

Thesis Title Effects of Expression Profile Data Variation on
Boolean Gene Regulatory Network Predictions

Author Ms. Kanda Euatham

Degree Master of Science (Bioinformatics)

Thesis Advisor Dr. Natee Tongsir

ABSTRACT

Reconstructing the regulatory relationships between genes using multiple time point expression profile data (EPD) is a powerful computational method to gain insight into gene networks. One such method uses binary on/off relationships to characterize the under- and over-expression of genes acting in unison. This approach yields results as to which genes are co-regulated, regulating or being regulated by other genes using only the relative expression levels of the genes of interest at multiple time points. One aspect of the EPD these methods often fail to account for is the inherent variability in the measurements of the gene expression levels. Here we characterize a measure of the variability in expression levels for a single time point as opposed to between different time points as is normally done. Using this measure of the inherent variability in a dataset for a single EPD time point we generate multiple

new expression profile data samples from the original data and measured variability. These new datasets are then binarized to test whether the gene network relationships change due to the random sampling. This also allows us to test different variation magnitudes to set limits on how large the inherent variability should be to yield reproducible results for the binary gene network method. We find that the current variabilities in EPD are in general too large by a factor of 2.x to yield reproducible gene regulatory networks, but that the data for some particular genes are sufficient to generate reproducible binarizations.

ชื่อเรื่องวิทยานิพนธ์

ผลจากการมีข้อมูลเอ็กซ์เพรสชัน โปรไฟล์ที่หลากหลายต่อ
การทำนายเครือข่ายการควบคุมยีนแบบบูลีน

ผู้เขียน

นางสาวกานดา เอื้อธรรม

ปริญญา

วิทยาศาสตรมหาบัณฑิต (ชีวสารสนเทศศาสตร์)

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ดร. นที ทองศิริ

บทคัดย่อ

การหาความสัมพันธ์ในเครือข่ายการควบคุมยีนโดยใช้ข้อมูลไมโครอะเรย์การแสดงออกของยีนแบบโทมัสรีส์ เป็นเครื่องมือที่มีประสิทธิภาพสูงที่ทำให้เราสามารถศึกษาเครือข่ายการควบคุมยีนได้อย่างลึกซึ้ง การโมเดลแบบบูลีนใช้การแทนค่าข้อมูลการแสดงออกของยีน 2 แบบคือเปิดกับปิด เพื่อวิเคราะห์ว่ายีนมีการแสดงออกมากขึ้นหรือน้อยลงหรือแสดงออกร่วมกัน ด้วยวิธีนี้จะวิเคราะห์ว่ายีนมีการแสดงออกสัมพันธ์กันอย่างไรจากระดับของการแสดงออกของยีนจากหลาย ๆ เวลาที่สนใจ ข้อสังเกตอย่างหนึ่งจากข้อมูลโปรไฟล์การแสดงออกของยีนมักจะมีหลากหลายจากการวัดอยู่เสมอ การศึกษาครั้งนี้จะทำการหาค่าความหลากหลายของข้อมูลนั้นแต่ละเวลา จากนั้นใช้ค่าที่หาได้จากข้อมูลต้นฉบับ นำไปสร้างข้อมูลใหม่ซึ่งมีคุณสมบัติเหมือนข้อมูลต้นฉบับ และหาค่าความหลากหลายจากข้อมูลที่สร้างขึ้นมา ข้อมูลชุดใหม่นี้จะถูกไปนาไรซ์เพื่อเทียบว่ามี การเปลี่ยนแปลงไปจากข้อมูลต้นฉบับอย่างไร รวมทั้งข้อมูลที่สร้างมาใหม่จะถูกเพิ่มค่าความหลากหลายต่าง ๆ เพื่อทดสอบว่า ข้อมูลสามารถมีความหลากหลายได้มากที่สุดเท่าไรที่ยังจะสามารถนำไปสร้างเครือข่ายการควบคุมยีนที่ยังได้ผลคืออยู่ พบว่าข้อมูลต้นฉบับที่นำมาใช้ศึกษายังมีความหลากหลายที่มากเกินไปถึงสองเท่าที่จะนำมาใช้สร้างแบบจำลองเครือข่ายการควบคุมยีนให้ได้ผลดี มีเพียงบางยีนจากชุดข้อมูลต้นฉบับที่สามารถเลือกนำมาสร้างแบบจำลองที่ได้ผลดี