บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

ปัจจุบันองค์การทะเบียนสารเคมีระหว่างชาติ ภายใต้การดำเนินงานของโครงการสิ่งแวดล้อมแห่งสหประชาชาติได้รายงานว่ามีสารเคมีจำนวนมากประมาณกว่า 6 ล้านชนิดเกิดขึ้นในโลกทั้งโดยธรรมชาติและโดยการสังเคราะห์ขึ้น ในจำนวนนี้ 60,000 ชนิดเป็นสารเคมีที่มนุษย์นำมาใช้ในกิจการต่าง ๆ สำหรับการดำเนินชีวิตประจำวัน คือนำมาใช้ในการเกษตรกรรม อุตสาหกรรม เป็นส่วนประกอบในอาหาร เครื่องอุปโภคบริโภค และสาธารณูปโภครวมทั้งยารักษาโรคเพื่อความเป็นอยู่ที่ดี ในทุก ๆ ปีจะมีสารเคมีถูกแนะนำสู่ตลาดเพิ่มขึ้นประมาณ 1,000 ชนิด สารเคมีทั้งหมดนี้ได้รับการพิสูจน์แล้วว่าเป็นพิษต่อมนุษย์และสิ่งแวดล้อมได้ทั้งทางตรงและทางอ้อม คือ จากการบริโภค การสัมผัสโดยตรงและ/หรือออกมากับของเสียในกระบวนการผลิตทางอุตสาหกรรมต่าง ๆ เข้าสู่สิ่งแวดล้อมและห่วงโซ่อาหาร ความเป็นพิษของสารเคมีเกิดขึ้นได้ทั้งแบบเฉียบพลันและเรื้อรัง ทั้งนี้ ขึ้นอยู่กับปริมาณที่สัมผัส ลักษณะการสัมผัส ช่วงเวลา และคุณสมบัติความเป็นพิษของสารเคมีตัวนั้น ซึ่งจะส่งผลให้เกิดอันตรายถึงชีวิต การเจ็บป่วยอย่างรุนแรงหรือเรื้อรังแล้วแต่กรณี และยังเป็นสาเหตุหนึ่งของการเปลี่ยนแปลงทางพันธุกรรม (Genetic) และความพิกลพิการของร่างกายด้วย [1] การศึกษาวิจัยเพื่อให้ทราบถึงผลกระทบที่เกิดขึ้นจากสารเคมีแต่ละชนิดเป็นสิ่งจำเป็นที่จะช่วยในการกำหนดมาตรการป้องกัน หรือแนวทางในการรักษาเมื่อมีการสัมผัสสารเคมีดังกล่าว ซึ่งการศึกษาผลกระทบจากสารเคมีที่มีต่อการเปลี่ยนแปลงทางพันธุกรรมกำลังเป็นแนวทางการวิจัยที่ได้รับความสนใจเป็นอย่างมากในปัจจุบัน [1-4] ทั้งนี้เพราะจะทำให้สามารถทราบถึงต้นตอของความผิดปกติที่เกิดขึ้นในร่างกายจากการสัมผัสกับสารเคมี วิธีการที่สามารถใช้ทำการวิเคราะห์ผลกระทบของสารเคมีลึกลงไปในระดับพันธุกรรมมีด้วยกันหลายวิธี วิธีหนึ่งคือการใช้เทคนิคไมโครอาร์เรย์ (Microarray) ซึ่งเป็นเทคนิควิเคราะห์ถึงผลกระทบจากสารเคมีในระดับพันธุกรรมด้วยการตรวจวัดระดับการแสดงออกของยีน (Gene Expression) โดยใช้หลักการตรวจสอบการทำงานของยีนซึ่งขึ้นอยู่กับปริมาณโปรตีนที่ยีนแต่ละตัวสามารถแปลงได้ หรือที่เรียกว่าค่าการแสดงออกของยีน ค่าดังกล่าวเป็นข้อมูลสำคัญที่ช่วยให้เข้าใจลักษณะการทำงานต่าง ๆ ของยีนว่าสารเคมีนั้น ๆ ส่งผลต่อยีนกลุ่มไหนในร่างกาย และเป็นแบบยับยั้งการทำงานของยีน หรือกระตุ้นการ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ทำงานของยีน โดยการได้มาแต่ละหน่วยตัวอย่างจะอาศัยการทำการทดลองด้วยแผ่นสไลด์ (Array Slide) ซึ่งในแต่ละแผ่นสไลด์ จะบรรจุไปด้วยยีนจำนวนมาก (จำนวนยีนของสิ่งมีชีวิต มีประมาณ 20,000 ถึง 25,000 ยีน [5]) ภายหลังใส่กรรมวิธี (Treatment) ลงไป สามารถใช้เครื่องมือในการวัดระดับการแสดงออกของยีนแต่ละตัวได้ แต่เนื่องด้วยค่าใช้จ่ายในการทดลองโดยใช้เทคนิคไมโครอาร์เรย์แต่ละครั้งค่อนข้างสูงมาก ทำให้มักจะได้ตัวอย่างหรือแผ่นสไลด์เป็นจำนวนน้อยในการทดลอง

