

ซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการแพทย์แม่นยำของโรคมะเร็ง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

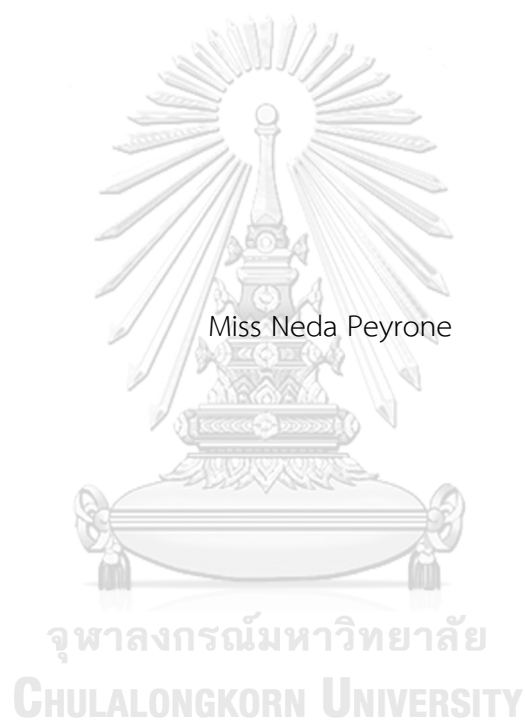
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2562

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

An Integrative Software Platform for Cancer Precision Medicine



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	ซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการแพทย์แม่นยำ ของโรคมะเร็ง
โดย	น.ส.เนต้า เปอีโรเน
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร.ดวงดาว วิชาตากุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

.....	คณบดีคณะวิศวกรรมศาสตร์ (ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)
คณะกรรมการสอบวิทยานิพนธ์ ประธานกรรมการ (รองศาสตราจารย์ ดร.วิวัฒน์ วัฒนาวุฒิ)
..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก (อาจารย์ ดร.ดวงดาว วิชาตากุล)
..... กรรมการ (รองศาสตราจารย์ ดร.เกริก ภิรมย์โสภา)
..... กรรมการภายนอกมหาวิทยาลัย (ดร.ศิษณุศ ทองสีมา)

เนต้า เปอีโรเน : ซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการแพทย์แม่นยำของโรคมะเร็ง. (An Integrative Software Platform for Cancer Precision Medicine)
 อ.ที่ปรึกษาหลัก : อ. ดร.ดวงดาว วิชาดากุล

การแพทย์แม่นยำคือ กลยุทธ์ในการวินิจฉัยโรคและการตัดสินใจเลือกแนวทางการรักษาจากข้อมูลพันธุกรรมเฉพาะบุคคล ด้วยการพัฒนาความสามารถด้านเทคโนโลยีโอมิคส์ไปอย่างรวดเร็ว เช่น การหาลำดับเบสดีเอ็นเอ การหาระดับการแสดงออกของยีนด้วยเทคโนโลยีอาร์เอ็นเอซีควนซิง เป็นต้น จากผลลัพธ์ของกระบวนการตัวอย่างข้างต้น นำมาวิเคราะห์เพื่อตรวจหาลักษณะจำเพาะของยีนกลายพันธุ์ในผู้ป่วยแต่ละราย หรือวัดระดับการแสดงออกของยีนที่ตอบสนองต่อยาเพื่อช่วยแพทย์เฉพาะทางในการออกแบบการรักษาและการเลือกใช้ยาที่เหมาะสมกับผู้ป่วยแต่ละราย ดังนั้น นอกเหนือจากข้อมูลทางคลินิกแล้ว ข้อมูลโอมิคส์จึงกลายเป็นสิ่งจำเป็นในการวินิจฉัยโรคได้อย่างแม่นยำจำเพาะต่อตัวบุคคล โดยวิทยานิพนธ์ฉบับนี้นำเสนอซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการแพทย์แม่นยำของโรคมะเร็ง ซึ่งมีชื่อว่ารันองโค (RUN-ONCO) โดยมีวัตถุประสงค์เพื่อสนับสนุนให้แพทย์เฉพาะทางและนักวิจัยสามารถดูแลจัดการข้อมูลที่หลากหลายได้โดยง่าย และสามารถใช้อ้างอิงข้อมูลที่มีอยู่ได้อย่างมีประสิทธิภาพและก่อให้เกิดประโยชน์แพลตฟอร์มให้บริการจัดเก็บข้อมูลทางคลินิก ข้อมูลชีววัตถุ และข้อมูลโอมิคส์ ที่สนับสนุนการบูรณาการข้อมูลเข้าสู่ระบบได้ รวมทั้งสามารถดำเนินการเชื่อมโยงข้อมูลระหว่างฐานข้อมูลสาธารณะ เช่น ฐานข้อมูลสตริง ฐานข้อมูลองโคเคบี เป็นต้น เนื่องจากงานวิจัยในด้านการวิเคราะห์ข้อมูลโอมิคส์ร่วมกับไลบรารีจาวาสคริปต์ที่ช่วยในการวิเคราะห์และแสดงผลเชิงรูปภาพมีจำนวนเพิ่มมากขึ้นเรื่อย ๆ รันองโคจึงถูกออกแบบให้มีความยืดหยุ่นสูง โดยออกแบบซอฟต์แวร์แพลตฟอร์มภายใต้แนวคิดสถาปัตยกรรม 3-เทียร์ และสถาปัตยกรรมเชิงคอมโพเนนท์ นอกจากนี้ยังมีการนำแบบรูปต่าง ๆ มาประยุกต์ใช้ เพื่อสนับสนุนการเพิ่มกระบวนการวิเคราะห์และแสดงผลเชิงรูปภาพที่หลากหลายได้โดยง่าย

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์

ปีการศึกษา 2562

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาหลัก

6070931421 : MAJOR COMPUTER SCIENCE

KEYWORD: precision medicine, cancer, highly extensible plugin, analysis plugin,
software platform, visualization plugin

Neda Peyrone : An Integrative Software Platform for Cancer Precision
Medicine. Advisor: Duangdao Wichadakul, Ph.D.

