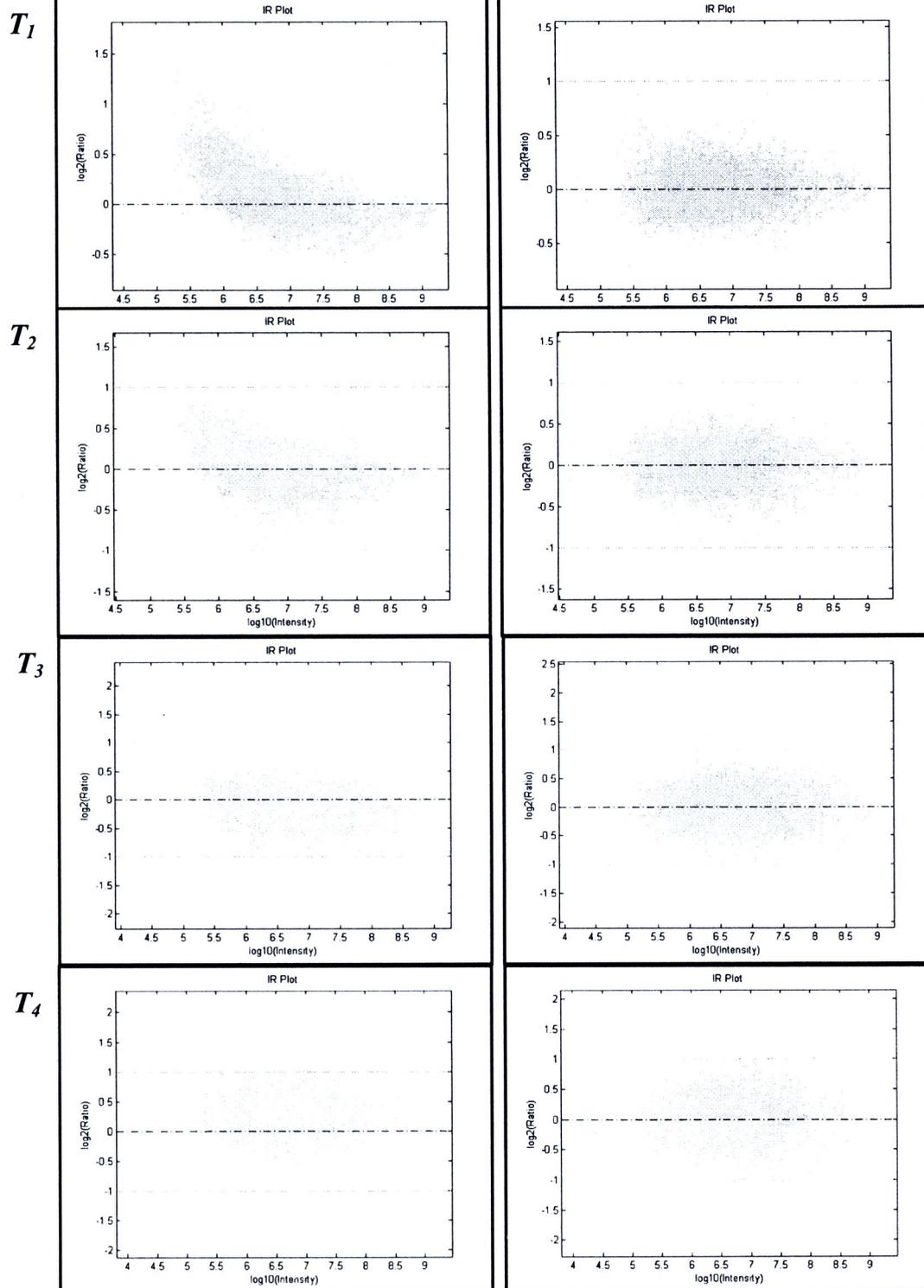
รูปที่ 3.1 การกระจายตัวของข้อมูลการแสดงผลออกของยีนของทุกช่วงเวลา (T_1-T_7) ในรูปแบบ \log_2 ratio



รูปที่ 3.2 การกระจายตัวของข้อมูลการแสดงผลออกของยีนของทุกช่วงเวลา (T_1-T_7) ที่ผ่านการ normalization ด้วย background correction แล้ว

Background correction (before normalization)

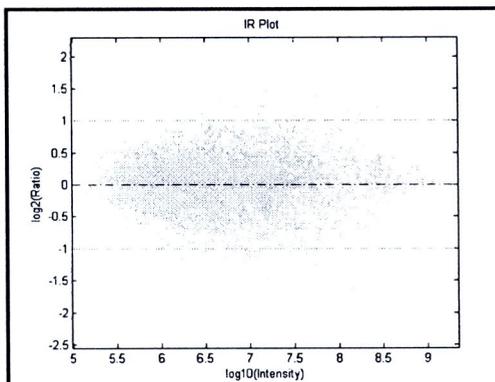
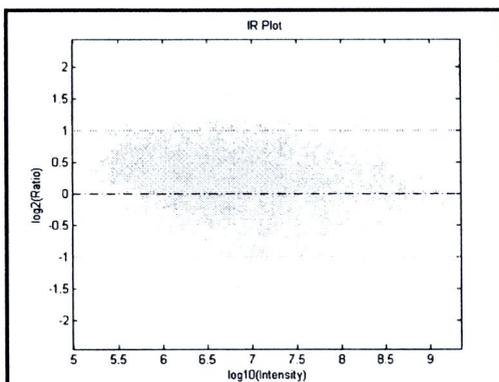
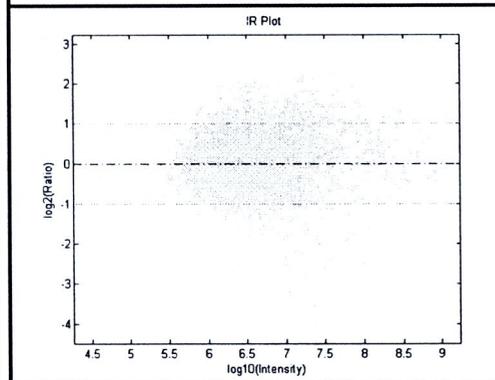
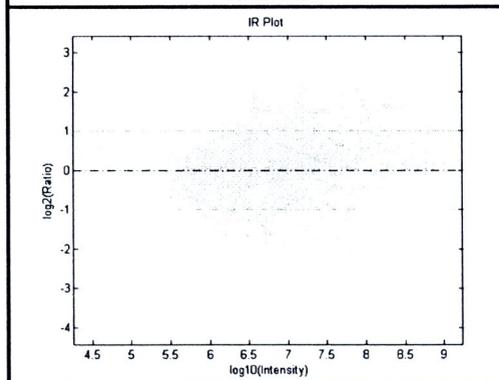
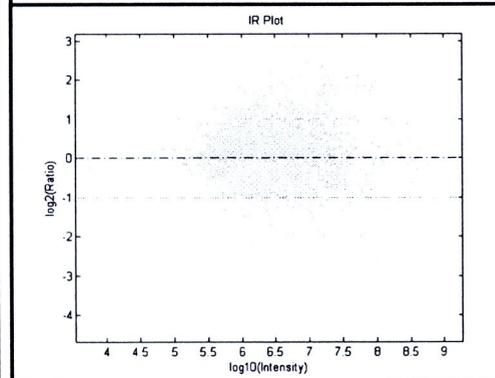
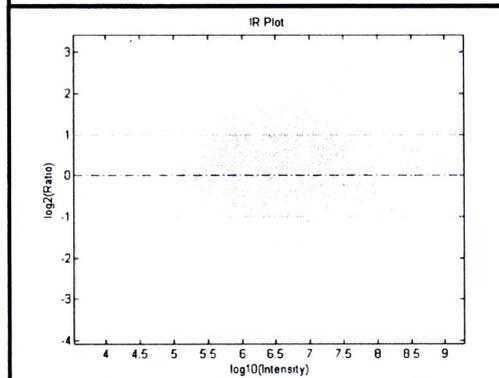
After normalization



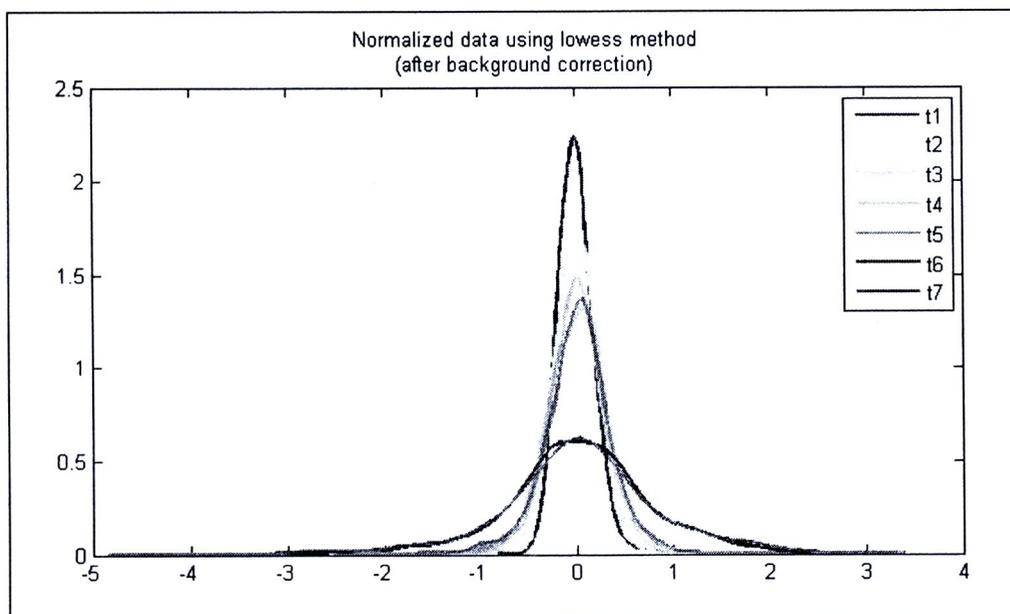
รูปที่ 3.3 เปรียบเทียบ IR plot (ใช้ Bioinformatics Toolbox ของ MATLAB) ของข้อมูลการแสดงออกของยีนแต่ละช่วงเวลา (T_1 - T_4) ในรูปแบบ $\log_2\text{ratio}$ ที่ผ่าน background correction และ ที่ผ่านการ normalization ด้วยวิธี lowess

Background correction (before normalization)

After normalization

 T_5  T_6  T_7 

รูปที่ 3.4 เปรียบเทียบ IR plot (ใช้ Bioinformatics Toolbox ของ MATLAB) ของข้อมูลการแสดงผลออกของยีนแต่ละช่วงเวลา (T_5 - T_7) ในรูปแบบ $\log_2\text{ratio}$ ที่ผ่าน background correction และที่ผ่านการ normalization ด้วยวิธี *lowess*



รูปที่ 3.5 การกระจายตัวของข้อมูลการแสดงออกของยีนของของทุกช่วงเวลา (T_1 - T_7) ที่ผ่านการ normalization ด้วยวิธี *lowess*

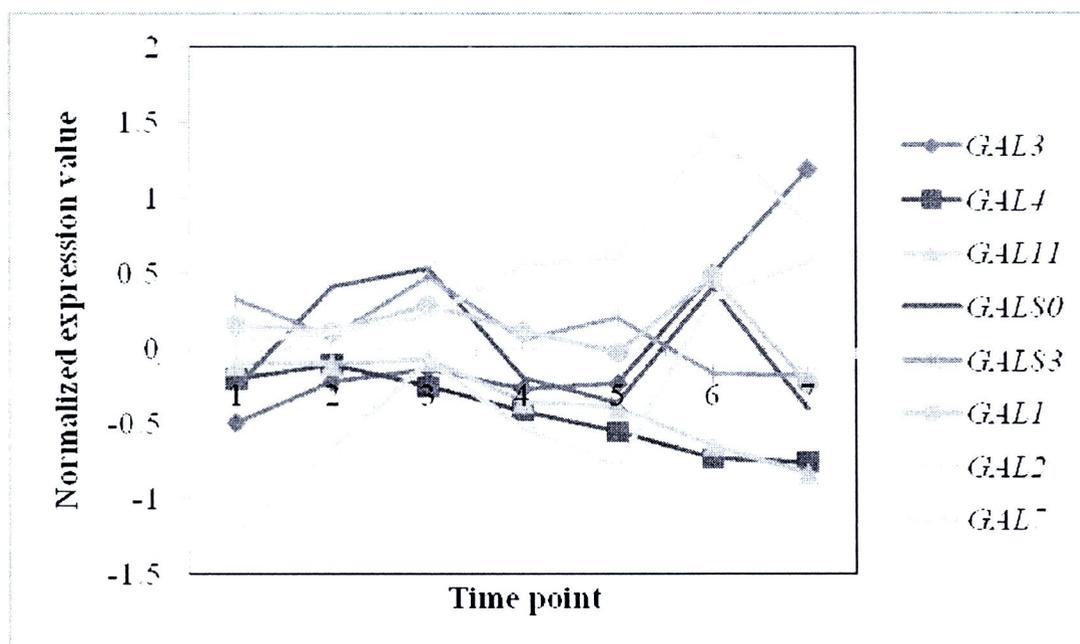
ส่วนข้อมูล Affymetrix GeneChips เป็นข้อมูลการแสดงออกของยีนของพืช *Arabidopsis* ภายใต้อาการกลางวันและกลางคืน มีทั้งหมด 11 ช่วงเวลา (time points; T) คือ ณ เวลาเริ่มต้นในช่วงเวลา (0 ชั่วโมง; T_1) 1 ชั่วโมง (T_2), 2 ชั่วโมง (T_3), 4 ชั่วโมง (T_4), 8 ชั่วโมง (T_5), 12 ชั่วโมง (T_6) 13 ชั่วโมง (T_7), 14 ชั่วโมง (T_8), 16 ชั่วโมง (T_9) 20 ชั่วโมง (T_{10}) และ 24 ชั่วโมง (T_{11}) โดยช่วงเวลา $T_1 - T_5$ คือ ช่วงเวลากลางคืน $T_6 - T_{10}$ คือ ช่วงเวลากลางวัน ส่วน T_6 เป็นช่วงเวลาซ้อนทับช่วงเวลา T_1 เตรียมข้อมูลการแสดงออกของยีนชุดนี้ใช้ภาษา R [<http://www.r-project.org/>] ใน Bioconductor [<http://www.bioconductor.org>] [17, 18] โดยเตรียมข้อมูลการแสดงออกของยีนให้อยู่ในรูปแบบ \log_2 intensity และหาอินที่มีการแสดงออกอย่างมีนัยสำคัญด้วยโปรแกรม EDGE [19] เวอร์ชัน 1.1.175 [<http://faculty.washington.edu/jstorey/edge/>] ก่อนนำไปวิเคราะห์สร้างเครือข่ายควบคุมการแสดงออกของยีน

3.3 แปลงค่าการแสดงออกของยีนและการประเมินวิธีการที่ใช้ในการแปลงข้อมูลการแสดงออกของยีน (Data Discretization)

ข้อมูลที่ผ่านการ Normalization แล้วจะนำมาแปลงค่าการแสดงออกของยีนเป็น 2 ระดับ (binary values) คือ 0 และ 1 ซึ่งหมายถึงไม่มีการแสดงออกของยีน และมีการแสดงออกของยีน ตามลำดับ เพื่อสะดวกต่อการวิเคราะห์ด้วยเทคนิคบูลีนในขั้นตอนต่อไป ขั้นตอนนี้เรียกว่า “discretization” ในงานวิจัยในสนใจการวิเคราะห์การแสดงออกของยีนที่เกี่ยวข้องกับ galactose pathway โดยเลือกเอาเฉพาะกลุ่มยีนที่เกี่ยวข้องกับ galactose pathway (ตารางที่ 3.2) ที่มีข้อมูลของการควบคุมเครือข่ายการแสดงออกของยีน [15] มาวิเคราะห์ต่อไป เพราะจะได้สามารถประเมินประสิทธิภาพของงานวิจัยนี้ได้ ข้อมูลการแสดงออกของกลุ่มยีนดังกล่าว ดังรูปที่ 3.6

ตารางที่ 3.2 ยีนที่เกี่ยวข้องกับ galactose pathway ที่ใช้ในงานวิจัยนี้ [15]

| Gene name | Protein | Enzyme | Function |
|--------------|---------------|-------------------|--|
| <i>GAL2</i> | <i>gal2p</i> | permease | Transports galactose into the cell |
| <i>GAL1</i> | <i>gal1p</i> | Galactokinase | Conversion of intracellular galactose |
| <i>GAL7</i> | <i>gal7p</i> | Uridyltransferase | Conversion of intracellular galactose |
| <i>GAL10</i> | <i>gal10p</i> | Epimerase | Conversion of intracellular galactose |
| <i>GAL4</i> | <i>gal4p</i> | Regulatory gene | Promotes transcription of <i>GAL</i> genes |
| <i>GAL80</i> | <i>gal80p</i> | Regulatory gene | Binds <i>gal4p</i> and inhibits its activity (absent of galactose) |
| <i>GAL3</i> | <i>gal3p</i> | Regulatory gene | Associate with <i>gal80p</i> to release its repression of <i>gal4p</i> |



รูปที่ 3.6 การแสดงออกของกลุ่มยีนที่เกี่ยวข้องกับ galactose pathway ในทุกช่วงเวลา (T_1 - T_7) ที่ผ่านการ normalization ด้วยวิธี *lowess*

วิธีที่ใช้ในการแปลงค่าการแสดงออกของยีนให้เป็น 0 หรือ 1 (discretization method) ในงานวิจัยนี้คือ

- Mean คือ ค่าเฉลี่ยของการแสดงออกของยีนของทุกๆช่วงเวลา ถ้าค่าการแสดงออกของยีนนั้นมีค่ามากกว่า Mean ก็มีค่าเท่ากับ 1 ถ้าน้อยกว่ามีค่าเท่ากับ 0
- Max-x%Max โดย Max คือ ค่าการแสดงออกของยีนแต่ละยีนที่มีค่ามากที่สุดของทุกๆช่วงเวลา (time points), x คือ 10, 20, 30, ..., 90 ถ้าค่าการแสดงออกของยีนนั้นมีค่ามากกว่า Max-x%Max ก็มีค่าเท่ากับ 1 ถ้าน้อยกว่ามีค่าเท่ากับ 0

- Sign of \log_2 ratio คือ ถ้าค่าการแสดงออกของยีนมีค่าน้อยกว่าศูนย์จะให้เท่ากับ 0 ถ้ามีค่ามากกว่าศูนย์ จะให้เท่ากับ 1

จะเห็นว่าเรามีการพิจารณาวิธีที่ใช้แปลงค่าการแสดงออกของยีนหลายวิธี ดังนั้น จึงมีการวิเคราะห์เพื่อหาวิธีที่เหมาะสมที่สุดที่นำมาใช้แปลงค่าการแสดงออกของยีนที่ผ่านการ normalization (raw data) มาเป็นข้อมูล binary values (0 หรือ 1) ก่อนนำไปวิเคราะห์ต่อไปด้วยเทคนิคบูลีน ข้อมูลดิบที่ผ่านการ normalization (raw data) และข้อมูล binary values ของยีนที่เกี่ยวข้องกับ galactose pathway โดยใช้วิธีการ Mean Max-x%Max และ Sign of \log_2 ratio แสดงดังตารางที่ 3.3-3.6 ตามลำดับ