การวิเคราะห์ค่าการแสดงออกของยีนจากเทคนิคไมโครอาร์เรย์ มีเป้าหมายหนึ่งก็คือการวิเคราะห์เพื่อระบุยีนที่มีระดับการแสดงออกแตกต่าง (Differentially Genes Expression) อย่างมีนัยสำคัญ เมื่อได้รับกรรมวิธีที่แตกต่างกัน [6-7] โดยอาศัยวิธีการทดสอบทางสถิติ เปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยในแต่ละกลุ่มกรรมวิธี ในกรณีทำการทดสอบเพื่อศึกษาถึงยีนตัวหนึ่งว่ามีความแตกต่างในค่าการแสดงออกของยีนเมื่อได้รับกรรมวิธีที่แตกต่างกัน เราสามารถใช้วิธีการทดสอบทางสถิติ เช่นการทดสอบ *t (t-test)* การทดสอบ *Z (z-test)* หรือ การทดสอบ *F (F-test)* เพื่อทำการหาข้อสรุป แต่ภายใต้ลักษณะเฉพาะของข้อมูลที่ได้จากเทคนิคไมโครอาร์เรย์ ซึ่งมียีนที่ศึกษาเป็นจำนวนมาก ทำให้ต้องทำการทดสอบเพื่อเปรียบเทียบความแตกต่างค่าเฉลี่ยระดับการแสดงออกของยีนระหว่างกลุ่มกรรมวิธีมากครั้ง การทดสอบที่มีลักษณะเช่นนี้ เป็นรูปแบบการวิเคราะห์ที่เรียกว่า การทดสอบพหุคูณ (Multiple Testing) นอกจากนี้ด้วยการที่ข้อมูลจากเทคนิคไมโครอาร์เรย์มักจะมีตัวอย่างขนาดเล็ก ทำให้การใช้สถิติทดสอบที่กล่าวไว้ข้างต้น ได้รับผลกระทบ กล่าวคือ ให้อำนาจการทดสอบ (Power of the Test) ต่ำ จึงมีงานวิจัยหลายงาน [8-11] ที่พยายามพัฒนาเทคนิค และตัวสถิติที่สามารถแก้ปัญหาดังกล่าว หนึ่งในนั้นก็คือ เทคนิค Microarray Analysis of Variance (MAANOVA) [10-12] โดยใช้หลักการสร้างตัวประมาณความแปรปรวน (Variance Estimator) เพื่อทำการประมาณค่าความแปรปรวนและให้เป็นไปตามข้อตกลงเบื้องต้นว่าประชากรที่สุ่มตัวอย่างมาแต่ละชุด จะมีความแปรปรวนเท่ากับความแปรปรวนที่ประมาณได้จากตัวประมาณดังกล่าว โดยวิธีการนี้จะช่วยเพิ่มอำนาจการทดสอบได้ เทคนิค MAANOVA ประกอบด้วยสถิติทดสอบ 4 ชนิด คือ 1) วิธี *F-test* 2) วิธี Hybrid Statistic (F_2) ที่สร้างตัวสถิติภายใต้หลักการใช้ค่าความแปรปรวนที่หาได้จากค่าเฉลี่ย (Average) ระหว่างค่าความแปรปรวนของยีนแต่ละตัวและค่าความแปรปรวนรวม 3) วิธี Pooled-varianceF Statistic (F_3) โดยใช้การหาค่าความแปรปรวนรวม (Pooled Variance) และให้ยีนทุกตัวมีค่าความแปรปรวนเท่ากับค่าความแปรปรวนรวมดังกล่าว และ 4) วิธี F_s - test (F_s) ซึ่งเป็นวิธีล่าสุดที่ได้พัฒนาขึ้นมาโดยใช้หลักการประมาณค่าความแปรปรวนด้วยหลักของวิธี Shrinkage Estimator และไม่ได้มีการกำหนดเงื่อนไขของข้อตกลงเบื้องต้นเกี่ยวกับการเท่ากันของความแปรปรวน [10] ผลจากงานวิจัยที่ได้นำเทคนิค MAANOVA ไป

