การประยุกต์ใช้เทคโนโลยีการหาลำดับสารพันธุกรรมรุ่นใหม่ในงานนิติพันธุศาสตร์

กรเกียรติ วงศ์ไพศาลสิน

หน่วยนิติเซโรวิทยาและดีเอ็นเอ ภาควิชานิติเวชศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร 10332

Email: kornkiat.v@chula.ac.th

บทนำ (Introduction)

เทคโนโลยีในการตรวจดีเอ็นเอทางนิติพันธุศาสตร์นั้นมีประโยชน์ในการตรวจเพื่อพิสูจน์เอกลักษณ์บุคคล และการตรวจเพื่อพิสูจน์ยืนยันความสัมพันธ์ระหว่างบิดา มารดา และบุตร อีกทั้งสามารถใช้เปรียบเทียบดีเอ็นเอเพื่อค้นหากรณีบุคคลสูญหายจากฐานข้อมูลดีเอ็นเอ โดยในปัจจุบัน การตรวจทางห้องปฏิบัติการด้วยเทคนิค PCR (polymerase chain reaction) และการวิเคราะห์เพื่อคัดแยกความหลากหลายของตำแหน่งตรวจด้วยวิธี Capillary Electrophoresis นับเป็นมาตรฐานที่ใช้อย่างกว้างขวางในงานนิติวิทยาศาสตร์ โดยตำแหน่งที่มีความหลากหลายชนิด short tandem repeat (STR) ได้ถูกศึกษาและนำมาใช้ประโยชน์ในการทำฐานข้อมูลความถี่อัลลีล (allele frequency) ในประชากรกลุ่มต่างๆทั่วโลก ซึ่งข้อมูลดังกล่าวจะถูกนำมาใช้ในการคำนวณความค่าทางสถิติประชากร และความน่าจะเป็นเพื่อการพิสูจน์เอกลักษณ์บุคคล และพิสูจน์ความสัมพันธ์ เป็นต้น มีการศึกษาเพื่อค้นหาตำแหน่ง STR ที่มีประโยชน์ในการพิสูจน์บุคคล ซึ่งทางตำรวจสากลของประเทศสหรัฐอเมริกาได้กำหนดให้ใช้ 13 ตำแหน่งหลัก (13 core loci) และนับเป็นตำแหน่งมาตรฐานที่ประกอบในชุดการตรวจสำเร็จรูป อีกทั้งการสร้างฐานข้อมูลดีเอ็นเอระดับประเทศเพื่อใช้ในกระบวนการยุติธรรม ได้แสดงให้เห็นถึงความสำเร็จในการเชื่อมโยงตัวบุคคลต้องสงสัยในคดีต่างๆ จึงทำให้มีการศึกษาจำนวนมากเพื่อค้นหาตำแหน่งตรวจที่มีความจำเพาะสูงในการพิสูจน์บุคคลทางนิติเวช และเพื่อวัตถุประสงค์ต่างๆ เช่น การศึกษาเพื่อค้นหาตำแหน่งเพิ่มเติมจาก 13 core STR ที่เหมาะสมในแต่ละประเทศหรือกลุ่มประชากร, การศึกษาตำแหน่งที่มีความสามารถในการสืบค้นถึงบรรพบุรุษ (ancestry marker) เช่น การตรวจหาตำแหน่ง single nucleotide polymorphism (SNP) บนโครโมโซมเพศชาย (Y chromosome) และการศึกษาลำดับสารพันธุกรรมในสายไมโตคอนเดรียล (Mitochondrial DNA) เป็นต้น ทำให้เทคโนโลยีการหาลำดับสารพันธุกรรมรุ่นใหม่ (Next generation sequencing, NGS) นั้นถูกนำมาใช้ประยุกต์เพื่อวัตถุประสงค์ต่างๆดังที่กล่าวมาข้างต้น