ตารางที่ 3.3 ข้อมูลการแสดงออกของยีนที่เกี่ยวข้องกับ galactose pathway ที่ผ่านกระบวนการ normalization แล้ว

| | T_1 | T_2 | T_3 | T_4 | T_5 | T_6 | T_7 |
|-------|----------|----------|----------|----------|----------|----------|----------|
| GAL3 | -0.49411 | -0.21759 | -0.13606 | -0.26882 | -0.23447 | 0.4957 | 1.18269 |
| GAL4 | -0.20091 | -0.1047 | -0.25154 | -0.41504 | -0.55639 | -0.73697 | -0.76121 |
| GAL11 | -0.1047 | -0.1047 | -0.074 | -0.35845 | -0.39593 | -0.64386 | -0.8365 |
| GAL80 | -0.25154 | 0.41143 | 0.53605 | -0.18442 | -0.35845 | 0.40054 | -0.39593 |
| GAL83 | 0.32193 | 0.08406 | 0.47508 | 0.05658 | 0.20163 | -0.16812 | -0.18442 |
| GAL1 | 0.15056 | 0.11103 | 0.28688 | 0.11103 | -0.02915 | 0.47508 | -0.23447 |
| GAL2 | 0.05658 | 0.11103 | 0.21412 | 0.55582 | 0.60407 | 1.43296 | 0.83996 |
| GAL7 | -1.25154 | -0.66658 | -0.074 | -0.53533 | -0.78588 | 0.34483 | 0.57531 |
| GAL10 | -0.55639 | -0.39593 | 0.66903 | -0.13606 | -0.76121 | 0.64155 | 0.78241 |

ตารางที่ 3.4 ข้อมูล Binary values ของยีนที่เกี่ยวข้องกับ galactose pathway ด้วยวิธี Mean

| | T_1 | T_2 | T_3 | T_4 | T_5 | T_6 | T_7 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| GAL3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| GAL4 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| GAL11 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| GAL80 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| GAL83 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| GAL1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| GAL2 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| GAL7 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| GAL10 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |

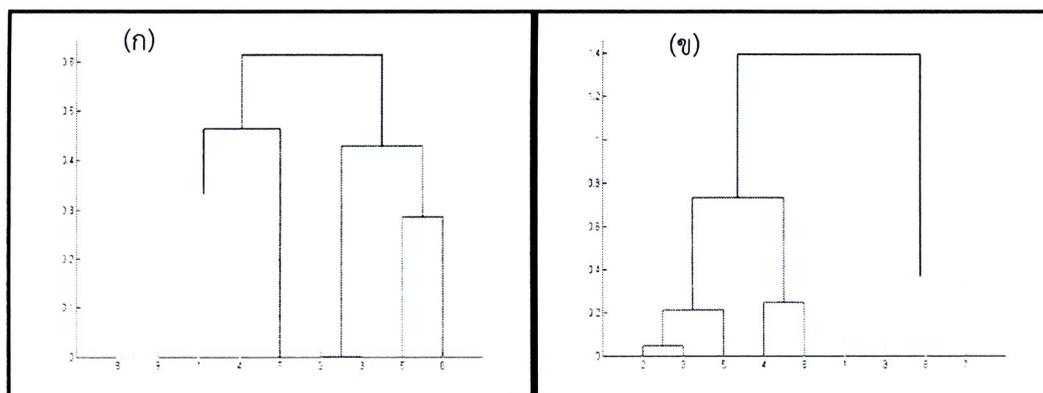
ตารางที่ 3.5 ข้อมูล Binary values ของยีนที่เกี่ยวข้องกับ galactose pathway ด้วยวิธี Max-x%Max

| | T_1 | T_2 | T_3 | T_4 | T_5 | T_6 | T_7 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| GAL3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| GAL4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GAL11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GAL80 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| GAL83 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| GAL1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| GAL2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| GAL7 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| GAL10 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |

ตารางที่ 3.6 ข้อมูล Binary values ของยีนที่เกี่ยวข้องกับ galactose pathway ด้วยวิธี Sign of \log_2 ratio

| | T_1 | T_2 | T_3 | T_4 | T_5 | T_6 | T_7 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| GAL3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| GAL4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GAL11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GAL80 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| GAL83 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| GAL1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| GAL2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| GAL7 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| GAL10 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |

เพื่อให้ได้เครือข่ายการควบคุมการแสดงออกของยีนที่สอดคล้องกับข้อมูลในห้องปฏิบัติการมากที่สุด งานวิจัยนี้ได้การเปรียบเทียบค่าความเหมือน (Similarity score) ของเดนโดแกรม (dendrogram) ของการจัดกลุ่มความเหมือนของกลุ่มยีนที่เลือกมาในชุดข้อมูล raw data และเดนโดแกรมของข้อมูล binary values ว่ามีความเหมือนหรือต่างกันเพียงใด โดยการสร้างเดนโดแกรมด้วยวิธี hierarchical clustering โดยเขียนโปรแกรมเพื่อคำนวณหา similarity score ใน MATLAB [ภาคผนวก ก] ตัวอย่างการเปรียบเทียบความเหมือนของเดนโดแกรมของข้อมูลทั้งสองแบบ โดยมีสมมติฐานว่า discretization method ที่สามารถเปลี่ยนข้อมูลดิบเป็น binary values แล้วทำให้ข้อมูลทั้งสองชุดนี้คล้ายกันด้วยค่า similarity score สูงสุด น่าจะเป็นวิธีการที่เหมาะสมที่สุดสำหรับแปลงค่าการแสดงออกของยีน เพื่อวิเคราะห์ด้วยเทคนิคบูลีนแบบมีเงื่อนไข (Constraint-based Boolean network) เพราะให้ค่าสอดคล้องกับความเป็นจริง (raw data) มากที่สุด ตัวอย่างเดนโดแกรมของ raw data และ binary profiles แสดงในรูปที่ 3.7

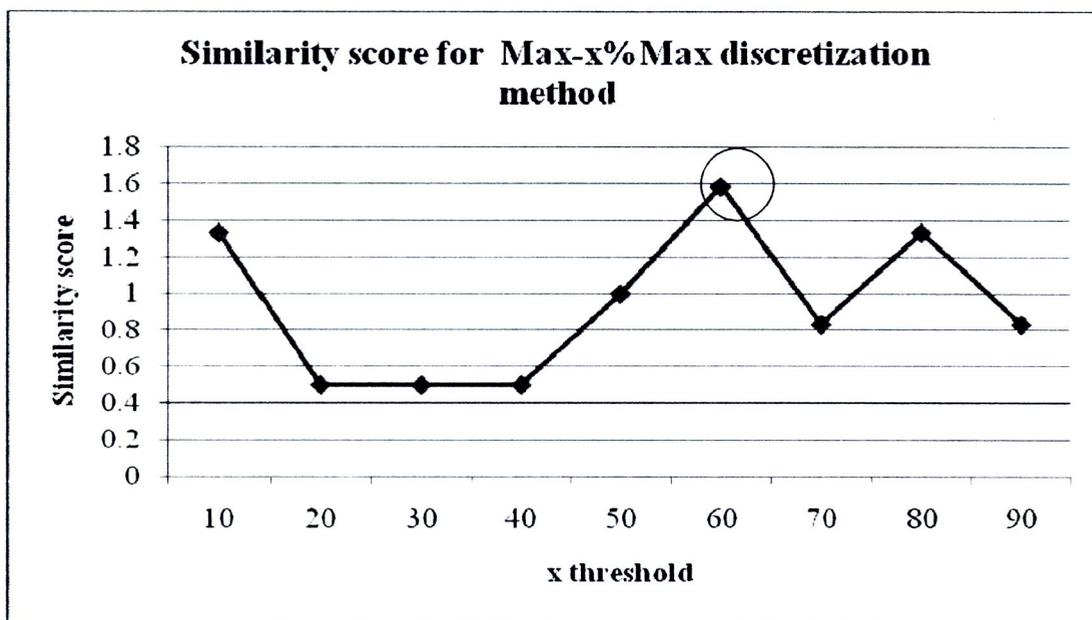


รูปที่ 3.7 การเปรียบเทียบความเหมือนกันระหว่างเดนโดแกรมของข้อมูล (ก) raw data/reference และ (ข) ข้อมูล binary profiles ที่ใช้วิธี Mean มาแปลงค่าการแสดงผลของยีนเป็น 0 หรือ 1

จากรูปที่ 3.7 ลำดับเลขด้านล่างของเดนโดแกรมแต่ละรูปคือ ยีน ตัวอย่างเช่น รูป 3.7ก เมื่อเปรียบเทียบคู่ยีนในข้อมูล raw data พบว่ายีน 2 และ ยีน 3 มีความใกล้เคียงกันมากที่สุด เมื่อเทียบกับคู่ยีนคู่อื่นๆ ก็จะถูกจัดอยู่ในกลุ่มเดียวกันก่อน และไล่จัดกลุ่มต่อไปตามหลักการของ hierarchical clustering เมื่อได้เดนโดแกรมของ raw data ก็จะมาเปรียบเทียบกับเดนโดแกรมของ binary values ที่แปลงโดย discretization method ต่างๆ (ในรูป 3.7 นี้ใช้วิธีการ mean) (รูป 3.7ข) ด้วยค่า similarity score และจากการเปรียบเทียบค่า similarity score พบว่า Max-60%Max นั้นให้ค่า similarity score สูงสุด ดังตารางที่ 3.7 และรูปที่ 3.8 ดังนั้น จึงเลือกวิธี Max-60%Max เพื่อไปแปลงข้อมูลการแสดงผลของยีนเป็น binary values ก่อนการสร้างเครือข่ายควบคุมการแสดงออกของยีน ด้วยเทคนิคบูลีนแบบมีเงื่อนไข (ที่พัฒนาขึ้น) ต่อไป

ตารางที่ 3.7 เปรียบเทียบ Similarity score ของการเปรียบเทียบเดนโดแกรมของ reference/reference กับ reference -binary profiles ที่ได้จาก discretization methods ทั้ง 3 วิธี

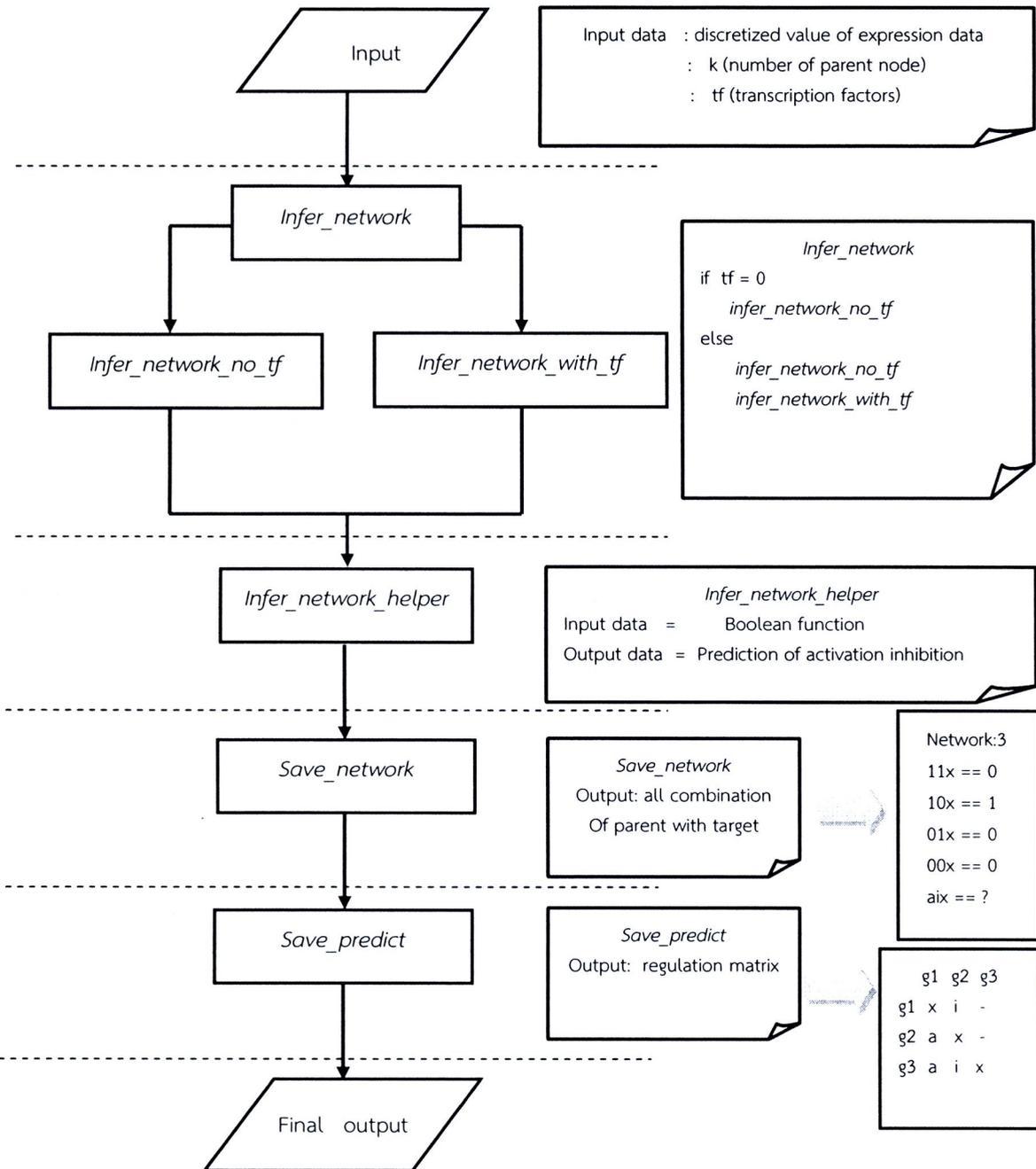
| Comparing dendogram | Similarity score | Normalized similarity score |
|---|------------------|-----------------------------|
| Reference/reference | 2.6167 | 1 |
| Reference/Mean | 1.3333 | 0.51 |
| Reference/Sig of Log ₂ Ratio | 0.8333 | 0.32 |
| Reference/Max-60%Max | 1.5833 | 0.61 |



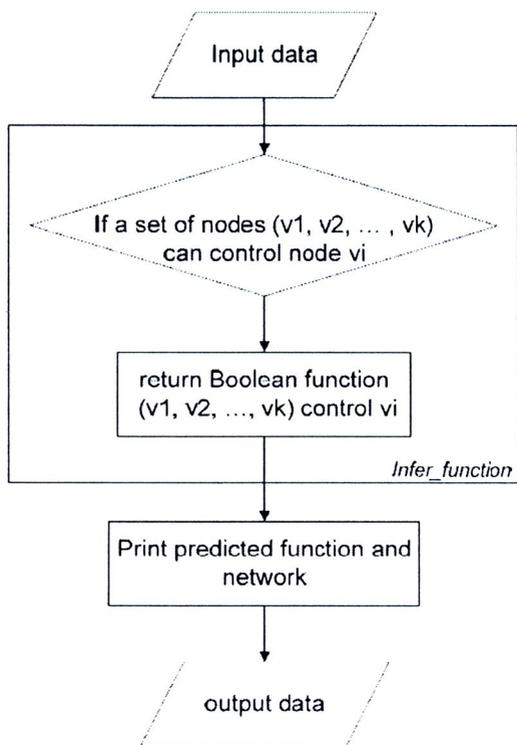
รูปที่ 3.8 Similarity score ของการเปรียบเทียบเดนโดแกรมของ raw data/reference กับ binary values ที่ได้จากวิธีการ Max-x%Max เมื่อ $x = 10, 20, 30, \dots, 90$

3.4 พัฒนาโปรแกรมที่ใช้ในการสร้างเครือข่ายการแสดงออกของยีนแบบมีเงื่อนไข (Constraint-based Boolean network)

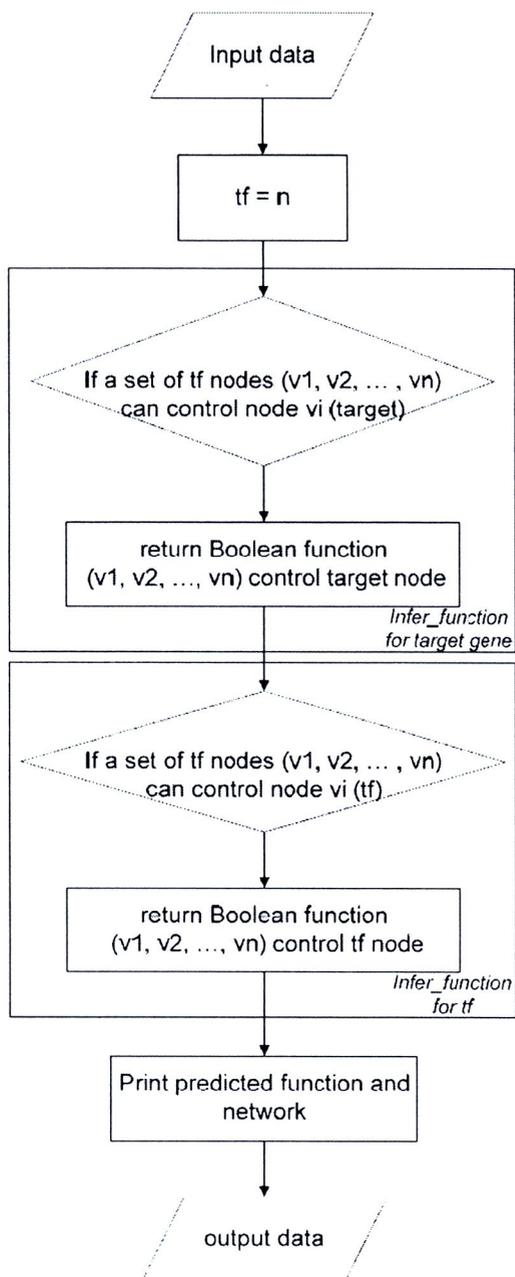
การจะศึกษาเครือข่ายการแสดงออกของยีนโดยใช้ Boolean network ที่มีการแปลงข้อมูลเพียง 2 ระดับ คือ 0 และ 1 ทำให้ได้เครือข่ายควบคุมการแสดงออกของยีนที่มีความซับซ้อน เนื่องจากได้ false positive มาก ดังนั้น จึงพัฒนาโปรแกรมเพื่อสร้างเครือข่ายควบคุมการแสดงออกของยีน โดยเพิ่มเงื่อนไขทางชีววิทยา (biological constraint) บนพื้นฐานของหลักการ genetic circuit และข้อมูลทางห้องปฏิบัติการ เพื่อลดความผิดพลาดและความซับซ้อนของเครือข่ายที่สร้างขึ้น ซึ่งได้มีการเพิ่มเงื่อนไขทางชีววิทยาว่าเอนไซม์ไม่สามารถควบคุมโปรตีนควบคุม (transcription factors) ได้ เพื่อเป็นการลดความซับซ้อนของ network ลง ในงานวิจัยนี้ ใช้ภาษา C++ ในการพัฒนา Constraint-based Boolean network โดยใช้ข้อมูล galactose pathway [15] ที่ผ่านขั้นตอน 3.3 ขั้นตอนการทำงานของโปรแกรมที่พัฒนาขึ้นแสดงดังรูปที่ 3.9-3.10 ส่วนโค้ดสคริป C++ สำหรับสร้าง gene regulatory network ของกลุ่มยีนที่สนใจด้วยวิธี Constraint-based Boolean network [ภาคผนวก ข]