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ทดลองใช้เพื่อดูอำนาจการทดสอบในข้อมูลที่ทำการจำลอง (Simulation) ขึ้นมา 3 กรณี คือ กรณีที่ 1 เป็นการจำลองข้อมูลภายใต้รูปแบบคาโนนิคอล (Canonical Form) $\hat{\theta}_{g,t} = \theta_{g,t} + \varepsilon_{g,t}$ โดยที่ $\theta_{g,t}$ คือค่าการแสดงออกของยีนตัวที่ g ภายหลังได้รับกรรมวิธีที่ t และ $\varepsilon_{g,t}$ คือค่า Gene-Specific Residual Error โดย $\varepsilon_{g,t} \sim N(0, \sigma_g^2)$ ทำการทดสอบภายใต้ความหลากหลายของความแปรปรวน (Variance Heterogeneity) โดยการทดลองปรับเปลี่ยนค่า Coefficient of Variation (*CV*) และองศาความเป็นอิสระ (Degree of Freedom) ต่าง ๆ กัน ทำการทดสอบด้วยสถิติ 4 ตัว คือ F_1 (*F-test*), F_2, F_3 และ F_s ได้ผลว่า วิธี F_1 มีอำนาจการทดสอบ ที่ดีเฉพาะกรณีที่องศาอิสระสูง ($\nu > 6$) วิธี F_3 มีอำนาจการทดสอบที่ดีเฉพาะกรณีที่ค่า CV ต่ำ (CV<1) วิธี F_2 ให้ผลคล้ายคลึงกับวิธี F_3 แต่มีความแกร่ง (Robust) มากกว่า ค่าอำนาจการทดสอบทั้งสองวิธีนี้จะยิ่งลดลงเมื่อค่า CV เพิ่มขึ้น (แปรผกผันกับค่า CV) ในขณะที่วิธี F_s เปรียบเทียบกับทุกวิธีแล้ว มีความแกร่งมากกว่า และในเกือบทุกสถานการณ์ จะให้อำนาจการทดสอบที่ดีที่สุด โดยเฉพาะกรณีที่องศาอิสระต่ำ กรณีที่ 2 เป็นการจำลองข้อมูลภายใต้ลักษณะของข้อมูลที่ได้จากการทดลองจริง ซึ่งเป็นการทดลองกับยีนมะเร็งรังไข่ 3 ชนิด เทคนิคไมโครอาร์เรย์ที่ใช้เป็นแบบสลับสี (Dye Swap) ทำการทดลองด้วยแผนการทดลองแบบ Loop Design ภายใต้ตัวแบบ $y_{ijk} = \mu + S_i + D_j + A_k + \varepsilon_{ijk}$ โดย μ คือค่าเฉลี่ยการแสดงออกของยีน (Gene Mean) S_i คือผลกระทบจากตัวอย่าง (Sample Effect) D_j คือผลกระทบจากการย้อมสี (Dye Effect) A_k คือผลกระทบจากแผ่นสไลด์ (Array Effect) และ ε_{ijk} คือค่าเรสซิดวล (Residual) ทำการจำลองข้อมูล 10 ชุด แต่ละชุดประกอบไปด้วยยีนที่ไม่แตกต่างในการแสดงออก 1,000 ยีน และ ยีนที่มีการแสดงออกแตกต่าง 1,000 ยีน โดยทำการทดลองปรับเปลี่ยนค่า *CV*=1.2, 1.8 และ 3.7 ผลที่ได้คือ วิธี F_s ให้ค่าอำนาจการทดสอบที่ดีกว่า วิธี F_1 และ F_s กรณีที่ 3 เป็นการจำลองข้อมูลภายใต้ลักษณะของข้อมูลที่ได้จากการทดลองจริง ซึ่งมีการทำซ้ำในหน่วยทดลอง โดยใช้แผนการทดลองแบบ Reference Design ภายใต้ตัวแบบ $y_{ijklh} = \mu + A_i + D_j + T_k + M_l + R_h + \varepsilon_{ijklh}$ โดย T_k คือผลกระทบจากกรรมวิธี (Treatment Effect) M_l คือผลกระทบจากหน่วยทดลอง (Mouse Effect) และ R_h คือผลกระทบที่เกิดจากตัวชี้วัดการเปรียบเทียบระหว่างตัวควบคุมกับหน่วยทดลอง (Indicator of Reference versus Tissue Sample) ทำการจำลองข้อมูล 10 ชุด แต่ละชุดประกอบไปด้วยยีนที่ไม่แตกต่างในการแสดงออก 1,000 ยีน และ ยีนที่มีการแสดงออกแตกต่าง 1,000 ยีน ได้ผลว่า วิธี F_s ให้อำนาจการทดสอบดีกว่า วิธี F_1 และเมื่อเปรียบเทียบกับทุกวิธีแล้ว จะยิ่งเห็นชัดว่าภายใต้สถานการณ์ที่หน่วยทดลองส่งผลกับตัวแบบ วิธี F_s-*test* จะให้ผลดีกว่าวิธีทดสอบอื่น ๆ [10-11]

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ในการศึกษาครั้งนี้สนใจศึกษาถึงผลกระทบของสารโลหะหนัก (Heavy Metals) และอนุภาคซิลเวอร์นาโน (Silver Nanoparticles) ในระดับพันธุกรรม โดยสารโลหะหนักที่ศึกษาคือ อาเซนิก (Arsenic) แคดเมียม (Cadmium) และ โครเมียม (Chromium) พบได้ในยาฆ่าแมลง ยาเบื่อหนู แบตเตอร์รี่ พลาสติกพีวีซี เป็นต้น เป็นสารเคมีที่องค์กร ATSDR (The Agency for Toxic Substances and Disease Registry) ซึ่งเป็นองค์กรที่ก่อตั้งขึ้นเพื่อศึกษาผลกระทบต่อสุขภาพและคุณภาพชีวิตที่ลดลงจากสารอันตรายต่าง ๆ ได้จัดไว้เป็นสารอันตรายอันดับต้น ๆ ที่เมื่อถูกดูดซึมเข้าสู่ร่างกายมนุษย์แล้ว จะมีการสะสมและตรงเข้าทำลายเนื้อเยื่อในอวัยวะภายในที่สำคัญ โดยการเปลี่ยนแปลงโครงสร้างและการทำงานของเซลล์ ทำความเสียหายต่อโครโมโซมซึ่งเป็นปัจจัยทางพันธุกรรม เป็นตัวชักนำให้เกิดมะเร็ง หรืออาจทำให้เซลล์ตาย ส่วนอนุภาคซิลเวอร์นาโน ซึ่งปัจจุบันพบได้มากขึ้นในผลิตภัณฑ์อุปโภคบริโภคต่าง ๆ เช่น การบรรจุหีบห่ออาหาร สิ่งทอที่ต้านทานการเกิดกลิ่น อุปกรณ์เครื่องใช้ในบ้าน และอุปกรณ์ทางการแพทย์ รวมถึงผ้าปิดแผล ในระยะเวลาไม่กี่เดือนมานี้เริ่มมีการตระหนักถึงความเสี่ยงของอนุภาคซิลเวอร์นาโนต่อการเป็นพิษต่อสุขภาพของมนุษย์และสิ่งแวดล้อม มีความเป็นไปได้ที่อนุภาคซิลเวอร์นาโนจะส่งผลกระทบต่อแบคทีเรียที่มีประโยชน์ในสิ่งแวดล้อม โดยเฉพาะอย่างยิ่งในดินและน้ำ [2] หรือทำให้แบคทีเรียที่มีโทษเกิดการต้านทานในการยับยั้งต่ออนุภาคซิลเวอร์นาโน [1] โดยในงานวิจัยครั้งนี้จะอาศัยข้อมูลไมโครอาร์เรย์ของงานวิจัยเรื่อง Analysis of Mixture Toxicity of Dual Combinations for Silver Nanoparticles and Heavy Metals using Toxicogenomics Approach [1] ที่ทำการทดลองในเซลล์มะเร็งตับด้วยกรรมวิธี 2 ประเภท คือ อนุภาคซิลเวอร์นาโน และ โลหะหนัก (คือ อาเซนิก แคดเมียม และโครเมียม) มาทำการวิเคราะห์ค้นหายีนที่มีการแสดงออกอย่างมีนัยสำคัญ โดยอาศัยวิธีการทางสถิติที่เหมาะสม ซึ่งพิจารณาจากการตรวจสอบตัวสถิติที่มีการพัฒนาขึ้นมาสำหรับข้อมูลไมโครอาร์เรย์โดยเฉพาะ คือ เทคนิค MAANOVA โดยเลือก สถิติทดสอบวิธี *F-test* (สำหรับการวิเคราะห์ความแปรปรวน) และวิธี F_s *- test* (F_s) ใช้วิธีการจำลองข้อมูล ภายใต้สถานการณ์ที่สอดคล้องกับข้อมูลไมโครอาร์เรย์ที่ใช้ในการศึกษาครั้งนี้เพื่อทำการตรวจสอบสถิติทดสอบที่เหมาะสมกับข้อมูลที่สุด

