



การใช้เทคโนโลยี Next Generation Sequencing ในการวินิจฉัย รักษาและวิจัยโรคมะเร็ง

ณัฐณี จินาวัฒน์^{1,3*} และ ภูมิ สุขธิติพัฒน์^{2,3**}

¹สำนักงานวิจัยวิชาการและนวัตกรรม คณะแพทยศาสตร์ ร.พ. รามาธิบดี;

²ภาควิชาชีวเคมี คณะแพทยศาสตร์ศิริราชพยาบาล; ³ศูนย์ปฏิบัติการ

กลางด้านชีววิทยาศาสตร์คอมพิวเตอร์เพื่อการวิจัยแบบบูรณาการ

มหาวิทยาลัยมหิดล กรุงเทพฯ 10700

Email: *jnatinini@hotmail.com; **bhoom.suk@mahidol.ac.th



ภูมิหลัง

โรคมะเร็งเป็นสาเหตุการเสียชีวิตในอันดับต้นๆ ของประชากรไทยติดต่อกันมานานหลายสิบปี จากสถิติล่าสุดในปีพุทธศักราช 2554 ของกระทรวงสาธารณสุข มีรายงานว่ามียุติเสียชีวิต 61,082 ราย เฉลี่ยชั่วโมงละเกือบ 7 ราย เป็นชาย 35,437 ราย และหญิง 25,645 ราย และคาดการณ์ว่าจะมีแนวโน้มการเพิ่ม จำนวนของผู้ป่วยรายใหม่มากขึ้นเรื่อยๆ อย่างต่อเนื่อง สืบเนื่องมาจากการใช้ชีวิต สภาพแวดล้อมภายนอก ที่เป็นปัจจัยเสี่ยง เช่น อาหาร สารเคมี รวมทั้งปัจจัยภายใน เช่น พันธุกรรม ดังนั้น โรคมะเร็งจึงถือเป็นปัญหาสาธารณสุขที่สำคัญของประเทศ ซึ่งนอกจากจะส่งผลเสียต่อเศรษฐกิจและสังคมในระดับมหภาคแล้ว ยังขัดขวาง ขีดจำกัดความสามารถในการพัฒนาประเทศ เนื่องจากในแต่ละปีต้องสูญเสียทรัพยากรบุคคลอันทรงคุณค่าไปเป็นจำนวนมากอีกด้วย

ทั้งนี้การรักษาโรคมะเร็งในปัจจุบันได้เปลี่ยนแปลงไปจากอดีตมาก โดยเน้นไปที่การรักษาเฉพาะบุคคล (Personalized cancer medicine) เช่น การใช้ยารักษามะเร็งกลุ่มใหม่ที่มีความเฉพาะเจาะจงต่อเซลล์มะเร็ง (targeted therapy) และใช้วิธีการตรวจระดับโมเลกุลเพื่อช่วยในการวินิจฉัย ติดตามโรค และเลือกยาที่ใช้ในการรักษา ซึ่งส่งผลให้สามารถเลือกวิธีการรักษาที่เหมาะสมกับผู้ป่วยแต่ละคน ตามสาเหตุและกระบวนการก่อโรค (molecular pathway) ได้ นำมาสู่ความสำเร็จในการตอบสนองต่อการรักษาที่ดีขึ้น ทำให้ผู้ป่วยมีคุณภาพชีวิตที่ดีขึ้น ลดการสูญเสียทั้งทรัพยากรมนุษย์ และงบประมาณของประเทศชาติได้อย่างมีประสิทธิภาพ

ปัจจุบันการศึกษาหาลำดับเบสทั้งหมดของสารพันธุกรรมดีเอ็นเอ (DNA sequence) ถือได้ว่าเป็นหนึ่งในองค์ประกอบที่สำคัญที่สุดในการศึกษาทางด้านจีโนมิกส์ (Genomics) ซึ่งก็คือการศึกษาจีโนม (genome) ทั้งหมดของสิ่งมีชีวิตแต่ละชนิดนั่นเอง ตัวอย่างของการศึกษา DNA sequence ได้แก่ การศึกษาการกลายพันธุ์ของยีนที่ทำให้เกิดโรคหรือมีผลต่อการรักษาโรคในมนุษย์ หนึ่งในวิธีที่เป็นที่นิยมที่สุดคือ การหาลำดับเบสดีเอ็นเอแบบ Sanger (Sanger sequencing) ซึ่งถูกค้นพบในปี 1977 ซึ่งใช้วิธีการหาลำดับเบสที่เรียกว่า dideoxy หรือ chain termination อย่างไรก็ตาม Sanger sequencing ถือเป็นวิธีที่มีประสิทธิภาพ (output) ไม่สูง ถึงแม้ว่าจะมีความถูกต้อง (accuracy) สูงก็ตาม เนื่องจากการเรียงลำดับเบสได้ครั้งละ 1 ตัวอย่างส่งตรงเท่านั้น ทำให้วิธีนี้เป็นวิธีที่สิ้นเปลืองทรัพยากรและมีค่าใช้จ่ายสูง ดังนั้นในช่วง 10 ปีที่ผ่านมา จึงมีการพัฒนาและนำเอาวิธีการจัดลำดับเบสแบบใหม่มาใช้ ซึ่งวิธีใหม่นี้มีความสามารถในการหาลำดับเบสในปริมาณที่มากกว่า (high-throughput) ทำได้หลายตัวอย่างในเวลาเดียวกัน (multiplexing) และ มีความไวในการตรวจหาการกลายพันธุ์ของ DNA สูงกว่า (high sensitivity) ซึ่งวิธีนี้เรียกรวมๆว่า Next Generation Sequencing (NGS)

NGS เป็นเทคโนโลยีการหาลำดับเบสที่พัฒนาขึ้นใหม่ มีความสามารถในการเรียงลำดับเบสจำนวนมากในการทำงานของเครื่องเพียงครั้งเดียว โดยสามารถใช้หาลำดับเบสของจีโนมมนุษย์ (Human genome) ซึ่งมีความยาว 3 พันล้านเบส การหาลำดับเบสเอ็กโซมมนุษย์ (Exome Sequencing) ซึ่งมีความยาว 50 ล้านเบส หรือ การหาลำดับเบสของบริเวณเฉพาะของจีโนมหลายๆบริเวณหรือยีนหลายยีนที่สนใจ (Panel or Targeted sequencing) จากหลายตัวอย่างที่มีความยาวรวมกันเป็นล้านเบสได้ โดยในปัจจุบันเทคโนโลยี

NGS ที่เป็นที่นิยมและใช้กันมากที่สุดได้แก่ sequencing by synthesis (SBS) technology ของบริษัท Illumina และ semiconductor sequencing technology ของบริษัท Life Technologies NGS ได้ถูกนำมาใช้กันอย่างแพร่หลายในโรงพยาบาลหลายแห่งและห้องปฏิบัติการชั้นนำของโลก เพื่อช่วยในการวินิจฉัยและการวิจัยโรคต่างๆ รวมถึงโรคมะเร็ง ประโยชน์หลักของการใช้ NGS เมื่อเทียบกับวิธีเก่าคือ ค่าใช้จ่ายที่ลดลงในระยะยาวและข้อมูลที่เพิ่มขึ้นต่อตัวอย่าง ซึ่งค่าใช้จ่ายเฉลี่ยของการหาลำดับเบสโดยใช้ NGS ต่อตัวอย่างจะยิ่งต่ำลงอีกเมื่อมีจำนวนตัวอย่างที่มากขึ้นต่อการทำงานของเครื่องหนึ่งครั้ง แต่ปริมาณข้อมูลที่มากขึ้นก็อาจทำให้เกิดปัญหาในการวิเคราะห์ข้อมูลและแปลผลได้

โดยในต่างประเทศ ได้มีการนำ NGS มาใช้ในงานทางคลินิก ดังต่อไปนี้

- Noninvasive prenatal testing (NIPT) การวินิจฉัยก่อนคลอดโดยใช้ cell-free fetal DNA ที่ปนอยู่ในเลือดของมารดา
- Comprehensive cancer panels การหาลำดับเบสของยีนที่เกี่ยวข้องกับมะเร็งหลายๆ ยีนที่มีผลต่อการรักษา (clinically actionable) ในการตรวจครั้งเดียว เพื่อช่วยในการวินิจฉัย เลือกให้ยา ติดตามผลการรักษา หรือทำนายการดำเนินโรค
- Assessment of diseases of unknown etiology การหาที่มาของโรคซึ่งไม่ทราบสาเหตุ เช่น โรคทางพันธุกรรมหายาก (rare disease) ต่างๆ
- Infectious disease identification and surveillance การตรวจหาเชื้อโรคที่เป็นสาเหตุของโรคติดเชื้อใหม่ๆ

NGS กับ การศึกษาทางด้านจีโนมิกส์ของโรคมะเร็ง

เป็นที่ทราบกันว่า โรคมะเร็งจัดเป็นโรคทางพันธุกรรมอย่างหนึ่งซึ่งเกิดจากการกลายพันธุ์ของสารพันธุกรรม DNA ในเซลล์มะเร็ง การศึกษาการกลายพันธุ์ของ DNA ทั้ง 2 ชนิด (somatic และ germline mutation) ในโรคมะเร็ง จึงมีความสำคัญทั้งในแง่การวิจัยเพื่อหาวิธีการรักษาใหม่ๆ หรือเพื่อช่วยให้อุบัติการณ์โรคได้เร็วขึ้น รวมถึงเพื่อพัฒนาการติดตาม และทำนายผลการรักษาโรค นอกจากนี้การรักษาโรคมะเร็งในปัจจุบันเน้นการรักษาเชิงบุคคล (Personalized cancer medicine) โดยให้ยารักษาโรคมะเร็งที่ต่างกันตามลักษณะการกลายพันธุ์ของ DNA ในผู้ป่วยมะเร็งแต่ละคน ตัวอย่างที่เห็นได้ชัดคือการตรวจหา somatic mutation ของ ยีน KRAS และ BRAF จากชิ้นเนื้อมะเร็งลำไส้ใหญ่ เพื่อช่วยในการเลือกให้ยา Cetuximab ซึ่งก็คือ monoclonal antibody ต่อ Epidermal Growth Factor Receptor (EGFR) แก่ผู้ป่วย โดยจะเลือกให้ยานี้กับผู้ป่วยที่ไม่มีการกลายพันธุ์ของยีน 2 ตัวนี้เท่านั้น (wild type) เนื่องจากถ้าผู้ป่วยมีการกลายพันธุ์ของยีนที่กล่าวมาตัวใดตัวหนึ่ง ก็จะมีผลการตอบสนองต่อการรักษาด้วย Cetuximab ที่ไม่ดี ดังนั้นจึงจะเห็นได้ว่าการนำเทคโนโลยี NGS มาใช้ในการตรวจรักษาโรคมะเร็งจะเป็นประโยชน์ต่อผู้ป่วยอย่างมาก โดยเฉพาะเมื่อมียาที่จำเพาะเจาะจงต่อการกลายพันธุ์ (targeted therapy) ของยีนที่เกี่ยวข้องกับโรคมะเร็งตัวใหม่ๆ ออกมาอย่างต่อเนื่องในปัจจุบัน

ปัจจุบัน บริษัทผู้ผลิตมีการพัฒนาชุดตรวจหาการกลายพันธุ์ของ DNA สำหรับโรคมะเร็งโดยใช้เทคโนโลยี NGS ทั้งในระดับ germline และ somatic ออกมาอย่างแพร่หลาย โดยบางชุดตรวจสามารถให้กับชิ้นเนื้อมะเร็งที่เก็บในรูปแบบ Formalin-Fixed Paraffin-Embedded (FFPE) ซึ่งเป็นวิธีการเก็บรักษาชิ้นเนื้อเพื่อการวินิจฉัยที่พยาธิแพทย์ทั่วโลกใช้กันแพร่หลายที่สุดได้ด้วย เป็นที่ทราบกันว่า DNA ที่ได้จาก FFPE tissue มีคุณภาพต่ำและมีความยาวสั้น ซึ่งถือเป็นปัญหาสำคัญต่อการนำไปใช้กับ NGS นอกจากนี้ จีโนมของมะเร็ง (cancer genome) โดยทั่วไป มีความซับซ้อนสูงมาก ประกอบไปด้วยการเปลี่ยนแปลงทางพันธุกรรม (genetic alteration) หลายชนิด ทั้งการกลายพันธุ์ระดับยีน (point mutation, small indel, loss of heterozygosity) การกลายพันธุ์ระดับโครโมโซมทั้งโครงสร้าง (structural variation) และจำนวน

(aneuploidy) ที่สำคัญขึ้นเนื้อมะเร็งประกอบไปด้วยทั้งเซลล์ปกติและเซลล์มะเร็ง (mosaicism) และบริเวณต่างๆ ของก้อนมะเร็งก็อาจมีลักษณะการกลายพันธุ์ที่ไม่เหมือนกัน (tumor heterogeneity) ซึ่งปัจจัยต่างๆ ที่กล่าวมานี้ทำให้การวิเคราะห์ข้อมูลลำดับเบสของมะเร็งทำได้ค่อนข้างยาก และจำเป็นต้องใช้เทคนิคทางชีวสารสนเทศและระบบคอมพิวเตอร์ประสิทธิภาพสูงในการวิเคราะห์ข้อมูล

ข้อมูลที่ได้จากเทคโนโลยี NGS

เทคโนโลยี NGS ในปัจจุบัน สามารถตรวจลำดับเบสได้อย่างรวดเร็ว โดยใช้เทคนิคที่เรียกว่า short-gun sequencing เริ่มต้นจากการเตรียมแม่แบบของการ sequencing (template library preparation) โดยแบ่ง genomic DNA เป็นชิ้นเล็กๆ (DNA fragment) จำนวนนับพันล้านชิ้น โดยความยาวของลำดับเบสของชิ้นส่วน DNA เหล่านี้จะมีตั้งแต่สายสั้นๆ ขนาด 25-1,000 เบสขึ้นกับชนิดของเทคโนโลยีที่ใช้ ในขั้นตอนถัดไปชิ้นส่วนของ DNA เหล่านี้จะถูกเชื่อมเข้ากับอะแดปเตอร์ (adapter) โดยอะแดปเตอร์เหล่านี้จะจับกับ universal primer ซึ่งจะถูกใช้เพื่อเพิ่มจำนวนชิ้นส่วน DNA และใช้ตรวจลำดับเบสโดยวิธี SBS โดยชิ้นส่วนนับล้านนี้จะถูกตรวจหาลำดับเบสพร้อมๆ กัน หลังจากนั้นข้อมูลลำดับเบสของชิ้นส่วน DNA แต่ละชิ้นจะถูกนำมาประกอบกันใหม่ (assemble) เพื่อให้ทราบลำดับเบสรวมทั้งหมดของ sequence ที่ต้องการศึกษา

ในการวิเคราะห์ข้อมูลจีโนมมะเร็งนั้น เพื่อให้สามารถแยกจีโนมปกติ ออกจากจีโนมมะเร็งได้ จำเป็นจะต้องมีข้อมูลจีโนมปกติจากผู้ป่วยรายเดียวกัน (matched normal genome) มาเปรียบเทียบกับข้อมูลจีโนมของเนื้อเยื่อมะเร็ง โดยข้อมูลของจีโนมปกติสำหรับการศึกษามะเร็งทางเลือดส่วนใหญ่มาจากเซลล์ผิวหนัง (skin biopsy) หรือ สำหรับจากเม็ดเลือดขาวในเลือดในกรณีศึกษามะเร็งชนิดอื่นๆ เพื่อลดโอกาสการปนเปื้อนของจีโนมปกติในการตรวจลำดับเบสของเซลล์มะเร็ง [Mwenifumbo and Marra, 2013] ในกรณีที่ไม่มีข้อมูลของจีโนมปกติจากผู้ป่วยรายเดียวกันให้เปรียบเทียบ อาจใช้ข้อมูลที่มีอยู่ในฐานข้อมูลความหลากหลายทางพันธุกรรมในคนปกติเช่น dbSNP จาก US National Center for Biotechnology (www.ncbi.nlm.nih.gov/SNP) หรือ ข้อมูลจาก โครงการ 1000 Genome Project (www.1000genomes.org)

เพื่อให้การตรวจการกลายพันธุ์มีความถูกต้องมากขึ้น เทคโนโลยี NGS ที่ใช้ในการศึกษาเนื้อเยื่อมะเร็งจำเป็นต้องมีข้อมูลของการกลายพันธุ์แต่ละตำแหน่ง (coverage) ซ้ำๆ กันจำนวนมาก ทั้งนี้ลำดับความซ้ำซ้อนของข้อมูลที่ยอมรับกันว่าน่าเชื่อถือสำหรับข้อมูล NGS ที่ควรใช้สำหรับการวินิจฉัยทางคลินิก และการรักษาเชิงบุคคล อย่างต่ำควรอยู่ที่ 50x [Ajay, 2011] ซึ่งสูงกว่าตัวเลขเดิมที่ 30x ทั้งนี้เนื่องจากการกลายพันธุ์ที่เกิดขึ้นในชิ้นเนื้อมะเร็งอาจเป็นสัดส่วนที่ต่างๆ กันตั้งแต่ 0% ถึง 100% ดังนั้นเพื่อให้สามารถตรวจพบการกลายพันธุ์ในเนื้อเยื่อมะเร็งในปริมาณต่ำๆ เช่น 1-2 % จึงอาจจำเป็นต้องใช้ coverage ในระดับสูงมากกว่า 400x เพื่อให้มีความมั่นใจว่าการกลายพันธุ์ที่ตรวจพบเกิดขึ้นจริง [Mwenifumbo and Marra, 2013] การใช้เทคนิคเช่น microdissection, cell sorting หรือสร้าง low-passage cell line เพื่อเลือกเฉพาะเซลล์มะเร็งให้มีสัดส่วนมากกว่า 80% ของเซลล์ทั้งหมดมาสกัดสารพันธุกรรม ก็จะช่วยลดโอกาสที่จะตรวจไม่พบการกลายพันธุ์ในเนื้อเยื่อมะเร็งที่ศึกษา (false negative) โดยถ้าเลือกใช้การตรวจแบบ paired-end แทน single read ถึงแม้จะมีราคาสูงกว่า ก็จะสามารถช่วยให้ตรวจพบการเปลี่ยนแปลงขนาดใหญ่ของ DNA (structural variation) หรือ indels ได้ถูกต้องสูงขึ้น

ปัจจัยที่สำคัญที่สุด ก่อนที่จะสรุปว่าการเปลี่ยนแปลงทางพันธุกรรมที่พบสัมพันธ์กับโรคมะเร็งที่ทำการศึกษาคือ การตรวจเช็คโดยใช้เทคนิคอื่นเช่น PCR และ Sanger sequencing เพื่อตรวจสอบว่าการเปลี่ยนแปลงทางพันธุกรรมนั้นเกิดขึ้นจริง และ ไม่ได้เป็นความผิดพลาดในการตรวจจาก NGS เช่น จาก sequencing error หรือ sequencing bias หรือ alignment ผิดตำแหน่ง

การแปลผลข้อมูลที่ได้จากเทคโนโลยี NGS

ผลจากการที่ NGS แพร่หลายมากขึ้น ความเร็วในการตรวจลดลง และราคาตรวจต่อผู้ป่วยหนึ่งรายเริ่มถูกลง ทำให้แพทย์เริ่มหันมาสนใจส่งตรวจโรคทางพันธุกรรมหลายๆอย่างโดยใช้ NGS เพิ่มมากขึ้น ไม่นับรวมถึงการตรวจทางพันธุกรรมที่มุ่งเป้ากลุ่มผู้ป่วยโรคโดยตรงโดยอีกหลายบริษัท ปัจจุบันเมื่อเราตรวจพบการเปลี่ยนแปลงรหัสทางพันธุกรรมเช่น Single Nucleotide Polymorphisms (SNPs) ข้อมูลส่วนใหญ่ที่ใช้สำหรับการแปลความหมายของรหัสพันธุกรรมเหล่านี้ เป็นข้อมูลที่ได้มาศึกษาหาความสัมพันธ์ระหว่าง SNPs กับโรคจาก genome-wide association study (GWAS) ในประชากรต่างชาติ เช่น ยุโรป หรือ อเมริกา หากเป็นการกลายพันธุ์ที่มีผลกระทบต่อการทำงานของโปรตีน หรือ เอนไซม์โดยตรง ถ้าพบการกลายพันธุ์นั้นในประชากรในประเทศไทยด้วย การแสดงออกของการกลายพันธุ์นั้นน่าจะมีผลไม่ต่างกัน ทั้งนี้การนำเอาผลการศึกษาจากต่างชาติมาใช้กับประชากรไทยโดยตรง ควรต้องใช้ด้วยความระมัดระวัง ถ้าไม่มีการศึกษาเพื่อตรวจสอบผลของการเปลี่ยนแปลงทางพันธุกรรมเหล่านี้ในประชากรไทยก่อน

ปัญหาหลักที่ตามมาคือจำนวนข้อมูลที่ได้มามีจำนวนมาก การวิเคราะห์ข้อมูลที่ได้จาก โครงการวิจัยที่ใช้ NGS จึงจำเป็นต้องใช้ระบบคอมพิวเตอร์ที่มีประสิทธิภาพสูง (high-performance computing resource) ถึงแม้ว่าในปัจจุบันจะมีคอมพิวเตอร์ระบบคลัสเตอร์ขนาดใหญ่ (computing cluster) เช่น โครงการ Extreme Science and Engineering Discovery Environment (www.xsede.org) ที่สนับสนุนโดย National Science Foundation สหรัฐอเมริกา ซึ่งสนับสนุนให้มีการแชร์ข้อมูลระหว่างกลุ่มนักวิทยาศาสตร์ และ สนับสนุนให้การวิเคราะห์ข้อมูล NGS เป็นอย่างโปร่งใส และสามารถทำซ้ำได้ (reproducible) แต่การใช้ระบบ cloud computing สาธารณะ ก็ยังเป็นปัญหาที่ทางจริยธรรมที่ยังเป็นที่ถกเถียงกันอยู่ว่าทำอย่างไรจึงจะให้เกิดความสมดุลระหว่างการรักษาความเป็นส่วนตัว (privacy) และ เพื่อเป็นประโยชน์ต่อความก้าวหน้าทางวิทยาศาสตร์ [Greenbaum *et al*, 2011]

นอกเหนือจากทรัพยากรทางคอมพิวเตอร์ที่จำเป็นต้องใช้ในการวิเคราะห์ข้อมูล NGS แล้ว ทรัพยากรบุคคลที่มีความรู้และความเข้าใจในการวิเคราะห์ข้อมูลที่ได้จาก NGS ก็เป็นอีกปัญหาหนึ่งที่สำคัญสำหรับประเทศไทย ทั้งในด้าน bioinformatics, genetic epidemiology, statistical genetics และสาขาที่เกี่ยวข้อง นอกจากนี้แพทย์ในสาขาที่เกี่ยวข้องก็ควรมีความรู้ความเข้าใจเกี่ยวกับการส่งตรวจและแปลผลข้อมูลที่ได้จาก NGS

เอกสารอ้างอิง

- Mwenifumbo JC, Marra MA. Cancer genome-sequencing study design. *Nat Rev Genet.* 2013 May; 14(5):321-32.
- Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. Accurate and comprehensive sequencing of personal genomes. *Genome Res.* 2011 Sep; 21(9):1498-505.
- Greenbaum D, Sboner A, Mu XJ, Gerstein M. Genomics and privacy: implications of the new reality of closed data for the field. *PLoS Comput Biol.* 2011 Dec; 7(12): e1002278.