รูปที่ 3.9 ขั้นตอนการพัฒนาโปรแกรมการสร้างเครือข่ายการแสดงออกของยีนแบบมีเงื่อนไข



Non-constraint-based Boolean Network
(classical Boolean network) [1]



Constraint-based Boolean network

รูปที่ 3.10 เปรียบเทียบ Non-constraint-based Boolean Network (classical Boolean network) และ Constraint-based Boolean network

จากรูป 3.9 โปรแกรมจะรับค่าข้อมูล binary values ผ่านทาง command line โดยการเรียกใช้โปรแกรมด้วยคำสั่ง

```
>./boolean.out หรือ
>./boolean.out <k-value> <tf-value> <input> <output>
```

โดย <k-value> <tf-value> <input> <output> เป็นพารามิเตอร์ที่ผู้วิเคราะห์กำหนด มีรายละเอียดดังนี้

- *k-value* คือ จำนวนยีนที่มาควบคุม (parent genes) กำหนดมากที่สุด คือ 5
- *tf-value* คือ โปรตีนควบคุม (Transcription factors: TF) ซึ่งจำเป็นต้องเป็นเงื่อนไขทางชีววิทยา จากตัวอย่างไฟล์ข้อมูลดังตารางที่ 3.8 ถ้ากำหนด TF = 2 นั่นคือ g1 และ g2 สามารถควบคุม g3 แต่ g3 ไม่สามารถควบคุม g1 และ g2 ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนอาจมี TF หรือไม่มีก็ได้ หากไม่มี TF โปรแกรมก็จะสร้างเครือข่ายควบคุมการแสดงออกของยีนโดยไม่ได้คำนึงถึงเงื่อนไขทางชีววิทยาที่ใส่เข้าไป แต่หากมีการใส่ TF ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนด้วย โปรแกรมจะมีการพิจารณาว่า เอนไซม์ ไม่ควรควบคุมการแสดงออกของ TF ได้ ดังนั้น โปรแกรมจะไม่พิจารณาฟังก์ชันบูลีนของเอนไซม์ที่มีต่อ TF นั้นๆ
- *input* คือ ชื่อไฟล์ข้อมูล binary values (.csv) ตัวอย่างไฟล์ข้อมูลดังตารางที่ 3.8 โดยจัด TF ไว้ column ต้นๆ แล้วตามด้วยเอนไซม์ ดังนั้น หากไฟล์ข้อมูล มี TF 2 ตัว เช่น g1 และ g2 ก็ต้องจัด g1 ไว้คอลัมน์แรก แล้วตามด้วย g2 ที่คอลัมน์สอง ส่วนคอลัมน์สามเป็นต้นไป คือ เอนไซม์ โปรแกรมจะรู้อัตโนมัติว่า TF อยู่ตรงตำแหน่งคอลัมน์ใดจากการกำหนดจำนวน TF ให้โปรแกรม
- *output* คือ ชื่อไฟล์ผลที่ได้จากการวิเคราะห์ด้วย Constraint-based Boolean network (.csv) ตัวอย่างไฟล์ผลที่ได้จากการวิเคราะห์ด้วย Constraint-based Boolean network ดังตารางที่ 3.9 โดยแถว คือ ยีนหรือโปรตีนควบคุม (parent genes) ส่วนคอลัมน์ คือ ยีนเป้าหมาย (target genes) “i” คือ inhibition เป็นความสัมพันธ์ระหว่างคู่ยีนแบบยับยั้ง (inhibition) ส่วน “a” คือ activation เป็นความสัมพันธ์ระหว่างคู่ยีนแบบกระตุ้น “-” คือ ไม่มีความสัมพันธ์ระหว่างคู่ยีน โดยการศึกษาการพิจารณาการควบคุมยีนตัวเองออก ซึ่งก็คือ “x” ในผล จากตัวอย่างในตาราง 3.9 จะเห็นว่าเอนไซม์ g3 ถูกยับยั้งด้วยยีน g1 และกระตุ้นด้วยยีน g2 ดังนั้นฟังก์ชันบูลีนของความสัมพันธ์นี้คือ $g3 = g2 \text{ AND NOT } g1$ ซึ่งผลที่ได้จะแสดงเป็นเครือข่ายควบคุมการแสดงออกของยีนด้วยโปรแกรม Cytoscape [20]

จากการเปรียบเทียบขั้นตอนการทำงานโปรแกรม Non-constraint-based Boolean Network (classical Boolean network) [1] และ Constraint-based Boolean network ที่พัฒนาขึ้นมาในงานวิจัยนี้ แสดงดังรูปที่ 3.10

ตารางที่ 3.8 ตัวอย่างไฟล์ข้อมูล binary values (.csv) ที่ใช้วิเคราะห์ด้วย Constraint-based Boolean network

| g1 | g2 | g3 |
|----|----|----|
| 0 | 1 | 1 |
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

ตารางที่ 3.9 ตัวอย่างไฟล์ผล (.csv) จากการวิเคราะห์ด้วย Constraint-based Boolean network

| g1 | g2 | g3 |
|----|----|----|
| g1 | x | i |
| g2 | a | x |
| g3 | a | i |

3.5 การทดสอบและประเมินโปรแกรม (Model validation)

หลังจากการพัฒนาโปรแกรมคอมพิวเตอร์เพื่อสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไขแล้ว โปรแกรมที่พัฒนาขึ้นนั้นจำเป็นต้องมีการทดสอบเพื่อวัดประสิทธิภาพของโปรแกรม Constraint-based Boolean network ที่พัฒนาขึ้น ทั้งนี้ มีการประเมินความถูกต้องของเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway ที่สร้างขึ้นด้วยข้อมูล binary values จากทุกวิธีคือวิธี Mean Max-x%Max และ Sign of \log_2 ratio เพื่อทดสอบสมมติฐานที่ตั้งไว้ว่า discretization method ที่ให้ค่า similarity score ของการเปรียบเทียบเดนไดรแกรมของ binary values เทียบกับ raw data สูงที่สุดจะเป็นวิธีที่เหมาะสมที่สุดในการสร้างเครือข่ายควบคุมการแสดงออกของยีนด้วยเทคนิคบูลีน ข้อมูลที่ได้จากโปรแกรมที่พัฒนาขึ้นจากภาษา C++ นั้น ผลที่ได้จากโปรแกรมจะสามารถระบุรูปแบบการควบคุมกันของยีนว่า ยีนหนึ่งสามารถควบคุมแบบการแสดงออกของยีนอื่นแบบยับยั้ง (i; inhibition) หรือแบบกระตุ้น (a; activation) ส่วน x คือ ไม่มีการกระตุ้น/ยับยั้งกันระหว่างคู่อินนั้นๆ ท้ายสุดจะนำผลเครือข่ายควบคุมการแสดงออกของยีนที่สร้างจาก Constraint-based Boolean network ที่ผ่านขั้นตอน discretization ด้วยวิธีการต่างๆ มาเปรียบเทียบกับเครือข่ายควบคุมการแสดงออกของยีนอ้างอิง (reference network) แสดงในตารางที่ 3.10 ซึ่งเป็นข้อมูลทางห้องปฏิบัติการหรือเป็นข้อมูลที่มีการกล่าวอ้างมาก่อน โดยประเมินประสิทธิภาพของโปรแกรม Constraint-based Boolean network ด้วยค่าทางสถิติ [21] ดังนี้

1. False positive rate (FPR) คือ ค่าอัตราส่วนการที่โปรแกรมทำนายว่ามีความสัมพันธ์กันของคู่ยีนทั้งที่ความจริงไม่มี ต่อความสัมพันธ์ของคู่ยีนที่ไม่มีใน reference network คำนวณได้จาก $FP/(TN+FP)$
2. False negative rate (FNR) คือ ค่าอัตราส่วนการที่โปรแกรมทำนายว่าไม่มีความสัมพันธ์กันของคู่ยีนที่ถูกต้องจริงๆ ต่อในความสัมพันธ์ของคู่ยีนที่สามารถทำนายได้ทั้งหมด คำนวณได้จาก $FN/(TP+FN)$
3. False discovery rate (FDR) คือ อัตราส่วนการที่โปรแกรมทำนายผิดพลาด จากความสัมพันธ์ทั้งหมดที่ทำนายได้ คำนวณได้จาก $FP/(TP+FP)$
4. ค่าความถูกต้อง (accuracy) คือ ค่าความถูกต้องของโปรแกรม คำนวณได้จาก $(TP+TN)/(TP+TN+TP+FN)$
5. ความแม่นยำ (precision) คือ ค่าความแม่นยำของโปรแกรมในการสร้างเครือข่ายควบคุมการแสดงออกของยีน คำนวณได้จากรูปแบบความสัมพันธ์ของการ regulation คือ $TR/(TR+FR+FI)$
6. ความจำเพาะ (specificity) คือ ค่าความจำเพาะต่อของการทำนายความสัมพันธ์ในเครือข่ายที่ไม่มีอยู่จริง คำนวณได้จากรูปแบบความสัมพันธ์ของการ regulation คือ $TZ/(TZ+FR)$
7. ความไว (sensitivity) คือ สัดส่วนของการทำนายความสัมพันธ์ที่มีอยู่จริง ในการทำนายหรือสร้างเครือข่ายควบคุมการแสดงออกของยีน คำนวณได้จากรูปแบบความสัมพันธ์ของการ regulation คือ $TR/(TR+FZ+FI)$

โดยคำนวณจากค่าการวัดผลลัพธ์ออกมาเป็นคลาส positive หรือ negative ต่างๆ ดังนี้

- True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าเป็นการควบคุมแบบ (i; inhibition) หรือกระตุ้น (a; activation) และสอดคล้องกับ reference network ว่าเป็นแบบยับยั้ง (i; inhibition) หรือกระตุ้น (a; activation) เช่นเดียวกัน
- True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่มีความสัมพันธ์กันระหว่างยีนและใน reference network ก็ไม่มีความสัมพันธ์ระหว่างยีนเช่นเดียวกัน
- False Positive (FP) คือ สิ่งที่โปรแกรมไม่สามารถทำนายความสัมพันธ์ระหว่างคู่ยีนได้ นั่นคือ ไม่มีความสัมพันธ์กันของคู่ยีน แต่ใน reference network มีความสัมพันธ์ระหว่างคู่ยีนเกิดขึ้นไม่ว่าจะเป็นการยับยั้งหรือกระตุ้น
- False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่ามีความสัมพันธ์ระหว่างคู่ยีนว่ามีความสัมพันธ์กันระหว่างคู่ยีน แต่ใน reference network ไม่มีความสัมพันธ์ระหว่างคู่ยีนนั้น หรือ ความสัมพันธ์ของคู่ยีนแตกต่างจากใน reference network เช่น โปรแกรมทำนายว่าความสัมพันธ์เป็นการกระตุ้น แต่ใน reference network มีความสัมพันธ์ระหว่างคู่ยีนเป็นแบบยับยั้ง หรือโปรแกรมทำนายว่าความสัมพันธ์เป็นแบบยับยั้ง แต่ใน reference network มีความสัมพันธ์ระหว่างคู่ยีนเป็นการกระตุ้น

รูปแบบความสัมพันธ์ของการ regulation ในเครือข่ายโดยพิจารณาความสัมพันธ์เป็น activation หรือ inhibition โดยสามารถแบ่งการวัดค่าผลลัพธ์ของการ regulation ในเครือข่ายได้เป็น

- True regulation (TR) คือ โปรแกรมทำนายว่าเป็นการควบคุมแบบ (i; inhibition) หรือกระตุ้น (a; activation) และสอดคล้องกับ reference network ว่าเป็นแบบยับยั้ง (i; inhibition) หรือกระตุ้น (a; activation) เช่นเดียวกัน
- True zero (TZ) คือ โปรแกรมทำนายว่าไม่มีความสัมพันธ์กันระหว่างคู่ยีน สอดคล้องกับใน reference network ว่าไม่มีความสัมพันธ์ระหว่างคู่ยีนนั้นเช่นเดียวกัน
- False regulation (FR) คือ โปรแกรมทำนายว่ามีความสัมพันธ์กันระหว่างคู่ยีน ไม่ว่าจะเป็นการ activation หรือ inhibition แต่ไม่มีความสัมพันธ์นั้นใน reference network
- False zero (FZ) โปรแกรมทำนายว่าไม่มีความสัมพันธ์กันระหว่างคู่ยีน แต่ใน reference network มีความสัมพันธ์เป็นแบบ activation
- False interaction (FI) โปรแกรมทำนายว่ามีความสัมพันธ์กันระหว่างคู่ยีนแบบ inhibition แต่ใน reference network มีความสัมพันธ์เป็นแบบ activation

ผลของการวัดผลลัพธ์ออกมาเป็นคลาส positive หรือ negative ของเครือข่ายควบคุมการแสดงออกของยีนด้วย Constraint-based Boolean network โดยใช้วิธี Max-x%Max Mean และ Sign of \log_2 ratio ดังตารางที่ 3.11-3.13 ตามลำดับ เครือข่ายควบคุมการแสดงออกของยีนด้วยโปรแกรม Constraint-based Boolean network โดยใช้ข้อมูล binary values จากทั้ง 3 วิธีการ แสดงดังรูปที่ 3.11 และผลการประเมินประสิทธิภาพของโปรแกรม Constraint-based Boolean network ในการสร้างส่วนเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway แสดงในรูป 3.12

ตารางที่ 3.10 Reference network ของ galactose pathway [16]

| | | <i>Regulator genes</i> | | | | | | |
|---------------------|--------------|------------------------|-----------------|--------------|-------------|-----------------|-------------|--------------|
| | | <i>GAL</i> 3 | <i>GAL</i> 4 | <i>GAL80</i> | <i>GAL1</i> | <i>GAL</i> 2 | <i>GAL7</i> | <i>GAL10</i> |
| <i>Target genes</i> | <i>GAL3</i> | x | a | - | - | - | - | - |
| | <i>GAL4</i> | - | x | i | - | - | - | - |
| | <i>GAL80</i> | i | a | x | - | - | - | - |
| | <i>GAL1</i> | a | a | i | x | - | - | - |
| | <i>GAL2</i> | a | a | i | - | x | - | - |
| | <i>GAL7</i> | a | a | i | - | - | X | - |
| | <i>GAL10</i> | a | a | i | - | - | - | x |

ตารางที่ 3.11 คลาส positive และ negative prediction ของการประเมินประสิทธิภาพของโปรแกรม ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway จากวิธีการ Max-x%Max

| | <i>GAL3</i> | <i>GAL4</i> | <i>GAL80</i> | <i>GAL1</i> | <i>GAL2</i> | <i>GAL7</i> | <i>GAL10</i> |
|--------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|
| <i>GAL3</i> | TN | FN | TN | TN | TN | TN | TN |
| <i>GAL4</i> | FP | TN | TP | TN | TN | TN | TN |
| <i>GAL80</i> | FN | FN | TN | TN | TN | TN | TN |
| <i>GAL1</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL2</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL7</i> | TP | TP | FN | TN | TN | TN | TN |
| <i>GAL10</i> | FN | FN | FN | TN | TN | TN | TN |

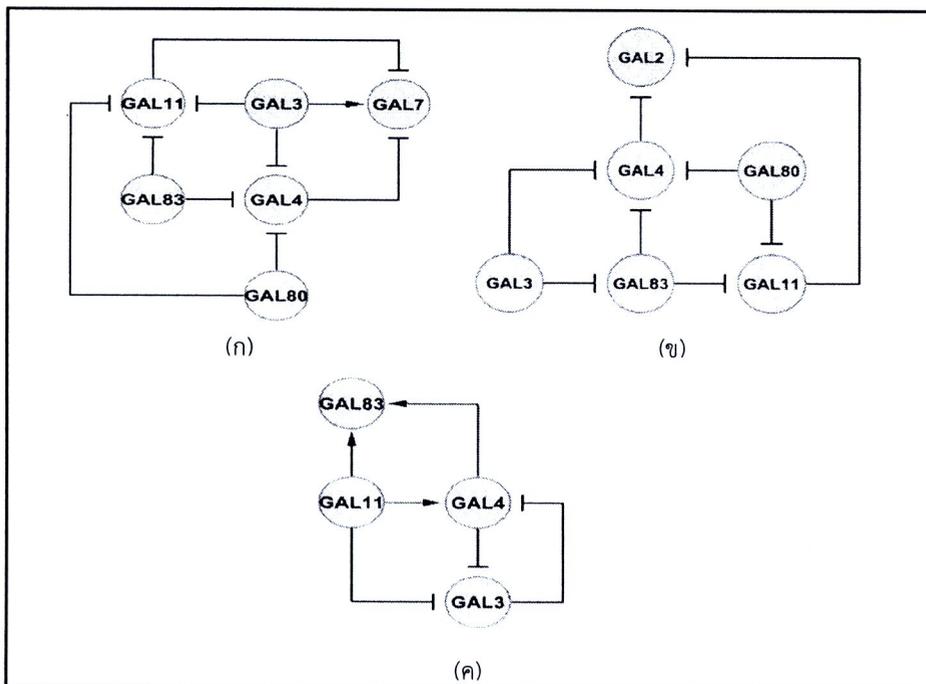
ตารางที่ 3.12 คลาส positive และ negative prediction ของการประเมินประสิทธิภาพของโปรแกรม ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway จากวิธีการ Mean

| | <i>GAL3</i> | <i>GAL4</i> | <i>GAL80</i> | <i>GAL1</i> | <i>GAL2</i> | <i>GAL7</i> | <i>GAL10</i> |
|--------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|
| <i>GAL3</i> | TN | TP | FP | TN | TN | TN | TN |
| <i>GAL4</i> | TN | TN | FN | TN | TN | TN | TN |
| <i>GAL80</i> | FN | FN | TN | TN | TN | TN | TN |
| <i>GAL1</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL2</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL7</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL10</i> | FN | FN | FN | TN | TN | TN | TN |

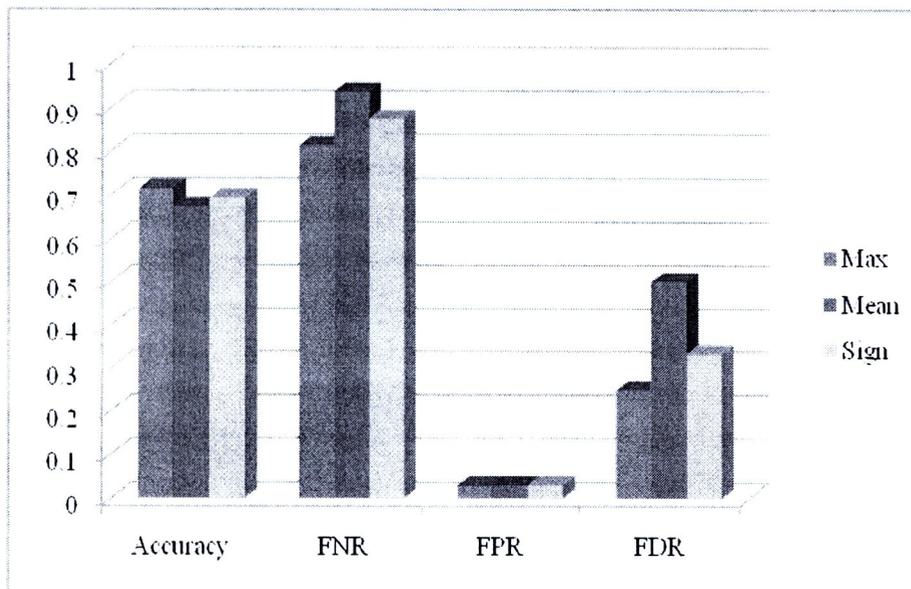
ตารางที่ 3.13 คลาส positive และ negative prediction ของการประเมินประสิทธิภาพของโปรแกรม ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway จากวิธีการ Sign of \log_2 ratio

| | <i>GAL3</i> | <i>GAL4</i> | <i>GAL80</i> | <i>GAL1</i> | <i>GAL2</i> | <i>GAL7</i> | <i>GAL10</i> |
|--------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|
| <i>GAL3</i> | TN | FN | TN | TN | TN | TN | TN |
| <i>GAL4</i> | FP | TN | TP | TN | TN | TN | TN |
| <i>GAL80</i> | FN | FN | TN | TN | TN | TN | TN |
| <i>GAL1</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL2</i> | FN | TP | FN | TN | TN | TN | TN |
| <i>GAL7</i> | FN | FN | FN | TN | TN | TN | TN |
| <i>GAL10</i> | FN | FN | FN | TN | TN | TN | TN |

จากรูปที่ 3.12 จะเห็นได้ว่า discretization method ที่ให้ค่า similarity score ของการเปรียบเทียบเดนไดรแกรมของ binary values เทียบกับ raw data สูงที่สุด นั่นคือ Max-60%Max ไม่ใช่วิธีที่เหมาะสมที่สุดในการสร้างเครือข่ายควบคุมการแสดงออกของยีนด้วยเทคนิคบูลีน เพราะไม่สามารถสร้างเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway ที่ถูกต้องแม่นยำที่สุด แต่กลับพบว่าการใช้วิธี Mean แปลงค่าการแสดงออกของยีนเป็น 0 หรือ 1 นั้นสามารถช่วยให้การสร้างเครือข่ายด้วย Constraint-based Boolean network ได้ถูกต้องแม่นยำกว่าวิธีการอื่น นอกจากนี้ เมื่อลองเปรียบเทียบประสิทธิภาพของการเครือข่ายควบคุมการแสดงออกของยีนที่สร้างจากโปรแกรม Constraint-based Boolean network ที่พัฒนาขึ้นเปรียบเทียบกับ Non-Constraint-based Boolean network (classical Boolean algorithm) [1] โดยการใช้ข้อมูล binary values จากวิธี mean พบว่าการเพิ่มเงื่อนไขทางชีววิทยาใน Constraint-based Boolean network สามารถเพิ่มความถูกต้อง (accuracy) โดยลดความผิดพลาด (FDR) ของการทำนายลง ดังตารางที่ 3.14 รวมทั้งช่วยเพิ่มความแม่นยำ (precision) ความจำเพาะ (specificity) แม้ว่าความไวของโปรแกรม (sensitivity) ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนของ Non-Constraint-based Boolean network (classical Boolean algorithm) [1] และ Constraint-based Boolean network จะไม่แตกต่างกันก็ตาม ดังรูปที่ 3.13 และรูปแบบความสัมพันธ์ของการ regulation ของเครือข่ายควบคุมการแสดงออกของยีนด้วย Mean Non-Constraint-based Boolean network (classical Boolean algorithm) [1] และ Constraint-based Boolean network แสดงดังตารางที่ 3.15-3.16



รูปที่ 3.11 เครือข่ายควบคุมการแสดงออกของยีนด้วยโปรแกรม Constraint-based Boolean network ด้วยข้อมูล binary values จากวิธีการ Max-x%Max (ข) Sign of \log_2 ratio และ (ค) Mean



รูปที่ 3.12 ผลของการประเมินประสิทธิภาพของโปรแกรม Constraint-based Boolean network ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนด้วยวิธี Mean Max-x%Max และ Sign of \log_2 ratio

ตารางที่ 3.14 การเปรียบเทียบประสิทธิภาพของการเครือข่ายควบคุมการแสดงออกของยีนที่สร้างจากโปรแกรม Constraint-based Boolean network และ Classical Boolean algorithm [1]

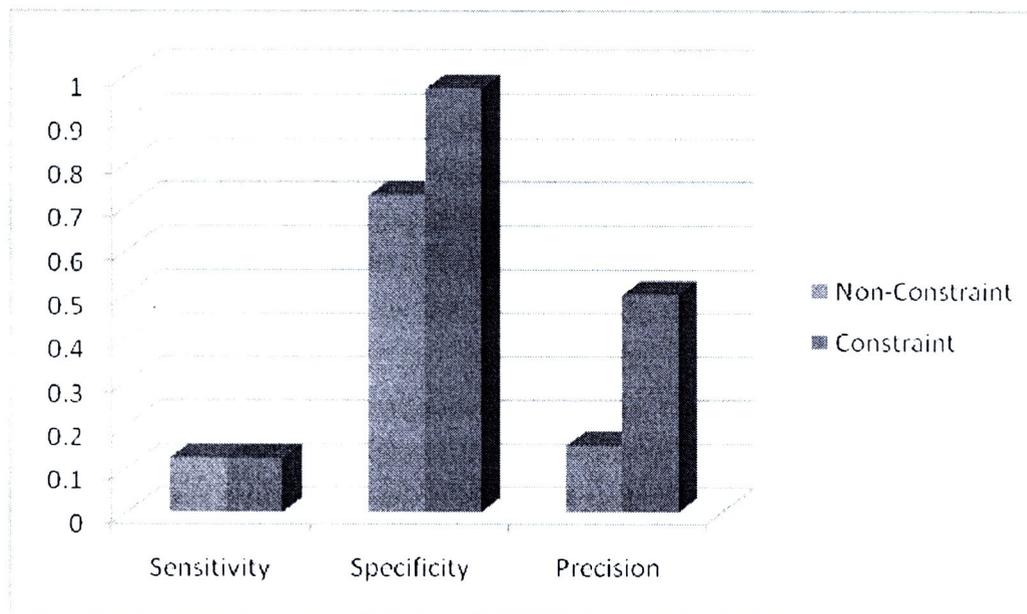
| | Non-Constraint | Constraint-based |
|----------|----------------|------------------|
| Accuracy | 0.57 | 0.71 |
| FNR | 0.75 | 0.81 |
| FPR | 0.27 | 0.03 |
| FDR | 0.69 | 0.25 |

ตารางที่ 3.15 รูปแบบความสัมพันธ์ของการ regulation จากการประเมินประสิทธิภาพของโปรแกรม Constraint-based Boolean network ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway จากวิธีการ Mean

| | <i>GAL3</i> | <i>GAL4</i> | <i>GAL80</i> | <i>GAL1</i> | <i>GAL2</i> | <i>GAL7</i> | <i>GAL10</i> |
|--------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|
| <i>GAL3</i> | TZ | FZ | TZ | TZ | TZ | TZ | TZ |
| <i>GAL4</i> | FR | TZ | TR | TZ | TZ | TZ | TZ |
| <i>GAL80</i> | FZ | FZ | TZ | TZ | TZ | TZ | TZ |
| <i>GAL1</i> | FZ | FZ | FZ | TZ | TZ | TZ | TZ |
| <i>GAL2</i> | FZ | FZ | FZ | TZ | TZ | TZ | TZ |
| <i>GAL7</i> | TR | FI | FZ | TZ | TZ | TZ | TZ |
| <i>GAL10</i> | FZ | FZ | FZ | TZ | TZ | TZ | TZ |

ตารางที่ 3.16 รูปแบบความสัมพันธ์ของการ regulation จากการประเมินประสิทธิภาพของโปรแกรม Classical Boolean algorithm [1] ในการสร้างเครือข่ายควบคุมการแสดงออกของยีนใน galactose pathway จากวิธีการ Mean

| | <i>GAL3</i> | <i>GAL4</i> | <i>GAL80</i> | <i>GAL1</i> | <i>GAL2</i> | <i>GAL7</i> | <i>GAL10</i> |
|--------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|
| <i>GAL3</i> | FR | FI | TZ | TZ | FR | FR | TZ |
| <i>GAL4</i> | FR | TZ | TR | FR | FR | FR | FR |
| <i>GAL80</i> | FZ | FZ | TZ | TZ | TZ | TZ | TZ |
| <i>GAL1</i> | FZ | FZ | FZ | TZ | TZ | TZ | TZ |
| <i>GAL2</i> | FZ | FZ | FZ | TZ | TZ | TZ | TZ |
| <i>GAL7</i> | TR | FI | FZ | TZ | FR | TZ | TZ |
| <i>GAL10</i> | FZ | FZ | FZ | TZ | TZ | TZ | TZ |



รูปที่ 3.13 การเปรียบเทียบประสิทธิภาพของการเครือข่ายควบคุมการแสดงออกของยีนที่สร้างจากโปรแกรม Constraint-based Boolean network ที่พัฒนาขึ้นเปรียบเทียบกับ Non-Constraint-based Boolean network (classical Boolean algorithm) [1]

