



รายงานวิจัยฉบับสมบูรณ์

โครงการ

“การพัฒนาฐานข้อมูลและโครงสร้างพื้นฐานสำหรับการศึกษาปฏิสัมพันธ์
ระหว่างไมโครอาร์เอ็นเอและโปรโมเตอร์จีโนมมนุษย์”

**Development of an integrated infrastructure for studying microRNA-
promoter interactions in human genome**

โดย ดร. จิตติมา พิริยะพงศา และคณะ

ธันวาคม 2556

รายงานวิจัยฉบับสมบูรณ์

โครงการ

“การพัฒนาฐานข้อมูลและโครงสร้างพื้นฐานสำหรับการศึกษา
ปฏิสัมพันธ์ระหว่างไมโครอาร์เอ็นเอและโปรโมเตอร์จีโนมมนุษย์”

**Development of an integrated infrastructure for studying
microRNA-promoter interactions in human genome**

คณะผู้วิจัย

สังกัด

- | | |
|---------------------------|--|
| 1. ดร. จิตติมา พิริยะพงศา | ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ |
| 2. ดร. ศิษฏาศ ทองสิมา | ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ |
| 3. นาย ชุมพล งามผิว | ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ |
| 4. นาย ชัยวัฒน์ บุตรไชย | ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ |

**สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัยและ
ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ**

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. และศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ
ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณสำนักงานกองทุนสนับสนุนการวิจัยและศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ที่ได้มอบทุนสนับสนุนการทำวิจัยในโครงการนี้จนประสบผลสำเร็จ นอกจากนี้ ขอขอบคุณศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติที่ได้ให้การสนับสนุนในส่วนของระบบคอมพิวเตอร์สมรรถนะสูงที่ใช้ในการวิเคราะห์ข้อมูล รวมทั้งอนุญาตให้จัดตั้งฐานข้อมูลบนเครื่องคอมพิวเตอร์แม่ข่ายที่มีอยู่ คณะผู้วิจัยยังขอขอบคุณ ศ.ดร.นพ.อภิวัฒน์ มุทิรางกูร สำหรับการสนับสนุนและความคิดเห็นในการพัฒนาฐานข้อมูลในโครงการนี้ และขอขอบคุณ Dr. Philip J. Shaw สำหรับข้อคิดเห็นที่เป็นประโยชน์ต่อโครงการ

จิตติมา พิริยะพงศา
หัวหน้าโครงการ

รหัสโครงการ: TRG5580011

ชื่อโครงการ: การพัฒนาฐานข้อมูลและโครงสร้างพื้นฐานสำหรับการศึกษาปฏิสัมพันธ์ระหว่างไมโครอาร์เอ็นเอและโปรโมเตอร์จีโนมมนุษย์

ชื่อนักวิจัย: ดร. จิตติมา พิริยะพงศา / ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ

E-mail Address: jittima.pir@biotec.or.th

ระยะเวลาโครงการ: 1 ปี 6 เดือน (2 กรกฎาคม 2555 ถึง 1 มกราคม 2557)

งานวิจัยนี้มุ่งเน้นการจัดตั้งฐานข้อมูลสำหรับการศึกษาวิจัยด้านกลไกที่พบใหม่ของไมโครอาร์เอ็นเอ ในการควบคุมการแสดงออกของยีนผ่านการจับกับตำแหน่งเป้าหมายบนยีนโปรโมเตอร์ คณะผู้วิจัยได้พัฒนาฐานข้อมูลสาธารณะในรูปแบบเว็บไซต์ที่ชื่อว่า microPIR (microRNA-Promoter Interaction Resource) ซึ่งแสดงข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอที่อยู่บนลำดับเบสโปรโมเตอร์ทั่วจีโนมมนุษย์ ที่ได้มาจากการทำนายด้วยวิธีทางคอมพิวเตอร์โดยอาศัยข้อมูลการทดลองที่มีมาก่อนในการสร้างกฎการทำนาย รวมทั้งได้รวมเอาข้อมูลการทดลองอื่นๆที่เกี่ยวข้องเข้าไว้ในฐานข้อมูล เพื่อเป็นข้อมูลสนับสนุนให้นักวิจัยสามารถประเมินคุณภาพและพิจารณาเลือกตำแหน่งเป้าหมายที่น่าสนใจได้ นอกจากนี้ ยังได้เชื่อมโยงข้อมูลจีโนมต่างๆที่สำคัญจำนวนมากมาไว้ด้วยกัน เพื่อให้ผู้ใช้สามารถศึกษาความสัมพันธ์ระหว่างตำแหน่งเป้าหมายที่ได้จากการทำนายและคุณลักษณะจีโนมอื่นๆ ระบบฐานข้อมูลยังประกอบไปด้วยการแสดงผลในรูปแบบกราฟิกผ่านจีโนมเบราว์เซอร์ในตัว ซึ่งสามารถให้บริการมุมมองที่ครอบคลุมข้อมูลที่มีหลายมิติในจีโนมบริเวณเดียวกัน ลักษณะทั้งหมดนี้ทำให้ฐานข้อมูล microPIR ถือได้ว่าเป็นคลังข้อมูลแบบบูรณาการแหล่งแรกของข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนลำดับเบสโปรโมเตอร์ ฐานข้อมูลนี้เป็นแหล่งข้อมูลที่มีประโยชน์ซึ่งสามารถส่งเสริมแนวคิดในการพัฒนาเทคนิคใหม่ในการรักษาโรค ผ่านการควบคุมการแสดงออกของยีนที่เกี่ยวข้องกับการเกิดโรคดังกล่าวได้ในอนาคต และยังเป็นองค์ความรู้ที่สามารถนำมาใช้ในการระบุกลไกทางชีววิทยาที่ไม่ทราบแน่ชัดที่ส่งผลต่อลักษณะฟีโนไทป์ของมนุษย์ นอกจากนี้ การคัดกรองตำแหน่งเป้าหมายที่ตรงกับความสนใจของนักวิจัยแต่ละคนโดยใช้ฐานข้อมูลนี้ก่อนการศึกษาเชิงลึก ยังสามารถช่วยลดเวลาและค่าใช้จ่าย รวมทั้งเพิ่มโอกาสความสำเร็จในการดำเนินงานวิจัย ฐานข้อมูล microPIR เปิดให้ใช้งานได้ที่ <http://www4a.biotec.or.th/micropir> ปัจจุบันอยู่ในระหว่างการพัฒนาฐานข้อมูลเวอร์ชันใหม่ ซึ่งรวมเอาข้อมูลตำแหน่งเป้าหมายในจีโนมหนูและโมเดลการเปรียบเทียบตำแหน่งเป้าหมายระหว่างจีโนมมนุษย์และหนูไว้ด้วย

คำหลัก: ไมโครอาร์เอ็นเอ, ตำแหน่งเป้าหมายบนยีนโปรโมเตอร์, ฐานข้อมูล, มนุษย์

Project Code: TRG5580011

Project Title: Development of an integrated infrastructure for studying microRNA-promoter interactions in human genome

Investigator: Jittima Piriyaongsa, PhD / National Center for Genetic Engineering and Biotechnology (BIOTEC)

E-mail Address: jittima.pir@biotec.or.th

Project Period: 1.5 years (2 July 2012 - 1 January 2014)

This research focused on the construction of an integrated database to facilitate the exploration of a novel microRNA regulatory mechanism through promoter targeting. We developed an open-access web-based database hosting predicted microRNA target sites located within human promoter sequences and their associated genomic features, called microPIR (microRNA-Promoter Interaction Resource). A computational method was used to identify all potential target sites based on the prediction rules created from the previously reported experimental information. Various experimentally confirmed data are presented as supporting information for researchers to appraise the quality of predicted target sites. The database integrates considerable resources of annotated genomic data offering users the facility to extensively explore relationships among target sites and other genomic features. The system is also equipped with the built-in genome browser, which provides a comprehensive view of multidimensional genomic data. All these features make the microPIR database the first comprehensive repository of microRNA promoter target site information. This database could be a valuable resource which provides insights into currently uncharacterized mechanisms for some observed phenotypes and the development of a novel therapeutic technique for various diseases through controlling the expression of disease-associated target genes. The customizable filtering tool assists scientists to identify good target candidates that fit with their interest, thus saving tremendous amount of time and expenses and facilitating the success of experimental validation. The database is freely available at <http://www4a.biotec.or.th/micropir>. The new version of microPIR, which includes the target information for mouse genome and the comparative features of human-mouse target sites, is in development.

Keywords: microRNA, promoter target site, database, human

Executive Summary

ไมโครอาร์เอ็นเอ เป็นโมเลกุลอาร์เอ็นเอสายสั้นๆในร่างกาย ซึ่งทำหน้าที่ควบคุมการทำงานของยีน ผ่านการจับกับตำแหน่งเป้าหมายที่อยู่บริเวณ 3'-UTR ของเมสเซนเจอร์อาร์เอ็นเอของยีน อย่างไรก็ตาม กลไกซึ่งเป็นที่รู้จักกันดีนี้ ไม่สามารถใช้อธิบายการทำงานของไมโครอาร์เอ็นเอได้ในทุกกรณีที่พบ การศึกษาที่ผ่านมาระบุว่า ไมโครอาร์เอ็นเออาจจะสามารถทำงานผ่านกลไกอื่น ผ่านการจับกับลำดับเบสโปรโมเตอร์ของยีนได้ อย่างไรก็ดี ในปัจจุบัน ยังไม่ค่อยมีรายงานการศึกษาในหัวข้อนี้มากนัก ทั้งนี้อาจเนื่องมาจากการทดลองทางห้องปฏิบัติการใช้เวลาและค่าใช้จ่ายค่อนข้างสูง รวมทั้งยังขาดข้อมูลที่มีประสิทธิภาพเพื่อใช้ในการช่วยคัดเลือกตำแหน่งเป้าหมายและกลุ่มยีนที่น่าสนใจจากยีนทั้งหมดกว่าสามหมื่นยีนในร่างกาย เพื่อที่จะนำมาศึกษาต่อไปในเชิงลึก

งานวิจัยนี้มุ่งเน้นการจัดตั้งฐานข้อมูลสาธารณะในรูปแบบเว็บไซต์ โดยมีวัตถุประสงค์เพื่อใช้เป็นแหล่งข้อมูลสำหรับการศึกษาวิจัยทางด้านกลไกการทำงานของไมโครอาร์เอ็นเอผ่านยีนโปรโมเตอร์ โดยเริ่มจากการใช้วิธีทางคอมพิวเตอร์ในการทำนายตำแหน่งการจับของไมโครอาร์เอ็นเอบนยีนโปรโมเตอร์ทั่วจีโนมมนุษย์ ซึ่งอาศัยข้อมูลการทดลองทางห้องปฏิบัติการที่มีมาก่อนในการสร้างหลักการทำนายและประมวลผลด้วยเครื่องคอมพิวเตอร์สมรรถนะสูง ข้อมูลตำแหน่งเป้าหมายที่ได้จากการทำนายดังกล่าวถูกเชื่อมโยงเข้ากับข้อมูลสาธารณะต่างๆที่ได้จากการทดลองทางห้องปฏิบัติการและข้อมูลคุณลักษณะจีโนมอื่นๆที่เกี่ยวข้องจำนวนมาก เพื่อวิเคราะห์และสร้างความสัมพันธ์ ทำให้ฐานข้อมูลนี้เป็นคลังข้อมูลแบบบูรณาการแหล่งแรกของข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนยีนโปรโมเตอร์ ฐานข้อมูลนี้สร้างขึ้นเป็นแหล่งข้อมูลสำคัญสำหรับศึกษาการทำงานของไมโครอาร์เอ็นเอ ซึ่งสามารถนำไปสู่แนวคิดในการพัฒนาเทคนิคใหม่ในการรักษาโรคต่างๆเช่น มะเร็ง ผ่านการควบคุมการแสดงออกของยีนที่เกี่ยวข้องกับการเกิดโรคมะเร็งได้ในอนาคต และยังเป็นองค์ความรู้ที่สามารถนำมาใช้ในการระบุกลไกทางชีววิทยาที่ไม่ทราบแน่ชัดของฟีโนไทป์ในมนุษย์ นอกจากนี้ การคัดกรองตำแหน่งเป้าหมายที่ตรงกับความต้องการของนักวิจัยแต่ละคนก่อนการศึกษาเชิงลึก ยังสามารถช่วยลดเวลาและค่าใช้จ่าย รวมทั้งเพิ่มโอกาสความสำเร็จในการดำเนินงานวิจัย

ผลงานที่ได้จากงานวิจัยในปัจจุบันคือ ฐานข้อมูลชื่อ microPIR (microRNA-Promoter Interaction Resource) ซึ่งเปิดให้ใช้งานฟรีที่ <http://www4a.biotech.or.th/micropir> โดยในปัจจุบันมีจำนวนผู้ใช้งานฐานข้อมูลแล้วประมาณ 3500 คน ผลงานดังกล่าวได้รับการตีพิมพ์ในวารสารนานาชาติ PLoS ONE และได้รับเลือกเป็น Research Highlight บนเว็บไซต์ A-IMBN (Asia-Pacific International Molecular Biology Network) research และยังได้รับรางวัลทุนวิจัย ลอริอัล ประเทศไทย “เพื่อสตรีในงานวิทยาศาสตร์” สาขาวิทยาศาสตร์ชีวภาพ ประจำปี 2556 นอกจากนี้ ยังมีการนำเสนอผลงานนี้ในงานประชุมวิชาการนานาชาติ ปัจจุบันกำลังอยู่ในระหว่างการพัฒนาฐานข้อมูลเวอร์ชัน 2 โดยได้มีการเพิ่มข้อมูลตำแหน่งเป้าหมายของจีโนมหนู และมีการเพิ่มโมดูลในการเปรียบเทียบตำแหน่งเป้าหมายระหว่างมนุษย์และหนู (comparative view) โดยแสดงผลทั้งในรูปแบบรายการผลลัพธ์และภาพกราฟิก เพื่อช่วยในการเปรียบเทียบข้อมูลและความอนุรักษ์ของตำแหน่งเป้าหมาย และยังมีการเพิ่มข้อมูลประกอบอื่นๆ เช่น ข้อมูลความสัมพันธ์ระหว่างไมโครอาร์เอ็นเอกับโรค, ข้อมูลยีนที่มีบรรพบุรุษร่วมกันระหว่างมนุษย์และหนู, ข้อมูล subcellular organization ของไมโครอาร์เอ็นเอที่ได้จากการทดลอง เป็นต้น ทั้งนี้ เพื่อเพิ่มข้อมูลที่เป็นประโยชน์ต่อผู้ใช้ในการคัดเลือกตำแหน่งเป้าหมายที่ตรงกับความต้องการและสมมติฐานที่ตั้งไว้ได้ดียิ่งขึ้น

สารบัญ

	หน้าที่
บทนำ.....	10
วิธีดำเนินการวิจัย.....	12
ฐานข้อมูล microPIR.....	12
การทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ.....	12
แหล่งข้อมูลและการวิเคราะห์ข้อมูลสำหรับฐานข้อมูล.....	14
การออกแบบและสร้างฐานข้อมูล.....	17
การทดสอบการใช้งานฐานข้อมูล.....	19
ฐานข้อมูล microPIR เวอร์ชัน 2.....	23
การทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ.....	23
แหล่งข้อมูลและการวิเคราะห์ข้อมูลสำหรับฐานข้อมูล.....	23
การออกแบบและสร้างฐานข้อมูล.....	25
ผลการวิจัย.....	26
ฐานข้อมูล microPIR.....	26
โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ (Search).....	26
โมดูลสำหรับการแสดงผลแบบกราฟิก (Genome browser).....	27
โมดูลสำหรับการแสดงผลสถิติของฐานข้อมูล (Database statistics).....	29
ฐานข้อมูล microPIR เวอร์ชัน 2.....	30
โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ (Search).....	30
โมดูลสำหรับการเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนู (Human-mouse comparative view).....	33
โมดูลสำหรับการแสดงผลแบบกราฟิก (Genome Browser).....	33
โมดูลสำหรับการแสดงผลสถิติของฐานข้อมูล (Database statistics).....	34
โมดูลการดาวน์โหลดข้อมูล (Download)	34
บทวิจารณ์.....	36
บรรณานุกรม.....	40
ผลงานที่ได้จากโครงการ.....	44
ภาคผนวก.....	46

สารบัญตาราง

	หน้า
ตารางที่ 1 ข้อมูลสารธารณะที่รวมอยู่ในฐานข้อมูล microPIR.....	15
ตารางที่ 2 รายการของตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนลำดับเบสโปรโมเตอร์ ซึ่งได้รับการรายงานมาก่อนในการศึกษาอื่น.....	19
ตารางที่ 3 ข้อมูลสารธารณะของคุณลักษณะจีโนมที่อยู่ในฐานข้อมูล micorPIR เวอร์ชัน 2.....	25

สารบัญภาพ

หน้าที่

รูปที่ 1 ตัวอย่างค่าความถี่ไดนิวคลีโอไทด์ของลำดับเบสบนโปรโมเตอร์แสดงในไฟล์ที่ใช้เป็นอินพุตสำหรับโปรแกรม RNAhybrid.....	13
รูปที่ 2 ขั้นตอนการหาค่า location และ scale parameters.....	13
รูปที่ 3 ตัวอย่างผลลัพธ์ของตำแหน่งเป้าหมายที่ได้จากการทำนาย.....	14
รูปที่ 4 ลักษณะการคาบเกี่ยวกันระหว่างตำแหน่งเป้าหมาย (ts) และคุณลักษณะอื่นๆของจีโนม (a).....	17
รูปที่ 5 ภาพรวมการออกแบบฐานข้อมูล microPIR แสดงถึงแหล่งข้อมูลและลักษณะของเว็บไซต์ที่ติดต่อกับผู้ใช้.....	18
รูปที่ 6 ผลลัพธ์การค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอในยีนต่างๆซึ่งได้รับการรายงานมาก่อน.....	20
รูปที่ 7 ขั้นตอนสำหรับกระบวนการวิเคราะห์ข้อมูลของ microPIR เวอร์ชัน 2.....	24
รูปที่ 8 ภาพรวมการออกแบบฐานข้อมูล microPIR เวอร์ชัน 2.....	26
รูปที่ 9 โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ.....	28
รูปที่ 10 ตัวอย่างการแสดงผลลัพธ์ของโมดูลการแสดงผลแบบกราฟิกบนจีโนมเบราร์เซอร์.....	29
รูปที่ 11 ตัวอย่างข้อมูลสถิติในโมดูลสำหรับการแสดงผลสถิติของฐานข้อมูล	30
รูปที่ 12 หน้าเพจการกรอกอินพุตสำหรับการค้นหาตำแหน่งเป้าหมายของ microPIR เวอร์ชัน 2.....	31
รูปที่ 13 หน้าเพจสำหรับโมดูลการค้นหาตำแหน่งเป้าหมายของ microPIR เวอร์ชัน 2.....	32
รูปที่ 14 ส่วนแสดงผลลัพธ์ตามชื่อยีนซึ่งพัฒนาขึ้นมาใหม่ใน microPIR เวอร์ชัน 2.....	32
รูปที่ 15 ส่วนแสดงผลลัพธ์ของรายการตำแหน่งเป้าหมายบนยีนเป้าหมายแต่ละตัว.....	33
รูปที่ 16 หน้าเพจแสดงผลลัพธ์การค้นหาตำแหน่งเป้าหมายเปรียบเทียบระหว่างมนุษย์และหนู.....	34
รูปที่ 17 ภาพกราฟิกแสดงการเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนูในยีนที่มีบรรพบุรุษร่วมกัน	35
รูปที่ 18 รายละเอียดของตำแหน่งเป้าหมายบนหน้ากราฟิกแสดงการเปรียบเทียบระหว่างมนุษย์และหนู.....	35
รูปที่ 19 ตัวอย่างการแสดงผลลัพธ์ของโมดูลการแสดงผลแบบกราฟิกบนจีโนมเบราร์เซอร์ของ microPIR เวอร์ชัน 2.....	36

คำอธิบายสัญลักษณ์และคำย่อที่ใช้ในการวิจัย

3'-UTR	3' Untranslated Region
5'-UTR	5' Untranslated Region
AGO	Argonaute
<i>CDH1</i>	cadherin 1
CDS	Coding DNA Sequence
CLIP	Cross-Linking and Immunoprecipitation
<i>CSDC2</i>	cold shock domain containing C2
EST	Expressed Sequence Tag
MFE	Minimum Free Energy
miRNP	microribonucleoprotein
<i>PGR</i>	progesterone receptor
<i>POLR3D</i>	polymerase (RNA) III (DNA directed) polypeptide D, 44kDa
SNP	Single Nucleotide Polymorphism
SVG	Scalable Vector Graphics
TFBS	Transcription Factor Binding Site

บทนำ

ไมโครอาร์เอ็นเอ เป็นอาร์เอ็นเอขนาดเล็กประมาณ 22 นิวคลีโอไทด์ ที่ทำหน้าที่ควบคุมการแสดงออกของยีนหลังกระบวนการคัดลอกยีน (post-transcriptional regulation) ผ่านการจับกับตำแหน่งเป้าหมายบนยีน โดยจะส่งผลทำให้เกิดการยับยั้งกระบวนการสังเคราะห์โปรตีน หรือเกิดการทำลายเมสเซนเจอร์อาร์เอ็นเอของยีนเป้าหมาย (BARTEL 2004; ENGELS and HUTVAGNER 2006; RUVKUN 2008) ไมโครอาร์เอ็นเอมีบทบาทสำคัญในการควบคุมกระบวนการทางชีววิทยาหลายกระบวนการ เช่น กระบวนการพัฒนาของสิ่งมีชีวิต (development), การเพิ่มจำนวนและการเปลี่ยนแปลงของเซลล์เพื่อไปทำหน้าที่ต่างๆ (cell proliferation and differentiation), เมตาบอลิซึม (metabolism), การตายของเซลล์ (apoptosis), และวัฏจักรของเซลล์ (cell cycle) ทำให้ไมโครอาร์เอ็นเอมีความสำคัญทั้งทางด้านชีววิทยาและการแพทย์ โดยพบว่าความผิดปกติในการทำงานของไมโครอาร์เอ็นเอมีความเกี่ยวข้องกับโรคต่างๆ เช่น มะเร็ง อีกด้วย

โดยทั่วไป การระบุหน้าที่ของไมโครอาร์เอ็นเอแต่ละตัวนั้นจะพิจารณาจากชนิดของยีนเป้าหมายและผลของไมโครอาร์เอ็นเอที่มีต่อการแสดงออกของยีนนั้น ดังนั้นการระบุเป้าหมายของไมโครอาร์เอ็นเอจึงมีความสำคัญอย่างยิ่ง โดยการระบุเป้าหมายในระดับจีโนมนั้นต้องอาศัยการทำนายด้วยวิธีทางคอมพิวเตอร์ และการวิเคราะห์ทางชีวสารสนเทศเป็นหลัก วิธีการการทำนายที่มีความถูกต้องสูงจึงเป็นเครื่องมือที่สำคัญที่ช่วยให้การระบุหน้าที่ของไมโครอาร์เอ็นเอที่มีอยู่ในจีโนมเป็นไปได้อย่างรวดเร็วยิ่งขึ้น

ในสิ่งมีชีวิตจำพวกสัตว์นั้น เป็นที่เข้าใจกันว่า ไมโครอาร์เอ็นเอจะควบคุมการแสดงออกของยีนผ่านการจับกับตำแหน่งเป้าหมายที่อยู่ในบริเวณ 3'-UTR ของเมสเซนเจอร์อาร์เอ็นเอซึ่งมีลำดับเบสที่เป็นคู่สมกับไมโครอาร์เอ็นเอ (AMBROS 2008; BARTEL 2009) ดังนั้น การทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอจึงมักจำกัดอยู่ในบริเวณ 3'-UTR นี้ ถึงแม้ว่า จะมีการรายงานความเป็นไปได้ที่ไมโครอาร์เอ็นเอจะจับกับลำดับเบสบริเวณอื่นด้วยเช่น 5'-UTR และ CDS (FORMAN *et al.* 2008; LYTLE *et al.* 2007; TAY *et al.* 2008) อย่างไรก็ตาม ข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอที่มีอยู่ในปัจจุบันยังไม่สามารถใช้อธิบายกลไกและหน้าที่ของไมโครอาร์เอ็นเอได้อย่างสมบูรณ์ โดยเฉพาะไมโครอาร์เอ็นเอที่ยังไม่สามารถระบุตำแหน่งเป้าหมายได้

การศึกษาที่ผ่านมาได้แสดงให้เห็นว่า อาร์เอ็นเอสายคู่ขนาดเล็กซึ่งมีลำดับเบสที่เป็นคู่สมกับยีนโปรโมเตอร์สามารถทำหน้าที่เป็นตัวยับยั้งหรือตัวกระตุ้นการแสดงออกของยีนได้ (HAWKINS *et al.* 2009; HUANG *et al.* 2010; JANOWSKI *et al.* 2005; JANOWSKI *et al.* 2007; LI *et al.* 2006; MORRIS *et al.* 2004; NAPOLI *et al.* 2009; SUZUKI *et al.* 2005; TING *et al.* 2005) โดยอาร์เอ็นเอขนาดเล็กดังกล่าวสามารถจับกับอาร์เอ็นเอที่ไม่ถอดรหัสเป็นโปรตีน (noncoding RNA transcript) ที่มีส่วนประกอบของลำดับเบสโปรโมเตอร์ (HAN *et al.* 2007; MORRIS 2009; SCHWARTZ *et al.* 2008) และมีการนำส่งโปรตีน AGO ไปจับยังโปรโมเตอร์เป้าหมาย (CHU *et al.* 2010; JANOWSKI *et al.* 2006; KIM *et al.* 2006) โปรตีน AGO นั้นมีบทบาทสำคัญในกลไกการทำงานของไมโครอาร์เอ็นเอ โดยเป็นส่วนประกอบหลักของ miRNP (microribonucleoprotein) ซึ่งเป็นคอมเพล็กซ์ระหว่างไมโครอาร์เอ็นเอและโปรตีนต่างๆ ที่จะไปจับกับตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ (FILIPOWICZ *et al.* 2008) การศึกษาที่ผ่านมาเหล่านี้แสดงให้เห็นถึงการทำงานของกลไกในร่างกาย ที่อาร์เอ็นเอขนาดเล็กสามารถจับกับยีนโปรโมเตอร์ได้ ซึ่งอาร์เอ็นเอขนาดเล็กดังกล่าวอาจรวมไปถึงไมโคร

