

บทที่ 2

ความรู้พื้นฐาน และเอกสาร/งานวิจัยที่เกี่ยวข้อง

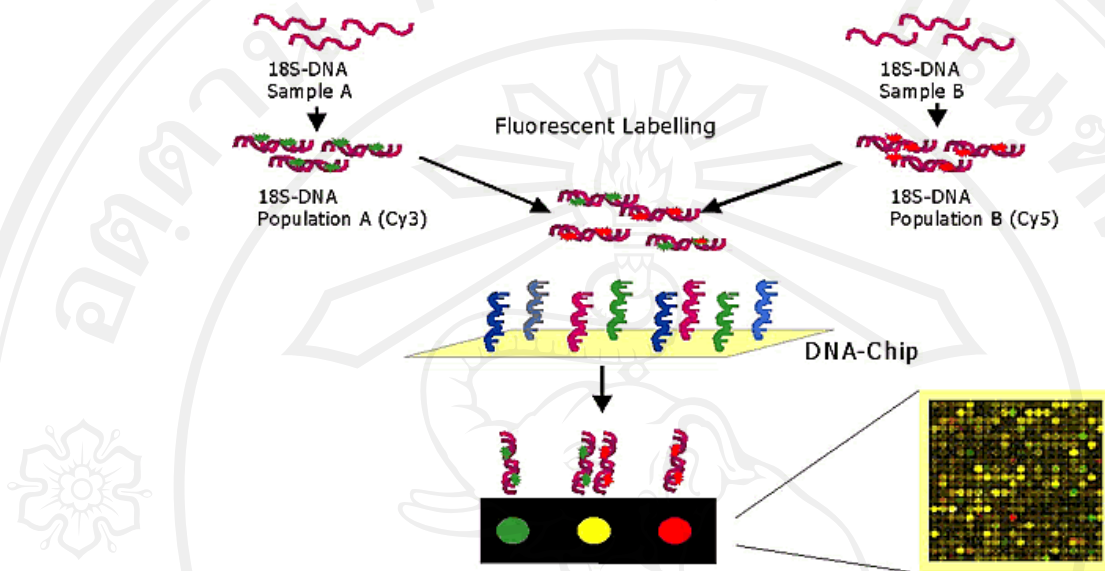
ในการวิเคราะห์ค่าการแสดงออกของยีนที่ได้จากเทคนิคไมโครอาร์เรย์ จำเป็นต้องใช้การทดสอบทางสถิติที่เหมาะสมกับลักษณะของข้อมูล ได้มีผู้นำเสนอวิธีการทดสอบทางสถิติต่าง ๆ ที่เหมาะสม และมีประสิทธิภาพแตกต่างกันไป ตามลักษณะของสภาพแวดล้อมของข้อมูล

เนื้อหาในบทนี้ จะกล่าวถึงความรู้พื้นฐานของเทคนิคไมโครอาร์เรย์ การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) รวมถึงปัญหาที่เกิดจากลักษณะของข้อมูลที่ได้จากเทคนิคไมโครอาร์เรย์ที่เรียกว่า การทดสอบพหุคูณ (Multiple Testing) ทฤษฎีที่เกี่ยวข้อง และเทคนิควิธีการทดสอบทางสถิติที่ได้มีการพัฒนาขึ้นเพื่อแก้ปัญหาที่เกิดขึ้น คือ เทคนิค MAANOVA ประกอบไปด้วยสถิติทดสอบ 3 ชนิด คือ วิธี Hybrid Statistic (F_2), วิธี Pooled-variance F Statistic (F_3) และ วิธี F_s -test (F_s) รวมไปถึงการทบทวนวรรณกรรมที่เกี่ยวข้อง ซึ่งจะรวมอยู่ในส่วนท้ายของบท

2.1 เทคนิคไมโครอาร์เรย์

2.1.1 หลักการพื้นฐาน

ยีนมีหน้าที่ควบคุมการทำงานของเซลล์โดยควบคุมให้เกิดการสร้างโปรตีนเพื่อใช้ในกิจกรรมต่าง ๆ เช่น โปรตีนบางชนิดทำหน้าที่เป็นโครงสร้างของเซลล์ โปรตีนบางชนิดทำหน้าที่เป็นเอนไซม์เป็นต้น กระบวนการสร้างโปรตีนชนิดหนึ่ง ๆ เริ่มต้นจากการถอดรหัสยีนหรือดีเอ็นเอ (Gene or DNA Transcription) จากนั้นรหัสของยีนจะถูกส่งออกมาเป็นโมเลกุลในรูปของเมสเซนเจอร์อาร์เอ็นเอ (Messenger RNA) หรือเอ็มอาร์เอ็นเอ (mRNA) จากนั้นจะมีกระบวนการแปลรหัส (Translation) เพื่อที่จะสร้างโปรตีนต่อไป ดังนั้นการตรวจพบเอ็มอาร์เอ็นเอของยีนใดย่อมแสดงว่าในขณะนั้นเซลล์กำลังมีการสร้างโปรตีนที่ถูกควบคุมโดยยีนนั้น เทคนิคไมโครอาร์เรย์ใช้ในการศึกษาการแสดงออกของยีนโดยตรวจสอบว่าในเวลานั้น ๆ มีเอ็มอาร์เอ็นเอชนิดใดอยู่ในเซลล์นอกจากใช้ตรวจสอบอาร์เอ็นเอแล้วยังใช้ตรวจสอบจีโนมิกดีเอ็นเอ (Genomic DNA) เช่นการตรวจการกลายพันธุ์ของดีเอ็นเอ (DNA Mutation) และตรวจการเปลี่ยนแปลงของยีนในโครโมโซมได้



รูปที่ 2.1 หลักการทำงานของเทคนิคไมโครอาร์เรย์ [14]

หลักการของเทคนิคไมโครอาร์เรย์ (แสดงในรูปที่ 2.1) อาศัยการจับกันอย่างจำเพาะเจาะจงของสายดีเอ็นเอสองสายที่เป็นคู่กัน (Complementary Strands of DNA) ดีเอ็นเอหรือยีนที่ทราบบทบาทหน้าที่แล้วจะถูกทำให้ติดอยู่บนฐานซึ่งเป็นแผ่นกระจกที่เคลือบด้วยแผ่นกระจกสไลด์ แผ่นกระจกนี้ได้ปรับให้มีคุณสมบัติที่ยอมให้ดีเอ็นเอเกาะติดที่ผิวหน้าได้ สามารถบรรจุชิ้นดีเอ็นเอได้เป็นจำนวนมากและอาจได้ถึง 30,000 ชิ้น ดังนั้นกระจกแผ่นนี้จึงเปรียบเสมือนแหล่งรวมของดีเอ็นเอหรือยีน เรียกแผ่นกระจกนี้ว่า “แผ่นไมโครอาร์เรย์ (Microarray Slide: DNA-Chip)” ซึ่งเมื่อนำดีเอ็นเอไม่ว่าจะเป็นจีโนมิกดีเอ็นเอหรือคอมพลีเมนต์ดีเอ็นเอ (Complementary DNA: cDNA) ที่สังเคราะห์จากเอ็มอาร์เอ็นเอมาทำปฏิกิริยาเพื่อให้จับคู่กัน (Hybridization) ดีเอ็นเอที่มีลำดับนิวคลีโอไทด์เข้าคู่กับดีเอ็นเอที่ติดอยู่บนแผ่นไมโครอาร์เรย์จะสามารถจับคู่กันได้ กล่าวคือดีเอ็นเอชนิดเดียวกันกับดีเอ็นเอที่อยู่บนแผ่นไมโครอาร์เรย์จะถูกจับติดอยู่บนแผ่นไมโครอาร์เรย์นั้น แต่ดีเอ็นเอที่มีลำดับนิวคลีโอไทด์แตกต่างออกไปจะไม่สามารถยึดจับกับดีเอ็นเอบนแผ่นไมโครอาร์เรย์ได้ จากนั้นนำแผ่นไมโครอาร์เรย์นี้ไปเข้าเครื่องอ่านเนื่องจากได้มีการติดฉลากเช่นแถบสีเรืองแสงไว้ที่ดีเอ็นเอเครื่องจึงสามารถตรวจวัดสีเรืองแสงได้ หากดีเอ็นเอถูกจับคู่

ติดอยู่กับดีเอ็นเอบนแผ่นไมโครอาร์เรย์ และความเข้มของสีก็สามารถบอกปริมาณดีเอ็นเอที่ถูกจับไว้ได้ [15]

2.1.2 การศึกษาการทำงานของยีนด้วยไมโครอาร์เรย์ของดีเอ็นเอ

ไมโครอาร์เรย์ของดีเอ็นเอเป็นเครื่องมือศึกษาการแสดงออกของยีนทั้งในขั้นตอนการสร้างเอ็มอาร์เอ็นเอและจีโนมิกดีเอ็นเอ ซึ่งการเลือกศึกษาในขั้นตอนใดจะต้องเตรียมแผ่นไมโครอาร์เรย์ให้เหมาะสม โดยสามารถแบ่งกลุ่มการศึกษาได้เป็นการศึกษาระดับการแสดงออกหรือการทำงานของยีนจำนวนยีน และการกลายพันธุ์ของดีเอ็นเอดังรายละเอียดต่อไปนี้

2.1.2.1 ศึกษาการแสดงออกหรือการทำงานของยีน (Study of Gene Expression Level or Gene Function)

โดยอาจศึกษาเปรียบเทียบระดับการแสดงออกของยีนระหว่างเซลล์ปกติกับเซลล์ที่ผิดปกติ ทำให้ทราบได้ว่าในสถานะที่เซลล์เกิดความผิดปกติ มียีนใดแสดงออกมากน้อยต่างกันอย่างไร แผ่นไมโครอาร์เรย์ที่ใช้ในการศึกษาจะต้องประกอบด้วยคอมพลีเมนต์อาร์ดีเอ็นเอที่สังเคราะห์จากอาร์เอ็นเอในเซลล์นั้น ๆ และเมื่อทำการไฮบริไดเซชันด้วยคอมพลีเมนต์อาร์ดีเอ็นเอ ที่ได้จากอาร์เอ็นเอของเซลล์แต่ละชนิด ผลที่ปรากฏจะบอกให้ทราบถึงระดับการแสดงออกของยีนแต่ละชนิดได้ เช่นยีนบางชนิดอาจแสดงออกน้อยลงเมื่อเซลล์ไม่ปกติ ในขณะที่ยีนบางชนิดกลับแสดงออกในทางกลับกัน ซึ่งสามารถนำข้อมูลที่ได้ไปใช้ในการวินิจฉัยโรค ตลอดจนเป็นแนวทางในการค้นคว้าตัวยาใหม่ ๆ ได้ นอกจากนี้ยังใช้ศึกษารูปแบบการแสดงออกของยีน (Gene Expression Pattern) ในช่วงอายุต่าง ๆ ของเซลล์เนื่องจากยีนมีหน้าที่ควบคุมการเจริญเติบโตของเซลล์ ดังนั้นยีนบางชนิดจะทำงานในช่วงอายุของเซลล์เท่านั้น ซึ่งการศึกษาในลักษณะเช่นนี้จะเป็แนวทางในการตรวจวินิจฉัยแนวโน้มในการเกิดมะเร็ง [1]

2.1.2.2 ศึกษาจำนวนยีน (Study of Gene Numbers)

การทำงานของเซลล์ถูกควบคุมโดยยีน ดังนั้นการที่ยีนถูกทำลายจะมีผลให้การทำงานของเซลล์ผิดปกติและมีโอกาสกลายเป็นเซลล์มะเร็งได้ อย่างไรก็ตามร่างกายมีกระบวนการซ่อมแซมดีเอ็นเอที่ถูกทำลาย โดยมียีนกลุ่มหนึ่งคอยควบคุมอยู่เรียกว่า “DNA Repair Genes” ซึ่งอาจมีผลให้ยีนบางส่วนบนโครโมโซมหายไปหรือมียีนบางส่วนเพิ่มจำนวน ซึ่งไมโครอาร์เรย์สามารถตรวจวัดการเพิ่มขึ้นหรือลดลงของจำนวนยีนบนโครโมโซมได้ โดยการใช้แผ่นไมโครอาร์เรย์ที่เป็นจีโนมิกดีเอ็นเอโดยเตรียมดีเอ็นเอจากโครโมโซมชุดใดชุดหนึ่งผลการตรวจสอบจะบอกได้ว่าบนโครโมโซมนั้น ๆ มียีนใดหายไปหรือมียีนใดเพิ่มจำนวน

2.1.2.3 ศึกษาการกลายพันธุ์ของดีเอ็นเอ (Study of DNA Mutation)

ไมโครอาร์เรย์สามารถใช้ตรวจสอบการกลายพันธุ์ของยีนและตรวจสอบภาวะพหุสัณฐานของยีนได้ ยีนที่ก่อกลายพันธุ์และยีนที่มีภาวะพหุสัณฐานเป็นยีนที่มีนิวคลีโอไทด์ในบางตำแหน่งต่างไปจากกลุ่ม ซึ่งความแตกต่างนี้อาจมีผลให้ยีนมีการแสดงออกต่างกัน ผลที่ปรากฏเหมือนเช่นการที่คนกลุ่มหนึ่งมีการตอบสนองต่อยาต่างจากคนอีกกลุ่มหนึ่ง การตรวจสอบด้วยไมโครอาร์เรย์ทำได้โดยการเตรียมแผ่นไมโครอาร์เรย์ด้วยจีโนมิกดีเอ็นเอ ซึ่งเมื่อทำการไฮบริไดเซชันด้วยจีโนมิกดีเอ็นเอที่สกัดจากเซลล์ดีเอ็นเอที่มีภาวะพหุสัณฐานเหมือนกันจะจับคู่กันและปรากฏเห็นสีที่สามารถตรวจวัดได้ [16]

2.1.3 การประยุกต์ใช้เทคนิคดีเอ็นเอไมโครอาร์เรย์กับงานทางด้านพิษวิทยา

Hitoshi [17] ได้ทำงานวิจัยที่พบ Synergistic จากการผสมแคดเมียมและไทอยูเรม แล้วใส่สารพิษกลุ่มผสมนี้ลงไปในปีศาจ จากการทดลองพบว่าความเป็นพิษได้เพิ่มมากขึ้นในสารกลุ่มผสมเมื่อเทียบกับการใส่แคดเมียมหรือไทอยูเรมเพียงอย่างเดียวลงในเซลล์ ทั้งนี้เนื่องจากสารกลุ่มผสมสามารถทำลายการทำงานของไมโทคอนเดรีย แต่ไม่พบการทำลายไมโทคอนเดรียในเซลล์ที่ใส่แคดเมียมหรือไทอยูเรมเพียงอย่างเดียว โดยการค้นพบนี้ได้ใช้เทคนิคดีเอ็นเอไมโครอาร์เรย์มาเป็นเครื่องมือหลักในการวิจัยซึ่งทำให้เห็นถึงความเป็นไปได้ในการนำเทคนิคดีเอ็นเอไมโครอาร์เรย์มาประยุกต์ใช้กับงานทางด้านพิษวิทยา

2.2 ทฤษฎีที่เกี่ยวข้อง

2.2.1 ทฤษฎีการขยายพจน์ของเทเลอร์ (Theory of the Taylor Expansion) [18]

ฟังก์ชัน $f(x)$ ประกอบไปด้วยพจน์ $1, 2, \dots, n+1$ จากหลักการการปริพันธ์จะได้ว่า

$$f(x) - f(a) = \int_a^x f'(t) dt$$

กฎการหาผลคูณของอนุพันธ์ สำหรับฟังก์ชันใด ๆ u กับ v คือ

$$(uv)' = u'v + uv'$$

เมื่อทำการปริพันธ์ จะได้ว่า

$$\int u dv = uv - \int v du.$$

หากกำหนดให้

$$u = f'(t), \text{ and } dv = dt,$$

จะทำให้ได้ว่า

$$du = f''(t)dt, \text{ and } v = t - x,$$

ดังนั้นการปริพันธ์เทียบกับ t จะทำให้ได้ในแต่ละส่วนดังนี้

$$\int_a^x f'(t)dt = \left[(t-x)f'(t) \right]_{t=a}^{t=x} + \int_a^x f''(t)(x-t)dt$$

ซึ่งหมายถึง

$$f(x) = f(a) + f'(a)(x-a) + \int_a^x f''(t)(x-t)dt.$$

จึงได้ฟังก์ชัน $f(x)$ ในรูปของ $f(a)$ ณ จุด $x = a$

ในการทำอนุพันธ์อันดับสอง จะสามารถใช้หลักการรูปแบบเดียวกัน โดยการทำการปริพันธ์ในแต่ละส่วน จาก

$$u = f''(t), \text{ and } dv = (x-t)dt$$

เมื่อทำการอนุพันธ์ค่า u เทียบกับค่า t และปริพันธ์ค่า dv เทียบกับ t จะได้

$$du = f^{(3)}(t)dt, \text{ and } v = -\frac{1}{2}(t-x)^2,$$

(สัญลักษณ์ $f^{(n)}(x)$ หมายถึงเทอมที่ n ของการอนุพันธ์ฟังก์ชัน $f(x)$)

ดังนั้นพจน์การกระจายตัวของ $f(x)$ จึงเปลี่ยนไปเป็น

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{1}{2} \int_a^x f^{(3)}(t)(x-t)^2 dt$$

หากทำการปริพันธ์แต่ละส่วน เป็นจำนวน p รอบ ก็จะได้ผลลัพธ์คือ

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt,$$

โดย

$$n! = n (n-1) (n-2) \dots 3 2 1$$

การกระจายพจน์ของฟังก์ชัน $f(x)$ ข้างต้น ได้ถูกเรียกว่า สูตรคิดเครื่องหมายปริพันธ์ของเทเลอร์ [18] (Taylor's Formula with Integral Remainder) หากให้ $n \rightarrow \infty$ จะสามารถเขียนสมการข้างต้นได้อีกรูปคือ

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \dots,$$

หรือ

$$f(x) = \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!} f^{(n)}(a),$$

ทั้งนี้ พจน์ $n!$ จะมีค่าเท่ากับ 1 เมื่อ $n = 0$

2.2.2 ตัวประมาณที่ได้จากการหดพจน์ของเจมส์-สตีน์ (James-Stein Shrinkage Estimator)

การหดพจน์ของเจมส์-สตีน์ [19] นั้นเป็นการวิเคราะห์เพื่อให้พจน์ใด ๆ ก็ตามที่มีจำนวนมิติที่ซับซ้อน หรือมีหลายมิติให้มีความเรียบง่ายขึ้น ตามหลักการแล้วเหมาะสมกับการวิเคราะห์ที่มีจำนวนตัวอย่างน้อย ในตัวประมาณดั้งเดิมตั้งแต่แรกเริ่มนั้น เป็นการพิจารณาที่ $n = 1$ แต่ก็มีการพัฒนาให้มีความหลากหลายมากขึ้นในเวลาต่อ ๆ มา

การหดพจน์ของเจมส์-สตีน์ มีพื้นฐานมาจากโมเดลตั้งแต่ 2 โมเดลขึ้นไปที่มีความแตกต่างกันอย่างมาก เช่น โมเดลที่มีความหลากหลายในมิติสูงมาก ประกอบกับมี Bias ต่ำ และ ความแปรปรวนสูง กับ โมเดลที่มีความหลากหลายในมิติต่ำกว่า แต่มี Bias สูงกว่าในขณะที่มีความแปรปรวนต่ำกว่า

พิจารณา

$$\hat{\theta}_k^{Shrink} = \lambda t_k + (1 - \lambda) \hat{\theta}_k^{ML}$$

ซึ่ง $\lambda \in [0,1]$ เป็นความหนาแน่นของการหด (Shrinkage Intensity) ซึ่งมีค่าระหว่าง 0 (แปลว่าไม่มีการหด) ถึง 1 (แปลว่าหดเต็มรูปแบบ) และ t_k เป็นเป้าหมายในการหด (Shrinkage Target) t_k มีการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) พิจารณา ค่าเอนเอียง (Bias) $(\hat{\theta}_k^{ML}) = 0$ โดยการใช้ตัวประมาณที่ไม่เอนเอียง (Unbiased Estimator) $Var(\hat{\theta}_k^{ML}) = \frac{\hat{\theta}_k^{ML}(1 - \hat{\theta}_k^{ML})}{n-1}$ เราจะได้ความหนาแน่นของการหดดังนี้

$$\hat{\lambda}^* = \frac{\sum_{k=1}^p Var(\hat{\theta}_k^{ML})}{\sum_{k=1}^p (t_k - \hat{\theta}_k^{ML})^2} = \frac{1 - \sum_{k=1}^p (\hat{\theta}_k^{ML})^2}{(n-1) \sum_{k=1}^p (t_k - \hat{\theta}_k^{ML})^2}$$

2.2.3 The Lowess Normalization

วิธีการ Normalization มีได้มากมายหลายวิธีแต่ในงานวิจัยครั้งนี้เลือกใช้วิธี LOWESS (Locally Weighted Scatterplot Smoothing) ซึ่งมีพื้นฐานมาจากเทคนิค Regression ทำการสร้างวิธีการ Normalization เพื่อข้อมูลแบบไมโครอาร์เรย์โดยเฉพาะ [21] โดยค่าที่นำมาทำการปรับให้เป็นการแจกแจงปกตินั้น หากเป็นการทดลองไมโครอาร์เรย์แบบสองสี จะใช้ค่าเฉลี่ยการแสดงออกของยีนของสีทั้งสอง (สีแดง:R และสีเขียว:G) โดยกำหนดให้สีแดงหนึ่งเป็นสีของ Treatment Dye และให้อีกสีเป็น Control Dye แต่ถ้าหากการทดลองเป็นแบบสีเดียว ก็จะใช้ค่าเฉลี่ยการแสดงออกของยีนควบคุม กับยีนที่ผ่านกรรมวิธีมาทำการคำนวณ โดยมีสูตรการคำนวณสำหรับแบบสีเดียวคือ

$$\bar{x}'_i \rightarrow \log(\bar{x}_i) - \frac{\log(\bar{x}_c / \bar{x}_i)}{\log(\sqrt{\bar{x}_c} \times \bar{x}_i)}$$

เมื่อ $\bar{x}_i$ เป็นค่าเฉลี่ยการแสดงออกของยีนของแต่ละยีนภายหลังจากได้รับกรรมวิธีก่อน Normalized และ $\bar{x}_c$ เป็นค่าเฉลี่ยการแสดงออกของยีนของสไลด์ควบคุม

2.3 การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) [13]

การวิเคราะห์ค่าการแสดงออกของยีนจากเทคนิคไมโครอาร์เรย์ มีเป้าหมายหนึ่งก็คือการวิเคราะห์เพื่อระบุยีนที่มีระดับการแสดงออกแตกต่าง (Differentially Genes Expression) อย่างมีนัยสำคัญเมื่อได้รับกรรมวิธีที่แตกต่างกัน [6-7] โดยอาศัยวิธีการทดสอบทางสถิติ เปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยในแต่ละกลุ่มกรรมวิธี ในกรณีทำการทดสอบเพื่อศึกษาถึงยีนตัวหนึ่งว่ามีความแตกต่างในค่าการแสดงออกของยีนเมื่อได้รับกรรมวิธีที่แตกต่างกัน เราสามารถใช้วิธีการทดสอบทางสถิติ เช่น การทดสอบ t (t -test) การทดสอบ Z (z -test) สำหรับการทดสอบระหว่างกรรมวิธี 2 กลุ่มแต่หากเป็นการทดสอบระหว่างกรรมวิธีมากกว่า 2 กลุ่มขึ้น การใช้การทดสอบ t และการทดสอบ Z จะทำให้ต้องทำการทดสอบทีละคู่ ซึ่งจะทำให้เกิดปัญหาความยุ่งยากและอำนาจการทดสอบ (Power of the Test) ก็จะมีค่าต่ำ ทำให้จำเป็นต้องใช้วิธีการทางสถิติที่เรียกว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) ซึ่งข้อมูลที่จะนำมาวิเคราะห์ด้วยวิธีการนี้ จะต้องเป็นข้อมูลที่สอดคล้องกับข้อสมมติต่าง ๆ คือ ประชากรที่ต้องการทดสอบทุกกลุ่ม มีการแจกแจงแบบปกติ และมีความแปรปรวนเท่ากัน เมื่อต้องการทดสอบกรรมวิธีต่าง ๆ k วิธี โดยทำการทดลองกับหน่วยทดลองที่มีลักษณะคล้ายคลึงกัน (Homogeneous) ซึ่งในงานวิจัยครั้งนี้ก็เทียบได้กับการที่ยีนตัวที่ g จากทั้งหมด G ยีนในแต่ละแผ่นสไลด์ที่ใช้ในการทดลองด้วยเทคนิคไมโครอาร์เรย์ จะต้องมีความเหมือนกันทุกประการและทำการทดสอบด้วยกรรมวิธีต่าง ๆ กัน k วิธี และทำซ้ำในแต่ละกรรมวิธีเป็นจำนวน n ซ้ำ (n สไลด์) การทดสอบจะใช้แผนการทดลองที่เรียกว่า แผนการทดลองแบบสุ่มตลอด (Completely Randomized design: CRD) โดยมีสมมติฐานที่จะทดสอบสำหรับยีนแต่ละตัวคือ

สมมติฐานหลัก H_0 : การแสดงออกของยีนไม่แตกต่างกันระหว่างกรรมวิธีทั้ง k วิธี

สมมติฐานรอง H_1 : มีการแสดงออกของยีนที่ได้รับกรรมวิธี อย่างน้อย 2 กลุ่ม แตกต่างกัน

ตารางที่ 2.1 ค่าสังเกตของยีนตัวที่ g จากทั้งสิ้น N ยีน

	กรรมวิธีที่				
	1	2	...	k	
ทำซ้ำครั้งที่ 1	x_{11}	x_{12}		x_{1k}	
ทำซ้ำครั้งที่ 2	x_{21}	x_{22}		x_{2k}	
ทำซ้ำครั้งที่ ...	...	...	x_{ij}	...	
ทำซ้ำครั้งที่ n	x_{n1}	x_{n2}		x_{nk}	
รวม	$T_{.1}$	$T_{.2}$	$T_{.j}$	$T_{.k}$	$T_{..}$
ค่าเฉลี่ย	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.j}$	$\bar{x}_{.k}$	$\bar{x}_{..}$

การแยกผลบวกกำลังสอง

ในการวิเคราะห์เพื่อเปรียบเทียบอิทธิพลของกรรมวิธีต่าง ๆ ที่มีต่อผลการทดลองนั้นจะพิจารณาจากความแปรผันของข้อมูลโดยส่วนรวม ซึ่งเรียกว่า ผลบวกกำลังสองของการทดลอง หรือของทั้งหมด (Sum Square Total: SST) นิยามดังนี้

$$SST = \sum_{i=1}^{n_j} \sum_{j=1}^k (x_{ij} - \bar{x}_{..})^2$$

โดยทฤษฎีทางสถิติ สามารถแยก SST ออกเป็น 2 ส่วนได้ ดังนี้

$$\sum_{i=1}^{n_j} \sum_{j=1}^k (x_{ij} - \bar{x}_{..})^2 = \sum_j^k n_j (\bar{x}_{.j} - \bar{x}_{..})^2 + \sum_i^{n_j} \sum_j^k (x_{ij} - \bar{x}_{.j})^2$$

พิจารณาพจน์ต่าง ๆ ดังต่อไปนี้

$\sum_j^k n_j (\bar{x}_{.j} - \bar{x}_{..})^2$: เป็นผลรวมของความแตกต่างกำลังสองระหว่างค่าเฉลี่ยของกรรมวิธีที่ j กับค่าเฉลี่ยทั้งหมด เป็นค่าที่วัดความแปรผันของอิทธิพลอันเนื่องมาจากกรรมวิธีต่อไปนี้จะเรียกพจน์ดังกล่าวว่า ผลบวกกำลังสองของกรรมวิธี (Sum Square Treatment: SSTr)

$\sum_i^{n_j} \sum_j^k (x_{ij} - \bar{x}_{.j})^2$: เป็นผลรวมของความแตกต่างกำลังสอง ระหว่างค่าสังเกตภายในกรรมวิธีกับค่าเฉลี่ยของกรรมวิธีนั้น ๆ เป็นค่าที่ใช้วัดความแปรผันของความคลาดเคลื่อนของการทดลอง ต่อไปนี้จะเรียกพจน์ดังกล่าวว่า ผลบวกกำลังสองของความคลาดเคลื่อนของแผนการทดลองแบบสุ่มตลอด (Sum Square of Error: *SSE*)

ดีกรีความเป็นอิสระ (Degree of Freedom: *df*) ของความแปรผันแต่ละชนิด

ความแปรผันของการทดลอง $df = n-1$

ความแปรผันของกรรมวิธี $df = k-1$

ความแปรผันของความคลาดเคลื่อน $df = n-k$

ซึ่งดีกรีความเป็นอิสระข้างต้น มีความสัมพันธ์กันดังนี้

$$n-1 = (k-1) + (n-k)$$

โดย n คือ จำนวนการทำซ้ำในแต่ละกรรมวิธี และ k คือจำนวนกรรมวิธีที่ใช้ในการทดลอง

เมื่อผลการบวกกำลังสองของความแปรผันดังกล่าวด้วยดีกรีความเป็นอิสระของความแปรผันนั้น ๆ จะได้ความแปรผันเฉลี่ย (Mean Square: *MS*) ซึ่งก็คือ ความแปรปรวน

ความแปรผันเฉลี่ยของการทดลอง (Mean Square Total: *MST*)

$$MST = \frac{SST}{n-1}$$

ความแปรผันเฉลี่ยของกรรมวิธี (Mean Square Treatment: *MStr*)

$$MStr = \frac{SStr}{k-1}$$

ความแปรผันเฉลี่ยของความคลาดเคลื่อน (Mean Square Error: *MSE*)

$$MSE = \frac{SSE}{n-k}$$

โดยจะได้สถิติทดสอบ F ดังนี้

$$F = \frac{MStr}{MSE}$$

โดยจะทำการปฏิเสธ H_0 เมื่อ $F > F_{1-\alpha, k-1, n-k}$

หรือทำการกำหนดค่าระดับนัยสำคัญ α จากนั้นพิจารณาว่า P-Value จากตารางโดยจะยอมรับสมมติฐาน

หลัก ถ้า P-Value มากกว่า α และปฏิเสธ ถ้า P-Value เท่ากับหรือน้อยกว่า

ตารางวิเคราะห์ความแปรปรวนเป็นดังนี้

ตารางที่ 2.2 ตารางวิเคราะห์ความแปรปรวน

แหล่งของความแปรผัน	Degree of Freedom	SS	MS	F-ratio
กรรมวิธี	$k-1$	$SStr$	$MStr$	$F = \frac{MStr}{MSE}$
ความคลาดเคลื่อน	$n-k$	SSE	MSE	
รวม	$n-1$	SST		

2.4 เทคนิค MAANOVA

MAANOVA ย่อมาจาก Microarray Analysis of Variance เป็นเทคนิคทางสถิติที่ถูกพัฒนาขึ้นโดย Xiangqin Cui และ Gary A. Churchill [10] จากพื้นฐานของการวิเคราะห์ความแปรปรวน (ANOVA: Analysis of Variance) สำหรับใช้ในการวิเคราะห์ถึงการแสดงออกที่แตกต่างกันของยีนเริ่มจากสถิติทดสอบ F -test และพัฒนาต่อมาโดยใช้หลักของความแปรปรวนร่วม (Pool Variance) ได้เป็นสถิติทดสอบ F_2 -test และ F_3 -test ซึ่งได้รับการทดสอบว่าในหลายกรณีมีประสิทธิภาพดีกว่า F -test ล่าสุดผู้สร้างได้พัฒนาตัวสถิติ F_s -test ซึ่งใช้หลักการของตัวประมาณด้วยการหด (Shrinkage Estimator) ทำให้ได้ตัวสถิติที่มีประสิทธิภาพดีกว่าตัวสถิติอื่น ๆ ในหลายกรณี

สถิติทดสอบ F_s -test (F_s) ใน MAANOVA

วิธีทดสอบ F_s เป็นวิธีทดสอบที่มีการพัฒนาเพื่อใช้วิเคราะห์ข้อมูลที่ได้จากเทคนิคไมโครอาร์เรย์วิธีล่าสุด เป็นการเปลี่ยนแปลงในส่วนของตัวส่วนในสูตรสถิติทดสอบ โดยใช้หลักการของ Shrinkage Estimator [19]

ให้ X_g แทนค่าของผลบวกกำลังสองของความคลาดเคลื่อนของแผนการทดลองแบบสุ่มตลอด (SSE) และ σ_g^2 แทนค่าของความแปรปรวนที่แท้จริงในยีน g ($g = 1, 2, \dots, G$) จากการ Assume ให้ X_g / σ_g^2 มีความเป็นอิสระ (Independent) แต่ละตัวมีการแจกแจงแบบไคสแควร์ (Chi-squared Distribution) ด้วยองศาความเป็นอิสระ ν จึงทำให้ได้ว่า

$$\frac{X_g}{\sigma_g^2} \sim \chi_\nu^2 \quad \text{ดังนั้น} \quad X_g \sim \sigma_g^2 \chi_\nu^2$$

ทำการใส่ค่า $\log$ ทั้งสองข้างสมการ จะได้ว่า

$$\ln \frac{X_g}{\nu} \sim \ln \sigma_g^2 + \ln \frac{\chi_\nu^2}{\nu} \quad (1)$$

ให้ m แทนค่าเฉลี่ยของ $\ln \frac{\chi_v^2}{v}$ ทำการลบด้วย m ทั้งสองข้างจะทำให้ได้

$$X'_g \sim \ln \sigma_g^2 + \varepsilon'_g$$

โดยที่ $X'_g = \ln \frac{X_g}{v} - m$ และ $\varepsilon'_g = \ln \frac{\chi_v^2}{v} - m$

ให้ v แทน ความแปรปรวนของ ε'_g จากทฤษฎีการขยายพจน์ของเทเลอร์ [18] จะทำให้ได้ว่า

$$\text{Var}\left(\ln \frac{\chi_v^2}{v}\right) \approx \text{Var}\left(\frac{\chi_v^2}{v}\right) = \frac{2}{v}$$

จากทฤษฎีตัวประมาณค่าแบบหดของ James-Stein-Lindley กล่าวว่าเมื่อทำการหดเข้าสู่

Common Mean, $\bar{X}' = \sum X'_g - \bar{X}'/G$ จะทำให้ได้ $\bar{X}' + \left(1 - \frac{(G-3)V}{\sum (X'_g - \bar{X}')^2}\right)_+ \times (X'_g - \bar{X}')$

โดย $\left(1 - \frac{(G-3)V}{\sum (X'_g - \bar{X}')^2}\right)_+$ หมายถึงค่าดังกล่าวจะไม่สามารถเป็นค่าลบ

จากนั้นทำการ Transformation กลับไปสู่ Scale เดิมจึงทำให้ได้สถิติทดสอบคือ

$$F_s = \frac{\left(\sum_{i=1}^{n_i} \sum_{j=1}^k (x_{ij} - \bar{x}_{..})^2 - \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2\right) \div (k-1)}{\left(\prod_{g=1}^G (X_g / v)^{1/G} B \times \exp \left[\left(1 - \frac{(G-3)V}{\sum (\ln X_g - \ln X_g)^2}\right)_+ \times (\ln X_g - \ln X_g) \right] \right)}$$

โดย G = จำนวนยีนทั้งหมด

$$X_g = \text{ค่า SSE หรือ } \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2$$

v = องศาความเป็นอิสระของ SSE หรือ $n-k$

ค่า B และ V จะเป็นไปตามตารางที่ 2.1 ซึ่งสร้างขึ้นโดย Cui และ Churchill [10] ดังนี้

ตารางที่ 2.3 ค่า B และ V ในรูปฟังก์ชันขององศาความเป็นอิสระ ν ของ Cui และ Churchill

ν	B	$V/(2/\nu)$	ν	B	$V/(2/\nu)$
1	3.53	2.45	13	1.08	1.08
2	1.77	1.64	14	1.08	1.08
3	1.44	1.39	15	1.07	1.07
4	1.31	1.27	16	1.07	1.06
5	1.24	1.22	17	1.06	1.06
6	1.19	1.18	18	1.06	1.06
7	1.16	1.15	19	1.06	1.05
8	1.14	1.13	20	1.05	1.05
9	1.12	1.12	25	1.04	1.04
10	1.11	1.11	30	1.04	1.03
11	1.10	1.10	40	1.03	1.03
12	1.09	1.09	50	1.02	1.02

2.5 เอกสาร และงานวิจัยที่เกี่ยวข้อง

Wongta [1] ได้ทำการทดลองโดยเทคนิคเอนเอไมโครอาร์เรย์ในการวิเคราะห์ความเป็นพิษของสารพิษต่าง ๆ ดังนี้ สารซิลเวอร์นาโน อาเซนิก แคดเมียม และโครเมียม โดยสารพิษทั้งสี่ชนิดนี้ได้ถูกนำมาผสมกันเป็นคู่ คือ ซิลเวอร์นาโนผสมกับอาเซนิก (AgNPs-As) ซิลเวอร์นาโนผสมกับแคดเมียม (AgNPs-Cd) และซิลเวอร์นาโนผสมกับโครเมียม (AgNPs-Cr) จากนั้นได้ทำการตรวจสอบความเป็นพิษต่อเซลล์ HepG2 ด้วยเทคนิค Cytotoxicity และทำการตรวจสอบการแสดงออกของยีนด้วยเทคนิคเอนเอไมโครอาร์เรย์ ผลการทดลองพบว่า การผสมกันของซิลเวอร์นาโนกับอาเซนิก และซิลเวอร์นาโนกับแคดเมียม มีพิษต่อเซลล์มากกว่าการใส่สารพิษเพียงตัวเดียวลงไปในเซลล์การทดลองในครั้งนี้ใช้ Human Genome Focus Array ที่ประกอบไปด้วยยีนของมนุษย์ทั้งหมด 8793 ยีน จากบริษัท Affymetrix และ ทำการสแกน Probe Arrays ด้วยเครื่อง Genechip Ayatem Confocal Scanner (Affymetrix, Santa Clara, CA) โดยในทุกๆ ชุดทดลองจะทำซ้ำ สามซ้ำ จากนั้นข้อมูลดิบที่ได้จากการสแกนจะถูกนำไปวิเคราะห์ทางสถิติด้วยโปรแกรม Gene Chip Operating Software (GCOS) (Affymetrix, Santa Clara, CA) ซึ่งจะทำได้ข้อมูลออกมาในรูปแบบของ “CEL File” และนำไฟล์ที่ได้ไปวิเคราะห์หา ยีนที่มีการเปลี่ยนแปลงการแสดงออกอย่างมีนัยสำคัญด้วยโปรแกรม Avadis 4.2 Prophetic (Strand Genomics, Redwood City, CA) ซึ่ง MAS5 Algorithm ได้ถูกนำมาใช้ในการ Scale และ Normalize ข้อมูลดิบที่มาจาก การสแกนซึ่งเรียกว่า ค่า Signal Intensity ขั้นตอนสุดท้ายคือ ยีนที่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญ โดยมีค่า FC Value > 2 และ $p < 0.05$ จะถูกนำไปวิเคราะห์หา GO Terms of Biological Process at Level 6 ตัวสถิติที่ใช้ในการทดลองนี้คือ แบบทดสอบ t (t -test) ผลการทดลองทั้งหมดสรุปได้ว่าการผสมสารซิลเวอร์นาโนกับโลหะหนักสามชนิดคือ อาเซนิก แคดเมียมและโครเมียมนั้น ส่งผลให้เซลล์ HepG2 มีการแสดงออกของยีนที่เปลี่ยนไปเมื่อเทียบกับการใส่สารพิษเพียงตัวเดียวลงไปในเซลล์ และสารคู่ผสมระหว่างซิลเวอร์นาโนกับอาเซนิก และซิลเวอร์นาโนกับแคดเมียมนั้นมีความเป็นพิษสูงขึ้นเนื่องจากสารคู่ผสมสามารถชักนำให้เกิดการตายของเซลล์เพิ่มมากขึ้นโดยผ่านกลไกที่ซับซ้อน เช่น กระบวนการสร้างไขมันในเซลล์ และกระบวนการเพิ่มจำนวนของเซลล์เป็นต้น ทั้งนี้หากต้องการทราบกลไกที่ชัดเจนจะต้องทำการทดลองเชิงลึกในระดับของโปรตีนและชีวเคมีต่อไป

Cui and Churchill [10] ทำการพัฒนาวิธีการวิเคราะห์ข้อมูลทางสถิติ เพื่อการทดสอบสมมติฐานค่าการแสดงออกของยีนที่มีความแตกต่างอย่างมีนัยสำคัญ สำหรับข้อมูลที่ได้จากเทคนิคไมโครอาร์เรย์ โดยใช้พื้นฐานแนวคิดจากเทคนิคของ James-Stein Shrinkage ทำให้ได้วิธีการทางสถิติวิธี

ใหม่ที่เรียกว่าวิธี F_5 -test โดยผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของตัวสถิติตัวใหม่ กับวิธี F_1 (The Gene-specific F test), F_2 (Hybrid Statistic), F_3 (The Pooled-variances F Statistic), The Regularized t -statistic, The Posterior Odds Statistic B และ SAM t -test ซึ่งให้ผลสรุปว่า ประสิทธิภาพของวิธี F_5 ให้กำลังในการทดสอบหาขึ้นที่มีการแสดงออกที่แตกต่างอย่างมีนัยสำคัญดีที่สุดที่สุดในเกือบทุกกรณี วิธีทดสอบดังกล่าว ยังมีความแกร่ง (Robust) และสามารถประยุกต์ใช้ในการวิเคราะห์ข้อมูลที่ได้จากเทคนิคไมโครอาร์เรย์ในหลายกรณี และมีข้อจำกัดน้อยกว่ากรณีทั่วไป

Cui and Churchill [11] ได้เสนอวิธีการทางสถิติสำหรับการทดสอบสมมติฐานค่าการแสดงออกของยีนที่มีความแตกต่างอย่างมีนัยสำคัญ สำหรับข้อมูลที่ได้จากเทคนิคไมโครอาร์เรย์ โดยพัฒนามาจากแนวคิดของ สถิติทดสอบ t (t -test) และ Mixed ANOVA Model ทำให้ได้สถิติ F_2 (Hybrid Statistic) และ F_3 (The Pooled-variances F Statistic) สำหรับทดสอบค่าการแสดงออกของยีน โดยทั้งสองวิธีมีประสิทธิภาพในการทดสอบที่ดีกว่าสถิติทดสอบ F -test (สำหรับการวิเคราะห์ความแปรปรวน)