1.2 วัตถุประสงค์ของการศึกษา

1.2.1) เพื่อวิเคราะห์เปรียบเทียบสมรรถนะของตัวสถิติทดสอบ *F-test* และ F_s*-test* โดยการจำลองข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

1.2.2) เพื่อหากลุ่มยีนที่มีความแตกต่างในการแสดงออกของยีนอย่างมีนัยสำคัญทางสถิติภายหลังจากได้รับกรรมวิธี 2 ประเภท คือ อนุภาคซิลเวอร์นาโน และ โลหะหนัก (คือ อาเซนิก แคดเมียม และโครเมียม)

1.3 ประโยชน์ที่ได้รับจากการศึกษา

เมื่อทราบถึงกลุ่มยีนที่มีความแตกต่างการแสดงออกเมื่อได้รับสารโลหะหนักและอนุภาคนาโนซิลเวอร์ จะทำให้สามารถทำความเข้าใจกลไกการเปลี่ยนแปลงกระบวนการต่าง ๆ ที่เกิดขึ้นภายในเซลล์เมื่อเซลล์ได้รับสารพิษแบบผสมดียิ่งขึ้น และสามารถเป็นข้อมูลพื้นฐานเพื่อนำไปศึกษาการเปลี่ยนแปลงของเซลล์ในเชิงกลไกทางด้านชีววิทยาต่อไป

1.4 ขอบเขตการศึกษา

1.4.1 ข้อมูลที่ใช้ในการศึกษาครั้งนี้ เป็นข้อมูลทุติยภูมิ จากโครงการวิจัยเรื่อง Analysis of Mixture Toxicity of Dual Combinations for Silver Nanoparticles and Heavy Metals using Toxicogenomics Approach [1] ซึ่งได้ทำการทดลองด้วยเทคนิคไมโครอาร์เรย์ โดยใช้เซลล์มะเร็งตับเป็นเซลล์ทดลอง และใช้กรรมวิธีทั้งหมด 2 ประเภท คือ อนุภาคซิลเวอร์นาโน และโลหะหนัก (คือ อาเซนิก แคดเมียม และโครเมียม)

1.4.2 ตัวสถิติที่ใช้ในการทดสอบสมมติฐานของงานวิจัยครั้งนี้คือ

1. สถิติทดสอบ *F-test* [13]

เป็นตัวสถิติที่ใช้สำหรับเปรียบเทียบค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่มขึ้นไปพร้อม ๆ กันซึ่งเรียกว่าเป็นการวิเคราะห์ความแปรปรวน (ANOVA) โดยความแตกต่างระหว่างกลุ่มดังกล่าวจะถูกวัดในรูปของส่วนเบี่ยงเบนมาตรฐานหรือความแปรปรวน ซึ่งในที่นี้ก็คือค่าเฉลี่ยของความแปรผัน (Mean Squares) โดยที่ความแปรผันเกิดจากผลรวมยกกำลังสองของความแตกต่างระหว่างค่าแต่ละค่าของค่าเฉลี่ย การวิเคราะห์ความแปรปรวนจึงเป็นการเปรียบเทียบความแตกต่างระหว่างกลุ่มกับความแตกต่างภายในกลุ่มในกรณีที่ประชากรมีมากกว่า 2 กลุ่มในลักษณะของความแปรผัน โดยมีเงื่อนไขว่าข้อมูลที่ได้จากกลุ่มประชากรที่นำมาทดสอบความแตกต่างของค่าเฉลี่ยของข้อมูลของแต่ละกลุ่มจะต้องมีการแจกแจงแบบปกติเท่านั้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

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

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

โดยที่ SStr คือ $\sum_{j}^{k} n_{.j}\left(\bar{x}_{.j} - \bar{x}_{..}\right)^2$

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

โดยที่ SSE คือ $\sum_{i}^{n_j}\sum_{j}^{k}\left(x_{ij} - \bar{x}_{.j}\right)^2$

โดย n คือ จำนวนการทำซ้ำ (ขนาดการทดลอง) ในแต่ละกรรมวิธี

และ k คือ จำนวนกรรมวิธีที่ใช้ในการทดลอง

2. สถิติทดสอบ F_s-*test* [10]

ใช้หลักการประมาณค่าความแปรปรวนด้วยวิธี Shrinkage Estimator และไม่ได้มีการกำหนดเงื่อนไขของข้อตกลงเบื้องต้นเกี่ยวกับการเท่ากันของความแปรปรวน

$$F_s = \frac{MStr_g}{\tilde{\sigma}_g^2}$$

และ $\tilde{\sigma}_g^2 = \left(\prod_{g=1}^{G}\left(X_g/\upsilon\right)^{1/g}\right) B \times \exp\left[\left(1 - \frac{(G-3)V}{\sum\left(\ln X_g - \overline{\ln X_g}\right)^2}\right)_{+} \times \left(\ln X_g - \overline{\ln X_g}\right)\right]$

โดยที่ $\overline{\ln X_g} = \frac{1}{G}\sum \ln\left(X_g\right)$, $X_g = SSE$, υ คือ degree of freedom

G คือ จำนวนยีนทั้งหมด, V คือ Variance ของ ε_g', m คือ ค่าเฉลี่ยของ $\ln\frac{x_\upsilon^2}{\upsilon}$

และ $B = \exp(-m)$

ผู้วิจัยจะทำการเปรียบเทียบอำนาจการทดสอบของตัวสถิติ F-*test* และ F_s-*test* โดยใช้วิธีการจำลองข้อมูล (Simulation) เพื่อทำการตรวจสอบสถิติทดสอบที่เหมาะสมกับข้อมูลที่สุด โดยโปรแกรมทางสถิติที่ใช้ในงานวิจัยจะอาศัยโปรแกรมทางสถิติ R Programming Environments ในการประมวลผลข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved