

การทำนายรูปแบบการแปรผันทางพันธุกรรมของยีนที่มีผลต่อการตอบสนองต่อยา 51 ยีนใน
ประชากรไทยโดยใช้รหัสพันธุกรรมทั้งจีโนม



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวสารสนเทศศาสตร์และชีววิทยาเชิงคอมพิวเตอร์ (สหสาขาวิชา) สหสาขาวิชาชีวสารสนเทศ
ศาสตร์และชีววิทยาทางคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

GENETIC VARIATION PROFILING PREDICTION OF 51 PHARMACOGENES IN A THAI
POPULATION USING WHOLE GENOME SEQUENCING



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Bioinformatics and Computational Biology
Inter-Department of Bioinformatics and Computational Biology

GRADUATE SCHOOL

Chulalongkorn University

Academic Year 2021

Copyright of Chulalongkorn University

| | |
|---------------------------------|---|
| หัวข้อวิทยานิพนธ์ | การทำนายรูปแบบการแปรผันทางพันธุกรรมของยีนที่มีผลต่อการตอบสนองต่อยา 51 ยีนในประชากรไทยโดยใช้รหัสพันธุกรรมทั้งจีโนม |
| โดย | น.ส.ณัฐนิชา วันแก้ว |
| สาขาวิชา | ชีวสารสนเทศศาสตร์และชีววิทยาเชิงคอมพิวเตอร์ (สหสาขาวิชา) |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | ผู้ช่วยศาสตราจารย์ ดร.มนนันทน์ พงษ์พานิช |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม | ผู้ช่วยศาสตราจารย์ แพทย์หญิง ดร.ปาจรีย์ จรรย์วิไลสกุล ศาสตราจารย์ นายแพทย์วรศักดิ์ โชติเลอศักดิ์ |

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

| | |
|---|---------------------------------|
| | คณบดีบัณฑิตวิทยาลัย |
| (รองศาสตราจารย์ ดร.ยุทธนา ฉัพพรรณรัตน์) | |
| คณะกรรมการสอบวิทยานิพนธ์ | ประธานกรรมการ |
| | |
| (ศาสตราจารย์ ดร.ศุภจิตรา ชัชวาลย์) | อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก |
| | |
| (ผู้ช่วยศาสตราจารย์ ดร.มนนันทน์ พงษ์พานิช) | อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม |
| | |
| (ผู้ช่วยศาสตราจารย์ แพทย์หญิง ดร.ปาจรีย์ จรรย์วิไลสกุล) | อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม |
| | |
| (ศาสตราจารย์ นายแพทย์วรศักดิ์ โชติเลอศักดิ์) | กรรมการ |
| | |
| (อาจารย์ เกสัชกร ดร.ณัฐพล พรพุทธพงศ์) | กรรมการภายนอกมหาวิทยาลัย |
| | |
| (เกสัชกรหญิง ดร.จิตติมา พิริยะพงศา) | |

ณัฐณิชา วันแก้ว : การทำนายรูปแบบการแปรผันทางพันธุกรรมของยีนที่มีผลต่อการตอบสนองต่อยา 51 ยีนในประชากรไทยโดยใช้รหัสพันธุกรรมทั้งจีโนม. (GENETIC VARIATION PROFILING PREDICTION OF 51 PHARMACOGENES IN A THAI POPULATION USING WHOLE GENOME SEQUENCING) อ.ที่ปรึกษาหลัก : ผศ. ดร.มนันท์ พงษ์พานิช, อ.ที่ปรึกษาร่วม : ผศ. พญ. ดร. ปาจารย์ จรรย์วิไลสกุล,ศ. นพ.วรศักดิ์ โชติเลอศักดิ์

การแปรผันทางพันธุกรรมในยีนที่เกี่ยวข้องต่อการตอบสนองต่อยามีบทบาทสำคัญต่อความหลากหลายของการตอบสนองต่อยาที่แตกต่างกันในแต่ละบุคคล รูปแบบของการแปรผันทางพันธุกรรมเป็นที่รู้จักกันในชื่อของ star allele เป็นสิ่งที่ได้รับการศึกษาในหลาย ๆ ประชากร แต่อย่างไรก็ตามโพรไฟล์ของ star allele มีรูปแบบที่แตกต่างกันออกไปขึ้นอยู่กับกลุ่มประชากร ซึ่งในกลุ่มประชากรเอเชียยังมีการศึกษาไม่มากนัก การศึกษาของเราได้อธิบายความถี่แอลลีล ความถี่จีโนไทป์ ลักษณะของฟีโนไทป์ที่ถูกทำนายในยีนที่เกี่ยวข้องกับการตอบสนองต่อยา 51 ยีน ในคนไทยที่ไม่มีความเกี่ยวข้องกัน 171 คน และเปรียบเทียบความถี่ของการแปรผันทางพันธุกรรมที่มีผลกระทบอย่างสูงต่อการแสดงออกของยีนระหว่างประชากรไทยและประชากรชาติต่าง ๆ งานวิจัยนี้ใช้โปรแกรม Stargazer เพื่อทำนาย star allele และลักษณะของฟีโนไทป์โดยใช้ข้อมูลจากรหัสพันธุกรรมทั้งจีโนม จากผลการศึกษาแสดงให้เห็นว่าในคนหนึ่งคนมีอย่างน้อย 3 ยีนที่มีฟีโนไทป์แบบที่ไม่ใช่ฟีโนไทป์ปกติ เราพบว่า 40 จาก 51 ยีนจะมีอย่างน้อยหนึ่งคนที่มีฟีโนไทป์แบบที่ไม่ใช่ฟีโนไทป์ปกติ และมีอย่างน้อย 25% ของกลุ่มตัวอย่างที่มีฟีโนไทป์แบบที่ไม่ใช่ฟีโนไทป์ปกติใน 13 ยีน ได้แก่ *SLCO1B3* (97.08%), *CYP3A5* (88.3%), *CYP2C19* (60.82%), *CYP2A6* (60.2%), *SULT1A1* (56.14%), *G6PD* (54.39%), *CYP4B1* (50.00%), *CYP2D6* (48.65%), *CYP2F1* (46.41%), *NAT2* (40.35%), *SLCO2B1* (28.95%), *UGT1A1* (28.07%) และ *SLCO1B1* (26.79%) นอกจากนี้เรายังระบุการแปรผันทางพันธุกรรม 20 ตัวที่มีผลกระทบต่อการทำงานของยีนในระดับสูง ซึ่งยังไม่มีรายการงานในฐานข้อมูลใด ๆ ภูมิทัศน์ทางเภสัชพันธุศาสตร์ในคนไทยจากการศึกษานี้อาจมีส่วนช่วยให้คำแนะนำสำหรับนโยบายการทำจีโนไทป์ในประเทศไทย และประเทศในภูมิภาคเอเชียตะวันออกเฉียงใต้ ซึ่งเป็นการนำพาประเทศเข้าใกล้สู่การรักษาแบบจำเพาะรายบุคคลไปอีกขั้น

| | | |
|------------|------------------------------|----------------------------------|
| สาขาวิชา | ชีวสารสนเทศศาสตร์และชีววิทยา | ลายมือชื่อนิสิต |
| | เชิงคอมพิวเตอร์ (สหสาขาวิชา) | |
| ปีการศึกษา | 2564 | ลายมือชื่อ อ.ที่ปรึกษาหลัก |
| | | ลายมือชื่อ อ.ที่ปรึกษาร่วม |
| | | ลายมือชื่อ อ.ที่ปรึกษาร่วม |

6280144020 : MAJOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

KEYWORD: whole genome sequencing; pharmacogenes; star alleles; variants frequencies;
personalized medicine

Natnicha Wankaew : GENETIC VARIATION PROFILING PREDICTION OF
51 PHARMACOGENES IN A THAI POPULATION USING WHOLE GENOME SEQUENCING.
Advisor: Asst. Prof. MONNAT PONGPANICH, Ph.D. Co-advisor: Asst. Prof. PAJAREE
CHARIYAVILASKUL, Ph.D., Prof. VORASUK SHOTELERSUK

Genetic variations in pharmacogenes play an important role in differences of drug response in individuals. Patterns of these variants generally known as star alleles have been studied in many populations. However, the star allele profiles differ between ethnicities and Asian population are understudied. This report describes allele frequencies, genotype frequencies, phenotype prediction of 51 pharmacogenes in 171 unrelated Thai individuals and compares high-impact variants frequencies between Thai and other populations. We used stargazer to call star alleles and predict phenotype from whole genome sequencing data. Our results showed that one individual had at least three genes with altered phenotype. We found 40 of the 51 genes had at least one individual with altered phenotype. At least 25% of individuals had altered phenotype in thirteen genes: *SLCO1B3* (97.08%), *CYP3A5* (88.3%), *CYP2C19* (60.82%), *CYP2A6* (60.2%), *SULT1A1* (56.14%), *G6PD* (54.39%), *CYP4B1* (50.00%), *CYP2D6* (48.65%), *CYP2F1* (46.41%), *NAT2* (40.35%), *SLCO2B1* (28.95%), *UGT1A1* (28.07%), and *SLCO1B1* (26.79%). In addition, we identified twenty high impact variants that have not previously been reported. Pharmacogenetic landscape in Thai from this study may provide guidance for genotyping policy in Thailand and Southeast Asian countries moving the nation a step closer to personalized medicine

| | | |
|-----------------|---|------------------------------|
| Field of Study: | Bioinformatics and Computational Biology | Student's Signature |
| Academic Year: | 2021 | Advisor's Signature |
| | | Co-advisor's Signature |
| | | Co-advisor's Signature |

กิตติกรรมประกาศ

การศึกษาในครั้งนี้สำเร็จลงได้ด้วยความช่วยเหลือ แนะนำ ให้คำปรึกษา ให้กำลังใจ รวมทั้งตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ ที่เกิดขึ้นตลอดการวิจัย ด้วยความเอาใจใส่อย่างดียิ่งจาก ผู้ช่วยศาสตราจารย์ ดร. มนันท์ พงษ์พานิช, ผู้ช่วยศาสตราจารย์ แพทย์หญิง ดร.ปาจริย์ จรรย์วิลาศกุล และศาสตราจารย์ นายแพทย์วรศักดิ์ โชติเลอศักดิ์

ขอขอบพระคุณอาจารย์ ดร.จรีรัตน์ โพธิ์แก้ว และนางสาววรรณนา เชษฐเรืองชัย คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้ความสะดวกในการใช้เครื่อง High Performance Computing (HPC) และให้สิทธิ์ในการเข้าถึงข้อมูลรหัสพันธุกรรมทั้งจีโนม โปรแกรม และทรัพยากรการต่าง ๆ ที่มีประโยชน์ต่องานวิจัยในครั้งนี้

ขอขอบพระคุณนางสาวอัจฉิมา อัครพิทักษ์สกุล และนายเฉลิมพล ศรีจอมทอง คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ช่วยสืบค้น และตรวจสอบข้อมูลภูมิปัญญาของกลุ่มอาสาสมัครที่ใช้ในการศึกษาครั้งนี้

ขอขอบพระคุณนางสาวธารทิพย์ อริกุล และนายวรัตน์ เครือสุวรรณ จากหน่วยชีวสารสนเทศและการจัดการข้อมูลวิจัย คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล ที่ให้ความรู้เกี่ยวกับเทคโนโลยีการหาลำดับนิวคลีโอไทด์ (sequencing) และช่วยตรวจสอบการเขียนความเรียงให้มีความกระชับและถูกต้องตามหลักไวยากรณ์ภาษาไทย

นอกจากนี้ยังมีผู้ให้ความร่วมมือช่วยเหลืออีกหลายท่าน ซึ่งผู้เขียนไม่สามารถกล่าวนามในที่นี้ได้ทั้งหมด จึงขอขอบคุณทุกท่านเหล่านั้นไว้ ณ โอกาสนี้ด้วย

ณัฐนิชา วันแก้ว

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

| | หน้า |
|--|------|
| | ค |
| บทคัดย่อภาษาไทย..... | ค |
| | ง |
| บทคัดย่อภาษาอังกฤษ | ง |
| กิตติกรรมประกาศ | จ |
| สารบัญ..... | ฉ |
| สารบัญตาราง..... | ญ |
| สารบัญภาพ..... | ฎ |
| บทที่ 1..... | 1 |
| บทนำ | 1 |
| 1.1 ที่มาและความสำคัญของงานวิจัย | 1 |
| 1.2 วัตถุประสงค์การวิจัย..... | 2 |
| 1.3 ขอบเขตการวิจัย..... | 2 |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ..... | 3 |
| 1.5 นิยามศัพท์เฉพาะ | 3 |
| บทที่ 2..... | 6 |
| ทบทวนวรรณกรรม | 6 |
| 2.1 เทคโนโลยีการหาลำดับเบส (Sequencing technology)..... | 6 |
| 2.1.1 การหาลำดับเบสด้วยวิธี pyrosequencing..... | 7 |
| 2.1.2 การหาลำดับเบสด้วยวิธีอิลลูมินา (Illumina sequencing)..... | 10 |
| 2.1.3 การหาลำดับเบสด้วยวิธีไลเกชัน (Sequencing by ligation)..... | 12 |

| | |
|--|----|
| 2.2 การประยุกต์ใช้ข้อมูลจากเทคโนโลยีเอ็นจีเอส | 15 |
| 2.2.1 Whole Genome Sequencing (WGS)..... | 15 |
| 2.2.2 Whole Exome Sequencing (WES)..... | 15 |
| 2.2.3 Targeted panel sequencing..... | 15 |
| 2.2.4 Transcriptome sequencing (RNA-seq)..... | 16 |
| 2.2.5 Epigenomic sequencing | 16 |
| 2.2.6 Metagenome Sequencing..... | 16 |
| 2.3 การแปรผันทางพันธุกรรม (Genetic variation)..... | 16 |
| 2.3.1 การเปลี่ยนแปลงของนิวคลีโอไทด์เพียงตำแหน่งเดียว | 17 |
| 2.3.2 การแปรผันทางพันธุกรรมเชิงโครงสร้าง..... | 17 |
| 2.3.2.1 การแทรกและการลบของลำดับนิวคลีโอไทด์ขนาดเล็ก (Small Insertion deletion; indel)..... | 17 |
| 2.3.2.2 การแปรผันของจำนวนชุดของยีน (Copy number variant; CNV) | 18 |
| 2.3.2.3 การย้ายชิ้นส่วนของโครโมโซม (Translocation) | 18 |
| 2.4 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา (Pharmacogene) | 19 |
| 2.5 ฐานข้อมูลทางเภสัชพันธุศาสตร์ (Pharmacogenomics database)..... | 23 |
| 2.6 การวิเคราะห์ข้อมูลทางเภสัชพันธุศาสตร์ (Pharmacogenomics analysis)..... | 27 |
| 2.7 Star allele nomenclature | 30 |
| 2.8 การแสดงออกทางฟีโนไทป์ของยีนที่ข้องกับการตอบสนองต่อยา (metabolizing function) | 31 |
| 2.9 การศึกษาก่อนหน้านี้..... | 31 |
| บทที่ 3..... | 35 |
| วิธีการดำเนินงานวิจัย..... | 35 |
| 3.1 ตัวอย่างที่ใช้ในงานวิจัย | 35 |

| | |
|---|----|
| 3.2 คำแถลงจริยธรรม | 35 |
| 3.3 การคำนวณขอบของความผิดพลาด (Margin of error calculation) | 35 |
| 3.4 การวิเคราะห์ผล Star allele | 35 |
| 3.4.1 ขั้นตอนการวิเคราะห์ข้อมูล (pre-analysis)..... | 35 |
| 3.4.1.1 การตรวจสอบคุณภาพของข้อมูล (Data quality checking)..... | 36 |
| 3.4.1.2 การเรียกการแปรผันทางพันธุกรรม (Variant calling)..... | 42 |
| 3.4.1.3 การคำนวณความลึกของ read (GATK depth of coverage calculation)..... | 48 |
| 3.4.2 ขั้นตอนวิเคราะห์ข้อมูล (data analysis) | 49 |
| 3.4.2.1 การทำนาย Star allele | 49 |
| 3.4.2.2 การคำนวณหาความถี่แอลลีล (allele frequency calculation) | 51 |
| 3.4.2.3 การหาการแปรผันทางพันธุกรรมที่ก่อให้เกิดโรค (Pathogenic variants analysis) | 52 |
| บทที่ 4..... | 53 |
| ผลการวิจัย | 53 |
| 4.1 ตัวอย่างที่ใช้ในการวิจัย..... | 53 |
| 4.2 คุณภาพของลำดับนิวคลีโอไทด์ และค่าความลึกของ read (read depth) | 54 |
| 4.3 ยีนควบคุม..... | 55 |
| 4.4 การทำนายความถี่ของแอลลีล จีโนไทป์ และฟีโนไทป์ | 56 |
| 4.5 การเปรียบเทียบความถี่ของแอลลีลในยีนที่เกี่ยวข้องกับการตอบสนองต่อยา..... | 61 |
| บทที่ 5..... | 64 |
| อภิปรายผลการวิจัย..... | 64 |
| บทที่ 6..... | 72 |
| สรุปผลการวิจัย | 72 |
| บรรณานุกรม..... | 73 |

ประวัติผู้เขียน..... 97



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

| | หน้า |
|---|------|
| ตารางที่ 1 คำจำกัดความของคำอธิบายประกอบทางคลินิกระดับต่าง ๆ ตามระบบ CPIC..... | 20 |
| ตารางที่ 2 ประเภทของยีนที่เกี่ยวข้องกับการตอบสนองต่อยา | 22 |
| ตารางที่ 3 สรุปข้อมูลเกี่ยวกับฐานข้อมูลทางด้านเภสัชพันธุศาสตร์ | 23 |
| ตารางที่ 4 สรุปเครื่องมือและแผนงานชีวสารสนเทศต่าง ๆ สำหรับการวิเคราะห์ข้อมูลทางเภสัชพันธุศาสตร์..... | 27 |
| ตารางที่ 5 รายชื่อโรคและยาที่เกี่ยวข้องกับยีนที่มีผลต่อการตอบสนองต่อยา 26 ยีนที่จะทำการศึกษาเพิ่มเติม | 32 |
| ตารางที่ 6 ข้อมูลภูมิสำเนาของกลุ่มอาสาสมัครจำนวน 171 คน | 53 |
| ตารางที่ 7 ตารางแสดงผลกระทบระดับต่าง ๆ ของการแปรผันทางพันธุกรรมที่พบในกลุ่มตัวอย่าง 61 | |

สารบัญภาพ

| | หน้า |
|--|------|
| ภาพที่ 1 การหาลำดับด้วยวิธี pyrosequencing | 9 |
| ภาพที่ 2 ขั้นตอนการหาลำดับด้วยวิธีอิลลูมินา | 11 |
| ภาพที่ 3 การเตรียมดีเอ็นเอต้นแบบสำหรับการหาลำดับด้วยวิธีไลเกชัน | 13 |
| ภาพที่ 4 ตัวอย่างการตั้งชื่อตามระบบ star allele..... | 30 |
| ภาพที่ 5 แผนงานการทำ data quality checking และ sequence alignment | 36 |
| ภาพที่ 6 แผนงาน GATK workflow best practice..... | 42 |
| ภาพที่ 7 แผนภาพวิธีการทำงานของ Stargazer | 49 |
| ภาพที่ 8 ค่าเฉลี่ยความลึกของ read..... | 54 |
| ภาพที่ 9 เปรียบเทียบผลการเรียก star allele จากการใช้ยีนควบคุมทั้ง 3 ยีน | 55 |
| ภาพที่ 10 แผนภูมิแท่งแสดงยีนและจำนวนบุคคลที่มีพีนไทป์ที่ไม่ใช่พีนไทป์แบบปกติ | 57 |
| ภาพที่ 11 การเปรียบเทียบความถี่แอลลีลใน 35 การแปรผันทางพันธุกรรมที่ส่งผลกระทบต่อระดับสูง | 63 |

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของงานวิจัย

ความแตกต่างในการตอบสนองต่อยาในแต่ละบุคคล ส่วนหนึ่งนั้นมีผลมาจากการแปรผันทางพันธุกรรม สำหรับการรักษาด้วยยาชนิดเดียวกันเดียวกันบางคนอาจตอบสนองได้ดี ในขณะที่บางคนตอบสนองได้ไม่ดี หรือเกิดภาวะที่เป็นพิษต่อร่างกาย ยกตัวอย่างเช่น ในกรณีผู้ป่วยที่มี human leukocyte antigen class B (HLA-B)*15:02 ควรหลีกเลี่ยงการรับประทานยา carbamazepine (CBZ) และยาที่มีส่วนผสมของ CBZ เนื่องจากมีความเสี่ยงที่จะเป็นโรค Stevens-Johnson syndrome (SJS) และ toxic epidermal necrolysis (TEN) มากขึ้น (Fang et al., 2019; Lochareernkul et al., 2010) หรือ กรณีทารกแรกเกิดมีภาวะติดยาหลังการคลอด เนื่องจากยีน *CYP2D6* ของมารดาที่มีความไวต่อการตอบสนองต่อยาระงับปวดได้ในระดับดีมาก สามารถสร้างอนุพันธ์ย่อยของยาระงับปวด อย่างเช่น มอร์ฟีน (morphine) ได้ในปริมาณมาก ทำให้อนุพันธ์เหล่านี้ถูกส่งมายังทารก (Madadi et al., 2007) และสร้างผลข้างเคียงที่เป็นพิษรุนแรงต่อร่างกายของทารก จากตัวอย่างดังกล่าวชี้ให้เห็นว่า การแปรผันทางพันธุกรรม ในแต่ละบุคคลมีความจำเพาะอย่างมากต่อการตอบสนองต่อยา อีกทั้งข้อมูลการแปรผันทางพันธุกรรมนี้ยังสามารถใช้เป็นตัวบ่งชี้ถึงการใช้เพื่อให้เกิดประสิทธิผลทางการรักษาสูงสุด หรือ ใช้ประกอบการตัดสินใจการปรับเพิ่มหรือลดปริมาณยา หรือ ป้องกันภาวะเป็นพิษที่เกิดขึ้นจากการใช้ยา (Oates & Lopez, 2018)

งานวิจัยที่ศึกษาเกี่ยวกับ ยีนที่เกี่ยวข้องต่อการตอบสนองต่อยา ส่วนใหญ่มักจะศึกษายีนในกลุ่ม Cytochrome P450 (CYP) ซึ่งเป็นกลุ่มยีนหลักที่รับผิดชอบต่อการเกิดเมแทบอลิซึมของยาและสารต่างๆ โดยเฉพาะอย่างยิ่งการศึกษาในยีน *CYP2D6* และ *CYP2C19* พบว่ามากกว่า 20% ของยาที่มนุษย์รับประทานเข้าไป จะถูกยีนเหล่านี้เมแทบอลิซึมและสังเคราะห์สารต่าง ๆ ที่มีความสำคัญต่อร่างกาย เช่น คลอเรสเตอรอล และกรดไขมัน ประกอบกับความหลากหลายของการแปรผันทางพันธุกรรมที่มีความแตกต่างกันอย่างจำเพาะในแต่ละกลุ่มประชากร ดังนั้นการเลือกกลุ่มประชากรเพื่อใช้เป็นตัวอย่งสำหรับการวิจัยจึงมีความสำคัญอย่างมาก (Fricke-Galindo et al., 2016; Gaedigk et al., 2017; Petrović et al., 2020; Zhou et al., 2017) การศึกษายีนที่เกี่ยวข้องต่อการตอบสนองต่อยาในประชากรเอเชียยังมีไม่มากนัก ดังที่เห็นได้จากตัวอย่างของการศึกษาการแปรผันทางพันธุกรรมในยีนกลุ่ม CYP ในประชากรกว่า 65,945 คน ที่มีเชื้อชาติและภูมิภาคแตกต่างกัน มีเพียง 7.6% ของกลุ่มตัวอย่างเท่านั้นที่เป็นประชากรเอเชีย (Zhou et al., 2017) ในขณะที่อีกงานวิจัยหนึ่ง ของสหรัฐอเมริกาที่ทำการสำรวจการแปรผันทางพันธุกรรมในยีน *CYP2D6* มีเพียง 0.2% ของกลุ่มตัวอย่างเท่านั้นที่เป็นประชากรเอเชีย (Del Tredici et al., 2018) งานวิจัยทั้งสองนี้บ่งชี้ให้เห็นว่า ในหลาย ๆ งานวิจัย มีการศึกษาประชากรเอเชียเป็น

สัดส่วนที่น้อยมากเมื่อเทียบกับจำนวนประชากรในกลุ่มตัวอย่างทั้งหมด ดังนั้นข้อมูลเหล่านี้ อาจไม่ใช่ข้อมูลที่ครบถ้วนที่สุด สำหรับใช้อ้างอิงถึงรูปแบบการแปรผันทางพันธุกรรม ในยีนที่มีผลต่อการตอบสนองต่อยาในประชากรเอเชียหรือประชากรไทย

ในปัจจุบันเทคโนโลยีการหาลำดับนิวคลีโอไทด์ (sequencing) มีพัฒนาการมากขึ้น เทคโนโลยีการหาลำดับที่เป็นที่นิยมอย่าง High throughput sequencing ถูกนำมาใช้ในการรักษาระดับคลินิกมากขึ้น เนื่องจากให้ผลการวิเคราะห์จีโนมที่ถูกต้อง รวดเร็วและแม่นยำ อีกทั้งยังสามารถวิเคราะห์หลายยีนได้ในเวลาเดียวกัน ทำให้สามารถมองเห็นปัจจัยที่มีผลต่อการเกิดโรคได้ครอบคลุมขึ้น และยังเป็นการเพิ่มความน่าเชื่อถือให้กับการรักษา (Ingelman-Sundberg et al., 2018; Profaiser et al., 2020)

อย่างไรก็ตาม แม้ว่าในประเทศไทยจะมีการนำเทคโนโลยีนี้มาใช้ในระดับคลินิกมากขึ้น แต่งานวิจัยเกี่ยวกับการทำนายรูปแบบการแปรผันทางพันธุกรรมของยีนที่มีผลต่อการตอบสนองต่อยาในประชากรไทยกลับพบเพียงแค่นี้นึ่งงานวิจัย และเป็นงานวิจัยที่ศึกษารูปแบบการแปรผันและผลกระทบของลักษณะดังกล่าวใน 25 ยีนที่ทำการศึกษาในกลุ่มประชากรที่เป็นโรคไหลตาย (Buganda's syndrome) (Mauleekoonphairoj et al., 2020) ซึ่งโรคนี้เป็นโรคที่พบได้ทั่วไปในกลุ่มประชากรที่มีถิ่นกำเนิดเดิมอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย กลุ่มตัวอย่างเพียงกลุ่มตัวอย่างเดียวอาจไม่ครอบคลุมประชากรจากทุกภาคส่วนของประเทศ และอาจได้ข้อมูลการทำนายรูปแบบการแปรผันทางพันธุกรรมที่ไม่สามารถเป็นตัวแทนของประชากรไทยในองค์รวมทั้งหมด ดังนั้นทางผู้วิจัยจึงต้องการศึกษารูปแบบการแปรผันทางพันธุกรรมในยีนที่เกี่ยวข้องต่อการตอบสนองต่อยาใน 51 ยีน เพิ่มเติมจากงานวิจัยก่อนหน้านี้อีก 26 ยีน ในกลุ่มประชากรไทยที่ไม่เป็นโรค และมาจากทั่วทุกภาคส่วนของประเทศ เพื่อสร้างองค์ความรู้ให้ครอบคลุมและสามารถใช้อ้างอิงเกี่ยวกับความหลากหลายต่อการตอบสนองต่อยาที่พบในประชากรไทย

1.2 วัตถุประสงค์การวิจัย

- 1.2.1 เพื่อระบุรูปแบบการแปรผันทางพันธุกรรมในยีนที่มีผลต่อการตอบสนองต่อยา 51 ยีน ในประชากรไทย 171 คน ที่ไม่เป็นโรค และมีภูมิลำเนาจากทุกภูมิภาคทั่วประเทศไทย
- 1.2.2 เพื่อเปรียบเทียบความถี่แอลลีลระหว่างประชากรไทย และประชากรจากชาติอื่น ๆ
- 1.2.3 เพื่อค้นหารูปแบบการแปรผันทางพันธุกรรมแบบใหม่ที่มีความสำคัญและจำเพาะในประชากรไทย

1.3 ขอบเขตการวิจัย

- 1.3.1 ทำนายรูปแบบการแปรผันทางพันธุกรรมในยีนที่มีผลต่อการตอบสนองต่อยา 51 ยีน ได้แก่ *CACNA1S*, *CFTR*, *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP2A6*, *CYP2A13*, *CYP2B6*, *CYP2C8*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP2E1*, *CYP2F1*, *CYP2J2*, *CYP2R1*, *CYP2S1*, *CYP2W1*, *CYP3A4*, *CYP3A5*, *CYP3A7*, *CYP3A43*, *CYP4B1*, *CYP26A1*, *CYP4F2*, *CYP19A1*, *DPYD*,

G6PD, GSTM1, GSTP1, GSTT1, IFNL3, NAT1, NAT2, NUDT15, POR, RYR1, SLC15A2, SLC22A2, SLCO1B1, SLCO1B3, SLCO2B1, SULT1A1, TBXAS1, TPMT, UGT1A1, UGT1A4, UGT2B7, UGT2B15, UGT2B17 และ *VKORC1*

- 1.3.2 ตัวอย่างที่ใช้ในการวิจัยเป็นลำดับสารพันธุกรรมทั้งจีโนม (whole genome sequencing) ของชายหรือหญิงที่ไม่เป็นโรคจำนวน 171 คน ที่มีภูมิลำเนามาจาก ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคตะวันออก ภาคตะวันตกและภาคใต้ ของประเทศไทย โดยตัวอย่างดังกล่าวได้รับมาจากโรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย
- 1.3.3 ข้อมูลเกี่ยวกับการแปรผันทางพันธุกรรมอ้างอิงจาก Database of Genomic Structural Variation (dbVar) และ Database of Single Nucleotide Polymorphisms (dbSNP) อ้างอิงจาก NCBI (RRID:SCR_006472)
- 1.3.4 ความถี่แอลลีลของประชากรอื่น ๆ อ้างอิงจาก 1000 Genomes Project (1KGP) (RRID:SCR_006828)
- 1.3.5 ข้อมูลเกี่ยวกับยา, ยีนที่มีผลต่อการตอบสนองต่อยา และความรุนแรงของการแปรผันทางพันธุกรรม อ้างอิงจาก DrugBank (RRID:SCR_002700), PharmGKB (RRID:SCR_002689) และ Variant Effect Predictor (RRID:SCR_007931) ตามลำดับ

1.4 ประโยชน์ที่คาดว่าจะได้รับ

การศึกษาในครั้งนี้จะสามารถสร้างกระบวนการการศึกษาเพื่อทำนายรูปแบบการแปรผันทางพันธุกรรมใน 51 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา สำหรับประชากรไทยได้ และสามารถนำข้อมูลนี้อ้างอิงถึงผลกระทบจากความหลากหลายของการตอบสนองต่อยา เพื่อนำไปใช้เป็นข้อพึงระวังสำหรับการรักษา หรือ เพื่อหลีกเลี่ยงผลกระทบที่ไม่พึงประสงค์จากการใช้ยา ตลอดจนสามารถสร้างประโยชน์ให้กับการพัฒนายา การจ่ายยา และอาจเป็นข้อมูลประกอบสำหรับการกำหนดนโยบายการตรวจยีนที่เกี่ยวข้องกับการตอบสนองของยาในประเทศไทยได้ ซึ่งจะเป็นการเพิ่มประสิทธิภาพในการรักษาโรคในบุคคลให้ถูกต้องและแม่นยำมากยิ่งขึ้น ซึ่งองค์ความรู้นี้จะเป็นรากฐานของการพัฒนาคุณภาพของระบบสาธารณสุขของประเทศได้อย่างยั่งยืน

1.5 นิยามศัพท์เฉพาะ

- 1.5.1 Pharmacogene (PGx gene หรือ PGx) หมายถึง ยีนที่มีผลต่อการตอบสนองต่อยา
- 1.5.2 Genetic variation หรือ variant หมายถึง การแปรผันทางพันธุกรรม
- 1.5.3 Star allele nomenclature หมายถึง วิธีการกำหนดชื่อรูปแบบของการแปรผันทางพันธุกรรมที่เกิดขึ้นบนแอลลีลหนึ่ง ๆ โดยใช้เครื่องหมาย * (star) เป็นตัวคั่นระหว่างชื่อยีนและรูปแบบของการแปรผันทางพันธุกรรม เช่น *CYP2D6**2 เป็นต้น โดย *CYP2D6* หมายถึงชื่อยีน และ *2 หมายถึงชื่อของรูปแบบการ

แปรผันทางพันธุกรรม (star allele) ซึ่งในที่นี้จะใช้คำทับว่า star allele โดยแต่ละ star allele อาจมีการแปรผันทางพันธุกรรมได้มากกว่าหนึ่งประเภท (Sangkuhl et al., 2020)

- 1.5.4 Drug-Gene Interaction (DGI) หมายถึง ความสัมพันธ์กันระหว่างยาและการแปรผันทางพันธุกรรม
- 1.5.5 FASTQ file หมายถึง ไฟล์ข้อความที่บรรจุข้อมูลลำดับนิวคลีโอไทด์ ประกอบด้วยโครงสร้าง 4 บรรทัด ได้แก่
- บรรทัดที่ 1: จุดเริ่มต้นของข้อมูลที่ขึ้นต้นด้วย @
 - บรรทัดที่ 2: ลำดับเบสในนิวคลีโอไทด์
 - บรรทัดที่ 3: เครื่องหมาย + บ่งบอกการเว้นวรรค
 - บรรทัดที่ 4: คุณภาพของเบสในนิวคลีโอไทด์ที่อ่านค่าได้ (quality score)
- 1.5.6 FASTA file หมายถึง ไฟล์ข้อความที่อาจบรรจุข้อมูลของลำดับเบสในนิวคลีโอไทด์ หรือ ลำดับกรดอะมิโนก็ได้ โดยประกอบด้วยโครงสร้าง 2 บรรทัด ได้แก่
- บรรทัดที่ 1: จุดเริ่มต้นของข้อมูลที่ขึ้นต้นด้วย >
 - บรรทัดที่ 2: ลำดับเบสในนิวคลีโอไทด์ หรือ ลำดับกรดอะมิโน
- 1.5.7 Sequence หมายถึง ลำดับ ซึ่งในการศึกษาครั้งนี้จะหมายถึง ลำดับเบสในนิวคลีโอไทด์ (Deoxyribonucleic acid; DNA) และจะเรียกแทนว่าลำดับเบส
- 1.5.8 Sequence alignment หมายถึง การปรับแนวลำดับเบส
- 1.5.9 SAM file (Sequence Alignment/Map format) หมายถึง ไฟล์ที่บรรจุข้อมูลการปรับแนวลำดับเบส โดยระหว่างคอลัมน์ถูกคั่นด้วยเครื่องหมาย tab (tab-delimited text file)
- 1.5.10 BAM file (Binary Sequence Alignment/Map format: *.bam) หมายถึง ไบนารีไฟล์ (binary file) ของ SAM file
- 1.5.11 Read หมายถึง ลำดับเบสสายสั้น ซึ่งในที่นี้จะใช้คำทับศัพท์ว่า read
- 1.5.12 Read depth of coverage (จำนวนความลึกของ read) หมายถึง จำนวนครั้งที่เบสหนึ่ง ๆ ในจีโนมถูกอ่าน โดยถ้าค่าความลึกของ read ยิ่งมาก ก็แสดงว่าเรามีจำนวน read ที่ตำแหน่งนั้น ๆ มาก (McKenna et al., 2010)
- 1.5.13 GDF file (GATK depth of coverage file format) หมายถึง ไฟล์ข้อความที่บรรจุข้อมูลเกี่ยวกับความลึกของ read ในแต่ละตำแหน่ง (Lee et al., 2019(b); McKenna et al., 2010)
- 1.5.14 gVCF (genomic Variant Call Format file) และ VCF file (Variant Call Format file) หมายถึง ไฟล์ที่เก็บข้อมูลการแปรผันทางพันธุกรรมในแต่ละบุคคล
- 1.5.15 Markduplication หมายถึง ขั้นตอนในการระบุหรือตัดแท็ก read เดียวกันที่มีลำดับเบสเหมือนกัน
- 1.5.16 Base quality score หมายถึง คะแนนคุณภาพของเบสในนิวคลีโอไทด์นั้น ๆ โดยคะแนนนี้คิดมาจากความผิดพลาดในการอ่านต่อหนึ่งเบสในนิวคลีโอไทด์ที่ประมาณโดยเครื่องอ่านลำดับเบส (sequencer หรือ

- sequencing machine) ซึ่งจะเป็นค่าที่บอกระดับความเชื่อมั่นในการอ่านเบสในนิวคลีโอไทด์ต่อหนึ่งครั้ง (McKenna et al., 2010)
- 1.5.17 Base Quality Score Recalibration (BQSR) หมายถึง ขั้นตอนการคำนวณค่าคะแนนคุณภาพของเบสในนิวคลีโอไทด์ใหม่ โดยการนำค่าคะแนนคุณภาพของเบสในนิวคลีโอไทด์และค่าความผิดพลาด (error) มาทำ machine learning เพื่อสร้างเป็นแบบ (model) สำหรับคำนวณค่าคะแนนคุณภาพของเบสในนิวคลีโอไทด์ใหม่ (McKenna et al., 2010)
- 1.5.18 Haplotypecaller หมายถึงขั้นตอนในการเรียก (calling) การแปรผันทางพันธุกรรมที่เกิดขึ้นทั้งหมดของลำดับเบส (Van der Auwera et al., 2013)
- 1.5.19 Haplotype หมายถึง รูปแบบของการแปรผันทางพันธุกรรมที่พบบนแอลลีลหนึ่ง ๆ ที่มีแนวโน้มที่จะถูกถ่ายทอดทางพันธุกรรมจากรุ่นสู่รุ่น ซึ่งในที่นี้จะใช้คำทับศัพท์ว่า haplotype
- 1.5.20 Haplotype estimation หรือ haplotype phasing หรือ phasing หมายถึง กระบวนการทางคอมพิวเตอร์และสถิติที่มีส่วนช่วยในการทำนาย haplotype ที่สามารถถ่ายทอดทางพันธุกรรมจากรุ่นสู่รุ่นได้ (Browning & Browning, 2011)
- 1.5.21 Homozygous หมายถึง การมีลักษณะของยีนหนึ่ง ๆ ที่เหมือนกันบนคู่อัลลีลเดียวกัน
- 1.5.22 Heterozygous หมายถึง การมีลักษณะของยีนหนึ่ง ๆ ที่ต่างกับบนคู่อัลลีลเดียวกัน
- 1.5.23 QualByDepth (QD) หมายถึงค่าความเชื่อมั่นของการแปรผันทางพันธุกรรมที่ถูกทำให้เป็นมาตรฐานด้วยค่าความลึกของการแปรผันทางพันธุกรรมของแต่ละตัวอย่างที่ไม่ถูกรอง
- 1.5.24 FisherStrand หมายถึง การประมาณค่าความเอนเอียงของทิศทางของสาย DNA เช่น จากทิศ 3' ไป 5' (ทิศทางไปข้างหน้า) หรือ 5' ไป 3' (ทิศทางไปข้างหลัง) โดยใช้สถิติทดสอบแบบ Fisher's exact test (FS)
- 1.5.25 RMSMappingQuality หมายถึง ค่าเฉลี่ยกำลังสองของค่าคุณภาพการจัดตำแหน่งของ read ในตัวอย่างต่าง ๆ (MQ)
- 1.5.26 MappingQualityRankSumTest หมายถึง การทดสอบผลรวมของลำดับสำหรับคุณภาพการจัดตำแหน่งของ read อ้างอิง (reference read; REF read) และ read ที่แตกต่างจาก read อ้างอิง (alternative read; ALT read)
- 1.5.27 ReadPosRankSumTest หมายถึง การทดสอบผลรวมของลำดับสำหรับค่าตำแหน่งสัมพัทธ์ของแอลลีลอ้างอิง (REF) และแอลลีลที่แตกต่างกับแอลลีลอ้างอิง (ALT) ภายใน reads (ReadPosRankSum)