อาร์เอ็นเอด้วย ผลการศึกษาทางห้องปฏิบัติการเมื่อไม่นานมานี้ ได้นำเสนอกกลไกใหม่ของไมโครอาร์เอ็นเอในการควบคุมการแสดงออกของยีนผ่านการจับกับตำแหน่งเป้าหมายที่อยู่บริเวณโปรโมเตอร์ในเซลล์มนุษย์ โดย Place et al. (PLACE et al. 2008) นำเสนอหลัก ฐานจากไมโครอาร์เอ็นเอ miR-373 ซึ่งมีลำดับเบสที่เป็นคู่สมกับลำดับเบสโปรโมเตอร์ของยีน *CDH1* และ *CSDC2* โดย miR-373 สามารถจับกับลำดับเบสโปรโมเตอร์และส่งผลกระทบต่อแสดงออกของยีนดังกล่าว โดยทำให้ระดับการแสดงออกของยีนสูงขึ้น ในการศึกษาของ Kim et al. (Kim et al. 2008) ก็ได้รายงานถึงบทบาทของ miR-320 ในการจับกับตำแหน่งเป้าหมายที่อยู่บริเวณเดียวกันกับตำแหน่งบนจีโนมของตัวมันเอง (cis-regulation) โดย miR-320 นั้นอยู่ในบริเวณโปรโมเตอร์ของยีน *POLR3D* การจับของ miR-320 กับตำแหน่งเป้าหมายบนโปรโมเตอร์ส่งผลให้การแสดงออกของยีน *POLR3D* ลดลง โดยพบว่า กลไกนี้อาศัยโปรตีน AGO1 นอกจากนี้ รายงานล่าสุดโดย Younger และ Corey (YOUNGER and COREY 2011) แสดงให้เห็นว่า ไมโครอาร์เอ็นเอสามารถยับยั้งการแสดงออกของยีนโดยจับกับลำดับเบสโปรโมเตอร์ของยีนที่อยู่ห่างออกไปได้ (trans-regulation) รวมทั้งยังแสดงให้เห็นว่า การจับของไมโครอาร์เอ็นเอที่มีลำดับเบสเป็นคู่สมแบบไม่สมบูรณ์กับตำแหน่งเป้าหมายต้องอาศัยโปรตีน AGO2 แทนที่จะเป็นโปรตีน AGO1 การศึกษาเหล่านี้ชี้ให้เห็นถึงความเป็นไปได้ของการใช้กลไกที่แตกต่างกันในการยับยั้งการแสดงออกของยีนสำหรับแต่ละคู่ของไมโครอาร์เอ็นเอและยีนเป้าหมาย นอกจากนี้ ยังแสดงให้เห็นว่า การควบคุมกระบวนการคัดลอกยีนโดยการจับกับตำแหน่งเป้าหมายบนลำดับเบสโปรโมเตอร์ อาจจะเป็นกลไกทั่วไปอีกกลไกหนึ่งที่ไมโครอาร์เอ็นเอใช้ในการควบคุมการทำงานของยีน โดยการจับของไมโครอาร์เอ็นเอกับตำแหน่งเป้าหมายบนโปรโมเตอร์อาจส่งผลเพิ่มการแสดงออกของยีน (PLACE et al. 2008) ในลักษณะที่คล้ายกับกระบวนการ RNA activation (RNAa) ซึ่งพบในการศึกษาที่ใช้อาร์เอ็นเอสายคู่ที่สังเคราะห์ขึ้นมาให้จับกับยีนโปรโมเตอร์ หรือส่งผลยับยั้งการแสดงออกของยีนด้วยกระบวนการ Transcriptional Gene Silencing (TGS) (JANOWSKI et al. 2007; Li et al. 2006) โดยปัจจัยและกลไกที่กำหนดทางเลือกของกระบวนการดังกล่าว นั้น ยังไม่ทราบแน่ชัดและยังไม่มีการศึกษามากนัก ความเข้าใจในกลไกการทำงานนี้ของไมโครอาร์เอ็นเออาจนำไปสู่การพัฒนาแนวทางใหม่ในการรักษาโรค ผ่านการควบคุมการแสดงออกของยีนเป้าหมายที่มีความเกี่ยวข้องกับโรค รวมทั้งการปรับเปลี่ยนลักษณะฟีโนไทป์ต่างๆ โดยที่รูปแบบการแสดงออกของยีนที่เกี่ยวข้องกับฟีโนไทป์และรูปแบบการแสดงออกของไมโครอาร์เอ็นเอที่สามารถจับกับโปรโมเตอร์ของยีนนั้นมีความสัมพันธ์กัน

ในปัจจุบัน ยังมีเพียงไม่กี่ปฏิสัมพันธ์ระหว่างไมโครอาร์เอ็นเอและยีนโปรโมเตอร์ที่ได้รับการยืนยันจากการทดลองทางห้องปฏิบัติการ ทั้งนี้อาจเนื่องมาจากการทดสอบทางห้องปฏิบัติการใช้เวลาและค่าใช้จ่ายค่อนข้างสูง รวมทั้งยังขาดข้อมูลที่มีประสิทธิภาพเพื่อใช้ในการช่วยคัดเลือกตำแหน่งเป้าหมายและกลุ่มยีนที่น่าสนใจจากยีนทั้งหมดในร่างกาย เพื่อที่จะนำมาศึกษาต่อไปในเชิงลึกทางห้องปฏิบัติการ ดังนั้น การจัดตั้งแหล่งข้อมูลสาธารณะของข้อมูลตำแหน่งเป้าหมายที่เป็นไปได้ทั้งหมดบนยีนโปรโมเตอร์ ซึ่งได้จากการทำนายด้วยวิธีทางคอมพิวเตอร์ในระดับจีโนม จะสามารถช่วยสนับสนุนงานวิจัยทางด้านนี้ให้เป็นไปอย่างรวดเร็วยิ่งขึ้น โดยเฉพาะอย่างยิ่ง เมื่อข้อมูลการทดลองในเรื่องนี้ยังมีย่อยอย่างจำกัด ถึงแม้ว่าในปัจจุบันจะมีฐานข้อมูลของตำแหน่งเป้าหมายของไมโครอาร์เอ็นเออยู่จำนวนหนึ่งแล้ว เช่น ฐานข้อมูล miRDB (WANG 2008), miRNomeMap (Hsu et al. 2008), miRigator (NAM et al. 2008) และ miRecords (XIAO et al. 2009) เป็นต้น แต่ฐานข้อมูลเหล่านี้ยังไม่มีข้อมูลตำแหน่งเป้าหมายที่อยู่บนยีนโปรโมเตอร์ จากการสำรวจ พบว่า มีเพียงฐานข้อมูลที่พัฒนาเมื่อไม่นานมานี้ที่ชื่อว่า miRWalk (DWEET et al. 2011) ที่มีข้อมูลดังกล่าว อย่างไรก็ตาม ข้อมูลตำแหน่งเป้าหมายที่มีอยู่เป็นเพียงตำแหน่งเป้าหมายที่อยู่บนสายดีเอ็นเอที่มีการถอดรหัสพันธุกรรม (sense

strand) เท่านั้น รวมทั้งไม่มีการให้ข้อมูลอื่นๆซึ่งช่วยในการคัดกรองตำแหน่งเป้าหมายที่ตรงกับความสนใจ เพื่อนำไปศึกษาทางห้องปฏิบัติการต่อไป

ในโครงการนี้ คณะผู้วิจัยได้พัฒนาฐานข้อมูลสาธารณะแบบบูรณาการแห่งแรกสำหรับข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนยีนโปรโมเตอร์ เพื่อสนับสนุนการศึกษาวิจัยในหัวข้อดังกล่าวในจีโนมมนุษย์ ลักษณะความแตกต่างอีกอย่างหนึ่งของฐานข้อมูลนี้กับฐานข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอทั่วไปที่มีอยู่คือ มีการจัดการข้อมูลตำแหน่งเป้าหมายที่ได้จากการทำนายให้อยู่ในลักษณะที่ผู้ใช้สามารถสืบค้นได้อย่างยืดหยุ่นเพื่อให้ตรงกับสมมติฐานของผู้ใช้มากที่สุด รวมทั้งการมีภาพกราฟิกที่แสดงความสัมพันธ์ระหว่างตำแหน่งเป้าหมายกับคุณลักษณะจีโนมอื่นๆได้

วิธีดำเนินการวิจัย

ฐานข้อมูล microPIR

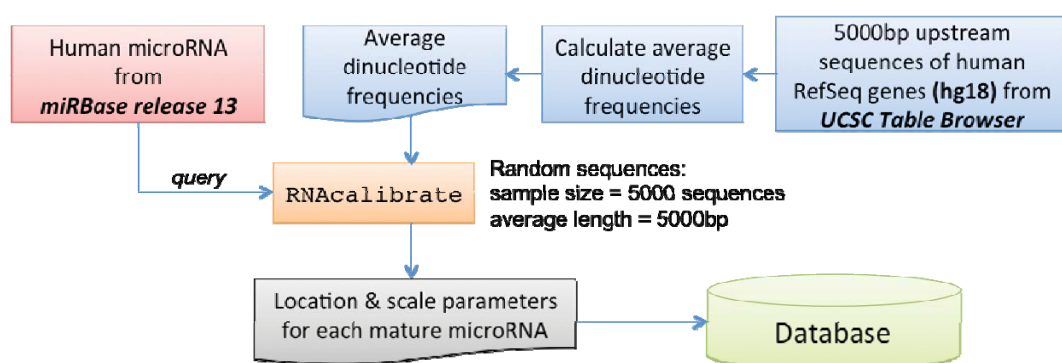
การทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ

ในขั้นตอนนี้จะใช้โปรแกรม RNAhybrid (KRUGER and REHMSMEIER 2006) version 2.1 ในการทำนายตำแหน่งเป้าหมายบนลำดับเบสโปรโมเตอร์ทั่วจีโนมมนุษย์ โดยโปรแกรมจะทำการจับคู่เบสที่ดีที่สุดของไมโครอาร์เอ็นเอกับลำดับเบสเป้าหมาย และจะคำนวณหาค่าพลังงานอิสระที่น้อยที่สุด (MFE: Minimum Free Energy) และค่า p -value ของตำแหน่งเป้าหมายที่ทำนายได้ รวมทั้งตำแหน่งการปฏิสัมพันธ์และรูปแบบการจับคู่กันระหว่างลำดับเบสเป้าหมายและไมโครอาร์เอ็นเอ โดยการทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอนั้น อาศัยข้อมูลลำดับเบสโปรโมเตอร์ที่ได้จาก UCSC Table Browser (KAROLCHIK *et al.* 2004) โดยเลือกใช้ลำดับเบส 5000 bp ก่อนตำแหน่งเริ่มต้นคัดลอกยีน (transcription start site) ของ RefSeq gene ทั่วจีโนมมนุษย์ (version hg18) ส่วนข้อมูลไมโครอาร์เอ็นเอทั้งหมดของจีโนมมนุษย์ได้มาจากฐานข้อมูล miRBase (GRIFFITHS-JONES *et al.* 2008) release 13

ก่อนการรันโปรแกรมจะมีการเตรียมข้อมูล 2 ชุด ชุดแรกคือ ไฟล์ความถี่ไดนิวคลีโอไทด์ (dinucleotide frequencies) ของลำดับเบสโปรโมเตอร์ทั้งจีโนม ตัวอย่างของไฟล์ความถี่ของไดนิวคลีโอไทด์แสดงในรูปที่ 1 และข้อมูลชุดที่สอง คือ ค่า location และ scale parameters สำหรับไมโครอาร์เอ็นเอแต่ละตัว ซึ่งใช้สำหรับช่วยในการแปลงค่า MFE ที่คำนวณได้จาก RNAhybrid ให้เป็นค่ามาตรฐานที่ถูกต้อง โดยการคำนวณค่าดังกล่าวทำโดยการสร้างกลุ่มของลำดับเบสแบบสุ่ม (random sequences) ที่มีค่าต่างๆตามที่ผู้ใช้กำหนดคือ จำนวนตัวอย่างลำดับเบส (sample size), พารามิเตอร์การกระจายตัวของความยาวลำดับเบส (length distribution parameters) และความถี่ของไดนิวคลีโอไทด์ (dinucleotide frequency) และหากการกระจายตัวของค่า MFE ของลำดับเบสแบบสุ่มดังกล่าว ค่า location และ scale parameters ที่คำนวณได้จากลำดับเบสแบบสุ่ม จะนำไปใช้เพื่อคำนวณค่า p -value ของ MFE ของตำแหน่งเป้าหมายที่ทำนายได้ต่อไป รูปที่ 2 แสดงขั้นตอนในการหาค่า location และ scale parameters

```
# average dinucleotide frequencies of all antisense promoter sequences [promoter-as.fa]
AA: 0.08240681
AC: 0.05003577
AG: 0.07305698
AT: 0.06096324
CA: 0.07201229
CC: 0.06809808
CG: 0.020398255
CT: 0.07274926
GA: 0.06025722
GC: 0.0551597
GG: 0.06797777
GT: 0.05012143
TA: 0.051799558
TC: 0.05996681
TG: 0.072040774
TT: 0.082519904
```

รูปที่ 1 ตัวอย่างค่าความถี่ไดนิวคลีโอไทด์ของลำดับเบสบนโปรโมเตอร์แสดงในไฟล์ที่ใช้เป็นอินพุตสำหรับโปรแกรม RNAhybrid

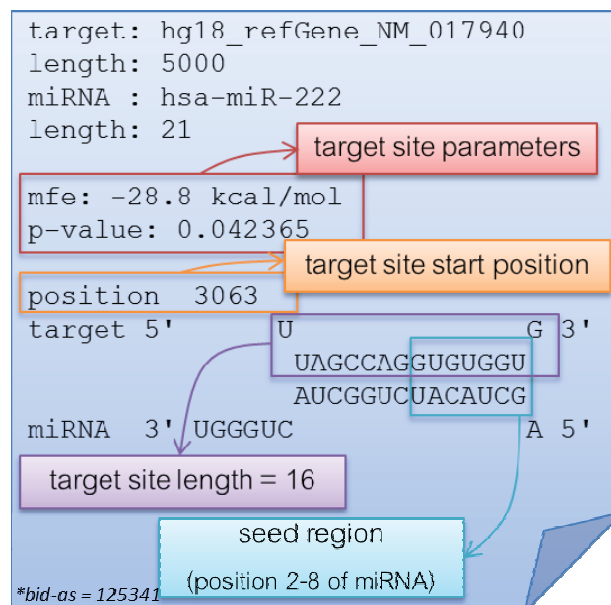


รูปที่ 2 ขั้นตอนการหาค่า location และ scale parameters

การคำนวณค่าพารามิเตอร์ทั้ง 2 ค่าของไมโครอาร์เอ็นเอแต่ละตัวนั้น ได้กำหนดจำนวนลำดับเบสแบบสุ่มที่สร้างเป็น 5000 ตามค่ามาตรฐานที่ตั้งไว้ และค่าความยาวเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของลำดับเบสแบบสุ่มเป็น 5000 bp และ 0 bp ตามลำดับ ซึ่งค่า location และ scale parameters ของไมโครอาร์เอ็นเอแต่ละตัวจะถูกจัดเก็บลงในฐานข้อมูลเพื่อให้เรียกใช้งานได้ง่ายขึ้นในการทำนายตำแหน่งเป้าหมาย โดยในการทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ คณะผู้วิจัยได้กำหนดคุณสมบัติของตำแหน่งเป้าหมายที่ต้องการคือ เป้าหมายต้องจับกับตำแหน่งที่ 2-8 ของไมโครอาร์เอ็นเอซึ่งเป็นบริเวณ seed sequence ตามที่ได้มีรายงานไว้ โดยจะหาตำแหน่งเป้าหมายที่อยู่บริเวณโปรโมเตอร์บนดีเอ็นเอทั้งสองสาย คือ บริเวณลำดับเบส 5000 เบสก่อนตำแหน่งเริ่มต้นของยีน (sense promoter) และลำดับเบสในตำแหน่งเดียวกันแต่อยู่บนดีเอ็นเออีกสายหนึ่ง (antisense promoter) รวมทั้งพลังงานในการจับระหว่างไมโครอาร์เอ็นเอและตำแหน่งเป้าหมาย ≤ -20 Kcal/mol และความยาวของตำแหน่งเป้าหมาย \leq ความยาวของไมโครอาร์เอ็นเอ + 10 ตัวอย่างของผลลัพธ์จากการทำนายดังแสดงในรูปที่ 3 จากรูปที่ 3 ผลลัพธ์จะอธิบายรายละเอียดของไมโครอาร์เอ็นเอและเป้าหมาย ได้แก่ ค่า MFE, *p*-value, ตำแหน่งเริ่มต้นของเป้าหมาย (target start position) และรูปของการจับกันระหว่างไมโครอาร์เอ็นเอและเป้าหมาย โดยที่ตำแหน่งเริ่มต้นของเป้าหมายและรูปของการจับกับเป้าหมายจะถูกนำไปคำนวณหาตำแหน่งที่แท้จริงของเป้าหมายบนจีโนมต่อไป

แหล่งข้อมูลและการวิเคราะห์ข้อมูลสำหรับฐานข้อมูล

ข้อมูลหลักที่มีอยู่ในฐานข้อมูล microPIR คือ ตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอที่ได้จากการทำนายด้วยกระบวนการทางคอมพิวเตอร์ ส่วนข้อมูลอื่นๆ ได้แก่ ข้อมูลคุณลักษณะต่างๆของจีโนมมนุษย์ที่เกี่ยวข้อง โดยกระบวนการวิเคราะห์และจัดเตรียมข้อมูลสำหรับฐานข้อมูลประกอบไปด้วยขั้นตอนต่างๆ ได้แก่



รูปที่ 3 ตัวอย่างผลลัพธ์ของตำแหน่งเป้าหมายที่ได้จากการทำนาย

1) การเตรียมข้อมูลสำหรับวิเคราะห์

คณะผู้วิจัยได้รวมเอาข้อมูลสาธารณะของคุณลักษณะจีโนมที่สำคัญของจีโนมมนุษย์ดังแสดงในตารางที่ 1 มาไว้ในฐานข้อมูล microPIR สำหรับเป็นข้อมูลสนับสนุนในการพิจารณาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอที่น่าสนใจ โดยได้จัดเก็บข้อมูลตำแหน่งเริ่มต้นและตำแหน่งสิ้นสุดบนจีโนมของคุณลักษณะต่างๆ เหล่านี้และข้อมูลที่สำคัญอื่นๆ ลงในฐานข้อมูล MySQL เพื่อให้ง่ายในการค้นหาและนำไปตรวจสอบการคาบเกี่ยวกันระหว่างตำแหน่งเหล่านี้กับตำแหน่งเป้าหมายที่ทำนายได้ต่อไป

2) การประมวลผลและระบุตำแหน่งเป้าหมายบนจีโนม

ตำแหน่งเป้าหมายที่ได้จากการทำนายเป็นตำแหน่งบนลำดับเบสโปรโมเตอร์ ซึ่งตำแหน่งดังกล่าวยังไม่ใช่ตำแหน่งที่แท้จริงบนจีโนม เมื่อทราบจุดเริ่มต้น (promoter_start) และจุดสิ้นสุด (promoter_end) ของลำดับเบสโปรโมเตอร์บนจีโนม, ทิศทางของลำดับเบส (strand) และความยาวของตำแหน่งเป้าหมาย (length) สามารถหาตำแหน่งเริ่มต้น (genome_start) และตำแหน่งสิ้นสุด (genome_end) ของเป้าหมายบนจีโนมจากอัลกอริทึมดังนี้

```
if strand = forward then
    genome_start = promoter_start + target_site_start - 1
    genome_end   = genome_start + length - 1
else
    genome_end   = promoter_end - target_site_start + 1
    genome_start = genome_end - length + 1
```

ตารางที่ 1 ข้อมูลสาธารณะที่รวมอยู่ในฐานข้อมูล microPIR

Type	Public database	Version
Human microRNA	miRBase (GRIFFITHS-JONES <i>et al.</i> 2008)	release 13
AGO binding sites	CLIPZ (KHORSHID <i>et al.</i> 2010)	-
Human gene upstream sequences	UCSC Table Browser (KAROLCHIK <i>et al.</i> 2004)	hg18 (build 36)
Transcription Factor Binding Site (TFBS)		
GenBank human EST		
Transposable Element (TE)		
CpG Island		
Conservation score of 17 vertebrate species	UCSC Genome Browser (FUJITA <i>et al.</i> 2011)	Mar 2006 (hg18)
SNP	NCBI dbSNP (SHERRY <i>et al.</i> 2001)	build 129
	HapMap (CONSORTIUM 2005)	release 27
	JSNP (HIRAKAWA <i>et al.</i> 2002)	release 35
	ThaiSNP	release 2

ตัวอย่างเช่น ยีน NBPF1 (NM_017940) ที่อยู่บนโครโมโซม 1 มีตำแหน่งเริ่มต้นของโปรโมเตอร์บนจีโนม (promoter_start) คือ 16812570 มีตำแหน่งเป้าหมายอยู่บน forward strand จากผลลัพธ์ของตำแหน่งเป้าหมายในรูปที่ 3 มีจุดเริ่มต้นของตำแหน่งเป้าหมาย (target_start_position) คือ 3063 และมีความยาวของตำแหน่งเป้าหมาย (length) คือ 16 เบส เราสามารถหาตำแหน่งเริ่มต้น (genome_start) และตำแหน่งสิ้นสุด (genome_end) ของตำแหน่งเป้าหมายบนจีโนมได้ดังนี้

```
genome_start = 16812570 + 3063 - 1 = 16815632
genome_end   = 16815632 + 16 - 1   = 16815647
```

ดังนั้นจุดเริ่มต้นและจุดสิ้นสุดของตำแหน่งเป้าหมายบนโครโมโซม 1 คือ 16815632 - 16815647

การประมวลผลในขั้นตอนนี้ใช้โปรแกรมที่เขียนโดย Java version 1.6 เพื่อหาตำแหน่งของเป้าหมายบนจีโนม จากนั้นจะนำข้อมูลตำแหน่งบนจีโนมของเป้าหมายทั้งหมดจัดเก็บลงในฐานข้อมูล รวมทั้งนำไปแปลงให้อยู่ในรูปแบบไฟล์ที่สามารถแสดงผลในรูปแบบกราฟิกบนจีโนมเบราว์เซอร์ (Genome Browser) ต่อไป

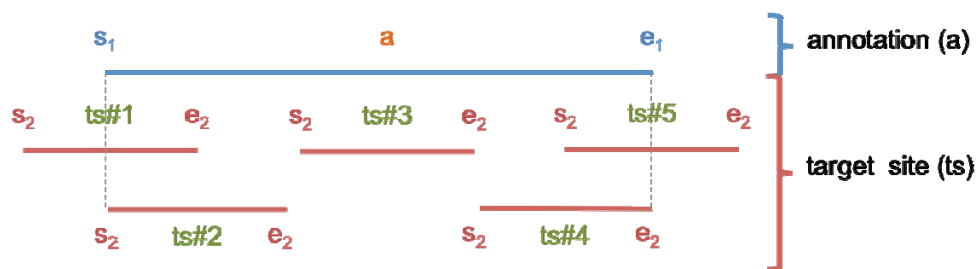
3) การตรวจสอบการคาบเกี่ยวกันของตำแหน่งเป้าหมายกับคุณลักษณะจีโนมอื่นๆ

ขั้นตอนนี้เป็นการนำข้อมูลตำแหน่งบนจีโนมของคุณลักษณะต่างๆ ที่เก็บไว้ในขั้นตอนที่ 1) และตำแหน่งบนจีโนมของตำแหน่งเป้าหมายที่เก็บไว้ในขั้นตอนที่ 2) มาตรวจสอบการคาบเกี่ยวกันของตำแหน่ง ซึ่งความเป็นไปได้ของการคาบเกี่ยวกันระหว่างตำแหน่งเป้าหมายและคุณลักษณะอื่นๆของจีโนมแสดงในรูปที่ 4 โดยสามารถเขียนแสดงอัลกอริทึมแทนรูปที่ 4 ได้ดังนี้

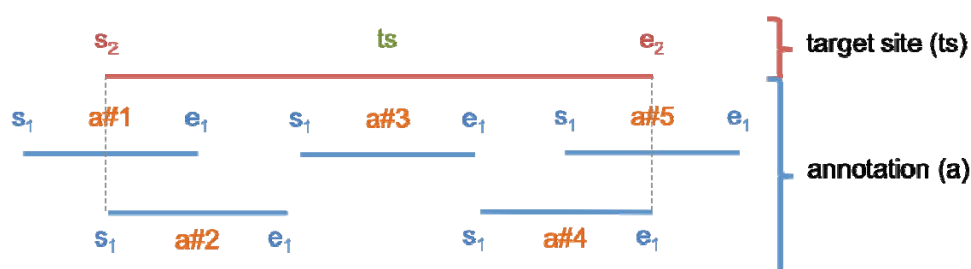
Define:

```
s1, e1: start position of genomic feature, end position of genomic feature
s2, e2: start position of target site, end position of target site
if target_site_length <= genomic_feature_length then
    if s2 >= s1 && e2 <= e1 then
        return true
    else
        if s1 >= s2 && e1 <= e2 then
            return true
```

ยกตัวอย่างในรูปที่ 4(A) ts#2 มีจุดเริ่มต้น (s₂) ที่พอดีกับจุดเริ่มต้นของคุณลักษณะจีโนม (s₁) และมีจุดสิ้นสุด (e₂) อยู่ไม่เกินจุดสิ้นสุดของคุณลักษณะจีโนม (e₁) ดังนั้น ts#2 จึงมีตำแหน่งคาบเกี่ยวกันกับคุณลักษณะจีโนมที่นำมาพิจารณา โดยจะคำนวณการคาบเกี่ยวกันนี้สำหรับคุณลักษณะจีโนมทุกชนิดในตารางที่ 1 ผลลัพธ์ทั้งหมดที่ได้จะถูกเก็บลงในฐานข้อมูล



(A)



(B)

รูปที่ 4 ลักษณะการคาบเกี่ยวกันระหว่างตำแหน่งเป้าหมาย (ts) และคุณลักษณะอื่นๆของจีโนม (a)

(A) ในกรณีที่ขนาดของคุณลักษณะจีโนมใหญ่กว่าขนาดของตำแหน่งเป้าหมาย

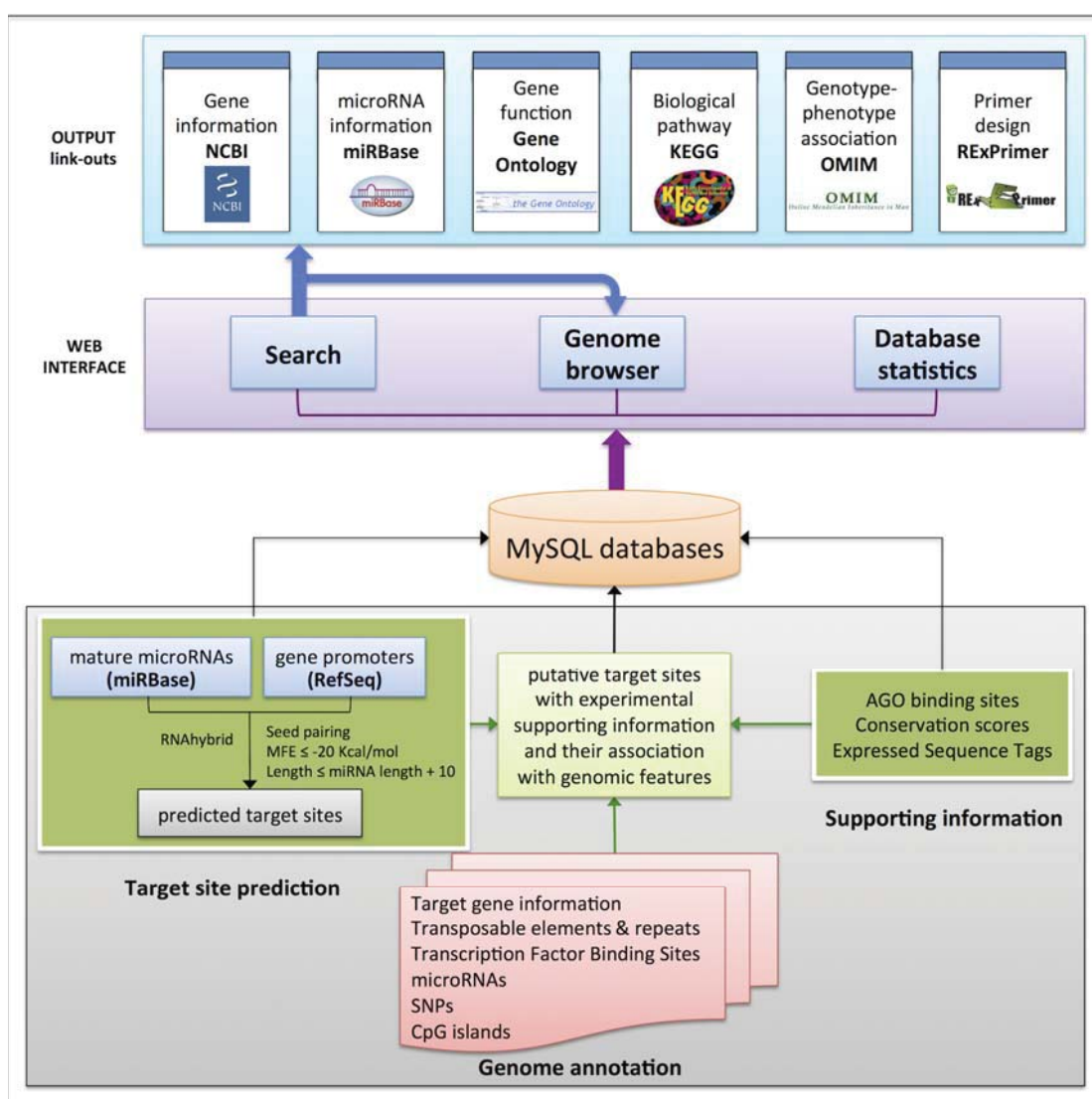
(B) ในกรณีที่ขนาดของตำแหน่งเป้าหมายใหญ่กว่าขนาดของคุณลักษณะจีโนม