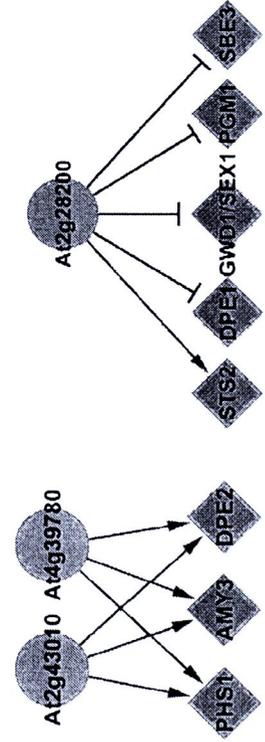
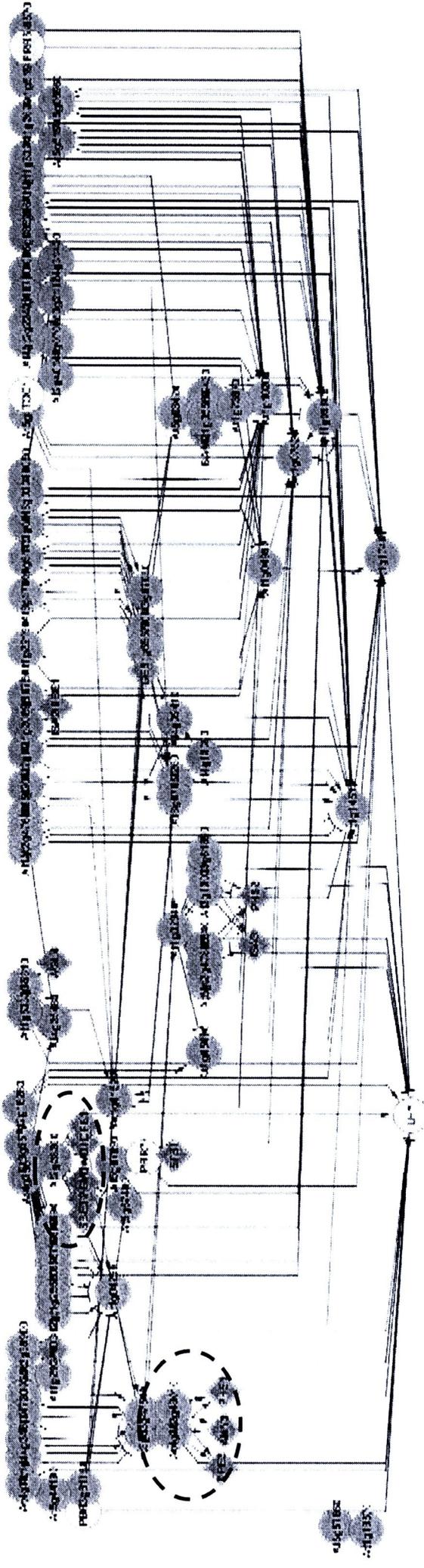
3.6 การประยุกต์ใช้โปรแกรมเพื่อสร้างและศึกษาเครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสังเคราะห์แป้ง (case study)

ในขั้นการทดลองนี้ จะนำเอา Constraint-based Boolean network ที่พัฒนาขึ้นมาใช้สร้างเครือข่ายควบคุมการทำงานของยีนในกระบวนการสังเคราะห์แป้ง เพื่อทำความเข้าใจเกี่ยวกับกลไกการทำงานของกระบวนการสังเคราะห์แป้งมากขึ้น โดยใช้ข้อมูลการแสดงออกของยีนในพืชต้นแบบ นั่นคือ *Arabidopsis thaliana* ภายใต้สภาวะการเลี้ยงแบบ diurnal cycle นั่นคือ ภายใต้สภาวะการเลี้ยงแบบ

กลางวัน-กลางคืน (12-12 ชั่วโมง) หลังจากผ่านขั้นตอนการ pre-process [17, 18] ด้วยภาษา R [<http://www.r-project.org/>] ใน Bioconductor [<http://www.bioconductor.org>] ในการจัดการข้อมูล โดยเตรียมข้อมูลการแสดงออกของยีนให้อยู่ในรูปแบบ \log_2 intensity และหายีนที่มีการแสดงออกอย่างมีนัยสำคัญด้วยโปรแกรม EDGE [19] เวอร์ชัน 1.1.175 [<http://faculty.washington.edu/jstorey/edge/>] หลังจากนั้น จึงเลือกกลุ่มยีนทำหน้าที่เป็นเอนไซม์ที่เกี่ยวข้องกับการสังเคราะห์แป้ง (Starch genes) โดยอ้างอิง [14] และโปรตีนควบคุม (Transcription factors; TF) และยีนที่เกี่ยวข้องการเวลา (clock genes) มาสร้างเครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสังเคราะห์แป้ง พบว่า มี Starch genes 21 ยีน [ภาคผนวก ค1] และ TF และ clock genes รวม 113 ยีน [ภาคผนวก ค2] ดังนั้น กลุ่มยีนที่จะนำไปสร้างเครือข่ายควบคุมการแสดงออกของยีนด้วย Constraint-based Boolean network มีทั้งหมด 134 ยีน โดยใช้ข้อมูล binary values ที่ได้จากวิธีการ Sign of \log_2 Ratio (ค่า similarity score สูงสุด) แสดงในรูปที่ 3.14

เครือข่ายควบคุมการแสดงออกของยีนของกระบวนการสังเคราะห์แป้งในพืช *Arabidopsis* ที่สร้างขึ้น ประกอบด้วย 112 nodes และ 234 directed edges โดย nodes แสดงถึงยีน และ directed edges คือ รูปแบบความสัมพันธ์ระหว่างยีนที่ได้จาโปรแกรม Constraint-based Boolean network ซึ่งเป็นได้ทั้งแบบกระตุ้นหรือยับยั้ง โดยความสัมพันธ์แบบกระตุ้นแสดงเป็นเส้นสีแดง หรือยับยั้งแสดงเป็นเส้นสีน้ำเงิน โดยเครือข่ายที่สร้างขึ้นประกอบด้วย starch genes จำนวน 16 ยีน (สี่เหลี่ยมขนมเปียกปูน สีเขียว) transcription factors (TF) จำนวน 91 ยีน (วงกลมสีส้ม) และ clock genes จำนวน 5 ยีน (วงกลมสีเหลือง) เมื่อพิจารณาเครือข่ายย่อย (sub-network) ระหว่าง starch genes และ TF พบว่า มี 2 เครือข่ายย่อย ดังรูปที่ 3.13 โดยเครือข่ายแรกเป็นความสัมพันธ์ของเอนไซม์ที่เกี่ยวข้องกับการสลายแป้ง (starch degradation) 3 เอนไซม์ คือ A cytosolic disproportionating enzyme like protein (*At2g40840*; *DPE2*) Alpha-amylase like 3 (*At1g69830*; *AMY3*) และ A plastidic alpha-glucan phosphorylase (*At3g29320*; *PHS1*) โดยทั้ง 3 เอนไซม์นี้เกี่ยวข้องกับ 2 โปรตีนควบคุม (TF) คือ *At4g39780* และ *At2g43010* ซึ่งเป็นโปรตีนควบคุมที่มี AP2 domain

ส่วน sub-network ที่สองเป็นความสัมพันธ์ของเอนไซม์ 5 เอนไซม์ คือ phosphoglucosyltransferase (*At5g51820*; *PGM1*) Starch synthase II (*At3g01180*; *STS2*) Starch branching enzyme III (*At2g36390*; *SBE3*) Glucan water dikinase 1 (*At1g10760*; *GWD1/SEX1*) และ Glucanotransferase (*At5g64860*; *DPE1*) โดยทั้ง 5 เอนไซม์นี้เกี่ยวข้องกับโปรตีนควบคุม 1 ตัว คือ *At2g28200* which has zinc-finger domain โดย *PGM1* *STS2* และเอนไซม์ *SBE3* เป็นเอนไซม์ที่เกี่ยวข้องกับการสร้างแป้ง (Starch biosynthesis) โดยเฉพาะ amylopectin *PGM1* เป็นเอนไซม์ที่เปลี่ยน alpha-D-glucose-6-phosphate ไปเป็น alpha-D-glucose-1-phosphate เพื่อเปลี่ยนเป็น ADP-glucose ต่อไปด้วยเอนไซม์ ADP-glucose pyrophosphorylase (*AGPase*) หลังจากนั้น Starch synthase II (*STSII*) จะต่อเชื่อม ADP-glucose เป็นสายยาวด้วยการเชื่อมพันธะชนิดอัลฟา-1-4 (1->4)-alpha-D-glucosyl) ส่วนเอนไซม์ *SBE3* จะเป็นเอนไซม์ที่สร้างกิ่งด้วยพันธะ ชนิดอัลฟา-1-6 (1->6)-alpha-D-glucosyl) ในขณะที่ *GWD1/SEX1* และ *DPE1* เป็นเอนไซม์ที่สำคัญกระบวนการสลายแป้งเพื่อเปลี่ยนเป็น maltose แล้วลำเลียงออกนอก cytosol อย่างไรก็ตาม แม้ว่าจะสามารถสร้างเครือข่ายควบคุมการแสดงออกของ starch genes และ transcription factors ได้ แต่ข้อมูลดังกล่าวยังต้องพิสูจน์ในห้องปฏิบัติการว่าเอนไซม์เหล่านี้เกี่ยวข้อง หรือถูกควบคุมการแสดงออกด้วย transcription factors เหล่านี้หรือไม่



รูปที่ 3.14 (ก) เครือข่ายควบคุมการแสดงออกของยีนในการควบคุมการสังเคราะห์แป้งซึ่งประกอบด้วยยีน starch genes จำนวน 16 ยีน (สีเหลี่ยมขนมเปียกนูนสีเขียว) transcription factors (TF) จำนวน 91 ยีน (วงกลมสีส้ม) และ clock genes จำนวน 5 ยีน (วงกลมสี่เหลี่ยม) เส้นสีน้ำเงิน คือ ความสัมพันธ์แบบยับยั้ง (inhibition) เส้นสีแดง คือ ความสัมพันธ์แบบกระตุ้น วงกลมเส้นประนั้น คือ sub-network ของ starch-TF network (ข) Sub-network ของ starch-TF gene network