<u>ข้อจำกัดของฐานข้อมูลสารพันธุกรรม ในการหาความสัมพันธ์ทางสถิติในฐานข้อมูลสารพันธุกรรม</u>

ผู้แต่ง : ERIN E. MURPHY

พิมพ์เมื่อ : 8/ส.ค./2558

บทความในหนังสือ : Inside the Cell: The Dark side of Forensic DNA

โดย : นายวิศรุต ถนอมทรัพย์ ตำแหน่ง นักนิติวิทยาศาสตร์ชำนาญการ กลุ่มตรวจพิสูจน์อัตลักษณ์บุคคล กองพัฒนาระบบการติดตามคนหายและการพิสูจน์ศพนิรนาม

ในปี พ.ศ.2544 ผู้ตรวจวิเคราะห์ข้อมูลสารพันธุกรรม ณ หน่วยสารพันธุกรรมของห้องปฏิบัติการด้านอาชญากรรมของรัฐแอริโซน่า ประเทศสหรัฐอเมริกา ได้มีการแจ้งเตือนสิ่งที่น่าสนใจคือ พบการแชร์ข้อมูลสารพันธุกรรมของคนผิวขาวและคนผิวดำที่ไม่มีความสัมพันธ์ทางสายโลหิตกัน โดยพบการซ้ำกันของตำแหน่งที่ตรวจหาข้อมูลสารพันธุกรรม 9 ตำแหน่ง ใน 13 ตำแหน่ง และในการซ้ำกันนั้นพบการซ้ำบน marker ทั้ง 2 marker ซึ่งถือว่าพบได้ยากมาก (ในการตรวจหาข้อมูลสารพันธุกรรม จะมีการใช้น้ำยา หรือ marker ในการตรวจและแปลผลเป็นตัวเลข โดยจะตรวจบน 13 ตำแหน่งบนโครโมโซม ดังนั้นการแปลผลจะแปลเป็นตัวเลขทั้งหมด 13 คู่ ในกรณีนี้พบการซ้ำกันทั้งหมด 9 คู่)

ตามมาตรฐานการคำนวณ การมีข้อมูลสารพันธุกรรมที่เหมือนกันของบุคคลสองคนที่ไม่มีความเกี่ยวข้องกันทางสายโลหิต โดยการเปรียบเทียบข้อมูลสารพันธุกรรมนั้นกระทำในฐานข้อมูลสารพันธุกรรมที่มีจำนวนมาก เรียกความน่าจะเป็นที่พบนี้ว่า “***Random match probability***” การค้นพบการเหมือนกันของข้อมูลสารพันธุกรรมของบุคคลที่ไม่มีความเกี่ยวข้องกันนี้ สามารถเจอได้ในอัตราส่วน 1 คน ต่อ 750 ล้านคน ในประชากรคนขาวที่ไม่มีเชื้อสายละติน และเจอได้ในอัตราส่วน 1 คน ต่อ 561 พันล้านคนในประชาการคนอเมริกันผิวดำ และการค้นพบ “Random match probability” จะเพิ่มขึ้นเป็นสองเท่าเมื่อฐานข้อมูลนั้นมีจำนวนน้อยกว่า 100,000 ข้อมูล และเป็นข้อมูลที่ถูกจัดเก็บจากกลุ่มประชากรหลากหลายเชื้อชาติ

ผู้ตรวจวิเคราะห์สารพันธุกรรมได้รายงานสิ่งที่ค้นพบและส่งผลดังกล่าวไปยังการประชุมใหญ่นิติพันธุศาสตร์นานาชาติ ข้อสังเกตของผู้ตรวจวิเคราะห์รายนี้ได้รับความสนใจจากทนายของรัฐซานฟรานซิสโกรายหนึ่ง ที่กำลังศึกษาในระดับปริญญาโท สาขาพันธุศาสตร์ โดยทนายผู้นี้เคยร่วมอยู่ในเหตุการแก้ต่างของ นาย John Puckett หรือที่รู้จักกันในนาม “California man” ชายที่เคยเป็นผู้ต้องหาในคดีข่มขืนและฆ่าเมื่อหลาย 10 ปีก่อน โดยในปี พ.ศ. 2515 ตำรวจได้มีการเก็บหลักฐานในคดีพยาบาลถูกกระทำผิดทางเพศและโดนมีดทำร้ายร่างกาย คดีนี้ค้างอยู่ระหว่างการสอบสวน จน 30 ปีให้หลังได้มีการรื้อฟื้นคดีและพนักงานสอบสวนได้ส่งหลักฐาน คือ คราบอสุจิที่ถูกพบบนร่างของเหยื่อไปตรวจพิสูจน์ หลังจากการตรวจพิสูจน์พบว่า สารพันธุกรรมมีความเสื่อมสภาพและพบความสัมพันธ์แบบบางส่วนหรือ Partial match กับฐานข้อมูลสารพันธุกรรมของรัฐแอริโซน่า โดยพบความสัมพันธ์กับนาย John Puckett ชายวัย 70 ปีที่นั่งบนรถเข็น หลังจากที่มีรายงานผลความสัมพันธ์นี้เกิดขึ้น นาย John Puckett ถูกจับกุมโดยเจ้าหน้าที่ตำรวจในข้อหาเป็นฆาตกร

ทนายของ นาย John Puckkett ได้ทำการร้องของข้อมูลการค้นพบความสัมพันธ์นี้จากห้องปฏิบัติการของรัฐแอริโซน่า แต่ถูกปฏิเสธจากหัวหน้าห้องปฏิบัติการ หลังจากนั้นศาลจึงมีคำสั่งให้ห้องปฏิบัติการเปิดเผยข้อมูลตามการร้องขอของทนาย หลังจากมีคำสั่งศาลผู้ตรวจวิเคราะห์แจ้งว่าพบความสัมพันธ์แบบ 9 ตำแหน่งใน 13 ตำแหน่งบนโครโมโซมในคนอีก 90 คนในฐานข้อมูลของรัฐแอริโซน่า เมื่อห้องปฏิบัติการไม่สามารถอธิบายถึงการพบความสัมพันธ์ของข้อมูลสารพันธุกรรมที่เหมือนกันในอัตราส่วน 1 ใน 1 ล้านล้านได้ ทางศาลจึงมีคำสั่งให้ห้องปฏิบัติการตรวจสอบกับฐานข้อมูลผู้ต้องหาทั้งหมดของรัฐแอริโซน่าใหม่อีกครั้ง และรายงานผลกลับมายังศาล

ท้ายที่สุดแล้วผลจากห้องปฏิบัติการแสดงว่า พบความสัมพันธ์ของข้อมูลในฐานข้อมูลสารพันธุกรรมลักษณะเดียวกันค่อนข้างมาก โดยฐานข้อมูลสารพันธุกรรมผู้ต้องหาของรัฐแอริโซน่า มีจำนวนข้อมูลสารพันธุกรรมทั้งหมด 65,493 ข้อมูล เป็นข้อมูลที่จัดเก็บจากเพศชายและหญิง ที่ตรวจวิเคราะห์บนโครโมโซมจำนวน 13 ตำแหน่ง จากฐานข้อมูลพบว่ามี 122 ข้อมูลสารพันธุกรรม ที่มีข้อมูลสารพันธุกรรมเหมือนกัน 9 ตำแหน่ง บางข้อมูลมีข้อมูลสารพันธุกรรมเหมือนกัน 10 11 หรือ 12 ตำแหน่ง ดังนั้นแล้วก็สมเหตุสมผลที่จะระบุตัวผู้กระทำผิดได้เช่นเดียวกัน

จากการพบความสัมพันธ์ของข้อมูลสารพันธุกรรม ทำให้เป็นที่สนใจของเหล่าทนายความในเมืองต่างๆ และพากันตั้งคำถามขึ้นว่า ในฐานข้อมูลสารพันธุกรรมจำนวน 65,000 ข้อมูล พบความสัมพันธ์ถึง 122 ความสัมพันธ์ หากในฐานข้อมูลสารพันธุกรรมของประเทศที่มีถึง 11 ล้านข้อมูล จะค้นพบความสัมพันธ์รูปแบบนี้อีกมากน้อยเพียงใด แต่แทนที่จะกระตุ้นเตือนให้ทาง FBI ตระหนัก และรวบรวมข้อมูลเพื่อใช้ประกอบการสอบสวน ทาง FBI กลับตำหนิทางห้องปฏิบัติการของรัฐแอริโซนา และกล่าวว่าเป็นการระเมิดข้อตกลงกับทาง FBI และยังขู่จะตัดการเข้าถึงข้อมูลของห้องปฏิบัติการอิสระต่างๆ ที่จะนำข้อมูลไปศึกษาต่อ

จากการค้นพบของห้องปฏิบัติการของรัฐแอริโซน่า ส่งผลให้ต้องมีการปรับตัวหรือทำความเข้าใจกันระหว่างโลกทางวิทยาศาสตร์และโลกทางกฎหมาย

ในขณะที่ห้องปฏิบัติการของรัฐแอริโซน่าค้นพบนั้น ฐานข้อมูลสารพันธุกรรมของรัฐและของชาติกำลังขยายตัว ประชาชนส่วนใหญ่คิดว่าผลการตรวจสารพันธุกรรมเป็นเครื่องมือหลักที่ใช้พิสูจน์ตัวคนร้ายในการกระทำผิด แต่จะทำอย่างไรให้เกิดความน่าเชื่อถือมากที่สุด จึงมีแนวคิดเรื่องฐานข้อมูลสารพันธุกรรมขนาดใหญ่ ที่ใช้ระบบคอมพิวเตอร์มาดำเนินการค้นหา เหมือนกับการดำเนินการค้นหาลายพิมพ์นิ้วมือในฐานข้อมูลซึ่งระบบฐานข้อมูลลายพิมพ์นิ้วมือนี้สามารถดำเนินการได้สมบูรณ์ในปี พ.ศ.2542

ทาง FBI ได้มีการจัดตั้งระบบฐานข้อมูลสารพันธุกรรมขึ้นมา โดยเรียกระบบนี้ว่า “Combined DNA Index System” หรือ “CODIS” เอง เพื่อช่วยให้การดำเนินการของคดี ที่เรียกว่า “cold hit” สามารถได้รับการแจ้งเตือนเมื่อพบความสัมพันธกับข้อมูลใหม่ที่ถูกบันทึกในฐานข้อมูลสารพันธุกรรม

บางคดีที่เป็น cold hit case สามารถเปลี่ยนสภาพเป็น hot case ได้ในการสอบสวน เมื่อมีวัตถุพยานที่ไม่ใช่วัตถุพยานด้านสารพันธุกรรมถูกค้นพบ ส่วนคดีของ นาย Puckett ที่ยังคงเป็น cold case อยู่ เนื่องจากไม่พบลายพิมพ์นิ้วมือใดๆในที่เกิดเหตุ จึงไม่สามารถนำมาตรวจเปรียบเทียบกับฐานข้อมูลลายพิมพ์นิ้วมือได้

จากเหตุการณ์เหมือนพนักงานสอบสวนมุ่งใช้ผลจากการตรวจพิสูจน์สารพันธุกรรมเพียงอย่างเดียว แต่ที่จริงแล้วนาย Puckett เองก็มีข้อมูลที่เข้ากันได้กับประจักษ์พยานที่ให้ปากคำ คือข้อมูลด้าน อายุ เพศ

เชื้อชาติ การปรากฏตัวในท้องที่ที่เกิดเหตุและ นาย Puckett เคยมีประวัติการกระทำผิดทางเพศกับหญิงสาว 3 ราย ในช่วงเวลาและท้องที่ใกล้เคียงกัน

คดีของนาย Puckett แสดงให้เห็นความสำคัญของการค้นพบของห้องปฏิบัติการของรัฐแอริโซน่าเมื่อปี พ.ศ.2544 และยังแสดงให้เห็นความสำคัญของผลจากการอภิปรายของนักคณิตศาสตร์ นักกฎหมายและนักนิติวิทยาศาสตร์ การค้นพบความสัมพันธ์ของสารพันธุกรรมที่ไม่น่าจะเป็นไปได้นี้ สามารถสังเกตเห็นได้จากผู้เชี่ยวชาญทางคณิตศาสตร์และนิติวิทยาศาสตร์ แต่จะยากต่อการสังเกตเห็นได้โดย ตำรวจ พนักงานสอบสวน และผู้ตรวจวิเคราะห์ในห้องปฏิบัติการ จากการค้นพบดังกล่าวยังแสดงให้เห็นว่าการค้นพบความสัมพันธ์ของข้อมูลสารพันธุกรรมจำนวน 9 ตำแหน่งเป็นสิ่งที่สามารถค้นพบได้ในฐานข้อมูลขนาดใหญ่

ในส่วนของคดีนาย Puckett แสดงให้เห็นว่า ก่อนการไต่สวนของลูกขุนพนักงานสอบสวนได้มีการเสนอต่อลูกขุนในเรื่องของ “Random match probability” ถึงผลการคำนวณที่แสดงค่า 1 ใน 1.1 ล้าน ส่วนทนายของนาย Puckett ก็มีการร้องขอต่อศาลให้อนุญาตให้มีการให้การต่อศาลถึงวิธีการคำนาณทางสถิติแนวทางอื่นๆ ที่ไม่ใช่การคำนวณรูปแบบ “Random match probability”

การคำนวณแนวทางอื่นที่จะใช้ในการแก้ต่างคือวิธี “Database match probability (DMP)” แนวทางการแก้ต่างต้องการแสดงให้เห็นถึงการคำนวณที่แตกต่างระหว่างกลุ่มประชากร 2 กลุ่ม คือ กลุ่มที่แสดงความสัมพันธ์จากการซุ่ม และ กลุ่มที่แสดงความสัมพันธ์จากกลุ่มประชากรที่กำหนดอย่างแน่ชัด ดังเช่นในคดีของนาย Puckett หากมีการระบุกลุ่มที่ชัดเจนคำถามที่ตั้งขึ้นมาใหม่ คือ จะมีจำนวนประชากรเพศชายที่อาศัยอยู่ในเมืองที่เกิดเหตุ มีอายุที่เข้ากันได้กี่ราย ที่จะแสดงความสัมพันธ์กับหลักฐานที่พบในที่เกิดเหตุ จากการคำนวณโดยสถิติที่ชื่อว่า “n*p” ผลที่ได้คือ มีอย่างน้อยสองรายที่จะพบความสัมพันธ์กับวัตถุพยานในที่เกิดเหตุ

แต่ละวิธีทางสถิติที่ใช้ในการคำนาณนั้น จะมีกระบวนการแปลผลที่แตกต่างกันออกไป การเลือกใช้ก็ต้องเลือกใช้วิธีให้เหมาะสมกับการพิจารณาในกระบวนการยุติธรรม เคยมีการโต้เถียงกันของทนายความว่าควรเสนอให้ศาลปฏิเสธเรื่อง “Database match probability” เสนอให้มีการค้นหาข้อมูลสนับสนุนการทดสอบความถี่ในการเกิด และนำเสนอความขัดแย้งของการคำนาณทางสถิติ

ในขณะที่ฐานข้อมูลสารพันธุกรรมของสหรัฐอเมริกากำลังขยายตัว ประเด็นความไม่สอดคล้องทางสถิตินี้ก็กลายเป็นประเด็นสำคัญในการเปรียบเทียบฐานข้อมูลสารพันธุกรรม จนในปี พ.ศ.2557 สถาบันเครือข่ายนิติวิทยาศาสตร์แห่งทวีปยุโรป (European Network of Forensic Science Institutes) มีการรายงานว่า การพบความสัมพันธ์ที่ผิดปกตินี้จะพบมากขึ้นเมื่อปริมาณข้อมูลในฐานข้อมูลสารพันธุกรรมเพิ่มมากขึ้น และเป็นการตรวจเปรียบเทียบกับข้อมูลสารพันธุกรรมที่เป็นแบบไม่สมบูรณ์และแบบผสม ดังนั้นทางผู้จัดการฐานข้อมูลควรมีการบันทึกปริมาณการพบความสัมพันธ์ที่ไม่คาดคิด รวมถึงบันทึกจำนวนฐานข้อมูลและครั้งที่พบความสัมพันธ์ ณ ขณะนั้นด้วย

ณ ปัจจุบันนี้การมีการยอมรับจากผู้ใช้กฎหมายแล้วว่า การพบความสัมพันธ์ของข้อมูลสารพันธุกรรมในคดี cold case ที่ผิดพลาดนี้เป็นความสะเพร่าในการดำเนินคดี มีคดีที่เกิดขึ้นในลักษณะนี้ในประเทศอังกฤษที่จับผู้ร้ายโดยใช้การค้นหาจากฐานข้อมูลสารพันธุกรรม โดยสุดท้ายแล้วผู้ที่ถูกจับกุมนั้นพ้นจากความผิดเนื่องจากผู้ที่ถูกจับกุมมีอาการของโรคพาร์กินสัน ไม่สามารถที่จะกระทำผิดในคดีลักทรัพย์ที่ถูกกล่าวหาได้

ข้อสรุปคดีของนาย Puckett คือ เค้าถูกพิพากษาว่ากระทำความผิด เนื่องจากคณะลูกขุนรับฟังเพียงเหตุผลจากพนักงานสอบสวน ที่แจ้งว่า จะมีบุคคล อีก 1 คน ใน 1.1 ล้านคน ที่จะมีข้อมูลสารพันธุกรรมเดียวกัน ลูกขุนไม่ได้รับฟังเหตุผลเรื่อง Database match probability ไม่ได้รับฟังเหตุผลการค้นพบจากห้องปฏิบัติการของรัซแอริโซน่า มิได้รับฟังว่ามีชายอีก 40 คนในรัฐแคลิฟอร์เนียที่พบความสัมพันธ์กับวัตถุพยานที่ตรวจเทียบ และไม่รับฟังว่ามีชายอีก 2 คนที่อยู่ในบริเวณใกล้เคียงที่พบความสัมพันธ์กับวัตถุพยานที่ตรวจเทียบ

หลังจากการสอบสวน 5 ปี ทาง FBI ได้ประกาศว่า พบความผิดพลาดในการคำนวณทางสถิติหลังจากนำคำถามจากการค้นพบของห้องปฏิบัติการของรัฐแอริโซน่ามาประกอบ การพบข้อผิดพลาดน้อยดังกล่าวจะพบน้อยลงในสิ่งส่งตรวจที่มีข้อมูลสารพันธุกรรมคุณภาพดีและปริมาณมาก ในปี พ.ศ.2558 ศาลในรัฐเท็กซัสได้รายงานว่า การคำนวณความน่าจะเป็นในการเจอความสัมพันธ์ของข้อมูลสารพันธุกรรมสามารถพบความผิดพลาดเกิดขึ้นได้ 1 ในพันล้าน ส่วนค่าความถูกต้องในการเลือกวิธีการคำนวณที่ถูกต้องคือ 1 ใน 100 แต่ข่าว

นี้มันก็สายเกินไปแล้วกับนาย Puckett ภายหลัง FBI ยอมรับในความผิดพลาดที่เกิดขึ้น และพยายามเริ่มช่วยเหลือนาย Puckett เรื่องการพิสูจน์ตัวเอง

จากบทความนี้แสดงให้เห็นว่า การตรวจพิสูจน์บุคคลนั้นต้องใช้ศาสตร์หลายแขนงในการตรวจพิสูจน์ ผลการตรวจพิสูจน์สารพันธุกรรมเพียงอย่างเดียวไม่สามารถใช้เป็นตัวตัดสินเพื่อยืนยันตัวบุคคลได้ ต้องใช้ข้อมูลแวดล้อมอื่นๆประกอบกัน เช่น ข้อมูลจากการสัมภาษณ์ญาติ ข้อมูลรูปพรรณสัณฐาน ข้อมูลสายวงศ์ตระกูล ข้อมูลทางการแพทย์ ข้อมูลการตรวจพิสูจน์ลายพิมพ์นิ้วมือ ข้อมูลทางทันตกรรม และข้อมูลอื่นๆที่มีประโยชน์ต่อการพิสูจน์บุคคล ดังนั้นผู้ตรวจพิสูจน์ทางนิติวิทยาศาสตร์ ด้านการตรวจพิสูจน์บุคคลต้องมีความรอบคอบ ช่างสังเกต และมีความรู้รอบด้าน เพื่อประโยชน์ในการตรวจพิสูจน์บุคคลที่ถูกต้องและสามารถมีเหตุผลประกอบการตรวจพิสูจน์ได้

เอกสารอ้างอิง

1. ERIN E. MURPHY. (2015). The Dark side of Forensic DNA Databases. Retrieved 8 August 2015 from https://www.theatlantic.com/science/archive/2015/10/the-dark-side-of-dna-database/408709/
2. พลตำรวจเอก อรรถพล แช่มสุวรรณวงศ์ และคณะ. (2544). นิติวิทยาศาสตร์ 3 เพื่อการสืบสวนสอบสวน (พิมพ์ครั้งที่ 1). กรุงเทพ: บริษัท ดาวฤกษ์ จำกัด.
3. แพทย์หญิงคุณหญิงพรทิพย์ โรจนสุนันท์. (2547). นิติเวชศาสตร์ การชันสูตรศพ. (พิมพ์ครั้งที่ 6). กรุงเทพ: บริษัท สำนักพิมพ์วิญญูชน จำกัด.
4. John M. Butler. (2001). FORENSIC DNA TYPING. Spain: Grafos SA Arte Sobre Papel.
5. J Robertson,/A. M. Ross,/and L. A. Burgoyne. (1990). DNA IN FORENSIC SCIENCE. England: Ellis Horwood Limited.
6. John M. Butler. (2005). FORENSIC DNA TYPING. (Secound edition). USA: Elsevier Academic press.