โดย s_1 และ e_1 คือ ตำแหน่งเริ่มต้นและตำแหน่งสิ้นสุดของคุณลักษณะจีโนม, s_2 และ e_2 คือ ตำแหน่งเริ่มต้นและตำแหน่งสิ้นสุดของตำแหน่งเป้าหมาย

การออกแบบและสร้างฐานข้อมูล

ภาพรวมการออกแบบฐานข้อมูล microPIR แสดงในรูปที่ 5 ซึ่งในส่วนล่างของรูปแสดงถึงข้อมูล 3 ชนิด ได้แก่ ตำแหน่งเป้าหมายที่ได้จากการทำนาย (Target site prediction), ข้อมูลคุณลักษณะจีโนมที่ช่วยในการพิจารณาคุณภาพของตำแหน่งเป้าหมายที่ทำนายได้ (Supporting information), และข้อมูลคุณลักษณะจีโนมอื่นๆที่เกี่ยวข้อง (Genome annotation) ข้อมูลเหล่านี้จะถูกวิเคราะห์ รวบรวม และจัดเก็บลงในฐานข้อมูล MySQL version 5.5.1 จากนั้นเพื่อให้ผู้ใช้ทั่วไปสามารถเข้าถึงฐานข้อมูลได้ จึงพัฒนาส่วนติดต่อผู้ใช้กับฐานข้อมูล (web interface) รวมทั้งการแสดงผลแบบกราฟิกด้วยโปรแกรมที่เขียนโดยภาษา Python และ toolkits จาก Python Webware (<http://www.webwareforpython.org>) ในส่วนติดต่อกับผู้ใช้ผ่านเว็บ (web interface) นั้นออกแบบให้มีฟังก์ชัน 3 อย่าง อย่างแรก คือ ฟังก์ชันในการค้นหาตำแหน่งเป้าหมาย อินพุตจะถูกส่งผ่านเว็บ ซึ่งจะเปลี่ยนเป็น SQL query ไปยังฐานข้อมูล MySQL โดยใช้ Python script คณะผู้วิจัยใช้ Python MySQL module ในการติดต่อ (รับและส่ง SQL query/output) กับฐานข้อมูล MySQL ฟังก์ชันที่สอง คือ การแสดงผลด้วยภาพกราฟิก ทำให้ผู้ใช้สามารถเห็นภาพของตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบน

จีโนมจากการใช้จีโนมเบราว์เซอร์ซึ่งอาศัย GBrowse (STEIN *et al.* 2002) version 1.69 ฟังก์ชันที่สาม จะเป็น การสรุปผลทางสถิติของข้อมูลที่มีอยู่ฐานข้อมูล ในส่วนของ OUTPUT link-outs นั้น มีไว้ให้ผู้ใช้สามารถตรวจสอบข้อมูลในฐานข้อมูลอื่นๆที่เกี่ยวข้อง และยังมีโมดูลสำหรับการออกแบบไพรเมอร์เพื่ออำนวยความสะดวก ในการทดสอบตำแหน่งเป้าหมายที่สนใจทางห้องปฏิบัติการอีกด้วย โดยระบบฐานข้อมูล microPIR ทั้งหมดจะ ปฏิบัติการอยู่บน 12-core database server (2 AMD 6-core 2.8 GHz processors, 64 GB RAM, 2TB disk space)



รูปที่ 5 ภาพรวมการออกแบบฐานข้อมูล microPIR แสดงถึงแหล่งข้อมูลและลักษณะของเว็บอินเทอร์เฟซที่ติดต่อกับผู้ใช้

การทดสอบการใช้งานฐานข้อมูล

ฐานข้อมูล microPIR เป็นฐานข้อมูลที่สามารถเข้าถึงด้วยอินเทอร์เน็ต ซึ่งใช้งานผ่านทางเว็บเบราว์เซอร์ทั่วไป โดยเข้าไปใช้งานได้ที่ <http://www4a.biotech.or.th/micropir> คณะผู้วิจัยได้จัดทำคู่มือการใช้งานฐานข้อมูลไว้บริการที่ http://www4a.biotech.or.th/micropir/help/Basic_Search.html และได้ทดสอบความถูกต้องของฐานข้อมูลโดยตรวจสอบผลลัพธ์ที่ได้จากการทำนายในฐานข้อมูลเปรียบเทียบกับผลการศึกษาทดลองทางห้องปฏิบัติการที่มีการรายงานมาก่อน (Kim *et al.* 2008; PLACE *et al.* 2008; YOUNGER and COREY 2011) ซึ่งแสดงในตารางที่ 2 จากการทดสอบพบว่า สามารถค้นพบตำแหน่งเป้าหมายทั้งหมดตามรายงานในตารางที่ 2 ได้ในฐานข้อมูล microPIR โดยมีตำแหน่งและรูปแบบการปฏิสัมพันธ์ตรงตามงานวิจัยดังกล่าว ดังผลที่แสดงในรูปที่ 6

ตารางที่ 2 รายการของตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนลำดับเบสโปรโมเตอร์ซึ่งได้รับการรายงานมาก่อนในการศึกษาอื่น

mature miRNA	Promoter Gene	Properties of target site					Reference
		Binding pattern	Direction	Length	MFE	p-value	
hsa-miR-373	CDH1	target 5' U A CU G 3' CAC CCUGAAUUC AGCACUUU GUG GGGUUUUG UCCUGAAG miRNA 3' U CU 5'	Sense	25	-29.20	0.08679	Place et al.
hsa-miR-373	CSDC2	target 5' C G C G G 3' ACA CUUGA AGU CCGAGCACUUC UGU GGGUU UUA GCUUCGUGAAG miRNA 3' G 5'	Sense	27	-29.00	0.09459	
hsa-miR-320a	POLR3D	target 5' U C 3' UCCGCCUCUCAACCCAGCUUUU AGCGGGAGAGUUGGGUCGAAAA miRNA 3' 5'	Sense	24	-48.30	0.00002	Kim et al.
hsa-miR-320b	POLR3D	target 5' C C 3' GCCUCUCAACCCAGCUUUU CGGGAGAGUUGGGUCGAAAA miRNA 3' AA 5'	Sense	22	-45.30	0.00008	
hsa-miR-320c	POLR3D	target 5' C C 3' GCCUCUCAACCCAGCUUUU UGGGAGAGUUGGGUCGAAAA miRNA 3' 5'	Sense	22	-42.70	0.00005	
hsa-miR-320d	POLR3D	target 5' C C 3' CCUCUCAACCCAGCUUUU GGAGAGUUGGGUCGAAAA miRNA 3' A 5'	Sense	20	-38.00	0.00018	
hsa-miR-1278	PGR	target 5' A UGA U 3' GUAGA GAUCU ACAGUACU UAUCU CUUAU UGUCAUGA miRNA 3' A CG U 5'	Antisense	23	-23.80	0.04844	Younger and Corey
hsa-miR-595	PGR	target 5' U UG C G 3' GGAUGU UGUG CACACUUC UCUGUG GUGC GUGUGAAG miRNA 3' UG C 5'	Antisense	23	-22.40	0.84333	
hsa-miR-532-3p	PGR	target 5' G ACA AAA A A 3' UG AAG UUUUGG GUGGGAG AC UUC GAACCC CACCCUC miRNA 3' G G A C 5'	Antisense	27	-29.90	0.06508	

miRNA binding site information :	
miRNA accession	: MI0000781
miRNA ID	: hsa-mir-373
Mature miRNA accession	: MIMAT0000726
Mature miRNA ID	: hsa-miR-373
Target gene and isoform	: CDH1 (Gene ID : 999) mRNA : NM_004360
Direction	: Sense
Chromosome position	: chr16:67328051-67328075 (+)
Length	: 25 nt
GBrowse view	: View
Position	: -621
Minimum Free Energy (MFE)	: -29.20 Kcal/mol
p-value	: 0.086786
Binding pattern	: <div> target 5' U A CU G 3' <div> CAC CCUGAAAUC AGCACUUU GUG GGGUUUUAG UCGUGAAG </div> miRNA 3' U CU 5' </div>

(A)

miRNA binding site information :	
miRNA accession	: MI0000781
miRNA ID	: hsa-mir-373
Mature miRNA accession	: MIMAT0000726
Mature miRNA ID	: hsa-miR-373
Target gene and isoform	: CSDC2 (Gene ID : 27254) mRNA : NM_014460
Direction	: Sense
Chromosome position	: chr22:40286225-40286251 (+)
Length	: 27 nt
GBrowse view	: View
Position	: -709
Minimum Free Energy (MFE)	: -29.00 Kcal/mol
p-value	: 0.094590
Binding pattern	: <div> target 5' C G C G G 3' <div> ACA CUUGA AGU CGGAGCACUUC UGU GGGUU UUA GCUUCGUGAAG </div> miRNA 3' G 5' </div>

(B)

miRNA binding site information :	
miRNA accession	: MI0000542
miRNA ID	: hsa-mir-320a
Mature miRNA accession	: MIMAT0000510
Mature miRNA ID	: hsa-miR-320a
Target gene and isoform	: POLR3D (Gene ID : 661) mRNA : NM_001722
Direction	: Sense
Chromosome position	: chr8:22158432-22158455 (+)
Length	: 24 nt
GBrowse view	: View
Position	: -109
Minimum Free Energy (MFE)	: -48.30 Kcal/mol
p-value	: 0.000019
Binding pattern	: <div> target 5' U C 3' <div> UCGCCCUUCUACCCAGCUUUU AGCGGGAGAGUUGGUCGAAAA </div> miRNA 3' 5' </div>

(C)

รูปที่ 6 ผลลัพธ์การค้นหาลำดับเป้าหมายของไมโครอาร์เอ็นเอในยีนต่างๆที่ได้รับการรายงานมาก่อน (A) hsa-miR-373 ในยีน *CDH1* (B) hsa-miR-373 ในยีน *CSDC2* (C) hsa-miR-320a ในยีน *POLR3D* (D) hsa-miR-320b ในยีน *POLR3D* (E) hsa-miR-320c ในยีน *POLR3D* (F) hsa-miR-320d ในยีน *POLR3D* (G) hsa-miR-1278 ในยีน *PGR* (H) hsa-miR-595 ในยีน *PGR* (I) hsa-miR-532 ในยีน *PGR*

miRNA binding site information :	
miRNA accession	: MI0006425
miRNA ID	: hsa-mir-1278
Mature miRNA accession	: MIMAT0005936
Mature miRNA ID	: hsa-miR-1278
Target gene and isoform	: PGR (Gene ID : 5241) mRNA : NM_000926
Direction	: Antisense
Chromosome position	: chr11:100506384-100506406 (+)
Length	: 23 nt
GBrowse view	: View
Position	: -4349
Minimum Free Energy (MFE)	: -23.80 Kcal/mol
p-value	: 0.048441
Binding pattern	: <div> target 5' A UGA U 3' <div> GUAGA GAUGU ACAGUACU UAUCU CUAUA UGUCAUGA </div> miRNA 3' A CG U 5' </div>

(G)

miRNA binding site information :	
miRNA accession	: MI0003607
miRNA ID	: hsa-mir-595
Mature miRNA accession	: MIMAT0003263
Mature miRNA ID	: hsa-miR-595
Target gene and isoform	: PGR (Gene ID : 5241) mRNA : NM_000926
Direction	: Antisense
Chromosome position	: chr11:100506261-100506283 (+)
Length	: 23 nt
GBrowse view	: View
Position	: -4472
Minimum Free Energy (MFE)	: -22.40 Kcal/mol
p-value	: 0.843331
Binding pattern	: <div> target 5' U UG C G 3' <div> GGAUGU UGUG CACACUUC UCUGUG GUGC GUGUGAAG </div> miRNA 3' UG C 5' </div>

(H)

miRNA binding site information :	
miRNA accession	: MI0003205
miRNA ID	: hsa-mir-532
Mature miRNA accession	: MIMAT0004780
Mature miRNA ID	: hsa-miR-532-3p
Target gene and isoform	: PGR (Gene ID : 5241) mRNA : NM_000926
Direction	: Antisense
Chromosome position	: chr11:100506203-100506229 (+)
Length	: 27 nt
GBrowse view	: View
Position	: -4526
Minimum Free Energy (MFE)	: -29.90 Kcal/mol
p-value	: 0.065082
Binding pattern	: <div> target 5' G ACA AAA A A 3' <div> UG AAG UUUGGG GUGGGAG AC UUC GAACCC CACCCUC </div> miRNA 3' G G A C 5' </div>

(I)

รูปที่ 6 ผลลัพธ์การค้นหาลำดับเป้าหมายของไมโครอาร์เอ็นเอในยีนต่างๆซึ่งได้รับการรายงานมาก่อน (A) hsa-miR-373 ในยีน *CDH1* (B) hsa-miR-373 ในยีน *CSDC2* (C) hsa-miR-320a ในยีน *POLR3D* (D) hsa-miR-320b ในยีน *POLR3D* (E) hsa-miR-320c ในยีน *POLR3D* (F) hsa-miR-320d ในยีน *POLR3D* (G) hsa-miR-1278 ในยีน *PGR* (H) hsa-miR-595 ในยีน *PGR* (I) hsa-miR-532 ในยีน *PGR* (ต่อ)

ฐานข้อมูล microPIR เวอร์ชัน 2

หลังจัดตั้งฐานข้อมูล microPIR เวอร์ชันแรกเสร็จสมบูรณ์ คณะผู้วิจัยได้ดำเนินงานวิจัยต่อยอดเป็น การพัฒนาฐานข้อมูล microPIR เวอร์ชัน 2 เพื่อเพิ่มฟังก์ชันการทำงานของฐานข้อมูลให้ใช้งานได้ดีและครอบคลุมยิ่งขึ้น โดยส่วนหลักๆที่ได้เพิ่มเติมในฐานข้อมูลเวอร์ชันใหม่นี้ ได้แก่ การปรับปรุงข้อมูลไมโครอาร์เอ็นเอ, คุณลักษณะจีโนม และตำแหน่งเป้าหมายในจีโนมมนุษย์ให้เป็นปัจจุบัน และเพิ่มการทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนโปรโมเตอร์ของจีโนมหนู และพัฒนาโมดูลการเปรียบเทียบตำแหน่งเป้าหมายระหว่าง จีโนมมนุษย์และหนู (Human-mouse comparative view) รวมทั้งเพิ่มข้อมูลสนับสนุนซึ่งเป็นผลการทดลองทางห้องปฏิบัติการที่เกี่ยวข้อง โดยมีรายละเอียดวิธีดำเนินงานวิจัยตามขั้นตอนต่างๆ ดังต่อไปนี้

การทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ

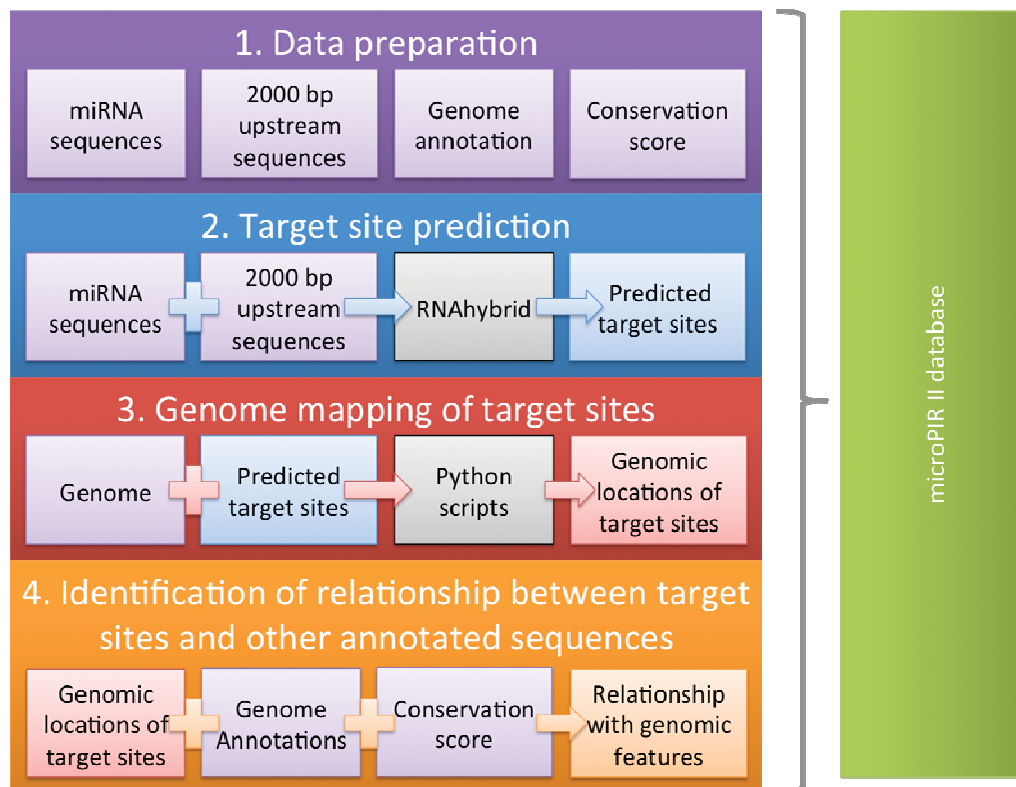
ในขั้นตอนนี้ ใช้โปรแกรม RNAhybrid (KRUGER and REHMSMEIER 2006) version 2.1 ในการทำนายตำแหน่งเป้าหมายบนลำดับเบสโปรโมเตอร์ทั้งในจีโนมมนุษย์และหนู ด้วยวิธีเดียวกันกับในฐานข้อมูล เวอร์ชันแรก ในฐานข้อมูลเวอร์ชัน 2 นี้ ได้คัดเลือกตำแหน่งเป้าหมายของมนุษย์และหนูซึ่งอยู่ภายใน 2000 เบสก่อนตำแหน่งเริ่มต้นของยีน (upstream sequences) และจัดเก็บลงในฐานข้อมูลที่แยกกันระหว่างมนุษย์ และหนู โดยในสิ่งมีชีวิตแต่ละชนิดจะแยกเก็บตำแหน่งเป้าหมายตามโครโมโซมที่พบ จากผลการทำนาย พบ ตำแหน่งเป้าหมายในมนุษย์และหนูจำนวนประมาณ 80 ล้านและ 40 ล้านตำแหน่ง ตามลำดับ โดยพบจำนวน ที่ใกล้เคียงกันบน sense และ antisense direction

แหล่งข้อมูลและการวิเคราะห์ข้อมูลสำหรับฐานข้อมูล

ในขั้นตอนนี้แบ่งเป็น 2 ส่วนย่อย ได้แก่ การปรับปรุงข้อมูลตำแหน่งเป้าหมายบนจีโนมมนุษย์ให้เป็น ปัจจุบัน และการทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนโปรโมเตอร์ของจีโนมหนู ซึ่งมีวิธีการจัด เตรียมและวิเคราะห์ข้อมูลในลักษณะเดียวกันกับในจีโนมมนุษย์ โดยกระบวนการจัดเตรียมและวิเคราะห์ข้อมูล สำหรับฐานข้อมูลประกอบไปด้วยขั้นตอนดังแสดงในรูปที่ 7

1) การเตรียมข้อมูลสำหรับวิเคราะห์

สำหรับฐานข้อมูล microPIR เวอร์ชัน 2 คณะผู้วิจัยได้รวบรวมข้อมูลสาธารณะของคุณลักษณะ จีโนมที่สำคัญของจีโนมมนุษย์และหนู ดังแสดงในตารางที่ 3 มาจัดเก็บไว้ในฐานข้อมูล สำหรับใช้เป็นข้อมูล สนับสนุนในการพิจารณาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ โดยได้จัดเก็บข้อมูลตำแหน่งเริ่มต้นและ ตำแหน่งสิ้นสุดบนจีโนมของคุณลักษณะต่างๆ รวมทั้งข้อมูลประกอบอื่นๆ ลงในฐานข้อมูล MySQL เพื่อให้ ง่ายต่อการค้นหาและการประมวลผล ในส่วนของข้อมูลในมนุษย์ ได้มีการใช้ข้อมูลไมโครอาร์เอ็นเอชุดล่าสุด คือ ข้อมูล release 19 ของ miRbase (GRIFFITHS-JONES *et al.* 2008) ซึ่งมีจำนวนไมโครอาร์เอ็นเอเพิ่มเป็น 2042 ตัว (จากเดิม 987 ตัวใน release 13) ทำให้ข้อมูลตำแหน่งเป้าหมายในจีโนมมนุษย์ของฐานข้อมูล เวอร์ชันใหม่นี้มีขนาดใหญ่กว่าในเวอร์ชันแรกมาก และมีขนาดประมาณสองเท่าของข้อมูลของหนูซึ่งมีไมโคร อาร์เอ็นเอ 1281 ตัว



รูปที่ 7 ขั้นตอนสำหรับกระบวนการวิเคราะห์ข้อมูลของ microPIR เวอร์ชัน 2

2) การประมวลผลและระบุตำแหน่งเป้าหมายบนจีโนม

ในขั้นตอนนี้จะใช้ python scripts ในการแปลงตำแหน่งเป้าหมายซึ่งเป็นตำแหน่งบนลำดับเบสโปรโมเตอร์ไปเป็นตำแหน่งที่แท้จริงบนจีโนม ได้แก่ ตำแหน่งเริ่มต้น (genome_start), ตำแหน่งสิ้นสุด (genome_end) และทิศทางของลำดับเบส (strand) ของตำแหน่งเป้าหมายบนจีโนม และจัดเก็บลงในฐานข้อมูล MySQL รวมทั้งแปลงข้อมูลตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆในตารางที่ 3 ให้อยู่ในรูปแบบ General Feature Format (GFF) file เพื่อใช้สำหรับการแสดงผลของตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆบนจีโนมเบราว์เซอร์

3) การตรวจสอบการคาบเกี่ยวกันของตำแหน่งเป้าหมายกับคุณลักษณะจีโนมอื่นๆ

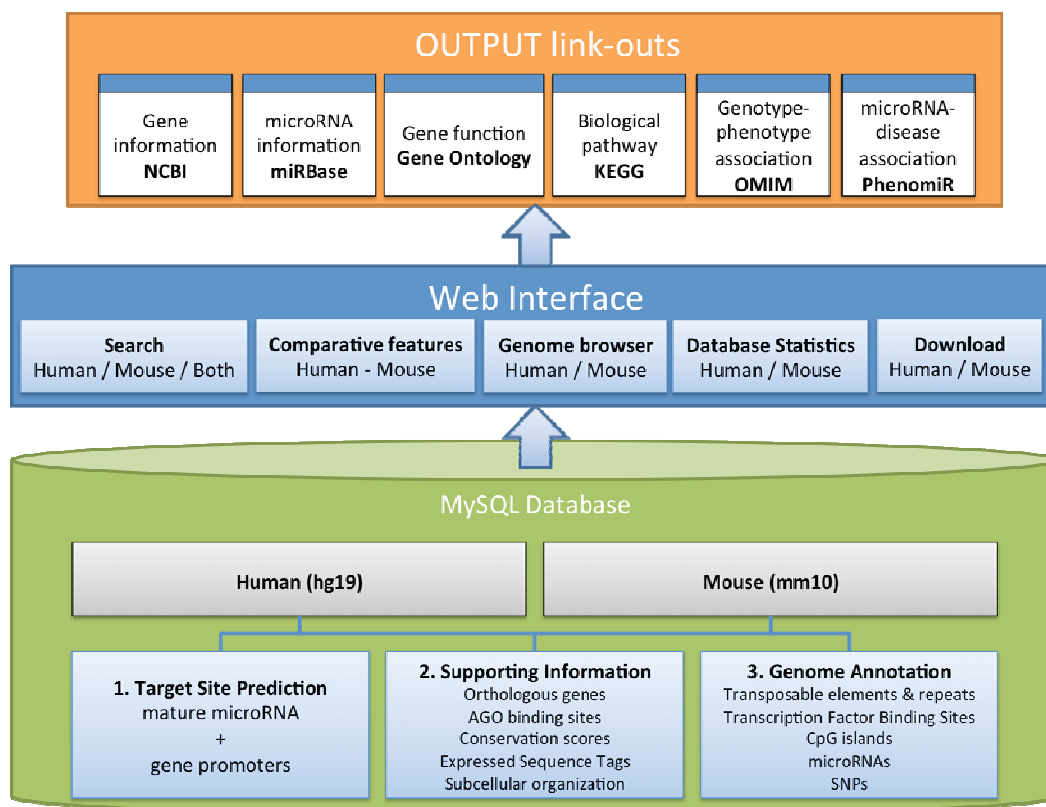
ขั้นตอนนี้จะนำข้อมูลตำแหน่งบนจีโนมของคุณลักษณะต่างๆ ที่เก็บไว้ในขั้นตอนที่ 1) และตำแหน่งบนจีโนมของตำแหน่งเป้าหมายที่เก็บไว้ในขั้นตอนที่ 2) มาตรวจสอบการคาบเกี่ยวกันของตำแหน่งเพื่อระบุความสัมพันธ์ที่เป็นไปได้ระหว่างตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆที่สนใจ โดยกำหนดให้มีการจัดเก็บข้อมูลความสัมพันธ์ที่มีขนาดการคาบเกี่ยวของตำแหน่งเป้าหมายกับคุณลักษณะจีโนมอย่างน้อย 80%

ตารางที่ 3 ข้อมูลสาธารณะของคุณลักษณะจีโนมที่อยู่ในฐานข้อมูล micorPIR เวอร์ชัน 2

Type	Public database/ version	
	Human	Mouse
MicroRNA	miRBase (GRIFFITHS-JONES <i>et al.</i> 2008)/ release 19	miRBase/ release 19
Gene's upstream sequences	UCSC Table Browser (KAROLCHIK <i>et al.</i> 2004)/ hg19	UCSC Table Browser/ mm10
Expressed Sequence Tag (EST)		
Transposable Element (TE)		
CpG Islands		
Single-nucleotide polymorphism (SNP)	UCSC Table Browser/ hg19-snp137	UCSC Table Browser/ mm10-snp137
Transcription Factor Binding Site (TFBS)	UCSC Table Browser/ hg19	ECRBase (LOOTS and OVCHARENKO 2007)/ mm9
AGO binding sites	CLIPZ (KHORSHID <i>et al.</i> 2010)/ hg18 (liftover to hg19)	StarBase (YANG <i>et al.</i> 2011)/ mm9
Human-mouse ortholog	MGI (http://www.informatics.jax.org/)/ updated on Apr 4, 2013	
Conservation score	UCSC Genome Browser (FUJITA <i>et al.</i> 2011)/ hg19 phastCons46way	UCSC Genome Browser/ mm10 phastCons60way

การออกแบบและสร้างฐานข้อมูล

ภาพรวมของการออกแบบฐานข้อมูล micorPIR เวอร์ชัน 2 แสดงดังในรูปที่ 8 โดยแบ่งเป็น 3 ส่วนหลักคือ ส่วนข้อมูลที่จัดเก็บ (MySQL Database), ส่วนติดต่อกับผู้ใช้ผ่านเว็บ (Web Interface) และส่วนเชื่อมโยงกับข้อมูลภายนอก (OUTPUT link-outs) โดยในส่วนของฐานข้อมูล (ด้านล่างของรูป) ยังคงมีโครงสร้างเช่นเดียวกับฐานข้อมูลเวอร์ชันแรก คือ แบ่งข้อมูลเป็น 3 ประเภท ได้แก่ ตำแหน่งเป้าหมายที่ได้จากการทำนาย (target site prediction), ข้อมูลคุณลักษณะจีโนมที่ช่วยในการพิจารณาคุณภาพของตำแหน่งเป้าหมายที่ทำนายได้ (supporting information) และข้อมูลคุณลักษณะจีโนมอื่นๆที่เกี่ยวข้อง (genome annotation) โดยมีข้อมูลของยีนที่มีบรรพบุรุษร่วมกันระหว่างมนุษย์และหนู (orthologous genes) สำหรับใช้ในการเชื่อมโยงและเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนู ซึ่งจะถูกวางอยู่ในส่วนของข้อมูลคุณลักษณะจีโนมที่ช่วยในการพิจารณาคุณภาพของตำแหน่งเป้าหมายที่ทำนายได้ ข้อมูลทั้งหมดจะถูกวิเคราะห์ รวบรวม และจัดเก็บลงในฐานข้อมูล MySQL version 5.5.1 สำหรับในส่วนติดต่อผู้ใช้กับฐานข้อมูลนั้น ได้รับการพัฒนาด้วยโปรแกรมที่เขียนโดยภาษา Python และ toolkits จาก Python webware และในส่วนการแสดงผลแบบกราฟิกของจีโนมจะอาศัย GBrowse (STEIN *et al.* 2002) version 1.69 ในการแสดงผลข้อมูล



รูปที่ 8 ภาพรวมการออกแบบฐานข้อมูล microPIR เวอร์ชัน 2

ผลการวิจัย

ฐานข้อมูล microPIR

ฐานข้อมูล microPIR แสดงข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอผ่านทางส่วนติดต่อกับผู้ใช้ผ่านเว็บ ซึ่งผู้ใช้สามารถค้นหา, ดึงข้อมูล และดูภาพกราฟิกของตำแหน่งเป้าหมายได้ โดยส่วนติดต่อกับผู้ใช้ผ่านเว็บ ประกอบไปด้วย 3 โมดูล ดังแสดงในรูปที่ 5 ได้แก่ 1) โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ (Search), 2) โมดูลสำหรับการแสดงผลแบบกราฟิก (Genome browser) และ 3) โมดูลสำหรับการแสดงผลสถิติของฐานข้อมูล (Database statistics)

โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ (Search)

ในโมดูลนี้จะมีตัวเลือกสำหรับการค้นหาตำแหน่งเป้าหมาย 2 แบบ คือ การค้นหาแบบธรรมดา (basic search) และการค้นหาแบบพิเศษ (advanced search) สำหรับการค้นหาแบบธรรมดา จะค้นหาได้จากชื่อยีนเป้าหมายหรือไมโครอาร์เอ็นเอเท่านั้น ซึ่งการค้นหาด้วยยีนเป้าหมายผู้ใช้สามารถค้นหาด้วยการระบุ HUGO gene name หรือ NCBI gene ID และการค้นหาด้วยไมโครอาร์เอ็นเอผู้ใช้สามารถค้นหาด้วยการระบุ

miRNA accession หรือ miRNA ID โดยการค้นหาแบบธรรมดาจะไม่สามารถแก้ไขค่าพารามิเตอร์และลักษณะต่างๆของตำแหน่งเป้าหมายที่ค้นหาได้ ผลลัพธ์ที่ได้จะเป็นรายการตำแหน่งเป้าหมายทั้งหมดตามค่าพารามิเตอร์เริ่มต้นที่กำหนดไว้ หากต้องการปรับเปลี่ยนพารามิเตอร์อื่นๆ สามารถเลือกเป็นการค้นหาแบบพิเศษ (advanced search)

สำหรับการค้นหาแบบพิเศษ มีลักษณะของเว็บอินเตอร์เฟซตามรูปที่ 9(A) โดยจะแบ่งออกเป็น 3 ส่วน ได้แก่ 1) ชื่อไมโครอาร์เอ็นเอหรือยีนเป้าหมาย, 2) ข้อมูลลักษณะและพารามิเตอร์ของตำแหน่งเป้าหมาย และ 3) ความสัมพันธ์กับคุณลักษณะอื่นๆของจีโนม ซึ่งในส่วนที่ 1) จะเหมือนกับการค้นหาแบบธรรมดา ในส่วนที่ 2) ผู้ใช้สามารถกำหนดค่าพารามิเตอร์หรือคุณสมบัติเฉพาะของตำแหน่งเป้าหมายที่ต้องการ ได้แก่ ทิศทางของตำแหน่งเป้าหมาย (strand orientation), ระยะตำแหน่งบนโปรโมเตอร์ที่ต้องการค้นหาเป้าหมาย (ค่ามากที่สุด 5000 เบส), ค่าเฉลี่ยการอนุรักษ์ลำดับเบสของตำแหน่งเป้าหมาย (average conservation score), ค่า MFE, ค่า *p*-value เป็นต้น ในส่วนที่ 3) ผู้ใช้สามารถเพิ่มเติมคุณสมบัติของตำแหน่งเป้าหมายตามความสัมพันธ์กับคุณลักษณะอื่นๆของจีโนม นั่นคือ สามารถระบุว่ามี/ไม่มีตำแหน่งคาบเกี่ยวกับคุณลักษณะต่างๆของจีโนม ได้แก่ ตำแหน่งการจับกับโปรตีน AGO, repetitive element, TFBS, EST และ SNP

จากรูปที่ 9(B) ในส่วนของผลลัพธ์ของตำแหน่งเป้าหมายที่ได้จากการค้นหา จะแสดงรายละเอียดของไมโครอาร์เอ็นเอและตำแหน่งเป้าหมายในเรื่องของทิศทางของตำแหน่งเป้าหมายที่สัมพันธ์กับยีน, ตำแหน่งบนจีโนม, สายดีเอ็นเอที่มีตำแหน่งเป้าหมายนั้น, ความยาวของตำแหน่งเป้าหมาย, ตำแหน่งของเป้าหมายเทียบกับจุดเริ่มต้นของยีน, ค่า MFE, และคะแนนการอนุรักษ์ของลำดับเบสในสัตว์ที่มีกระดูกสันหลัง นอกจากนี้ยังแสดงจำนวนของคุณลักษณะต่างๆ ของจีโนมที่คาบเกี่ยวกับเป้าหมายในแต่ละตำแหน่ง รวมถึงการให้บริการลิงค์เชื่อมโยงไปยังแหล่งฐานข้อมูลที่เกี่ยวข้องเพื่ออธิบายรายละเอียดเพิ่มเติมของยีนเป้าหมายและไมโครอาร์เอ็นเอ เช่น รายละเอียดของยีนจากฐานข้อมูล NCBI, ข้อมูลหน้าที่ของยีนจาก Gene Ontology (ASHBURNER *et al.* 2000), วิถีทางชีววิทยาของยีนจาก KEGG (KANEHISA *et al.* 2010), ข้อมูลลักษณะการแสดงออกทางกายภาพที่เกี่ยวข้องกับยีนจากฐานข้อมูล OMIM (<http://www.ncbi.nlm.gov/omim>) และข้อมูลไมโครอาร์เอ็นเอจาก miRBase ผู้ใช้ยังสามารถเลือกดูรายละเอียดของรูปแบบการจับกับตำแหน่งเป้าหมายและพารามิเตอร์เพิ่มเติม โดยคลิกที่แสดงรายละเอียด (detail) และในหน้ารายละเอียดจะมีลิงค์เชื่อมโยงเพื่อให้ผู้ใช้สามารถออกแบบไพรเมอร์ที่เกี่ยวข้องด้วยโปรแกรม RExPrimer (PIRIYAPONGSA *et al.* 2009) ซึ่งช่วยอำนวยความสะดวกในการออกแบบการทดสอบทางห้องปฏิบัติการของตำแหน่งเป้าหมายต่อไป

โมดูลสำหรับการแสดงผลแบบกราฟิก (Genome browser)

โมดูลนี้ใช้สำหรับการแสดงผลจากการค้นหาในโมดูลที่แล้ว โดยการคลิกลิงค์ของตำแหน่งเป้าหมายในผลลัพธ์หรือโดยผู้ใช้ใส่ตำแหน่งบนโครโมโซม (Chromosome position) ของบริเวณที่สนใจ ภาพของตำแหน่งเป้าหมายและคุณลักษณะของจีโนมบริเวณใกล้เคียงจะถูกแสดงในรูปแบบกราฟิกบนจีโนมเบราว์เซอร์ ดังตัวอย่างในรูปที่ 10 ซึ่งแสดงตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ miR-34c ผู้ใช้ยังสามารถขยับเลื่อนหรือขยายเพื่อเลือกดูตำแหน่งที่อยู่ใกล้เคียงได้เองด้วย การแสดงผลแบบกราฟิกนี้ทำให้ผู้ใช้มีมุมมองที่กว้างขึ้น มองเห็นตำแหน่งเป้าหมายที่อยู่ใกล้เคียง ทั้งยังสามารถดูรายละเอียดต่างๆ จากการนำเมาส์เลื่อนไปบนบริเวณที่ต้องการ และยังเห็นความสัมพันธ์และการคาบเกี่ยวกันของตำแหน่งเป้าหมายกับคุณลักษณะอื่นๆของจีโนมอีกด้วย

จากรูปที่ 10 ผลลัพธ์ของตำแหน่งเป้าหมายแสดงในรูปแบบกราฟิกบนเบราว์เซอร์พร้อมข้อมูลสนับสนุนอื่นๆ ซึ่งมีคำอธิบายจีโนมประกอบ โดยในมุมมองขอบเขต (region view) จะแสดงการกระจายของตำแหน่งเป้าหมายในบริเวณที่ผู้ใช้กำหนด ผู้ใช้สามารถเลือก (แสดงด้วยสีเหลือง) ในบริเวณสีเทาเพื่อขยายดูรายละเอียดในบริเวณที่สนใจ ตำแหน่งเป้าหมายที่ผู้ใช้เลือกจะแสดงอยู่ในแถบสีส้มที่มีกรอบสีเขียว และคะแนนการอนุรักษ์ลำดับเบสของแต่ละนิวคลีโอไทด์ซึ่งมีค่าอยู่ในช่วงตั้งแต่ 0 ถึง 1 จะแสดงในแถบสีน้ำเงิน กลุ่มของตำแหน่งการจับของโปรตีน AGO บนจีโนมซึ่งแสดงด้วยสีเขียวอ่อน เป็นการกระจายตัวของจำนวน sequence read ที่เกี่ยวข้องในช่วงตำแหน่งดังกล่าว โดยในภาพนี้แสดงตำแหน่งของไมโครอาร์เอ็นเอ miR-34c ในแถบสีแดง ซึ่งอยู่ในบริเวณเดียวกันกับตำแหน่งเป้าหมายนั่นเอง ดังนั้นภาพนี้จึงแสดงให้เห็นถึงบทบาทการควบคุมการแสดงออกของยีนแบบ cis ของไมโครอาร์เอ็นเอตัวนี้

A

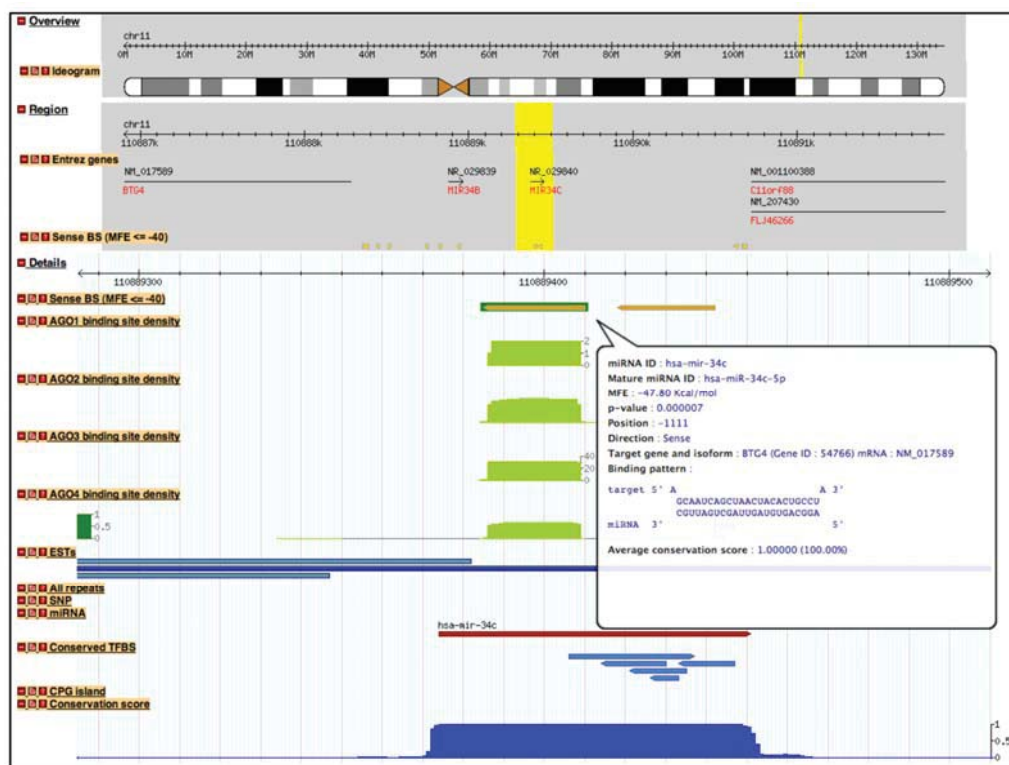
B

No.	mature miRNA ID	miRNA ID	Direction	Chr. position	Strand	Length	Position	MFE	Conserved score (bp)	Overlapping features								
										AGO1	AGO2	AGO3	AGO4	EST	CPC	REP	TFBS	SNP
1. detail	hsa-mir-34c-3p	hsa-mir-34c	Sense	Chr11:110889418-110889461	-	24	-1144	-48.20	1.0000 (24)	0	0	0	1	1	0	0	2	0
2. detail	hsa-mir-34c-5p	hsa-mir-34c	Sense	Chr11:110889381-110889409	-	25	-1111	-47.80	1.0000 (25)	1	1	0	1	1	0	0	0	0
3. detail	hsa-mir-34b	hsa-mir-34b	Sense	Chr11:110888921-110888944	-	24	-647	-44.50	0.9995 (24)	1	0	0	0	0	2	0	0	0
4. detail	hsa-mir-1181	hsa-mir-1181	Sense	Chr11:110888958-110888978	-	21	-684	-39.70	0.6097 (21)	0	0	1	0	2	0	0	0	0
5. detail	hsa-mir-34a	hsa-mir-34a	Sense	Chr11:110888888-110888909	-	24	-612	-39.50	0.9935 (24)	0	1	1	1	2	0	0	1	0
6. detail	hsa-mir-449b	hsa-mir-449b	Sense	Chr11:110889381-110889405	-	21	-1111	-35.10	1.0000 (21)	0	1	1	1	1	0	0	0	0
7. detail	hsa-mir-34c-5p	hsa-mir-34c	Antisense	Chr11:110888921-110888942	+	22	-4333	-32.50	0.9995 (22)	1	0	0	0	2	0	0	0	0

รูปที่ 9 โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ

(A) ส่วนติดต่อกับผู้ใช้สำหรับการค้นหาตำแหน่งเป้าหมายแบบพิเศษ (Advanced search)

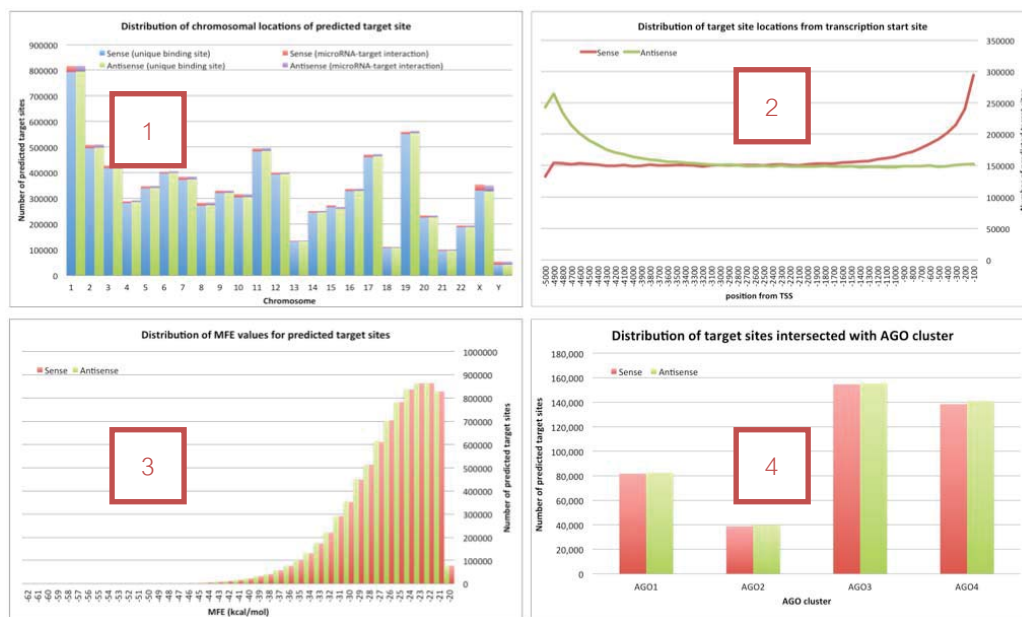
(B) ผลลัพธ์ที่ได้จากการค้นหาตำแหน่งเป้าหมายตามค่าพารามิเตอร์และลักษณะที่กำหนด



รูปที่ 10 ตัวอย่างการแสดงผลลัพธ์ของโมดูลการแสดงผลแบบกราฟิกบนจีโนมเบราร์เซอร์

โมดูลสำหรับการแสดงผลสถิติของฐานข้อมูล (Database statistics)

โมดูลนี้ทำหน้าที่แสดงรายละเอียดโดยสรุปของข้อมูลที่ใช้ในการพัฒนาฐานข้อมูล โดยนำเสนอเป็นข้อมูลสถิติดังตัวอย่างในรูปที่ 11 เช่น ข้อมูลรายละเอียดของไมโครอาร์เอ็นเอที่ใช้, ข้อมูลรายละเอียดของลำดับเบสโปรโมเตอร์ที่ใช้, ข้อมูลการกระจายของตำแหน่งเป้าหมายบนโครโมโซม, ข้อมูลการกระจายของค่า MFE กับจำนวนตำแหน่งเป้าหมาย รวมถึงข้อมูลสรุปความสัมพันธ์ระหว่างตำแหน่งเป้าหมายกับคุณลักษณะจีโนมต่างๆ โดยข้อมูลเหล่านี้จะเป็นแนวทางให้นักวิจัยใช้ตัดสินใจเลือกใช้พารามิเตอร์สำหรับการค้นหาตำแหน่งเป้าหมายได้เป็นอย่างดี ซึ่งโมดูลนี้จะมีการปรับปรุงข้อมูลอย่างอัตโนมัติให้เป็นข้อมูลปัจจุบันเมื่อมีการปรับปรุงฐานข้อมูล microPIR



รูปที่ 11 ตัวอย่างข้อมูลสถิติในโมดูลสำหรับการแสดงผลสถิติของฐานข้อมูล

- 1) ข้อมูลการกระจายของตำแหน่งเป้าหมายบนโครโมโซมต่างๆ
- 2) ข้อมูลการกระจายของจุดเริ่มต้นของตำแหน่งเป้าหมายบนโปรโมเตอร์
- 3) ข้อมูลการกระจายของค่า MFE ของการจับระหว่างไมโครอาร์เอ็นเอและตำแหน่งเป้าหมาย
- 4) ข้อมูลการกระจายของจำนวนตำแหน่งเป้าหมายที่คาบเกี่ยวกับตำแหน่งการจับบนจีโนมของโปรตีน AGO ชนิดต่างๆ

ฐานข้อมูล microPIR เวอร์ชัน 2

ส่วนติดต่อกับผู้ใช้งานเว็บของฐานข้อมูล microPIR เวอร์ชัน 2 ประกอบไปด้วย 5 โมดูล ดังแสดงในรูปที่ 8 โดยมีหนึ่งโมดูลเพิ่มขึ้นมาจากฐานข้อมูลเวอร์ชันแรก คือ โมดูลสำหรับการเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนู (Human-mouse comparative view)

โมดูลสำหรับการค้นหาตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ (Search)

สำหรับโมดูลนี้ยังคงมีฟังก์ชันการใช้งานคล้ายคลึงกับเวอร์ชันเดิม แต่ในเวอร์ชันเดิมเป็นการค้นหาในจีโนมมนุษย์เท่านั้นและแบ่งรูปแบบการค้นหาเป็นการค้นหาแบบธรรมดาและการค้นหาแบบพิเศษที่มีการระบุพารามิเตอร์ในการค้นหา ในเวอร์ชันใหม่นี้เปลี่ยนแปลงรูปแบบการค้นหาโดยแบ่งเป็น 2 รูปแบบ คือ 1) การค้นหาในสิ่งมีชีวิตชนิดเดียว (single organism) ได้แก่ มนุษย์ หรือ หนู และ 2) การค้นหาตำแหน่งเป้าหมายทั้งในมนุษย์และหนูในเวลาเดียวกัน (Human-mouse comparative view) โดยการใช้ชื่อยีนที่มีบรรพบุรุษร่วมกัน (orthologous genes) ของสิ่งมีชีวิตสองชนิดในการค้นหา การค้นหาทั้งสองรูปแบบสามารถทำได้ผ่านหน้าเว็บเพจเดียวกัน ดังรูปที่ 12A และ 12B ตามลำดับ และการค้นหาทั้งสองรูปแบบสามารถระบุ

พารามิเตอร์ของตำแหน่งเป้าหมายที่ต้องการได้เช่นเดียวกับการค้นหาแบบพิเศษในฐานข้อมูลเวอร์ชันเดิม โดยได้นำส่วนดังกล่าวมารวมไว้ในหน้าเพจการค้นหาด้วย ดังรูปที่ 13B

การค้นหาโดยใช้ชื่อของยีนเป้าหมายมีการปรับปรุงให้สามารถค้นหาจากรายการชื่อยีน จากเดิมที่ค้นหาได้ทีละยีน ผลลัพธ์ที่ได้จากการค้นหาในสิ่งมีชีวิตชนิดเดียวจะแสดงตามรายการยีนเป้าหมายดังในรูปที่ 14 ซึ่งมีลิงค์ที่สามารถเชื่อมโยงไปยังรายละเอียดของกลุ่มไมโครอาร์เอ็นเอที่มีตำแหน่งเป้าหมายบนยีนดังกล่าว (รูปที่ 15) สำหรับการค้นหาโดยใช้ชื่อไมโครอาร์เอ็นเอ ผลลัพธ์ที่ได้จะมีการแสดงข้อมูลเพิ่มเติมจากเวอร์ชันเดิมในส่วนของคุณสมบัติความสัมพันธ์กับโรคที่มีการรายงานว่าเกี่ยวข้องกับไมโครอาร์เอ็นเอดังกล่าวจากฐานข้อมูล PhenomiR (RUEPP *et al.* 2010) รวมทั้งข้อมูล subcellular organization ของไมโครอาร์เอ็นเอ ซึ่งบอกตำแหน่งของไมโครอาร์เอ็นเอในเซลล์ ข้อมูลเหล่านี้ช่วยเสริมการตัดสินใจเลือกไมโครอาร์เอ็นเอที่น่าสนใจเพื่อนำไปทดลองในห้องปฏิบัติการต่อไป โดยคณะผู้วิจัยนำข้อมูล subcellular organization มาจากสามการศึกษา (JEFFRIES *et al.* 2011; LIAO *et al.* 2010; TAFT *et al.* 2010) ซึ่งมีการทดลองเปรียบเทียบปริมาณของไมโครอาร์เอ็นเอที่พบภายในส่วนนิวเคลียสและไซโทพลาซึม การที่ไมโครอาร์เอ็นเอพบภายในส่วนนิวเคลียสเป็นส่วนใหญ่ เป็นข้อมูลหนึ่งที่ช่วยสนับสนุนความเป็นไปได้ในการจับกันของไมโครอาร์เอ็นเอกับตำแหน่งเป้าหมายที่อยู่ภายในลำดับเบสโปรโมเตอร์ได้ สำหรับการค้นหาแบบเปรียบเทียบในมนุษย์และหนู ผลลัพธ์ที่ได้จะมีการเชื่อมโยงไปยังโมดูลการเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนูต่อไป

A

B

รูปที่ 12 หน้าเพจการกรอกอินพุตสำหรับการค้นหาตำแหน่งเป้าหมายของ microPIR เวอร์ชัน 2

- (A) การค้นหาตำแหน่งเป้าหมายในสิ่งมีชีวิตชนิดเดียว
- (B) การค้นหาตำแหน่งเป้าหมายเปรียบเทียบในมนุษย์และหนู

microPIR [microRNA-promoter interactions resource]

Home About Login Search Genome browser Download Database statistics User Guide

1. Select miRNA/target gene

☒ Single organism ☐ Human ☐ Mouse

☒ A list of target genes (gene symbol, NCBI gene ID), one per line. See [example](#).