ในปี ค.ศ.1977 Frederick Sanger ได้นำเสนอเทคนิคในการตรวจลำดับสารพันธุกรรม โดยการใช้วิธี electrophoresis ซึ่งนับเป็นมาตรฐานแรกในการตรวจ (standard Sanger technique) เพื่อตรวจหาลำดับสารพันธุกรรมทั้งหมดของมนุษย์ภายใต้โครงการศึกษาร่วมที่มีชื่อว่า Human Genome Project แต่ด้วยข้อจำกัดของเทคนิคการตรวจ ที่ใช้เวลาตรวจทางห้องปฏิบัติการที่นาน และค่าใช้จ่ายในการตรวจค่อนข้างสูง จึงทำให้ในเวลาต่อมาได้มีการพัฒนาการใช้สารเรืองแสง (fluorescent) เพื่อติดตามสายดีเอ็นเอที่ศึกษา อีกทั้งการคัดแยกขนาดสายดีเอ็นเอโดยการใช้ Capillary Electrophoresis ถือได้ว่าเป็นยุคแรกของการศึกษาลำดับสารพันธุกรรม (first-generation sequencing) แต่อย่างไรก็ตาม การศึกษานั้นยังมีค่าใช่จ่ายในการศึกษาที่สูงและมีข้อจำกัดของขนาดสายดีเอ็นเอที่ศึกษา

ส่วนเทคโนโลยีการตรวจลำดับสารพันธุกรรมรุ่นใหม่ (NGS) นั้นทำให้สามารถตรวจลำดับสารพันธุกรรมที่ยาวมากขึ้น และครอบคลุมตำแหน่งดีเอ็นเอที่สนใจได้ เช่น การตรวจ whole-genome sequencing (WGS) เป็นต้น ทั้งนี้ผู้ศึกษาสามารถออกแบบตำแหน่งตรวจที่เหมาะสมกับวัตถุประสงค์ในงานนิติพันธุศาสตร์ เช่น สารมารถออกแบบตำแหน่งตรวจทั้งแบบ autosomal STR, autosomal SNP, Y or X

chromosome STR, Y SNPs หรือ mitochondrial sequence ในชุดการตรวจเดียวได้ (ตารางที่ 1) อีกทั้ง ด้วยเทคโนโลยีและความสามารถในการออกแบบเลือกใช้ตำแหน่งตรวจที่มากขึ้น ในตัวอย่างที่เสื่อมสภาพหรือกรณี ตัวอย่างจากสถานที่เกิดเหตุที่มีจำนวนปริมาณตัวอย่างที่จำกัด ก็จะสามารถเพิ่มโอกาสในการได้ข้อมูลทางพันธุกรรมที่เป็นประโยชน์ในการพิสูจน์บุคคลอีกด้วย

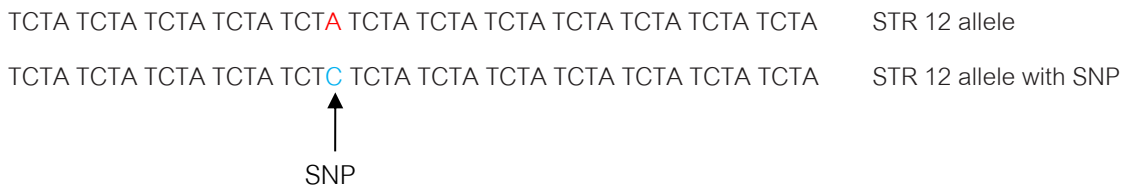
ตารางที่ 1 ตำแหน่งการตรวจทางนิติพันธุศาสตร์ที่สามารถนำมาประยุกต์ใช้ในการตรวจลำดับสารพันธุกรรมรุ่นใหม่ (next-generation sequencing)

13 CODIS Autosomal STRs	Identity SNPs
Worldwide Autosomal STRs	Ancestry SNPs
CODIS Y STRs	Phenotypic SNPs
Worldwide Y STRs	Mitochondrial D-loop
Y or X Haplotype	Complete Mitochondrial sequence
Reference: Targeted Next-Generation Sequencing for Forensic Genomics by Illumina	