บทที่ 2

ทบทวนวรรณกรรม

2.1 เทคโนโลยีการหาลำดับเบส (Sequencing technology)

หลังจากการค้นพบโครงสร้างดีเอ็นเอในปี ค.ศ. 1953 โดย เจมส์ วัตสัน (James Watson) และฟรานซิส คริก (Francis Crick) (Watson & Crick, 1953) ทำให้เกิดความเข้าใจเกี่ยวกับพันธุศาสตร์และการเกิดโรคมามากขึ้น นำไปสู่การพัฒนาทางเทคโนโลยีและสิ่งประดิษฐ์มากขึ้นอย่างต่อเนื่อง เพื่อไขคำตอบเกี่ยวกับความซับซ้อนและความหลากหลายทางพันธุกรรมของสิ่งมีชีวิต และพัฒนาเทคโนโลยีการหาลำดับเบสให้มีความรวดเร็วและแม่นยำมากขึ้น จนนำไปสู่ความสำเร็จในการหาลำดับเบสของจีโนมของมนุษย์ (Nurk et al., 2022)

ปัจจุบันเทคโนโลยีการหาลำดับเบสมีความก้าวหน้าอย่างมาก มีการพัฒนาเทคนิคใหม่ ๆ ที่ทำให้เกิดความรวดเร็วและมีต้นทุนค่าใช้จ่ายที่ต่ำ แตกต่างจากการหาลำดับเบสโดยวิธีดั้งเดิมหรือการหาลำดับเบสโดยวิธีแซงเกอร์ (Sanger sequencing) ที่ถูกพัฒนาขึ้นในปี ค.ศ. 1977 โดยเฟรดอริก แซงเกอร์ ซึ่งถูกจัดว่าเป็นเทคโนโลยีการหาลำดับเบสที่หนึ่ง เทคโนโลยีการหาลำดับเบสนี้อาศัยหลักการของ dideoxynucleotide chain termination ด้วยการใช้นิวคลีโอไทด์ที่ติดหมู่ฟอสเฟต (dideoxy nucleotide triphosphates หรือ ddNTPs) เป็นตัวหยุดการสังเคราะห์สายดีเอ็นเอ กระบวนการนี้จะแบ่งปฏิกิริยาการสังเคราะห์ดีเอ็นเอออกเป็นสี่หลอด ในแต่ละปฏิกิริยาไพรเมอร์จะเข้าจับกับดีเอ็นเอต้นแบบและเอนไซม์ DNA polymerase จะทำหน้าที่สร้างสายดีเอ็นเอเส้นใหม่ขึ้น โดยใช้ deoxynucleotide triphosphates (dNTPs) และในแต่ละหลอดของปฏิกิริยาจะถูกเติมด้วย ddNTPs (ddATP, ddCTP, ddGTP, ddTTP) คนละชนิดกันกระบวนการการสังเคราะห์สายดีเอ็นเอจะยุติลงเมื่อมี ddNTP เข้ามาจับ สายดีเอ็นเอที่เป็นผลผลิตจากทั้งสี่ปฏิกิริยาจะถูกนำมารวมกัน และใช้สำหรับวิเคราะห์หาลำดับเบส วิธีนี้มีข้อจำกัดคือสามารถอ่านลำดับเบสได้ในช่วงสั้น ๆ (ประมาณ 100–1,000 คู่เบส) และใช้เวลานานในการหาลำดับ อีกทั้งเมื่อเสร็จสิ้นกระบวนการ ข้อมูลลำดับเบสที่ได้มีจำนวนน้อยและมีต้นทุนต่อกระบวนการสูง ทำให้ไม่คุ้มค่าสำหรับการหาลำดับเบสในสิ่งมีชีวิตที่มีจีโนมขนาดใหญ่ (Sanger et al., 1977) อาทิ จีโนมมนุษย์

ด้วยข้อจำกัดในการอ่านลำดับเบสโดยเทคโนโลยีการจัดลำดับเบสที่หนึ่งวิธีการหาลำดับแบบ shotgun ถูกพัฒนาขึ้นเพื่อลดข้อจำกัดของการอ่านลำดับเบสที่มีความยาวเกิน 1,000 คู่เบส โดยการตัดดีเอ็นเอเป็นชิ้นสั้น ๆ แล้วทำการจัดเรียงใหม่ให้กลับเป็นเส้นดีเอ็นเอที่สมบูรณ์ โดยใช้โปรแกรม sequence assembler วิธีนี้ให้ลำดับเบสที่มีขนาดยาวขึ้นเมื่อเทียบกับการจัดลำดับแบบแซงเกอร์ อีกทั้งข้อมูลที่ได้ยังสามารถใช้เพื่อทำนายโครงสร้างของยีนหรือตำแหน่งของยีนบนจีโนมได้ (Anderson, 1981) จากแนวคิดการหาลำดับเบสแบบ shotgun นำไปสู่การริเริ่ม

เทคโนโลยีใหม่สำหรับหาลำดับเบสแบบ high throughput ที่เรียกว่า Next generation sequencing หรือเอ็นจีเอส (NGS) ซึ่งในที่นี้หมายถึงเทคโนโลยีการจัดลำดับยุคที่สอง เทคโนโลยีเอ็นจีเอสอาศัยการอ่านดีเอ็นเอต้นแบบแบบสุ่มทั้งจีโนม โดยตัดชิ้นส่วนดีเอ็นเอเป็นชิ้นสั้น ๆ จากนั้นนำชิ้นส่วนดังกล่าวเชื่อมต่อเข้ากับอะแดปเตอร์ (adaptor) เพื่อสังเคราะห์ดีเอ็นเอสายใหม่ ด้วยวิธีนี้สายดีเอ็นเอปริมาณมากจะถูกสร้างขึ้นในเวลาเดียวกัน (massive parallel sequencing) แม้ว่าข้อมูลลำดับเบสที่ได้จากเทคโนโลยีเอ็นจีเอสจะมีความยาวประมาณ 150–300 คู่เบส แต่ทุกชุดข้อมูลที่ได้ต่างก็เป็นข้อมูลที่มีคุณภาพดีและมีความถูกต้องแม่นยำสูง (Goodwin et al., 2016) ซึ่งการพัฒนาอย่างต่อเนื่องของเทคโนโลยีเอ็นจีเอสทำให้การหาลำดับเบสสามารถทำได้รวดเร็ว และมีต้นทุนที่ถูกลง ซึ่งสามารถประยุกต์ใช้ในการรักษาระดับคลินิกได้มากขึ้น (Twesigomwe et al., 2021; Twist et al., 2016)

เทคโนโลยีเอ็นจีเอสที่นิยมใช้กันในปัจจุบัน เป็นที่รู้จักกันในชื่อเทคโนโลยีการหาลำดับเบสยุคที่สองหรือ Second generation sequencing ซึ่งมีด้วยกัน 3 รูปแบบหลัก ๆ คือ 1) การหาลำดับเบสด้วยวิธี pyrosequencing 2) การหาลำดับเบสด้วยวิธีอิลลูมินา (illumina sequencing) และ 3) การหาลำดับเบสด้วยวิธีไลเกชัน (ligation sequencing) ซึ่งแต่ละรูปแบบก็เหมาะต่อการนำไปใช้ในงานวิจัยที่แตกต่างกัน

2.1.1 การหาลำดับเบสด้วยวิธี pyrosequencing

เทคโนโลยีการวิเคราะห์ลำดับเบสด้วยเครื่อง 454/Roche genome sequencer หรือที่รู้จักกันในชื่อเทคนิคการหาลำดับเบสไพโร จัดเป็นรูปแบบแรกของวิธีการหาลำดับเบสยุคใหม่ การหาลำดับเบสไพโรเป็นการตรวจวัดสารประกอบไพโรฟอสเฟต (PPI) จากปฏิกิริยา chemiluminescence สัญญาณแสงจากเบสที่เกิดขึ้นระหว่างการหาลำดับเบสโดยการสังเคราะห์ (sequencing by synthesis) จะถูกตรวจจับโดยกล้องและถูกแปลงค่าเป็นลำดับเบส (Liu et al., 2015) ประกอบด้วย 3 ขั้นตอน ดังนี้

a) การเตรียมดีเอ็นเอต้นแบบ (DNA library preparation)

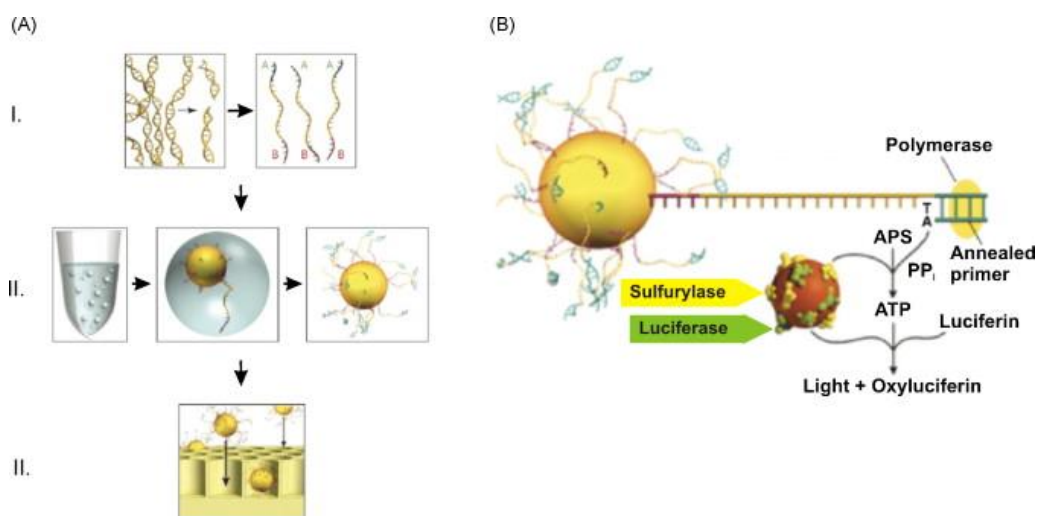
เริ่มต้นจากการตัดดีเอ็นเอสายคู่ให้มีขนาดประมาณ 400–600 คู่เบส จากนั้นเชื่อมต่อชิ้นดีเอ็นเอด้วยอะแดปเตอร์ที่แตกต่างกัน ที่ถูกเรียกว่า อะแดปเตอร์ A และ B ที่ปลายชิ้นดีเอ็นเอ โดยที่บริเวณส่วนปลายของอะแดปเตอร์ B ที่ปลาย 5' จะมีโมเลกุล biotin จะเข้าจับกับเม็ดบีดซึ่งถูกฉาบด้วยstreptavidin ส่วนอะแดปเตอร์ที่ไม่สามารถจับได้จะถูกกำจัดออก อะแดปเตอร์ที่ตรึงบนเม็ดบีดได้บางสายจะมีลำดับเบสโอลิโกนิวคลีโอไทด์ที่ไม่มีการ phosphorylate ทำให้มีช่องว่างในสายดีเอ็นเอคู่ นั้น ซึ่งเอนไซม์ที่มีคุณสมบัติเป็น strand displacing DNA polymerase จะเข้าซ่อมแซมและเติมเบสส่วนที่ขาดไปทำให้ได้เป็นดีเอ็นเอสายคู่ที่สมบูรณ์อีกครั้ง จากนั้นดีเอ็นเอสายคู่ที่ไม่ได้ตรึงกับเม็ดบีดจะหลุดออกและถูกแยกออกเป็นสายเดี่ยว โดยสุดท้ายจะมีเพียงสายดีเอ็นเอที่มีอะแดปเตอร์ A ติดที่ปลาย 5' และอะแดปเตอร์ B ติดที่ปลาย 3' เท่านั้นก่อนที่จะนำไปประเมินคุณภาพด้วยเครื่อง Agilent 2100 BioAnalyzer สำหรับสายดีเอ็นเอที่ผ่านการตรวจสอบคุณภาพจะถูกนำมาใช้เป็นดีเอ็นเอต้นแบบ ในขั้นตอนถัดไป (Liu et al., 2015) (ภาพที่ 1 (A))

b) การเพิ่มปริมาณดีเอ็นเอ (DNA amplification)

การหาลำดับเบสไฟโร จะใช้วิธีการเพิ่มปริมาณดีเอ็นเอบนเม็ดบีด (emulsion-based clonal amplification; emPCR) โดยนำสายดีเอ็นเอต้นแบบที่ผ่านการประเมินคุณภาพมาตรึงบนเม็ดบีด (immobilization) ที่มีไพรเมอร์ติดอยู่ จากนั้นจะเติมหยดน้ำมัน (emulsion oil) ที่มีสารสำหรับทำ polymerase chain reaction (PCR) ในที่นี้เราจะใช้คำทับศัพท์ว่า PCR จากนั้นอะแดปเตอร์บนสายดีเอ็นเอต้นแบบจะเข้ากับไพรเมอร์บนเม็ดบีดแล้วเริ่มต้นการเพิ่มปริมาณดีเอ็นเอ และเมื่อสิ้นสุดปฏิกิริยา PCR ในหนึ่งหยดน้ำมันจะมีดีเอ็นเอต้นแบบที่เหมือนกันเป็นล้านชุดติดอยู่ที่เม็ดบีด ซึ่งจะถูกนำไปหาลำดับเบส ในขั้นตอนถัดไป (Liu et al., 2015) (ภาพที่ 1(AII))

c) การหาลำดับเบสของดีเอ็นเอ (DNA sequencing)

เม็ดบีดแต่ละเม็ดที่ได้หลังจากเสร็จสิ้นปฏิกิริยา PCR จะถูกบรรจุลงใน PicoTiterPlate™ (PTP) ของเครื่องอ่านลำดับ 454/Roche เพื่อหาลำดับเบสโดยการสังเคราะห์ ซึ่งภายใน PTP ประกอบด้วยหลุม (well) จำนวนมาก โดยแต่ละหลุมจะบรรจุเม็ดบีดได้เพียงหนึ่งเม็ด (ภาพที่ 2.1(AIII)) ทำให้การอ่านลำดับเบสที่เกาะอยู่บนเม็ดบีดมีค่าเท่ากับข้อมูลลำดับเบสจากหนึ่งชุดเท่านั้น หลังจากทีบีดถูกบรรจุลงหลุมแล้ว เครื่อง 454/Roche จะสังเคราะห์สายดีเอ็นเอ โดยที่นิวคลีโอไทด์ทั้งสี่ชนิด คือ dATP, dCTP, dGTP และ dTTP จะถูกเติมเข้าไปทีละตัวเพื่อเข้าคู่กับสายดีเอ็นเอต้นแบบด้วยเอนไซม์ DNA polymerase ระหว่างการเกิดปฏิกิริยาจะมีการปลดปล่อยไพโรฟอสเฟตออกมา และเอนไซม์ sulfurylase จะเร่งปฏิกิริยาระหว่างไพโรฟอสเฟต กับ adenosine 5'-phosphosulfate (APS) ได้ ATP เป็นผลิตภัณฑ์ ซึ่งจะใช้ในปฏิกิริยาการเปลี่ยน luciferin ให้เป็น oxyluciferin โดยการเร่งปฏิกิริยาของเอนไซม์ luciferase ในปฏิกิริยาดังกล่าวจะเกิดการเรืองแสงขึ้น ซึ่งใช้เป็นสัญญาณในการอ่านลำดับเบส (Liu et al., 2015) (ภาพที่ 1(B))



ภาพที่ 1 การหาลำดับด้วยวิธี pyrosequencing (A) ขั้นตอนการหาลำดับด้วยวิธี pyrosequencing (I) การเตรียม ดีเอ็นเอต้นแบบ และการติดอะแดปเตอร์ A และ B ที่ปลายทั้งสองข้างของสายดีเอ็นเอ (II) การตรึงดีเอ็นเอบนเม็ด ปิด (immobilization) และการเพิ่มปริมาณดีเอ็นเอโดยใช้เม็ดปิดที่บรรจุที่ประกอบด้วยสารสำหรับการทำปฏิกิริยา พีซีอาร์อยู่ภายใน (droplet emulsion PCR) (III) การบรรจุเม็ดปิดลงในหลุม picotiter plate (PTP) (B) ปฏิกิริยา เคมีที่เกิดขึ้นในการหาลำดับเบสโดยการสังเคราะห์ (Ansorge, 2009)

2.1.2 การหาลำดับเบสด้วยวิธีอิลลูมินา (Illumina sequencing)

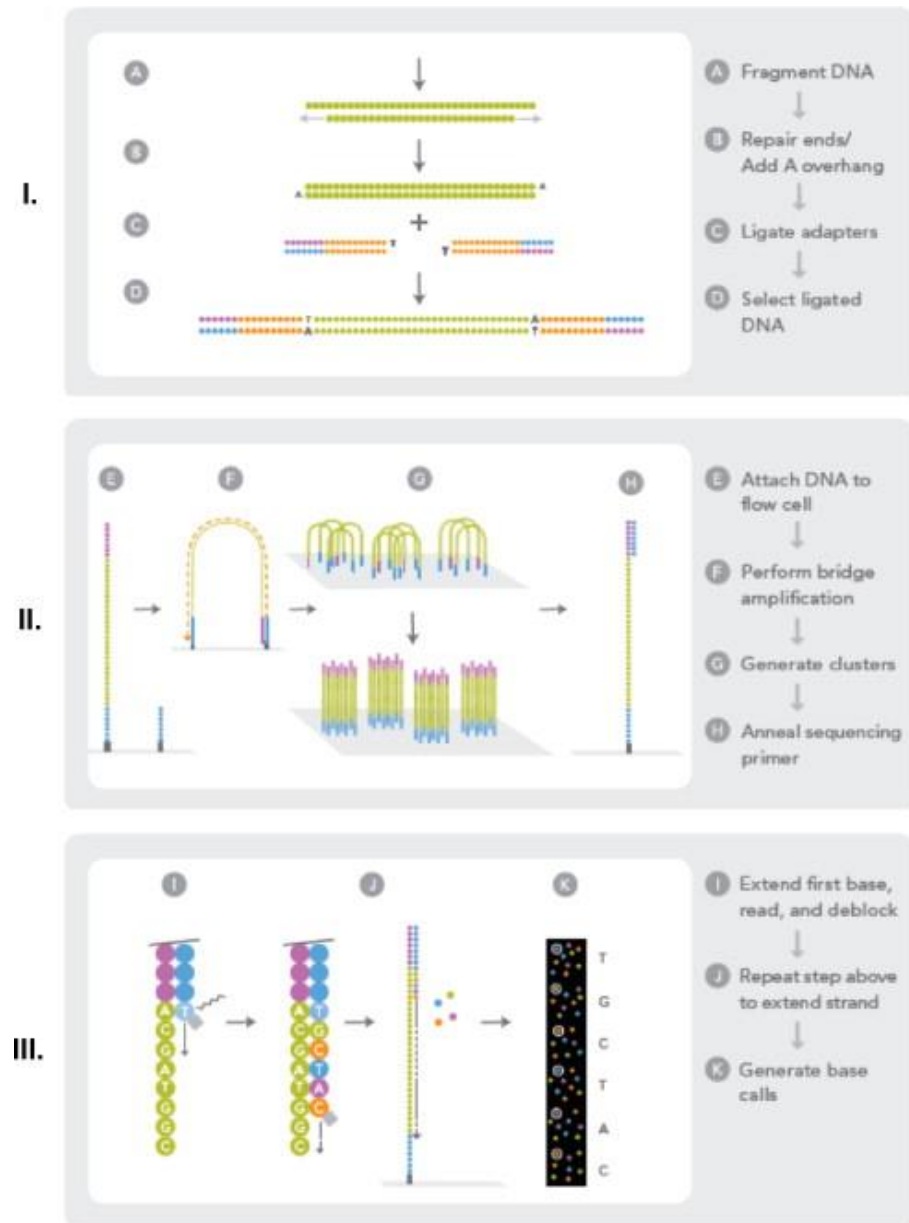
เทคโนโลยีการหาลำดับเบสด้วยวิธีอิลลูมินา อาศัยหลักการของการสังเคราะห์สายดีเอ็นเอ ร่วมกับการใช้นิวคลีโอไทด์ที่มีการดัดแปลงให้มีคุณสมบัติหยุดการสังเคราะห์สายดีเอ็นเอในการหาลำดับเบส โดยเริ่มจากการตัดสายดีเอ็นเอเป็นสายสั้น ๆ แล้วจึงหาลำดับเบสจากชิ้นดีเอ็นเอสายสั้น ๆ (DNA fragment) โดยวิธีการสังเคราะห์ ร่วมกับการใช้เทคนิคการติดฉลากสารเรืองแสงกับเบสทั้งสี่ชนิด (A, C, G และ T) ทำให้ได้ค่าสีที่แตกต่างกันในแต่ละชนิดของเบส (Fluorescent reversible terminator) เมื่อเริ่มต้นกระบวนการ สายดีเอ็นเอจะถูกตรึงเข้ากับอะแดปเตอร์ที่อยู่บนแผ่นสถานะของแข็ง (Solid surface) และทำการเพิ่มปริมาณดีเอ็นเอโดยการเชื่อมต่อบนสะพาน (Bridge amplification) (Ansoerge, 2009) โดยมีรายละเอียดดังนี้

a) การเตรียมดีเอ็นเอต้นแบบ (DNA template)

สายดีเอ็นเอจะถูกตัดเป็นสายสั้น ด้วย nebulizer ซึ่งความหนืดและความร้อนของสารละลายภายในเครื่อง nebulizer ทำให้สายดีเอ็นเอฉีกขาดและเกิดเป็นสายสั้น ๆ ที่มีขนาดสั้นกว่า 800 คู่เบส ส่วนปลายของสายดีเอ็นเอที่ได้รับความเสียหายจะถูกซ่อมแซมด้วยเอนไซม์ T4 DNA polymerase, Klenow enzyme และ T4 poly nucleotide kinase และเติมเบส A ที่บริเวณปลาย 3' เพื่อเพิ่มประสิทธิภาพในการเข้าจับกับอะแดปเตอร์ (ยาวประมาณ 66 คู่เบส) แล้วเชื่อมต่อกับเบส T อีกครั้งที่ปลายทั้งสองข้าง สายดีเอ็นเอที่มีความยาวประมาณ 150-200 คู่เบส ถูกคัดแยกแล้วนำไปเพิ่มปริมาณด้วยเทคนิค PCR (ภาพที่ 2 (I)) (Wilantho et al., 2012)

b) การเพิ่มปริมาณดีเอ็นเอ (Bridge amplification)

หลังจากแยกขนาดดีเอ็นเอแล้ว สายดีเอ็นเอจะถูกไหลลงบนกระจกสไลด์ (flow cell) ที่ฉาบด้วยกับอะแดปเตอร์คู่สม (complementary adapters) โดยที่อะแดปเตอร์ที่อยู่บนสายดีเอ็นเอต้นแบบ จะโค้งงอเข้าจับกับอะแดปเตอร์คู่สมมีลักษณะคล้ายสะพาน (Bridge) ซึ่งต่อมาจะถูกแยกให้เป็นดีเอ็นเอสายเดี่ยว เพื่อใช้เป็นต้นแบบสำหรับการเพิ่มปริมาณดีเอ็นเอด้วยเทคนิค PCR เมื่อเสร็จสิ้นกระบวนการ จะได้กลุ่มของดีเอ็นเอ (cluster) มากกว่า 50 ล้านกลุ่ม ซึ่งแต่ละกลุ่มจะมีดีเอ็นเอประมาณ 1,000 สาย และถูกนำไปหาลำดับเบสต่อไป (Ansoerge, 2009) (ภาพที่ 2 (II))



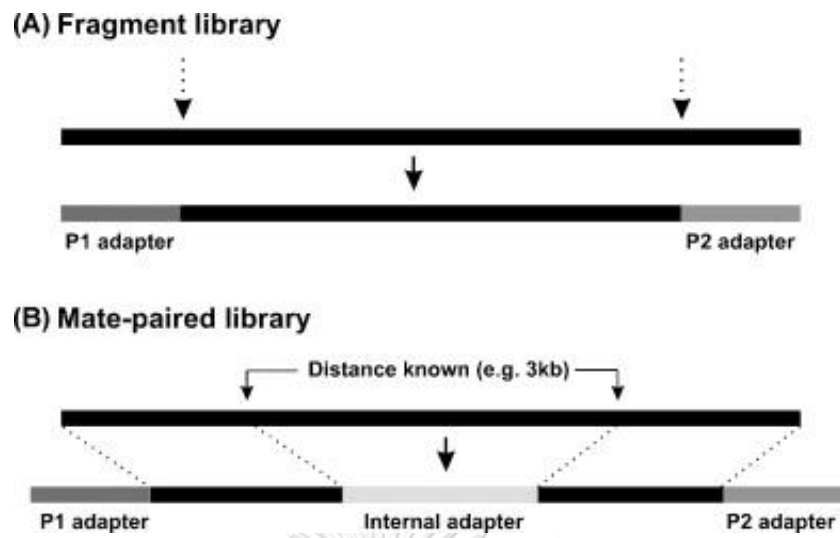
ภาพที่ 2 ขั้นตอนการหาลำดับด้วยวิธีอิลูมินา (I) สายดีเอ็นเอถูกตัดออกเป็นสายสั้น ๆ และเชื่อมด้วยอะแดปเตอร์ที่ปลายทั้งสองข้าง (adapter ligation) (II) สายดีเอ็นเอต้นแบบที่ติดอะแดปเตอร์จะถูกตรึงไว้กับแผ่นสถานะของแข็งเพื่อเพิ่มปริมาณดีเอ็นเอโดยการเชื่อมต่อแบบสะพาน (bridge amplification) เกิดเป็นกลุ่มของดีเอ็นเอสายคู่ (cluster) จากนั้นดีเอ็นเอสายคู่จะถูกแยกเป็นสายเดี่ยวและถูกติดด้วยไพรเมอร์ (sequencing primer annealing) (III) การเติมนิวคลีโอไทด์ชนิดที่ติดฉลากด้วยสารเรืองแสงเพื่อต่อสายดีเอ็นเอให้ยาวขึ้น โดยที่นิวคลีโอไทด์ที่ติดฉลากด้วยสารเรืองแสงจะมีคุณสมบัติขัดขวางปลาย 3'OH ของดีเอ็นเอทำให้หยุดการสังเคราะห์สายดีเอ็นเอ (Ansorge, 2009)

c) การหาลำดับเบส (DNA sequencing)

เมื่อได้ดีเอ็นเอต้นแบบแล้ว เครื่องอัลลูมินาจะสังเคราะห์สายดีเอ็นเอ โดยการเติมไพรเมอร์ และนิวคลีโอไทด์ทั้งสี่ชนิดที่ติดฉลากด้วยสารเรืองแสงที่มีสีต่างกัน รวมถึงเอนไซม์ DNA polymerase ในปฏิกิริยา การเชื่อมต่อของนิวคลีโอไทด์ที่ติดสารเรืองแสงนี้จะทำหน้าที่เป็นตัวขัดขวางปลาย 3'OH (reversible terminator) ของดีเอ็นเอ ทำให้เอนไซม์ DNA polymerase ไม่สามารถสังเคราะห์สายดีเอ็นเอต่อได้ และนิวคลีโอไทด์ที่ไม่สามารถเข้าคู่ได้จะถูกล้างออกไป เครื่องอัลลูมินาจะทำการบันทึกภาพการปลดปล่อยแสงจากสารเรืองแสงของนิวคลีโอไทด์ในแต่ละกลุ่มบนแผ่นกระจกสไลด์ จากนั้นมีการตัดหมู่เคมี ที่ป้องกันการต่อสายที่ปลาย 3' OH และสารเรืองแสงออก ทำให้นิวคลีโอไทด์กลับมาเป็นนิวคลีโอไทด์ปกติ หลังจากนั้นมีการเติมนิวคลีโอไทด์ชุดใหม่เพื่อเริ่มการหาลำดับเบสในรอบถัดไป กระบวนการนี้จะถูกทำซ้ำไปเรื่อย ๆ โดยแต่ละรอบ จะเป็นตัวกำหนดลำดับเบสที่อ่านได้ในแต่ละครั้ง การอ่านลำดับเบสด้วยวิธีนี้จะให้ read ที่มีความยาวประมาณ 100 คู่เบส และมีความถูกต้องถึง 99.5 % (Ansoerge, 2009) (ภาพที่ 2 (III))

2.1.3 การหาลำดับเบสด้วยวิธีไลเกชัน (Sequencing by ligation)

เทคโนโลยีการหาลำดับเบสด้วยวิธีไลเกชัน อาศัยกระบวนการเชื่อมต่อ (ligation) เพื่อการหาลำดับเบส โดยใช้ไพรเมอร์สองเบส (di-base probe) ที่ติดฉลากด้วยสารเรืองแสงการหาลำดับเบสด้วยวิธีนี้มีลักษณะคล้ายกับวิธีการหาลำดับเบสโดยวิธี pyrosequencing ตรงที่จะทำการตรึง DNA เข้ากับเม็ดปิดในสถานะที่เป็นน้ำมันเพื่อเพิ่มปริมาณเอ็นเอ (emulsion-based clonal amplification) (Ansoerge, 2009) โดยมีขั้นตอนการเตรียมดีเอ็นเอต้นแบบได้ 2 วิธีดังแสดงตามภาพที่ 3



ภาพที่ 3 การเตรียมดีเอ็นเอต้นแบบสำหรับการหาลำดับด้วยวิธีไลเกชัน (A) การเตรียมดีเอ็นเอต้นแบบโดยใช้วิธี sequencing fragment (B) การเตรียม DNA ต้นแบบโดยใช้วิธี mate-paired library (Ansorge, 2009)

a) การเตรียมดีเอ็นเอต้นแบบ (DNA template preparation)

สามารถเลือกทำได้ 2 วิธี ได้แก่

1. การเตรียมดีเอ็นเอ DNA ต้นแบบโดยวิธี Fragment library วิธีนี้จะเป็นการเชื่อมต่ออะแดปเตอร์ P1 และ P2 ที่มีความยาว 25 คู่เบส เข้าที่ปลาย 5' และ 3' ของสายดีเอ็นเอต้นแบบ (Ansorge, 2009) (ภาพที่ 3 (A))
2. การเตรียมดีเอ็นเอต้นแบบโดยวิธี Mate-paired library เป็นการใช้ internal adapter เพื่อเชื่อมสายดีเอ็นเอต้นแบบที่มาจากคนละตำแหน่งเข้าไว้ด้วยกัน จากนั้นจึงทำการติดอะแดปเตอร์ P1 และ P2 เข้าที่ปลาย 5' และ 3' ของสายดีเอ็นเอดังกล่าว (Ansorge, 2009) (ภาพที่ 3 (B))

b) การเพิ่มปริมาณดีเอ็นเอ (DNA amplification)

การเพิ่มปริมาณดีเอ็นเอสำหรับวิธีการหาลำดับด้วยวิธีไลเกชัน จะมีความคล้ายกับการเพิ่มปริมาณดีเอ็นเอในการหาลำดับแบบ pyrosequencing ที่จะทำการตรึงดีเอ็นเอบนเม็ดปิด ที่บรรจุสารสำหรับการทำ PCR ในสถานะที่เป็นน้ำมัน (emulsion-based clonal amplification) จากนั้นตรึงเม็ดปิดบนกระจกสไลด์เพื่อหาลำดับเบสต่อไป (Ansorge, 2009)

c) การหาลำดับเบส (Sequencing by ligation)

การหาลำดับเบสด้วยวิธีไลเกชัน อาศัยเครื่องหาลำดับ SOLiD/ABI เพื่ออ่านตำแหน่งเบสบนดีเอ็นเอ ต้นแบบทีละสองเบส (Di-base sequencing) ไพร์เมอร์ที่ใช้เพื่อเริ่มต้นปฏิกิริยาจะจับกับอะแดปเตอร์บนสายดีเอ็นเอต้นแบบ จากนั้นเครื่องหาลำดับจะทำการเชื่อม di-based probe ที่ติดฉลากสี (dye-labeled oligonucleotide) ขนาด 8-mer เข้าที่ปลายของโอลิโกนิวคลีโอไทด์ที่เข้าคู่กับดีเอ็นเอต้นแบบ โดยอาศัยเอนไซม์ ligase ในการเชื่อมต่อ หลังจากขั้นตอนการเชื่อมส่วนของ probe ที่ไม่สามารถเข้าคู่กับดีเอ็นเอต้นแบบได้อย่างจำเพาะจะถูกกำจัดออก จากนั้นบันทึกภาพของสารเรืองแสงซึ่งจะเรียกขั้นตอนเหล่านี้ว่า 1 ligation cycle ต่อมาจะเกิดปฏิกิริยาเคมีตัด probe ที่ตำแหน่งระหว่างเบสที่ห้าและหกออกและเข้าสู่การเชื่อมต่อ di-based probe ขนาด 8-mer ใหม่ใน ligation cycle ถัดไป ในการหาลำดับเบสจะต้องลดความยาวของไพร์เมอร์ลงไปที่ละหนึ่งเบส (n-1) ด้วย และภาพของสารเรืองแสงที่ได้ในแต่ละ ligation cycle จะถูกบันทึกเพื่อแปลงผลเป็นค่าลำดับเบส ซึ่งข้อมูลของลำดับเบสในแต่ละรอบของการลดความยาวของไพร์เมอร์ลงไปทีละหนึ่งเบสจะมีความเหลื่อมซ้อนกัน ทำให้ได้ข้อมูลที่มีคุณภาพดีมาก ๆ มีความถูกต้องมากกว่า 99.94% ขณะที่ความยาวของลำดับเบสน้อยกว่าเมื่อเทียบกับวิธีการอื่น ๆ อีกทั้งวิธีการเตรียมดีเอ็นเอต้นแบบยังมีความซับซ้อนและใช้เวลานาน (Ansorge, 2009)

2.2 การประยุกต์ใช้ข้อมูลจากเทคโนโลยีเอ็นจีเอส

เทคโนโลยีเอ็นจีเอสเป็นเทคโนโลยีที่มีบทบาทสำคัญในการไขความลับทางชีววิทยา โดยมีเทคนิคการหาลำดับเบสได้หลายรูปแบบดังนี้

2.2.1 Whole Genome Sequencing (WGS)

เป็นเทคนิคที่ใช้ในการศึกษาหาลำดับพันธุกรรมทั้งจีโนมของสิ่งมีชีวิตที่สนใจ วิธีนี้เป็นประโยชน์อย่างมากต่อการศึกษาความหลากหลายของโครงสร้างจีโนมของสิ่งมีชีวิต หรืองานวิจัยเกี่ยวกับการแปรผันทางพันธุกรรม (Ansorge, 2009) อาทิ การศึกษาเกี่ยวกับโรคต่าง ๆ ที่มีสาเหตุมาจากการเปลี่ยนแปลงทางพันธุกรรม (Hannan, 2018) การศึกษาการแปรผันทางพันธุกรรมที่ส่งผลกระทบต่อการทำงานของยีน (Mauleekoonphairoj et al., 2020) การศึกษาความสัมพันธ์ระหว่างลำดับเบสที่แตกต่างกันในกลุ่มประชากรหนึ่ง ๆ ต่อลักษณะทางฟีโนไทป์ในกลุ่มประชากร (Genome-wide association study; GWAS) (Stahl et al., 2019) หรือประยุกต์ใช้ในการศึกษาลำดับอนุกรมวิธานและกระบวนการเกิดสปีชีส์ใหม่ (speciation) (Chueca et al., 2021) เป็นต้น

2.2.2 Whole Exome Sequencing (WES)

เป็นเทคนิคที่ใช้ในการศึกษาหาลำดับพันธุกรรมในส่วนที่สามารถแปลรหัสเป็นโปรตีน หรือตำแหน่งที่เรียกว่า coding region วิธีนี้เหมาะที่จะใช้ศึกษาโรคที่มีสาเหตุทางพันธุกรรมจากตำแหน่ง coding region หรืองานวิจัยที่ศึกษาเกี่ยวกับการกลายพันธุ์ของยีน (Ansorge, 2009) อาทิ การศึกษาโรคทางพันธุกรรมที่เกิดจากความผิดปกติจากยีนเดี่ยว (single gene disorder) เช่น โรคซิสติกไฟโบรซิส (Cystic fibrosis; CF) (Yang et al., 2021) หรือ ใช้ตรวจหาสาเหตุทางพันธุกรรมในกลุ่มโรคที่หาสาเหตุการเกิดโรคได้ยาก เช่น การเจริญผิดปกติของโครงกระดูก (Skeletal dysplasia) (Sangsin et al., 2016) เป็นต้น

2.2.3 Targeted panel sequencing

เป็นการศึกษาพันธุกรรมพื้นฐานของมนุษย์ เพื่อให้เข้าใจถึงการเกิดโรค โดยจะทำการ screen หาลำดับนิวคลีโอไทด์ที่มีการกลายพันธุ์ไปจากเดิมด้วยกลุ่มยีน (gene sets) ที่มีข้อมูลเกี่ยวกับโรคอยู่ เพื่อให้ทราบถึงความหลากหลายของพันธุกรรม เช่น การศึกษาโรคมะเร็งทรวงอก (breast cancer) (Easton et al., 2015), การศึกษาโรคมะเร็งปอด (lung cancer) (E. H. Kim et al., 2014) และ การศึกษาโรคมะเร็งเม็ดเลือดขาว (leukemia cancer) (Engvall et al., 2020) เป็นต้น

2.2.4 Transcriptome sequencing (RNA-seq)

เป็นการศึกษาชุดคำสั่งของอาร์เอ็นเอ (RNA transcript) ได้แก่ mRNA และ non-coding RNA เพื่อศึกษาเกี่ยวกับโครงสร้างของยีนและปริมาณการแสดงออกของยีนในสภาวะหนึ่ง ๆ ที่สนใจ อาทิ การศึกษาปริมาณการแสดงออกของยีนที่แตกต่างกันภายใต้สภาวะเค็มและสภาวะปกติในข้าว (Sonsungsan et al., 2021) การศึกษารูปแบบการแสดงออกไอโซฟอร์มของยีน HLA (Boegel et al., 2012) รวมถึงสามารถใช้เพื่อศึกษาการรวมกันของยีน (gene fusion) ที่มีส่วนเกี่ยวข้องกับการเกิดเนื้องอกและมะเร็ง (Maher et al., 2009) เป็นต้น

2.2.5 Epigenomic sequencing

เป็นการศึกษาการเปลี่ยนแปลงที่เกี่ยวข้องกับการแสดงออกหรือกิจกรรมของยีนหรือพีโนไทป์ที่มีสาเหตุมาจากการเปลี่ยนแปลงของกลไกต่าง ๆ ในระดับที่เหนือกว่าระดับดีเอ็นเอ (Ansorge, 2009) ซึ่งส่วนใหญ่มีสาเหตุจากการเปลี่ยนแปลงของ DNA methylation หรือ DNA acetylation ตัวอย่างการประยุกต์ใช้เทคโนโลยีเอ็นจีเอสกับงานกลุ่มนี้ อาทิ การใช้เทคนิค Bisulfite sequencing เพื่อทำนายการเกิดมะเร็งเต้านมระยะแพร่กระจาย (metastatic breast cancer; MBC) ในผู้ป่วยมะเร็งเต้านมที่ได้รับการรักษาแบบเคมีบำบัด (chemotherapy) และเพื่อบ่งชี้ความเสี่ยงในการกลับมาเกิดโรคซ้ำอีกครั้ง (Maher et al., 2009) เป็นต้น