BTG4
ATM
OLD
AP3A1
CHUK
CASC
SOS2

☐ Mature miRNA/miRNA (ID, accession)

(Example : Human : hsa-miR-34c-3p, hsa-miR-100, MIMAT0004677 Mouse : mmu-miR-34c-3p, mmu-miR-707, MIMAT0004500)

☐ Disease/Phenotype (Example : Cancer, Alzheimer)

☐ Human-mouse comparative view Search from ☐ Human genes or ☐ Mouse genes

2. Set binding site characteristics

upstream region bp strand (maximum 2000 bp)

☒ Minimum Free Energy (MFE) \leq kcal/mol

☒ p-value \leq

☐ average conservation score \geq with available position-specific score \geq % of target length

☐ maximum number of unpaired nucleotides

☐ maximum number of G-U base pairs in seed region

☐ maximum internal loop size

☐ maximum bulge size

Overlap with supporting experimental data

☐ Any AGO cluster ☐ AGO1 cluster ☐ AGO2 cluster ☐ AGO3 cluster ☐ AGO4 cluster

☐ EST

Checking with the annotated genomic features

☐ Check ☐ with or ☐ without repetitive element

☐ Any repetitive element ☐ LINE ☐ SINE ☐ LTR ☐ DNA ☐ Simple repeat ☐ Low complexity ☐ Satellite ☐ RNA ☐ Other repeat ☐ Unknown repeat

☐ Check ☐ with or ☐ without Transcription Factor Binding Site (TFBS)

☐ Check ☐ with or ☐ without SNP

☐ Check ☐ with or ☐ without CpG island

รูปที่ 13 หน้าเพจสำหรับโมดูลการค้นหาตำแหน่งเป้าหมายของ microPIR เวอร์ชัน 2

(A) ส่วนการระบุรูปแบบการค้นหาและอินพุตที่ใช้ในการค้นหา

(B) ส่วนการระบุพารามิเตอร์ที่ใช้ในการค้นหาตำแหน่งเป้าหมาย

microPIR [microRNA-promoter interactions resource]

Home About Login Search Genome browser Download Database statistics User Guide

Gene information

1. Human official gene symbol : BTG4 Official full name : protein BTG4 Gene ID : 54766 OMIM ID : 605673

Other aliases : PC3B

Orthologous mouse genes : [Btg4](#) [Gene ID : 56057]

Transcripts, products and genomic position, [Click to show/hide genomic position](#)

Transcript	Product	Position	Strand	Size (bp)
NM_017599	NP_060059	Chr11:111338256 - 111363079	-	44824

Gene Ontology, [Click to show/hide GO](#)

Target sites : Sense direction **93** targets, Antisense direction **21** targets

2. Human official gene symbol : ATM Official full name : serine-protein kinase ATM Gene ID : 472 OMIM ID : 607585

Other aliases : AT1, ATA, ATC, ATD, ATDC, ATE, TEL1, TEL1

Orthologous mouse genes : [Atm](#) [Gene ID : 13920]

Transcripts, products and genomic position, [Click to show/hide genomic position](#)

Transcript	Product	Position	Strand	Size (bp)
NM_000051	NP_000042	Chr11:108093259 - 108128826	+	146268

KEGG pathway : [hsa04110](#), [hsa04115](#), [hsa04210](#)

Gene Ontology, [Click to show/hide GO](#)

Target sites : Sense direction **13** targets, Antisense direction **12** targets

3. Human official gene symbol : OLD Official full name : dihydrolipoil dehydrogenase, mitochondrial precursor Gene ID : 1738 OMIM ID : 238331

Other aliases : OLD0, OLD0, L3, GCSL, LAD, PHE3

Orthologous mouse genes : [Old](#) [Gene ID : 13082]

Transcripts, products and genomic position, [Click to show/hide genomic position](#)

Transcript	Product	Position	Strand	Size (bp)
NM_000106	NP_000099	Chr7:107531586 - 107561643	+	30058

KEGG pathway : [hsa00210](#), [hsa00260](#), [hsa00280](#), [hsa00630](#), [hsa01100](#)

Gene Ontology, [Click to show/hide GO](#)

Target sites : Sense direction **6** targets, Antisense direction **22** targets

รูปที่ 14 ส่วนแสดงผลลัพธ์ตามข้อยีนซึ่งพัฒนาขึ้นมาใหม่ใน microPIR เวอร์ชัน 2

Gene information

Human official gene symbol : [BTG4](#)

Other aliases : PC38

Orthologous mouse genes : [Btg4](#) (Gene ID : 56057)

Official full name : protein BTG4

Gene ID : 54766

OMIM ID : [605673](#)

Transcripts, products and genomic position, [Click to show/hide genomic position](#)

Transcript	Product	Position	Strand	Size (bp)
NM_017589	NP_060059	Chr11:111338256 - 111383079	-	44824

Gene Ontology, [Click to show/hide GO](#)

miRNA target sites

Export

Format

Tab-delimited

Sorted by

Position

No	mature miRNA ID	Direction	Position	Strand	Length	MFE	Conserved score (% pos.)	Overlapping genomic features								
								AGO1	AGO2	AGO3	AGO4	EST	CpG	REP	TFBS	SNP
1. detail	hsa-miR-1184	Sense	-5	-	27	-31.50	0.0906 (100)	0	0	0	0	2	0	0	0	0
2. detail	hsa-miR-3615	Sense	-17	-	24	-32.50	0.0037 (100)	0	0	0	0	2	0	0	0	1
3. detail	hsa-miR-5193	Sense	-27	-	21	-39.70	0.0181 (100)	0	0	0	0	2	0	1	0	1
4. detail	hsa-miR-877-3p	Sense	-27	-	26	-39.50	0.1362 (100)	0	0	0	0	2	0	1	0	1
5. detail	hsa-miR-1236-3p	Sense	-28	-	24	-37.50	0.1064 (100)	0	0	0	0	2	0	1	0	1
6. detail	hsa-miR-3679-3p	Sense	-28	-	21	-31.80	0.0221 (100)	0	0	0	0	2	0	1	0	1
7. detail	hsa-miR-4667-3p	Sense	-30	-	19	-38.20	0.0235 (100)	0	0	0	0	2	0	1	0	1
8. detail	hsa-miR-4446-5p	Sense	-31	-	20	-33.70	0.0767 (100)	0	0	0	0	2	0	1	0	1
9. detail	hsa-miR-4723-3p	Sense	-32	-	22	-39.30	0.2027 (100)	0	0	0	0	2	0	1	0	1
10. detail	hsa-miR-211-5p	Sense	-33	-	26	-35.00	0.2021 (100)	0	0	0	0	2	0	1	0	1
11. detail	hsa-miR-1237-3p	Sense	-34	-	22	-34.30	0.2381 (100)	0	0	0	0	2	0	1	0	1
12. detail	hsa-miR-6515-3p	Sense	-34	-	22	-32.40	0.2381 (100)	0	0	0	0	2	0	1	0	1
13. detail	hsa-miR-1470	Sense	-35	-	30	-33.40	0.1746 (100)	0	0	0	0	2	0	1	0	1
14. detail	hsa-miR-2116-3p	Sense	-35	-	20	-30.50	0.2614 (100)	0	0	0	0	2	0	1	0	1
15. detail	hsa-miR-1247-5p	Sense	-38	-	18	-31.60	0.2903 (100)	0	0	0	0	2	0	1	0	0

รูปที่ 15 ส่วนแสดงผลลัพธ์ของรายการตำแหน่งเป้าหมายบนยีนเป้าหมายแต่ละตัว

โมดูลสำหรับการเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนู (Human-mouse comparative view)

การใช้งานโมดูลนี้เริ่มต้นจากการเลือกการค้นหาลักษณะเปรียบเทียบจากรูปที่ 12B ผลลัพธ์ที่ได้จากการค้นหาแสดงในรูปที่ 16 ซึ่งคล้ายคลึงกับการแสดงผลจากการค้นหาในสิ่งมีชีวิตชนิดเดียว แต่ประกอบด้วยข้อมูลเพิ่มเติมของสิ่งมีชีวิตทั้งสองชนิดรวมทั้งข้อมูลยีนที่มีบรรพบุรุษร่วมกัน ในส่วนนี้ ผู้ใช้สามารถพิจารณา ยีนรวมทั้งจำนวนตำแหน่งเป้าหมาย และคลิกลิงค์เพื่อเข้าสู่หน้าเพจที่แสดงผลกราฟิกจากการใช้ Scalable Vector Graphics (SVG) ผู้ใช้สามารถเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ของมนุษย์และหนูรวมทั้งความสัมพันธ์กับคุณลักษณะจีโนมอื่นๆ ดังแสดงในรูปที่ 17

จากรูปที่ 17 บริเวณในส่วนสี่เหลี่ยมแสดงตำแหน่งของตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆ ที่เกี่ยวข้องในจีโนมมนุษย์ ในขณะที่บริเวณสี่เหลี่ยมแสดงตำแหน่งเป้าหมายและคุณลักษณะจีโนมบนยีนที่มีบรรพบุรุษร่วมกันในจีโนมหนู หากผู้ใช้ต้องการทราบรายละเอียดเพิ่มเติมเกี่ยวกับตำแหน่งเป้าหมายหรือคุณลักษณะจีโนม ผู้ใช้สามารถคลิกที่ตำแหน่งที่ต้องการบนรูปเพื่อแสดงรายละเอียดเพิ่มเติมได้ ดังรูปที่ 18

โมดูลสำหรับการแสดงผลแบบกราฟิก (Genome Browser)

โมดูลนี้ใช้สำหรับการแสดงรายละเอียดผลลัพธ์แบบกราฟิกจากการค้นหาในสิ่งมีชีวิตชนิดเดียว โดยการคลิกที่แสดงตำแหน่งบนจีโนมของตำแหน่งเป้าหมายในผลลัพธ์ หรือโดยผู้ใช้ใส่ตำแหน่งบนโครโมโซม (Chromosome position) ของบริเวณที่สนใจ ภาพของตำแหน่งเป้าหมายและคุณลักษณะของจีโนมอื่นๆ จะถูกแสดงในรูปแบบกราฟิกบนจีโนมเบราร์เซอร์ จุดประสงค์หลักในโมดูลส่วนนี้ คือ เพื่อให้ผู้ใช้สามารถมองเห็นรายละเอียดของตำแหน่งเป้าหมายและคุณลักษณะจีโนมในบริเวณใกล้เคียงในแต่ละสิ่งมีชีวิต ซึ่งทำให้ทราบ


ถึงความสัมพันธ์ที่เป็นไปได้และการคาบเกี่ยวกันระหว่างตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆ ดังแสดงตัวอย่างในรูปที่ 19

โมเดลสำหรับการแสดงผลสถิติของฐานข้อมูล (Database Statistics)

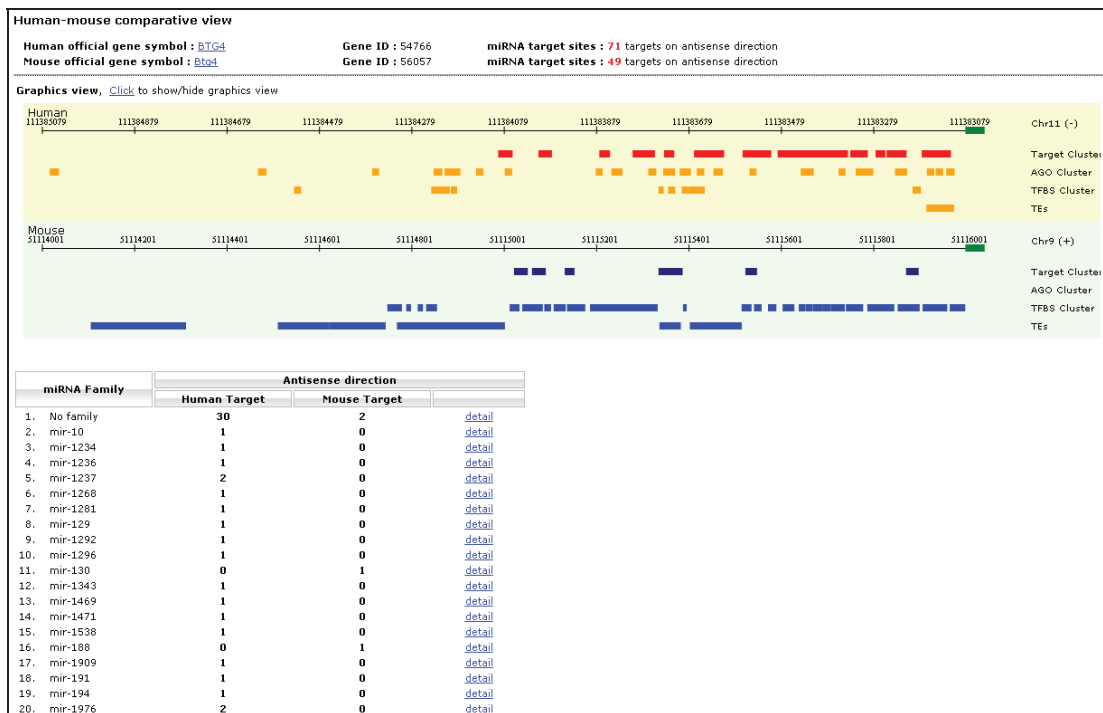
โมเดลนี้ทำหน้าที่แสดงรายละเอียดโดยสรุปของข้อมูลที่ใช้ในการพัฒนาฐานข้อมูล เช่น ข้อมูลรายละเอียดของไมโครอาร์เอ็นเอที่ใช้, ข้อมูลรายละเอียดของลำดับเบสโปรโมเตอร์ที่ใช้, ข้อมูลการกระจายของตำแหน่งเป้าหมายบนโครโมโซม, ข้อมูลการกระจายของค่า MFE กับจำนวนตำแหน่งเป้าหมาย รวมถึงข้อมูลสรุปความสัมพันธ์ระหว่างตำแหน่งเป้าหมายกับคุณลักษณะจีโนมต่างๆ โดยข้อมูลเหล่านี้จะใช้เป็นแนวทางให้นักวิจัยตัดสินใจเลือกใช้พารามิเตอร์สำหรับการค้นหาตำแหน่งเป้าหมายที่ต้องการได้

โมเดลการดาวน์โหลดข้อมูล (Download)

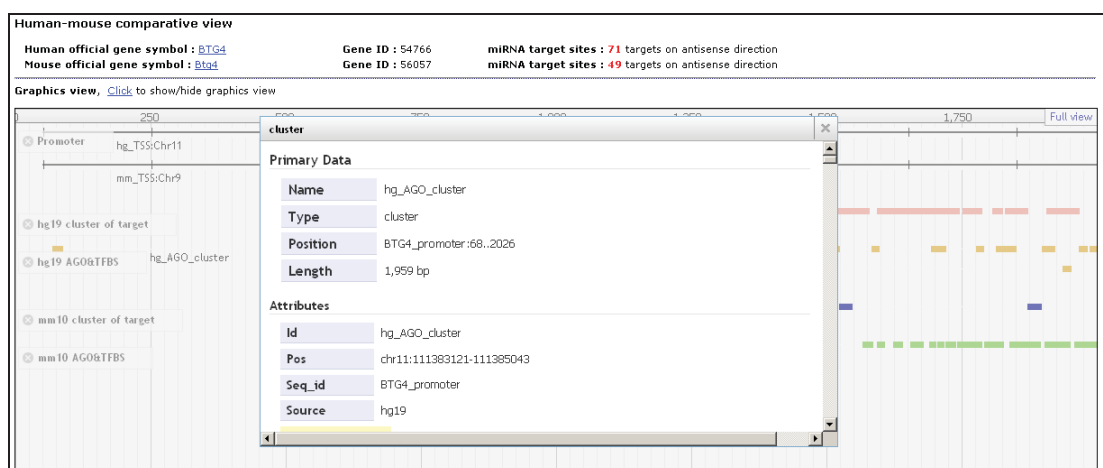
คณะผู้วิจัยเปิดโอกาสให้ผู้ใช้งานสามารถดาวน์โหลดข้อมูลตำแหน่งเป้าหมายที่ได้จากการทำนายทั้งหมดในรูปแบบของ BED file เพื่อให้สามารถนำข้อมูลดังกล่าวไปใช้ในการวิเคราะห์ระดับจีโนม (genome-wide analysis) ในหัวข้อวิจัยที่สนใจได้สะดวก โดยเปิดให้บริการสำหรับผู้ใช้งานที่ลงทะเบียนเท่านั้น

 [microRNA-promoter interactions resource]			
Home About Login Search Genome browser Download Database statistics User Guide			
Human-mouse comparative view			
1. Human official gene symbol : BTG4 Official full name : protein BTG4 Gene ID : 54766 QMID ID : 605673			
Other aliases : PC3B			
Transcripts, products and genomic position. Click to show/hide genomic position			
Transcript	Product	Position	Strand Size (bp)
NM_017589	NP_060059	Chr11:11139256 - 11139302	- 44624
Gene Ontology. Click to show/hide GO			
Mouse official gene symbol : Btg4 Official full name : protein BTG4 Gene ID : 56057 QMID ID : -			
Other aliases : C86116, PC3B			
Transcripts, products and genomic position. Click to show/hide genomic position			
Transcript	Product	Position	Strand Size (bp)
NM_019493	NP_062366	Chr2:51116001 - 51116200	+ 3700
KEGG pathway : mmu03038			
Gene Ontology. Click to show/hide GO			
miRNA target sites:			
Organism	Sense target sites	Antisense target sites	
Homo sapiens	93	71	
Mus musculus	5	49	
Visualization: sense Visualization: antisense			
2. Human official gene symbol : ATM Official full name : serine-protein kinase ATM Gene ID : 472 QMID ID : 607585			
Other aliases : AT1, ATA, ATC, ATD, ATDC, ATE, TEL1, TEL101			
Transcripts, products and genomic position. Click to show/hide genomic position			
Transcript	Product	Position	Strand Size (bp)
NM_000051	NP_000042	Chr11:108095559 - 108229826	+ 146268
KEGG pathway : hsa04110, hsa04115, hsa04210			
Gene Ontology. Click to show/hide GO			
Mouse official gene symbol : Atm Official full name : serine-protein kinase ATM Gene ID : 11920 QMID ID : -			
Other aliases : A1256621, C030026E19Rik			
Transcripts, products and genomic position. Click to show/hide genomic position			
Disclaimer Copyright © 2015 Bioinformatics and Informatics Laboratory, Genome Institute, B10TEC, NSTDA, Thailand			

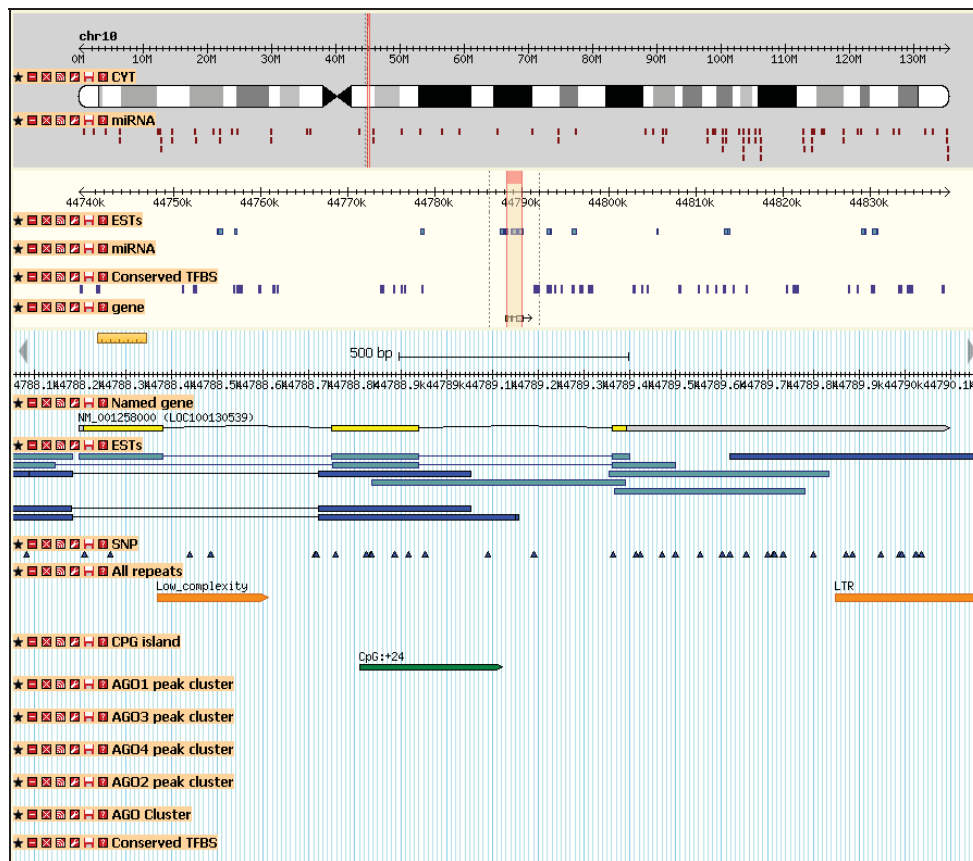
รูปที่ 16 หน้าเพจแสดงผลผลลัพธ์การค้นหาตำแหน่งเป้าหมายเปรียบเทียบระหว่างมนุษย์และหนู



รูปที่ 17 ภาพกราฟิกแสดงการเปรียบเทียบตำแหน่งเป้าหมายบนโปรโมเตอร์ระหว่างมนุษย์และหนูในยีนที่มีบรรพบุรุษร่วมกัน



รูปที่ 18 รายละเอียดของตำแหน่งเป้าหมายบนหน้ากราฟิกแสดงการเปรียบเทียบระหว่างมนุษย์และหนู



รูปที่ 19 ตัวอย่างการแสดงผลลัพธ์ของโมดูลการแสดงผลแบบกราฟิกบนจีโนมเบราว์เซอร์ของ microPIR เวอร์ชัน 2

บทวิจารณ์

โครงการนี้เป็นการพัฒนาฐานข้อมูลใหม่ที่ชื่อว่า microPIR (microRNA-Promoter Interaction Resource) โดยมีวัตถุประสงค์ให้เป็นแหล่งข้อมูลสาธารณะสำหรับการศึกษากลไกการทำงานของไมโครอาร์เอ็นเอผ่านยีนโปรโมเตอร์ของมนุษย์ คณะผู้วิจัยได้พัฒนาฐานข้อมูล microPIR เป็นเว็บแอปพลิเคชันที่สามารถให้บริการค้นหาและแสดงผลแบบกราฟิกของผลการทำนายตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอ โดยสามารถใช้งานผ่านเครือข่ายอินเทอร์เน็ตแบบออนไลน์ได้ที่ <http://www4a.biotech.or.th/micropir>

ข้อมูลหลักที่อยู่ในฐานข้อมูลนี้ ได้แก่ ข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอที่อยู่บนลำดับเบสโปรโมเตอร์ซึ่งได้มาจากการทำนายด้วยวิธีทางคอมพิวเตอร์ ซึ่งได้ถูกเชื่อมโยงกับข้อมูลลำดับเบสและข้อมูลจีโนมต่าง ๆ จากฐานข้อมูลสาธารณะที่มีอยู่ ได้แก่ ข้อมูลตำแหน่งการจับของโปรตีน AGO ในจีโนมซึ่งได้มาจากการทดลองทางห้องปฏิบัติการด้วยเทคนิค CLIP, ข้อมูลของยีนและ EST, ข้อมูลการอนุรักษ์ของลำดับเบส (conservation) ระหว่างกลุ่มสัตว์ที่มีกระดูกสันหลัง, ข้อมูลความหลากหลายทางพันธุกรรมแบบ SNP, ข้อมูล repetitive elements และข้อมูลที่เกี่ยวข้องกับการควบคุมการทำงานของยีน

(TFBS, CpG island, ไมโครอาร์เอ็นเอ) การเชื่อมโยงข้อมูลทั้งหมดเหล่านี้เข้ากับข้อมูลที่มีประโยชน์อื่นๆ ได้แก่ ข้อมูลหน้าที่ของยีน (gene ontology), ข้อมูลวิถีทางชีววิทยาจาก KEGG pathways และข้อมูลความสัมพันธ์ระหว่างยีนกับโรคและลักษณะการแสดงออกทางกายภาพอื่นๆ จากฐานข้อมูล OMIM ทำให้ฐานข้อมูล microPIR เป็นฐานข้อมูลแบบบูรณาการที่มีความครอบคลุมมากที่สุดสำหรับการศึกษาดำเนินเป้าหมายของไมโครอาร์เอ็นเอบนลำดับเบสโปรโมเตอร์

ฐานข้อมูล microPIR เวอร์ชันแรก ประกอบด้วยตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอซึ่งอยู่ภายใน 5000 เบสก่อนตำแหน่งเริ่มต้นของยีน (upstream sequences) โดยมีจำนวนมากกว่า 15 ล้านตำแหน่ง ซึ่งในจำนวนนี้ รวมทั้งตำแหน่งเป้าหมายที่อยู่บนสายดีเอ็นเอที่มีการถอดรหัสพันธุกรรม (sense strand) และสายที่ไม่ได้มีการถอดรหัสพันธุกรรม (antisense strand) ของยีนมนุษย์ โดยมีจำนวนใกล้เคียงกันทั้งสองสาย และอยู่บนโปรโมเตอร์ของ 22,357 ยีน ตำแหน่งเป้าหมายเหล่านี้คือตำแหน่งเป้าหมายทั้งหมดที่ผ่านกฎการทำนายที่ตั้งไว้ ได้แก่ 1) ตำแหน่งเป้าหมายมีลำดับเบสคู่สมแบบสมบูรณ์กับตำแหน่งที่ 2 ถึง 8 ของไมโครอาร์เอ็นเอ (seed region), 2) พลังงานในการจับระหว่างไมโครอาร์เอ็นเอและตำแหน่งเป้าหมาย ≤ -20 Kcal/mol, 3) ความยาวของตำแหน่งเป้าหมาย \leq ความยาวของไมโครอาร์เอ็นเอ+10 โดยพบตำแหน่งเป้าหมายทั้งหมดที่ได้มีการทดสอบยืนยันทางห้องปฏิบัติการแล้ว (Kim *et al.* 2008; PLACE *et al.* 2008; YOUNGER and COREY 2011) รวมอยู่ในกลุ่มที่ผ่านกฎการทำนายนี้ด้วย การที่พบตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนโปรโมเตอร์ของยีนส่วนใหญ่ ชี้ให้เห็นว่า กลไกการทำงานของไมโครอาร์เอ็นเอผ่านยีนโปรโมเตอร์อาจจะเป็นกลไกที่พบได้บ่อยกว่าที่คาดไว้

นอกจากข้อมูลตำแหน่งเป้าหมายจากการทำนายทางคอมพิวเตอร์แล้ว microPIR ยังแสดงข้อมูลจีโนม 3 ชนิดที่ใช้สนับสนุนการพิจารณาผลการทำนายตำแหน่งเป้าหมาย ได้แก่ ข้อมูลตำแหน่งการจับของโปรตีน AGO ในจีโนม, ข้อมูล EST และข้อมูลการอนุรักษ์ของลำดับเบสระหว่างกลุ่มสัตว์ที่มีกระดูกสันหลัง มีการรายงานของสมมติฐานที่ว่า ไมโครอาร์เอ็นเอที่มีกลไกการทำงานผ่านยีนโปรโมเตอร์ จะควบคุมการทำงานของยีนโดยการจับกับอาร์เอ็นเอที่ไม่ถอดรหัสเป็นโปรตีน (non-coding RNA transcript) ซึ่งมีส่วนประกอบของลำดับเบสโปรโมเตอร์ (YOUNGER and COREY 2011) เนื่องจากในฐานข้อมูล EST นั้น นอกจากจะมียีนที่ถอดรหัสเป็นโปรตีน (protein-coding gene) แล้ว ยังประกอบไปด้วยอาร์เอ็นเอจำนวนมากที่ไม่ถอดรหัสเป็นโปรตีน (noncoding transcript) (CLAVERIE 2005) ซึ่งในจำนวนนี้ มีอาร์เอ็นเอจำนวนมากที่มีส่วนประกอบของลำดับเบสโปรโมเตอร์ด้วย ดังนั้น การที่พบตำแหน่งเป้าหมายบนยีนโปรโมเตอร์ซึ่งได้จากการทำนายอยู่ในฐานข้อมูล EST จึงเป็นการสนับสนุนอย่างหนึ่งว่า ลำดับเบสของตำแหน่งเป้าหมายดังกล่าวมีการแสดงออกในเซลล์และน่าจะป็นตำแหน่งเป้าหมายที่แท้จริง นอกจากนี้ มีการรายงานในสองการศึกษาว่า พบโปรตีน AGO1 และ AGO2 จับที่บริเวณตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอบนยีนโปรโมเตอร์ ซึ่งโปรตีนดังกล่าวอาจมีส่วนร่วมในกลไกการทำงานนี้ของไมโครอาร์เอ็นเอ (Kim *et al.* 2008; YOUNGER and COREY 2011) ดังนั้น การจับของโปรตีน AGO บนตำแหน่งเป้าหมายที่ได้จากการทำนาย สามารถใช้เป็นหลักฐานหนึ่งในการสนับสนุนการเลือกตำแหน่งเป้าหมายที่น่าสนใจ อีกหลักฐานหนึ่งที่สำคัญ ได้แก่ การอนุรักษ์ระหว่างสิ่งมีชีวิตของตำแหน่งเป้าหมาย เนื่องจากโดยทั่วไปลำดับเบสบริเวณโปรโมเตอร์มีระดับการอนุรักษ์ที่ต่ำกว่าลำดับเบสที่ถอดรหัสเป็นโปรตีน (BLANCO *et al.* 2006; CONSORTIUM 2004; WATERSTON *et al.* 2002) ดังนั้น ตำแหน่งเป้าหมายบนโปรโมเตอร์ซึ่งประกอบด้วยลำดับเบสที่มีระดับการอนุรักษ์ระหว่างสิ่งมีชีวิตสูงจึงมีโอกาที่จะเป็นตำแหน่งเป้าหมายที่แท้จริง ข้อมูลต่างๆ ดังกล่าวมามีประโยชน์ในการจัดลำดับตำแหน่งเป้าหมายที่น่าสนใจสำหรับการนำไปศึกษาทางห้องปฏิบัติการต่อไป ยกตัวอย่างเช่น จากการใช้ชุดของค่า

พารามิเตอร์ที่จำเพาะสำหรับการจับระหว่างไมโครอาร์เอ็นเอและตำแหน่งเป้าหมาย ได้แก่ $MFE \leq -30$ Kcal/mol และ $p\text{-value} \leq 0.005$ และพิจารณาคัดกรองเฉพาะตำแหน่งเป้าหมายที่ไม่มี repetitive sequence เป็นส่วนประกอบ, มีระดับการอนุรักษ์ของลำดับเบสที่สูง (เลือกที่มีคะแนนเฉลี่ยของระดับการอนุรักษ์ ≥ 0.8), และมีผลการทดลองสนับสนุนคือ มีหลักฐานการจับของโปรตีน AGO1 หรือ AGO2 ที่ตำแหน่งเป้าหมาย รวมทั้งอยู่ในบริเวณของจีโนมที่มีการคัดลอก/แสดงออกตั้งแต่แสดงจากข้อมูล EST พารามิเตอร์ดังกล่าวจะสามารถลดจำนวนตำแหน่งเป้าหมายที่เป็นไปได้ลงเหลือเพียง 402 ตำแหน่งใน 297 ยีน ซึ่งในจำนวนนี้มี 211 ยีนที่พบในฐานข้อมูล OMIM ซึ่งแสดงความสัมพันธ์ระหว่างยีนกับโรคและลักษณะการแสดงออกทางกายภาพอื่นๆ

ฐานข้อมูล microPIR ได้รับการพัฒนามาเพื่อเป็นแหล่งข้อมูลใหม่สำหรับนักวิจัยด้านไมโครอาร์เอ็นเอ ความแตกต่างระหว่างรูปแบบของฐานข้อมูล microPIR และรูปแบบของฐานข้อมูลตำแหน่งเป้าหมายไมโครอาร์เอ็นเอแหล่งอื่นๆ คือ microPIR เปิดให้มีบริการค้นหาตำแหน่งเป้าหมายที่สามารถปรับเปลี่ยนได้อย่างยืดหยุ่น ทำให้ครอบคลุมรูปแบบต่างๆ ของปฏิสัมพันธ์ระหว่างไมโครอาร์เอ็นเอและตำแหน่งเป้าหมายที่ผู้ใช้สนใจ ยกตัวอย่างเช่น ถ้าต้องการคัดเลือกเฉพาะตำแหน่งเป้าหมายที่มีความน่าเชื่อถือสูง ผู้ใช้สามารถเลือกตำแหน่งเป้าหมายที่ประกอบด้วยลำดับเบสที่เป็นคู่สมแบบสมบูรณ์กับไมโครอาร์เอ็นเอ โดยกำหนดจำนวนมากที่สุดของเบสที่ไม่เป็นคู่สมกับไมโครอาร์เอ็นเอให้เป็นศูนย์ และเนื่องจากเป็นที่ยอมรับกันว่า ไมโครอาร์เอ็นเอสามารถจับกับตำแหน่งเป้าหมายที่มีลำดับเบสที่ไม่สมบูรณ์กับไมโครอาร์เอ็นเอได้ ผู้ใช้จึงสามารถปรับระดับของพารามิเตอร์ให้สอดคล้องกับตำแหน่งเป้าหมายแบบที่ต้องการ บริการค้นหาตำแหน่งเป้าหมายได้อย่างยืดหยุ่นนี้ทำให้ผู้ใช้สามารถสร้างสมมติฐานเกี่ยวกับความสัมพันธ์ระหว่างไมโครอาร์เอ็นเอ, ตำแหน่งเป้าหมาย และคุณลักษณะจีโนมอื่นๆ เช่น SNP, repetitive sequence, TFBS, CpG island ได้ นอกจากนี้จีโนมเบราร์เซอร์สามารถช่วยในการศึกษาเรื่องที่ผู้ใช้มีความสนใจ เช่น นักวิจัยที่สนใจในบทบาทของไมโครอาร์เอ็นเอที่อยู่บริเวณโปรโมเตอร์ ซึ่งจับกับตำแหน่งเป้าหมายที่อยู่ในบริเวณเดียวกัน (cis-regulation) สามารถที่จะใช้ภาพกราฟิกที่แสดงในจีโนมเบราร์เซอร์ในการสำรวจไมโครอาร์เอ็นเอที่มีตำแหน่งเป้าหมายของมันอยู่ในโปรโมเตอร์ของยีนเดียวกัน ดังตัวอย่างที่แสดงในรูปที่ 10

ในปัจจุบัน ข้อมูลการศึกษาทดลองทางห้องปฏิบัติการที่เกี่ยวกับปฏิสัมพันธ์ระหว่างไมโครอาร์เอ็นเอและลำดับเบสโปรโมเตอร์ ยังคงมีไม่มากพอที่จะนำมาใช้ในการกำหนดคุณสมบัติของตำแหน่งเป้าหมายในการทำนายทางคอมพิวเตอร์ได้อย่างละเอียดชัดเจน กฎในการทำนายตำแหน่งเป้าหมายที่มีอยู่จึงยังขึ้นอยู่กับผลการทดลองที่มีจำกัด และคาดว่าข้อมูลดังกล่าวจะยังไม่พร้อมในระยะเวลาอันใกล้ จากข้อจำกัดดังกล่าว คณะผู้วิจัยได้พัฒนาต่อยอดฐานข้อมูล microPIR เป็นเวอร์ชัน 2 โดยประเด็นหลักที่มีการพัฒนาเพิ่มจากฐานข้อมูลเดิมได้แก่ การเพิ่มข้อมูลตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆ ของจีโนมหนู จากเดิมที่มุ่งเน้นเฉพาะการศึกษาในจีโนมมนุษย์ รวมทั้งมีการปรับปรุงข้อมูลของจีโนมมนุษย์ให้เป็นปัจจุบัน โดยข้อมูลล่าสุดพบตำแหน่งเป้าหมายที่เป็นไปได้จำนวนมากกว่า 80 ล้านและ 40 ล้านตำแหน่งในจีโนมมนุษย์และหนูตามลำดับ นอกจากนี้ ในเวอร์ชันนี้ มีการเพิ่มโมดูลในการเปรียบเทียบตำแหน่งเป้าหมายระหว่างสิ่งมีชีวิตทั้งสองชนิด (comparative view) โดยแสดงผลทั้งในรูปแบบรายการผลลัพธ์และภาพกราฟิก เพื่อช่วยในการเปรียบเทียบข้อมูลและความอนุรักษ์ของตำแหน่งเป้าหมาย นอกจากนี้ ยังมีการเพิ่มข้อมูลประกอบอื่นๆ เพื่อช่วยในการตัดสินใจเลือกตำแหน่งเป้าหมายที่น่าสนใจ ได้แก่ ข้อมูลความสัมพันธ์ระหว่างไมโครอาร์เอ็นเอกับโรค, ข้อมูลยีนที่มีบรรพบุรุษร่วมกันระหว่างมนุษย์และหนู, ข้อมูล subcellular organization ของไมโครอาร์เอ็นเอที่ได้จากการทดลอง เป็นต้น ในอนาคตหากมีผลการทดลองทางห้องปฏิบัติการที่เกี่ยวข้องกับ

ปฏิสัมพันธ์ระหว่างไมโครอาร์เอ็นเอและลำดับเบสโปรโมเตอร์เพิ่มขึ้น คณะผู้วิจัยสามารถนำผลวิจัยใหม่ดังกล่าวมาใช้ทดสอบเปรียบเทียบเพิ่มเติม รวมทั้งนำมาใช้พิจารณาเป็นกฎการทำนายตำแหน่งเป้าหมายให้สมบูรณ์ยิ่งขึ้น คณะผู้วิจัยยังสามารถนำข้อมูลจีโนมใหม่ที่เกี่ยวข้องมาเชื่อมโยงกับข้อมูลในฐานข้อมูลเวอร์ชัน 2 นี้ในอนาคต เพื่อปรับปรุงฐานข้อมูลให้มีประสิทธิภาพสูงขึ้นต่อไป

บรรณานุกรม

- AMBROS, V., 2008 The evolution of our thinking about microRNAs. *Nat Med* **14**: 1036-1040.
- ASHBURNER, M., C. A. BALL, J. A. BLAKE, D. BOTSTEIN, H. BUTLER *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25-29.
- BARTEL, D. P., 2004 MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.
- BARTEL, D. P., 2009 MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215-233.
- BLANCO, E., X. MESSEGUER, T. F. SMITH and R. GUIGO, 2006 Transcription factor map alignment of promoter regions. *PLoS Comput Biol* **2**: e49.
- CHU, Y., X. YUE, S. T. YOUNGER, B. A. JANOWSKI and D. R. COREY, 2010 Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Res* **38**: 7736-7748.
- CLAVERIE, J. M., 2005 Fewer genes, more noncoding RNA. *Science* **309**: 1529-1530.
- CONSORTIUM, I. C. G. S., 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695-716.
- CONSORTIUM, I. H., 2005 A haplotype map of the human genome. *Nature* **437**: 1299-1320.
- DWEEP, H., C. STICHT, P. PANDEY and N. GRETZ, 2011 miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform* **44**: 839-847.
- ENGELS, B. M., and G. HUTVAGNER, 2006 Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* **25**: 6163-6169.
- FILIPOWICZ, W., S. N. BHATTACHARYYA and N. SONENBERG, 2008 Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**: 102-114.
- FORMAN, J. J., A. LEGESSE-MILLER and H. A. COLLIER, 2008 A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* **105**: 14879-14884.
- FUJITA, P. A., B. RHEAD, A. S. ZWEIG, A. S. HINRICHS, D. KAROLCHIK *et al.*, 2011 The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* **39**: D876-882.
- GRIFFITHS-JONES, S., H. K. SAINI, S. VAN DONGEN and A. J. ENRIGHT, 2008 miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**: D154-158.
- HAN, J., D. KIM and K. V. MORRIS, 2007 Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**: 12422-12427.
- HAWKINS, P. G., S. SANTOSO, C. ADAMS, V. ANEST and K. V. MORRIS, 2009 Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. *Nucleic Acids Res* **37**: 2984-2995.

- HIRAKAWA, M., T. TANAKA, Y. HASHIMOTO, M. KURODA, T. TAKAGI *et al.*, 2002 JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* **30**: 158-162.
- HSU, S. D., C. H. CHU, A. P. TSOU, S. J. CHEN, H. C. CHEN *et al.*, 2008 miRNAmap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* **36**: D165-169.
- HUANG, V., Y. QIN, J. WANG, X. WANG, R. F. PLACE *et al.*, 2010 RNAa is conserved in mammalian cells. *PLoS One* **5**: e8848.
- JANOWSKI, B. A., K. E. HUFFMAN, J. C. SCHWARTZ, R. RAM, D. HARDY *et al.*, 2005 Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol* **1**: 216-222.
- JANOWSKI, B. A., K. E. HUFFMAN, J. C. SCHWARTZ, R. RAM, R. NORDSELL *et al.*, 2006 Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**: 787-792.
- JANOWSKI, B. A., S. T. YOUNGER, D. B. HARDY, R. RAM, K. E. HUFFMAN *et al.*, 2007 Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* **3**: 166-173.
- JEFFRIES, C. D., H. M. FRIED and D. O. PERKINS, 2011 Nuclear and cytoplasmic localization of neural stem cell microRNAs. *RNA* **17**: 675-686.
- KANEHISA, M., S. GOTO, M. FURUMICHI, M. TANABE and M. HIRAKAWA, 2010 KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* **38**: D355-360.
- KAROLCHIK, D., A. S. HINRICHS, T. S. FUREY, K. M. ROSKIN, C. W. SUGNET *et al.*, 2004 The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**: D493-496.
- KHORSHID, M., C. RODAK and M. ZAVOLAN, 2010 CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res* **39**: D245-252.
- KIM, D. H., P. SAETROM, O. SNOVE, JR. and J. J. ROSSI, 2008 MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**: 16230-16235.
- KIM, D. H., L. M. VILLENEUVE, K. V. MORRIS and J. J. ROSSI, 2006 Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**: 793-797.
- KRUGER, J., and M. REHMSMEIER, 2006 RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* **34**: W451-454.
- LI, L. C., S. T. OKINO, H. ZHAO, D. POOKOT, R. F. PLACE *et al.*, 2006 Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* **103**: 17337-17342.
- LIAO, J. Y., L. M. MA, Y. H. GUO, Y. C. ZHANG, H. ZHOU *et al.*, 2010 Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS One* **5**: e10563.

- LOOTS, G., and I. OVCHARENKO, 2007 ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics* **23**: 122-124.
- LYTLE, J. R., T. A. YARIO and J. A. STEITZ, 2007 Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104**: 9667-9672.
- MORRIS, K. V., 2009 RNA-directed transcriptional gene silencing and activation in human cells. *Oligonucleotides* **19**: 299-306.
- MORRIS, K. V., S. W. CHAN, S. E. JACOBSEN and D. J. LOONEY, 2004 Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**: 1289-1292.
- NAM, S., B. KIM, S. SHIN and S. LEE, 2008 miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res* **36**: D159-164.
- NAPOLI, S., C. PASTORI, M. MAGISTRI, G. M. CARBONE and C. V. CATAPANO, 2009 Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J* **28**: 1708-1719.
- PIRIYAPONGSA, J., C. NGAMPHIW, A. ASSAWAMAKIN, P. WANGKUMHANG, P. SUWANNASRI *et al.*, 2009 RExPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation. *BMC Genomics* **10 Suppl 3**: S4.
- PLACE, R. F., L. C. LI, D. POOKOT, E. J. NOONAN and R. DAHIYA, 2008 MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* **105**: 1608-1613.
- RUEPP, A., A. KOWARSCH, D. SCHMIDL, F. BUGGENTHIN, B. BRAUNER *et al.*, 2010 PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol* **11**: R6.
- RUVKUN, G., 2008 The perfect storm of tiny RNAs. *Nat Med* **14**: 1041-1045.
- SCHWARTZ, J. C., S. T. YOUNGER, N. B. NGUYEN, D. B. HARDY, B. P. MONIA *et al.*, 2008 Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol* **15**: 842-848.
- SHERRY, S. T., M. H. WARD, M. KHOLODOV, J. BAKER, L. PHAN *et al.*, 2001 dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**: 308-311.
- STEIN, L. D., C. MUNGALL, S. SHU, M. CAUDY, M. MANGONE *et al.*, 2002 The generic genome browser: a building block for a model organism system database. *Genome Res* **12**: 1599-1610.
- SUZUKI, K., T. SHIJUUKU, T. FUKAMACHI, J. ZAUNDERS, G. GUILLEMIN *et al.*, 2005 Prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region. *J RNAi Gene Silencing* **1**: 66-78.
- TAFT, R. J., C. SIMONS, S. NAHKURI, H. OEY, D. J. KORBIE *et al.*, 2010 Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nat Struct Mol Biol* **17**: 1030-1034.

- TAY, Y., J. ZHANG, A. M. THOMSON, B. LIM and I. RIGOUTSOS, 2008 MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**: 1124-1128.
- TING, A. H., K. E. SCHUEBEL, J. G. HERMAN and S. B. BAYLIN, 2005 Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet* **37**: 906-910.
- WANG, X., 2008 miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* **14**: 1012-1017.
- WATERSTON, R. H., K. LINDBLAD-TOH, E. BIRNEY, J. ROGERS, J. F. ABRIL *et al.*, 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-562.
- XIAO, F., Z. ZUO, G. CAI, S. KANG, X. GAO *et al.*, 2009 miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* **37**: D105-110.
- YANG, J. H., J. H. LI, P. SHAO, H. ZHOU, Y. Q. CHEN *et al.*, 2011 starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res* **39**: D202-209.
- YOUNGER, S. T., and D. R. COREY, 2011 Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Res.*

ผลงานที่ได้จากโครงการ

1. ฐานข้อมูลสาธารณะในรูปแบบเว็บไซต์สำหรับการศึกษาดำเนินเป้าหมายของไมโครอาร์เอ็นเอบนยีนโปรโมเตอร์ในจีโนมมนุษย์ ชื่อว่า microPIR (microRNA-Promoter Interaction Resource) ซึ่งเปิดให้ใช้งานฟรีที่ <http://www4a.biotec.or.th/micropir> โดยข้อมูลตำแหน่งเป้าหมายของไมโครอาร์เอ็นเอในฐานข้อมูลได้รับความสนใจจากนักวิจัยในสาขาที่เกี่ยวข้อง ในปัจจุบันมีจำนวนผู้ใช้งานฐานข้อมูลแล้วประมาณ 3500 คนจากประเทศต่างๆทั่วโลก และคณะผู้วิจัยได้เปิดให้มีการลงทะเบียนสำหรับนักวิจัยที่ต้องการดาวน์โหลดข้อมูลตำแหน่งเป้าหมายทั้งหมดในระดับจีโนม โดยมีผู้ลงทะเบียนในระบบแล้วมากกว่า 50 คน เพื่อนำไปใช้ในโครงการวิจัยต่างๆ ปัจจุบันอยู่ในระหว่างพัฒนาฐานข้อมูล microPIR เวอร์ชัน 2 ให้เสร็จสมบูรณ์ โดยประเด็นหลักที่มีการพัฒนาเพิ่มจากฐานข้อมูลเดิม ได้แก่ การเพิ่มข้อมูลตำแหน่งเป้าหมายและคุณลักษณะจีโนมอื่นๆของจีโนมหนู จากเดิมที่มุ่งเน้นการศึกษาในจีโนมมนุษย์เท่านั้น รวมทั้งการปรับปรุงข้อมูลของจีโนมมนุษย์ให้เป็นปัจจุบัน นอกจากนี้ มีการเพิ่มโมดูลในการเปรียบเทียบตำแหน่งเป้าหมายระหว่างสิ่งมีชีวิตทั้งสองชนิด (Human-mouse comparative view) รวมทั้งเพิ่มข้อมูลสนับสนุนซึ่งเป็นผลการทดลองทางห้องปฏิบัติการที่เกี่ยวข้อง ทั้งนี้เพื่อช่วยให้การคัดเลือกตำแหน่งเป้าหมายของผู้ใช้มีประสิทธิภาพสูงขึ้น
2. บทความตีพิมพ์ในวารสารนานาชาติ PLoS ONE (Impact Factor ปี 2012 = 3.730)
Piriyapongsa J, Bootchai C, Ngamphiw C, Tongsimma S. microPIR: an integrated database of microRNA target sites within human promoter sequences. *PLoS ONE*. 2012, 7(3): e33888. (เอกสารตีพิมพ์ในภาคผนวก)
3. ผลงานวิจัยได้รับคัดเลือกเป็น Research Highlight ของ A-IMBN (Asia-Pacific International Molecular Biology Network) Research website (<http://www.natureasia.com/A-IMBN>) ในหัวข้อ "Mining and sharing microRNA data: An open-access bioinformatics database developed in Thailand will aid investigations of poorly understood gene regulation mechanisms" (เอกสารที่เกี่ยวข้องดังแสดงในภาคผนวก)
4. งานวิจัยได้รับรางวัลทุนวิจัย ลอริอัล ประเทศไทย "เพื่อสตรีในงานวิทยาศาสตร์" สาขาวิทยาศาสตร์ชีวภาพ ประจำปี 2556 จากผลงานวิจัยเรื่อง "การใช้เทคนิคชีวสารสนเทศในการศึกษาบทบาทและกลไกใหม่ของไมโครอาร์เอ็นเอในการจับกับตำแหน่งเป้าหมายบนยีนโปรโมเตอร์" โดยมีพิธีมอบทุนวิจัยลอริอัล ประเทศไทย "เพื่อสตรีในงานวิทยาศาสตร์" ประจำปี 2556 ณ โรงแรม ดับเบิลยู สาทร์ วันที่ 10 ตุลาคม 2556 (เอกสารที่เกี่ยวข้องดังแสดงในภาคผนวก)
5. การนำเสนอผลงานแบบโปสเตอร์ในงานประชุมวิชาการนานาชาติ The 7th Princess Chulabhorn International Science Congress (PC VII) CANCER: FROM BASIC RESEARCH TO CURE ในระหว่างวันที่ 29 พฤศจิกายน - 3 ธันวาคม พ.ศ. 2555 ที่โรงแรมแชงกรีล่า กรุงเทพฯ (บทความย่อแสดงในภาคผนวก)

6. รำบบทความของฐานข้อมูล microPIR เวอร์ชัน 2 เพื่อส่งเข้าพิจารณาตีพิมพ์ในวารสารนานาชาติ (อยู่ในระหว่างการเตรียม)

ภาคผนวก

microPIR: An Integrated Database of MicroRNA Target Sites within Human Promoter Sequences

Jittima Piriyaopongsa*, Chaiwat Bootchai, Chumpol Ngamphiw, Sissades Tongsima*

Genome Institute, National Center for Genetic Engineering and Biotechnology, Pathumthani, Thailand

Abstract

Background: microRNAs are generally understood to regulate gene expression through binding to target sequences within 3'-UTRs of mRNAs. Therefore, computational prediction of target sites is usually restricted to these gene regions. Recent experimental studies though have suggested that microRNAs may alternatively modulate gene expression by interacting with promoters. A database of potential microRNA target sites in promoters would stimulate research in this field leading to more understanding of complex microRNA regulatory mechanism.

Methodology: We developed a database hosting predicted microRNA target sites located within human promoter sequences and their associated genomic features, called microPIR (microRNA-Promoter Interaction Resource). microRNA seed sequences were used to identify perfect complementary matching sequences in the human promoters and the potential target sites were predicted using the RNAhybrid program. >15 million target sites were identified which are located within 5000 bp upstream of all human genes, on both sense and antisense strands. The experimentally confirmed argonaute (AGO) binding sites and EST expression data including the sequence conservation across vertebrate species of each predicted target are presented for researchers to appraise the quality of predicted target sites. The microPIR database integrates various annotated genomic sequence databases, e.g. repetitive elements, transcription factor binding sites, CpG islands, and SNPs, offering users the facility to extensively explore relationships among target sites and other genomic features. Furthermore, functional information of target genes including gene ontologies, KEGG pathways, and OMIM associations are provided. The built-in genome browser of microPIR provides a comprehensive view of multidimensional genomic data. Finally, microPIR incorporates a PCR primer design module to facilitate experimental validation.

Conclusions: The proposed microPIR database is a useful integrated resource of microRNA-promoter target interactions for experimental microRNA researchers and computational biologists to study the microRNA regulation through gene promoter. The database can be freely accessed from: <http://www4a.biotech.or.th/micropir>.

Citation: Piriyaopongsa J, Bootchai C, Ngamphiw C, Tongsima S (2012) microPIR: An Integrated Database of MicroRNA Target Sites within Human Promoter Sequences. PLoS ONE 7(3): e33888. doi:10.1371/journal.pone.0033888

Editor: I. King Jordan, Georgia Institute of Technology, United States of America

Received: December 26, 2011; **Accepted:** February 19, 2012; **Published:** March 16, 2012

Copyright: © 2012 Piriyaopongsa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JP is supported by the new researcher grant from the Thailand Research Fund (TRF) and National Center for Genetic Engineering and Biotechnology (BIOTEC). CB, CN, and ST were supported by the National Center for Genetic Engineering and Biotechnology (BIOTEC). ST would like to acknowledge the TRF grant number RSA5480026 and the chair professor grant by the National Science and Technology Development Agency (NSTDA), which partially support this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jittima.pir@biotech.or.th (JP); sissades@biotech.or.th (ST)

Introduction

MicroRNAs (miRNAs) are small ~22 nt regulatory RNA molecules that are involved in post-transcriptional regulation of gene expression through inhibition of translation initiation or targeting messenger RNAs for degradation [1,2,3]. miRNAs play important roles in regulation of numerous biological processes, such as development, cell proliferation and differentiation, metabolism, apoptosis, and the cell cycle [1,4,5,6,7]; hence, miRNAs are of medical importance. Moreover, alterations in miRNA regulation have been reported to be associated with several diseases such as cancer [8,9,10,11].

Generally, the function of a miRNA is defined by its target genes and its effect on gene expression; thus, identification of potential miRNA targets is of highest importance. Since many potential miRNA regulated genes exist in a genome, accurate bioinformatic prediction is an essential tool for identifying targets.

In animals, it is well known that miRNAs control gene expression through recognition of target sequences in the 3'-UTRs of mRNAs [4,12]. Hence, efforts to predict target sites have often been limited to 3'-UTR regions, although occasionally other gene regions, e.g. 5'-UTR, CDS, were also considered. However, the currently available target information is not enough to fully explain the complete mechanism and function of miRNAs, especially the ones with unknown targets.

Recent experimental studies have suggested an alternative miRNA mechanism for modulating gene expression by targeting outside gene bodies through promoter recognition in human cells. Place et al. [13] provided evidence of promoter-targeting miRNA by showing that introduction of miR-373 induced expression of *CDH1* and *CSDC2* containing complementary promoter sequences. Kim et al. [14] reported the *cis*-regulatory role of miR-320 in targeting its own genomic location, which resulted in transcriptional silencing of an adjacent downstream gene, *POLR3D* through

an AGO1-dependent mechanism. The recent report by Younger and Corey [15] demonstrated that miRNAs can silence gene expression by targeting gene promoters *in trans*. They showed that miRNAs with incomplete complementarity to their targets require AGO2 instead of AGO1, suggesting the possibility of different silencing mechanisms for different miRNA/target interactions. These studies indicated that the control of gene transcription by promoter targeting may be a general mechanism of miRNA gene regulation. Through promoter targeting, miRNAs may positively regulate gene expression [13] in a manner similar to RNA activation (RNAa) observed in synthetic dsRNAs targeting promoters [16,17] or direct nuclear transcriptional gene silencing (TGS) [14,15]. The underlying mechanism of regulation is still poorly understood and has not much been explored. Understanding this miRNA mechanism could lead to the development of a novel therapeutic approach for specifically controlling the expression of disease-associated target genes including the modulation of other phenotypes with associated expression profiles to promoter-targeting miRNAs.

So far, only a few miRNA-promoter interactions have been verified, possibly because there are few good candidate targets. A public resource providing genome-wide computationally identified potential miRNA binding sites on gene promoters could assist identification of biologically interesting candidates for validation [18], which is important given the paucity of experimental data for miRNA promoter targets. Despite a number of existing miRNA target databases such as miRDB [19], miRNAmap [20], miRGator [21], miRecords [22], the data of promoter targets are not readily available. To our knowledge, only recently developed miRWalk database [23] offers such data. However, the putative targets were only identified for sense strand and the detailed information of the predicted targets, which enables the screening of interesting candidates for experimental verification, is not provided.

We developed the microPIR (microRNA-Promoter Interaction Resource) database to facilitate comprehensive exploration of putative miRNA binding sites within gene promoters of the human genome. In contrast to most currently available miRNA target databases, the predicted target data are managed in a way that enables the interactive and flexible exploration of putative binding sites that are most suitable with users' hypotheses, especially in the global view along with other associated genomic features and functions. The experimentally verified binding sites of AGO proteins, which have been reported to be involved in the mechanism of promoter-targeting miRNAs [14,15], are incorporated as supporting information to enhance the effective screening of predicted target sites. The associated EST expression data including the sequence conservation data across vertebrate species of each predicted target are also incorporated into the database for the same purpose. Moreover, microPIR integrates various annotated genomic sequences, such as transposable elements (TEs) and repeats, transcription factor binding sites (TFBSs), CpG islands, and SNPs, allowing researchers to study relationships among target sites with these genomic features. All information can be visualized in a genome browser view format [24]. A local primer designing tool is incorporated into microPIR to assist experimental design of miRNA target validation. Target gene functional information is also available as link-outs from the microPIR search outputs.

Results

microPIR presents miRNA target sites on promoter regions through a web interface with supporting features not only

searching and retrieving data but also the visualizations of corresponding target sites. The web interface comprises three modules: 1) query, 2) genome browser, and 3) database statistics (Figure 1).

Query module

Two search options, basic and advanced, are provided for users to access target site information associated with the specified target gene or miRNA. The basic search supports inputs with the following types of identifiers: HUGO gene name, NCBI gene ID, miRNA accession and miRNA ID. Users can locate target sites within a certain gene by supplying a gene ID/symbol to the search form. Alternatively, a miRNA accession number/ID can be entered in order to view the list of target genes of this miRNA. This search option displays all target sites that pass the default parameter settings (see on the basic search page of web interface). The target site parameters can be customized to be more stringent with the advanced search option, which is designed to support diverse query types according to user's specific needs (Figure 2A). The adjustable parameters include strand orientation of target sites, the length of upstream region (default to 5000 bp), average conservation level of binding sites, miRNA-target binding pattern and parameters (Minimum Free Energy (MFE), *p*-value, number of unpaired nucleotides, G-U base pairs in seed region, internal loop size, and bulge size). The target output (s) can also be filtered based on the existence of supporting experimental information, AGO binding sites and/or EST, on predicted targets. The presence/absence of other annotated genomic sequences (TE and repetitive sequences, TFBSs, SNPs, CpG islands) in association with a target site can also be used as extra screening criteria as well.

The results obtained from the query page include the listing of target sites matched with the search criteria used (Figure 2B). This report shows the information of miRNA and target gene, genomic locations and length of the target sites, their positions relative to the transcription start sites of the target gene, strand (sense/antisense), MFE of miRNA-target binding and conservation level of target site. In addition, the presence/absence of different annotated sequences on each target is displayed. In the case of shared locations among different predicted targets of the same miRNA, the sites with the lowest MFE will be shown. The additional corresponding information is provided as hypertext link-outs, including gene information from NCBI gene database, gene function from Gene Ontology [25], gene biological pathways from KEGG [26], associated phenotypes from OMIM database (<http://www.ncbi.nlm.nih.gov/omim>), and miRNA information from miRBase [27]. More details on binding site including binding pattern and parameters are given with a link to the primer design service RExPrimer [28], which facilitates the design of corresponding primer pairs for the specific target site. The output can be sorted and exported as a text or BED file.

Genome browser graphical representation

Each target site can be visualized on a local genome browser (see Figure 3) implemented using a GBrowse generic browser service [24]. Through the customized viewing of the genome browser, users can visualize the position of a binding site within a gene promoter, as well as the spatial relationship between binding sites and other annotated sequences. Furthermore, the resolution of the genomic locus graphic can be adjusted from <100 bp to 1 Mbp. Certain supporting information can be displayed along with the location of target site (s). These features can be turned on/off at will.

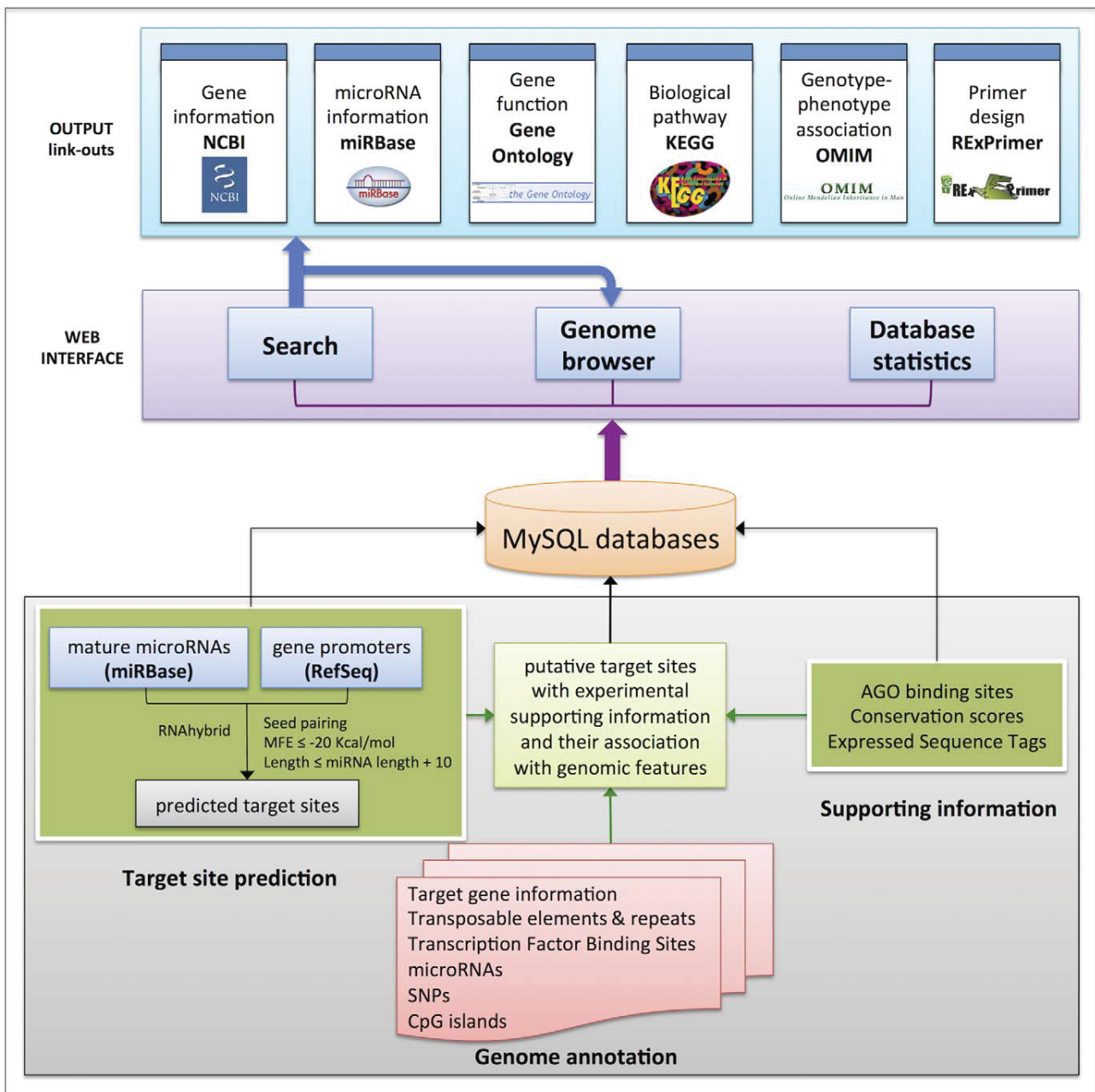


Figure 1. System overview of microPIR database. This is a three-tier system overview of the microPIR database displaying the data sources and web interface features.

doi:10.1371/journal.pone.0033888.g001

Database statistics

The summary information and initial analysis of data deposited in the microPIR database are displayed on the statistics page. The information includes build/release of miRNA/promoter sequences, number of unique targets, number of miRNA-target interactions, distribution of binding sites in terms of chromosomal locations, positions on the upstream regions, MFE, and *p*-value. The number of target sites found to be associated with each type of annotated sequence is also shown here. This information could guide researchers to make their decisions on searching parameters. This module will be automatically updated with the new release of microPIR.

Discussion

We constructed a novel database which facilitates a comprehensive exploration of miRNA-promoter interactions. The genome-wide promoter target sites of miRNAs were predicted, organized, and integrated with a number of publicly available annotated sequence databases including the database of CLIP-generated AGO binding sites, gene and EST databases, database of vertebrate sequence conservation scores, genetic variation (SNP, repeat) databases, and gene regulation (TFBS, miRNA, CpG islands) databases. By incorporating these databases together with functional annotation, microPIR offers

A

B

No	mature miRNA ID	miRNA ID	Direction	Chr. position	Strand	Length	Position	MFE	Conserved score (bp)	AGO1	AGO2	AGO3	AGO4	EST	CPG	REP	TFBS	SNP
1. detail	hsa-mir-34c-3p	hsa-mir-34c	Sense	Chr11:110889418-110889441	-	24	-1144	-48.20	1.0000 (24)	0	0	0	1	1	0	0	2	0
2. detail	hsa-mir-34c-5p	hsa-mir-34c	Sense	Chr11:110889385-110889409	-	25	-1111	-47.80	1.0000 (25)	1	1	0	1	1	0	0	0	0
3. detail	hsa-mir-34b	hsa-mir-34b	Sense	Chr11:110888921-110888944	-	24	-647	-44.50	0.9995 (24)	1	0	1	0	2	0	0	0	0
4. detail	hsa-mir-1181	hsa-mir-1181	Sense	Chr11:110888958-110888978	-	21	-684	-39.70	0.6097 (21)	0	0	1	0	2	0	0	0	0
5. detail	hsa-mir-34a	hsa-mir-34a	Sense	Chr11:110888886-110888909	-	24	-612	-39.50	0.9935 (24)	0	1	1	1	2	0	0	0	1
6. detail	hsa-mir-449b	hsa-mir-449b	Sense	Chr11:110889385-110889405	-	21	-1111	-35.10	1.0000 (21)	0	1	1	1	1	0	0	0	0
7. detail	hsa-mir-34c-5p	hsa-mir-34c	Antisense	Chr11:110888921-110888942	+	22	-4333	-32.50	0.9995 (22)	1	0	0	0	2	0	0	0	0

Figure 2. Search input and report output from the advanced search module. (A) The advanced search page is separated into three main parts. 1) A target gene or associated miRNA is put as a query. 2) The choices of binding site parameter settings can be adjusted to fit the user's needs. 3) The overlap of target site with specified annotated sequence is allowed as additional search criteria. (B) The list of resulting target sites obtained from search page is displayed. The information of associated miRNA is presented with the direction, chromosomal location, strand, length, upstream location, MFE, and conservation score of each target site including the number of bases with available score data. The number of different annotated sequences overlapping with each predicted target site is also shown. The hypertext link-outs to original sources of gene/miRNA associated information are provided. More target detail which includes a link to primer design is provided on the detail page of each target. doi:10.1371/journal.pone.0033888.g002

the most comprehensive and integrated view of these novel target sites.

Currently, the microPIR database contains a total of about 15 million unique target sites, with an approximately similar number of sense and antisense targets associated with 22,357 annotated genes. These numbers represent all possible sites on promoter sequences which pass the default parameter settings (perfect miRNA seed pairing, binding free energy ≤ -20 Kcal/mol, and target length \leq miRNA length+10) before considering any other parameter cutoff values. All experimentally verified cases of

miRNA promoter targets [13,14,15] were detected and included in the database. The presence of putative miRNA targets on most of gene upstream sequences suggests that miRNA gene regulation through promoter targeting may be a more general event than expected. In stead of merely displaying all possible computational prediction results, microPIR supplies three major types of supporting genomic information, including genome map of experimentally verified AGO binding sites, ESTs and evolutionary conservation of miRNA target sites. It has been hypothesized that promoter-targeting miRNA regulated transcription occurs

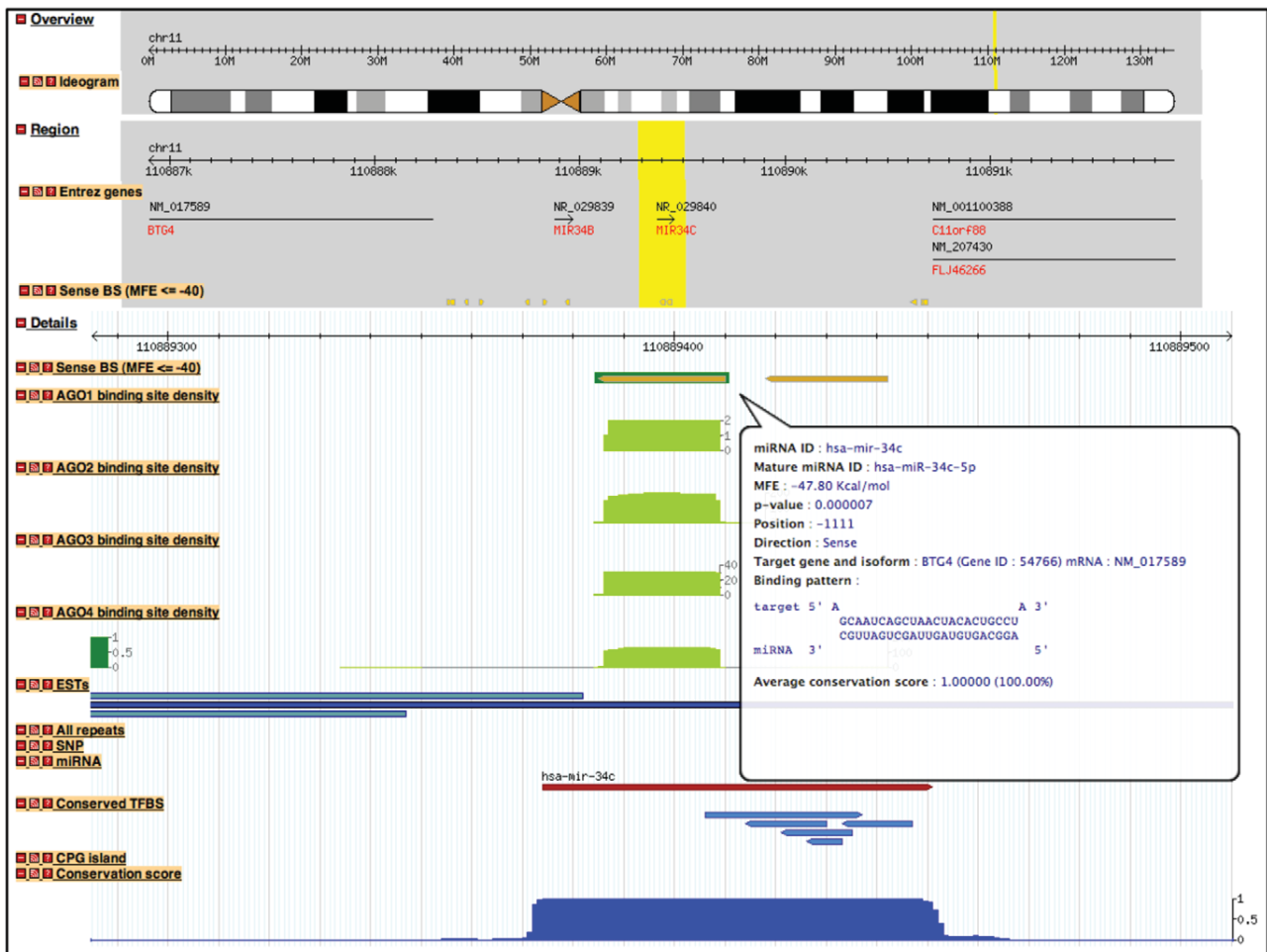


Figure 3. Genome browser displaying the resulting target sites with other genomic features. The position of target site is presented in an integrated view along with other supporting information and genomic annotations on a local genome browser. The region-level view shows the distribution of putative target sites located within a specified genomic region. Users can highlight (yellow color) to zoom in the interested location for the detail-level view. The target site is displayed in gold box with green line. The conservation score of each nucleotide position (blue color on the bottom) is displayed in the range from 0 to 1. AGO binding-site cluster is represented as distribution of read numbers along the cluster (green color). In this particular case, the presence of miRNA (red color) on the same locus as its binding site represents the *cis*-regulatory role of miRNA. doi:10.1371/journal.pone.0033888.g003

through recognition of noncoding RNA transcripts overlapping gene promoters [15]. In addition to protein-coding genes, a considerable number of noncoding transcripts originating from other genomic regions have been mapped in EST databases [29], and many of these noncoding transcripts overlap promoter regions. This observation supports the potentiality of predicted promoter targets presented in expressed sequences to be real target sites. AGO1/AGO2 proteins were demonstrated in two different studies to be recruited to the target promoter regions and were possibly involved in miRNA mechanism of action [14,15]. Thus, AGO occupancy on predicted target sites can be used as an additional supported evidence for selection of target candidates. Also, evolutionarily conserved sequences would be more likely to be biologically functional miRNA target sites, since sequence conservation in promoter regions is generally lower than in gene coding regions [30,31,32]. This information is useful to prioritize target site candidates for further experimental study. For instance, with a specific set of target binding parameters ($MFE \leq -30$ Kcal/mol, $p\text{-value} \leq 0.005$), a huge number of potential target sites can be narrowed down to 402 unique target candidates in 297 genes

(211 genes are found in OMIM database), when considering only non-repeat derived, highly conserved sites (average conservation score ≥ 0.8) accompanied with experimental support, containing AGO1 or AGO2 binding sites, and are located in the expressed genomic region, as shown by EST data.

microPIR database was developed with the main purpose to provide a new data resource for miRNA researchers. Unlike other miRNA target databases, microPIR caters for a broad range of adjustable search options, giving a better control over various kinds of interactions interested by a user. For example, highly confident sites can be selected with the perfectly complementary binding sites by setting maximum number of unpaired nucleotide to be zero. Since it is generally accepted that endogenous miRNAs do not require complete complementarity to achieve target recognition, the stringency can also be adjusted to produce the corresponding targets. These flexible search options also allow the generation of hypotheses about possible miRNA/target relationships with other genomic elements, e.g. sequence polymorphisms, repetitive elements, TFBSs, CpG islands. Moreover, the genome browser view supports users' specific interest, e.g. researchers who

are interested in the *cis*-regulatory role of promoter targeting miRNAs can look for target sites with their associated miRNA present in the same promoter region, as shown in Figure 3.

Future development

Currently, microPIR provides information of miRNA-promoter target interactions for human sequences for which experimentally determined data of AGO binding sites and other genomic data are available. In the future, we plan to include data from more organisms in order to permit the comparative features across different species. Since the knowledge on miRNA-promoter interaction is in its early stage, the computational procedures for target site prediction will soon be updated. Additional computational prediction algorithms are planned to be used in the future. The scoring system of predicted target sites will be implemented and tested once enough cases of experimentally confirmed miRNA-promoter interactions are reported. To enhance the complete picture of the miRNA regulatory network, previously determined target sites within the body of genes will be included for further analysis.

Conclusions

microPIR is the most comprehensive database specifically developed to provide an open access repository of information for miRNA target sites within human promoter sequences. Integration of considerable resources of genomic data and functional annotation allows researchers to investigate predicted miRNA-promoter interactions and generate new testable hypotheses and relevant biological insights of promoter-targeting miRNAs. The microPIR database is equipped with flexible search features including user-friendly GBrowse graphical interface and hypertext link-outs to various databases and a primer design service. We believe that microPIR will be a useful resource for the miRNA community to facilitate the discovery of new miRNA regulatory mechanism through promoter recognition.

Materials and Methods

microRNA target prediction

All known human miRNAs were downloaded from miRBase release 13.0 [27]. The 5000 bp upstream sequences of RefSeq genes were obtained from the UCSC Table Browser [33]. These sequences correspond with the hg18 assembly (human genome build 36). The promoter sequences were identified for the presence of putative miRNA target sites in both sense and antisense strands using the RNAhybrid program [34]. Complementarity to the miRNA seed sequence containing nucleotides 2–8 from the miRNA 5' end is well known as a major determinant for target recognition [1,35,36]. According to the most recent study [15], mutation of the seed sequence, but not other positions, abolished the silencing activity of promoter-targeting miRNA. As a minimum requirement, we set the desired target sequences to be completely complementary with the corresponding miRNA seed regions, the miRNA-target binding free energy ≤ -20 Kcal/mol, and the target length \leq miRNA length+10.

Data resources and processing

Crosslinking and immunoprecipitation (CLIP) coupled with deep sequencing is a high-throughput experimental technique which has been applied to AGO proteins for identification of miRNA-target interactions [37]. We incorporated the data of 999350 CLIP-determined AGO binding sequences obtained from the CLIPZ database [38] as supporting information for predicted target sites. These AGO binding sites were mapped to human

genome reference sequence hg18 (build 36.3). The sites that overlap by at least 1 bp were clustered, resulting in 1092430, 388859, 1966002 and 1317205 binding-site clusters for AGO1, AGO2, AGO3, and AGO4, respectively. The number of reads covering each nucleotide position in a cluster was counted. Each cluster is represented as the distribution of sequencing depth along the cluster.

Sequence conservation information of each predicted target site was also included into the database. Position-specific conservation scores, which were derived from multiple whole genome sequence alignments between the human and 16 other vertebrate genomes [39] were extracted from the UCSC Genome Browser [40]. These scoring values correspond to the posterior probability that a human genome site is conserved as computed by phastCons [41], and position-specific scores were averaged out across target sites to reflect the levels of evolutionary conservation among predicted target sites.

The GenBank human EST sequences were downloaded from the UCSC Table Browser [33]. In addition to transcript data, other genomic information was also included in microPIR. Transcription factor binding sites (6- to 12-nt consensus motif sequences) conserved in the human/mouse/rat alignment were retrieved from the UCSC Table Browser [33]. The genome locations and identities of transposable elements and repetitive elements were taken from annotations generated by the RepeatMasker program [42]. Finally, SNP data comprising more than nineteen million common and population-specific SNPs were obtained from various databases, namely NCBI dbSNP [43] build 129, HapMap [44] public release 27, JSNP [45] release 35, and ThaiSNP (<http://www.biotec.or.th/thaisnp>) release 2. The locations of CpG islands were taken from the UCSC Table Browser [33].

System design and implementation

The overall system design of microPIR is illustrated in Figure 1. From the bottom of this figure, three data sources 1) target site prediction, 2) supporting genomic information for predicted targets, and 3) other genome annotations were preprocessed/collected and incorporated into the local MySQL databases. MySQL version 5.5.1 was employed to manage all the predicted target entries as well as their supporting information. To make the database accessible for public use, the web interface, including submission forms and graphical outputs were constructed using Python scripts and toolkits from Python Webware (<http://www.webwareforpython.org>). The web interface offers three main functionalities. First, the search feature allows users to locate miRNA target sites of interest. Data can be queried through the webform, which formulates the corresponding SQL queries to MySQL using python scripting language. We used Python MySQLdb module to connect (sending and receiving SQL queries/outputs) to the MySQL database backend. Second, the viewing feature enables users to graphically view the location of miRNA target sites within a specific locus via a genome browser interface. Third, the statistics module summarizes the information contained within this database for the genomic region of interest. The link-out module is provided for users to cross-check with other related databases, e.g. gene ontology and original data sources. Furthermore, for the sake of convenience, a primer design section is included to assist in validating potential target genes. The whole microPIR framework is running on our 12-core database server (2 AMD 6-core (2.8 GHz) processors with 64 Gigabytes of RAM and 2 Terabytes of hard disk space).

Acknowledgments

The authors would like to thank Prof. Apiwat Mutirangura for his encouragement, discussion and support to the construction of the microPIR database and Dr. Philip J. Shaw for helpful remarks on the manuscript. We would like to acknowledge the National Center for Genetic Engineering and Biotechnology (BIOTEC), and the National Science and Technology Development Agency (NSTDA), Thailand for allowing us to host this database on the web/database server, and open for

public access. We also thank Uttapong Ruangrit for designing the microPIR web logo.

Author Contributions

Conceived and designed the experiments: JP. Performed the experiments: CB CN. Analyzed the data: JP CB CN ST. Contributed reagents/materials/analysis tools: JP ST. Wrote the paper: JP ST.

References

- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Engels BM, Hutvagner G (2006) Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* 25: 6163–6169.
- Ruvkun G (2008) The perfect storm of tiny RNAs. *Nat Med* 14: 1041–1045.
- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215–233.
- Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. *Science* 301: 336–338.
- Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9: 102–114.
- Ambros V (2004) The functions of animal microRNAs. *Nature* 431: 350–355.
- Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, et al. (2011) Biological functions of microRNAs: a review. *J Physiol Biochem* 67: 129–139.
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, et al. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* 37: D98–104.
- Zhang C (2008) MicroRNomics: a newly emerging approach for disease biology. *Physiol Genomics* 33: 139–147.
- Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, et al. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* 103: 2257–2261.
- Ambros V (2008) The evolution of our thinking about microRNAs. *Nat Med* 14: 1036–1040.
- Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* 105: 1608–1613.
- Kim DH, Saetrom P, Snove O, Jr., Rossi JJ (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* 105: 16230–16235.
- Younger ST, Corey DR (2011) Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Res*.
- Janowski BA, Younger ST, Hardy DB, Ram R, Hufman KE, et al. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* 3: 166–173.
- Li LC, Okino ST, Zhao H, Pookot D, Place RF, et al. (2006) Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* 103: 17337–17342.
- Younger ST, Pertsemidis A, Corey DR (2009) Predicting potential miRNA target sites within gene promoters. *Bioorg Med Chem Lett* 19: 3791–3794.
- Wang X (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* 14: 1012–1017.
- Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, et al. (2008) miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* 36: D165–169.
- Nam S, Kim B, Shin S, Lee S (2008) miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res* 36: D159–164.
- Xiao F, Zuo Z, Cai G, Kang S, Gao X, et al. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 37: D105–110.
- Dweep H, Sticht C, Pandey P, Gretz N (2011) miRWalk-database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J Biomed Inform* 44: 839–847.
- Stein LD, Mungall C, Shu S, Caudy M, Mangone M, et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* 12: 1599–1610.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38: D355–360.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36: D154–158.
- Piriyapongsa J, Ngamphiw C, Assawamakin A, Wangkumhang P, Suwannasri P, et al. (2009) REXPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation. *BMC Genomics* 10 Suppl 3: S4.
- Claverie JM (2005) Fewer genes, more noncoding RNA. *Science* 309: 1529–1530.
- Blanco E, Messegue X, Smith TF, Guigo R (2006) Transcription factor map alignment of promoter regions. *PLoS Comput Biol* 2: e49.
- Consortium ICGS (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, et al. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32: D493–496.
- Kruger J, Rehmsmeier M (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 34: W451–454.
- Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27: 91–105.
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115: 787–798.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141: 129–141.
- Khorshid M, Rodak C, Zavolan M (2010) CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res* 39: D245–252.
- Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, et al. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* 14: 708–715.
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, et al. (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 39: D876–882.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15: 1034–1050.
- Smit AFA, Hubley R, Green P RepeatMasker Open-3.0.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29: 308–311.
- Consortium IH (2005) A haplotype map of the human genome. *Nature* 437: 1299–1320.
- Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, et al. (2002) JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 30: 158–162.



Mining and sharing microRNA data

30 May 2012



An open-access bioinformatics database developed in Thailand will aid investigations of poorly understood gene regulation mechanisms

A comprehensive and freely available database for studies on how microRNAs regulate gene expression is now available, thanks to a research team led by Jittima Piriyaongsa and Sissades Tongsimas of the National Center for Genetic Engineering and Biotechnology, Thailand¹.

MicroRNAs are short nucleic acid molecules that control gene expression by binding to the ends of messenger RNA transcripts, then preventing their translation or promoting their degradation. They can also modulate gene expression by targeting regulatory sequences called promoters, but only a few microRNA-promoter interaction sites have been identified to date.

Piriyaongsa, Tongsimas and their colleagues used a bioinformatics approach to match microRNA sequences to their target sites in human promoter sequences; they identified more than 15 million target sites associated with about 22,300 genes. They then integrated these data with a number of other publicly available genome databases to allow exploration of the relationships among target sites and various genomic features.

The resulting database, called the microRNA-Promoter Interaction Resource (microPIR), is fully annotated with gene information and has an adjustable search, making it the most comprehensive open access repository of microRNA promoter target site information. The database should facilitate researchers' efforts to identify microRNA-promoter interaction sites and gain a better understanding of these new regulatory mechanisms.



1

Reference

1. Piriyaongsa, J.,¹ Bootchai, C.,¹ Ngamphiw, C.¹ & Tongsimas, S.¹ microPIR: An integrated database of microRNA target sites within human promoter sequences. *PLoS ONE* 7, e33888 (2012). | [article](#) |

Authors and Affiliations

1. Genome Institute, National Center for Genetic Engineering and Biotechnology, Pathumthani, Thailand

A-IMBN Research

© 2012 Asia-Pacific International Molecular Biology Network and Nature Japan K.K., trading as NPG Nature Asia-Pacific



L'OREAL THAILAND
And
The Thai National Commission for UNESCO

Presents this recognition to

Dr. Jittima Piriyaopongsa

For being selected
2013 Fellow in Life Science
On October 10th, 2013

May your research project be carried out successfully.
And may it yield much benefits to the society in the future.

Dr. Kopr Kritayakirana
President of Honorary Jury

Unesh Phadke
Managing Director L'Oréal Thailand

Headline	Grant to research scholarship		
MediaTitle	Daily News (Afternoon)		
Date	18 Oct 2013	Language	Thai
Circulation	900,000	Readership	2,700,000
Section	Social and Society	Color	Full Color
Page No	5	ArticleSize	16 Inch
Brand	Loreal FWIS 2013	AD Value	BTH 33,504



มอบทุนวิจัย จาตุรนต์ ฉายแสง รมว.ศึกษาธิการ เป็นประธานพิธีมอบทุนวิจัย ลอริอัล ประเทศไทย “เพื่อสตรีในงานวิทยาศาสตร์” ประจำปี 2556 ให้แก่ ผศ.ทพญ.ดร.ศรัณยา ตันเจริญ ดร.จิตติมา พิริยะพงศา ดร.วริดาพร บัวเจริญ และ ดร.ศรชล โยริยะ โดยมี มร.อูเมซ พัดเค และ ดร.กอบร กฤตยาภิรมณ์ ร่วมแสดงความยินดี ที่โรงแรมดับเบิลยู สาท

Headline	Two outstanding female researchers of 'BIOTEC'		
MediaTitle	Daily News (Afternoon)		
Date	17 Oct 2013	Language	Thai
Circulation	900,000	Readership	2,700,000
Section	ICT, IT and Communication	Color	Full Color
Page No	10	ArticleSize	58 Inch
Brand	Loreal FWIS 2013	AD Value	BTH 124,971

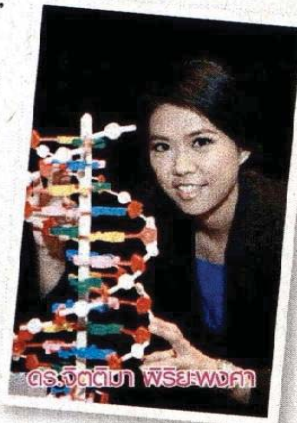
สองนักวิจัยหญิงเก่ง 'ไบโอเทค'



เพราะงานด้านวิทยาศาสตร์ไม่ได้เป็นข้อจำกัดความสามารถของผู้หญิง
วันนี้..จะพาไปรู้จักกับ 2 นักวิจัยหญิงเก่งแห่งไบโอเทคหรือศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ที่เพิ่งได้รับการคัดเลือกหมวด ๑ ให้ได้รับทุนในโครงการทุนวิจัย ลอริอัล ประเทศไทย “เพื่อสตรีในงานวิทยาศาสตร์” ประจำปีนี้

คนแรกก็คือ “ดร.จิตติมา พิริยะพงศ์” ซึ่งทำวิจัยเรื่อง “การใช้เทคนิคชีวสารสนเทศในการศึกษาบทบาทและกลไกใหม่ของไมโครอาร์เอ็นเอในการจับกับตำแหน่งเป้าหมายบนยีนโปรโมเตอร์”

ดร.จิตติมา บอกว่างานวิจัยนี้ได้รับการคัดเลือก



ดร.จิตติมา พิริยะพงศ์

ให้ได้รับทุนในสาขาวิทยาศาสตร์ชีวภาพ โดยเป็นการใช้เทคนิคชีวสารสนเทศ หรือการนำความรู้ด้านชีววิทยาและคอมพิวเตอร์มาใช้ร่วมกัน เพื่อทำนายหาตำแหน่งเป้าหมายของ

ไมโครอาร์เอ็นเอที่อยู่บนลำดับเบสโปรโมเตอร์ของยีนกว่าสามหมื่นชิ้นในจีโนมคน

และจัดตั้งเป็นฐานข้อมูลแบบบูรณาการแห่งแรกในโลกของข้อมูลตำแหน่งเป้าหมายดังกล่าว เผยแพร่ในรูปแบบเว็บไซต์เพื่อให้ผู้ใช้งานเข้าศึกษาและใช้งานได้ฟรี

ปัจจุบันมีการใช้งานแล้วกว่า 50 กลุ่มวิจัยจากทั่วโลก ซึ่งงานวิจัยดังกล่าว หากทดลองในห้องปฏิบัติการจะใช้



ดร.ธรรมาพร บัวเจริญ

เวลานานและมีค่าใช้จ่ายสูง ดังนั้นนักวิจัยสามารถใช้ฐานข้อมูลที่ทำคัด

กรองตำแหน่งเป้าหมาย และโฟกัสชิ้นที่สนใจได้ทันที ก่อนที่จะนำมาศึกษาต่อไปในเชิงลึก ซึ่งสามารถช่วยลดเวลาและค่าใช้จ่าย รวมทั้งเพิ่มโอกาสความสำเร็จในการดำเนินวิจัยอีกด้วย

สำหรับนักวิจัยหญิงเก่งอีกหนึ่งคนของไบโอเทคก็คือ “ดร.ธรรมาพร บัวเจริญ” ที่ได้รับทุนในสาขาวิทยาศาสตร์เคมี จากงานวิจัยเรื่อง “การค้นหายีนออกฤทธิ์ทางชีวภาพจากเชื้อจุลินทรีย์ที่พบในประเทศไทย”

ดร.ธรรมาพร อธิบายถึงงานวิจัยที่ทำอยู่ว่า



4 นักวิจัยสตรีเพื่อวิทยาศาสตร์ลอริอัล

Headline	Two outstanding female researchers of 'BIOTEC'		
MediaTitle	Daily News (Afternoon)		
Date	17 Oct 2013	Language	Thai
Circulation	900,000	Readership	2,700,000
Section	ICT, IT and Communication	Color	Full Color
Page No	10	ArticleSize	58 Inch
Brand	Loreal FWIS 2013	AD Value	BTH 124,971



เป็นการมุ่งพัฒนาจากสารออกฤทธิ์ทางชีวภาพของเชื้อจุลินทรีย์ที่ประเทศไทยมีความหลากหลายเป็นจำนวนมาก สารเหล่านี้สามารถนำมาพัฒนาเป็นยารักษาโรค เช่น ยับยั้งเชื้อมาลาเรีย วัณโรค เชื้อราและแบคทีเรียก่อโรคต่าง ๆ และยังสามารถใช้ประโยชน์ทางการเกษตร เช่น การป้องกัน การเกิดโรคในพืชได้อีกด้วย

ทั้งนี้จากการวิจัยที่ผ่านมา ทีมวิจัยสามารถค้นพบสารออกฤทธิ์ทางชีวภาพใหม่ ๆ ที่ไม่เคยมีมาก่อนหลายชนิด

สำหรับทุนวิจัยสตรีลอรีอัลปีนี้ ยังมีนักวิจัยหญิงเก่ง อีก 2 ท่านที่ได้รับทุนนี้ด้วยคือ ผศ.ทพญ.ดร.ศรัณยา ดันเจริญ จากภาควิชาเภสัชวิทยา คณะทันตแพทยศาสตร์ มหาวิทยาลัยมหิดล จากงานวิจัยเรื่อง “การผลิตโมโนโคลนอลแอนติบอดีที่จำเพาะกับเปปไทด์ขนาดเล็กสำหรับรักษาโรคปริทันต์อักเสบชนิดลุกลาม” และ ดร.ศรชล โยริยะ จากศูนย์เทคโนโลยีโลหะและวัสดุแห่งชาติ (เอ็มเทค) สวทช. จากงานวิจัยเรื่อง “โครงการการขึ้นรูปฟิล์มท่อนาโนไทเทเนียมและการประยุกต์ใช้งานด้านชีวการแพทย์”

ทุนดังกล่าวจัดอย่างต่อเนื่องมาเป็นปีที่ 11 ทำให้คนไทยได้มีโอกาสรู้จักนักวิจัยหญิงที่ทำงานด้านวิทยาศาสตร์ของประเทศมาแล้ว ถึง 44 คน

รางวัลนี้ แม้จะมีส่วนในการให้กำลังใจกับนักวิจัยสตรี แต่คงไม่ใช่แค่นั้น เพราะนางสดับ วัฒนกิจพิชญ์ ผู้อำนวยการฝ่ายสื่อสารองค์กร และองค์กรสัมพันธ์ บริษัท ลอรีอัล (ประเทศไทย) จำกัด บอกว่า จากการพูดคุยกับนักวิจัยที่ได้รับทุนที่ผ่านมา มักจะบอกว่ารางวัลนี้เป็น “จุดเปลี่ยนของชีวิต”

จากผู้ที่อยู่เบื้องหลัง ทำงานในห้องแล็บทดลอง กลายเป็นที่รู้จัก และได้รับโอกาสจากหน่วยงานต่าง ๆ มากขึ้น ซึ่งมีส่วนช่วยในการต่อยอดงานวิจัยให้เกิดการใช้งานจริงได้มากขึ้นอีกด้วย.

นัตตยา คชินทร
nattayap.k@gmail.com

Headline	L Oreal awards four Thai women researchers (Oct 31 to Nov 3)		
MediaTitle	Prachachart Turakij		
Date	31 Oct 2013	Language	Thai
Circulation	150,000	Readership	450,000
Section	Kid and Education	Color	Full Color
Page No	30	ArticleSize	71 Inch
Brand	Loreal FWIS 2013	AD Value	BTH 87,877



‘ลอรีอัล’ ชู 4 นักวิจัยสตรีไทย คิดค้นนวัตกรรมขึ้นใหม่ จุดกำเนิดคุณภาพชีวิต

ฟานพันไปไม่นานสำหรับการประกาศ
ผลผู้ที่ได้รับทุนจาก “โครงการทุน
วิจัย ลอรีอัล ประเทศไทย” เพื่อ
สตรีในงานวิทยาศาสตร์ (For Women
in Science) ประจำปี 2556 ซึ่งเกิดจาก
ความร่วมมือกันระหว่าง บริษัท ลอรีอัล
(ประเทศไทย) จำกัด และสำนักเลขาธิการ
แห่งชาติ ว่าด้วยการศึกษาวิทยาศาสตร์
และวัฒนธรรมแห่งสหประชาชาติ (ยูเนสโก)
การดำเนินการในครั้งนี้ถือเป็นหนึ่งใน
นโยบายของลอรีอัลทั่วโลก ที่ต้องการผลักดัน
และมีส่วนสนับสนุนให้นักวิจัยสตรีทุกคน
คิดค้นงานวิจัยวิทยาศาสตร์จนสามารถ
ต่อยอดไปสู่การสร้างผลงานที่มีคุณค่า
ต่อสังคมโดยรวม

“ระดับพัน คำฉันทิพย์” ผู้อำนวยการ
ฝ่ายสื่อสารองค์กรและองค์กรสัมพันธ์ บริษัท
ลอรีอัล (ประเทศไทย) จำกัด บอกว่า
ได้มีการสานต่อโครงการทุนวิจัย ลอรีอัล
ประเทศไทย มาอย่างต่อเนื่องถึงปีที่ 11

“เราคัดเลือกและส่งมอบทุนให้กับ

นักวิจัยสตรีไทย จำนวน 250,000 บาทต่อทุน
ซึ่งมีเกณฑ์ในการพิจารณาจากคุณค่าของ
งานวิจัยที่สามารถเกิดขึ้นกับสังคมได้ ภายใต้
หลักการ 3 ข้อ คือ 1.กระบวนการวิจัย
ถูกต้องและเหมาะสมกับวัตถุประสงค์ของ
โครงการ 2.จริยธรรมในการทำงานของ
นักวิจัย 3.ต้องเป็นที่ยอมรับในวงการนักวิจัย”
จากการคัดเลือกในครั้งนี้มีนักวิจัย
ไทยได้รับทุนทั้งหมด 4 คน ซึ่งแต่ละคน
มีผลงานวิจัยในแต่ละสาขาที่น่าสนใจ
แตกต่างกัน ดังนี้

ผลงานชิ้นแรก การผลิตโมโนโคลนอล
แอนติบอดีที่จำเพาะกับเปปไทด์ขนาดเล็ก
สำหรับรักษาโรคปริทันต์อักเสบชนิดลุกลาม
ของ “ผศ.ทพ.ญ.ดร.ศรัณยา ตันเจริญ”
สาขาวิทยาศาสตร์ชีวภาพ จากภาควิชา
เภสัชวิทยา คณะทันตแพทยศาสตร์
มหาวิทยาลัยมหิดล

“พบว่าผู้หญิงที่มีอายุตั้งแต่ 60 ปีขึ้นไป
มักเกิดโรคปริทันต์อักเสบถึง 80% ซึ่งขณะ
นี้ยังไม่มีวัคซีนที่ช่วยในการรักษาได้ งาน

วิจัยชิ้นนี้มุ่งไปที่การผลิตโมโนโคลนอล
แอนติบอดีที่จำเพาะกับเปปไทด์ขนาดเล็ก
และหากการวิจัยสำเร็จ เราจะสามารถผลิต
วัคซีนใหม่เพื่อป้องกันการเกิดโรคดังกล่าวนี้
ได้ในอนาคต”

ผลงานชิ้นถัดมาเกี่ยวข้องกับการใช้
เทคนิคชีวสารสนเทศ ในการศึกษาบทบาท
และกลไกใหม่ของไมโครอาร์เอ็นเอ ในการ
จับตำแหน่งเป้าหมายบนยีนโปรโมเตอร์
ของ “ดร.จิตติมา พิริยะพงศา” สาขา
วิทยาศาสตร์ชีวภาพ จากศูนย์พันธุวิศวกรรม
และเทคโนโลยีชีวภาพแห่งชาติ สำนักงาน
พัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ
(สวทช.)

โดยผลงานชิ้นนี้ “ดร.จิตติมา” ได้
ใช้เทคนิคชีวสารสนเทศ จากการนำเอา
ศาสตร์ทางชีววิทยาและคอมพิวเตอร์มาใช้
ร่วมกัน เพื่อทำนายหาตำแหน่งเป้าหมาย
ของไมโครอาร์เอ็นเอที่อยู่บนลำดับเบส
โปรโมเตอร์ของยีน และจัดตั้งฐานข้อมูลนี้
ไว้ในเว็บไซต์

Headline	L Oreal awards four Thai women researchers (Oct 31 to Nov 3)		
MediaTitle	Prachachart Turakij		
Date	31 Oct 2013	Language	Thai
Circulation	150,000	Readership	450,000
Section	Kid and Education	Color	Full Color
Page No	30	ArticleSize	71 Inch
Brand	Loreal FWIS 2013	AD Value	BTH 87,877



“ผู้สนใจสามารถเข้าศึกษาและใช้งานได้ฟรี ถือว่าเป็นฐานข้อมูลแบบบูรณาการแห่งแรกของข้อมูลตำแหน่งเป้าหมายก่อนจะนำมาศึกษาต่อในเชิงลึก ซึ่งสามารถช่วยลดเวลาและค่าใช้จ่าย รวมทั้งเป็นการเพิ่มโอกาสความสำเร็จในการดำเนินการวิจัยขึ้นอื่น ๆ ได้อีกทางหนึ่ง”

สำหรับผลงานชิ้นที่ 3 คือ การค้นหาสารออกฤทธิ์ทางชีวภาพจากเชื้อจุลินทรีย์ที่พบในประเทศไทย ของ “ดร.ธรรดาพร บัวเจริญ” สาขาวิทยาศาสตร์เคมี จาก

ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ สวทช. ได้ค้นพบว่าประเทศไทยมีความหลากหลายทางทรัพยากรชีวภาพ และจุลินทรีย์ค่อนข้างสูง

“สารออกฤทธิ์ของจุลินทรีย์บางชนิดมีประโยชน์ทางการแพทย์ การเกษตร อุตสาหกรรมอาหาร และสิ่งแวดล้อม หากเราได้ค้นคว้าวิจัย จะค้นพบสารออกฤทธิ์ทางชีวภาพใหม่ที่มีศักยภาพเพื่อนำไปพัฒนาต่อสำหรับเป็นยารักษาโรคหรือใช้ประโยชน์ทางการเกษตร

ซึ่งนอกจากจะช่วยเพิ่มคุณภาพชีวิตของคนไทยแล้ว ยังเป็นการช่วยเพิ่มมูลค่าให้กับจุลินทรีย์อีกด้วย”

และผลงานสุดท้าย โครงการการขึ้นรูปฟิล์มท่อนาโนไทเทเนียมและการประยุกต์ใช้งานด้านชีวการแพทย์ ของ “ดร.ศรชล โยริยะ” สาขาวัสดุศาสตร์ จากศูนย์เทคโนโลยีโลหะและวัสดุแห่งชาติ เกิดจากความสนใจในการศึกษาคุณสมบัติของท่อนาโนไทเทเนียม โดยออกแบบโครงสร้างขนาดและรูปร่างให้เหมาะกับการใช้งานในแต่ละด้าน

“การนำมาประยุกต์ใช้งานทางด้านชีวการแพทย์ เช่น ใช้เป็นวัสดุในการนำส่งตัวยา วัสดุห้ามเลือด และทำโครงสร้างรองรับเซลล์ในวิศวกรรมเนื้อเยื่อและกระดูก ซึ่งฟิล์มท่อนาโนที่ค้นพบนี้มีคุณสมบัติที่เข้ากับเลือดและร่างกายมนุษย์ สามารถส่งผ่านโมเลกุลยาหรือโปรตีนผ่านตัวท่อ เพื่อช่วยในการนำส่งตัวยาได้อย่างมีประสิทธิภาพ”

จากผลงานทั้งหมดนี้ “สตีบพิณ” เชื่อมั่นว่า การค้นคว้าวิจัยจะก่อให้เกิดความก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยี ซึ่งเป็นกุญแจดอกสำคัญสู่คุณภาพชีวิตที่สมดุล สมบูรณ์ และสวยงามยิ่งขึ้น และส่งผลสู่การสร้างสรรคคุณภาพชีวิตให้กับสังคมในอนาคต

EFFECT OF RUTHENIUM(II)-ARENE (PTA) ON SECONDARY STRUCTURE OF THE BRCA1 RING PROTEIN

PC-09

Pornwicha Temboot¹, Paul Joseph Dyson², Adisorn Ratanaphan¹

¹Laboratory of Pharmaceutical Biotechnology, Department of Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Prince of Songkla University, Hat-Yai, Songkhla 90110, Thailand, ²Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland

A novel approach for cancer therapy involves alterations to the DNA repair pathways by which cancerous cells with dysfunctional DNA repair pathways accumulate high levels of DNA damage that eventually result in major genomic instability and cell death. The breast cancer susceptibility protein 1 (BRCA1) participates in genomic integrity maintenance through DNA repair, cell cycle checkpoint, protein ubiquitination, and transcription regulation. The N-terminus of BRCA1 contains a RING domain plays an essential role in response to DNA damage. Preclinical and clinical studies have recently revealed that the dysfunction of BRCA1 in cancer cells leads to chemosensitivity. Therefore, approaching the BRCA1 RING domain as a potentially molecular target for a metal-based drug might be of interest in cancer chemotherapy. The ruthenium(II)-arene (PTA) or RAPTA complexes, (PTA = 1,3,5-triaza-7-phosphaadamantane) exhibit promising antitumor properties. However, the mechanisms of action of these complexes are unexplored. In this study, the *in vitro* ruthenation of the BRCA1 RING protein by the RAPTA complexes were observed. SDS-PAGE revealed that RAPTA complexes formed intermolecular BRCA1 adducts. Circular dichroism showed that the RAPTA complexes perturbed the secondary structure of the BRCA1 RING domain and underwent more folded structural rearrangement after increasing RAPTA complexes concentrations. Our findings demonstrated the first conformational evidence for a direct binding of the RAPTA complexes to the BRCA1 RING domain and might raise a possibility of selectively targeting of cancer with an improved efficacy and less toxicity.

PC-10

MicroPIR: AN INTEGRATED DATABASE OF microRNA-PROMOTER INTERACTIONS IN HUMAN GENOME

Jittima Piriyaopongsa, Chaiwat Bootchai, Chumpol Ngamphiw, Sissades Tongsim

National Center for Genetic Engineering and Biotechnology, Pathumthani, Thailand

Alterations in microRNA regulation have been reported to be associated with several diseases including cancer. Recent experimental studies have suggested that microRNAs may modulate gene expression by interacting with promoters. microRNAs may positively regulate gene expression in a manner similar to RNA activation (RNAa) or alternatively direct transcriptional gene silencing (TGS). The factors and promoter binding pattern which determine the choice of expression control are not clear. The exact understanding of this new microRNA regulatory mechanism could lead to the development of a novel therapeutic approach for specifically controlling the expression of cancer-associated target genes. A database of genome-wide potential microRNA target sites in promoters would stimulate research in this field and facilitate the discovery of target genes for cancer treatment. We developed a database hosting predicted microRNA target sites located within human promoter sequences and their associated genomic features, called microPIR (microRNA-Promoter Interaction Resource). The experimentally confirmed argonaute (AGO) binding sites and EST expression data including the sequence conservation of each predicted target are presented as supporting information for researchers to appraise the quality of predicted target sites. The system integrates considerable resources of annotated genomic data, e.g., repetitive elements, transcription factor binding sites, and SNPs, offering users the facility to extensively explore relationships among target sites and other genomic features. Importantly, functional information of target genes including gene ontologies, KEGG pathways, and OMIM associations are provided. The infrastructure is also equipped with flexible search features including user-friendly genome browser graphical interface which provides a comprehensive view of multidimensional genomic data. Finally, the database incorporates a PCR primer design module to support experimental validation. The database can be freely accessed from: <http://www4a.biotec.or.th/micropir>.

COMMITTEE

NATIONAL ORGANIZING COMMITTEE:

Professor Dr. HRH Princess Chulabhorn **Chairperson**

Charas Suwanwela

Khunying Mathuros Ruchirawat

Chirayu U. Auewarakul

Vice-Chairperson

Secretary, Basic Research Program

Secretary, Clinical Research Symposium Program

Members:

Chawat	Arthayukti	Somsak	Ruchirawat
Gen. Charn	Boonprasert	Jutamaad	Satayavivad
Sunathee	Isvarphornchai	Taveepong	Seniwong Na Ayudhaya
Sorajak	Kasemsuvan	Teerapong	Sodaari
Thiravud	Khuhaprema	M.R. Jisnuson	Svasti
<i>Khunying</i> Laxanachantorn	Laohaphan	Sumet	Tantivejkul
Piamsak	Milintachinda	Supanna	Techasakul
Thanuttkhul	Mongkolaussavarat	<i>Khunying</i> Charatsri	Teepirach
Skorn	Mongkolsuk	Pairash	Thajchayapong
Pittayapoom	Pattaranutaporn	Anurak	Thananan
Thakur	Phanit	<i>Thanphuying</i> Putrie	Viravaidya
Kavi	Ratanabanangkoon	Adm. Narong	Yuthavong
Damrong	Ratanapanich	Commissioner-General, Royal Thai Police	

INTERNATIONAL ADVISORY COMMITTEE:

Herman N. Autrup	(Denmark)	Leona D. Samson	(U.S.A.)
Frederick F. Becker	(U.S.A.)	Ram Sasisekharan	(U.S.A.)
Prapote Clongsusuek	(Thailand)	Niramol Savaraj	(U.S.A.)
John M. Essigmann	(U.S.A.)	William A. Suk	(U.S.A.)
Lynn G. Feun	(U.S.A.)	James E. Trosko	(U.S.A.)
Curtis C. Harris	(U.S.A.)	Christopher P. Wild	(France)
Teerawat Kulthanan	(Thailand)	Robert H. Wiltrout	(U.S.A.)
Enrico Mihich	(U.S.A.)	Gerald N. Wogan	(U.S.A.)
Sompone Punyagupta	(Thailand)	Harald zur Hausen	(Germany)

SCIENTIFIC PROGRAM COMMITTEE:

William W. Au	(P.R. China)	Kavi Ratanabanangkoon	(Thailand)
Norbert Frank	(Germany)	<i>Khunying</i> Mathuros Ruchirawat	(Thailand)
Skorn Mongkolsuk	(Thailand)	Somsak Ruchirawat	(Thailand)
Apiwat Mutirangura	(Thailand)	Vichien Srimuninnimit	(Thailand)
Xin Wei Wang	(U.S.A.)	M.R. Jisnuson Svasti	(Thailand)
Medhi Wangpaichitr	(U.S.A.)	Torpol Wattana	(Thailand)