Targeted Forensic NGS

ดีเอ็นเอทางนิติพันธุศาสตร์ ถูกศึกษาพัฒนาและนำมาใช้ในกระบวนการทางนิติวิทยาศาสตร์ ในการตรวจพิสูจน์เอกลักษณ์บุคคล การตรวจหาความสัมพันธ์ทางเครือญาติ การตรวจหาบรรพบุรุษร่วม เป็นต้น โดยรูปแบบของความหลากหลายของดีเอ็นเอที่นำมาใช้ในงานนิติพันธุศาสตร์นั้น ได้คัดเลือกจากลำดับดีเอ็นเอที่เรียงตัวกันอย่างมีรูปแบบจำเพาะเป็นหน่วยซ้ำ (repeat unit) และเรียงตัวกันอย่างต่อเนื่อง (tandem repeat) ด้วยเทคโนโลยีการแยกขนาดสายดีเอ็นเอชนิด STR บน capillary electrophoresis ซึ่งจะแปลผลตามขนาดความยาวของ STR ออกมาเป็นจำนวนซ้ำของ repeat unit เช่น 12,12 (homozygous ของ repeat unit ซ้ำกันอย่างต่อเนื่อง 12 ครั้ง) เป็นต้น แต่เนื่องจาก CE จะแปลผลของ STR ตามขนาดความยาวทำให้ไม่สามารถแยกความหลากหลายชนิด SNP ที่อาจเกิดขึ้นภายในลำดับดีเอ็นเอของ STR ได้ (รูปที่ 1) จึงเป็นข้อจำกัดหนึ่งของ CE ในขณะที่ NGS สามารถให้ข้อมูลในรายละเอียดของลำดับสารพันธุกรรม (Loman 2012)

ตำแหน่งของ Forensic STR core marker นั้นมีขนาดยาวประมาณ 250-300 เบส/ตำแหน่ง ผู้ศึกษาสามารถออกแบบชุดตำแหน่งตรวจที่ประกอบด้วยจำนวนตำแหน่งที่ต้องการ และจำนวนตัวอย่างที่จะใช้ในการตรวจและคำนวณค่าความครอบคลุมที่เหมาะสมกับงานที่ตรวจได้ (ตารางที่ 2) อีกทั้งจากการศึกษาทางพันธุศาสตร์ประชากร (Population Genetics) ได้ค้นหาตำแหน่ง SNPs ที่สามารถนำมาใช้ในการคัดแยกกลุ่มประชากร ซึ่งมีประโยชน์ทางนิติเวชในการบอกถึงข้อมูลกลุ่มประชากรที่สงสัย อีกทั้งสามารถคัดเลือกตำแหน่ง SNPs ที่สามารถนำมาใช้ในการพิสูจน์บุคคล (Individualized markers) ซึ่งตำแหน่งต่างๆเหล่านี้สามารถนำมาประกอบในชุดการตรวจด้วยเทคโนโลยีการลำดับสารพันธุกรรมรุ่นใหม่ได้ โดยที่ค่าความถูกต้องและแม่นยำของผลการตรวจนั้นมีความสำคัญในงานทางนิติพันธุศาสตร์ ซึ่งจะรายงานค่าความแม่นยำจากการคำนวณ Likelihood ของการอ่านค่าลำดับเบส โดยแปลออกมาเป็น Q score ซึ่งค่าที่สูงนั้นหมายความถึงคุณภาพการอ่านลำดับเบสที่มีความแม่นยำสูง เช่น Q30 หมายถึง การอ่านลำดับเบสที่ถูกต้องมีค่าเท่ากับ 99.9% เป็นต้น



รูปที่ 1 ตัวอย่างแสดงขนาดของสายดีเอ็นเอบนตำแหน่ง STR ที่มี 12 อัลลีล จะสามารถตรวจพบความหลากหลายชนิด SNP เกิดขึ้นภายในตำแหน่งของ repeat units.

ตารางที่ 2 แสดงการประมาณค่าความครอบคลุม (Sequencing coverage) บนเครื่อง MiSeq platform ซึ่งสามารถอ่านค่าลำดับสารพันธุกรรมได้ 14 ล้านเบส ต่อการใช้งานหนึ่งรอบ

จำนวนตำแหน่งที่ตรวจ	จำนวนตัวอย่างศึกษา			
	1	48	96	384
25	560,000	11,700	5,800	1,400
50	280,000	5,850	2,900	700
150	93,333	1,950	960	230
250	56,000	1,170	580	140
300	46,667	975	480	115

อ้างอิงจาก: Targeted Next-Generation Sequencing for Forensic Genomics by Illumina

หลักการของการศึกษาด้วยเทคโนโลยีการตรวจลำดับสารพันธุกรรมรุ่นใหม่ (Jason 2012)

เทคโนโลยีการหาลำดับสารพันธุกรรมรุ่นใหม่นั้น ประกอบด้วยขั้นตอนหลัก คือ (1) Template generation (2) Sequencing reaction and detection (3) Data analysis

Template generation

โดยทั่วไป ในการเริ่มต้นของกระบวนการจะต้องใช้ดีเอ็นเอสายคู่ (Double-stranded DNA) ซึ่งอาจจะมาจากแหล่งของตัวอย่างได้หลากหลายชนิด นำมาสร้างเป็น sequencing library ซึ่งประกอบด้วยขั้นตอนย่อย เช่น การตัดสายดีเอ็นเอออกเป็นท่อนๆ (DNA fragmentation), การคัดเลือกขนาดสายดีเอ็นเอ (size selection) และ การติด adapter (adapter ligation) โดยที่ปลายสายของดีเอ็นเอที่ถูกตัดเป็นท่อนๆนั้นจะถูกติดด้วย adapter ซึ่งทำหน้าที่เปรียบเสมือนเป็น primer สำหรับการเริ่มต้นปฏิกิริยาเพิ่มปริมาณดีเอ็นเอ และด้วยเทคโนโลยีของเครื่อง NGS ซึ่งออกแบบให้นักวิจัยสามารถเลือกการทำงานทั้งชนิด (1) Single-molecule templates (การอ่านวิเคราะห์ลำดับสารพันธุกรรมบนสาย DNA library โดยที่ไม่ต้องผ่านการเพิ่มจำนวนสายดีเอ็นเอก่อนการวิเคราะห์) ทำให้สามารถออกแบบการศึกษาได้แม้ในตัวอย่างที่มีปริมาณน้อยมากๆ จนถึงระดับ single molecule หรือ (2) Clonally amplified templates (จะทำการเพิ่มจำนวนจากสายดีเอ็นเอต้นแบบแล้วตามด้วยการอ่านวิเคราะห์ลำดับสารพันธุกรรม)

Sequencing reaction and detection

Sequencing by synthesis (SBS) คือ การสังเคราะห์ หรือสร้างสายดีเอ็นเอ และวิเคราะห์ไปพร้อมๆ กัน โดยในกระบวนการของการทำปฏิกิริยานั้นในแต่ละ Platforms อาจใช้สารเคมีที่แตกต่างกัน แต่ยังคงใช้เอนไซม์ DNA polymerase หรือ DNA ligase เป็นหลักในปฏิกิริยา ซึ่งหลักการของ SBS จะสามารถทำการสังเคราะห์ดีเอ็นเอ จากการเติม deoxyribonucleotide triphosphate (dNTP) และระบบจะทำการอ่านค่าการเรืองแสงฟลูออเรสเซนต์ทันที ทำให้เราสามารถออกแบบตำแหน่งที่ต้องการศึกษาในหลายๆตำแหน่งไปพร้อมๆ กันในรอบปฏิกิริยาเดียว

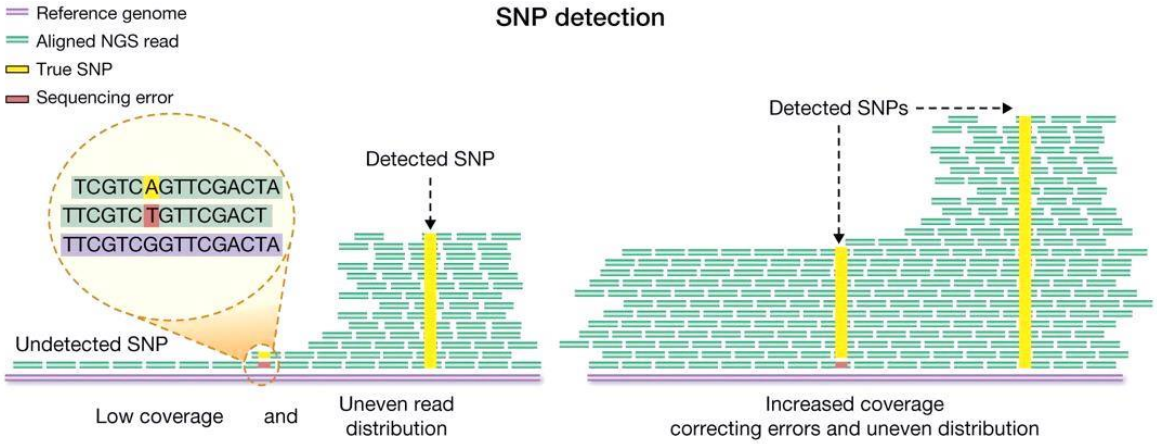
Data analysis

ข้อมูลทางพันธุกรรมที่ได้จาก NGS นั้นจะเป็นข้อมูลทางพันธุกรรมที่มีขนาดข้อมูลใหญ่ตั้งแต่ megabases (millions) จนถึง gigabases (billions) และต้องอาศัยการจัดการข้อมูลอย่างมีระบบตั้งแต่การตรวจคัดกรองข้อมูล (Data filtering and base calling) หลังจากนั้นข้อมูลลำดับสารพันธุกรรมต้องนำมาเทียบกับลำดับจีโนมอ้างอิง (Reference genome) โดยที่มีการศึกษาด้าน bioinformatic จำนวนมากที่พัฒนา software หรือชุดคำสั่งในการจัดการข้อมูลดังกล่าว อย่างไรก็ดี เทคนิคการหาลำดับสารพันธุกรรมรุ่นใหม่ (NGS) ยังมีข้อผิดพลาดที่อาจเกิดขึ้นในระบบได้เสมอ (Thompson 2011) เช่น ในช่วงของการเพิ่มจำนวนสายดีเอ็นเอ Clonally amplified DNA templates อาจทำให้เกิดการสร้างตำแหน่งดีเอ็นเอที่มี mutation และทำให้อ่านผลผิดและเข้าใจว่าเป็น sequence variation เป็นต้น

ในด้านปริมาณ และคุณภาพของผลลัพธ์ที่ได้จาก NGS นั้น ขึ้นอยู่กับการออกแบบชนิดของการศึกษา ซึ่งทั้งหมดจะเป็นตัวกำหนดความครอบคลุม และความแม่นยำของผลการตรวจ ตัวอย่างเช่น (1) การอ่านค่าลำดับค่าดีเอ็นเอผิดพลาด (calling error rates) เป็นปัญหาในเชิงคุณภาพ ซึ่งค่าความผิดพลาดนั้นมีค่าที่แตกต่างกัน ขึ้นอยู่กับเครื่องมือที่ใช้ในการตรวจ ซึ่งโดยทั่วไปจะมีการกำหนดค่าความเชื่อมั่น (Confidence scores) สำหรับทุกๆการอ่านค่าในแต่ละบุคคล (2) จำนวนข้อมูลลำดับสารพันธุกรรมสามารถกำหนดเป็นค่าความครอบคลุม (Coverage) และมีความหมายว่า ในแต่ละดีเอ็นเอนั้นถูกวิเคราะห์ และอ่านเป็นจำนวนกี่ครั้ง ซึ่งเป็นการตรวจสอบความถูกต้องของผลการตรวจนั่นเอง ในการค้นหาตำแหน่งบนลำดับสารพันธุกรรมที่มีความหลากหลายชนิด insertions, deletions และ translocations ด้วยเครื่องมือในปัจจุบัน ได้เสนอให้กำหนดค่าความครอบคลุมที่ 20x ถึง30x ส่วนการตั้งค่าความครอบคลุมที่สูงมากๆ จำเป็นสำหรับการอ่านค่า SNP เพื่อให้ได้ค่าที่แม่นยำ โดยกำหนดให้มีค่าความครอบคลุมที่ 30x ถึง 100x เป็นต้น (Nielsen 2011) (รูปที่ 2) ทั้งนี้ การเพิ่มค่าความครอบคลุมในการศึกษาจะส่งผลให้ค่าใช้จ่ายในการศึกษานั้นสูงขึ้น

บทสรุป (Conclusion)

ด้วยเทคโนโลยีของการหาลำดับสารพันธุกรรมรุ่นใหม่ (Next Generation Sequencing) ได้นำเสนอความสามารถที่มีประโยชน์ต่อการศึกษาทางนิติพันธุศาสตร์ และยังสามารถนำข้อมูล Forensic sequence Markers ที่ได้จากการศึกษาด้วยเทคโนโลยี NGS เข้ารวมกับข้อมูล Forensic STR เดิมบนเทคโนโลยี Capillary electrophoresis ซึ่งเป็นฐานข้อมูลประชากรขนาดใหญ่ที่ใช้ในกรณีศึกษาต่างๆทางนิติวิทยาศาสตร์ จึงทำให้ในอนาคตอันใกล้การศึกษาเพื่อค้นหา Forensic STR marker หรือ SNPs ตำแหน่งใหม่ๆ หรือตำแหน่งดีเอ็นเอที่เหมาะสมต่อประชากรเฉพาะกลุ่ม อีกทั้ง การศึกษาเพื่อค้นหา Phenotypic Marker เพื่อใช้ในการพิสูจน์บุคคล เหล่านี้นับเป็นข้อมูลทางพันธุศาสตร์ที่เป็นประโยชน์ต่อกระบวนการยุติธรรมและนิติวิทยาศาสตร์



รูปที่ 2 ตัวอย่างแสดงการศึกษาเพื่อค้นหาตำแหน่ง Single Nucleotide Polymorphism (SNP), (ขวา) ด้วยเทคนิคของ NGS ผู้ศึกษาสามารถออกแบบการศึกษาที่ใช้ตำแหน่งลำดับสารพันธุกรรมที่ครอบคลุม และมีการอ่านลำดับสารพันธุกรรมนั้นซ้ำๆ (high coverage) จะเพิ่มความถูกต้อง และแม่นยำในการค้นหาตำแหน่ง SNP, (ซ้าย) ในขณะที่หากการครอบคลุมนั้นไม่มากเพียงพอ (low coverage) อาจทำให้เกิดข้อผิดพลาดในการวินิจฉัยตำแหน่ง SNP (Jason 2012)

เอกสารอ้างอิง (References)

Jason M, Rizzo JB, Michael JB. Key principles and clinical applications of “Next-Generation” DNA sequencing. Cancer Prev Res 2012; 5(7): 890-900.

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, *et al*. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012;30: 434-439.

Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 2011; 12:443-51.

Thompson JF, Milos PM. The properties and applications of single molecule DNA sequencing. Genome Biol 2011; 12:217.