2.2.6 Metagenome Sequencing

เป็นการศึกษาจีโนมของสิ่งมีชีวิตทั้งหมดที่มีอยู่ในสิ่งแวดล้อมที่สนใจ เช่นกลุ่มจุลินทรีย์ในดิน จุลินทรีย์ในลำไส้ หรือจุลินทรีย์ในอุจจาระ อาทิ การศึกษาไมโครไบโอม (microbiome) ในกระเพาะและลำไส้ของคน (Kho & Lal, 2018) และสัตว์ (Bueno de Mesquita et al., 2021) เพื่อศึกษาความหลากหลายและความสัมพันธ์ของจุลินทรีย์ในแต่ละกลุ่มที่มีผลต่อสุขภาพและการเกิดโรค หรือ เป็นการศึกษาไมโครไบโอมในดินที่มีผลต่อการเจริญเติบโตและการอยู่รอดของพืช (Williams et al., 2015)

2.3 การแปรผันทางพันธุกรรม (Genetic variation)

การแปรผันทางพันธุกรรม คือ ความแตกต่างที่เกิดขึ้นในองค์ประกอบทางพันธุกรรม ซึ่งอาจเป็นความหลากหลายของลักษณะของสิ่งมีชีวิต หรือเป็นความแตกต่างของลำดับเบสระหว่างบุคคลหรือประชากรก็ได้ สาเหตุหลัก ๆ ของการแปรผันทางพันธุกรรมเกิดจากการกลายพันธุ์ (mutation) และการรวมกันใหม่ของยีน (gene recombination) แบ่งออกได้เป็น 2 ประเภทใหญ่ ๆ คือ การแปรผันของจีโนไทป์ (genotype) และการแปรผันของฟีโนไทป์ (phenotype) การแปรผันทั้งสองแบบต่างก็เป็นสิ่งที่เกี่ยวข้องซึ่งกันและกัน หากการเปลี่ยนแปลงที่เกิดขึ้นบนลำดับเบส ส่งผลต่อลำดับอะมิโนของโปรตีน ก็จะทำให้เกิดการแปรผันทางฟีโนไทป์ และส่งผลให้ยีนหรือเอ็มไซม์มีรูปร่าง ลักษณะและการทำงานที่เปลี่ยนแปลงไป หรือแม้แต่ลักษณะจีโนไทป์ของยีน ที่เกิดขึ้นจากการ

ปรับตัวเพื่อการดำรงชีวิตในสิ่งแวดล้อมหนึ่ง ๆ เมื่อเวลาผ่านไปมีการส่งต่อจีโนมไปข้างหน้าไปยังรุ่นถัดไปก็จะทำให้เกิดการแปรผันทางจีโนมได้เช่นเดียวกัน (Pavlopoulos et al., 2013)

ความหลากหลายในความแตกต่างของการแปรผันทางพันธุกรรมเหล่านี้ มีความสัมพันธ์ที่ซับซ้อนกับสุขภาพและโรคของมนุษย์ซึ่ง โดยทั่วไปเราแบ่งประเภทของการแปรผันทางพันธุกรรมเป็น 2 ประเภท ใหญ่ ๆ ได้แก่ การเปลี่ยนแปลงของนิวคลีโอไทด์เพียงตำแหน่งเดียว (single nucleotide variation; SNV) ในที่นี้เราจะใช้คำทับศัพท์ว่า SNV และการแปรผันทางพันธุกรรมเชิงโครงสร้าง (structural variation; SV)

2.3.1 การเปลี่ยนแปลงของนิวคลีโอไทด์เพียงตำแหน่งเดียว

หนึ่งในรูปแบบของการแปรผันทางพันธุกรรมที่มักพบมากที่สุด คือ single nucleotide polymorphisms (SNP) ในที่นี้เราจะใช้คำทับศัพท์ว่า SNP (สนิปส์) ซึ่งเป็นการเปลี่ยนแปลงของนิวคลีโอไทด์ใด ๆ ในจีโนมเพียงหนึ่งตำแหน่ง โดยการเปลี่ยนแปลงนี้จะถือว่าเป็น SNP ก็ต่อเมื่อมีการเปลี่ยนแปลงที่ตำแหน่งนั้น ๆ เกิดขึ้นอย่างจำเพาะเฉพาะในประชากรหนึ่ง ๆ และมีสัดส่วนของความถี่มากกว่า 1%

เนื่องจาก SNP เกิดขึ้นโดยเฉลี่ยทุก ๆ 300 เบสในนิวคลีโอไทด์ ดังนั้นในบุคคลหนึ่ง ๆ จะมี SNP ประมาณ 10 ล้านสนิปส์เกิดขึ้นในจีโนม ซึ่งสิ่งเหล่านี้ส่วนใหญ่ไม่เป็นอันตรายและไม่ก่อให้เกิดโรค อีกทั้งยังสามารถใช้เป็นตัวบ่งชี้ทางชีววิทยาในการตรวจสอบหาโรคที่เกี่ยวข้องกับโรคได้ อย่างไรก็ตาม SNP บางส่วนก็เป็นสาเหตุโดยตรงของการเกิดโรคเช่น ในที่พบส่วนใหญ่ในโรค Achondroplasia เป็นผลมาจากการเปลี่ยนแปลงของเบสเพียงหนึ่งตำแหน่งที่เฉพาะเจาะจงในยีน *FGFR3* ซึ่งนำไปสู่การเปลี่ยนแปลงของโปรตีนที่เกี่ยวข้องกับการเจริญเติบโตของกระดูก โดยยีนนี้ประกอบด้วย SNP ทั้งหมด 2,236 SNP และมีเพียงส่วนน้อยเท่านั้นที่ทำให้เกิดโรค (Boschiero et al., 2015)

2.3.2 การแปรผันทางพันธุกรรมเชิงโครงสร้าง

2.3.2.1 การแทรกและการลบของลำดับนิวคลีโอไทด์ขนาดเล็ก (Small Insertion deletion; indel)

การแทรกและการลบของลำดับนิวคลีโอไทด์ขนาดเล็ก ในที่นี้เราจะใช้คำทับศัพท์ว่า indel เป็นการแปรผันทางพันธุกรรมที่พบมากเป็นอันดับสองรองจากรองจาก SNP และเป็นการแปรผันเชิงโครงสร้างที่พบได้บ่อยที่สุด indel มีส่วนเกี่ยวข้องกับโรคและความผิดปกติทางพันธุกรรมได้หลากหลายประเภท ไม่เหมือนกับ SNP ซึ่งเกิดขึ้นจากการเปลี่ยนแปลงของเบสในนิวคลีโอไทด์เฉพาะตำแหน่ง โดยความผิดปกติทางพันธุกรรมดังกล่าวสามารถเกิดขึ้นได้จากหลายสาเหตุ เช่น การจำลองดีเอ็นเอแบบ slippage (replication slippage) การรวมตัวกันใหม่ของยีน (recombination) การเกิด crossing over ที่แบ่งชิ้นส่วนของโครโมโซมไม่เท่ากัน

การเกิดสำเนาเบสในนิวคลีโอไทด์แบบแทนเดม (tandem duplication) และจากการซ่อมแซม ดีเอ็นเอปลายคู่ที่ผิดปกติอย่างไม่สมบูรณ์ ด้วยเหตุนี้เองขนาดของ indel จึงมีความแตกต่างกัน ตั้งแต่ขนาดเล็กเพียง 1 เบสในนิวคลีโอไทด์ไปจนถึงหลายกิโลเบส (kb) (Boschiero et al., 2015)

2.3.2.2 การแปรผันของจำนวนชุดของยีน (Copy number variant; CNV)

การแปรผันประเภทนี้เป็นสิ่งที่ทำให้จีโนมของเรามีแตกต่างกันในเรื่องของจำนวนชุดของยีนอันเนื่องมาจากการทำซ้ำ (gene duplication/gene multiplication) หรือการขาดหายไปของยีน (gene deletion) จากการวิจัยทางวิทยาศาสตร์เมื่อเร็ว ๆ นี้บ่งชี้ให้เห็นว่ามากกว่า 9.5% ของจีโนมมนุษย์ อาจประกอบด้วย การแปรผันของจำนวนชุดของยีน (Zarrei et al., 2015) การมีอยู่ของการแปรผันทางพันธุกรรมประเภทนี้อาจเป็นสาเหตุของการเกิดโรค หรือ อาจไม่ส่งผลกระทบต่อ การเกิดโรคก็ได้ ตัวอย่างเช่น ในยีนอัลฟา-อะไมเลส 1 (AMY1A) ของมนุษย์ ซึ่งเป็นยีนที่เกี่ยวข้องกับการสร้างเอนไซม์ที่ย่อยสลายแป้งให้เป็นน้ำตาล ยีนนี้ถูกพบว่า สามารถมีได้หลายชุดในจีโนมของบุคคลหนึ่ง ๆ และมีรายงานว่าอาจพบจำนวนชุดของยีนได้มากถึง 14 ชุดในกลุ่มประชากรที่มีการบริโภคอาหารที่มีแป้งสูง (Zarrei et al., 2015)

2.3.2.3 การย้ายชิ้นส่วนของโครโมโซม (Translocation)

การย้ายชิ้นส่วนของโครโมโซม เป็นการแปรผันเชิงโครงสร้างชนิดหนึ่ง ที่เกิดจากการจัดเรียงใหม่ของชิ้นส่วนทางพันธุกรรมระหว่างแท่งโครโมโซมที่ต่างกัน หรือภายในแท่งโครโมโซมเดียวกัน

หนึ่งในรูปแบบที่เกิดขึ้นง่ายที่สุดของการย้ายชิ้นส่วนของโครโมโซม คือ การโยกย้ายตำแหน่งของยีนแบบสมดุล เป็นย้ายตำแหน่งของยีนที่มีการแลกเปลี่ยนชิ้นส่วนของโครโมโซมอย่างเท่าเทียมกัน ส่งผลให้ไม่เกิดการสูญเสียหรือได้รับข้อมูลทางพันธุกรรมที่มากกว่าปกติ การโยกย้ายที่สมดุลมักจะไม่ส่งผลร้ายต่อร่างกาย และผู้ที่เป็นพาหะของการแปรผันประเภทนี้ มักไม่ทราบว่าตนมีความผิดปกติชนิดดังกล่าวอยู่

ในทางตรงกันข้าม การโยกย้ายที่ไม่สมดุลอาจส่งผลให้เกิดภาวะแทรกซ้อนและปัญหาสุขภาพมากมาย รวมถึงอาจส่งผลให้เกิดข้อบกพร่องด้านพัฒนาการล่าช้าและเป็นสาเหตุของ มะเร็งหลาย ๆ ประเภท อย่างเช่น การจัดเรียงใหม่ของสารพันธุกรรมระหว่างบริเวณของโครโมโซม 9 และ 22 ซึ่งเป็นการแปรผันที่พบได้บ่อยในผู้ป่วยมะเร็งเม็ดเลือดขาวชนิดหนึ่ง เนื่องจากพบว่า การเคลื่อนย้ายของชิ้นส่วนของโครโมโซมนี้ทำให้เกิดการหลอมรวมของยีน 2 ยีน ที่ส่งผลให้มีการแบ่งตัวของเซลล์เพิ่มขึ้น (Perry et al., 2007)

2.4 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา (Pharmacogene)

มนุษย์เราแต่ละคนล้วนแต่มีการตอบสนองต่อยาชนิดเดียวกันที่แตกต่างกัน ความหลากหลายต่อการตอบสนองต่อยาชนิดเดียวกันเป็นความท้าทายที่สำคัญสำหรับการรักษาโรคให้เกิดประสิทธิภาพสูงสุด เภสัชพันธุศาสตร์ (Pharmacogenomics) เป็นศาสตร์ที่พยายามหาจุดเชื่อมโยงระหว่างการตอบสนองของยากับจีโนมของผู้ป่วยเพื่อการรักษาให้ดียิ่งขึ้น แต่เมื่อไม่กี่ปีมานี้งานวิจัยส่วนใหญ่ได้มุ่งเน้นและให้ความสำคัญเกี่ยวกับยีนที่มีบทบาทและความสำคัญทางเภสัชศาสตร์เป็นหลัก อีกทั้งยังพยายามศึกษาเกี่ยวกับความสัมพันธ์ระหว่างการแปรผันทางพันธุกรรม และการตอบสนองต่อยาที่พบในประชากรชาติต่าง ๆ เราจึงเรียก ยีนที่มีความสำคัญทางเภสัชศาสตร์และเกี่ยวข้องกับการตอบสนองต่อยา ว่า “pharmacogene” (Katara & Yadav, 2019)

ยีนที่เกี่ยวข้องกับการตอบสนองต่อยาถูกแบ่งประเภทตามหลักเภสัชจลนศาสตร์ซึ่งจะแบ่งได้เป็น 4 ประเภทตามกิจกรรมของยาที่ยีนนั้นเกี่ยวข้องด้วย ได้แก่ ยีนที่เกี่ยวข้องกับการดูดซึมของยา (absorption), ยีนที่เกี่ยวข้องกับการกระจายตัวของยา (distribution), ยีนที่เกี่ยวข้องกับการเมแทบอลิซึมของยา (metabolism) และยีนที่เกี่ยวข้องกับการกำจัดยา (excretion) (Doogue & Polasek, 2013) เนื่องจากมีการค้นพบยีนที่เกี่ยวข้องกับการตอบสนองต่อยามากขึ้น จึงมีการแบ่งประเภทอีกรูปแบบหนึ่งคือแบ่งตามระดับหลักฐานอ้างอิงทางคลินิก ซึ่งถูกใช้ในฐานข้อมูล Clinical Pharmacogenetics Implementation Consortium (CPIC) (Relling & Klein, 2011) โดยแสดงในตารางที่ 1

ตารางที่ 1 คำจำกัดความของคำอธิบายประกอบทางคลินิกในระดับต่าง ๆ ตามระบบ CPIC (Relling & Klein, 2011)

| ระดับ CPIC | คำอธิบายประกอบทางคลินิก | ระดับของหลักฐานอ้างอิงทางคลินิก | คำแนะนำในการสั่งยา |
|------------|---|---|---|
| A | ควรใช้ข้อมูลทางพันธุกรรมประกอบเพื่อการสั่งจ่ายยา | หลักฐานอ้างอิงทางคลินิกมีผลระดับสูงหรือปานกลางต่อการเปลี่ยนแปลงการสั่งจ่ายยา | แนะนำให้เปลี่ยนการสั่งจ่ายยา |
| A/B | การศึกษาเบื้องต้นที่บ่งชี้ว่ามีแนวโน้มว่าระดับ CPIC ขั้นสุดท้ายจะเป็น A หรือ B | จำเป็นต้องทบทวนหลักฐานทั้งหมดเพื่อประเมินระดับหลักฐาน แต่มีแนวโน้มว่าจะต้องดำเนินการได้จริง | ควรได้รับการทบทวนโดยกลุ่มผู้เชี่ยวชาญเพื่อกำหนดคำแนะนำการสั่งจ่ายยา |
| B | ข้อมูลทางพันธุกรรมสามารถใช้เพื่อเปลี่ยนแปลงการสั่งจ่ายยา เนื่องจากการรักษาหรือการให้ยาทางเลือกมีโอกาสสูงที่จะให้ประสิทธิภาพในการรักษาได้ปลอดภัยเท่ากับการให้ยาที่ไม่เกี่ยวข้องกัน | หลักฐานประกอบทางคลินิกมีน้อย และมีบางข้อมูลที่มีหลักฐานขัดแย้งกัน | แนะนำให้ใช้ยาทางเลือกอย่างน้อย 1 อย่าง (เปลี่ยนแปลงการสั่งจ่ายยา) |
| B/C | การศึกษาเบื้องต้นที่บ่งชี้ว่ามีแนวโน้มว่าระดับ CPIC ขั้นสุดท้ายจะเป็น B หรือ C | หลักฐานเกี่ยวกับการสั่งจ่ายยาที่เหมาะสมกับลักษณะทางพันธุกรรมไม่ชัดเจนหากไม่มีการตรวจสอบหลักฐานเพิ่มเติม | ควรได้รับการทบทวนโดยกลุ่มผู้เชี่ยวชาญเพื่อกำหนดคำแนะนำการสั่งจ่ายยา |
| C | มีการตีพิมพ์การศึกษาวิจัยในระดับต่าง ๆ ของหลักฐาน บางส่วนมีเหตุผลเชิงกลไก แต่ไม่แนะนำให้ดำเนินการใด ๆ เนื่องจาก | หลักฐานทางคลินิกมีความหลากหลาย | ไม่มีคำแนะนำสำหรับการสั่งจ่ายยา |

| | | | |
|-----|--|---------------------------------------|--|
| | <p>(ก) การให้ยาที่เหมาะสมกับลักษณะทางพันธุกรรม ไม่ได้ทำให้เกิดความแตกต่างที่น่าเชื่อ</p> <p>(ข) ทางเลือกการใช้ยาอื่น ยังไม่ชัดเจน และอาจมีประสิทธิภาบน้อยกว่า หรือมีพิษมากกว่า หรืออย่างอื่น หรือ ยังอยู่ในขั้นทดลอง</p> <p>(ค) เป็นการศึกษาที่มีการตีพิมพ์เพียงเล็กน้อยหรือหลักฐานที่อ่อนแอกว่าและมีการดำเนินการทางคลินิกที่ไม่ชัดเจน</p> | | |
| C/D | <p>การตรวจสอบเบื้องต้นบ่งชี้ว่ามีแนวโน้มว่าระดับ CPIC ขั้นสุดท้ายจะเป็น C หรือ D</p> | <p>หลักฐานทางคลินิกมีความหลากหลาย</p> | <p>ไม่มีคำแนะนำสำหรับการสั่งจ่ายยา</p> |
| D | <p>มีการศึกษาที่ตีพิมพ์เพียงไม่กี่ฉบับ การดำเนินการทางคลินิกไม่ชัดเจน มีข้อมูลเกี่ยวกับกลไกของยาเพียงเล็กน้อย</p> <p>หลักฐานส่วนใหญ่อ่อนหรือเป็นข้อมูลที่ขัดแย้งกันอย่างมาก หากเป็นยีนที่ไม่ได้อยู่ในการทดสอบทางเภสัชวิทยา ก็ไม่จำเป็นต้องให้ความสนใจ</p> | <p>หลักฐานทางคลินิกมีความหลากหลาย</p> | <p>ไม่มีคำแนะนำสำหรับการสั่งจ่ายยา</p> |

แต่สำหรับในการศึกษาครั้งนี้เราจะแบ่งยีนที่เกี่ยวข้องกับการตอบสนองต่อยาออกเป็น 4 ประเภท ตามตารางที่ 2 ตารางที่ 2 ประเภทของยีนที่เกี่ยวข้องกับการตอบสนองต่อยา

| ประเภทของยีน | ตัวอย่าง |
|--|--|
| ยีนที่เกี่ยวข้องกับการเมแทบอลิซึมของยา (metabolizing genes) ในที่นี้เราจะใช้คำทับศัพท์ว่า metabolizing genes ซึ่งแบ่งออกเป็น 2 กลุ่มได้แก่ Phase I metabolizing genes (Cytochrome P450 หรือ CYPs gene) และ Phase II metabolizing genes | <p>Phase I metabolizing genes: <i>CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2A13, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP2F1, CYP2J2, CYP2R1, CYP2S1, CYP2W1, CYP3A4, CYP3A5, CYP3A7, CYP3A43, CYP4B1, CYP26A1, CYP4F2, CYP19A1</i></p> <p>Phase II metabolizing genes: <i>GSTP1, GSTM1, GSTT1, IFNL3, NAT1, NAT2, SULT1A1, TPMT, UGT1A1, UGT1A4, UGT2B7, UGT2B15, UGT2B17</i></p> |
| ยีนที่เกี่ยวข้องกับการขนส่งยา (drug transporters) | <i>SLC15A2, SLC22A2, SLCO1B1, SLCO1B3, SLCO2B1</i> |
| ยีนที่กำหนดการสร้างโปรตีนบนผิวเซลล์ที่มีนิวเคลียสของร่างกาย หรือ MHC ยีน (Major histocompatibility complex genes; MHC genes) | <p>MHC class I: <i>HLA-A, HLA-B, HLA-C</i></p> <p>MHC classII: <i>HLA-DRB, HLA-DQ</i></p> |
| ยีนอื่น ๆ ที่เกี่ยวข้องกับกิจกรรมของยา (drug target genes) | <i>CACNA1S, CFTR, DPYD, G6PD, NUDT15, POR, TBXAS1, VKORC1</i> |

*งานวิจัยเล่มนี้เราไม่ได้ศึกษาการแปรผันทางพันธุกรรมใน MHC ยีน

2.5 ฐานข้อมูลทางเภสัชพันธุศาสตร์ (Pharmacogenomics database)

ข้อมูลจากการศึกษาเภสัชพันธุศาสตร์ในระดับประชากรได้ถูกรวบรวมและจัดการภายใต้กลุ่มและเครือข่ายขนาดใหญ่ เช่น Clinical Pharmacogenetic Consortium (CPIG), Pharmacogenomics Research Network (PGRN) หรือ South East Asian Pharmacogenomic Network (SEAPHARM) ซึ่งในปัจจุบัน ฐานข้อมูลจำนวนมากได้ถูกรวมเข้ากับฐานข้อมูลที่ใหญ่กว่า เช่น PharmGKB, Pharmacogene Variation (PharmVar), Pharmacogenomics Research Network (PGRN) และ Clinical Pharmacogenetics Implementation Consortium (CPIG) โดยฐานข้อมูลทั้งหมดนี้เป็นฐานข้อมูลที่เปิดให้ประชาชนสามารถเข้าใช้งานได้จากอินเทอร์เน็ตโดยไม่มีค่าใช้จ่าย นอกจากนี้เรายังได้ทำสรุปข้อมูลเกี่ยวกับฐานข้อมูลทางด้านเภสัชพันธุศาสตร์ (Tong et al., 2021) ดังนี้

ตารางที่ 3 สรุปข้อมูลเกี่ยวกับฐานข้อมูลทางด้านเภสัชพันธุศาสตร์ (Tong et al., 2021)

| ฐานข้อมูลและเครือข่ายเภสัชพันธุศาสตร์ | คำอธิบาย | URL |
|---------------------------------------|--|---|
| PharmGKB | PharmGKB เป็นฐานข้อมูลที่รวบรวมและตรวจสอบความรู้เกี่ยวกับการเปลี่ยนแปลงทางพันธุกรรมของมนุษย์ในยีนที่เกี่ยวกับการตอบสนองของยา ฐานข้อมูลนี้ให้ข้อมูลเกี่ยวกับคำอธิบายประกอบการใช้ยาจำนวน 706 รายการ และวิถิเมแทบอลิซึมที่ทราบเส้นทางแน่ชัดจำนวน 149 รายการ ปัจจุบันฐานข้อมูลนี้มีคำอธิบายประกอบการใช้ยาสำหรับการรักษาทางคลินิกจำนวน 155 รายการ, คำอธิบายประกอบสำหรับการติดตามยาจำนวน 753 รายการ ฐานข้อมูลนี้ยังมีคำอธิบายประกอบการรักษาทางคลินิก จำนวน 4,570 รายการและคำอธิบายเกี่ยวกับการแปรผันในยีนที่เกี่ยวข้องกับการตอบสนองต่อยา จำนวน 23,938 รายการ | https://www.pharmgkb.org/ |
| Clinical | เป็นองค์กรประเมินผลที่มีการค้นพบ | https://cpicpgx.org/ |

| | | |
|---|---|---|
| Pharmacogenetics Implementation Consortium (CPIC) | ทางคลินิกและห้องปฏิบัติการทางเภสัชพันธุศาสตร์ ล่าสุด CPIC ได้จัดทำคำอธิบายประกอบ 24 แนวทาง โดยครอบคลุมใน 20 ยีน และยา 62 ชนิด เพื่อที่จะแก้ไขปัญหาการใช้ยาแบบ "หนึ่งขนาดพอดีกับคนทุกคน" และเพื่อเพิ่มประสิทธิภาพยาในยาที่มีความแม่นยำต่อการรักษามากยิ่งขึ้น | |
| Dutch Pharmacogenetics Working Group (DPWG) | จุดมุ่งหมายของ DPWG คือการให้การทดสอบทางคลินิกในยีนที่เกี่ยวข้องกับการตอบสนองต่อยาทั่ว ๆ ไป เพื่อแปลผลจีโนไทป์เป็นฟีโนไทป์ ฐานข้อมูลนี้ให้แนวทางการรักษาทางคลินิกมากกว่า 90 ข้อคำอธิบาย ซึ่งคำอธิบายเหล่านี้ได้ผ่านการตรวจสอบโดย DPWG และจะถูกกำหนดระดับของสมมติฐานเพื่อสนับสนุนการดำเนินการรักษาทางคลินิก หรือ ใช้ในเภสัชจลนศาสตร์แบบชิลิโค | https://www.pharmgkb.org/page/dpwg |
| Ubiquitous Pharmacogenomics (U-PGx) | U-PGx เป็นฐานข้อมูลที่ก่อตั้งขึ้นโดยผู้เชี่ยวชาญชาวยุโรป เพื่อกำหนดระบบและวิธีการต่าง ๆ ทางเภสัชพันธุศาสตร์ ฐานข้อมูลนี้รวบรวมข้อมูลของยีนที่เกี่ยวข้องกับการตอบสนองต่อยา 13 ยีนที่มีการแปรผันมากกว่า 50 รูปแบบ ซึ่งสามารถช่วยในการศึกษาที่เกี่ยวข้องกับปัจจัยทางพันธุกรรมของผู้ป่วยที่มีอิทธิพลต่อการตอบสนองของยา โดยฐานข้อมูลนี้มีวัตถุประสงค์เพื่อปรับปรุงคุณภาพชีวิต และลดต้นทุนในการรักษาให้กับผู้ป่วย รวมถึงสร้างผลลัพธ์ทางการรักษาที่ดีขึ้นสำหรับผู้ป่วย | https://upgx.eu |
| Canadian Pharmacogenomics | CPNDS ก่อตั้งขึ้นในปี 2004 โดยมีเป้าหมายในการสร้างแนวทางการ | https://cpnds.ubc.ca/ |

| | | |
|--|---|---|
| Networks for Drug Safety (CPNDS) | รักษา (ปัจจุบันมี 8 แนวทาง; ปรับปรุงล่าสุด 07/03/2020) ที่เกี่ยวข้องกับ การตอบสนองของยีนที่เกี่ยวข้องกับ การตอบสนองต่อยา และการเกิดผลข้างเคียงจากการใช้ยา (Drug adverse reaction; ADR) ฐานข้อมูลนี้จะเรียนรู้และประเมินความเสี่ยงต่อ ปัจจัยทางพันธุกรรมเพื่อนำมาพัฒนา เครื่องมือสำหรับวิเคราะห์ยีนที่เกี่ยวข้องกับการตอบสนองต่อยาใน ระดับคลินิกสำหรับการสั่งจ่ายยา และเพิ่มประสิทธิภาพการใช้ยาใน รายบุคคล | |
| PharmVar | PharmVar เป็นฐานข้อมูลที่รวบรวม ข้อมูลเกี่ยวกับรูปแบบการแปรผันทาง พันธุกรรมที่สนับสนุน haplotype และแอลลีลที่ถูกละเมิด โดยจะให้ความ สันใจที่ยีนกลุ่ม P450 ของมนุษย์ และ <i>NUDT15</i> ฐานข้อมูลนี้ยังให้บริการ ข้อมูลจากฐานข้อมูล PharmGKB และ CPIC ด้วย | https://www.pharmvar.org/ |
| European Pharmacogenetics Implementation Consortium (EU-PIC) | ฐานข้อมูลนี้ให้ข้อมูลการดำเนินการ รักษาทางการแพทย์ของหลายประเทศ ในยุโรปเพื่อปรับปรุงการรักษาโดยใช้ แนวทางเภสัชพันธุศาสตร์สำหรับการ รักษาในระดับคลินิก. | https://www.eu-pic.net |
| Southeast Asian Pharmacogenomic Research Network (SEAPHARM) | SEAPHARM ก่อตั้งขึ้นเพื่อสนับสนุน การวิจัยทางด้านเภสัชพันธุศาสตร์ของ ประชากรชาติต่าง ๆ ในภูมิภาคเอเชีย ตะวันออกเฉียงใต้ ฐานข้อมูลนี้มี เป้าหมายสูงสุดเพื่อสร้างแนวทางการ รักษาโดยใช้ข้อมูลทางเภสัชพันธุ ศาสตร์ของประชากรในภูมิภาคเอเชีย ตะวันออกเฉียงใต้ | https://www.pharmagtc.org/seapharm/ |
| Database genomic | DGV ให้แหล่งเก็บข้อมูลเกี่ยวกับการ | https://www.ebi.ac.uk/dgva/ |

| | | |
|---------|--|---|
| variant | กระจายของตัวของการแปรผันทางพันธุกรรมเชิงโครงสร้างรูปแบบต่าง ๆ ที่พบบนจีโนมของสิ่งมีชีวิตทุกสายพันธุ์ | |
| dbSNP | ฐานข้อมูล dbSNP ให้ข้อมูลเกี่ยวกับ SNV, การเปลี่ยนแปลงของนิวคลีโอไทด์บน microsatellites และ indel พร้อมทั้งยังให้ข้อมูลเกี่ยวกับความถี่ แอลลีล, ลำดับโมเลกุลและรายละเอียดเกี่ยวกับยีน ซึ่งฐานข้อมูลนี้ได้รวมรวมการแปรผันทางพันธุกรรมทั้งแบบที่พบได้ทั่วไปและแบบที่เป็นกรณีศึกษาในการรักษาระดับคลินิก | https://www.ncbi.nlm.nih.gov/snp/ |

2.6 การวิเคราะห์ข้อมูลทางเภสัชพันธุศาสตร์ (Pharmacogenomics analysis)

จากความก้าวหน้าของเทคโนโลยีในการวิเคราะห์ข้อมูลขนาดใหญ่ เราจึงสามารถใช้ประโยชน์จากข้อมูลจีโนมเพื่อค้นหาต้นเหตุทางพันธุกรรมในยีนที่เกี่ยวข้องต่อการตอบสนองต่อยาที่มีรูปแบบจำเพาะในแต่ละบุคคลได้ เนื่องจากเทคโนโลยีการหาลำดับที่ถูกลงในปัจจุบัน ทำให้เราสามารถผลิตข้อมูลได้ปริมาณมาก จึงทำให้การวิเคราะห์ข้อมูลเป็นปัญหาคอขวด เนื่องจากเรามีข้อมูลและฐานข้อมูลจำนวนมาก แต่วิธีการวิเคราะห์ข้อมูลยังคงมีข้อจำกัด เราจึงได้รวบรวม และทำการสรุปเครื่องมือและแผนงานชีวสารสนเทศต่าง ๆ ที่ทันสมัยและเป็นที่ยอมรับในปัจจุบัน สำหรับการวิเคราะห์ข้อมูล (Tong et al., 2021) ดังนี้

ตารางที่ 4 สรุปเครื่องมือและแผนงานชีวสารสนเทศต่าง ๆ สำหรับการวิเคราะห์ข้อมูลทางเภสัชพันธุศาสตร์ (Tong et al., 2021)

| เครื่องมือวิเคราะห์การแปรผัน | คำอธิบาย | URL |
|------------------------------|---|---|
| Variant calling | | |
| GATK | เครื่องมือมาตรฐานสำหรับการเรียกการแปรผันทางพันธุกรรม ซึ่งสามารถใช้ได้ทั้งการหาลำดับแบบทั้งจีโนม, ทั้งเอ็กซีโซม, RNA-seq และการจัดลำดับแบบรู้เป้าหมาย (Targeted Sequencing) | https://gatk.broadststitute.org/hc/en-us |
| BWA | เครื่องมือมาตรฐานสำหรับการจัดตำแหน่งลำดับจีโนมขนาดสั้น เทียบกับลำดับอ้างอิงขนาดใหญ่อย่างเช่น ลำดับอ้างอิงของจีโนมมนุษย์ | http://bio-bwa.sourceforge.net/bwa.shtml |
| PGx allele prediction | | |
| Astrolabe | เครื่องมือสำหรับการเรียก star allele ซึ่งได้รับการพัฒนาครั้งแรกสำหรับการเรียก star allele ในยีน <i>CYP2D6</i> ต่อมาจึงได้รับการขยายขอบเขตไปยังยีน <i>CYP2C9</i> , <i>CYP2C19</i> และยีนอื่น ๆ | https://www.nature.com/articles/npjgenmed201639 |
| Stargazer | เครื่องมือสำหรับการเรียก star alleles (haplotypes) ในยีนที่เกี่ยวข้องกับการตอบสนองต่อยาโดยใช้ข้อมูลจาก NGS หรือ SNP array | https://stargazer.gs.washington.edu/stargazerweb/ |
| StellarPGX | เป็นเครื่องมือที่สร้างจาก nextflow ซึ่งเป็น | https://github.com/SBIMB/Stel |

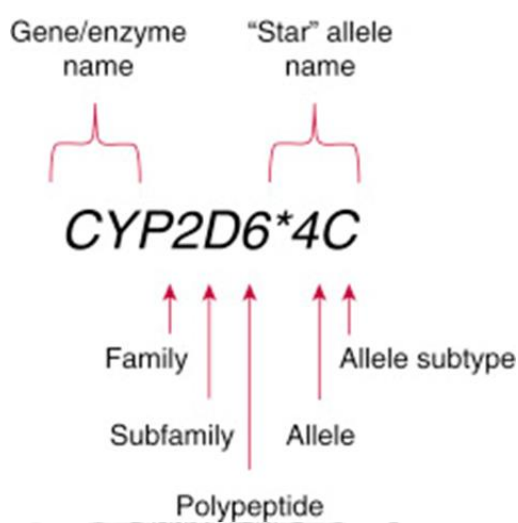
| | | |
|---------------------------|--|---|
| | ระบบการจัดการที่อำนวยความสะดวก สำหรับการทำงานแบบขนานกันในเวลา เดียวกัน การทำงานซ้ำในหลาย ๆ ตัวอย่าง และการทำงานกับตัวอย่างในปริมาณมาก อีกทั้ง ก็ยังทำให้ไปป์ไลน์สามารถใช้ได้ทุกที่ | larPGx |
| PharmVIP | เป็นเครื่องมือการวิเคราะห์แบบเว็บไซต์ ที่ใช้ สามารถวิเคราะห์และประมวลผลกระทบของ การแปรผันทางพันธุกรรม ซึ่งจะประกอบไป ด้วย 3 โมดูล คือ การทำนายรูปแบบการแปร ผันทางพันธุกรรมในยีนที่มีคำแนะนำการสั่ง จ่ายยาตามฐานข้อมูล CPIC, การทำนาย รูปแบบการแปรผันทางพันธุกรรมในแอลลีล HLA และการระบุและการทำนายผลกระทบ ที่เกิดจากการแปรผันทางพันธุกรรมในยีนที่ เกี่ยวข้องกับการตอบสนองต่อยา โดยผู้ใช้งาน สามารถเลือกได้มากกว่า 1 โมดูลสำหรับการ วิเคราะห์ได้ | https://pharmvip.nbt.or.th/ |
| Aldy | เป็นเครื่องมือที่ใช้ระบุจีโนไทป์ในยีนที่มีความ หลากหลายของรูปแบบการแปรผันทาง พันธุกรรมสูง โดย Aldy จะใช้ข้อมูลการ ทำนาย star allele จากฐานข้อมูลต่าง ๆ และรายงานผลการวิเคราะห์ของแต่ละ ตัวอย่าง (SNP, indel, การแปรผันทาง พันธุกรรมเชิงโครงสร้าง) ในรูปแบบของ haplotype ที่สามารถถ่ายทอดทาง พันธุกรรมจากรุ่นสู่รุ่นได้ (phased star- allele) | https://github.com/0xTCG/aldy |
| PharmCAT | เครื่องมือสำหรับการแปลจีโนไทป์เป็นฟีโน ไทป์โดยใช้ข้อมูลจีโนไทป์และการหาลำดับ ตามแนวทางของฐานข้อมูล CPIC | http://pharmcat.org |
| Variant annotation | | |
| VEP | เครื่องมือสำหรับการทำนายผลกระทบของ การแปรผันทางพันธุกรรม ไม่ว่าจะเป็น SNP, indel, CNV หรือ การแปรผันเชิงโครงสร้าง ที่ เกิดขึ้นบนจีโนม | https://asia.ensembl.org/info/docs/tools/vep/index.html |

| | | |
|----------------|---|---|
| Annovar | เครื่องมือสำหรับทำนายโปรตีนที่เปลี่ยนแปลงไปจากการเกิดการแปรผันทางพันธุกรรม เช่น SNV, CNV | http://wannovar.wglab.org |
| SnpEff/SnpSift | SnpEff เป็นเครื่องมือสำหรับให้คำอธิบาย (annotate) และทำนายผลกระทบที่เกิดขึ้นกับยีนและโปรตีนจากการแปรผันทางพันธุกรรม ในขณะที่ SnpSift เป็นเครื่องมือสำหรับให้ข้อมูลการแปรผันทางพันธุกรรมของจีโนมโดยใช้ข้อมูลอ้างอิงจากฐานข้อมูล โดยทั่วไปมักจะใช้ SnpSift หลังจากที่ใช้ SnpEff เพื่อหาการแปรผันทางพันธุกรรมที่มีนัยสำคัญทางสถิติ | https://pcingola.github.io/SnpEff/ |



2.7 Star allele nomenclature

ระบบการตั้งชื่อ star allele เป็นการอธิบายการแปรผันทางพันธุกรรมในรูปแบบของ haplotype ที่ระดับยีน รูปแบบของการแปรผันเหล่านี้มักเกี่ยวข้องกับระดับการแสดงออกของโปรตีน ในหนึ่ง haplotype สามารถมีรูปแบบการแปรผันทางพันธุกรรมได้มากกว่า 1 รูปแบบ/ชนิด ซึ่งอาจเป็น SNP, indel, CNV หรือการแปรผันทางพันธุกรรมเชิงโครงสร้างรูปแบบอื่น ๆ ก็ได้ การทราบรูปแบบการรวมกันของการแปรผันทางพันธุกรรมภายใน haplotype มีความสำคัญอย่างยิ่งในการศึกษาเภสัชจลนศาสตร์ของยา การตอบสนองต่อยา และปฏิกิริยาของยาที่ไม่พึงประสงค์ในรายบุคคล (Kalman et al., 2016)



ภาพที่ 4 ตัวอย่างการตั้งชื่อตามระบบ star allele (Lee et al., 2019(b))

2.8 การแสดงออกทางฟีโนไทป์ของยีนที่ข้องกับการตอบสนองต่อยา (metabolizing function)

เมื่อการทดสอบทางเภสัชพันธุศาสตร์ทางคลินิกเริ่มเป็นที่แพร่หลายขึ้น ระบบ star allele จึงเป็นสิ่งสำคัญที่จะอธิบายถึงลักษณะจีโนไทป์ของผู้ป่วยและคาดการณ์ลักษณะทางคลินิกของผู้ป่วยได้อย่างมีประสิทธิภาพ โดยทั่วไปการทดสอบยีนที่เกี่ยวข้องกับการตอบสนองต่อยามักรายงานผลการทดสอบเป็นคู่ของ haplotype ในที่นี้เราจะใช้คำทับศัพท์ว่า diplotype เนื่องจากยีนของมนุษย์มีจำนวนชุดอยู่สองชุด ยกเว้นยีนที่อยู่ในโครโมโซมเพศ (โครโมโซม X ในเพศหญิง และ Y ในเพศชาย) ผลรวมของการแสดงออกของ diplotype ส่งผลโดยตรงต่อการแสดงออกทางฟีโนไทป์ของยีนที่ข้องกับการตอบสนองต่อยา (Robarge et al., 2007) ซึ่งในการศึกษาคั้งนี้เราแบ่งการแสดงออกทางฟีโนไทป์ออกเป็น 6 ระดับตามความสามารถของเอนไซม์ได้แก่

- กลุ่มที่มีความสามารถในการย่อยสลายยาได้รวดเร็วมากที่สุด (ultrapid metabolizer)
- กลุ่มที่มีความสามารถในการย่อยสลายยาได้รวดเร็ว (rapid metabolizer)
- กลุ่มที่มีความสามารถในการย่อยสลายยาได้ในระดับปกติ (normal metabolizer)
- กลุ่มที่มีความสามารถในการย่อยสลายยาได้ปานกลาง (intermediate metabolizer)
- กลุ่มที่มีความสามารถในการย่อยสลายยาได้ช้า (slow metabolizer)
- กลุ่มที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย (poor metabolizer)

2.9 การศึกษาก่อนหน้านี้

การแปรผันทางพันธุกรรมมีบทบาทสำคัญอย่างมากต่อการตอบสนองต่อการใช้ยารายบุคคล และยังมี ความเชื่อมโยงกับยีนที่ทำงานเกี่ยวข้องกับการตอบสนองต่อยา ที่สามารถใช้เป็น biomarker สำหรับอธิบายถึงความ ไวต่อการตอบสนองต่อยาที่แตกต่างกันระหว่างการใช้ยา การปรับหาปริมาณการใช้ยาที่เหมาะสมเพื่อให้ได้การรักษา ที่มีประสิทธิภาพสูงสุด และยังใช้ประกอบการพิจารณาหาการรักษาทางเลือกที่เหมาะสมและจำเพาะสำหรับบุคคล นั้น ๆ (Oates & Lopez, 2018) การตอบสนองต่อยาอาจมีตั้งแต่ตอบสนองได้ดีมากไปจนถึงสามารถความเป็นพิษ ที่คุกคามชีวิตมนุษย์ได้ ตัวอย่างเช่น การใช้ยา fluvoxamine ในผู้ป่วยที่เป็นโรคซึมเศร้าและมีจีโนไทป์ของยีน *CYP2D6* แบบ GA (rs3892097) จะให้ประสิทธิภาพต่ำกว่า การใช้ยาตัวเดียวกันในผู้ป่วยโรคซึมเศร้าที่มีจีโนไทป์ ของยีน *CYP2D6* แบบปกติ (Zastrozhin et al., 2021) ผู้ป่วยติดเชื้อโควิด 19 ที่มีจีโนไทป์ของยีน *DRD3* แบบ AT (rs1486012) จะมีการเผาผลาญของ lopinavir / ritonavir ซึ่งเป็นหนึ่งในยาที่นำมาใช้เพื่อรักษา COVID-19 ที่ ลดลง (Takahashi et al., 2020) ส่งผลให้มีค่าความเข้มข้นของยา lopinavir / ritonavir ใน plasma เพิ่มขึ้น และ สร้างความเป็นพิษให้แก่ผู้ที่ใช้ได้ (López Aspiroz et al., 2019) อีกหนึ่งตัวอย่างที่น่าสนใจเกิดขึ้นกับการใช้ยา codeine เพื่อลดอาการเจ็บหลังการคลอดบุตรในแม่ที่มีการแสดงออกของยีน *CYP2D6* แบบที่มีความสามารถในการย่อยสลายยาได้รวดเร็วมากที่สุด ส่งผลให้ร่างกายเมื่อย่อยสลาย codeine ได้อย่างรวดเร็ว ทำให้มี morphine ปริมาณเหมาะสมในกระแสเลือด และเมื่อ morphine เหล่านี้ถูกส่งไปยังทารกแรกเกิดผ่านทางน้ำนม ทารก จึงเสียชีวิตในที่สุด (Madadi et al., 2007)

ในปัจจุบันเทคโนโลยีการทำ NGS เป็นที่แพร่หลายมากขึ้น อีกทั้งการหาลำดับนิวคลีโอไทด์โดยเทคนิคนี้สามารถทำให้เราได้ข้อมูลที่เกี่ยวข้องกับยีนที่ตอบสนองต่อยามากขึ้น ดังที่เห็นได้ในหลาย ๆ งานวิจัย เช่น การศึกษาลำดับพันธุกรรมใน 82 ยีนที่เกี่ยวข้องต่อการตอบสนองต่อยาในประชากรที่ไม่มีความสัมพันธ์ทางสายเลือดกันกว่า 5,632 คน ซึ่งในจำนวนนี้มีเพียง 2% เป็นข้อมูลของประชากรจากเอเชีย (Bush et al., 2016) หรือ ในการศึกษาการแปรผันทางพันธุกรรมใน 17 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา โดยใช้รหัสพันธุกรรมทั้งจีโนม (whole exome sequencing) ในตัวอย่างประชากรเมสติโซ (คนที่มีบรรพบุรุษผสมระหว่างชาวผิวขาวและชาวพื้นเมือง) และชาวพื้นเมืองแม็กซิโก (Gonzalez-Covarrubias et al., 2019) และการศึกษาล่าสุดในประเทศไทยที่ศึกษาลักษณะการแปรผันของยีนที่เกี่ยวข้องต่อการตอบสนองต่อยา 25 ยีนในกลุ่มผู้ป่วยโรคไหลตาย (Bruganda syndrome) จำนวน 291 โดยใช้รหัสพันธุกรรมทั้งจีโนม (Mauleekoonphairoj et al., 2020)

ในความพยายามที่จะกำหนดโปรไฟล์ star allele สำหรับรายการเภสัชพันธุศาสตร์ที่ครอบคลุมมากขึ้น การศึกษาจะตรวจสอบ 51 ยีนใน 171 คนจากประชากรไทยทั่วไป โดยตัวอย่างทั้งหมด ไม่มีความเกี่ยวข้องกันและมาจากภาคต่าง ๆ ของประเทศไทย เราพบว่า มี 25 ยีนที่มีการศึกษาในประเทศไทย เป็นกลุ่มยีนที่มีหลักฐานระดับ A ตามการแบ่งประเภทข้อมูลของฐานข้อมูล CPIC (Mauleekoonphairoj et al., 2020) ในการศึกษาครั้งนี้เราจะทำการศึกษาเพิ่มเติมในอีก 26 ยีน ที่ทำหน้าที่เกี่ยวข้องกับกิจกรรมต่าง ๆ ของยา ไม่ว่าจะเป็นการดูดซึม การกระจายตัว การเมแทบอลิซึม และการขจัดยา โดยเราพบว่ายีนเหล่านี้ต่างก็เกี่ยวข้องกับยาที่ถูกส่งจ่ายเพื่อรักษาโรคที่เป็นสาเหตุให้คนไทยเสียชีวิตมากที่สุด 10 อันดับแรก (<https://www.cdc.gov/globalhealth/ประเทศ/ประเทศไทย/>) ได้แก่ มะเร็งและภูมิคุ้มกันบกพร่อง โรคไม่ติดต่อ กลุ่มโรคความผิดปกติทางระบบประสาท, กลุ่มโรคความผิดปกติของกล้ามเนื้อและกระดูก, การจัดการความเจ็บปวดและการติดยาเสพติด, โรคที่ติดเชื้อมาจากไวรัสและจุลินทรีย์ (ตารางที่ 1) (<https://www.pharmgkb.org>; <https://www.drugbank.com>; <https://porta.fda.moph.go.th>) โดยรายชื่อยาเหล่านี้ล้วนได้รับการอนุมัติจากสำนักงานคณะกรรมการอาหารและยา กระทรวงสาธารณสุขประเทศไทย (<https://porta.fda.moph.go.th>)

ตารางที่ 5 รายชื่อโรคและยาที่เกี่ยวข้องกับยีนที่มีผลต่อการตอบสนองต่อยา 26 ยีนที่จะทำการศึกษาเพิ่มเติม

| โรค | ยีนที่เกี่ยวข้อง | ยา |
|--|--|--|
| กลุ่มโรคมะเร็งและภูมิคุ้มกันบกพร่อง (Cancer and immunology) | | |
| Neoplasms | <i>CYP1A1, CYP1A2, CYP2A6, CYP2E1, CYP19A1, CYP1B1, CYP26A1, CYP2F1, CYP2S1, CYP2W1, CYP4B1, SLCO1B3, SLC22A2, GSTM1, GSTP1, GSTT1</i> | capecitabine, erlotinib, docetaxel, tamoxifen, cyclophosphamide, letrozole, antrozoole, tretinoin, exemestane, fluorouracil, epirubicine, doxorubicin, |

| | | |
|---|--|--|
| | | sunitinib, cisplatin, paclitaxel, imatinib |
| Immunosuppressant | <i>SLCO1B3, CYP2J2, CYP3A7, GSTM1</i> | mycophenolic acid, tacrolimus, leflunomide |
| กลุ่มโรคที่ไม่ติดต่อ (Non-communicable diseases) | | |
| Cardiovascular diseases | <i>SLCO2B1, SLCO1B1, SLC15A2, POR, CYP3A7</i> | ticagrelor, warfarin, simvastatin, atorvastatin, rosuvastatin |
| Other non-communicable diseases | <i>SLCO2B1, SLC22A2, POR</i> | metformin, montelukast |
| กลุ่มโรคความผิดปกติทางระบบประสาท (Neurological disorders) | | |
| Psychiatric Disorder | <i>CYP1A1, CYP1A2, CYP3A7, CYP2A6, CYP3A43, UGT2B7, GSTM1, GSTT1</i> | olanzapine, paroxetine, bupropion, antipsychotics, clozapine, risperidone, ziprasidone |
| Seizures | <i>CYP1A1, UGT2B7</i> | carbamazepine, lamotrigine, valproic acid, diazepam |
| กลุ่มโรคความผิดปกติของกล้ามเนื้อและกระดูก (Musculoskeletal disorders) | | |
| Rheumatoid arthritis | <i>SULT1A1, UGT2B7</i> | sulfonamides, montelukast |
| การจัดการความเจ็บปวดและสารเสพติด (Management of pain and opioid addiction) | | |
| Management of pain | <i>CYP2A13, CYP2A6, RYR1, UGT2B15, UGT2B17, UGT2B7, TBXAS1</i> | aspirin, nicotin, oxycodone, morphine, codeine, isoflurane, sevoflurane, acetaminophen, exemestane |
| Management of opioid addiction | <i>UGT2B7</i> | methadone, fentanyl |
| โรคที่ติดเชื้อที่เกิดจากไวรัสและจุลินทรีย์ (Viral and microbial infection) | | |
| Antivirals | <i>CYP2A6, SLC22A2, CYP2R1, GSTM1</i> | efavirenz, lamivudine, zidovudine, ribavirin, nevirapine |
| Antibiotics | <i>CYP2A6, SLCO2B1, UGT2B7, GSTM1, GSTP1, GSTT1</i> | sulfonamides, metronidazole, streptomycin, rifampin, isoniazid |

จากการศึกษาพบว่า การแปรผันทางพันธุกรรมในยีนที่เกี่ยวข้องกับการตอบสนองต่อยามีความแตกต่างกันตามเชื้อชาติและกลุ่มประชากร (LL. et al., 2014; Zhang et al., 2019; Zhou et al., 2017) แม้แต่ในประชากรจากภูมิภาคเดียวกัน ก็ยังคงพบรูปแบบการแปรผันทางพันธุกรรมที่หลากหลาย (Petrović et al., 2020) ด้วยความก้าวหน้าของเทคโนโลยีการหาลำดับนิวคลีโอไทด์ และฐานข้อมูลทางเภสัชพันธุศาสตร์ ไม่ว่าจะเป็น PharmGKB หรือ PharmVar ทำให้เราสามารถตรวจสอบความถี่แอลลีล ความถี่จีโนไทป์ และความถี่ฟีโนไทป์ ของยีนเหล่านี้ได้เป็นจำนวนมากในเวลาเดียวกัน สำหรับเรื่องของประสิทธิภาพของขั้นตอน ก็มีการใช้เครื่องมือวิเคราะห์ที่ให้ผลแม่นยำและมีส่วนอย่างมากต่อการเพิ่มความถูกต้องให้กับผลลัพธ์ โดยปัจจุบันพบเครื่องมือที่ใช้ในการทำนาย star allele หลายเครื่องมือรวมถึง PharmCAT (Sanguhl et al., 2020), Astrolabe (Twist et al., 2016), Aldy (Numanagić et al., 2018) Stargazer (Lee et al., 2019(a)), StellarPGX (Twesigomwe et al., 2021) และ PharmVIP (Piriyaongsa et al., 2021)

ในหลาย ๆ ฐานข้อมูลยังไม่มีเครื่องมือสำหรับสำหรับการทำนายการแปรผันทางพันธุกรรมโดยตรง แต่มีการร่วมมือกันระหว่างฐานข้อมูลกับเครื่องมือ PharmCAT ซึ่งใช้วิธีการการทำนาย star allele สำหรับยีน *CYP2D6* โดยการยืมเครื่องมือการทำนาย จากโปรแกรม Astrolabe (Sanguhl et al., 2020; Twesigomwe et al., 2020) จากการศึกษเปรียบเทียบผลการทำนาย star allele โดยโปรแกรม Astrolabe, Aldy และ Stargazer พบว่าโปรแกรม Stargazer มีประสิทธิภาพสูงสุดในด้านการทำนายรูปแบบการแปรผันทางพันธุกรรมและลักษณะฟีโนไทป์ รองลงมาคือ Aldy และ Astrolabe ตามลำดับ โดยในการศึกษานี้ใช้ชุดข้อมูล WGS จากฐานข้อมูล Genetic Testing Reference Material Coordination Program (GeT-RM) (Twesigomwe et al., 2020) สำหรับโปรแกรม Stargazer จะใช้การแปลข้อมูลจากข้อมูลการแปรผันทางพันธุกรรมควบคู่ไปกับข้อมูลจำนวนความลึกของ read โดยใช้ตารางการแปลลักษณะของ haplotype (Lee et al., 2019(a); Lee et al., 2019(b)) โปรแกรม Stargazer รองรับการทำนายผลของยีนจำนวน 51 ยีน ในขณะที่ Aldy และ Astrolabe รองรับเพียงแค่ 25 และ 3 ยีนตามลำดับ (Numanagić et al., 2018; Sanguhl et al., 2020; Twist et al., 2016) นอกจากนี้ Stargazer ยังสามารถทำนาย star allele ที่ทราบรูปแบบอยู่แล้วของฐานข้อมูล GeT-RM ได้ถูกต้อง 100% อีกทั้งยังสามารถระบุ star allele เพิ่มเติมที่ไม่พบในรายงานก่อนหน้านี้ star allele ที่พบเพิ่มเติมนั้นมีทั้งที่เป็น star allele ที่ทราบรูปแบบการแปรผันอยู่แล้ว (know variant) และการแปรผันเชิงโครงสร้างรูปแบบใหม่ที่ไม่มีในฐานข้อมูล (novel SVs) (Lee et al., 2019(b))

บทที่ 3

วิธีการดำเนินงานวิจัย

3.1 ตัวอย่างที่ใช้ในงานวิจัย

ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลของรหัสพันธุกรรมทั้งจีโนม (whole genome sequencing) โดยตัวอย่างที่ใช้ในการวิจัย เป็นข้อมูลจีโนมของพ่อหรือแม่ของเด็กที่ป่วยเป็นโรคที่วินิจฉัยได้ยาก หรือโรคหายาก (rare disease) ที่มีบุตรธิดาเข้ารับการรักษาตัวอยู่ที่โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย โดยที่พ่อหรือแม่เป็นผู้ที่ไม่มีโรค ไม่มีความเกี่ยวข้องกันทางสายเลือด และมีภูมิลำเนาอยู่ในประเทศไทย

3.2 ค่าแถลงจริยธรรม

การศึกษานี้ได้รับการอนุมัติโดยคณะกรรมการพิจารณาจริยธรรมการวิจัยในคน คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เลข IRB 264/62 และผู้เข้าร่วมทุกท่านได้ให้ความยินยอมเป็นลายลักษณ์อักษร

3.3 การคำนวณขอบของความผิดพลาด (Margin of error calculation)

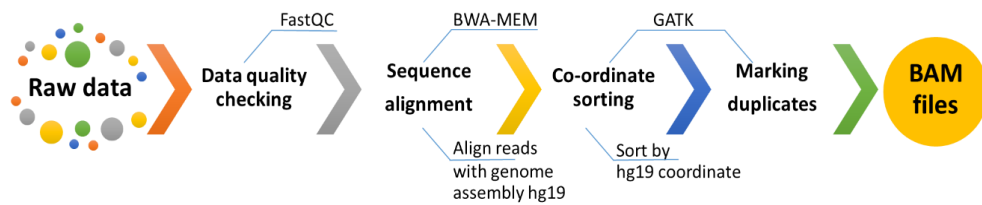
ขอบของความผิดพลาดถูกคำนวณตามสูตรขนาดตัวอย่างของ Cochran ที่ระดับความเชื่อมั่น 95% โดยกลุ่มตัวอย่างมีขนาด 171 คน และประชากรไทยมีขนาด (จำนวน) โดยประมาณ 67 ล้านคน ซึ่งพบว่าโดยสัดส่วนของประชากร ที่มีคุณลักษณะหนึ่ง ๆ เช่นการแสดงออกของยีนที่มีผลต่อการตอบสนองต่อยาอย่างปกติ (normal phenotype) ที่ค่า 0.5 ซึ่งค่านี้จะให้ขอบของความผิดพลาดสูงสุด

3.4 การวิเคราะห์ผล Star allele

การวิเคราะห์ผล star allele แบ่งออกเป็น 2 ขั้นตอนใหญ่ ๆ ได้แก่ ขั้นตอนการวิเคราะห์ข้อมูล (pre-analysis) และขั้นวิเคราะห์ข้อมูล (data analysis)

3.4.1 ขั้นก่อนการวิเคราะห์ข้อมูล (pre-analysis)

เป็นขั้นตอนเกี่ยวกับการตรวจสอบคุณภาพข้อมูล (Data quality checking) sequence alignment (ภาพที่ 5) การเรียกการแปรผันทางพันธุกรรม (Variant calling) (ภาพที่ 6) และการคำนวณความลึกของ read (GATK depth of coverage calculation)



ภาพที่ 5 แผนงานการทำ data quality checking และ sequence alignment

3.4.1.1 การตรวจสอบคุณภาพของข้อมูล (Data quality checking)

Fastq file ของกลุ่มตัวอย่างจำนวน 171 ไฟล์ จะถูกนำมาตรวจสอบคุณภาพของ sequence โดยใช้โปรแกรม FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) ตั้งคำสั่งด้านล่าง (ภาพที่ 5)

```
/path/to/fastqc [-o /path/to/outdir] [-f /path/to/fastq]
```

โดย -o คือ ที่อยู่ของโฟลเดอร์ที่จะทำการจัดเก็บผลลัพธ์ไฟล์จากการตรวจสอบคุณภาพ

-f คือ ที่อยู่ของไฟล์ FASTQ

จากนั้นนำไฟล์ที่ผ่านการตรวจสอบคุณภาพมาแปลงเป็นไฟล์ SAM ตั้งคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
```

```
FastqToSam\
```

```
-F1 /path/to/r1.fastq\
```

```
-F2 /path/to/r2.fastq\
```

```
-O /path/to/fastqtosam.bam\
```

```
-RG [readGroup_tag: WGS] -SM [sample name: HS0x] \
```

```
-LB [library name: WGS] -PL [platform: illumine] \
```

```
-CN EMA -DT [date: YYYY-MM-DD]
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-F1 คือ ที่อยู่ของไฟล์ FASTQ จากทิศทางการอ่านที่หนึ่ง

- F2 คือ ที่อยู่ของไฟล์ FASTQ จากทิศทางการอ่านที่สอง
- O คือ ที่อยู่ของไฟล์ BAM (*.bam) ซึ่งเป็นไฟล์ไบนารีที่ถูกแปลงมาจากไฟล์ SAM
- RG คือ ชื่อกลุ่มของ read
- SM คือ ชื่อตัวอย่าง
- LB คือ ชื่อ library ที่ใช้ในการเตรียมข้อมูล
- PL คือ ชื่อแพลตฟอร์มที่ใช้ในการจัดลำดับข้อมูล เช่น ILLUMINA, SOLID
- CN คือ ชื่อของที่มาของข้อมูล (ใส่เป็นตัวอักษร)
- DT คือ วันที่ที่จัดทำไฟล์

ไฟล์ BAM ที่ได้จะถูกนำมาประมวลผลของ illumina (ภาพที่ 5) ดังคำสั่งด้านล่าง

```
java -Xmx8G -jar /data/users/natnicha/tools/picard-tools-1.119/MarkIlluminaAdapters.jar \
```

```
I= /path/to/fastqtosam.bam \
```

```
O= /path/to/markilluminaadapters.bam \
```

```
M= /path/to/metrics.txt
```

โดย -Xmx8G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 8G

I คือ ที่อยู่ของไฟล์ BAM

O คือ ที่อยู่ของไฟล์ BAM ที่ระบุแอดปเตอร์แล้ว

M คือ ที่อยู่ของไฟล์เมตริกซ์ที่ใช้ในการจัดลำดับรหัสพันธุกรรม

ไฟล์ BAM ที่ระบุแอดปเตอร์แล้ว จะถูกนำมาแปลงเป็นไฟล์ FASTQ อีกครั้ง (ภาพที่ 3.1) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /data/users/natnicha/tools/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar SamToFastq \
```

```
-I /path/to/markilluminaadapters.bam \
```

```
-F /path/to/interleave.fq -CLIP_ATTR XT -CLIP_ACT 2 -INTER true --
NON_PF true
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-I คือ ที่อยู่ของไฟล์ BAM ที่ทำการระบุอะแดปเตอร์แล้ว

-F คือ ที่อยู่ของไฟล์ FASTQ ที่ถูกแปลงจากไฟล์ BAM

-CLIP_ATTR คือ ชื่อแอตทริบิวต์ที่เก็บค่าตำแหน่งเบสที่ไฟล์ SAM ว่าควรจะทำกา
ตัดทิ้ง

-CLIP_ACT คือ คำสั่งให้ทำการตัด read โดยจำนวนเต็มใด ๆ หมายถึง ค่าคุณภาพ
ของเบสที่จะถูกตั้งค่าให้ตัดทิ้ง

-INTER คือ การระบุให้สร้าง ไฟล์ interleaved.fastq ที่เป็นไฟล์ FASTQ ที่รวม
read จากทิศทางไปข้างหน้าและไปข้างหลัง โดย true หมายถึง สั่งให้สร้างไฟล์
interleaved.fastq และ false หมายถึง ไม่ต้องสร้างไฟล์ interleaved.fastq

--NON_PF คือ การระบุให้รวม read ที่ไม่มีคุณภาพ โดย true หมายถึง ให้รวม read
ที่ไม่มีคุณภาพ และ false หมายถึง ไม่รวม read ที่ไม่มีคุณภาพ

จากนั้นเราจะนำไฟล์ interleaved.fastq มาใช้ในการจัดตำแหน่ง รหัสพันธุกรรมทั้งจีโนมจะถูก
map เทียบกับจีโนมอ้างอิง (reference genome) hg19 ด้วยโปรแกรม Burrows-Wheeler
Alignment Tool (BWA) ซึ่งทำ local alignment แบบ maximal exact matches (MEMs)
หรือ BWA-MEM model (Li & Durbin, 2009) (ภาพที่ 5) ตามคำสั่งด้านล่าง

```
/path/to/bwa-0.7.17/bwa mem -M -t 7 -p\
```

```
/path/to/ref/ucsc.hg19.fasta\
```

```
/path/to/interleave.fq > /path/to/bwa_mem_out.bam
```

โดย /path/to/ref/ucsc.hg19.fasta คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

/path/to/interleave.fq คือ ที่อยู่ของไฟล์ interleaved.fq

/path/to/bwa_mem_out.bam คือ ที่อยู่ของไฟล์ BAM ที่บรรจุรหัสพันธุกรรมที่
ทำการจัดตำแหน่งเทียบกับจีโนมอ้างอิงแล้ว

จากนั้นเราจะทำการรวมข้อมูลต่าง ๆ ของตัวอย่างจากไฟล์ BAM ที่ได้รับการจัดลำดับเทียบกับ จีโนมอ้างอิงแล้ว และไฟล์ BAM ที่ยังไม่ได้ทำการจัดอันดับ (ดั้งเดิม) ตามคำสั่งด้านล่าง (ภาพที่ 5)

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
MergeBamAlignment\
```

```
-ALIGNED /path/to/bwa_mem_out.bam\
```

```
-UNMAPPED /path/to/fastqtosam.bam\
```

```
-O /path/to/mergebamalignment.bam\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
```

```
--CREATE_INDEX true -MC true --CLIP_ADAPTERS false --
CLIP_OVERLAPPING_READS true\
```

```
--INCLUDE_SECONDARY_ALIGNMENTS true -MAX_GAPS -1 --
PRIMARY_ALIGNMENT_STRATEGY MostDistant\
```

```
--ATTRIBUTES_TO_RETAIN XS
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ที่ได้มากที่สุด 32G

-ALIGNED คือ ที่อยู่ของไฟล์ BAM ที่ถูกจัดตำแหน่งแล้ว

-UNMAPPED คือ ที่อยู่ของไฟล์ BAM ที่ยังไม่ได้ทำการจัดตำแหน่ง

-O คือ ที่อยู่ของไฟล์ BAM ที่บรรจุข้อมูลทั้งหมดของตัวอย่าง

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

--CREATE_INDEX คือ การสั่งให้สร้างดัชนีของไฟล์ BAM ที่บรรจุข้อมูลทั้งหมดของ ตัวอย่าง โดย true หมายถึง สั่งให้สร้าง และ false หมายถึง ไม่ต้องสร้าง

-MC คือ การสั่งให้ใส่แถบระบุเลข CIGAR โดย true หมายถึง สั่งให้ใส่ และ false หมายถึง ไม่ต้องใส่

--CLIP_ADAPTERS คือ การสั่งให้ตัด adapter โดย true หมายถึง สั่งให้ตัด adapter และ false หมายถึง ไม่ต้องตัด adapter

--CLIP_OVERLAPPING_READS เป็นคำสั่งสำหรับ paired end read โดยทำการตัดเบสที่บริเวณปลาย 3' แล้วเก็บเบสไว้ (soft clip) เพื่อไม่ให้ยาวเกินปลาย 5' ของสายคู่สม adapter โดย true หมายถึง สั่งให้ทำ และ false หมายถึง ไม่ต้องทำ

--INCLUDE_SECONDARY_ALIGNMENTS คือ การสั่งให้เขียนข้อมูลของการเกิดการจับลำดับที่มากกว่าหนึ่งตำแหน่ง (secondary alignment) โดย true หมายถึง สั่งให้เขียน และ false หมายถึง ไม่ต้องเขียน

-MAX_GAPS คือ จำนวนของ insertion หรือ deletion ที่มากที่สุดที่อนุญาตให้เกิดในการจัดลำดับ

--PRIMARY_ALIGNMENT_STRATEGY คือ ชื่อของวิธีการเลือกการจัดลำดับที่บริเวณหลัก (primary alignment)

--ATTRIBUTES_TO_RETAIN คือ แท็กที่จะระบุข้อมูลที่สงวนไว้ (ที่ขึ้นต้นด้วย X, Y หรือ Z) เป็นตำแหน่งข้อมูลที่จะเลือกมาจากข้อมูลที่ถูกรวบรวมแล้ว

เมื่อทำการรวมข้อมูลเสร็จแล้วไฟล์ BAM ที่ได้จะถูกนำมาเรียงลำดับข้อมูลตามลำดับโครโมโซม (co-ordinate sorting) (ภาพที่ 5) ตามคำสั่ง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar SortSam\
-l /path/to/mergebamalignment.bam\
-O /path/to/sorted.bam\
-SO coordinate
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-l คือ ที่อยู่ของไฟล์ BAM

-O คือ ที่อยู่ของไฟล์ BAM ที่ทำการเรียงลำดับแล้ว

-SO คือ วิธีการจัดลำดับ โดย coordinate หมายถึง ให้เรียงลำดับตามลำดับโครโมโซม

ไฟล์ BAM ที่ถูกเรียงลำดับตามลำดับโครโมโซมแล้ว จะถูกนำมาแท็ก หรือ ทำเครื่องหมายระบุการซ้ำกันของ read (mark duplicate) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
```

```
MarkDuplicates\
```

```
-I /path/to/sorted.bam\
```

```
-O /path/to/dedup.bam\
```

```
-M /path/to/dedup_metrics.txt\
```

```
--OPTICAL_DUPLICATE_PIXEL_DISTANCE 2500\
```

```
--CREATE_INDEX true --REMOVE_DUPLICATES true
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-I คือ ที่อยู่ของไฟล์ BAM

-O คือ ที่อยู่ของไฟล์ BAM ที่ถูกระบุการซ้ำกันของ read แล้ว

-M คือ ที่อยู่ของไฟล์เมตริกซ์ที่ใช้ในการระบุการซ้ำกันของ read

--OPTICAL_DUPLICATE_PIXEL_DISTANCE คือ ระยะห่างสูงสุดระหว่าง 2 คลัสเตอร์ที่ซ้ำกัน ใช้เป็นเกณฑ์พิจารณาว่าคลัสเตอร์ทั้งสองเป็นคลัสเตอร์ที่ซ้ำกัน

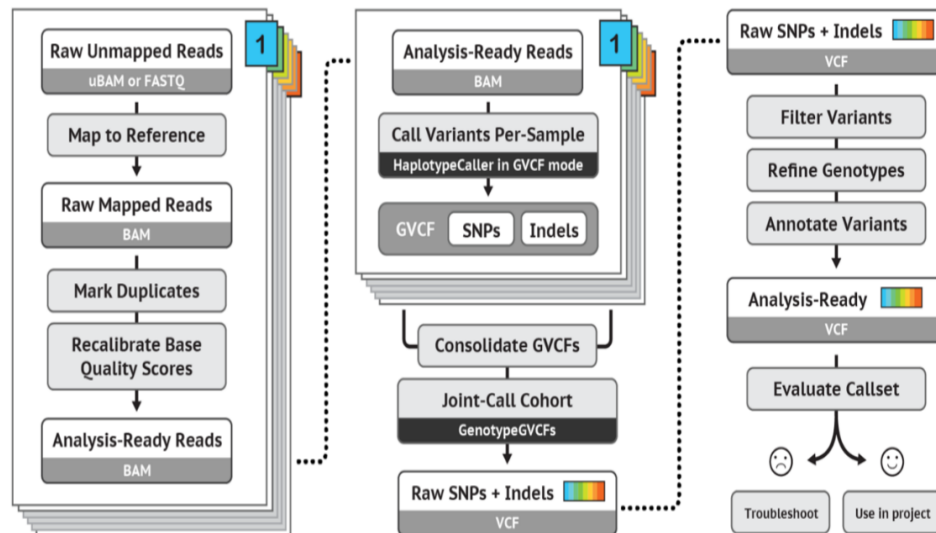
--CREATE_INDEX คือ การสั่งให้สร้างดัชนีของไฟล์ BAM ที่บรรจุข้อมูลทั้งหมดของตัวอย่าง โดย true หมายถึง สั่งให้สร้าง และ false หมายถึง ไม่ต้องสร้าง

--REMOVE_DUPLICATES คือ การสั่งให้ไม่เขียน read ที่มีนิวคลีโอไทด์ซ้ำกันลงในไฟล์ BAM ที่ถูกระบุการซ้ำกันของ read โดย true หมายถึง ไม่เขียนข้อมูล และ false หมายถึง จะทำการเก็บค่าที่ซ้ำกันในรูปของเซตข้อมูล

ทั้งนี้เพื่อความรวดเร็วในการวิเคราะห์ข้อมูล เราได้เลือกข้อมูลนิวคลีโอไทด์มาแค่ในส่วนของ 51 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา ซึ่งได้แก่ *CACNA1S*, *CFTR*, *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP2A6*, *CYP2A13*, *CYP2B6*, *CYP2C8*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP2E1*, *CYP2F1*, *CYP2J2*, *CYP2R1*, *CYP2S1*, *CYP2W1*, *CYP3A4*, *CYP3A5*, *CYP3A7*, *CYP3A43*, *CYP4B1*, *CYP26A1*, *CYP4F2*, *CYP19A1*, *DPYD*, *G6PD*, *GSTM1*, *GSTP1*, *GSTT1*, *IFNL3*, *NAT1*, *NAT2*, *NUDT15*, *POR*, *RYR1*, *SLC15A2*, *SLC22A2*, *SLCO1B1*, *SLCO1B3*, *SLCO2B1*, *SULT1A1*, *TBXAS1*, *TPMT*, *UGT1A1*, *UGT1A4*, *UGT2B7*, *UGT2B15*, *UGT2B17* และ *VKORC1*

3.4.1.2 การเรียกการแปรผันทางพันธุกรรม (Variant calling)

ในขั้นตอนนี้เราได้ใช้ GATK workflow best practice เวอร์ชัน 4.1.3.0 (DePristo et al., 2011; Van der Auwera et al., 2013) เพื่อการวิเคราะห์การแปรผันทางพันธุกรรมที่เกิดขึ้นในแต่ละบุคคล



ภาพที่ 6 แผนงาน GATK workflow best practice เพื่อการทำ variant calling
(<https://gatk.broadinstitute.org/hc/en-us>)

ไฟล์ BAM ที่ถูกทำเครื่องหมายระบุการซ้ำกันของ read และไฟล์เมตริกซ์ จะถูกนำมาใช้เพื่อการปรับคะแนนคุณภาพของเบส โดยการคำนวณ base quality score ใหม่อีกครั้ง (Base Quality Score Recalibration: BQSR) (ภาพที่ 6) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
```

```
BaseRecalibrator \
```

```
-R /path/to/ref/ucsc.hg19.fasta \
```

```
-I /path/to/dedup.bam \
```

```
--known-sites /path/to/dbsnp_138.hg19.vcf \
```

```
--known-sites
```

```
/path/to/Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \
```

```
-O /path/to/recal.table
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

-I คือ ที่อยู่ของไฟล์ BAM

--known-sites คือ ที่อยู่ของไฟล์ฐานข้อมูลที่ใช้ในการเรียกการแปรผันทางพันธุกรรม โดยอาจใช้ได้มากกว่าหนึ่งฐานข้อมูลในการระบุการแปรผันทางพันธุกรรมก็ได้

-O คือ ที่อยู่ของไฟล์คะแนนคุณภาพเบสหลังจากปรับคะแนนแล้ว โดยจะถูกเก็บในรูปแบบของตาราง (*.table)

ตารางคะแนนคุณภาพที่ได้รับการคำนวณใหม่แล้ว จะถูกนำมาใช้ในการทำ BAM ไฟล์ (ภาพที่ 6) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar ApplyBQSR\
  -R /path/to/ref/ucsc.hg19.fasta\
  -I /path/to/dedup.bam\
  -bqsr /path/to/recal.table\
  -L /path/to/snpCalling_regions.bed\
  -O /path/to/bqsr.bam
```

โดย Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

-I คือ ที่อยู่ของไฟล์ BAM

-bqsr คือ ที่อยู่ของไฟล์คะแนนคุณภาพเบสหลังจากปรับคะแนนแล้ว โดยจะถูกเก็บในรูปแบบของตาราง (*.table)

-L คือ ที่อยู่ของไฟล์ BED ของตำแหน่งการแปรผันทางพันธุกรรม

-O คือ ที่อยู่ของไฟล์ BAM ที่ได้รับการปรับคะแนนคุณภาพเบสใหม่แล้ว

เมื่อผ่านการทำ BQSR แล้ว ไฟล์ BAM ของแต่ละคนจะถูกนำมาเรียกหาการแปรผันทางพันธุกรรม (HaplotypeCaller) โดยใช้ข้อมูลจากจีโนมอ้างอิง (hg19) ประกอบกับ ฐานข้อมูล SNPs และ INDELS (ภาพที่ 6) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/tools/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
HaplotypeCaller\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
-l /path/to/bqsr.bam\
-O /path/to/vcf/sample.g.vcf\
-ERC GVCF\
-L /path/to/snpCalling_regions.bed
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

-l คือ ที่อยู่ของไฟล์ BAM

-O คือ ที่อยู่ของไฟล์ VCF

-ERC คือ โหมดสำหรับเขียนคะแนนความเชื่อมั่นอ้างอิง (reference confidence scores)

-L คือ ที่อยู่ของไฟล์ BED ของตำแหน่งการแปรผันทางพันธุกรรม

หลังจากการทำ HaplotypeCaller จะได้ gVCF file สำหรับแต่ละคน ซึ่งจะถูกรวมเข้าไว้เป็นไฟล์เดียวกัน (ภาพที่ 6) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
CombineGVCFs\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
--variant /path/to/vcf/sample1.g.vcf\
--variant /path/to/vcf/sample2.g.vcf\
-O /path/to/vcf/CombineVCF.g.vcf
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

--variant คือ ที่อยู่ของไฟล์ VCF ของแต่ละตัวอย่าง

-O คือ ที่อยู่ของไฟล์ VCF ที่รวมของแต่ละตัวอย่างเข้าไว้ด้วยกัน

ไฟล์ VCF ที่รวมข้อมูลการแปรผันทางพันธุกรรมของแต่ละคนเข้าไว้ด้วยกันแล้ว จะถูกนำมาให้ข้อมูลจากการแปรผันทางพันธุกรรมที่ทราบอยู่แล้ว (ภาพที่ 6) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
GenotypeGVCFs\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
```

```
-V /path/to/vcf/CombineVCF.g.vcf\
```

```
-O /path/to/vcf/raw.vcf\
```

```
-D /path/to/dbsnp_138.hg19.vcf
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

-V คือ ที่อยู่ของไฟล์ VCF ที่รวมของแต่ละตัวอย่างเข้าไว้ด้วยกัน

-O คือ ที่อยู่ของไฟล์ VCF ที่ถูกให้ข้อมูลการแปรผันทางพันธุกรรมที่ทราบอยู่แล้ว

-D คือ ที่อยู่ของไฟล์ฐานข้อมูลที่ใช้ในการเรียกการแปรผันทางพันธุกรรม

จากนั้น เราจะทำการแยกการแปรผันทางพันธุกรรม จากไฟล์ VCF ที่ได้มาจากขั้นก่อนหน้า โดยแยกออกเป็นตำแหน่งที่เกิดการแปรผันทางพันธุกรรมแบบ SNP (ไฟล์แรก) และ indel (ไฟล์ที่สอง) (ภาพที่ 6) ตามคำสั่งด้านล่าง

คำสั่งสำหรับการแยกการแปรผันทางพันธุกรรมประเภท SNP

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
SelectVariants\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
```

```
-V /path/to/vcf/raw.vcf \
```

```
-select-type SNP\
```

```
-O /path/to/vcf/raw.SNP.vcf
```

คำสั่งสำหรับการแยกการแปรผันทางพันธุกรรมประเภท indel

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
SelectVariants\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
```

```
-V /path/to/vcf/raw.vcf\
```

```
-select-type INDEL\
```

```
-O /path/to/vcf/raw.INDEL.vcf
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

-V คือ ที่อยู่ของไฟล์ VCF

-select-type หมายถึง รูปแบบการแปรผันทางพันธุกรรมที่ต้องการเลือก หากใช้ฐานข้อมูล SNP ให้ใช้ -select-type แบบ SNP ถ้าหาก ใช้ฐานข้อมูล Indel ให้ใช้ -select-type แบบ INDEL

-O คือ ที่อยู่ของไฟล์ VCF หลังทำการเลือกประเภทรูปแบบการแปรผันทางพันธุกรรม

จากนั้นจะนำไฟล์การแปรผันทางพันธุกรรมแบบ SNP และแบบ indel มาทำการคัดเลือก (กรอง) การแปรผันทางพันธุกรรมที่มีความน่าเชื่อถือสำหรับการวิเคราะห์ในลำดับถัดไป (downstream process) โดยการแปรผันทางพันธุกรรมที่ไม่มีคุณสมบัติต่อไปนี้จะถูกเก็บเอาไว้

SNP: QualByDepth (QD) <2.0, FisherStrand (FS) >60.0, RMSMappingQuality (MQ) <40, MappingQualityRankSumTest (MQRankSum) <12.5, และ ReadPosRankSumTest (ReadPosRankSum) <-8.0

Indel: QD <2.0, FS >200.0, and ReadPosRankSum <-20.0

โดยเราจะทำการคัดเลือกการแปรผันทางพันธุกรรม (ภาพที่ 6) ตามคำสั่งด้านล่าง

คำสั่งสำหรับการคัดเลือก (กรอง) การแปรผันทางพันธุกรรมประเภท SNP

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
VariantFiltration\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
-V /path/to/vcf/raw.SNP.vcf\
-filter "QD < 2.0 || FS > 60.0 || MQ < 20.0 || MQRankSum < -12.5 ||
ReadPosRankSum < -8.0"\
--filter-name "SNP_Filter"\
-O /path/to/vcf/filtered_snps.vcf
```

คำสั่งสำหรับการคัดเลือก (กรอง) การแปรผันทางพันธุกรรมประเภท indel

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar
VariantFiltration\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
-V /path/to/vcf/raw.INDEL.vcf\
-filter "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0"\
--filter-name "INDEL_Filter"\
-O /path/to/vcf/filtered_indels.vcf
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ที่ได้มากที่สุด 32G

-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง

-V คือ ที่อยู่ของไฟล์ VCF

-filter คือ เงื่อนไขที่ใช้ในการคัดเลือกการแปรผันทางพันธุกรรม

-filter-name คือ ชื่อที่จะใช้สำหรับรายการตัวกรอง

-O คือ ที่อยู่ของไฟล์ VCF ที่บรรจุข้อมูลการแปรผันทางพันธุกรรมที่ผ่านการคัดเลือกแล้ว

ไฟล์ VCF ทั้งสองที่ได้จากการกรอง SNP และ indel จะถูกนำมาเขียนข้อมูลรวมกัน (ภาพที่ 6) ตามคำสั่งด้านล่าง

```
java -Xmx32G -jar /path/to/gatk-4.1.3.0/gatk-package-4.1.3.0-local.jar MergeVcfs\
```

```
-R /path/to/ref/ucsc.hg19.fasta\
-D /path/to/dbsnp_138.hg19.vcf\
-l /path/to/vcf/filtered_snps.vcf\
-l /path/to/vcf/filtered_indels.vcf\
-O /path/to/vcf/output.vcf
```

โดย -Xmx32G คือ การอนุญาตให้ใช้ทรัพยากรสำหรับการวิเคราะห์ได้มากที่สุด 32G

```
-R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง
-D คือ ที่อยู่ของไฟล์ฐานข้อมูลการแปรผันทางพันธุกรรม
-l คือ ที่อยู่ของไฟล์ VCF ที่ผ่านการกรอง SNP
-l คือ ที่อยู่ของไฟล์ VCF ที่ผ่านการกรอง Indel
-O คือ ที่อยู่ของไฟล์ VCF ที่รวมข้อมูลการแปรผันทางพันธุกรรมทั้งหมดที่เกิดขึ้นใน
ทุกตัวอย่าง
```

3.4.1.3 การคำนวณความลึกของ read (GATK depth of coverage calculation)

เนื่องจาก Stargazer เป็นโปรแกรมที่รับข้อมูลความลึกของ read ตามแบบฟอร์มของ GATK เราจึงใช้ GATK-DepthOfCoverage เวอร์ชัน 3.8.1 ในการอ่านค่า BAM และคำนวณจำนวน read ที่ปกคลุมในแต่ละตำแหน่งของยีนตามคำสั่งด้านล่าง โดย Stargazer เรียกไฟล์ที่บรรจุข้อมูลของความลึกของ read ว่าไฟล์ GDF (GATK-DepthOfCoverage Format) (ภาพที่ 7)

```
java -jar /path/to/GenomeAnalysisTK-3.8-1-0/GenomeAnalysisTK.jar \
```

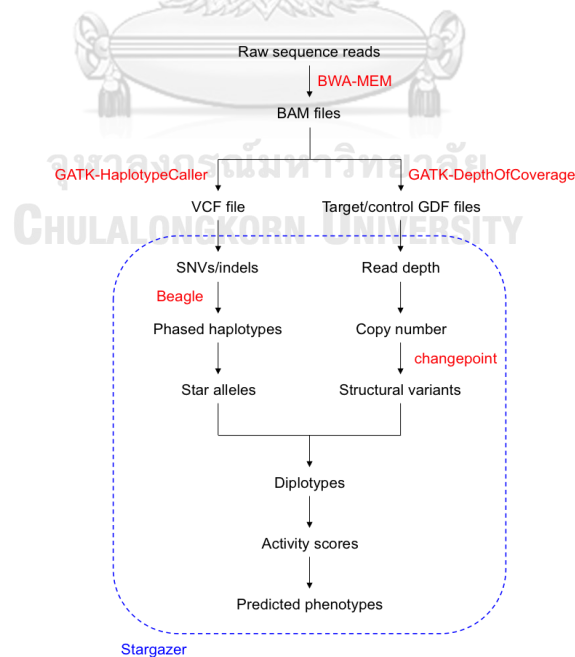
```
-T DepthOfCoverage \
-R /path/to/ref/ucsc.hg19.fasta \
-o /path/to/out.gdf \
-l /path/to/bam.list \
-geneList:REFSEQ /path/to/refseq.txt \
-L /path/to/gene_location.list
```

- โดย
- T คือ โหมด โดย DepthOfCoverage หมายถึง โหมดที่ใช้ในการคำนวณ depth of coverage
 - R คือ ที่อยู่ของไฟล์จีโนมอ้างอิง
 - o คือ ที่อยู่ของไฟล์ GDF
 - l คือ ที่อยู่ของไฟล์ BAM
 - geneList:REFSEQ คือ ที่อยู่ของไฟล์ที่บรรจุข้อมูลต่าง ๆ ของยีนอ้างอิง (reference sequence gene)
 - L คือ ที่อยู่ของไฟล์ข้อมูลเกี่ยวกับที่อยู่ของยีน

3.4.2 ชั้นวิเคราะห์ข้อมูล (data analysis)

3.4.2.1 การทำนาย Star allele

สำหรับการวิเคราะห์ข้อมูลใน 51 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา เราจะใช้โปรแกรม Stargazer เวอร์ชัน 1.0.8 (Lee et al., 2019(b)) เพื่อทำนายรูปแบบการแปรผันทางพันธุกรรม



ภาพที่ 7 แผนภาพวิธีการทำงานของ Stargazer (Lee et al., 2019(b))

จากแผนภาพ จะเห็นได้ว่าตัวโปรแกรม Stargazer จะเรียกโปรแกรม Beagles เพื่อทำการแยก haplotype ของแต่ละแอลลีล (haplotype phasing) (Browning et al., 2018; Browning & Browning, 2007) โดย Stargazer จะระบุให้ Beagles ใช้ reference panel จากฐานข้อมูล 1000 genomes project (1KGP) ซึ่งในการศึกษาครั้งนี้เราได้ปรับเปลี่ยน reference panel ให้ใช้เพียงข้อมูลจากกลุ่มประชากรเอเชียและเอเชียตะวันออกเฉียงใต้ เพื่อเพิ่มความถูกต้องของการทำ haplotype phasing ของกลุ่มตัวอย่าง ที่ซึ่งเป็นประชากรไทย (ภาพที่ 7)

Stargazer จะทำการทำนายรูปแบบการแปรผันทางพันธุกรรมโดยอาศัยข้อมูลของไฟล์ BAM GDF และ VCF ซึ่งเราสามารถเลือกที่จะใช้ข้อมูลความลึกของ read ของยีนควบคุม (control gene) *EGFR* *VDR* หรือ *RYS1* ยีนใดยีนหนึ่ง ซึ่งจะเป็ค่าสำหรันำไปเปรียบเทียบเพื่อกำหนดจำนวนชุด (copy number) ของยีนเป้าหมาย (target gene) จากนั้น Stargazer จะเรียกโปรแกรม changepoint มาเพื่อกำหนดหาจุดเปลี่ยนของข้อมูล เพื่อตัดสินใจว่าภายในยีนเป้าหมาย เกิดการแปรผันทางพันธุกรรมเชิงโครงสร้างหรือไม่ ซึ่งการคำนวณนี้อาศัยข้อมูลความลึกของ read ของยีนควบคุม และยีนเป้าหมาย ที่บรรจุอยู่ใน GDF file เดียวกัน ในการศึกษาครั้งนี้เราได้ทำการทดลองโดยใช้ยีนควบคุมทั้งสามยีนทีละยีน เพื่อเปรียบเทียบผลการทำนายความแตกต่างของรูปแบบการแปรผันทางพันธุกรรมจากการเลือกยีนควบคุมที่แตกต่างกัน (Lee et al., 2019(b))

หลังจากที่ Stargazer ได้ผลการระบุ star allele และรูปแบบการแปรผันทางพันธุกรรมเชิงโครงสร้าง ของแต่ละ haplotype บนแต่ละแอลลีลได้แล้ว จะนำมาเขียนรวมกันเป็น diplotype แล้วจึงทำการแปลงค่า diplotype ให้เป็นคะแนนของการแสดงออกของยีน (Activity scores) เพื่อตัดสินใจว่าท้ายที่สุดแล้ว star allele ที่ทำนายได้จะมีการแสดงออกทางฟีโนไทป์แบบใด (ภาพที่ 7) โดยใช้คำสั่งด้านล่าง

```
python3 /path/to/stargazer.py genotype \
```

```
-o /path/to/outdir \
```

```
-t [target gene] \
```

```
-c [control gene] \
```

```
--vcf /path/to/file.vcf \
```

```
--gdf /path/to/file.gdf \
```

-d wgs

โดย -O คือ ที่อยู่ของโฟลเดอร์สำหรับผลการวิเคราะห์

-t คือ ยีนเป้าหมาย (target gene)

-c คือ ยีนควบคุม (control gene)

--vcf คือ ที่อยู่ของไฟล์ VCF

--gdf คือ ที่อยู่ของไฟล์ GDF

-d คือ ลักษณะข้อมูล

3.4.2.2 การคำนวณหาความถี่แอลลีล (allele frequency calculation)

หลังจากที่ทำการวิเคราะห์ star allele แล้ว ไฟล์ VCF จะถูกนำมาคำนวณหาค่าสถิติ missingness ของแต่ละ SNV โดยใช้โปรแกรม PLINK เวอร์ชัน 1.7 หาก SNV ใด ๆ มีค่าสถิติ missingness ของจีโนไทป์ มากกว่า 0.2 จะถูกลบทิ้ง ตามคำสั่งด้านล่าง

```
/path/to/plink --vcf /path/to/vcf/output.vcf --geno 0.2 --make-bed --out sample_output
```

โดย --vcf คือ ที่อยู่ของไฟล์ VCF

--geno คือ คำสั่งให้ลบ SNV ที่มีค่า missingness rate มากกว่าค่าที่กำหนด

--make-bed คือ คำสั่งให้สร้างไฟล์ binary fileset สำหรับบรรจุข้อมูลของการแปรผันทางพันธุกรรมที่มีค่าสถิติผ่านเกณฑ์ที่กำหนด

--out คือ ชื่อไฟล์ผลการวิเคราะห์ข้อมูล

ไฟล์ sample_output จะถูกนำมาคำนวณหาความถี่แอลลีลตามคำสั่งด้านล่าง

```
/path/to/plink --bfile /path/to/sample_output --freq --out MAF_check
```

โดย --freq คือ คำสั่งคำนวณความถี่แอลลีล (minor allele frequency; MAF)

นอกจากนี้ การแปรผันทางพันธุกรรมทุกตำแหน่งที่พบบนไฟล์ VCF จะถูกนำไปหาค่าความถี่แอลลีลของกลุ่มประชากรชาติต่าง ๆ จากฐานข้อมูล 1000 genomes อีกทั้งการแปรผันทาง

พันธุกรรมเหล่านี้จะถูกนำไปให้ข้อมูล ระดับความรุนแรงของการแปรผันทางพันธุกรรมที่มีผลต่อการทำงานของโปรตีนซึ่งจะถูกทำนายโดยใช้ Ensemble variant effect predictor web tools (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP) ถ้าหาก SNP และ indel ที่พบถูกระบุว่าเป็นการแปรผันทางเภสัชพันธุศาสตร์ที่รู้จักอยู่แล้ว หรือเคยมีการรายงานในฐานข้อมูล NCBI (<https://www.ncbi.nlm.nih.gov/snp>) หรือ PharmGKB (<https://www.pharmgkb.org>) เราจะระบุว่าเป็น known pharmacogenetic variant

3.4.2.3 การหาการแปรผันทางพันธุกรรมที่ก่อให้เกิดโรค (Pathogenic variants analysis)

ความถี่ของการแปรผันทางพันธุกรรมที่มีความรุนแรงในระดับสูง (high impact) จะถูกนำมาเปรียบเทียบกันระหว่างประชากรไทย (ที่พบ) และประชากรชาติต่าง ๆ จากฐานข้อมูล 1KGP ได้แก่ กลุ่มประชากรแอฟริกา (AFR), กลุ่มประชากรอเมริกา (AMR), กลุ่มประชากรเอเชีย (EAS), กลุ่มประชากรยุโรป (EUR), และกลุ่มประชากรเอเชียใต้ (SAS) ความแตกต่างกันของความถี่แอลลีลระหว่างแต่ละกลุ่มประชากร ถูกเปรียบเทียบโดยใช้การทดสอบสถิติ chi-square ด้วยโปรแกรม R เวอร์ชัน 4.0.3 (Lenth, 2016) เราควบคุมอัตราการค้นพบที่ผิดพลาด (false discovery rate) ที่ 0.05 โดยขั้นตอน Benjamini-Hochberg (Benjamini et al., 2001) และใช้ค่า p-value ที่ระดับนัยสำคัญเท่ากับ 0.01

บทที่ 4

ผลการวิจัย

4.1 ตัวอย่างที่ใช้ในการวิจัย

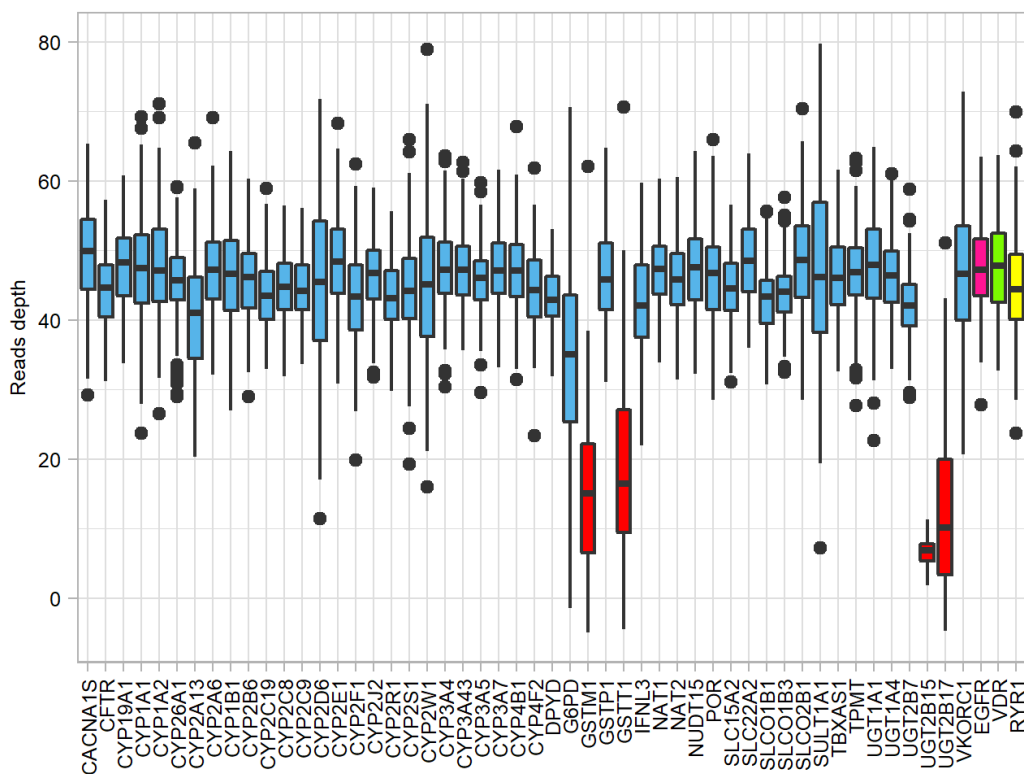
เราวิเคราะห์ข้อมูลทางพันธุกรรมของประชากรไทยที่ไม่เป็นโรค จำนวน 171 คน ประกอบด้วย เพศชาย 91 คน และเพศหญิง 80 คน ซึ่งเราทราบข้อมูลภูมิภาค 65.5% ของตัวอย่างโดยมีภูมิลำเนาจากภาคเหนือ 14 คน ภาคตะวันออกเฉียงเหนือ 15 คน ภาคกลาง 64 คน ภาคตะวันออก 11 คน ภาคใต้และภาคตะวันตก 8 คน และไม่ทราบภูมิลำเนาแน่ชัดจำนวน 59 คน (ตารางที่ 4.1) ซึ่งจะเห็นว่ากลุ่มตัวอย่างมาจากทุกภาคในประเทศไทย ด้วยจำนวนตัวอย่าง 171 คน เราจะได้ว่าขอบของความผิดพลาดสูงสุดสำหรับสัดส่วนของประชากรที่มีค่า star allele หนึ่ง ๆ มีค่าประมาณ 7.5% ที่ระดับความเชื่อมั่น 95%

ตารางที่ 6 ข้อมูลภูมิลำเนาของกลุ่มอาสาสมัครจำนวน 171 คน

| ภูมิลำเนา | เพศหญิง | เพศชาย | รวม |
|-----------------------|---------|--------|-----|
| ภาคเหนือ | 3 | 11 | 14 |
| ภาคตะวันออกเฉียงเหนือ | 10 | 5 | 15 |
| ภาคกลาง | 33 | 31 | 64 |
| ภาคตะวันออก | 5 | 6 | 11 |
| ภาคตะวันตกและภาคใต้ | 3 | 5 | 8 |
| ไม่ระบุ | 26 | 33 | 59 |
| รวม | 80 | 91 | 171 |

4.2 คุณภาพของลำดับนิวคลีโอไทด์ และค่าความลึกของ read (read depth)

ความยาว read ในแต่ละบุคคลมีค่าโดยเฉลี่ย 100 เบส มีปริมาณเบส G + เบส C อยู่ที่ 40-45% และมีคะแนนคุณภาพของลำดับเบส (base quality score) เฉลี่ยทุกเบสมากกว่า 30 ในทุกตัวอย่าง จำนวนความลึกของ read ที่พบในทุกลำดับยีนมีค่ามากกว่า 40x ยกเว้น *GSTM1*, *GSTT1*, *UGT2B15* และ *UGT2B17* ซึ่งมีค่าเฉลี่ยของความลึกของ read น้อยกว่า 40x (ภาพที่ 8) จากการเปรียบเทียบจำนวนค่าเฉลี่ยของความลึกของ read ของยีนควบคุม (*EGFR*, *VDR* และ *RYR1*) พบว่า *EGFR* มีค่าเฉลี่ยของความลึกของ read สูงที่สุด (ค่าเฉลี่ย \pm ส่วนเบี่ยงเบนมาตรฐาน) อยู่ที่ 43.91 ± 7.45 รองลงมาคือ *VDR* (43.30 ± 7.76) และ *RYR1* (40.39 ± 7.82)



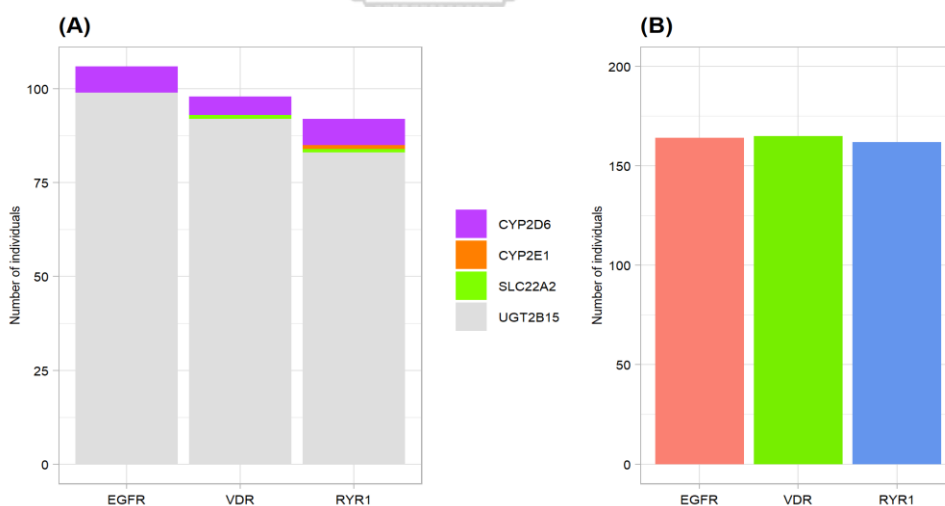
ภาพที่ 8 ค่าเฉลี่ยความลึกของ read เป็นภาพที่แสดงค่าเฉลี่ยของความลึกของ read ใน 51 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา (สีน้ำเงินและสีแดง) และยีนควบคุมทั้ง 3 ได้แก่ *EGFR* (สีชมพู), *VDR* (สีเขียว) และ *RYR1* (สีเหลือง) ในยีนที่แทนด้วยสีแดง เป็นยีนที่มีค่าเฉลี่ยของความลึกของ read น้อยกว่า 20 ได้แก่ *GSTM1*, *GSTT1*, *UGT2B15* และ *UGT2B17*

4.3 ยีนควบคุม

จากการเปรียบเทียบผลการทำนายรูปแบบการแปรผันทางพันธุกรรม ซึ่งรองรับในโปรแกรม Stargazer พบว่าทั้ง 3 ยีนควบคุม ให้ผลลัพธ์ที่ใกล้เคียงกันถึง 99.08% ผลการทำนายรูปแบบการแปรผันทางพันธุกรรม ระหว่าง *EGFR* เทียบกับ *VDR*, *EGFR* เทียบกับ *RYR1* และ *VDR* เทียบกับ *RYR1* พบว่ามีความเหมือนกัน 99.75%, 99.09% และ 99.31% ตามลำดับ

เราพบว่าเมื่อใช้ *EGFR* เป็นยีนควบคุม พบว่ามียีนเป้าหมาย 2 ยีนที่ไม่สามารถทำนายรูปแบบการแปรผันทางพันธุกรรมได้ ได้แก่ *UGT2B15* (n=99) และ *CYP2D6* (n=7) เมื่อเปลี่ยนมาใช้ *VDR* เป็นยีนควบคุม พบว่ามียีนเป้าหมาย 3 ยีน ที่ไม่สามารถทำนายรูปแบบการแปรผันทางพันธุกรรมได้ ได้แก่ *UGT2B15* (n=92), *CYP2D6* (n=5) และ *SLC22A2* (n=1) และเมื่อเปลี่ยนมาใช้ *RYR1* เป็นยีนควบคุม พบว่ามียีนเป้าหมาย 4 ยีน ที่ไม่สามารถทำนายรูปแบบการแปรผันทางพันธุกรรมได้ ได้แก่ *UGT2B15* (n=83), *CYP2D6* (n=8), *SLC22A2* (n=2) และ *CYP2E1* (n=1) (ภาพที่ 9A)

Stargazer สามารถทำนายรูปแบบการแปรผันทางพันธุกรรม ได้ใน 164, 165, และ 162 คน เมื่อใช้ *EGFR*, *VDR* และ *RYR1* เป็นยีนควบคุมตามลำดับ (ภาพที่ 9B) จากผลการทำนายพบว่า ยีน *UGT2B15* เป็นยีนที่มีจำนวนคนที่ไม่สามารถทำนายจีโนไทป์ได้มากที่สุด รองลงมาคือ *CYP2D6*, *SLC22A2* และ *CYP2E1* ผลลัพธ์ที่แสดงในรายงานเล่มนี้จะใช้ *EGFR* เป็นยีนควบคุม เนื่องจากมีจำนวนยีนที่สามารถทำนายได้มากที่สุด

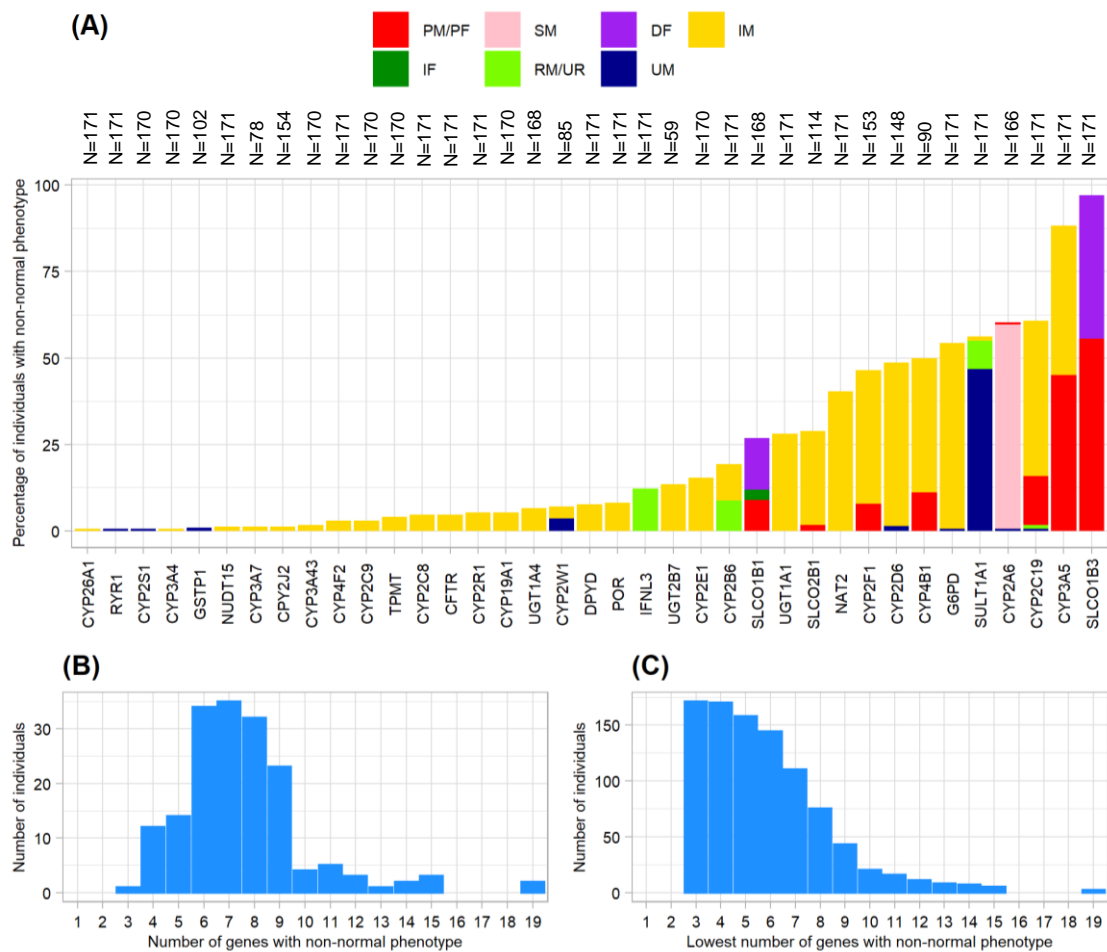


ภาพที่ 9 เปรียบเทียบผลการเรียก star allele จากการใช้ยีนควบคุมทั้ง 3 ยีน (*EGFR*, *VDR* และ *RYR1*) (A) แสดงจำนวนบุคคลที่ไม่สามารถทำนาย star allele ได้จากการใช้ยีนควบคุมแต่ละตัว ซึ่งพบในยีน *CYP2D6*, *CYP2E1*, *SLC22A2* และ *UGT2B15*. (B) แสดงจำนวนบุคคลที่สามารถทำนาย star allele ได้โดยใช้ยีนควบคุมแต่ละตัว (ไม่รวม *UGT2B15*) โดยที่สีชมพู สีเขียว สีน้ำเงิน หมายถึงยีนควบคุม *EGFR*, *VDR* และ *RYR1* ตามลำดับ

4.4 การทำนายความถี่ของแอลลีล จีโนไทป์ และฟีโนไทป์

ในภาพรวม Stargazer สามารถทำนายรูปแบบของ star allele ได้ทั้งหมด 196 star allele ซึ่งเป็นการแปรผันทางพันธุกรรมในกลุ่ม Phase I metabolizing genes (CYPs gene) จำนวน 105 แอลลีล, กลุ่ม Phase II metabolizing genes จำนวน 38 แอลลีล, กลุ่ม transporter gene จำนวน 19 แอลลีล และกลุ่ม drug target gene จำนวน 34 แอลลีล จากทั้งหมด 51 ยีน เราพบว่าประชากรมากกว่า 25% มีการแปรผันทางพันธุกรรมที่เปลี่ยนฟีโนไทป์จากปกติเป็นแบบอื่น ๆ ของยีนที่ตอบสนองต่อยาใน 13 ยีน ได้แก่ *SLCO1B3*, *CYP3A5*, *CYP2C19*, *CYP2A6*, *SULT1A1*, *G6PD*, *CYP4B1*, *CYP2D6*, *CYP2F1*, *NAT2*, *SLCO2B1*, *UGT1A1* และ *SLCO1B1* (ภาพที่ 10A) ในบางยีน stargazer ไม่สามารถสรุป star allele หรือไม่สามารถทำนายฟีโนไทป์ จากการแปรผันทางพันธุกรรมในบางคนได้ (unknown phenotype) ดังนั้นค่าเปอร์เซ็นต์ที่แสดงผลลัพธ์ในรูป 4.3 จะอ้างอิงจากจำนวนฟีโนไทป์ที่ให้ข้อมูลได้ (informative phenotype) กล่าวคือไม่นับจำนวนคนที่ไม่สามารถสรุปฟีโนไทป์หรือไม่ทราบฟีโนไทป์ได้





ภาพที่ 10 แผนภูมิแท่งแสดงยีนและจำนวนบุคคลที่มีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์แบบปกติ (A) เปอร์เซนต์ของบุคคลที่มีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์แบบปกติ. กลุ่มที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย (PM; สีแดง) หรือ กลุ่มที่มีเอนไซม์ที่ทำงานได้ไม่ดี (PF; สีแดง) ใน *SLCO1B1*, *SLCO1B3* และ *SLCO2B1*; กลุ่มที่มีความสามารถในการย่อยสลายยาได้ช้า (SM; สีชมพู); กลุ่มที่มีเอนไซม์ที่ลดการทำงานลง (DF; สีม่วง); กลุ่มที่มีความสามารถในการย่อยสลายยาได้ปานกลาง (IM; สีเหลือง); กลุ่มที่มีเอนไซม์ที่เพิ่มการทำงาน (IF; สีเขียวเข้ม); กลุ่มที่มีความสามารถในการย่อยสลายยาได้รวดเร็ว (RM; สีเขียว) หรือ กลุ่มที่มีเอนไซม์ที่มีการตอบสนองที่ไม่ปกติ (UR; สีเขียว) สำหรับ *IFNL3*; กลุ่มที่มีความสามารถในการย่อยสลายยาได้รวดเร็วมากที่สุด (UM; สีน้ำเงินเข้ม) (B) จำนวนบุคคลที่มีจำนวนยีนที่มีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์แบบปกติ จำนวนต่าง ๆ (C) จำนวนบุคคลที่มีจำนวนยีนที่มีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์แบบปกติ อย่างน้อย จำนวนต่าง ๆ

ในยีน *SLCO1B3* พบว่ามี 5 คน (2.9%) ที่มีการแสดงออกของยีนแบบปกติ (*SLCO1B3*1/*1*) ในขณะที่กลุ่มตัวอย่าง 166 (97.1%) คน มีการแปรผันทางพันธุกรรมแบบ *SLCO1B3*S1* ซึ่งเป็นแอลลีลที่ลดการทำงาน (decreased function allele) ในจำนวนนี้มี 71 (41.5%) คน ที่เป็น *SLCO1B3 *1/*S1* และมีลักษณะการทำงานของยีนแบบลดการทำงาน (decreased function) ในขณะที่อีก 95 (55.6%) คน มีการทำงานของยีนไม่ดี (poor function) (*SLCO1B3*S1/*S1* หรือ **S1/DEL*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 3)

*CYP3A5*3* ซึ่งเป็นแอลลีลแบบไม่ทำหน้าที่ (no-function allele) ในยีน *CYP3A5* ถูกพบใน 74 คน (43.3%) ซึ่งมีทีโนไทป์แบบที่มีความสามารถในการย่อยสลายยาได้ปานกลาง (*CYP3A5*1/*3*) และพบในอีก 77 คน (45.0%) ที่มีทีโนไทป์แบบมีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย (*CYP3A5*3/*3*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 1)

สำหรับยีน *CYP2C19* พบว่า 77 คน (45.0%) มีความสามารถในการย่อยสลายยาได้ปานกลาง ซึ่งมาจากการมีแอลลีลที่เกิดการขาดหายไปของยีน (**DEL*) หรือ แอลลีลที่ไม่ทำหน้าที่ *CYP2C19 *2, *3* หรือ **5* กลายเป็นจีโนไทป์ได้แก่ *CYP2C19*1/*2, CYP2C19*1/*3, CYP2C19*1/*5* และ *CYP2C19*1/*DEL* นอกจากนี้ 24 คน (14%) ถูกทำนายว่ามีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย โดยเกิดจากแอลลีลที่ไม่ทำหน้าที่ *CYP2C19 *2* และ **3* (*CYP2C19*2/*2, *2/*3* และ **3/*3*) สำหรับการแปรผันทางพันธุกรรมแบบ *CYP2C19 *17* มีผลเสริมการทำงานของแอลลีล (increased function allele) จากการทำนายของ Stargazer พบว่า มี 2 คนที่มีความสามารถในการย่อยสลายยาได้รวดเร็ว (*CYP2C19*1/*17*) และมีเพียง 1 คน ที่มีความสามารถในการย่อยสลายยาได้เร็วมากที่สุด (*CYP2C19*17/*17*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 1)

59% (n=98) ของกลุ่มตัวอย่างที่เราศึกษามีการแสดงออกของยีน *CYP2A6* แบบที่มีความสามารถในการย่อยสลายยาได้ช้า ซึ่งเกิดจากการมีแอลลีลที่ลดการทำงาน อย่างน้อย 1 แอลลีล (*CYP2A6*7, *9, *10, *11, *12, *19, *21* และ **35*) หรือ แอลลีลที่ไม่ทำหน้าที่ (*CYP2A6*4*) โดยจีโนไทป์ที่พบคือ *CYP2A6*1/*4, *1/*7, *1/*9, *1/*10, *1/*12, *1/*19, *1/*35, *4/*7, *4/*9, *4/*11, *4/*35, *7/*12, *7/*21, *10/*11* และ **10/*35* ในการศึกษาพบว่า มี 1 คนที่มีการแสดงออกของยีนแบบที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย (*CYP2A6*4/*4*) และอีก 1 คน ที่มีการแสดงออกของยีนแบบที่มีความสามารถในการย่อยสลายยาได้เร็วมากที่สุด (*CYP2A6*1/*1x2*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 1)

ในยีน *SULT1A1* พบว่ามีหลายแอลลีลที่มีจำนวนชุดของยีนมากกว่า 2 ชุด ได้แก่ *SULT1A1*1/*1x2, *1/*1x3, *1/*1x4, *1/*1x5, *1x2/*1x2, *1x2/*1x3, *1x2/*1x6, *1x2/*2x2, *1x2/*2x3, *1x3/*1x6* และ **1x3/*2x3* โดยพบว่ากลุ่มตัวอย่าง 80 คน (46.8%) ที่มีแอลลีลเหล่านี้ มีความสามารถในการย่อยสลายยาได้เร็วมากที่สุด มี 14 คน (8.2%) ถูกพบว่าแอลลีลแบบ *SULT1A1*1x2/*2* ซึ่งมีความสามารถในการย่อยสลายยาได้รวดเร็ว นอกจากนี้เราพบว่ามี 2 คน ที่มีความสามารถในการย่อยสลายยาได้ปานกลาง (*SULT1A1*1/*S1* และ **2/*2*) ที่ประกอบไปด้วยแอลลีลที่ไม่ทำหน้าที่ (*SULT1A1*S1*) และแอลลีลที่ลดการทำงาน (*SULT1A1*2*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 2)

แอลลีลที่มีการทำงานแบบปกติ ($G6PD^*$ 1) และแอลลีลที่ไม่ทำหน้าที่ ($G6PD^*$ DEL) เป็นรูปแบบการแปรผันทางพันธุกรรมส่วนใหญ่ของกลุ่มตัวอย่างนี้ โดยพบมากถึง 66.1% และ 26.6% ตามลำดับ จากการศึกษาพบ 92 (53.8%) คน มีการแสดงออกของยีน $G6PD$ แบบมีความสามารถในการย่อยสลายยาได้ปานกลาง ซึ่งเป็นผลมาจากแอลลีลที่ไม่ทำหน้าที่ ($G6PD^*$ DEL) หรือ แอลลีลที่ลดการทำงาน ($G6PD^*$ 8, *21, *28, *31,*50 และ*51) ซึ่งพบได้ใน $G6PD^*$ 1/*DEL, *21/*DEL, *28/*DEL, *51/*51, *51/*DEL และ *8/*DEL จากรูปแบบการแปรผันทางพันธุกรรมทั้งหมดเราพบว่ารูปแบบการแปรผันแบบ $G6PD^*$ 1/*DEL เป็นรูปแบบที่พบมากที่สุด (49.1%) รองลงมาคือ $G6PD^*$ 1/*1 (36.8%) นอกจากนี้เราพบว่ามี 1 คน มีความสามารถในการย่อยสลายยาได้รวดเร็วมากที่สุด ($G6PD^*$ 1/*1x2) และสืบเนื่องจากเพศชายเป็นเพศที่มีโครโมโซม X เพียงแท่งเดียว ดังนั้นเพศชายจึงถูกพบว่าเป็นเพศที่มีการแปรผันทางพันธุกรรมแบบ $G6PD^*$ 1/*DEL (ภาพที่ 10A, ภาคผนวก ภาพที่ 2)

ในยีน $CYP4B1$ พบว่ากลุ่มตัวอย่างจำนวน 35 คน (38.9%) มีความสามารถในการย่อยสลายยาได้ปานกลาง ($CYP4B1^*$ 1/*2) และ 10 คน (11.1%) มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย ($CYP4B1^*$ 2/*2 หรือ *S1/*2) (ภาพที่ 10A, ภาคผนวก ภาพที่ 4)

$CYP2D6$ เป็นอีกยีนหนึ่งที่มีความหลากหลายของรูปแบบการแปรผันทางพันธุกรรมสูงมาก ในการศึกษาครั้งนี้เราพบรูปแบบของแอลลีลที่ต่างกันมากถึง 19 รูปแบบ ซึ่งประกอบด้วย แอลลีลที่เกี่ยวข้องกับการเพิ่มจำนวนชุดของยีน ($CYP2D6^*$ 1x2, *2x2, *2x3, *10x2 และ *39x2), การขาดหายไปของจำนวนชุดของยีน ($CYP2D6^*$ 5) และ การจัดเรียงใหม่ของยีน (gene rearrangement) ($CYP2D6^*$ 36+10 และ *36x3+10) สำหรับแอลลีลที่ลดการทำงาน $CYP2D6^*$ 36+*10 และ $CYP2D6^*$ 10 เป็นแอลลีลที่ถูกพบมากที่สุดในกลุ่มตัวอย่าง มี 62 คนที่มีความสามารถในการย่อยสลายยาได้ปานกลาง ($CYP2D6^*$ 10/*10, *10/*36+*10, *10/*36x3+*10, *10/*41, *36+*10/*36+*10, *36+*10/*41, *4/*10, *4/*36+*10, *5/*10, และ *5/*36+*10) นอกจากนี้ยังพบความสามารถในการย่อยสลายยาได้ปานกลางในจีโนไทป์ $CYP2D6^*$ 1/*4, *1/*5, *2/*5, *4/*41, *5/*39 และ *5/*41 ซึ่งเกิดจากการมีแอลลีลที่ลดการทำงาน ($CYP2D6^*$ 41) และแอลลีลที่ไม่ทำหน้าที่ ($CYP2D6^*$ 4 และ $CYP2D6^*$ 5) ในกลุ่มตัวอย่าง 7 คน และมีอีก 2 คน ที่มีความสามารถในการย่อยสลายยาได้รวดเร็วมากที่สุด ($CYP2D6^*$ 1/*2x2 และ $CYP2D6^*$ 1/*2x3) (ภาพที่ 10A, ภาคผนวก ภาพที่ 1)

$CYP2F1^*$ 2 เป็นแอลลีลที่ไม่ทำหน้าที่ ส่งผลให้มีความสามารถในการย่อยสลายยาได้ปานกลาง ($CYP2F1^*$ 1/*2) ใน 59 คน (38.6%) และมีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย ($CYP2F1^*$ 2/*2) ใน 12 คน (7.8%) (ภาพที่ 10A, ภาคผนวก ภาพที่ 1)

ในยีน $NAT2$ เราพบว่าประชากรส่วนใหญ่ ($n=69$; 40.4%) มีความสามารถในการย่อยสลายยาได้ปานกลาง ($NAT2^*$ 1/*DEL, *5/*6, *5/*7, *5/*DEL, *6/*6, *6/*7, *6/*DEL, *7/*7, และ *7/*DEL) ซึ่งเกิดจากการมีแอลลีลที่ลดการทำงาน $NAT2^*$ 5, *6, *7 และ *DEL (ภาพที่ 10A, ภาคผนวก ภาพที่ 2)

กลุ่มตัวอย่างส่วนใหญ่ (71.1%) มีการทำงานของยีน $SLCO2B1$ แบบปกติ ($SLCO2B1^*$ 1/*1) ในขณะที่กลุ่มตัวอย่าง 57 คน พบว่ามีแอลลีลที่ไม่ทราบหน้าที่ (unknown function allele) ($SLCO2B1^*$ S464F) ที่ส่งผล

ให้มีฟีโนไทป์แบบที่ไม่ทราบการทำงาน (*SLCO2B1*1/*S464F*, **S1/*S464F* และ **S464F/*S464F*) เราพบว่ามี 31 (27.2%) คน ที่มีแอลลีลที่ไม่ทำหน้าที่ (*SLCO2B1*S1*) ซึ่งส่งผลให้มีความสามารถในการย่อยสลายยาได้ปานกลาง (*SLCO2B1*1/*S1*) และมี 2 (1.8%) คนที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย (*SLCO2B1*S1/*S1*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 3)

ในยีน *UGT1A1* พบแอลลีลที่ลดการทำงาน (*UGT1A1*6*, **7*, **27* และ **60*) ใน 48 คน (28%) และทำให้มีความสามารถในการย่อยสลายยาได้ปานกลาง (*UGT1A1*27/*27*, **27/*60*, **6/*27*, **6/*6*, **6/*60*, **6/*7* และ **60/*60*) (ภาพที่ 10A, ภาคผนวก ภาพที่ 2)

จากการศึกษาการทำงานของยีน *SLCO1B1* พบว่ามีแอลลีลที่ทำงานไม่ได้ดี , แอลลีลที่ลดการทำงาน, แอลลีลทำงานปกติ และแอลลีลที่เสริมการทำงาน โดยในกลุ่มตัวอย่างมี 15 (8.9%) คน มีแอลลีลที่ทำงานไม่ได้ดี (*SLCO1B1*15/*17*, **15/*DEL*, **17/*DEL* และ **1B/*DEL*) เนื่องจากมีแอลลีลที่ลดการทำงาน (*SLCO1B1*15* และ **17*) และแอลลีลที่ไม่ทำหน้าที่ (*SLCO1B1*DEL*) และมี 25 (14.3%) คน มียีนที่ลดการทำงาน (*SLCO1B1*1/*15*, **1/*17*, **1B/*15*, **1B/*17* และ **35/*DEL*) โดย *SLCO1B1*1B* เป็นแอลลีลที่ทำงานปกติ และ *SLCO1B1*35* เป็นแอลลีลที่เสริมการทำงาน นอกจากนี้เรายังพบว่ามี 5 (3%) คน มียีนที่เสริมการทำงาน (*SLCO1B1*1/*35*, **14/*35* และ **1B/*35*) โดย *SLCO1B1*14* เป็นแอลลีลที่ลดการทำงาน (ภาพที่ 10A, ภาคผนวก ภาพที่ 3)

ในยีน *GSTM1*, *GSTT1*, *UGT2B17* และ *UGT2B15* พบว่ามีความถี่แอลลีลของแอลลีลที่ไม่ทำหน้าที่สูงมาก โดยแอลลีลที่ไม่ทำหน้าที่เหล่านี้เกิดจากการขาดหายไปของยีน สำหรับลักษณะฟีโนไทป์ที่เกิดขึ้นจากแอลลีลที่ไม่ทำหน้าที่ในยีน *GSTM1*, *GSTT1* และ *UGT2B17* ถูกพบว่าเป็นแบบที่มีความสามารถในการย่อยสลายยาได้ปานกลาง และน้อยมากหรือไม่ได้เลย ในขณะที่รูปแบบของฟีโนไทป์ของแอลลีลที่พบในยีน *UGT2B15* เป็นแบบไม่ทราบการทำงานที่ชัดเจน และมีฟีโนไทป์แบบที่ไม่ทราบการทำงาน (ภาคผนวก ภาพที่ 2) จากการศึกษาของเราพบว่าทั้ง 4 ยีนนี้มีความถี่ของ read ต่ำมาก ๆ ซึ่งอาจเป็นคำอธิบายการพบความถี่ของแอลลีลที่เป็นการขาดหายไปของยีนในระดับสูง และควรระวังในการแปลผลสำหรับ 4 ยีนนี้ (ภาคผนวก ภาพที่ 2)

ใน *CACNA1S* พบว่ากลุ่มตัวอย่าง 171 คน (100%) มีความสามารถในการย่อยสลายยาได้ในระดับปกติ(*CACNA1S*1/*1*) (ภาคผนวก ภาพที่ 4) อีกทั้ง ในยีน *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP2A13* (ภาคผนวก ภาพที่ 1), *NAT1* (ภาคผนวก ภาพที่ 2), *SLC15A2*, *SLC22A2* (ภาคผนวก ภาพที่ 3) และ *TBXAS1* (ภาคผนวก ภาพที่ 4) พบว่าทุกคนมีฟีโนไทป์ที่เป็นปกติ หรือมีฟีโนไทป์แบบที่ไม่ทราบการทำงาน

ประชากรแต่ละคนส่วนใหญ่พบว่ามีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์ปกติใน 6-8 ยีนที่เกี่ยวข้องต่อการตอบสนองต่อยา (ภาพที่ 10A) โดยพบว่าในทุก ๆ คนจะมีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์แบบปกติอย่างน้อยที่สุด 3 ยีน และมากที่สุดไม่เกิน 19 ยีน เราพบว่ามี 158 จาก 171 คน มีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์ปกติอย่างน้อย 5 ยีนขึ้นไป

4.5 การเปรียบเทียบความถี่ของแอลลีลในยีนที่เกี่ยวข้องกับการตอบสนองต่อยา

เราสามารถระบุ SNV ได้ทั้งหมด 28,061 ตัว ในประชากร 171 คน โดยการแปรผันทางพันธุกรรมส่วนใหญ่ที่พบเป็นการแปรผันทางพันธุกรรมที่เกิดขึ้นในบริเวณ intron (intron variant) การแบ่งประเภทของการแปรผันทางพันธุกรรม ในการศึกษาที่นี้แบ่งการแปรผันทางพันธุกรรม ออกเป็นประเภทตามการส่งผลกระทบต่อโปรตีน เช่น frameshift, missense และ synonymous variant ตามตารางที่ 7

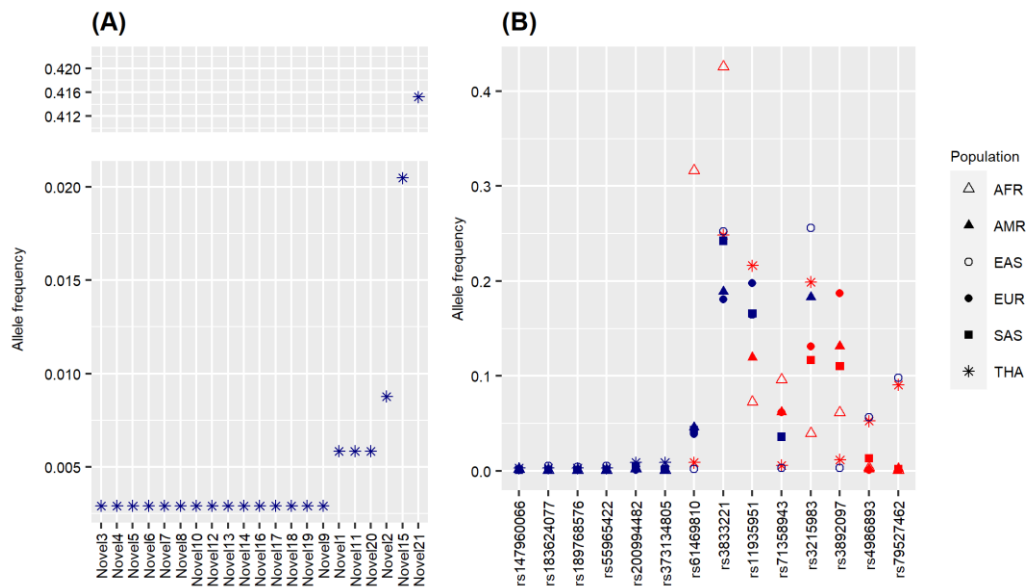
ตารางที่ 7 ตารางแสดงผลกระทบระดับต่าง ๆ ของการแปรผันทางพันธุกรรมที่พบในกลุ่มตัวอย่าง

| ระดับของผลกระทบต่อยีน | การแปรผันทางพันธุกรรมตามตำแหน่งที่เกิด | จำนวนการแปรผันที่ทราบ (known variant) | จำนวนการแปรผันที่ยังไม่มีการรายงาน (novel variants [#]) | รวม | เปอร์เซ็นต์ |
|--|--|---------------------------------------|---|--------|-------------|
| ระดับสูง (high impact) | frameshift_variant | 3 | 12 | 15 | 0.05 |
| | splice_acceptor_variant | 2 | 0 | 2 | 0.01 |
| | splice_donor_variant | 4 | 4 | 8 | 0.03 |
| | stop_gained | 5 | 5 ^{&} | 10 | 0.04 |
| ระดับปานกลาง (moderate) | inframe_deletion | 1 | 6 | 7 | 0.02 |
| | inframe_insertion | 1 | 3 | 4 | 0.01 |
| | missense_variant | 233 | 182 | 415 | 1.48 |
| ระดับต่ำ (low) | intron_variant | 34 | 31 | 65 | 0.23 |
| | stop_retained_variant | 0 | 1 | 1 | 0.00 |
| | synonymous_variant | 195 | 78 | 273 | 0.97 |
| ระดับที่ยากต่อการทำนายผลกระทบ (modifier) | 3_prime_UTR_variant | 155 | 147 | 302 | 1.08 |
| | 5_prime_UTR_variant | 46 | 31 | 77 | 0.27 |
| | downstream_gene_variant | 1,082 | 973 | 2,055 | 7.32 |
| | intron_variant | 12,067 | 10,634 | 22,701 | 80.90 |
| | non_coding_transcript_exon_variant | 3 | 3 | 6 | 0.02 |
| | upstream_gene_variant | 1,125 | 995 | 2,120 | 7.55 |

[#]novel = ไม่พบการรายงานใน 1000 Genomes Project Phase 3

[&] Novel6 เป็นการแปรผันทางพันธุกรรมที่ถูกรายงานอยู่บนฐานข้อมูล dbSNP ในประชากร Singaporean and Malaysian (https://www.ncbi.nlm.nih.gov/projects/SNP/snp_retrieve.cgi?subsnp_id=657184749) โดยมีหมายเลขคือ rs370320936

เราพบว่ามี SNV 322 ตัว ที่มีผลกระทบของความรุนแรงอยู่ในระดับสูง, ปานกลาง, และต่ำ นอกจากนี้เรายังพบว่ามี SNV 21 จาก 322 ตัว (Novel1 – Novel21) มีผลกระทบของความรุนแรงอยู่ในระดับสูง และยังไม่มีความรู้พื้นฐานการรายงานบนฐานข้อมูล 1000 Genomes Project Phase 3 มีเพียง Novel 6 ที่พบหลักฐานการรายงานบนฐานข้อมูล dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/snp_retrieve.cgi?subsnp_id=657184749) ในประชากร Singaporean and Malaysian ดังนั้นจึงมี SNV 20 ตัว ที่เป็นการแปรผันทางพันธุกรรมรูปแบบใหม่ โดยความถี่แอลลีลของ SNV 6 ตัว ซึ่งอยู่ในยีน *DPYD* (rs189768576), *CYP3A5* (rs373134805, rs55965422), *SLCO1B1* (rs200994482, rs183624077) และ *CYP2D6* (rs147960066) ถูกพบว่าไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติระหว่างแต่ละกลุ่มประชากร สำหรับ SNV 8 ตัว ที่เหลือ ความถี่แอลลีลถูกพบว่ามีความแตกต่างกันระหว่างประชากรไทย และประชากรกลุ่มอื่นๆ ในจำนวนนี้เราพบ SNV 3 ตัว (rs4986893, rs79527462 และ rs3892097) ที่ความถี่แอลลีลระหว่างประชากรเอเชีย (ประชากรไทย ประชากรจากเอเชียตะวันออกเฉียงและเอเชียใต้) ไม่มีความแตกต่างกัน แต่จะมีความแตกต่างกันกับประชากรจากชาติอื่น ๆ อีก 4 ชาติ (แอฟริกัน, ประชากรผสมในทวีปอเมริกา, ยุโรป) อีกหนึ่ง SNV (rs71358943) ถูกพบว่าไม่มีความแตกต่างระหว่างประชากรเอเชีย แต่มีความแตกต่างกันกับประชากรแอฟริกัน, ประชากรผสมในทวีปอเมริกา และยุโรป ในการศึกษาครั้งนี้เราพบ SNV 2 ตัว (rs61469810 และ rs3833221) ที่มีความถี่แอลลีลเท่ากันในทุก ๆ ประชากร ยกเว้นประชากรจากกลุ่มแอฟริกาที่พบว่ามีความถี่แอลลีลใน SNV 2 ตัว นี้สูงกว่ากลุ่มชาติพันธุ์อื่น สำหรับ rs3215983 พบว่าประชากรไทยมีความถี่แอลลีลแตกต่างจากกลุ่มประชากรแอฟริกัน เอเชียใต้และยุโรป ในขณะที่ความถี่แอลลีลของ rs11935951 มีแตกต่างกันระหว่างกลุ่มประชากรเอเชียกับกลุ่มประชากรแอฟริกาและประชากรผสมในทวีปอเมริกา (ภาพที่ 11)



ภาพที่ 11 การเปรียบเทียบความถี่แอลลีลใน 35 การแปรผันทางพันธุกรรมที่ส่งผลกระทบต่อระดับสูงระหว่างประชากรไทย (THA) ประชากรแอฟริกัน (AFR) ประชากรผสมในทวีปอเมริกา (AMR) ประชากรเอเชียตะวันออก (EAS) ประชากรยุโรป (EUR) และประชากรเอเชียใต้ (SAS) โดยอ้างอิงจากฐานข้อมูล 1000 Genome Project (A) ความถี่ allele ของการแปรผันทางพันธุกรรมที่ไม่ปรากฏในฐานข้อมูล 1000 Genome Project แต่พบในประชากรไทย แสดงด้วย Novel1-Novels21 (B) สัญลักษณ์ Δ , \blacktriangle , \circ , \bullet , \blacksquare , $*$ หมายถึง ความถี่อัลลีลของ AFR, AMR, EAS, EUR, SAS และ THA ตามลำดับ ความถี่แอลลีลที่แตกต่างกันอย่างมีนัยสำคัญ ($p < 0.05$) ระหว่าง THA และประชากรอื่น ๆ ถูกแสดงด้วยสีแดง (Δ , \blacktriangle , \circ , \bullet , \blacksquare , $*$) และความถี่แอลลีลที่ไม่มี ความแตกต่างทางสถิติ จะแสดงด้วยสีน้ำเงิน (Δ , \blacktriangle , \circ , \bullet , \blacksquare , $*$)

บทที่ 5

อภิปรายผลการวิจัย

การศึกษาในครั้งนี้ได้รายงานรูปแบบการแปรผันทางพันธุกรรม (star allele), diplotype, และพีโนไทป์ ใน 51 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยาโดยใช้รหัสพันธุกรรมทั้งจีโนมในประชากรไทยที่ไม่เป็นโรคจำนวน 171 คน เปรียบเทียบกับผลการศึกษาก่อนหน้านี้ในกลุ่มประชากรไทยที่เป็นโรคหลอดเลือด ซึ่งมักจะพบได้มากในแถบภาคตะวันออกเฉียงเหนือของไทย (ตารางที่ 4.1) การศึกษาของเราได้ผู้เข้าร่วมมาจากทั่วประเทศ แต่รับข้อมูลมาจากโรงพยาบาลกลางกรุงเทพมหานครเพียงแห่งเดียวเนื่องจากนักพันธุศาสตร์คลินิก 20 คนอยู่ที่กรุงเทพฯ ทั้งหมดยกเว้นเพียงหนึ่งคน ดังนั้นผู้เข้าร่วมการศึกษาซึ่งมาจากคลินิกพันธุศาสตร์ของโรงพยาบาลที่เป็นศูนย์ประสานงานรับและส่งต่อผู้ป่วยระดับตติยภูมิมาจากทั่วประเทศ (Shotelersuk et al., 2014; Shotelersuk et al., 2019) นอกจากนี้การศึกษาของเรายังศึกษาเพิ่มในอีก 26 ยีนที่ยังไม่มีการรายงานจากการศึกษาก่อนหน้านี้ (Mauleekoonphairoj et al., 2020) ซึ่งยีนทั้ง 26 ยีนนี้มีความเกี่ยวข้องกับยาที่ได้รับการอนุมัติโดยสำนักงานคณะกรรมการอาหารและยา (อย.), กระทรวงสาธารณสุขประเทศไทย (<https://porta.fda.moph.go.th>) และเป็นยาที่ถูกส่งจ่ายสำหรับรักษาโรคที่เป็นสาเหตุการตาย 10 ลำดับแรกของคนไทย (ตารางที่ 5)

จากการศึกษาอื่นที่เกี่ยวข้องกับการตอบสนองต่อยาทั้งหมด 51 ยีน พบว่า มี 40 ยีนที่มีอย่างน้อย 1 คน มีพีโนไทป์ที่ไม่ใช่พีโนไทป์ปกติ การค้นพบนี้แสดงให้เห็นถึงความต้องการความรู้ด้านการปรับปริมาณยา การรักษาที่แตกต่างกันในแต่ละบุคคลหรือทางเลือกการรักษาแบบอื่น ข้อมูลจากการศึกษาของเราชี้ให้เห็นว่าไม่มีใครเลยที่จะมียีนที่มีพีโนไทป์ปกติทั้ง 51 ยีน ซึ่งในประเด็นนี้ได้เน้นย้ำให้เห็นถึงความสำคัญของการทดสอบทางเภสัชพันธุศาสตร์ที่อาจเป็นประโยชน์ต่อการสร้างวิธีการรักษาที่จำเพาะรายบุคคลได้ นอกจากนี้เรายังได้รายงานความถี่แอลลีลที่มีผลกระทบระดับสูงต่อการทำงานของยีน โดยเราได้ทำการเปรียบเทียบความถี่แอลลีลนี้กับอีก 5 กลุ่มประชากร และเราพบว่าการแปรผันทางพันธุกรรม 20 ตัวที่ไม่เคยมีการรายงานที่ไหนดมาก่อน ดังนั้นการศึกษานี้จึงให้รายละเอียดเกี่ยวกับยีนที่ควรรวมอยู่ในรายชื่อยีนทางเภสัชพันธุศาสตร์เบื้องต้นที่ควรตรวจสอบสำหรับประชากรไทย

13 ยีนที่ 25% ของกลุ่มตัวอย่างเกิดการแปรผันที่ส่งผลต่อการแสดงออกทางพีโนไทป์ของยีน เป็นยีนที่เกี่ยวข้องกับ การดูดซึม, การแพร่กระจาย, การย่อยสลาย หรือ การขจัดยาต่าง ๆ ออกจากร่างกาย ดังนั้นความ

จำเป็นในการเปลี่ยนแปลงการจัดการทางคลินิกขึ้นกับข้อมูลเกี่ยวกับยีนและยา ในทางตรงกันข้ามก็มียีนบางยีน อย่างเช่น *CYP26A1*, *CACNA1S*, หรือ *TBXAS1* ที่กลุ่มที่ประชากรไทยส่วนใหญ่มีลักษณะทีโนไทป์แบบปกติ

ในกลุ่มประชากร 166 คน พบว่ามีการทำงานของแอลลีลแบบลดการทำงานลง และแบบที่ทำงานได้น้อยมากหรือแทบไม่ได้เลยใน *SLCO1B3* เนื่องจาก *SLCO1B3*51* โดยการแปรผันทางพันธุกรรมรูปแบบนี้มีตำแหน่ง SNP หลักของ *SLCO1B3*51* คือ rs7311358 (c.699G>A) มีการศึกษาได้วิเคราะห์ผล SNP เหล่านี้ร่วมกับ rs4149117 (c.334T>G) พบว่าจีโนไทป์ rs7311358 AA +AG และ rs4149117 GG+GT มีความเกี่ยวข้องกับความน่าจะเป็นที่สูงขึ้นที่จะไม่ตอบสนองต่อปริมาณยามาตรฐานของ imatinib ในการรักษาโรคมะเร็งเม็ดเลือดขาวชนิดมัยอีลอยด์เรื้อรัง (de Lima et al., 2015) นอกจากนี้ผู้ที่มีแฮพลอไทป์แบบพาหะของ rs4149117(G)-rs7311358(A) จะมีการดูดซึมยา mycophenolic acid glucuronide เข้าสู่เซลล์ได้น้อยลงอย่างมีนัยยะสำคัญทางสถิติ เมื่อเทียบกับผู้ที่มีแฮพลอไทป์ตามลำดับอ้างอิง (Picard et al., 2010)

กลุ่มยีน *CYP3A* เป็นหนึ่งในยีนตัวสำคัญของกลุ่มยีน *CYP* ยีนนี้จะถูกแปลรหัสไปเป็นเอนไซม์ที่เกี่ยวข้องกับการย่อยสลายของยา ไม่ว่าจะเป็น nifedipine, cyclosporine, tacrolimus, erythromycin, midazolam, alprazolam และ triazolam (Bjornsson et al., 2003; Thervet et al., 2003) สำหรับยีน *CYP3A5* การแปรผันทางพันธุกรรมแบบ *CYP3A5*3* ส่งผลให้โปรตีนไม่สามารถแสดงออกได้เนื่องจากเกิดการ splicing ที่ผิดปกติบน mRNA (Liu et al., 2019)ตามที่มีรายงาน *CYP3A5*3* (rs776746) พบมากในชาวคอเคเซียน (85–95%), ชาวเม็กซิกัน (75%), และชาวเอเชีย (65–85%) แต่พบได้น้อยชาวแอฟริกันอเมริกัน (27–55%) (Zhang et al., 2013) สำหรับข้อมูลของชาวเอเชียนั้นตรงกับผลการศึกษาของเรา (นั่นคือ 66.7%) ดังนั้นเพื่อที่จะให้ได้ปริมาณความเข้มข้นของยาที่ให้ผลการรักษาอย่างมีประสิทธิภาพ กลุ่มที่มีความสามารถในการย่อยสลายยาได้ปานกลาง เช่น *CYP3A5*1*3* นั้นถูกแนะนำให้ได้รับการจ่ายยาแบบเพิ่มปริมาณความเข้มข้นของยา tacrolimus (เพิ่มเป็น 1.5–2 เท่าของความเข้มข้นปกติ) ในทางกลับกัน กลุ่มที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย เช่น *CYP3A5*3*3* ควรได้รับการสั่งจ่ายยาในปริมาณความเข้มข้นมาตรฐานตามเอกสารกำกับยา tacrolimus (Birdwell et al., 2015)

CYP2C19 นั้นเป็นที่รู้จักกันว่าเกี่ยวข้องกับการเมแทบอลิซึมสารที่สำคัญบางตัว เช่น omeprazole, imipramine, และ diazepam รูปแบบการแปรผันทางพันธุกรรมแบบ *CYP2C19*17* นั้นส่วนใหญ่พบในชาวยุโรป, ชาวแอฟริกัน และประชากรผสมในทวีปอเมริกาในขณะที่ *CYP2C19*2* นั้น สามารถพบเจอได้ทั่วไปในชาวเอเชีย (Zhou et al., 2017) จากการศึกษาของเราพบว่าความถี่แอลลีล *CYP2C19*2* ของประชากรไทยมีค่าสูง ในผู้ป่วยที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย (รวมถึง *CYP2C19*2*2*) สามารถกำจัด *Heterobacter pylori* ได้อย่างมีประสิทธิภาพเมื่อรักษาด้วย omeprazole และ amoxicillin เพราะคนเหล่านี้มีปริมาณยาที่ยังอยู่ในกระแสเลือดได้ยาวนานกว่า คนที่มีความสามารถในการย่อยสลายยาได้ในระดับปกติ อีกทั้งคนเหล่านี้ยังสามารถหลีกเลี่ยงการดื้อยาในระหว่างการรักษาได้อีกด้วย (Sagar et al., 2000; Tanigawara et al.,

1999) สำหรับการสั่งจ่ายยา citalopram และ escitalopram ในคนที่มีความสามารถในการย่อยสลายยาได้ปานกลาง เช่นใน *CYP2C19*1/*2* มีคำแนะนำให้เริ่มการรักษาโดยระดับยาเริ่มต้น ในขณะที่คนที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย เช่น *CYP2C19*2/*2* ควรได้รับความเข้มข้นของยาเพียง 50% ของปริมาณยา ระดับเริ่มต้น หรือ ควรพิจารณาทางเลือกใหม่สำหรับการใช้ยาที่ไม่มีส่วนเกี่ยวข้องกับการทำงานของเอนไซม์ *CYP2C19* (Hicks et al., 2015)

สำหรับยีน *CYP2A6* นั้นเกี่ยวข้องกับการย่อยสารที่จำเป็นต่อร่างกาย (แต่ร่างกายไม่สามารถสร้างสารนั้นเองได้) เป็นจำนวนมาก และรับผิดชอบการย่อยยาประมาณ 3% ของยาที่ถูกย่อยโดยเอนไซม์กลุ่ม *CYP* แอลลีล *CYP2A6*4, *7, *10, *11, *19* และ **35* นั้นเป็นแอลลีลที่พบได้มากในชาวเอเชียตะวันออก ในขณะที่ *CYP2A6*14, *28* และ **34* พบได้มากในชาวคอเคเซียน (López-Flores et al., 2017) ผลการศึกษาของเราพบว่า 3-12% ของกลุ่มประชากรมีแอลลีลที่ลดการทำงาน (*CYP2A6*4, *7, *9* และ **10*) โดยคนที่มีแอลลีลแบบทั้ง homozygous และ heterozygous ถูกทำนายว่าเป็นกลุ่มที่มีความสามารถในการย่อยสลายยาได้น้อยมากหรือไม่ได้เลย คนที่มีแอลลีลแบบ *CYP2A6*4, *7, *9* และ **10* นั้นมักจะเกี่ยวข้องกับการลดความสามารถในการเมแทบอลิซึมโคดีน (Nakajima et al., 2006; Peamkrasatam et al., 2006; Yoshida et al., 2002) และลดความสามารถในการย่อย tegafur เมื่อเปรียบเทียบกับคนที่มีแอลลีลแบบปกติ เมื่อถูกรักษาด้วย S-1 (สารต้านมะเร็งแบบรับประทาน) (Fujita et al., 2008; Kim et al., 2011) นอกจากนี้ในผู้หญิงวัยหมดประจำเดือนที่สุขภาพดีที่มีแอลลีล *CYP2A6*4, *7, *9* จะมีความสามารถในการกำจัด letrozole น้อยกว่าผู้ที่มีแอลลีลแบบ *CYP2A6*1* (Tanii et al., 2011)

ยีน *SULT1A1* เป็นยีนที่มีความสำคัญในการเมแทบอลิซึม, การกระตุ้นทางชีวภาพ, กระบวนการขับสารพิษของสารก่อมะเร็ง, การยับยั้งสารสื่อประสาท catecholamines และการควบคุมอุณหภูมิของสเตรอยด์ฮอร์โมน อย่าง เอสโตรเจน เป็นต้น (Hildebrandt et al., 2009) ยีน *SULT1A1* อยู่บนบริเวณที่มีลำดับพันธุกรรมที่ซ้ำกัน และมีชิ้นส่วนของจำนวนชุดของยีนซ้ำกันมากบริเวณหนึ่ง (Martin et al., 2004) ส่งผลให้ในบริเวณนี้มักจะพบการแปรผันเชิงโครงสร้างแบบที่มีจำนวนชุดของยีนมากกว่า 2 ชุดขึ้นไป (0-6 ชุด) (Hebbring et al., 2007; Liu et al., 2017) โดยการศึกษาในชาวคอเคเซียนพบว่ามี 64% ของกลุ่มประชากรที่มีจำนวนชุดยีน 2 ชุด และ 32% ของประชากร มีจำนวนชุดของยีน ≥ 3 ชุด (Liu et al., 2017) การเพิ่มขึ้นของจำนวนชุดของยีน *SULT1A1* มีความสัมพันธ์เชิงลบต่ออัตราการเปลี่ยน estrogen-sulfate ไปเป็น estrone ในผู้ชาย (Liu et al., 2017) แอลลีล *SULT1A1*2* เป็นแอลลีลที่พบได้ทั่วไปในชาวคอเคเซียนและแอฟริกันอเมริกันแต่ไม่ค่อยพบในชาวจีน (ความถี่แอลลีล 0.332, 0.294, และ 0.08 ตามลำดับ) (Carlini et al., 2001) ข้อมูลของเราสอดคล้องกับการศึกษานี้โดยความถี่แอลลีลของ *SULT1A1*2* อยู่ที่ 0.07 และมีการแปรผันเชิงโครงสร้างแบบที่มีจำนวนชุดของยีน 2 ชุดอยู่ที่ 23% โดยประมาณ

สำหรับยีน *G6PD* เป็นยีนที่มีส่วนเกี่ยวข้องกับโรคพร่องเอนไซม์จีซิกพีดี เนื่องจากยีนนี้เป็นยีนที่พบบนโครโมโซมเพศในเพศหญิงจะพบยีนนี้ 2 ชุด ในขณะที่เพศชายจะพบเพียงหนึ่งชุดของยีนเท่านั้น ดังนั้นเพศชายจึงมีโอกาสที่จะเป็นโรคบกพร่องเอนไซม์จีซิกพีดีได้มากกว่าเพศหญิง เนื่องจากหากเพศชายได้รับการถ่ายทอดยีนกลายพันธุ์มาเพียงแค่อินเดียก็สามารถแสดงอาการของโรคออกมาได้ ในขณะที่เพศหญิงผู้หนึ่งต้องได้รับยีนที่ผิดปกติถึงสองยีนจึงจะแสดงอาการของโรค (Guindo et al., 2007) ภาวะบกพร่องเอนไซม์จีซิกพีดีเป็นโรคทางพันธุกรรมที่ส่งผลกระทบต่อความผิดปกติของเอนไซม์ ซึ่งเราพบว่าแอลลีล *G6PD* Viangchan (*G6PD**51) และแอลลีล *G6PD* Mahidol (*G6PD**31) เป็นแอลลีลที่พบมากในประเทศไทย (Tanigawara et al., 1999) สำหรับการแปรผันทางพันธุกรรมบนยีน *G6PD* รูปแบบอื่น ๆ ไม่ว่าจะเป็น *G6PD* Canton (*G6PD**8), *G6PD* Kaiping (*G6PD**28), *G6PD* Union (*G6PD**50), และ *G6PD* Gaohe (*G6PD**153) พบในชาวจีน, ชาวอินเดีย, และประชากรจากภูมิภาคเอเชียตะวันออกเฉียงใต้ (Jitueakul et al., 2018) จากการศึกษา เราพบว่าผู้ชาย 8 คนที่มีแอลลีล *G6PD**8, *21, *28 และ *51 ในขณะที่ผู้ชายส่วนใหญ่มีแอลลีลที่การทำงานเป็นปกติ โดยในแอลลีล *G6PD**8 มีตำแหน่ง SNP หลักคือ rs72554665 (g.153760484C>A) มีการรายงานก่อนหน้านี้ว่า พบกรณีเด็กชายลูกครึ่งชาวจีน-ญี่ปุ่น อายุ 26 เดือนที่มีการแปรผัน rs72554665(A) โดยหลังจากที่เด็กชายคนนี้รับประทานกินถั่วปากอ้าเข้าไป ส่งผลให้เกิดภาวะโลหิตจางอย่างฉับพลัน (Shibuya et al., 1999) นอกจากนี้ยาบางประเภทอย่าง rasburicase นั้นก็ถูกห้ามใช้ในผู้ชายที่มีความผิดปกติของแอลลีลแบบคลาส I, II, และ III เช่น แอลลีล *G6PD* Canton เนื่องจากส่งผลให้เกิดภาวะเม็ดเลือดแดงแตกฉับพลัน (Relling et al., 2014)

CYP4B1 เป็นอินเตอร์เฟสในกระบวนการเมแทบอลิซึมของสารประเภท xenobiotics เช่น 2-aminofluorene และสารประเภท endobiotic รวมถึง ligands ที่เอนไซม์กลุ่ม P450 รับผิดชอบในการดัดแปลงกระบวนการภายใน (Baer & Rettie, 2006) โดยทั่วไปยีน *CYP4B1* มักจะแสดงออกในเซลล์ปอดของมนุษย์เป็นหลัก และอาจส่งผลต่อการก่อตัวของมะเร็งได้ (Lim et al., 2020) จากการศึกษาพบว่า 20% ของกลุ่มตัวอย่างของเรามีแอลลีล *CYP4B1**2 การศึกษาในประเทศญี่ปุ่นพบว่ามีคนที่มีจีโนไทป์แบบ *CYP4B1**1/*2 หรือ *2/*2 มีความเสี่ยงที่จะเป็นมะเร็งกระเพาะปัสสาวะเพิ่มขึ้นถึง 1.75 เท่า (Sasaki et al., 2008) สำหรับความถี่แอลลีลของ *CYP4B1**2 ในประชากรญี่ปุ่นนั้นอยู่ที่ 0.328 และ 0.147 ในประชากรฝรั่งเศส (Hiratsuka et al., 2004)

ยีน *CYP2D6* และยีนคู่เสมือน *CYP2D7* เป็นยีนที่รับผิดชอบเกี่ยวกับการเมแทบอลิซึมและการกำจัดยามากถึง 20% ของยาทั้งหมดที่มนุษย์กำลังใช้กันอยู่ทั่วโลก ยีนนี้เป็นอีกยีนที่ความหลากหลายของการผันแปรทางพันธุกรรมมาก อีกทั้งแต่ละกลุ่มประชากรก็มีรูปแบบความหลากหลายของการแปรผันทางพันธุกรรมของยีนที่แตกต่างกัน (Leitão et al., 2020) รูปแบบการแปรผันทางพันธุกรรมแบบแอลลีล *CYP2D6**2 และ *CYP2D6**4 พบได้มากในชาวยุโรป ชาวแอฟริกัน ชาวเอเชียใต้ และประชากรผสมในทวีปอเมริกา ในขณะที่แอลลีล *CYP2D6**5 (การขาดหายไปของยีน) และ *CYP2D6**10 นั้นพบได้มากในประชากรเอเชียตะวันออกเฉียงใต้ (Zhou et al., 2017) นอกจากนี้แอลลีล *CYP2D6**41 ยังถูกพบว่ามีความถี่มากในประชากรเอเชียใต้ (Zhou et al., 2017) ในการศึกษาของเรา

พบว่า *CYP2D6**36+*10 เป็นอัลลีลที่มีความถี่สูงที่สุด ตามมาด้วยอัลลีล *CYP2D6**10 ในขณะที่อัลลีล *CYP2D6**2, *4, *5 และ *41 เป็นอัลลีลที่ถูกลบได้น้อย สำหรับคนที่มีความสามารถในการย่อยสลายยาได้ปานกลางซึ่งมีอัลลีล *CYP2D6**10 ประกอบอยู่ในจีโนมไทป์ เช่น *CYP2D6**10/*10 และ *CYP2D6**10/*41 คำแนะนำในการสั่งจ่ายยา atomoxetine สำหรับคนเหล่านี้ แนะนำให้เริ่มต้นใช้ยาที่ปริมาณ 40 mg/วัน และ ค่อยปรับเพิ่มเป็น 80 mg/วัน หลังจากใช้ยาได้สองสัปดาห์ถ้าหากไม่ค่อยมีการตอบสนองและไม่มีอาการที่ไม่พึงประสงค์หลังจากการใช้ยา (Brown et al., 2019)

การแสดงออกของยีน *CYP2F1* ในมนุษย์ ถูกพบอย่างจำเพาะในเนื้อเยื่อบริเวณปอดโดยยีนนี้ทำหน้าที่ในการย่อยสารประกอบไอระเหยในระบบทางเดินหายใจ พบอัลลีลที่ไม่ทำงาน *CYP2F1**2 ในความถี่ที่สูง ซึ่งอัลลีลนี้เป็นอัลลีลที่เกิดจากการเปลี่ยนแปลงของนิวคลีโอไทด์เพียงหนึ่งตำแหน่งและไม่ส่งผลกระทบต่อการทำงานของโปรตีน (silence mutation) ในขณะที่อัลลีลรูปแบบอื่นถูกพบได้ในส่วนที่น้อยมาก ๆ (0.6–7.2%) ในประชากรฝรั่งเศส (Tournel et al., 2007) และจากการศึกษาเราพบว่าอัลลีล *CYP2F1**2 นี้เป็นอัลลีลที่พบได้ทั่วไปในกลุ่มตัวอย่างของเรา

ในคนที่มีลักษณะจีโนมไทป์ของยีน *NAT2* แบบที่มีความสามารถในการการเติมหมู่อะซิetylได้ช้า (slow acetylator) ซึ่งอาจเป็นลักษณะของจีโนมไทป์แบบ homo- หรือ compound heterozygotes ของ *NAT2**5, *6 หรือ *7 มีความเกี่ยวข้องกับความเสี่ยงที่เพิ่มขึ้นของการบาดเจ็บที่ตับจากการใช้ยาต้านวัณโรค (Suvichapanich et al., 2018) ความเสี่ยงในการเกิดผลกระทบที่ไม่พึงประสงค์จากการใช้ยา cotrimoxazole ในผู้ป่วยโรคลูปัส (systemic lupus erythematosus) (Soejima et al., 2007), และความเสี่ยงต่อการเกิดภาวะพิษจากการใช้ยา sulfasalazine (Yee et al., 2020) อย่างไรก็ตามเกือบครึ่งหนึ่งของกลุ่มตัวอย่างที่เราใช้ในการศึกษาคั้งนี้เป็นบุคคลที่มีการทำงานของ *NAT2* แบบที่มีความสามารถในการการเติมหมู่อะซิetylได้ปานกลาง (intermediate acetylator)

ยีน *SLCO2B1* เป็นยีนที่มีการแสดงออกในเยื่อหุ้มเซลล์ของ enterocytes ที่อยู่ภายในช่องว่างของลำไส้เล็ก และอาจมีหน้าที่เกี่ยวกับการดูดซึมยาจากช่องว่างภายในลำไส้เข้าสู่ร่างกาย (Shitara et al., 2013) จากการศึกษาพบว่าลักษณะฟีโนไทป์แบบที่มีความสามารถในการย่อยสลายยาได้ปานกลาง และน้อยมากหรือไม่ได้เลย มาจากการที่มีอัลลีล *SLCO2B1**S1 เป็นส่วนประกอบของจีโนมไทป์ โดยอัลลีลนี้เป็นรูปแบบการแปรผันทางพันธุกรรมแบบ in-frame deletion และเป็นอัลลีลที่พบมากในประชากรเอเชียตะวันออกเท่านั้น (Lee et al., 2019(b))

การแปรผันทางพันธุกรรมในยีน *UGT1A1* นั้นมีส่วนเกี่ยวข้องกับโรคมากมายเช่น โรค Gilbert syndrome (*UGT1A1**6/*6) (Yamamoto et al., 1998), มะเร็งศีรษะและลำคอ (head and neck cancers)

(Lacko et al., 2010), มะเร็งลำไส้ (colorectal cancer) (Tang et al., 2005) และโรคหลอดเลือดหัวใจตีบตัน (Lingenhel et al., 2008) จากการศึกษาก่อนหน้านี้พบว่าแอลลีล *UGT1A1*60* เป็นแอลลีลที่พบในประชากรเอเชีย (เกาหลี, จีน และญี่ปุ่น), แอฟริกันอเมริกัน, และยุโรปอเมริกัน (J. Y. Kim et al., 2014) โดยแอลลีล *UGT1A1*6* เป็นแอลลีลที่พบได้ทั่วไปที่พบได้ในประชากรเอเชียเท่านั้น (Akaba et al., 1999) บุคคลที่มีแอลลีลนี้ไม่ว่าจะ 1 หรือ 2 แอลลีลเป็นผู้ที่มีความเสี่ยงในการเกิดภาวะเม็ดเลือดขาวต่ำเนื่องจากการใช้ยา irinotecan เมื่อเทียบกับคนที่มีการแสดงออกของยีนนี้แบบปกติ (Jada et al., 2007; Onoue et al., 2009) นอกจากนี้มีรายงานว่า ผู้ HIV ชาวไทยที่มีแอลลีล *UGT1A1*6* เมื่อได้รับการรักษาด้วยยา indinavir มีความเสี่ยงสูงที่จะเกิดภาวะตัวเหลืองจากการมีบิลิรูบินมากในกระแสเลือด (Boyd et al., 2006) จากการศึกษาของเราพบว่าแอลลีล *UGT1A1*1* และ *UGT1A1*60* เป็นแอลลีลที่มีความถี่ของแอลลีลสูงที่สุด ในขณะที่แอลลีล *UGT1A1*6* เป็นแอลลีลที่มีความถี่สูงเป็นลำดับรองลงมา

ยีน *SLCO1B1* เป็นยีนที่ทำหน้าที่เกี่ยวกับการขนส่งอินทรีย์สารประเภทแอนไอออน และยังเป็นที่ยูจกกัน ในอีกชื่อหนึ่งว่า *OATP1B1* ซึ่งเป็นโปรตีนที่มีบทบาทในการดูดซึมสารประกอบประเภท endogenous และ xenobiotic หลายตัวของตับ (Niemi et al., 2011) ในกลุ่มตัวอย่างที่เราใช้ศึกษาพบว่า กลุ่มตัวอย่างที่มีการแสดงออกของแอลลีลแบบลดการทำงานและมีแอลลีลที่ทำงานได้ซ้ำมักจะเป็นกลุ่มคนที่มีแอลลีลแบบ *SLCO1B1*15* หรือ *SLCO1B1*17* คำแนะนำสั่งจ่ายยา simvastatin ในบุคคลที่มีแอลลีลแบบ heterozygous (แอลลีลข้างหนึ่งเป็นแอลลีลที่ลดการทำงาน) เช่น *SLCO1B1*1B/*15*, **1B/*17* หรือ สำหรับบุคคลที่มีแอลลีลแบบ homozygous (เป็นแอลลีลที่ลดการทำงานทั้งสองข้าง) เช่น *SLCO1B1*15/*15*, **15/*17*, **17/*17* ระบุว่าควรเริ่มใช้ยาในปริมาณที่ต่ำ หรือ พิจารณาทางเลือกประเภท statin เพราะอาจมีความเสี่ยงต่อการเกิดความผิดปกติของกล้ามเนื้อ (myopathy) ในระดับปานกลาง สำหรับคนที่เป็น heterozygous และระดับสูง ในคนที่เป็น homozygous (Niemi et al., 2011) นอกจากนี้ *SLCO1B1*15* ยังถูกพบว่ามีมีความเกี่ยวข้องกับ การเพิ่มขึ้นของสาร pravastatin (Ide et al., 2009) ทำให้มีการสัมผัสกับสาร pravastatin มากขึ้น (Oh et al., 2013) ทำให้มีปริมาณความเข้มข้นของ repaglinide เพิ่มขึ้น (Pei et al., 2018) ทำให้เกิดความเสี่ยงของการบาดเจ็บที่ตับจากการใช้ยา rifampin (Li et al., 2012) และยังเกี่ยวเนื่องกับการการกำจัด olmesartan ที่ลดลง (Suwannakul et al., 2008) การเมแทบอลิซึมยา rosuvastatin (Choi et al., 2008) และการขนส่งยา atrasentan (Katz et al., 2006)

งานวิจัยของเราได้ยืนยันรูปแบบของ star allele ความถี่แอลลีล และลักษณะฟีโนไทป์ที่พบว่ามีลักษณะที่คล้ายคลึงกับงานวิจัยก่อนหน้านี้ที่ทำในกลุ่มประชากรไทย (Mauleekoonphairoj et al., 2020) ข้อมูลรูปแบบการแปรผันทางพันธุกรรมของยีน *CACNA1S* มีผลการศึกษาที่เหมือนกันกับการศึกษาครั้งก่อนหน้า เรายังพบถึง 19 ยีนที่มีรูปแบบการแปรผันทางพันธุกรรมคล้ายคลึงกับการศึกษาก่อนหน้านี้ (*CFTR*, *CYP2B6*, *CYP2C8*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *CYP3A5*, *CYP4F2*, *GSTP1*, *NAT1*, *NAT2*, *NUDT15*, *RYR1*, *SLCO1B1*, *TPMT*,

UGT1A1, *UGT1A4* และ *VKORC1*) ไม่ว่าจะเป็นเรื่องของความถี่แอลลีลหรือลักษณะฟีโนไทป์ อย่างไรก็ตาม เราพบว่า มี 3 ยีน (*DPYD*, *G6PD* และ *IFNL3*) ที่มีผลการศึกษาที่แตกต่างกัน ในการศึกษาของเราพบว่าแอลลีล *DPYD*512* เป็นแอลลีลที่มีความถี่สูงสุด และเป็นแอลลีลที่ทำให้เกิดฟีโนไทป์แบบที่มีความสามารถในการย่อยสลายยาได้ปานกลาง ในขณะที่ *DPYD*1* มีความถี่สูงสุด และแอลลีล *DPYD*512* เป็นแอลลีลที่ไม่ถูกพบในการศึกษาก่อนหน้า (Mauleekoonphairoj et al., 2020) ทำให้มีการแสดงออกทางฟีโนไทป์แบบที่มีความสามารถในการย่อยสลายยาได้ในระดับปกติ งานวิจัยเราพบ star alleles หลายรูปแบบในยีน *G6PD* ได้แก่ *G6PD*1x2*, **21*, **28*, **31*, **50*, **51*, **8* และ **DEL* ซึ่งแอลลีลทั้งหมดนี้ทำให้เกิดการแสดงออกทางฟีโนไทป์แบบที่มีความสามารถในการย่อยสลายยาได้ปานกลาง ในขณะที่การศึกษาก่อนหน้าพบเพียงแอลลีล *G6PD*1* แอลลีลเดียว (Mauleekoonphairoj et al., 2020) สำหรับยีน *IFNL3* ทั้งสองการศึกษาพบตรงกันว่าในกลุ่มประชากรของเรามีแอลลีลแบบ *IFNL3*1* และ *IFNL3*53* อย่างไรก็ตาม การศึกษาของเราพบแอลลีล *IFNL3*DEL* และฟีโนไทป์ที่ทำนายก็แตกต่างออกไป สำหรับการแสดงออกของยีน *IFNL3* การศึกษาของเราใช้คำว่า unfavorable response แทนคำว่า rapid metabolizer ซึ่งเป็นคำนิยามที่ใช้อธิบายลักษณะของฟีโนไทป์ในงานวิจัยขึ้นก่อนหน้า สำหรับผลการศึกษาในยีน *GSTM1* และ *UGT2B15* ในงานวิจัยของพวกเราที่มีความน่าเชื่อถือต่ำ ดังนั้นพวกเราจึงไม่รายงานการเปรียบเทียบผลการศึกษาระหว่างสองงานวิจัยในสองยีนนี้

เราพบว่าทั้งสามยีนควบคุมสามารถใช้แทนกันได้และให้ผลการทำนายรูปแบบการแปรผันทางพันธุกรรมที่คล้ายคลึงกันเนื่องจากทั้งสามยีนมีความลึกของ read ที่ใกล้เคียงกันในการศึกษาของเรา แต่สิ่งหนึ่งที่ควรทราบคือจำนวน read depth ของยีนควบคุมควรมีปริมาณที่สอดคล้องกันกับยีนเป้าหมาย มิเช่นนั้นอาจทำให้ได้ผลการวิเคราะห์ที่ผิดพลาดในเรื่องของการเกิดการแปรผันทางพันธุกรรมเชิงโครงสร้าง อย่างการขาดหายไป หรือการเพิ่มจำนวนของชุดของยีน โดยในยีนเป้าหมายอาจถูกรายงานว่าพบการแปรผันของจำนวนชุดของยีนอันเนื่องมาจากการทำซ้ำ หรือการขาดหายไปของยีน หากยีนควบคุมมีความลึกของ read ที่ต่ำไป/สูงไป ในการศึกษาของเรา เมื่อผลลัพธ์จาก 3 ยีนไม่สอดคล้องกัน การใช้ *RYR1* เป็นยีนควบคุมจะรายงานการแปรผันของจำนวนชุดของยีนอันเนื่องมาจากการทำซ้ำ ในขณะที่ยีนควบคุมอีกสองยีนรายงานผลว่ามีจำนวนชุดของยีนเป็นปกติ และเมื่อเราตรวจสอบความลึกของ read เราพบว่า *RYR1* มีจำนวนความลึกของ read ที่น้อยกว่าอีกสองยีน

ยีนเป้าหมายทั้ง 51 ยีนที่เราได้ทำการศึกษานั้น เป็นยีนทั้งหมดที่โปรแกรม Stargazer version 1.0.8 สามารถให้รูปแบบการแปรผันทางพันธุกรรมได้และเป็นจำนวนยีนที่สูงที่สุดที่โปรแกรมจะสามารถทำนายได้ในขณะนี้ เมื่อเปรียบเทียบกับเครื่องมืออื่น ๆ อย่างเช่น Aldy (Numanagic et al., 2018), StellarPGx (Twesigomwe et al., 2021), PharmVIP (Piriyaongsa et al., 2021) และ Astrolabe (Twist et al., 2016) ซึ่งโปรแกรมเหล่านี้สามารถทำนายได้ 35, 13, 17 (guideline module) และ 9 ยีนตามลำดับ เราพบว่ามี 16 ยีน ที่มีหลักฐานอ้างอิงทางคลินิกตามมาตรฐานของฐานข้อมูล PharmGKB ระดับ 1A (ภาคผนวก ตารางที่ 1) ยีนที่มีหลักฐานอ้างอิงทาง

คลินิกระดับ 3 หรือ 4 อาจมีความสำคัญเนื่องจากบาง variants พบได้บ่อยในกลุ่มตัวอย่างของเรามากกว่าในประชากรอื่น ๆ ดังนั้น คู่ของ variant-drug อาจถูกทำการศึกษาต่อไปในบริบทของเรา

ในการศึกษาทั้งหมดของเราข้อมูลที่น่ามาใช้ในการศึกษานั้นเป็นข้อมูลที่ได้มาจากการจัดลำดับพันธุกรรมแบบสั้นทั้งจีโนม (short-read sequencing technique) ซึ่งค่าความถูกต้องแม่นยำในการหาลำดับเบสที่เกิดการแปรผันเชิงโครงสร้าง อย่างเช่น ความผิดปกติของจำนวนชุดของยีน อาจไม่ได้ค่าที่ดีที่สุด สิ่งนี้อาจส่งผลกระทบต่อการทำนายรูปแบบการแปรผันทางพันธุกรรมในบางยีน อย่างเช่น *CYP2D6* ได้

การศึกษาของเราได้คัดเลือกพ่อแม่ที่ไม่เป็นโรคของเด็กที่เป็นโรคหายาก ดังนั้นจึงอาจเพิ่มโอกาสที่จะเจอการแปรผันที่เป็นสาเหตุของการเกิดโรคหายากได้มากขึ้น ดังที่พบในฐานข้อมูล Thai Reference Exome (T-REX) ซึ่งมีเกณฑ์การเลือกกลุ่มตัวอย่างที่คล้ายกัน (Shotelersuk et al., 2021) T-Rex ใช้กลุ่มตัวอย่างจากผู้ปกครองที่ไม่เป็นโรคแต่มีลูกเป็นโรคหายากและพบการแปรผันทางพันธุกรรมที่ก่อให้เกิดโรคหายากมากกว่าเมื่อเปรียบเทียบกับฐานข้อมูลอย่าง gnomAD อย่างไรก็ตามการศึกษานี้ครั้งนี้ไม่ได้รับสมัครผู้ป่วยที่มีโรคที่เกี่ยวข้องกับยีนที่เกี่ยวข้องกับการตอบสนองต่อยา อย่างเช่น โรคบกพร่องเอนไซม์ G6PD เป็นต้น เราตั้งใจที่จะใช้เฉพาะผู้ปกครองของผู้ป่วยที่เป็นโรคหายากที่ 1) มีการกลายพันธุ์แบบใหม่ อย่างเช่น ภาวะกระดูกอ่อนไม่เจริญ achondroplasia หรือ 2) โรคที่ไม่เกี่ยวข้องกับ 51 ยีนที่เกี่ยวข้องกับการตอบสนองต่อยา เช่น methylmalonic academia และโรคกล้ามเนื้อเสื่อมดูเชน (Duchenne muscular dystrophy) ดังนั้นการเพิ่มขึ้นของการแปรผันที่ทำให้เกิดโรคในผู้ปกครองที่ไม่เป็นโรคเหล่านี้หรือแม้แต่วผู้ป่วยเองไม่ควรส่งผลกระทบต่อบทสรุปของการศึกษาในครั้งนี้

บทที่ 6

สรุปผลการวิจัย

การศึกษาของเราตอบโจทย์วัตถุประสงค์งานวิจัยทั้ง 3 หัวข้อ ได้แก่

1. สามารถทำนายรูปแบบการแปรผันทางพันธุกรรมในยีนที่มีผลต่อการตอบสนองต่อยา 51 ยีน ในประชากรไทย 171 คน ที่ไม่เป็นโรค และมีภูมิลำเนาจากทุกภูมิภาคทั่วประเทศไทย
2. สามารถเปรียบเทียบความถี่แอลลีลระหว่างประชากรไทย และประชากรจากชาติอื่น ๆ
3. สามารถค้นพบรูปแบบการแปรผันทางพันธุกรรมแบบใหม่ที่มีความสำคัญและจำเพาะในประชากรไทย

การใช้โปรแกรม stargazer เพื่อทำนาย star allele และลักษณะของฟีโนไทป์โดยใช้ข้อมูลจากรหัสพันธุกรรมทั้งจีโนมจำนวน 171 คน พบว่าในคนหนึ่งคนมีอย่างน้อย 3 ยีนที่มีฟีโนไทป์แบบที่ไม่ใช่ฟีโนไทป์ปกติ และ 40 จาก 51 ยีนที่ศึกษา จะมีอย่างน้อยหนึ่งคนที่มีฟีโนไทป์แบบที่ไม่ใช่ฟีโนไทป์ปกติ โดยอย่างน้อย 25% ของกลุ่มตัวอย่างที่เราศึกษามีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์ปกติ ใน 13 ยีน ได้แก่ *SLCO1B3* (97.08%), *CYP3A5* (88.3%), *CYP2C19* (60.82%), *CYP2A6* (60.2%), *SULT1A1* (56.14%), *G6PD* (54.39%), *CYP4B1* (50.00%), *CYP2D6* (48.65%), *CYP2F1* (46.41%), *NAT2* (40.35%), *SLCO2B1* (28.95%), *UGT1A1* (28.07%) และ *SLCO1B1* (26.79%) อีกทั้งการศึกษาของเรายังพบการแปรผันทางพันธุกรรมที่มีผลกระทบต่อการทำงานของยีนในระดับสูงจำนวน 20 ตัว ซึ่งยังไม่พบการรายงานบนฐานข้อมูลใด ๆ

โดยสรุป การศึกษาของเราได้ให้แนวทางเภสัชพันธุศาสตร์ สำหรับการสั่งจ่ายยาและเป็นแนวทางนำร่องสำหรับนโยบายการตรวจจีโนไทป์เฉพาะประชากร เพื่อปรับปรุงการตอบสนองต่อยา หรือการป้องกันการผลกระทบบที่ไม่พึงประสงค์จากการใช้ยาสำหรับประชากรไทย นอกจากนี้การศึกษาในครั้งนี้ยังให้ข้อมูลของยีนที่มีสัดส่วนของประชากรที่มีฟีโนไทป์ที่ไม่ใช่ฟีโนไทป์ปกติสูง ซึ่งยีนเหล่านี้ควรจะถูกรวบรวมเข้าไปอยู่ในรายชื่อยีนทางเภสัชพันธุศาสตร์เบื้องต้นที่ควรตรวจสำหรับประชากรไทย

บรรณานุกรม

- Akaba, K., Kimura, T., Sasaki, A., Tanabe, S., Wakabayashi, T., Hiroi, M., Yasumura, S., Maki, K., Aikawa, S., & Hayasaka, K. (1999). Neonatal hyperbilirubinemia and a common mutation of the bilirubin uridine diphosphate-glucuronosyltransferase gene in Japanese. *J Hum Genet*, *44*(1), 22-25.
<https://doi.org/10.1007/s100380050100>
- Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res*, *9*(13), 3015-3027.
<https://doi.org/10.1093/nar/9.13.3015>
- Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N Biotechnol*, *25*(4), 195-203. <https://doi.org/10.1016/j.nbt.2008.12.009>
- Baer, B. R., & Rettie, A. E. (2006). CYP4B1: An Enigmatic P450 at the Interface between Xenobiotic and Endobiotic Metabolism. *Drug Metab*, *38*(3), 451-476.
<https://doi.org/10.1080/03602530600688503>
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*, *125*(1-2), 279-284.
[https://doi.org/10.1016/s0166-4328\(01\)00297-2](https://doi.org/10.1016/s0166-4328(01)00297-2)
- Birdwell, K. A., Decker, B., Barbarino, J. M., Peterson, J. F., Stein, C. M., Sadee, W., Wang, D., Vinks, A. A., He, Y., Swen, J. J., Leeder, J. S., van Schaik, R., Thummel, K. E., Klein, T. E., Caudle, K. E., & MacPhee, I. A. (2015). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. *Clin Pharmacol Ther*, *98*(1), 19-24.
<https://doi.org/10.1002/cpt.113>
- Bjornsson, T. D., Wagner, J. A., Donahue, S. R., Harper, D., Karim, A., Khouri, M. S., Murphy, W. R., Roman, K., Schneck, D., Sonnichsen, D. S., Stalker, D. J., Wise, S. D., Dombey, S., & Loew, C. (2003). A review and assessment of potential sources of ethnic differences in drug responsiveness. *J Clin Pharmacol*, *43*(9), 943-967.
<https://doi.org/10.1177/0091270003256065>
- Boegel, S., Löwer, M., Schäfer, M., Bukur, T., de Graaf, J., Boisguérin, V., Türeci, Ö., Diken,

- M., Castle, J. C., & Sahin, U. (2012). HLA typing from RNA-Seq sequence reads. *Genome Med*, 4(12), 102. <https://doi.org/10.1186/gm403>
- Boschiero, C., Gheyas, A. A., Ralph, H. K., Eory, L., Paton, B., Kuo, R., Fulton, J., Preisinger, R., Kaiser, P., & Burt, D. W. (2015). Detection and characterization of small insertion and deletion genetic variants in modern layer chicken genomes. *BMC Genom*, 16(1), 562. <https://doi.org/10.1186/s12864-015-1711-1>
- Boyd, M. A., Srasuebku, P., Ruxrungtham, K., Mackenzie, P. I., Uchaipichat, V., Stek, M., Jr., Lange, J. M., Phanuphak, P., Cooper, D. A., Udomuksorn, W., & Miners, J. O. (2006). Relationship between hyperbilirubinaemia and UDP-glucuronosyltransferase 1A1 (UGT1A1) polymorphism in adult HIV-infected Thai patients treated with indinavir. *Pharmacogenet Genomics*, 16(5), 321-329. <https://doi.org/10.1097/01.fpc.0000197465.14340.d4>
- Brown, J. T., Bishop, J. R., Sangkuhl, K., Nurmi, E. L., Mueller, D. J., Dinh, Jean C., Gaedigk, A., Klein, T. E., Caudle, K. E., McCracken, J. T., de Leon, J., & Leeder, J. S. (2019). Clinical Pharmacogenetics Implementation Consortium Guideline for Cytochrome P450 (CYP)2D6 Genotype and Atomoxetine Therapy. *Clin Pharmacol Ther*, 106(1), 94-102. <https://doi.org/https://doi.org/10.1002/cpt.1409>
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* 103(3), 338-348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5), 1084-1097. <https://doi.org/10.1086/521987>
- Browning, S. R., & Browning, B. L. (2011). Haplotype phasing: existing methods and new developments. *Nat Rev Genet*, 12(10), 703-714. <https://doi.org/10.1038/nrg3054>
- Bueno de Mesquita, C. P., Nichols, L. M., Gebert, M. J., Vanderburgh, C., Bocksberger, G., Lester, J. D., Kalan, A. K., Dieguez, P., McCarthy, M. S., Agbor, A., Álvarez Varona, P., Ayimisin, A. E., Bessone, M., Chancellor, R., Cohen, H., Coupland, C., Deschner, T., Egbe, V. E., Goedmakers, A., . . . Dunn, R. R. (2021). Structure of Chimpanzee Gut Microbiomes across Tropical Africa. *mSystems*, 6(3), e0126920.

<https://doi.org/10.1128/mSystems.01269-20>

Bush, W. S., Crosslin, D. R., Owusu-Obeng, A., Wallace, J., Almoguera, B., Basford, M. A., Bielinski, S. J., Carrell, D. S., Connolly, J. J., Crawford, D., Doheny, K. F., Gallego, C. J., Gordon, A. S., Keating, B., Kirby, J., Kitchner, T., Manzi, S., Mejia, A. R., Pan, V., . . . Ritchie, M. D. (2016). Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. *Clin Pharmacol Ther*, *100*(2), 160-169.

<https://doi.org/10.1002/cpt.350>

Carlini, E. J., Raftogianis, R. B., Wood, T. C., Jin, F., Zheng, W., Rebbeck, T. R., & Weinshilboum, R. M. (2001). Sulfation pharmacogenetics: SULT1A1 and SULT1A2 allele frequencies in Caucasian, Chinese and African-American subjects.

Pharmacogenetics, *11*(1), 57-68. <https://doi.org/10.1097/00008571-200102000-00007>

Choi, J., Lee, M., Cho, J.-Y., Lee, J.-E., Kim, K., & Park, K. (2008). Influence of OATP1B1 Genotype on the Pharmacokinetics of Rosuvastatin in Koreans. *Clin Pharmacol Ther*, *83*(2), 251-257. <https://doi.org/https://doi.org/10.1038/sj.clpt.6100267>

Chueca, L. J., Schell, T., & Pfenninger, M. (2021). Whole-genome re-sequencing data to infer historical demography and speciation processes in land snails: the study of two *Candidula* sister species. *Philos Trans R Soc Lond, B, Biol Sci*, *376*(1825), 20200156. <https://doi.org/doi:10.1098/rstb.2020.0156>

de Lima, L. T., Bueno, C. T., Vivona, D., Hirata, R. D. C., Hirata, M. H., Hungria, V. T. d. M., Chiattoni, C. S., Zanichelli, M. A., Chauffaille, M. d. L. L. F., & Guerra-Shinohara, E. M. (2015). Relationship between SLCO1B3 and ABCA3 polymorphisms and imatinib response in chronic myeloid leukemia patients. *Hematology (Amsterdam, Netherlands)*, *20*(3), 137-142.

<https://doi.org/10.1179/1607845414y.0000000181>

Del Tredici, A. L., Malhotra, A., Dedek, M., Espin, F., Roach, D., Zhu, G. D., Volland, J., & Moreno, T. A. (2018). Frequency of CYP2D6 Alleles Including Structural Variants in the United States. *Front Pharmacol*, *9*, 305.

<https://doi.org/10.3389/fphar.2018.00305>

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky,

- A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5), 491-498. <https://doi.org/10.1038/ng.806>
- Doogue, M. P., & Polasek, T. M. (2013). The ABCD of clinical pharmacokinetics. *Ther Adv Drug Saf*, 4(1), 5-7. <https://doi.org/10.1177/2042098612469335>
- Easton, D. F., Pharoah, P. D., Antoniou, A. C., Tischkowitz, M., Tavtigian, S. V., Nathanson, K. L., Devilee, P., Meindl, A., Couch, F. J., Southey, M., Goldgar, D. E., Evans, D. G., Chenevix-Trench, G., Rahman, N., Robson, M., Domchek, S. M., & Foulkes, W. D. (2015). Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med*, 372(23), 2243-2257. <https://doi.org/10.1056/NEJMSr1501341>
- Engvall, M., Cahill, N., Jonsson, B.-I., Höglund, M., Hallböök, H., & Cavelier, L. (2020). Detection of leukemia gene fusions by targeted RNA-sequencing in routine diagnostics. *BMC Med Genomics*, 13(1), 106. <https://doi.org/10.1186/s12920-020-00739-4>
- Fang, H., Xu, X., Kaur, K., Dedek, M., Zhu, G. D., Riley, B. J., Espin, F. G., Del Tredici, A. L., & Moreno, T. A. (2019). A Screening Test for HLA-B(*)15:02 in a Large United States Patient Cohort Identifies Broader Risk of Carbamazepine-Induced Adverse Events. *Front Pharmacol*, 10, 149. <https://doi.org/10.3389/fphar.2019.00149>
- Fricke-Galindo, I., Céspedes-Garro, C., Rodrigues-Soares, F., Naranjo, M. E., Delgado, Á., de Andrés, F., López-López, M., Peñas-Lledó, E., & A, L. L. (2016). Interethnic variation of CYP2C19 alleles, 'predicted' phenotypes and 'measured' metabolic phenotypes across world populations. *Pharmacogenomics J*, 16(2), 113-123. <https://doi.org/10.1038/tpj.2015.70>
- Fujita, K., Yamamoto, W., Endo, S., Endo, H., Nagashima, F., Ichikawa, W., Tanaka, R., Miya, T., Araki, K., Kodama, K., Sunakawa, Y., Narabayashi, M., Miwa, K., Ando, Y., Akiyama, Y., Kawara, K., Kamataki, T., & Sasaki, Y. (2008). CYP2A6 and the plasma level of 5-chloro-2, 4-dihydroxypyridine are determinants of the pharmacokinetic variability of tegafur and 5-fluorouracil, respectively, in Japanese patients with cancer given S-1. *Cancer Sci*, 99(5), 1049-1054. <https://doi.org/10.1111/j.1349-7006.2008.00773.x>

- Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Klein, T., & Leeder, J. S. (2017). Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med*, *19*(1), 69-76. <https://doi.org/10.1038/gim.2016.80>
- Gonzalez-Covarrubias, V., Morales-Franco, M., Cruz-Correa, O. F., Martínez-Hernández, A., García-Ortiz, H., Barajas-Olmos, F., Genis-Mendoza, A. D., Martínez-Magaña, J. J., Nicolini, H., Orozco, L., & Soberón, X. (2019). Variation in Actionable Pharmacogenetic Markers in Natives and Mestizos From Mexico. *Front Pharmacol*, *10*, 1169. <https://doi.org/10.3389/fphar.2019.01169>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, *17*(6), 333-351. <https://doi.org/10.1038/nrg.2016.49>
- Guindo, A., Fairhurst, R. M., Doumbo, O. K., Wellem, T. E., & Diallo, D. A. (2007). X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. *PLoS Med*, *4*(3), e66. <https://doi.org/10.1371/journal.pmed.0040066>
- Hannan, A. J. (2018). Tandem repeats mediating genetic plasticity in health and disease. *Nat Rev Genet*, *19*(5), 286-298. <https://doi.org/10.1038/nrg.2017.115>
- Hebbring, S. J., Adjei, A. A., Baer, J. L., Jenkins, G. D., Zhang, J., Cunningham, J. M., Schaid, D. J., Weinshilboum, R. M., & Thibodeau, S. N. (2007). Human SULT1A1 gene: copy number differences and functional implications. *Hum Mol Genet*, *16*(5), 463-470. <https://doi.org/10.1093/hmg/ddl468>
- Hicks, J. K., Bishop, J. R., Sangkuhl, K., Müller, D. J., Ji, Y., Leckband, S. G., Leeder, J. S., Graham, R. L., Chiulli, D. L., A, L. L., Skaar, T. C., Scott, S. A., Stingl, J. C., Klein, T. E., Caudle, K. E., & Gaedigk, A. (2015). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. *Clin Pharmacol Ther*, *98*(2), 127-134. <https://doi.org/10.1002/cpt.147>
- Hildebrandt, M., Adjei, A., Weinshilboum, R., Johnson, J. A., Berlin, D. S., Klein, T. E., & Altman, R. B. (2009). Very important pharmacogene summary: sulfotransferase 1A1. *Pharmacogenet Genomics*, *19*(6), 404-406. <https://doi.org/10.1097/FPC.0b013e32832e042e>

- Hiratsuka, M., Nozawa, H., Konno, Y., Saito, T., Konno, S., & Mizugaki, M. (2004). Human CYP4B1 Gene in the Japanese Population Analyzed by Denaturing HPLC. *Drug Metab Pharmacokinet*, 19(2), 114-119.
<https://doi.org/https://doi.org/10.2133/dmpk.19.114>
- Ide, T., Sasaki, T., Maeda, K., Higuchi, S., Sugiyama, Y., & Ieiri, I. (2009). Quantitative Population Pharmacokinetic Analysis of Pravastatin Using an Enterohepatic Circulation Model Combined With Pharmacogenomic Information on SLCO1B1 and ABCC2 Polymorphisms. *J Clin Pharmacol*, 49(11), 1309-1317.
<https://doi.org/https://doi.org/10.1177/0091270009341960>
- Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y., & Lauschke, V. M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum Genomics*, 12(1), 26. <https://doi.org/10.1186/s40246-018-0157-3>
- Jada, S. R., Lim, R., Wong, C. I., Shu, X., Lee, S. C., Zhou, Q., Goh, B. C., & Chowbay, B. (2007). Role of UGT1A1*6, UGT1A1*28 and ABCG2 c.421C>A polymorphisms in irinotecan-induced neutropenia in Asian cancer patients. *Cancer Sci*, 98(9), 1461-1467. <https://doi.org/10.1111/j.1349-7006.2007.00541.x>
- Jitueakul, S., Buncherd, H., Thawornpan, P., Tun, A., & Thanapongpichat, S. (2018). Characterization of G6PD genotypes in G6PD deficiency patients from Suratthani Hospital, Thailand. *J assoc med sci*, 51.
- Kalman, L., Agúndez, J., Appell, M. L., Black, J., Bell, G., Boukouvala, S., Bruckner, C., Bruford, E., Caudle, K., Coulthard, S., Daly, A., Tredici, A. D., den Dunnen, J., Drozda, K., Everts, R., Flockhart, D., Freimuth, R., Gaedigk, A., Hachad, H., . . . Zanger, U. (2016). Pharmacogenetic allele nomenclature: International workgroup recommendations for test result reporting. *Clin Pharmacol Ther*, 99(2), 172-185. <https://doi.org/https://doi.org/10.1002/cpt.280>
- Katara, P., & Yadav, A. (2019). Pharmacogenes (PGx-genes): Current understanding and future directions. *Gene*, 718, 144050.
<https://doi.org/https://doi.org/10.1016/j.gene.2019.144050>
- Katz, D. A., Carr, R., Grimm, D. R., Xiong, H., Holley-Shanks, R., Mueller, T., Leake, B., Wang, Q., Han, L., Wang, P. G., Edeki, T., Sahelijo, L., Doan, T., Allen, A., Spear, B. B., & Kim, R. B. (2006). Organic anion transporting polypeptide 1B1 activity

- classified by SLCO1B1 genotype influences atrasentan pharmacokinetics. *Clin Pharmacol Ther*, 79(3), 186-196. <https://doi.org/10.1016/j.clpt.2005.11.003>
- Kho, Z. Y., & Lal, S. K. (2018). The Human Gut Microbiome – A Potential Controller of Wellness and Disease [Review]. *Front Microbiol*, 9. <https://doi.org/10.3389/fmicb.2018.01835>
- Kim, E. H., Lee, S., Park, J., Lee, K., Bhak, J., & Kim, B. C. (2014). New lung cancer panel for high-throughput targeted resequencing. *Genomics Inform*, 12(2), 50-57. <https://doi.org/10.5808/gi.2014.12.2.50>
- Kim, J. Y., Cheong, H. S., Park, B. L., Kim, L. H., Namgoong, S., Kim, J. O., Kim, H. D., Kim, Y. H., Chung, M. W., Han, S. Y., & Shin, H. D. (2014). Comprehensive variant screening of the UGT gene family. *Yonsei Med J*, 55(1), 232-239. <https://doi.org/10.3349/ymj.2014.55.1.232>
- Kim, K. P., Jang, G., Hong, Y. S., Lim, H. S., Bae, K. S., Kim, H. S., Lee, S. S., Shin, J. G., Lee, J. L., Ryu, M. H., Chang, H. M., Kang, Y. K., & Kim, T. W. (2011). Phase II study of S-1 combined with oxaliplatin as therapy for patients with metastatic biliary tract cancer: influence of the CYP2A6 polymorphism on pharmacokinetics and clinical activity. *Br J Cancer*, 104(4), 605-612. <https://doi.org/10.1038/bjc.2011.17>
- Lacko, M., Roelofs, H. M., Te Morsche, R. H., Voogd, A. C., Ophuis, M. B., Peters, W. H., & Manni, J. J. (2010). Genetic polymorphism in the conjugating enzyme UGT1A1 and the risk of head and neck cancer. *Int J Cancer*, 127(12), 2815-2821. <https://doi.org/10.1002/ijc.25296>
- Lee, S. B., Wheeler, M. M., Patterson, K., McGee, S., Dalton, R., Woodahl, E. L., Gaedigk, A., Thummel, K. E., & Nickerson, D. A. (2019(a)). Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet Med*, 21(2), 361-372. <https://doi.org/10.1038/s41436-018-0054-0>
- Lee, S. B., Wheeler, M. M., Thummel, K. E., & Nickerson, D. A. (2019(b)). Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clin Pharmacol Ther*, 106(6), 1328-1337. <https://doi.org/10.1002/cpt.1552>
- Leitão, L. P. C., Souza, T. P., Rodrigues, J. C. G., Fernandes, M. R., Santos, S., & Santos, N. P. C. (2020). The Metabolization Profile of the CYP2D6 Gene in Amerindian Populations: A Review. *Genes*, 11(3), 262. <https://www.mdpi.com/2073->

[4425/11/3/262](https://doi.org/10.18637/journal.v1i1.4425)

- Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *J Stat Softw*, 69(i01).
<https://doi.org/http://hdl.handle.net/10.18637/journal.v1i1.4425>. (Journal of Statistical Software)
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.
<https://doi.org/10.1093/bioinformatics/btp324>
- Li, L. M., Chen, L., Deng, G. H., Tan, W. T., Dan, Y. J., Wang, R. Q., & Chen, W. S. (2012). SLCO1B1 *15 haplotype is associated with rifampin-induced liver injury. *Mol Med Rep*, 6(1), 75-82. <https://doi.org/10.3892/mmr.2012.900>
- Lim, S., Alshagga, M., Ong, C., Chieng, J., & Pan, Y. (2020). Cytochrome P450 4B1 (CYP4B1) as a target in cancer treatment. *Hum Exp Toxicol*, 39(6), 785-796.
<https://doi.org/10.1177/0960327120905959>
- Lingenhel, A., Kollerits, B., Schwaiger, J. P., Hunt, S. C., Gress, R., Hopkins, P. N., Schoenborn, V., Heid, I. M., & Kronenberg, F. (2008). Serum bilirubin levels, UGT1A1 polymorphisms and risk for coronary artery disease. *Exp Gerontol*, 43(12), 1102-1107. <https://doi.org/10.1016/j.exger.2008.08.047>
- Liu, H., Xu, Q., Huang, W., Zhao, Q., Jiang, Z., Kuang, X., Li, Z., Sun, H., & Qiu, X. (2019). CYP3A5 and CYP3A7 genetic polymorphisms affect tacrolimus concentration in pediatric patients with nephrotic range proteinuria. *Eur J Clin Pharmacol* 75(11), 1533-1540. <https://doi.org/10.1007/s00228-019-02726-w>
- Liu, J., Zhao, R., Ye, Z., Frey, A. J., Schriver, E. R., Snyder, N. W., & Hebbing, S. J. (2017). Relationship of SULT1A1 copy number variation with estrogen metabolism and human health. *J Steroid Biochem Mol Biol*, 174, 169-175.
<https://doi.org/10.1016/j.jsbmb.2017.08.017>
- Liu, Y., Liao, J., & Lu, Q. (2015). Chapter 2 - Laboratory Methods in Epigenetics. In Q. Lu, C. C. Chang, & B. C. Richardson (Eds.), *Epigenetics and Dermatology* (pp. 7-35). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-800957-4.00002-3>
- LL., A., Naranjo, M. E., Rodrigues-Soares, F., Penas, L. E. M., Fariñas, H., & Tarazona-Santos, E. (2014). Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin Drug*

Metab Toxicol, 10(11), 1569-1583.

<https://doi.org/10.1517/17425255.2014.964204>

Locharenkul, C., Shotelersuk, V., & Hirankarn, N. (2010). HLA-B* 1502 screening: time to clinical practice. *Epilepsia*, 51(5), 936-938. <https://doi.org/10.1111/j.1528-1167.2010.02549.x>

López-Flores, L. A., Pérez-Rubio, G., & Falfán-Valencia, R. (2017). Distribution of polymorphic variants of CYP2A6 and their involvement in nicotine addiction. *Excli j*, 16, 174-196. <https://doi.org/10.17179/excli2016-847>

López Aspiroz, E., Cabrera Figueroa, S. E., Valverde Merino, M. P., & Carracedo Álvarez, Á. (2019). Individualized Protease Inhibitor Monotherapy: The Role of Pharmacokinetics and Pharmacogenetics in an Aged and Heavily Treated HIV-Infected Patient. *Clin Drug Investig*, 39(11), 1125-1131. <https://doi.org/10.1007/s40261-019-00829-x>

Madadi, P., Koren, G., Cairns, J., Chitayat, D., Gaedigk, A., Leeder, J. S., Teitelbaum, R., Karaskov, T., & Aleksa, K. (2007). Safety of codeine during breastfeeding: fatal morphine poisoning in the breastfed neonate of a mother prescribed codeine. *Can Fam Physician*, 53(1), 33-35.

Maher, C. A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., & Chinnaiyan, A. M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 458(7234), 97-101. <https://doi.org/10.1038/nature07638>

Martin, J., Han, C., Gordon, L. A., Terry, A., Prabhakar, S., She, X., Xie, G., Hellsten, U., Chan, Y. M., Altherr, M., Couronne, O., Aerts, A., Bajorek, E., Black, S., Blumer, H., Branscomb, E., Brown, N. C., Bruno, W. J., Buckingham, J. M., . . . Pennacchio, L. A. (2004). The sequence and analysis of duplication-rich human chromosome 16. *Nature*, 432(7020), 988-994. <https://doi.org/10.1038/nature03187>

Mauleekoonphairoj, J., Chamnanphon, M., Khongphatthanayothin, A., Sutjaporn, B., Wandee, P., Poovorawan, Y., Nademanee, K., Pongpanich, M., & Chariyavilaskul, P. (2020). Phenotype prediction and characterization of 25 pharmacogenes in Thais from whole genome sequencing for clinical implementation. *Sci Rep*, 10(1), 18969. <https://doi.org/10.1038/s41598-020-76085-3>

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, *20*(9), 1297-1303.
<https://doi.org/10.1101/gr.107524.110>
- Nakajima, M., Fukami, T., Yamanaka, H., Higashi, E., Sakai, H., Yoshida, R., Kwon, J.-T., McLeod, H. L., & Yokoi, T. (2006). Comprehensive evaluation of variability in nicotine metabolism and CYP2A6 polymorphic alleles in four ethnic populations. *Clin Pharmacol Ther*, *80*(3), 282-297.
<https://doi.org/https://doi.org/10.1016/j.clpt.2006.05.012>
- Niemi, M., Pasanen, M. K., & Neuvonen, P. J. (2011). Organic Anion Transporting Polypeptide 1B1: a Genetically Polymorphic Transporter of Major Importance for Hepatic Drug Uptake. *Pharmacol Rev*, *63*(1), 157-181.
<https://doi.org/10.1124/pr.110.002857>
- Numanagić, I., Malikić, S., Ford, M., Qin, X., Toji, L., Radovich, M., Skaar, T. C., Pratt, V. M., Berger, B., Scherer, S., & Sahinalp, S. C. (2018). Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat Commun*, *9*(1), 828. <https://doi.org/10.1038/s41467-018-03273-1>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizkadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., . . . Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44-53. <https://doi.org/10.1126/science.abj6987>
- Oates, J. T., & Lopez, D. (2018). Pharmacogenetics: An Important Part of Drug Development with A Focus on Its Application. *Int J Biomed Investig*, *1*(2).
<https://doi.org/10.31531/2581-4745.1000111>
- Oh, E. S., Kim, C. O., Cho, S. K., Park, M. S., & Chung, J. Y. (2013). Impact of ABCC2, ABCG2 and SLCO1B1 polymorphisms on the pharmacokinetics of pitavastatin in humans. *Drug Metab Pharmacokinet*, *28*(3), 196-202.
<https://doi.org/10.2133/dmpk.dmpk-12-rg-068>
- Onoue, M., Terada, T., Kobayashi, M., Katsura, T., Matsumoto, S., Yanagihara, K.,

- Nishimura, T., Kanai, M., Teramukai, S., Shimizu, A., Fukushima, M., & Inui, K. (2009). UGT1A1*6 polymorphism is most predictive of severe neutropenia induced by irinotecan in Japanese cancer patients. *Int J Clin Oncol*, *14*(2), 136-142. <https://doi.org/10.1007/s10147-008-0821-z>
- Pavlopoulos, G. A., Oulas, A., Iacucci, E., Sifrim, A., Moreau, Y., Schneider, R., Aerts, J., & Iliopoulos, I. (2013). Unraveling genomic variation from next generation sequencing data. *BioData Min*, *6*(1), 13. <https://doi.org/10.1186/1756-0381-6-13>
- Peamkrasatam, S., Sriwatanakul, K., Kiyotani, K., Fujieda, M., Yamazaki, H., Kamataki, T., & Yoovathaworn, K. (2006). In vivo Evaluation of Coumarin and Nicotine as Probe Drugs to Predict the Metabolic Capacity of CYP2A6 Due to Genetic Polymorphism in Thais. *Drug Metab Pharmacokinet*, *21*(6), 475-484. <https://doi.org/https://doi.org/10.2133/dmpk.21.475>
- Pei, Q., Liu, J. Y., Yin, J. Y., Yang, G. P., Liu, S. K., Zheng, Y., Xie, P., Guo, C. X., Luo, M., Zhou, H. H., Li, X., & Liu, Z. Q. (2018). Repaglinide-irbesartan drug interaction: effects of SLCO1B1 polymorphism on repaglinide pharmacokinetics and pharmacodynamics in Chinese population. *Eur J Clin Pharmacol* *74*(8), 1021-1028. <https://doi.org/10.1007/s00228-018-2477-6>
- Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nat. Genet*, *39*(10), 1256-1260. <https://doi.org/10.1038/ng2123>
- Petrović, J., Pešić, V., & Lauschke, V. M. (2020). Frequencies of clinically important CYP2C19 and CYP2D6 alleles are graded across Europe. *Eur J Hum Genet*, *28*(1), 88-94. <https://doi.org/10.1038/s41431-019-0480-8>
- Picard, N., Yee, S. W., Woillard, J. B., Lebranchu, Y., Le Meur, Y., Giacomini, K. M., & Marquet, P. (2010). The role of organic anion-transporting polypeptides and their common genetic variants in mycophenolic acid pharmacokinetics. *Clin Pharmacol Ther*, *87*(1), 100-108. <https://doi.org/10.1038/clpt.2009.205>
- Piriyapongsa, J., Sukritha, C., Kaewprommal, P., Intarat, C., Triparn, K., Phornsiricharoenphant, K., Chaosrikul, C., Shaw, P. J., Chantratita, W., Mahasirimongkol, S., & Tongsimma, S. (2021). PharmVIP: A Web-Based Tool for

- Pharmacogenomic Variant Analysis and Interpretation. *J Pers Med*, 11(11).
<https://doi.org/10.3390/jpm11111230>
- Profaizer, T., Pole, A., Monds, C., Delgado, J. C., & Lázár-Molnár, E. (2020). Clinical utility of next generation sequencing based HLA typing for disease association and pharmacogenetic testing. *Hum Immunol*, 81(7), 354-360.
<https://doi.org/https://doi.org/10.1016/j.humimm.2020.05.001>
- Relling, M. V., & Klein, T. E. (2011). CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther*, 89(3), 464-467. <https://doi.org/10.1038/clpt.2010.279>
- Relling, M. V., McDonagh, E. M., Chang, T., Caudle, K. E., McLeod, H. L., Haidar, C. E., Klein, T., & Luzzatto, L. (2014). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for Rasburicase Therapy in the Context of G6PD Deficiency Genotype. *Clin Pharmacol Ther*, 96(2), 169-174.
<https://doi.org/https://doi.org/10.1038/clpt.2014.97>
- Robarge, J. D., Li, L., Desta, Z., Nguyen, A., & Flockhart, D. A. (2007). The Star-Allele Nomenclature: Retooling for Translational Genomics. *Clin Pharmacol Ther*, 82(3), 244-248. <https://doi.org/https://doi.org/10.1038/sj.clpt.6100284>
- Sagar, M., Bertilsson, L., Stridsberg, M., Kjellin, A., Mårdh, S., & Seensalu, R. (2000). Omeprazole and CYP2C19 polymorphism: effects of long-term treatment on gastrin, pepsinogen I, and chromogranin A in patients with acid related disorders. *Aliment Pharmacol Ther*, 14(11), 1495-1502. <https://doi.org/10.1046/j.1365-2036.2000.00835.x>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74(12), 5463-5467.
<https://doi.org/10.1073/pnas.74.12.5463>
- Sanguhl, K., Whirl-Carrillo, M., Whaley, R. M., Woon, M., Lavertu, A., Altman, R. B., Carter, L., Verma, A., Ritchie, M. D., & Klein, T. E. (2020). Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clin Pharmacol Ther*, 107(1), 203-210.
<https://doi.org/https://doi.org/10.1002/cpt.1568>
- Sangsin, A., Srichomthong, C., Pongpanich, M., & Suphapeetiporn, K. (2016). Whole-exome sequencing reveals a novel COL2A1 mutation in a patient with

spondyloepiphyseal dysplasia congenita. *Genet Mol Res* 15.

<https://doi.org/10.4238/gmr.15017624>

- Sasaki, T., Horikawa, M., Orikasa, K., Sato, M., Arai, Y., Mitachi, Y., Mizugaki, M., Ishikawa, M., & Hiratsuka, M. (2008). Possible Relationship Between the Risk of Japanese Bladder Cancer Cases and the CYP4B1 Genotype. *Jpn J Clin Oncol* 38(9), 634-640. <https://doi.org/10.1093/jjco/hyn081>
- Shibuya, A., Hirono, A., Ishii, S., Fujii, H., & Miwa, S. (1999). Hemolytic crisis after excessive ingestion of fava beans in a male infant with G6PD Canton. *Int J Hematol*, 70(4), 233-235.
- Shitara, Y., Maeda, K., Ikejiri, K., Yoshida, K., Horie, T., & Sugiyama, Y. (2013). Clinical significance of organic anion transporting polypeptides (OATPs) in drug disposition: their roles in hepatic clearance and intestinal absorption. *Biopharm. Drug Dispos*, 34(1), 45-78. <https://doi.org/10.1002/bdd.1823>
- Shotelersuk, V., Limwongse, C., & Mahasirimongkol, S. (2014). Genetics and genomics in Thailand: challenges and opportunities. *Mol Genet Genom* 2(3), 210-216. <https://doi.org/https://doi.org/10.1002/mgg3.83>
- Shotelersuk, V., Tongsimma, S., Pithukpakorn, M., Eu-ahsunthornwattana, J., & Mahasirimongkol, S. (2019). Precision medicine in Thailand. *Am J Med Genet C Semin Med Genet* 181(2), 245-253. <https://doi.org/https://doi.org/10.1002/ajmg.c.31694>
- Shotelersuk, V., Wichadakul, D., Ngamphiw, C., Srichomthong, C., Phokaew, C., Wilantho, A., Pakchuen, S., Nakhonsri, V., Shaw, P. J., Wasitthankasem, R., Piriyaongsa, J., Wangkumhang, P., Assawapitaksakul, A., Chetruengchai, W., Lapphra, K., Khuninthong, A., Makarawate, P., Suphapeetiporn, K., Mahasirimongkol, S., . . . Tongsimma, S. (2021). The Thai reference exome (T-REx) variant database. *Clinical Genetics*, 100(6), 703-712. <https://doi.org/https://doi.org/10.1111/cge.14060>
- Soejima, M., Sugiura, T., Kawaguchi, Y., Kawamoto, M., Katsumata, Y., Takagi, K., Nakajima, A., Mitamura, T., Mimori, A., Hara, M., & Kamatani, N. (2007). Association of the diplotype configuration at the N-acetyltransferase 2 gene with adverse events with co-trimoxazole in Japanese patients with systemic lupus erythematosus. *Arthritis Res Ther*, 9(2), R23. <https://doi.org/10.1186/ar2134>

- Songsungsan, P., Chantanakool, P., Suratane, A., Buaboocha, T., Comai, L., Chadchawan, S., & Plaimas, K. (2021). Identification of Key Genes in 'Luang Pratahn', Thai Salt-Tolerant Rice, Based on Time-Course Data and Weighted Co-expression Networks [Original Research]. *Front Plant Sci*, 12. <https://doi.org/10.3389/fpls.2021.744654>
- Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., . . . the Bipolar Disorder Working Group of the Psychiatric Genomics, C. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet*, 51(5), 793-803. <https://doi.org/10.1038/s41588-019-0397-8>
- Suvichapanich, S., Fukunaga, K., Zahroh, H., Mushiroda, T., Mahasirimongkol, S., Toyooka, L., Chaikledkaew, U., Jittikoon, J., Yuliwulandari, R., Yanai, H., Wattanapokayakit, S., & Tokunaga, K. (2018). NAT2 ultra-slow acetylator and risk of anti-tuberculosis drug-induced liver injury: a genotype-based meta-analysis. *Pharmacogenet Genomics*, 28(7), 167-176. <https://doi.org/10.1097/fpc.0000000000000339>
- Suwannakul, S., Ieiri, I., Kimura, M., Kawabata, K., Kusuhara, H., Hirota, T., Irie, S., Sugiyama, Y., & Higuchi, S. (2008). Pharmacokinetic interaction between pravastatin and olmesartan in relation to SLCO1B1 polymorphism. *Journal of Human Genetics*, 53(10), 899-904. <https://doi.org/10.1007/s10038-008-0324-9>
- Takahashi, T., Luzum, J. A., Nicol, M. R., & Jacobson, P. A. (2020). Pharmacogenomics of COVID-19 therapies. *NPJ Genom Med* 5, 35. <https://doi.org/10.1038/s41525-020-00143-y>
- Tang, K. S., Chiu, H. F., Chen, H. H., Eng, H. L., Tsai, C. J., Teng, H. C., & Huang, C. S. (2005). Link between colorectal cancer and polymorphisms in the uridine-diphosphoglucuronosyltransferase 1A7 and 1A1 genes. *World J Gastroenterol*, 11(21), 3250-3254. <https://doi.org/10.3748/wjg.v11.i21.3250>
- Tanigawara, Y., Aoyama, N., Kita, T., Shirakawa, K., Komada, F., Kasuga, M., & Okumura, K. (1999). CYP2C19 genotype-related efficacy of omeprazole for the treatment of

- infection caused by *Helicobacter pylori*. *Clin Pharmacol Ther*, 66(5), 528-534.
[https://doi.org/10.1016/s0009-9236\(99\)70017-2](https://doi.org/10.1016/s0009-9236(99)70017-2)
- Tanii, H., Shitara, Y., & Horie, T. (2011). Population pharmacokinetic analysis of letrozole in Japanese postmenopausal women. *Eur J Clin Pharmacol* 67(10), 1017-1025.
<https://doi.org/10.1007/s00228-011-1042-3>
- Thervet, E., Anglicheau, D., King, B., Schlageter, M. H., Cassinat, B., Beaune, P., Legendre, C., & Daly, A. K. (2003). Impact of cytochrome p450 3A5 genetic polymorphism on tacrolimus doses and concentration-to-dose ratio in renal transplant recipients. *Transplantation*, 76(8), 1233-1235.
<https://doi.org/10.1097/01.Tp.0000090753.99170.89>
- Tong, H., Phan, N. V. T., Nguyen, T. T., Nguyen, D. V., Vo, N. S., & Le, L. (2021). Review on Databases and Bioinformatic Approaches on Pharmacogenomics of Adverse Drug Reactions. *Pharmgenomics Pers Med*, 14, 61-75.
<https://doi.org/10.2147/pgpm.S290781>
- Tournel, G., Cauffiez, C., Billaut-Laden, I., Allorge, D., Chevalier, D., Bonnifet, F., Mensier, E., Lafitte, J. J., Lhermitte, M., Broly, F., & Lo-Guidice, J. M. (2007). Molecular analysis of the CYP2F1 gene: identification of a frequent non-functional allelic variant. *Mutat Res*, 617(1-2), 79-89.
<https://doi.org/10.1016/j.mrfmmm.2007.01.007>
- Twesigomwe, D., Drögemöller, B. I., Wright, G. E. B., Siddiqui, A., da Rocha, J., Lombard, Z., & Hazelhurst, S. (2021). StellarPGx: A Nextflow Pipeline for Calling Star Alleles in Cytochrome P450 Genes. *Clin Pharmacol Ther*, 110(3), 741-749.
<https://doi.org/https://doi.org/10.1002/cpt.2173>
- Twesigomwe, D., Wright, G. E. B., Drögemöller, B. I., da Rocha, J., Lombard, Z., & Hazelhurst, S. (2020). A systematic comparison of pharmacogene star allele calling bioinformatics algorithms: a focus on CYP2D6 genotyping. *npj Genomic Medicine*, 5(1), 30. <https://doi.org/10.1038/s41525-020-0135-2>
- Twist, G. P., Gaedigk, A., Miller, N. A., Farrow, E. G., Willig, L. K., Dinwiddie, D. L., Petrikin, J. E., Soden, S. E., Herd, S., Gibson, M., Cakici, J. A., Riffel, A. K., Leeder, J. S., Dinakarandian, D., & Kingsmore, S. F. (2016). Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene,

- CYP2D6, from whole-genome sequences. *npj Genomic Medicine*, 1(1), 15007.
<https://doi.org/10.1038/npjgenmed.2015.7>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 43(1110), 11.10.11-11.10.33.
<https://doi.org/10.1002/0471250953.bi1110s43>
- Watson, J. D., & Crick, F. H. (1953). The structure of DNA. *Cold Spring Harb Symp Quant Biol*, 18, 123-131. <https://doi.org/10.1101/sqb.1953.018.01.020>
- Wilantho, A., Praditsap, O., Charoenchim, W., Kulawonganchai, S., Assawamakin, A., & Tongsim, S. (2012). Next generation sequencing (NGS) technologies and their applications in omics-research. *Thai J Genet* 5.
- Williams, Elsie M., Little, Rory F., Mowday, Alexandra M., Rich, Michelle H., Chan-Hyams, Jasmine V. E., Copp, Janine N., Smalls, Jeff B., Patterson, Adam V., & Ackerley, David F. (2015). Nitroreductase gene-directed enzyme prodrug therapy: insights and advances toward clinical utility. *Biochem J* 471(2), 131-153.
<https://doi.org/10.1042/bj20150650>
- Yamamoto, K., Sato, H., Fujiyama, Y., Doida, Y., & Bamba, T. (1998). Contribution of two missense mutations (G71R and Y486D) of the bilirubin UDP glycosyltransferase (UGT1A1) gene to phenotypes of Gilbert's syndrome and Crigler-Najjar syndrome type II. *Biochim Biophys Acta*, 1406(3), 267-273. [https://doi.org/10.1016/s0925-4439\(98\)00013-1](https://doi.org/10.1016/s0925-4439(98)00013-1)
- Yang, B., Lei, C., Yang, D., Tan, Z., Guo, T., & Luo, H. (2021). Whole-Exome Sequencing Identified CFTR Variants in Two Consanguineous Families in China [Original Research]. *Front. Genet*, 12. <https://doi.org/10.3389/fgene.2021.631221>
- Yee, J., Kim, S. M., Han, J. M., Lee, N., Yoon, H. Y., & Gwak, H. S. (2020). The association between NAT2 acetylator status and adverse drug reactions of sulfasalazine: a systematic review and meta-analysis. *Sci Rep*, 10(1), 3658.
<https://doi.org/10.1038/s41598-020-60467-8>
- Yoshida, R., Nakajima, M., Watanabe, Y., Kwon, J.-T., & Yokoi, T. (2002). Genetic

polymorphisms in human CYP2A6 gene causing impaired nicotine metabolism. *Br J Clin Pharmacol* 54(5), 511-517.

<https://doi.org/https://doi.org/10.1046/j.1365-2125.2002.01667.x>

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nat Rev Genet*, 16(3), 172-183.

<https://doi.org/10.1038/nrg3871>

Zastrozhin, M., Skryabin, V., Smirnov, V., Zastrozhina, A., Grishina, E., Ryzhikova, K., Bure, I., Golovinskii, P., Koporov, S., Bryun, E., & Sychev, D. (2021). Effect of Genetic Polymorphism of the CYP2D6 Gene on the Efficacy and Safety of Fluvoxamine in Major Depressive Disorder. *Am J Ther*, 29(1), e26-e33.

<https://doi.org/10.1097/mjt.0000000000001388>

Zhang, H., De, T., Zhong, Y., & Perera, M. A. (2019). The Advantages and Challenges of Diversity in Pharmacogenomics: Can Minority Populations Bring Us Closer to Implementation? *Clin Pharmacol Ther*, 106(2), 338-349.

<https://doi.org/10.1002/cpt.1491>

Zhang, J.-J., Zhang, H., Ding, X.-L., Ma, S., & Miao, L.-Y. (2013). Effect of the P450 oxidoreductase *28 polymorphism on the pharmacokinetics of tacrolimus in Chinese healthy male volunteers. *Eur J Clin Pharmacol* 69(4), 807-812.

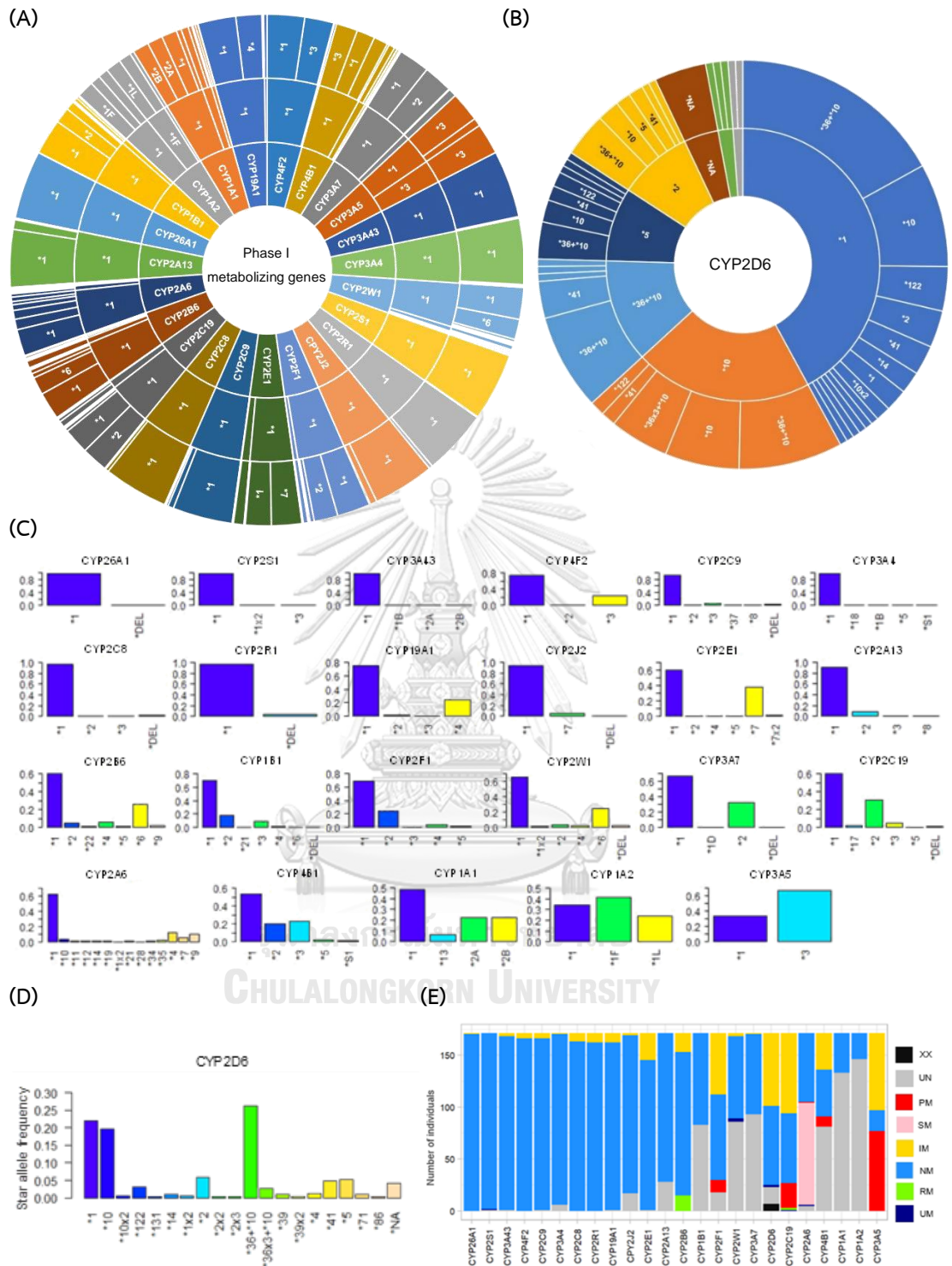
<https://doi.org/10.1007/s00228-012-1432-1>

Zhou, Y., Ingelman-Sundberg, M., & Lauschke, V. M. (2017). Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clin Pharmacol Ther*, 102(4), 688-700. <https://doi.org/10.1002/cpt.690>

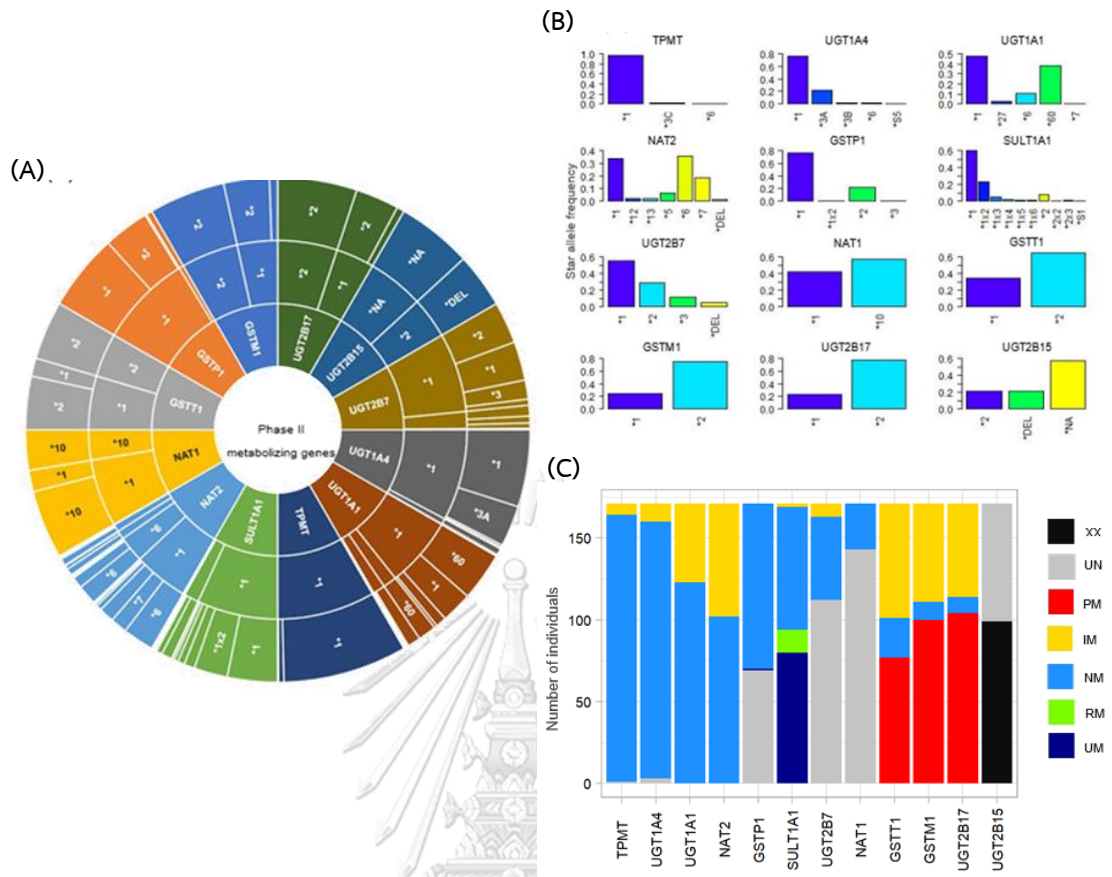


ภาคผนวก

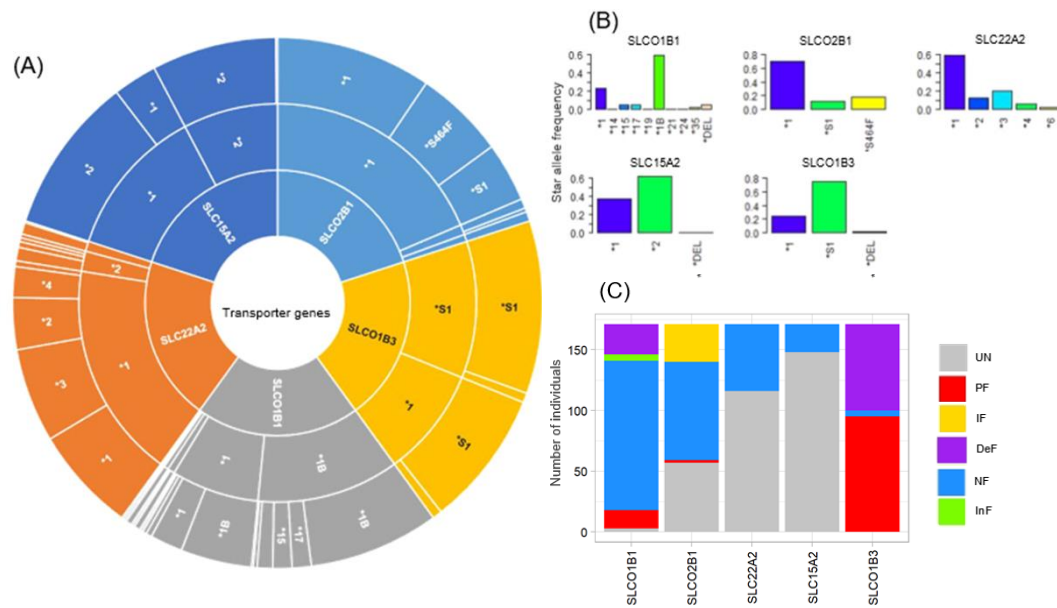
จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



ภาพที่ 1 ผลการทำนายรูปแบบการแปรผันทางพันธุกรรมในยีนที่เกี่ยวข้องกับการเมแทบอลิซึมเฟส 1. (A และ B) แผนภาพวงกลม โดยวงนอกสุดจะแสดงสัดส่วนของดีโพลไทป์ของแต่ละยีน และสำหรับยีน *CYP2D6* (C และ D) กราฟแท่งแสดงความถี่ของ star allele (E) จำนวนฟีโนไทป์ของแต่ละยีนที่พบในแต่ละคน โดย unpredictable diplotype (XX; สีดำ); unknown function (UN; สีเทา); poor metabolizer (PM; สีแดง); slow metabolizer (SM; สีชมพู); intermediate metabolizer (IM; สีเหลือง); normal metabolizer (NM; สีฟ้า); rapid metabolizer (RM; สีเขียว); ultra-rapid metabolizer (UM; สีน้ำเงิน).



ภาพที่ 2 ผลการทำนายรูปแบบการแปรผันทางพันธุกรรมในยีนที่เกี่ยวข้องกับการเมแทบอลิซึมเฟส 2. (A) แผนภาพวงกลม โดยวงนอกสุดจะแสดงสัดส่วนของดีโพลไทป์ของแต่ละยีน (B) กราฟแท่งแสดงความถี่ของ star allele (C) จำนวนฟีโนไทป์ของแต่ละยีนที่พบในแต่ละคน โดย unpredictable diplotype (XX; สีดำ); unknown function (UN; สีเทา); poor metabolizer (PM; สีแดง); intermediate metabolizer (IM; สีเหลือง); normal metabolizer (NM; สีฟ้า); rapid metabolizer (RM; สีเขียว); ultra-rapid metabolizer (UM; สีน้ำเงิน).



ภาพที่ 3 ผลการทำนายรูปแบบการแปรผันทางพันธุกรรมในยีนที่เกี่ยวข้องกับการขนส่งยา. (A) แผนภาพวงกลม โดยวงนอกสุดจะแสดงสัดส่วนของดีโพลไทป์ของแต่ละยีน (B) กราฟแท่งแสดงความถี่ของ star allele (C) จำนวนฟีโนไทป์ของแต่ละยีนที่พบในแต่ละคน โดย unknown function (UN; สีเทา); poor function (PF; สีแดง); intermediate function (IF; สีเหลือง); decrease function (DeF; สีม่วง); normal function (NF; สีฟ้า); increase function (InF; สีเขียว)

ตารางที่ 1 ตารางแสดงโปรไฟล์ของ star alleles ที่พบในยีนที่เกี่ยวข้องกับการตอบสนองต่อยา 51 ยีน และคำอธิบายประกอบทางคลินิกของยีนระดับต่าง ๆ ตามระบบ CPIC

| Genes | Star alleles (n=196) | Clinical annotation levels of evidence |
|---|---|--|
| Phase I metabolizing genes (Cytochrome P450) (n=105) | | |
| <i>CYP19A1</i> | *2, *3, *4 | 3 |
| <i>CYP1A1</i> | *13, *2A, *2B | 3 |
| <i>CYP1A2</i> | *1F, *1L | 3, 4 |
| <i>CYP1B1</i> | *2, *21, *3, *4, *6, *DEL | 3 |
| <i>CYP2A6</i> | *1x2(dup), *10, *11, *12, *14, *19, *21, *28, *34, *35, *4, *7, *9 | 1B, 3, 4 |
| <i>CYP26A1</i> | *DEL | No data |
| <i>CYP2A13</i> | *2, *3, *8 | 3 |
| <i>CYP2B6</i> | *2, *22, *4, *5, *6, *9 | 1A, 2A, 3, 4 |
| <i>CYP2C19</i> | *17, *2, *3, *5, *DEL | 1A, 2A, 3, 4 |
| <i>CYP2C8</i> | *2, *3, *DEL | 3, 4 |
| <i>CYP2C9</i> | *2, *3, *37, *8, *DEL | 1A, 1B, 2A, 3, 4 |
| <i>CYP2D6</i> | *1x2(dup), *10, *10x2(dup), *122, *131, *14, *2, *2x2(dup), *2x3(multi), *36+*10(rea), *36x3+*10(rea), *39, *39x2(dup), *4, *41, *5(del), *71, *86, *NA | 1A, 2A, 3, 4 |
| <i>CYP2E1</i> | *2, *4, *5, *7, *7x2(dup) | 3, 4 |
| <i>CYP2F1</i> | *2, *3, *4, *5 | 3 |
| <i>CYP2J2</i> | *7, *DEL | 3 |
| <i>CYP2R1</i> | *DEL | 3 |
| <i>CYP2S1</i> | *1x2(dup), *3 | No data |
| <i>CYP2W1</i> | *1x2(dup), *2, *4, *6, *DEL | No data |
| <i>CYP3A4</i> | *18, *1B, *5, *S1 | 1B, 2A, 3, 4 |
| <i>CYP3A43</i> | *1B, *2A, *2B | 3 |
| <i>CYP3A5</i> | *3 | 1A, 1B, 2A, 3, 4 |
| <i>CYP3A7</i> | *1D, *2, *DEL | 3 |
| <i>CYP4B1</i> | *2, *3, *5, *S1 | 3 |
| <i>CYP4F2</i> | *2, *3 | 1A, 2A, 3, 4 |
| Phase II metabolizing genes (n= 38) | | |
| <i>GSTM1</i> | *2(del) | 3, 4 |
| <i>GSTP1</i> | *1x2(dup), *2, *3 | 3, 4 |
| <i>GSTT1</i> | *2(del) | 3, 4 |
| <i>NAT1</i> | *10 | 3 |
| <i>NAT2</i> | *12, *13, *5, *6, *7, *DEL | 1B, 2A, 3 |
| <i>SULT1A1</i> | *1x2(dup), *1x3(multi), *1x4(multi), *1x5(multi), *1x6(multi), *2, *2x2(multi), *2x3(multi), *S1(del) | 3 |

| | | |
|---------------------------------|--|------------------|
| <i>TPMT</i> | *3C, *6 | 1A, 3, 4 |
| <i>UGT1A1</i> | *27, *6, *60, *7 | 1A, 1B, 2A, 3, 4 |
| <i>UGT1A4</i> | *3A, *3B, *6, *S5 | 3 |
| <i>UGT2B15</i> | *2, *DEL, *NA | 3 |
| <i>UGT2B17</i> | *2(del) | No data |
| <i>UGT2B7</i> | *2, *3, *DEL | 3, 4 |
| Transporter genes (n=19) | | |
| <i>SLC15A2</i> | *2, *DEL | 3 |
| <i>SLC22A2</i> | *2, *3, *4, *6 | 3, 4 |
| <i>SLCO1B1</i> | *14, *15, *17, *19, *1B, *21, *24, *35, *DEL | 1A, 2A, 3, 4 |
| <i>SLCO1B3</i> | *S1, *DEL | 3 |
| <i>SLCO2B1</i> | *S1, *S464F | 3 |
| Drug target genes (n=34) | | |
| <i>CACNA1S</i> | None | 1A |
| <i>CFTR</i> | *DEL | 1A, 3, 4 |
| <i>DPYD</i> | *5, *6, *9A, *S12, *S3, *S4, *S46, *S9, *DEL | 1A, 3, 4 |
| <i>G6PD</i> | *1x2(dup), *21, *28, *31, *50, *51, *8, *DEL | 1A, 3, 4 |
| <i>IFNL3</i> | *S3, *DEL | 1A, 1B, 2A, 3 |
| <i>NUDT15</i> | *2, *3, *5, *6 | 1A, 3 |
| <i>POR</i> | *28, *28x2(dup) | 3, 4 |
| <i>RYR1</i> | *1x2(dup) | 1A, 3 |
| <i>TBXAS1</i> | *2, *7 | 3 |
| <i>VKORC1</i> | *2, *2x2(dup), *3, *3x2(dup), *4 | 1A, 1B, 2A, 3, 4 |

*NA = ลักษณะดีโพลไทป์ที่ไม่สามารถทำนายได้, dup = gene duplication, del = gene deletion, multi = gene multiplication, rea = gene rearrangement

ประวัติผู้เขียน

| | |
|-------------------|---|
| ชื่อ-สกุล | Natnicha Wankaew |
| วัน เดือน ปี เกิด | 24 September 1995 |
| สถานที่เกิด | Chonburi, Thailand |
| ที่อยู่ปัจจุบัน | 118/22 Moo 9 Sukumwit Rd., Sattahip, Chonburi, Thailand 20180 |



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY