

ชีวสารสนเทศ: กระบวนทัศน์ใหม่ของการวิจัยยุคหลังจีโนม

ศาสตราจารย์วิชาติ สิริวราภรณ์

ภาควิชาชีวเคมี และหน่วยวิจัยชีวสารสนเทศและจีโนมประยุกต์

คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

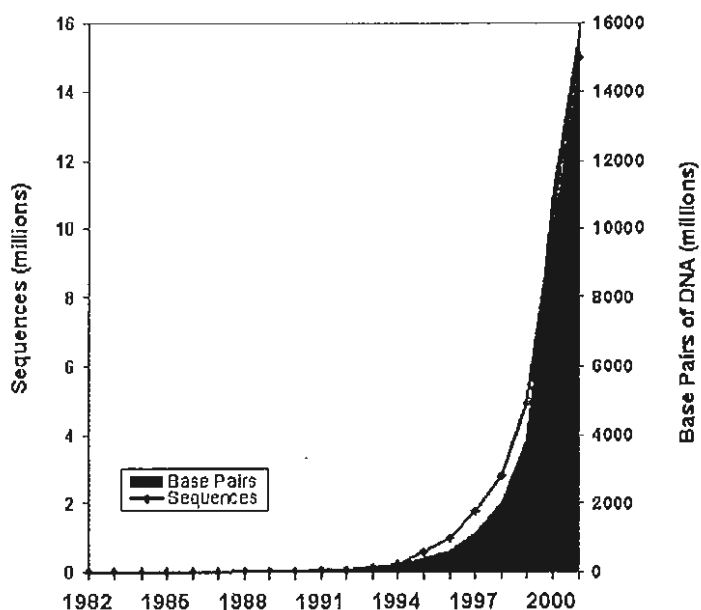
ชีวสารสนเทศคืออะไร?

ชีวสารสนเทศ (bioinformatics) เป็นศาสตร์ใหม่ที่กำลังเจริญก้าวหน้าไปอย่างรวดเร็วและมีผลกระทบต่อการวิจัยวิทยาศาสตร์ชีวภาพในทุกแขนง ชีวสารสนเทศเกิดจากการผสมผสานองค์ความรู้ทางชีววิทยาหลาย ๆ แขนงเข้ากับความก้าวหน้าทางด้านคอมพิวเตอร์ในการจัดเก็บรวบรวมข้อมูลทางชีววิทยาให้เป็นระบบในรูปแบบฐานข้อมูล (databases) และวิเคราะห์ข้อมูลเหล่านั้นด้วยเครื่องมือ (tools) หรือซอฟต์แวร์ต่างๆ ทั้งนี้เพื่อให้เกิดการเข้าใจในกระบวนการต่างๆ ในสิ่งมีชีวิต เข้าใจการทำงานของยีน ความสัมพันธ์ระหว่างยีน หน้าที่ของยีน ตลอดจนความบกพร่องของยีนที่มีผลต่อพยาธิสภาพหรือโรค อาจกล่าวได้ว่าเทคโนโลยีชีวสารสนเทศเกิดจากความก้าวหน้าของโครงการถอดรหัสพันธุกรรมมนุษย์ (Human Genome Project) ซึ่งตัวโครงการเองได้ถูกริเริ่มขึ้นเมื่อปี พ.ศ. 2528 แต่เริ่มต้นอย่างเป็นทางการในปี พ.ศ. 2533 (ค.ศ. 1990) โดยการนำของรัฐบาลสหรัฐอเมริกา โครงการถอดรหัสพันธุกรรมมนุษย์นี้มีเป้าหมายในการหาลำดับรหัสพันธุกรรมทั้งหมดของมนุษย์ ภายในระยะเวลา 15 ปี ใช้งบประมาณ 3 พันล้านเหรียญสหรัฐ และมีหน่วยงานหลายแห่งทั่วโลกเข้าร่วมโครงการนี้ แผนงานโครงการนี้ได้แบ่งงานการศึกษาลำดับรหัสพันธุกรรมออกตามโครโมโซม (22 คู่ + 2 ข้าง) โดยใช้วิธีสร้างแผนที่กายภาพ (physical map) ขึ้นก่อน จากนั้นจึงทำการศึกษาลำดับของรหัสพันธุกรรมโดยวิธี automated DNA sequencing ข้อมูลที่ได้จากโครงการนี้ได้ทยอยบรรจุในฐานข้อมูลสาธารณะที่สามารถเข้าถึงได้ทางอินเทอร์เน็ต และแล้วเมื่อเดือนกุมภาพันธ์ ปี พ.ศ. 2544 (ค.ศ. 2001) ก็ได้มีการประกาศว่าการศึกษาลำดับรหัสพันธุกรรมทั้งหมดของมนุษย์ได้สำเร็จเรียบร้อยแล้วทั้งนี้ได้มีการตีพิมพ์รายละเอียดในวารสารวิชาการชั้นนำทั้ง 2 ฉบับ คือ *Nature* และ *Science* พร้อมๆ กันในฉบับเดือนกุมภาพันธ์ ปี ค.ศ. 2001 และได้มอบข้อมูลลำดับพันธุกรรมให้เป็นสาธารณประโยชน์ในอินเทอร์เน็ตเพื่อให้ผู้สนใจสามารถ

download ไปใช้ต่อไปได้ ผลการศึกษาหาลำดับพันธุกรรมมนุษย์ทั้งสิ้นกว่า 3,000 ล้านเบส ในเบื้องต้นได้พบยีนเพียงประมาณ 39,000 ยีน และเชื่อกันว่ายังมียีนอีกจำนวนมากที่ยังไม่ทราบหน้าที่ ยังไม่มีผู้ใดศึกษา การศึกษาวิจัยหลังจากนี้เรียกว่าการวิจัยหลังจีโนม (post-genomics) และจะอาศัยฐานข้อมูลลำดับรหัสพันธุกรรมมนุษย์ที่กล่าวนี้เป็นพื้นฐาน ในปัจจุบันได้มีโครงการศึกษาลำดับรหัสพันธุกรรมของสิ่งมีชีวิตอื่นๆ อีกมากมาย ทั้งไวรัส แบคทีเรีย พืช และสัตว์ต่างๆ ที่มีส่วนเกี่ยวข้องกับมนุษย์เกิดขึ้น ผู้สนใจรายละเอียดเพิ่มเติมสามารถเข้าไปดูข้อมูลได้ที่ <http://www.ncbi.nlm.nih.gov/>

ปริมาณข้อมูลลำดับสารพันธุกรรมที่ได้เพิ่มจำนวนขึ้นอย่างมากมายอย่างรวดเร็ว ในระยะเวลาไม่กี่ปีที่ผ่านมา (รูปที่ 1) ทำให้นักวิทยาศาสตร์ต้องหาวิธีในการที่จะจัดการเก็บข้อมูลอย่างเป็นระบบ พัฒนาระบบข้อมูล ตลอดจนซอฟต์แวร์ที่ใช้เป็นเครื่องมือในการวิเคราะห์และแปลความหมายของลำดับดีเอ็นเอให้ออกมาในรูปของหน้าที่และการทำงานของยีน กลไกการทำงานในระดับเซลล์ จนถึงกลไกการทำงานในสิ่งมีชีวิต

Growth of GenBank



รูปที่ 1 ข้อมูลการเติบโตของ GenBank ระหว่างปี ค.ศ. 1982-2001 ตามรายงานของ National Center for Biotechnology Information (NCBI) ประเทศสหรัฐอเมริกา ณ เดือนสิงหาคม ปีค.ศ. 2002 มีจำนวนกว่า 16,000 ล้านเบสในฐานข้อมูล (แหล่งข้อมูลจาก <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>)

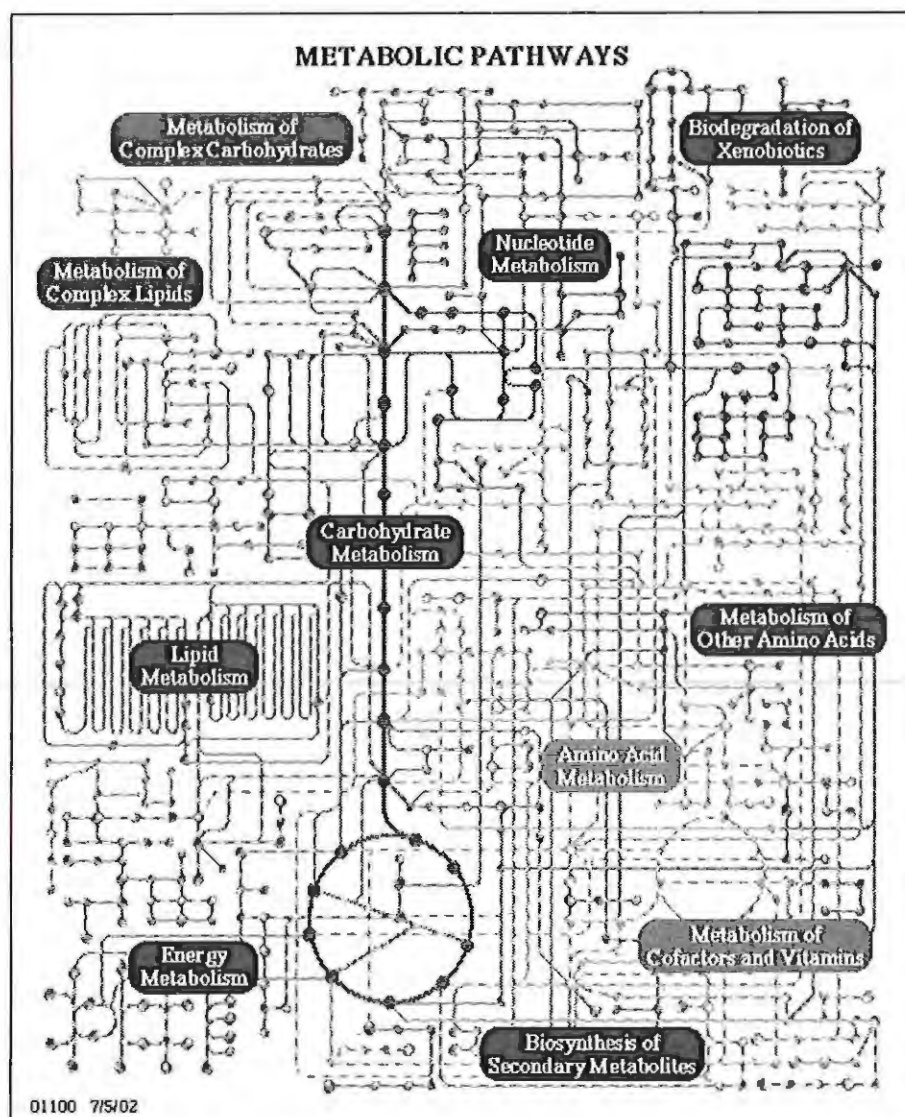
จะเห็นได้ว่างานหลักของชีวสารสนเทศคือการสืบค้นข้อมูลที่มีอยู่ในฐานข้อมูลชีววิทยา (biological databases) ต่างๆ ซึ่งบรรจุข้อมูลรหัสพันธุกรรมของมนุษย์ หรือรหัสพันธุกรรมของเชื้อโรคต่างๆ หรือพืชต่างๆ ข้อมูลการเรียงตัวของกรดนิวคลีอิก

(nucleotide sequence) การเรียงตัวของอะมิโนในสายโปรตีน (protein sequence) โครงสร้างของโปรตีน (protein structure) เป็นต้น ในการสืบค้นข้อมูลที่มีความหลากหลายเหล่านี้จำเป็นต้องอาศัยคอมพิวเตอร์ อาศัยวิธีการที่หลากหลาย ตลอดจนใช้ซอฟต์แวร์ที่เกี่ยวข้องมากมาย และที่สำคัญที่สุดคือ การที่ต้องมีคอมพิวเตอร์ที่เชื่อมต่ออย่างมีประสิทธิภาพกับเครือข่าย internet และ World Wide Web

คงจะหลีกเลี่ยงไม่ได้ว่าคอมพิวเตอร์ เครือข่าย internet และ World Wide Web (www) ได้กลายมาเป็นส่วนสำคัญในการวิเคราะห์ปัญหาและตอบคำถามวิทยาศาสตร์ชีวภาพในปัจจุบัน จะเห็นได้ว่าในระยะเวลาเกือบสองทศวรรษที่ผ่านมา รูปแบบของการวิจัยวิทยาศาสตร์ชีวภาพได้เปลี่ยนไปจากเดิม (paradigm shift) ซึ่งเป็นการศึกษายีนเดี่ยวๆ หรือโปรตีนเพียงตัวเดียว ไปเป็นความสามารถในการศึกษายีนหลายๆ ยีน หรือโปรตีนหลายๆ ชนิดพร้อมๆ กันทีเดียว เป็นการศึกษาทั้งจีโนม (whole genome analysis) รูปแบบของการวิจัยใหม่นี้ล้วนแต่เป็นผลจากความก้าวหน้าด้านเทคโนโลยีการวิจัยด้านยีนโนม เข้าใจว่ามีจุดเริ่มต้นจากโครงการถอดรหัสพันธุกรรมมนุษย์ซึ่งได้มีส่วนอย่างมากที่ก่อให้เกิดการปฏิวัติการวิจัยในรูปแบบใหม่ ไม่ว่าจะเป็นในด้านการพัฒนาวิธีการเพิ่มขีดความสามารถ (high-throughput) หรือการย่อระบบให้เล็กลง (miniaturization) เพื่อให้สามารถรองรับการเพิ่มจำนวนตัวอย่างที่จะทำการวิเคราะห์ต่อหน่วยเวลา หรือการพัฒนากระบวนการวิเคราะห์แบบอัตโนมัติ (automation) เพื่อให้สามารถทำงานวิเคราะห์ที่มีลักษณะซ้ำๆ กันให้มีประสิทธิภาพซึ่งจะเป็นการลดความผิดพลาดจากมนุษย์

จากยีนโนมไปสู่ยา

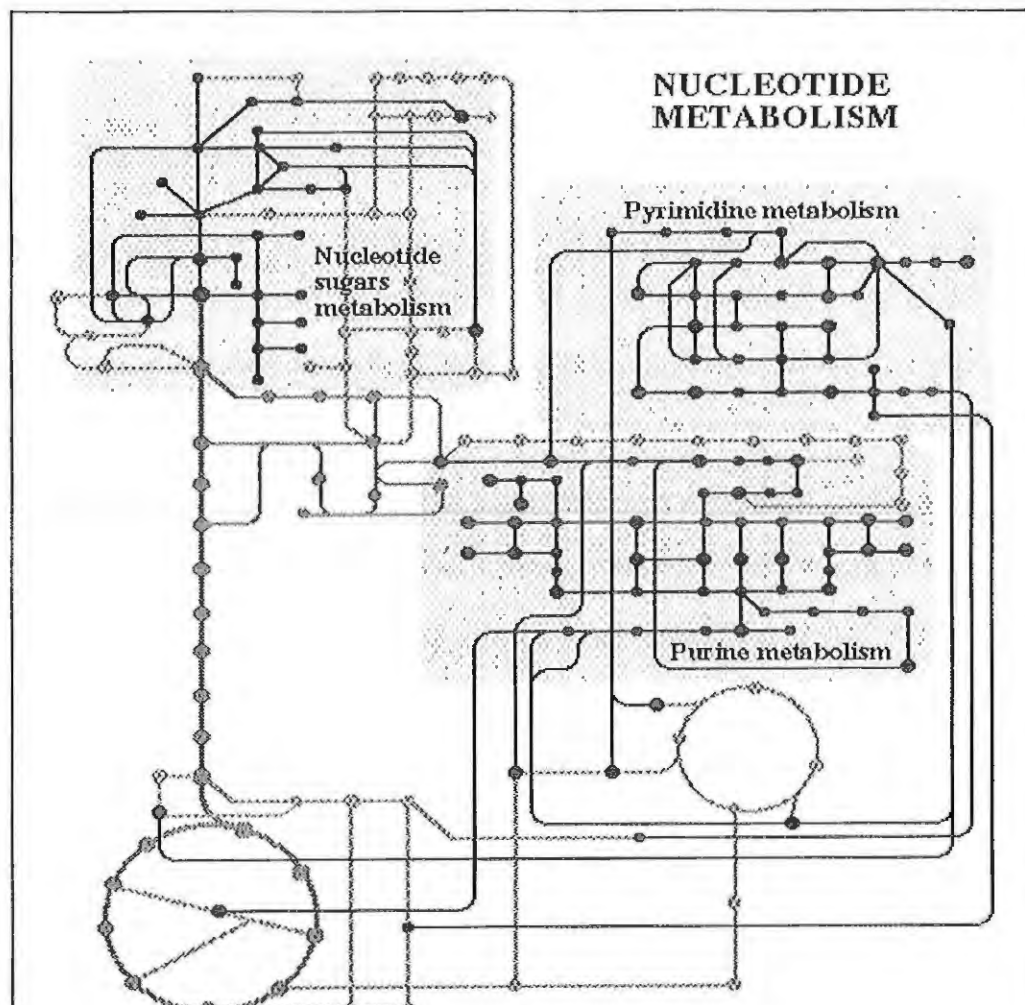
เป้าหมายสำคัญของการศึกษายีนคือการเข้าใจถึงหน้าที่ของยีน และกลไกการทำงานของยีนนั้นร่วมกับยีนที่เกี่ยวข้องกับกระบวนการอื่นๆ ที่เกิดขึ้นในเซลล์ แน่นอนเหลือเกินว่ากระบวนการเหล่านี้จะสลับซับซ้อนมากในสิ่งมีชีวิตชั้นสูงในลักษณะของ metabolic network ซึ่งมีลักษณะคล้ายๆ กับวงจรอิเล็กทรอนิกส์ (รูปที่ 2)



รูปที่ 2 แสดงวิถีเมตาบอลิซึม แบบวงจร metabolic network (ข้อมูลจาก <http://www.genome.ad.jp/kegg/metabolism.html>)

ลักษณะของ metabolic network ดังแสดงในรูปที่ 2 นี้ เป็นรูปแบบที่พัฒนาให้เห็นการเชื่อมโยงของกระบวนการเมตาบอลิซึมในรูปแบบของวงจรอิเล็กทรอนิกส์โดยกลุ่มนักวิจัยจากประเทศญี่ปุ่นซึ่งได้จัดทำ resources ด้าน bioinformatics ที่เรียกว่า KEGG (Kyoto Encyclopedia of Genes and Genomes) ในลักษณะวงจรอิเล็กทรอนิกส์เพื่อศึกษาความเชื่อมโยงของกระบวนการเมตาบอลิซึมและหน้าที่ของเอนไซม์ที่เกี่ยวข้อง เช่น หากสนใจ

nucleotide metabolism ก็สามารถ click ไปที่ box ที่เขียนว่า nucleotide metabolism ซึ่งจะสามารถได้ข้อมูลเฉพาะวงจรดังกล่าวดังแสดงในรูปที่ 3



รูปที่ 3 แสดงวิถีเมตาบอลิซึมของวงจร nucleotide metabolism

(ข้อมูลจาก <http://www.genome.ad.jp/kegg/pathway/map/map01140.html>)

หากต้องการรายละเอียดที่ลึกลงไปกว่านี้ก็สามารถ click ไปยัง metabolism ที่ต้องการ ซึ่งจะแสดงรายละเอียดของกระบวนการรวมทั้งรายละเอียดของเอนไซม์ต่างๆ ที่เกี่ยวข้องกับปฏิกิริยา เป็นต้น

แนวทางการประยุกต์ใช้ชีวสารสนเทศเป็นเครื่องมือในงานวิจัยวิทยาศาสตร์ชีวภาพที่กำลังเป็นที่สนใจ คือ การค้นหา protein targets และการค้นหา (สารยับยั้ง) ต่อ

targets เหล่านี้ วิธีการหนึ่งซึ่งได้รับการพัฒนาเพื่อค้นหา protein targets ใหม่ ๆ ในยุคหลังจีโนมคือการพัฒนาใช้เทคโนโลยี DNA chips เทคโนโลยีดังกล่าวสามารถใช้ติดตามการสร้างโปรตีนในระยะต่างๆ ของการเจริญเติบโตของเซลล์ที่ต้องการทำการศึกษา เทคโนโลยี DNA chip มีหลายลักษณะ เช่น ลักษณะที่ใช้ EST (Expressed Sequence Tags) เพื่อมาศึกษาข้อมูลการแสดงออกของยีน หรือลักษณะของ proteomics 2D-gel ในการศึกษาโปรตีนที่สร้างจากยีนในระยะใดระยะหนึ่งของการเจริญเติบโตของเซลล์ จะเห็นได้ว่าเทคโนโลยีเหล่านี้ล้วนแล้วแต่ต้องใช้ชีวสารสนเทศในการวิเคราะห์ เช่น sequence assembly การวิเคราะห์เชิงเปรียบเทียบของยีนสำหรับโปรตีนที่สนใจ การทำนายโครงสร้างของโปรตีน ตลอดจนการศึกษากฎสัมพันธ์ระหว่างโปรตีน ในด้านการค้นหาหรือสารยับยั้งต่อ targets นั้นได้มีความก้าวหน้าการสังเคราะห์ในลักษณะ combinatorial synthesis ผสมกับความก้าวหน้าในการศึกษาโครงสร้างโปรตีน ทำให้นักวิทยาศาสตร์สามารถทำนายโครงสร้างที่เหมาะสมของสารยับยั้งตามโครงสร้างบริเวณเร่งหรือบริเวณที่ใช้จับกับสารยับยั้ง (rational หรือ structure-based design)

บทบาทของชีวสารสนเทศในยุคหลังจีโนม

ความสำคัญของชีวสารสนเทศใช้ว่าจะอยู่ที่การพัฒนาระบบจัดเก็บ ฐานข้อมูล ตลอดจนวิเคราะห์ข้อมูลความเหมือนและความต่างระหว่างยีนของสิ่งมีชีวิตต่างๆ แต่เป้าหมายหลักที่สำคัญคือการใช้ชีวสารสนเทศในการสร้างความเข้าใจในความหมายของการเรียงตัวของลำดับเบสทั้ง 4 ชนิดในสายดีเอ็นเอ ความสามารถในการเข้าใจถึงหน้าที่ของยีน และกลไกการแสดงออกของยีน ตลอดจนหน้าที่ของผลิตภัณฑ์ของยีน ชีวสารสนเทศได้แทรกซึมเข้าไปมีบทบาทเกี่ยวข้องกับส่วนของวิธีการหลายส่วน อีกทั้งยังเกี่ยวข้องกับสาขาวิชาย่อย อีกหลายสาขาวิชา ไม่ว่าจะเป็นด้านการวิเคราะห์การเรียงลำดับของเบสในยีน ในเรื่อง base calling, physical mapping, fragment assembly การค้นหายีน (gene finding) ค้นหาและศึกษาความสำคัญของบริเวณที่เป็น non-coding regions ในจีโนม ซึ่งอาจเป็นบริเวณที่มีผลต่อการควบคุมการแสดงออกของยีน การเปรียบเทียบความเหมือนของลำดับกรดอะมิโนในสายโปรตีน ซึ่งจะนำไปสู่ความเข้าใจในหน้าที่ของโปรตีน และในทางกลับกันการศึกษาหน้าที่ของโปรตีนที่คล้ายคลึงกันจะเป็นทางนำไปสู่ความเข้าใจหน้าที่ของโปรตีนที่สนใจได้ นอกจากนี้ ยังรวมไปถึงการสร้างโครงสร้าง

จำลอง และศึกษาโครงสร้างสามมิติ การศึกษาโครงสร้างของ DNA และ RNA ตลอดจนโครงสร้างของชีวโมเลกุลอื่นๆ ซึ่งได้แก่ lipid และ carbohydrate และปฏิสัมพันธ์ระหว่างชีวโมเลกุลเหล่านี้

กล่าวโดยสรุป เทคโนโลยีชีวสารสนเทศเป็นเทคโนโลยียุคใหม่ที่เชื่อมโยงกับเทคโนโลยีชีวภาพ การพัฒนาซอฟต์แวร์สำหรับการวิเคราะห์ข้อมูลการเรียงลำดับรหัสพันธุกรรม หรือข้อมูลการเรียงลำดับกรดอะมิโน ประกอบกับการพัฒนา internet ให้สามารถเข้าถึงข้อมูลสาธารณะที่มีอยู่ในฐานข้อมูลปัจจุบันเป็นสิ่งสำคัญที่จะนำไปสู่การประยุกต์ใช้เทคโนโลยีสารสนเทศอย่างมีประสิทธิภาพและกว้างขวางมากยิ่งขึ้น จากความสำคัญของชีวสารสนเทศที่มีแนวโน้มที่จะมีบทบาทอย่างมากต่องานวิจัยทางด้านวิทยาศาสตร์ชีวภาพในยุคหลังจีโนม คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดลจึงได้ทำการจัดตั้งหน่วยวิจัยชีวสารสนเทศและจีโนมประยุกต์ (Center for Bioinformatics and Applied Genomics, CBAG) ซึ่งเป็นหน่วยงานในลักษณะเสริมสร้างศักยภาพ (Capability Building Research Unit) ของคณะวิทยาศาสตร์ มหาวิทยาลัยมหิดลขึ้น โดยในระยะเริ่มต้นจะทำหน้าที่เป็นหน่วยงานเสริมสร้างขีดความสามารถด้านชีวสารสนเทศ และเป็นศูนย์ฝึกอบรมระดับภูมิภาคเอเชียขององค์การอนามัยโลกในด้านชีวสารสนเทศ เมื่อมีบุคลากรที่เข้มแข็งขึ้นแล้วจะได้ประยุกต์ใช้เทคโนโลยีดังกล่าวในงานวิจัยวิทยาศาสตร์ชีวภาพในสาขาต่างๆ ต่อไป

เอกสารอ่านเพิ่มเติม

1. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, 2nd edition (Eds. Baxevanis, A. D. and Francis Ouellette, B. F.), 2001, John Wiley & Sons, Inc.
2. Bioinformatics: From Genomes to Drugs. Vol. 1: Basic Technologies (Ed. Thomas Lengauer), 2002, Wiley-VCH, Germany
3. Developing Bioinformatics Computer Skills. (Eds. Gibas, C. and Jambeck, P.) 2001, O' Reilly & Associates, Inc., California
4. Bioinformatics: Sequence and Genome Analysis (Ed. Mount, D. W.) Cold Spring Harbor Laboratory Press, New York