Precision medicine is a strategy to personalize disease identification and medical care decisions through genetics. The rapid development of -omics technologies e.g., DNA and RNA sequencing, which reveal specific gene mutations in a patient's tumor or profiling of gene expressions for drug responses, helps oncologists find effective treatments for individual patients based on their genetics. Hence, besides the clinical records, -omics data become essential for personalized diagnosis and treatments. In this paper, a web-based standalone software platform for cancer precision medicine, called RUN-ONCO, is proposed aiming to help oncologists and researchers manage and make use of the available clinical and -omics data easily and efficiently. The platform allows the management of clinical records, biospecimens, and -omics data and enables various integrative data analyses together with public databases such as STRING and OncoKB. With the increasing number of published methods for various -omics data analyses together with the availability of numerous javascript libraries for data visualization. RUN-ONCO was designed and developed as a web-based application with the 3-tier and component-based architecture. There were also several design patterns applied in this platform for enabling the highly extensible plugins for both visualizations and analyses.

Field of Study: Computer Science

Student's Signature

Academic Year: 2019

Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความอนุเคราะห์ของอาจารย์ ดร.ดวงดาว วิชิตากุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งสละเวลาให้คำปรึกษา ช่วยตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ จนทำให้การวิจัยครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณกรรมการสอบวิทยานิพนธ์ ดร.ศิษณุศ ทองสีมา รองศาสตราจารย์ ดร.เกริก ภิรมย์โสภา และรองศาสตราจารย์ ดร.วิวัฒน์ วัฒนาวุฒิ ที่กรุณาสละเวลาให้คำแนะนำ ตรวจสอบและแก้ไขวิทยานิพนธ์ ซึ่งเป็นประโยชน์ในการทำวิทยานิพนธ์ฉบับนี้อย่างยิ่ง

ขอขอบพระคุณอาจารย์ นพ.ไตรรักษ์ พิสิษฐ์กุล อาจารย์ ดร.ธรรมกร แซ่ตั้ง สรินยา คำปัญญา ที่สนับสนุนตัวอย่างข้อมูลโอมิกส์ต่าง ๆ พร้อมแนวทางการวิเคราะห์ข้อมูลและแปลผล นอกจากนี้ต้องขอขอบคุณ พญ.ศุภิพร อุตมาภินันท์ ที่กรุณาสละเวลาให้รายละเอียดข้อมูล กำหนดของการทำงานของซอฟต์แวร์รวมทั้งตัวอย่างข้อมูลทางคลินิก และแนวทางการเชื่อมโยงข้อมูลกับฐานข้อมูลสาธารณะ เช่น อองโคเคบี เพื่อการวินิจฉัยโรคมะเร็งที่แม่นยำยิ่งขึ้น

ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทความรู้ ตลอดจนให้ข้อคิดต่าง ๆ ซึ่งเป็นแรงผลักดันสำคัญ ให้มีความมุ่งมั่นตั้งใจทำงานวิจัยนี้ให้ดีที่สุด นอกจากนี้ต้องขอขอบคุณรองศาสตราจารย์ ดร.วรา วราวิทย์ ที่กรุณาสละเวลาให้คำปรึกษาและข้อคิดที่เป็นประโยชน์ในการทำวิทยานิพนธ์

ท้ายที่สุด ขอกราบขอบพระคุณ ครอบครัวสำหรับกำลังใจ รวมถึงขอขอบพระคุณผู้บังคับบัญชาในสายงาน เพื่อนร่วมงาน รุ่นพี่ รุ่นน้องและมิตรสหาย ที่คอยให้กำลังใจและช่วยเหลือในด้านต่าง ๆ

สุดท้ายนี้ ข้าพเจ้าหวังเป็นอย่างยิ่งว่า เนื้อหาในวิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์แก่ผู้อื่นไม่มากนัก

เนต้า เป็โรเน

สารบัญ

หน้า

.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ญ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตงานวิจัย	3
1.4 ขั้นตอนและวิธีการดำเนินการวิจัย.....	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.6 บทความวิชาการที่ได้รับการตีพิมพ์.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	5
2.1 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1.1 การหาลำดับเบสดีเอ็นเอ (DNA sequencing).....	5
2.1.2 การถอดรหัสจีโนม (Whole Genome Sequencing: WGS).....	6
2.1.3 การถอดรหัสเอ็กโซม (Whole Exome Sequencing : WES).....	6
2.1.4 การแปรผันทางพันธุกรรม (Genetic variation)	6

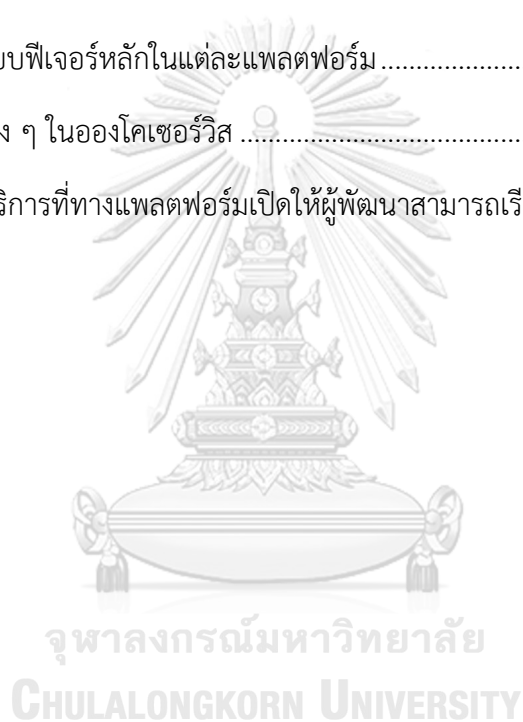
2.1.4.1	ซิงเกิลนิวคลีโอไทด์โพลีมอร์ฟิซึม (Single Nucleotide Polymorphisms)...	6
2.1.4.2	อินเดล (Insertions and deletions: Indels).....	7
2.1.4.3	การแปรผันเชิงโครงสร้าง (Structural variation)	7
2.1.5	รูปแบบไฟล์วีซีเอฟ (Variant Call Format: VCF).....	9
2.1.6	ไฟล์ที่เกี่ยวกับการแสดงออกของยีน หรือไฟล์อีเอ็กซ์พี (Expression Matrix Format: EXP)	10
2.1.7	อีทีแอล (Extract-Transform-Load: ETL).....	11
2.1.8	สถาปัตยกรรม 3-เทียร์ (3-Tier Architecture)	13
2.1.9	สถาปัตยกรรมเชิงคอมโพเนนต์ (Component-based Architecture: CBA)	13
2.1.10	แบบรูปพะชาด (Facade Pattern).....	14
2.1.11	แบบรูปดีเพนเดนซีอินเจกชัน (Dependency Injection: DI Pattern).....	14
2.1.12	โมเดล-วิว-คอนโทรลเลอร์ (Model-View-Controller: MVC)	15
2.1.13	ตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ (Object-Relational Mapping: ORM).....	16
2.1.14	วัตถุการเข้าถึงข้อมูล (Data Access Object: DAO).....	16
2.2	งานวิจัยที่เกี่ยวข้อง	17
2.2.1	จีโนมิกส์พอร์ทัลสำหรับโรคมะเร็งระบบศีรษะและลำคอ (GENIPAC: A Genomic Information Portal for Head and Neck Cancer Cell Systems).....	17
2.2.2	เว็บทีซีจีเอแพลตฟอร์มออนไลน์สำหรับการวิเคราะห์โรคมะเร็งแบบบูรณาการในระดับ โมเลกุล (Web-TCGA: an online platform for integrated analysis of molecular cancer data sets).....	20
2.2.3	เว็บพอร์ทัลแบบบูรณาการสำหรับจีโนมมะเร็ง (canEvolve: A Web Portal for Integrative Oncogenomics).....	22
2.2.4	เว็บและโมบายอินเตอร์เฟซสำหรับวิเคราะห์หาความสัมพันธ์ระหว่างข้อมูลทางคลินิก และ จีโนม (The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical–genomic driver associations).....	27

2.2.5	เรดแคป (Research Electronic Data Capture: REDCap)	30
2.2.6	จีโนมิกส์คอมมอนส์ของเอ็นซีไอ ฐานข้อมูลและเครื่องมือสำหรับการแพทย์แม่นยำ (NCI Genomic Data Commons (GDC) as an engine for precision medicine)	31
2.2.7	เปรียบเทียบงานวิจัยที่เกี่ยวข้อง.....	35
บทที่ 3	วิธีการดำเนินงานวิจัย	38
3.1	ภาพรวมการทำงานของระบบ	38
3.2	แผนภาพแสดงขั้นตอนการทำงานในระบบ	39
3.3	สถาปัตยกรรมของระบบ.....	47
3.4	การประยุกต์ใช้แบบรูปเพื่อการออกแบบระบบ.....	48
3.5	การวิเคราะห์และออกแบบเซอวิสและรูปแบบข้อมูลที่รับส่งภายในระบบ.....	50
3.6	วิเคราะห์และออกแบบฐานข้อมูลของระบบ.....	54
3.6.1	ศึกษาโครงสร้างข้อมูลเชิงคลินิก.....	54
3.6.2	ศึกษาโครงสร้างข้อมูลภายในไฟล์จีเอฟ.....	58
3.6.3	ศึกษาโครงสร้างข้อมูลภายในไฟล์อีเอ็กซ์พี.....	59
3.6.1	ศึกษาโครงสร้างข้อมูลของโคเคบี (OncoKB).....	62
3.6.2	แนวทางออกแบบตารางข้อมูลให้รองรับข้อมูลจากฐานข้อมูลสาธารณะ.....	62
3.6.3	แนวทางการออกแบบการกำหนดสิทธิ์การเข้าถึงหน้าจการทำงานและเมนูต่าง ๆ ของระบบ	63
3.6.4	แนวทางการออกแบบการกำหนดสิทธิ์การเข้าถึงปลั๊กอินการวิเคราะห์.....	64
3.7	วิเคราะห์และออกแบบระบบเพื่อสนับสนุนการเพิ่มลดเซอวิสที่ใช้ในการวิเคราะห์ข้อมูลและแสดงผลข้อมูล	65
3.7.1	การเพิ่มเซอวิสที่ใช้ในการแสดงผลข้อมูล.....	65
3.7.2	การเพิ่มเซอวิสที่ใช้ในการวิเคราะห์ข้อมูล.....	69
บทที่ 4	ผลการวิจัยและอภิปรายผล	72
4.1	ภาพรวมของระบบ	72

4.2 การจัดเก็บข้อมูลประวัติผู้ป่วย (Patient Information Management)	72
4.3 การจัดเก็บข้อมูลทางคลินิก (Clinical Data Management)	74
4.4 การจัดเก็บข้อมูลชีววัตถุ (Biospecimen Management).....	74
4.5 การจัดเก็บและการวิเคราะห์ข้อมูลโอมิกส์ (-Omics Data Management and Analysis). 76	
4.5.1 การเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างชีววัตถุ 2 ชนิด (Variant Comparison).....	76
4.5.2 การวิเคราะห์ระดับการแสดงออกของยีนมะเร็ง (Integrative Analysis of Gene Expression Profiles).....	77
4.6 พัฒนาเครื่องมือที่ช่วยสร้างเทมเพลตปลั๊กอินส่วนแสดงผลบนเว็บของโคโคโดยอัตโนมัติ.....	80
4.7 แนวทางการพัฒนาต่อยอดปลั๊กอินการวิเคราะห์ใหม่ ๆ บนแพลตฟอร์ม	81
4.8 เครื่องมือที่ใช้ในการออกแบบและพัฒนาระบบ	85
บทที่ 5 สรุปผลการวิจัย.....	87
5.1 สรุปผลการวิจัย.....	87
5.2 แนวทางวิจัยในอนาคต.....	87
ภาคผนวก.....	88
ภาคผนวก ก. ตารางฐานข้อมูล (Database) ของซอฟต์แวร์แพลตฟอร์ม RUN-ONCO	89
ภาคผนวก ข. ขั้นตอนการเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างชีววัตถุ 2 ชนิด	110
ภาคผนวก ค. รายการบริการที่ทางแพลตฟอร์มเปิดให้ผู้พัฒนาสามารถเรียกใช้งานได้	113
บรรณานุกรม.....	121
ประวัติผู้เขียน.....	127

สารบัญตาราง

	หน้า
ตารางที่ 1: พิลด์หลักในไฟล์วีซีเอฟ.....	10
ตารางที่ 2: พิลด์ต่าง ๆ ในไฟล์อีเอ็กซ์พี.....	11
ตารางที่ 3: ประเภทและจำนวนชุดข้อมูลในแคณวีโวลฟ์	23
ตารางที่ 4: เปรียบเทียบงานวิจัยที่เกี่ยวข้อง	36
ตารางที่ 5: เปรียบเทียบพีเจอร์หลักในแต่ละแพลตฟอร์ม	37
ตารางที่ 6: เลเยอร์ต่าง ๆ ในองโคเซอร์วิส	49
ตารางที่ 7: รายการบริการที่ทางแพลตฟอร์มเปิดให้ผู้พัฒนาสามารถเรียกใช้งานได้.....	81



สารบัญรูปภาพ

หน้า

รูปที่ 1: ตัวอย่างสนิปส์ ดัดแปลงจาก [18] โดยมีการเปลี่ยนจากเบส T ในลำดับเบสอ้างอิงเป็น A ในชุดของลำดับเบสที่ถอดรหัสมาได้จากรูปตัวอย่าง	7
รูปที่ 2: ตัวอย่างอินเดล ดัดแปลงจาก [20] โดยมีการแทรกเพิ่มเข้ามาของลำดับเบส GTA และมีการขาดหายไปของลำดับเบส CA	7
รูปที่ 3: รูปแบบการแปรผันเชิงโครงสร้างของจีโนมตัวอย่าง (ขวา) เมื่อเทียบกับจีโนมอ้างอิง (ซ้าย) ดัดแปลงจาก [22]	8
รูปที่ 4: โครงสร้างไฟล์วีซีเอฟ ดัดแปลงจาก [23]	9
รูปที่ 5: ตัวอย่างไฟล์วีซีเอฟ ดัดแปลงจาก [24]	9
รูปที่ 6: ตัวอย่างไฟล์อีเอ็กซ์พี	11
รูปที่ 7: โครงสร้างการทำงานของอีทีแอล	12
รูปที่ 8: โครงสร้างสถาปัตยกรรม 3-เทียร์	13
รูปที่ 9: ตัวอย่างโครงสร้างสถาปัตยกรรมคอมโพเนนท์จาก [29]	14
รูปที่ 10: ตัวอย่างคลาสไดอะแกรมของแบบรูปพะซาด [30]	14
รูปที่ 11: ตัวอย่างแบบรูปดีเพนเดนซีอินเจกชัน ดัดแปลงจาก [31]	15
รูปที่ 12: โครงสร้างโมเดล-วิว-คอนโทรลเลอร์ ดัดแปลงจาก [32]	16
รูปที่ 13: ตัวอย่างการทำงานของวัตถุการเข้าถึงข้อมูลที่เกี่ยวข้องกับตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเซสชัน (ดัดแปลงจากรูปที่ 9.1 ของ [34])	17
รูปที่ 14: หน้าจอค้นหาข้อมูลจีโนมสำหรับโรคมะเร็งระบบศีรษะและลำคอ (A) ตัวเลือกชุดข้อมูลทางพันธุกรรมเพื่อใช้ในการค้นหา ได้แก่ ซีรีส์โออาร์แอล โอฟีซี-22 และซีรีส์เอช (B) ลิงก์สรุปผลชุดข้อมูลทางพันธุกรรม ได้แก่ ซีรีส์โออาร์แอล โอฟีซี-22 และซีรีส์เอช (C) เมื่อกดลิงก์ชุดข้อมูลซีรีส์โออาร์แอล ระบบจะแสดงข้อมูลสรุปผลการศึกษาและส่วนข้อมูลเพิ่มเติมอื่น ๆ (D) พิมพ์ชื่อยีนในช่องค้นหา เช่น EGFR เป็นต้น และ (E) แท็บข้อมูลดาวนโหลด ผู้ใช้สามารถนำออกชุดข้อมูลที่สนใจ เพื่อวิเคราะห์ข้อมูลแบบออฟไลน์ได้ (จากรูปที่ 1(A), 1(B), 1(C), 1(D), 1(E) ดัดแปลงจากรูปที่ 1 ของ [35])	18

รูปที่ 15: การกลายพันธุ์ของยีน *TP53* ภายในจีโนมแพคด้วยการเปรียบเทียบกับฐานข้อมูลโปรตีนพีแพม (A) ตำแหน่งของกรดอะมิโนกลายพันธุ์ที่พบบ่อยในยีนนั้น ๆ ในรูปแบบกราฟลอลลีพล็อต (lollipop) แกน x คือตำแหน่งของกรดอะมิโน และแกน y คือความถี่ของการกลายพันธุ์ของยีนที่กรดอะมิโน ตำแหน่งนั้น ๆ (B) ตารางรายละเอียดเกี่ยวกับยีนนั้น ๆ เช่น ยีนอยู่ในชุดข้อมูลใด พบยีนในฐานข้อมูล องค์กรใดหรือไม่ ประเภทของการกลายพันธุ์ เป็นต้น (ที่มา: รูปที่ 3(A), 3(B) ของ [35]) 19

รูปที่ 16: ภูมิภาคของยีนมะเร็ง 2 ชนิด (A) เปรียบเทียบอัตราการกลายพันธุ์ของยีน เมื่อเทียบกับ โรคมะเร็งต่างชนิดกัน (B) และ (C) แสดงกราฟวงกลมตามสัดส่วนการกลายพันธุ์ประเภทต่าง ๆ ในยีน *TP53* และ *VHL* ตามลำดับ (ที่มา: รูปที่ 1(A), 1(B), 1(C) ของ [41])..... 21

รูปที่ 17: กราฟผลการวิเคราะห์การเกิดเมธิลเลชันของยีน (A) เปรียบเทียบผลการเกิดเมธิลเลชันใน บริเวณต่าง ๆ ของยีน *SFRP1* และ *SFRP4* (B) และ (C) กราฟแสดงค่าประมาณการกระจายตัวของ ค่าความแตกต่างของการเกิดเมธิลเลชันเปรียบเทียบระหว่างตัวอย่างปกติและตัวอย่างมะเร็งโดย พิจารณายีน *SFRP1* และ *SFRP4* ตามลำดับ (ที่มา: รูปที่ 2(A), 2(B), 2(C) ของ [41]) 22

รูปที่ 18: ผลการวิเคราะห์ข้อมูลเบื้องต้นเชิงรูปภาพ (A) 1) รายการชุดข้อมูลที่จัดเตรียมสำหรับการ วิเคราะห์ 2) ช่องค้นหาเพื่อระบุกลุ่มยีนของมะเร็งในพารามิเตอร์ที่สนใจ 3) รายการชื่อยีนที่สนใจ และ (B) แผนภูมิความร้อนเปรียบเทียบการแสดงออกของไมโครอาร์เอ็นเอในเซลล์ปกติกับมะเร็งเม็ดเลือด ขาวมัลติเพิลมายอีโลมา (ที่มา: รูปที่ 1(A), 1(B) ของ [42])..... 24

รูปที่ 19: ผลการวิเคราะห์ความสัมพันธ์ระหว่างการแสดงออกของยีนและระยะรอดชีพของผู้ป่วยมะ เริงด้วยชุดข้อมูล GSE6477 (A) กราฟแสดงความน่าจะเป็นของระยะรอดชีพเทียบกับระยะเวลา (เดือน) และ (B) กราฟแสดงความน่าจะเป็นของระยะรอดชีพเทียบกับระยะเวลา (วัน) (ที่มา: รูปที่ 2(A), 2(B) ของ [42])..... 25

รูปที่ 20: ตัวอย่างผลทำนายการเกิดมะเร็งเม็ดเลือดขาวมัลติเพิลมายอีโลมาโดยหาความสัมพันธ์ระหว่างทรานสคริปชันแฟคเตอร์และยีนเป้าหมายด้วยวิธีวิเคราะห์ความสำคัญของกลุ่มยีน (ที่มา: รูปที่ 7 ของ [42]) 26

รูปที่ 21: กระบวนการวิเคราะห์ข้อมูลทางคลินิกและจีโนมด้วยวิธีแบบอีลาสติกเน็ต (ที่มา: รูปที่ 1 ของ [46]) 27

รูปที่ 22: หน้าจอเพื่อการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทางคลินิก และจีโนมด้วยวิธีการอีลาส 29

รูปที่ 23: ผลลัพธ์ของการค้นหาความสัมพันธ์ระหว่างยีนและข้อมูลทางคลินิก (A) แสดงตารางผลการคำนวณค่าพีแวลูของยีน TP53 ส่งผลสำคัญต่อข้อมูลทางคลินิกตัวแปรตัวใดบ้าง โดยเรียงลำดับความสำคัญ (B) แสดงตารางสรุปจำนวนกลุ่มข้อมูลที่นำมาวิเคราะห์หาความสัมพันธ์ระหว่างมะเร็งลำไส้ใหญ่และไส้ตรงกับข้อมูลทางคลินิก และ (C) แสดงตารางสรุปจำนวนกลุ่มข้อมูลนำมาวิเคราะห์หาความสัมพันธ์ระหว่างยีนต่าง ๆ ที่เกี่ยวข้องกับมะเร็งลำไส้ใหญ่และไส้ตรงกับระยะของโรคตามลักษณะทางคลินิก (ที่มา: รูปที่ 4 ของ [46]).....	30
รูปที่ 24: กระบวนการกำหนดโครงสร้างข้อมูลในเรดแคป (ที่มา: รูปที่ 1 ของ [47]).....	31
รูปที่ 25: กระบวนการขออนุมัตินำข้อมูลเข้าระบบจีดีซี (ที่มา: รูปที่ 1 ของ [48]).....	32
รูปที่ 26: ภาพรวมกระบวนการวิเคราะห์ข้อมูลชีวสารสนเทศด้วยเทคนิคไปป์ไลน์ (ที่มา: รูปที่ 2 ของ [48]).....	33
รูปที่ 27: ตัวอย่างการแสดงผลเชิงรูปภาพด้วยจีดีซี อองโคกริด (คัดเลือกมาบางส่วนจากรูปที่ 4 ของ [48]).....	34
รูปที่ 28: ภาพรวมกระบวนการทำงานของระบบ.....	38
รูปที่ 29: แผนภาพกิจกรรมของส่วนจัดการข้อมูลประวัติผู้ป่วย	39
รูปที่ 30: แผนภาพกิจกรรมของส่วนจัดการข้อมูลทางคลินิก.....	41
รูปที่ 31: แผนภาพกิจกรรมของส่วนจัดการข้อมูลชีววัตถุ	42
รูปที่ 32: แผนภาพกิจกรรมของส่วนนำเข้าไฟล์วีซีเอฟ.....	43
รูปที่ 33: แผนภาพกิจกรรมของส่วนนำเข้าไฟล์อีเอ็กซ์พี.....	44
รูปที่ 34: แผนภาพแสดงลำดับขั้นตอนการทำงานของแบตช์สกัดข้อมูลจากไฟล์วีซีเอฟ	45
รูปที่ 35: แผนภาพแสดงลำดับขั้นตอนการทำงานของแบตช์เพื่อตรวจสอบยีนที่เกี่ยวข้องกับมะเร็งที่สามารถรักษาได้จากข้อมูลองโคเคปี.....	46
รูปที่ 36: แผนภาพแสดงลำดับขั้นตอนการทำงานของแบตช์สกัดข้อมูลจากไฟล์อีเอ็กซ์พี	47
รูปที่ 37: ภาพรวมโครงสร้างซอฟต์แวร์แพลตฟอร์ม RUN-ONCO.....	48
รูปที่ 38: โครงสร้างซอฟต์แวร์แพลตฟอร์ม RUN-ONCO ภายใต้แนวคิดสถาปัตยกรรม 3-เทียร์.....	49
รูปที่ 39: ตัวอย่างการประยุกต์ใช้แบบรูปในแพลตฟอร์ม.....	50
รูปที่ 40: แผนภาพแสดงลำดับของการรับส่งข้อมูลภายในเว็บองโค.....	52

รูปที่ 41: แผนภาพแสดงลำดับของการรับส่งข้อมูลภายในองโคเซอร์วิส	53
รูปที่ 42: ตัวอย่างโครงสร้างข้อมูลทางคลินิก	54
รูปที่ 43: ตัวอย่างโครงสร้างข้อมูลเนื้อเยื่อ	54
รูปที่ 44: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลผู้ป่วย	55
รูปที่ 45: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลทางคลินิก	57
รูปที่ 46: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลรายละเอียดชีวิต	58
รูปที่ 47: ตัวอย่างไฟล์วีซีเอฟที่นำมาศึกษา	59
รูปที่ 48: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอ	60
รูปที่ 49: ตัวอย่างไฟล์ข้อมูลที่เกี่ยวข้องกับการแสดงออกของยีน	61
รูปที่ 50: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลระดับการแสดงออกของยีนจากเทคโนโลยีอาร์เอ็นเอซีค	61
รูปที่ 51: ตัวอย่างไฟล์ข้อมูลองโคเคบิที่นำมาศึกษา	62
รูปที่ 52: โครงสร้างตารางจัดเก็บข้อมูลที่รองรับการเปลี่ยนแปลงเวอร์ชันของข้อมูลสาธารณะ	63
รูปที่ 53: โครงสร้างตารางจัดเก็บข้อมูลการกำหนดสิทธิ์การเข้าถึงหน้าจอการทำงานและเมนูต่าง ๆ ของระบบ	64
รูปที่ 54: โครงสร้างตารางจัดเก็บข้อมูลการกำหนดสิทธิ์การเข้าถึงปลั๊กอินการวิเคราะห์	64
รูปที่ 55: กลไกการทำงานของไดนามิกคอนเทนต์แฮนด์เลอร์ (A) แผนภาพการเชื่อมต่อระหว่างคอมโพเนนต์หลักและคลัสเตอร์แกรมเมอร์คอมโพเนนต์ด้วยไดนามิกคอนเทนต์แฮนด์เลอร์ (B) โครงสร้างภาพการเรียกใช้คลัสเตอร์แกรมเมอร์ในคอมโพเนนต์หลักผ่านไดนามิกคอนเทนต์แฮนด์เลอร์	66
รูปที่ 56: แผนภาพแสดงลำดับของไดนามิกคอนเทนต์แฮนด์เลอร์	66
รูปที่ 57: การกำหนดค่าในไฟล์ต่าง ๆ ที่จำเป็นสำหรับการติดตั้งปลั๊กอินแบบไดนามิกในแพลตฟอร์ม RUN-ONCO	67
รูปที่ 58: ตัวอย่างผลการปลั๊กอิน clustergrammer.js เข้ามาในคอมโพเนนต์หลัก (parent component) ผ่านไดนามิกคอนเทนต์แฮนด์เลอร์	68
รูปที่ 59: แผนภาพแสดงลำดับการรับส่งข้อมูลของส่วนประกอบต่าง ๆ ที่เกี่ยวข้องกับการวิเคราะห์การแสดงออกของยีนมะเร็ง	70

รูปที่ 60: ตำแหน่งที่เพิ่มไฟล์ clustergrammer_api.py ในดาต้าอานาไลซิสเอพีไอ.....	71
รูปที่ 61: ตัวอย่างการเขียนโปรแกรมเพิ่มส่วนการวิเคราะห์ข้อมูลในไฟล์ clustergrammer_api.py	71
รูปที่ 62: ภาพรวมของระบบ	73
รูปที่ 63: ขั้นตอนการจัดเก็บข้อมูลประวัติผู้ป่วย (A) กรอกประวัติผู้ป่วยใหม่ (B) เพิ่มประวัติการติดตามการมีชีวิตของผู้ป่วย โดยการคลิกที่ปุ่ม "+ Add" และ (C) ข้อมูลประวัติการติดตามการมีชีวิตของผู้ป่วยจะอัปเดตในตารางข้อมูลให้อัตโนมติ.....	73
รูปที่ 64: ขั้นตอนการจัดเก็บข้อมูลการวินิจฉัยทางพยาธิวิทยา (A) เพิ่มข้อมูลการวินิจฉัยโรคในหน้าจอ (B) เพิ่มการวินิจฉัยทางพยาธิวิทยา โดยการคลิกที่ปุ่ม "+ Add" และ (C) ข้อมูลการวินิจฉัยทางพยาธิวิทยาอัปเดตในตารางข้อมูลให้อัตโนมติ.....	75
รูปที่ 65: แบบฟอร์มกรอกข้อมูลระยะของโรคมะเร็ง.....	75
รูปที่ 66: แบบฟอร์มกรอกข้อมูลชีววัตถุ.....	76
รูปที่ 67: ผลการเปรียบเทียบการแปรผันของดีเอ็นเอจากชีววัตถุ 2 ชนิดของผู้ป่วยรายเดียวกัน (A) เลือกการวิเคราะห์ผล Variant Comparison และกรอกรหัสผู้ป่วย HN620001 เพื่อค้นหาชีววัตถุของผู้ป่วยรายดังกล่าว แล้วเลือกประเภทการวิเคราะห์ลำดับเบสเป็น WES โดยสามารถเลือกชีววัตถุหลากหลายชนิดในคราวเดียวกัน ได้แก่ 0001 (Tissue), 0002 (Blood) หลังจากนั้น (B) เลือกไฟล์วีซีเอฟดังนี้ 0001_WES และ 0002_WES และกดปุ่ม "Update" ผลลัพธ์จะปรากฏบนหน้าจอ และ (C) คลิกที่ไอคอนของอองโคเคปป์ สืบหาข้อมูลรายละเอียดการกลายพันธุ์ของยีน <i>MTOR</i> และยาที่เกี่ยวข้อง	78
รูปที่ 68: ผลการวิเคราะห์ระดับการแสดงออกของยีนมะเร็ง (A) เลือกการวิเคราะห์ผล clustergrammer (B) เลือกไฟล์อีเอ็กซ์พี RUN-CA05.exp, RUN-CA07.exp และเลือกกลุ่มยีนมะเร็งที่สนใจและกดปุ่ม "Update" ผลลัพธ์การจัดกลุ่มจะปรากฏบนหน้าจอ (C) เลือกแหล่งข้อมูลจากข้อมูลสตริง โดยสามารถเลือกแหล่งข้อมูลหลาย ๆ แหล่งในคราวเดียวกัน แล้วคลิก "Update" และ (D) กราฟแสดงการมีปฏิสัมพันธ์ระหว่างคู่โปรตีนของยีนก็จะปรากฏบนหน้าจอ	79
รูปที่ 69: แผนภาพแสดงลำดับการทำงานของเครื่องมือที่ช่วยสร้างเทมเพลตปลั๊กอินบนเว็บของโคอ์ตโนมติ	80
รูปที่ 70: ตำแหน่งที่เพิ่มไฟล์ใหม่ของดาต้าอานาไลซิสเอพีไอ.....	82

รูปที่ 71: ตัวอย่างการเขียนโปรแกรมเพิ่มส่วนการวิเคราะห์ข้อมูลในไฟล์ tmb_api.py..... 82

รูปที่ 72: การเรียกใช้เครื่องมือเพื่อสร้างคอมโพเนนท์ใหม่และติดตั้งบนแพลตฟอร์มให้โดยอัตโนมัติ
..... 82

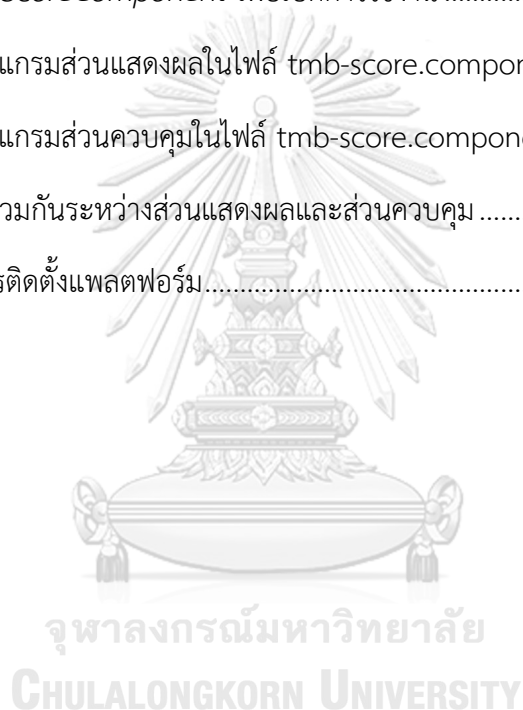
รูปที่ 73: ผลจากการใช้เครื่องมือ dynamic-component-jar-with-dependencies.jar โปรแกรม
ดังกล่าวจะสร้างคอมโพเนนท์ tmb-score และเพิ่มรายการ tmb-score เข้าไปในคลาส *Dynamic
ContentModule* โดยอัตโนมัติ (A) ไฟล์ต่าง ๆ จะอยู่ที่พาท `src/shared/components/dynamic-
content` ภายใต้โฟลเดอร์ tmb-score และ (B) ภายในคลาส *DynamicContentModule* จะถูก
แทรกด้วยคลาส *TmbScoreComponent* เพื่อเปิดการใช้งาน 83

รูปที่ 74: ตัวอย่างโปรแกรมส่วนแสดงผลในไฟล์ tmb-score.component.html 84

รูปที่ 75: ตัวอย่างโปรแกรมส่วนควบคุมในไฟล์ tmb-score.component.ts 84

รูปที่ 76: การทำงานร่วมกันระหว่างส่วนแสดงผลและส่วนควบคุม 85

รูปที่ 77: ภาพรวมการติดตั้งแพลตฟอร์ม..... 86



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โรคมะเร็งยังคงเป็นโรคที่คร่าชีวิตคนไทยเป็นอันดับต้น ๆ ในแต่ละปีมีผู้ป่วยรายใหม่กว่า 130,000 ราย และมีผู้เสียชีวิตมากกว่า 60,000 ราย นอกจากนี้ยังมีแนวโน้มที่จำนวนผู้ป่วยมะเร็งจะเพิ่มขึ้นจากสถิติสถาบันมะเร็งแห่งชาติ [1] ปัจจัยเสี่ยงที่ทำให้เกิดโรคคือ รูปแบบการใช้ชีวิต สภาพแวดล้อมภายนอกและกรรมพันธุ์ โรคมะเร็งถือเป็นปัญหาสาธารณสุขสำคัญของไทย เป็นอุปสรรคสำคัญในการพัฒนาประเทศ และส่งผลกระทบต่อสังคมและเศรษฐกิจโดยรวม

ความก้าวหน้าด้านการแพทย์นำไปสู่การรักษาโรคมะเร็งแบบแม่นยำและจำเพาะ (cancer precision medicine) โดยใช้วิธีการตรวจถึงระดับพันธุกรรม เพื่อช่วยในการวินิจฉัยติดตามโรค ซึ่งส่งผลให้สามารถเลือกวิธีการรักษาและยาที่เหมาะสมกับผู้ป่วยแต่ละคน ช่วยลดผลข้างเคียง ทุนเวลาและลดต้นทุนในการรักษารวมทั้งช่วยยกระดับคุณภาพชีวิตให้กับผู้ป่วย

การหาลำดับเบสดีเอ็นเอ (DNA sequencing) เป็นองค์ประกอบสำคัญในการศึกษาด้านจีโนมิกส์ (genomics) เพื่อช่วยหาบริเวณของยีนที่เกิดการกลายพันธุ์ ซึ่งอาจมีผลต่อการเกิดโรค รวมทั้งผลการรักษาเทคโนโลยีที่ได้รับความนิยมในการหาลำดับเบสดีเอ็นเอคือ เทคโนโลยีเอ็นจีเอส (next generation sequencing: NGS) ซึ่งสามารถหาลำดับเบสได้รวดเร็ว โดยในปัจจุบันมีต้นทุนค่าใช้จ่ายที่ต่ำลงต่อจำนวนข้อมูลลำดับเบสที่วิเคราะห์ได้เพิ่มขึ้น ด้วยวิธีการตัดสารพันธุกรรมหรือดีเอ็นเอ (DNA) จำนวนหลายพันล้านชิ้น เป็นเส้นสั้น ๆ ขนาด 25–1,000 เบส ขึ้นกับชนิดของเทคโนโลยี ผ่านกระบวนการเพิ่มจำนวน การอ่านลำดับเบส และนำมาเทียบเคียงกับจีโนมอ้างอิงเพื่อวิเคราะห์หาการแปรผันเชิงพันธุกรรมในลักษณะต่าง ๆ เป็นต้น

ในการวิเคราะห์ข้อมูลจีโนมมะเร็งจำเป็นต้องมีข้อมูลจีโนมปกติเปรียบเทียบกับข้อมูลจีโนมของเนื้อเยื่อมะเร็งจากผู้ป่วยรายเดียวกัน โดยทั้งสองจะถูกเปรียบเทียบกับข้อมูลที่มีอยู่ในฐานข้อมูลความหลากหลายทางพันธุกรรมในคนปกติ เช่น ฐานข้อมูลดีบีเอส (dbSNP) [2] ฐานข้อมูลคลินวาร (ClinVar) [3] ฐานข้อมูลคอสมิก (COSMIC) [4] และข้อมูลการแปรผันของรหัสพันธุกรรมจากโครงการ 1000 Genome Project [5] เป็นต้น ซึ่งการตรวจหาลำดับสารพันธุกรรมสำหรับการรักษาโรคมะเร็งเฉพาะจุด จะศึกษาหาความสัมพันธ์ระหว่างการแปรผันประเภทต่าง ๆ กับโรคมะเร็งโดยอาจครอบคลุมเฉพาะบริเวณที่เป็นยีน เนื่องจากมักมีข้อมูลทางคลินิกโดยผ่านการถอดรหัสพันธุกรรมเฉพาะส่วนที่เป็นเอ็กซอนทั้งหมดในจีโนม หรือที่เรียกว่าเอ็กโซม (exome) หรืออาจครอบคลุมรหัสพันธุกรรมทั้งจีโนม (genome) จีโนมของมะเร็งมีความซับซ้อน ขึ้นเนื้อมะเร็งประกอบไปด้วยทั้งเซลล์

ปกติและเซลล์มะเร็ง โดยบริเวณต่าง ๆ ของชิ้นเนื้อนี้อาจมีลักษณะการกลายพันธุ์ที่แตกต่างกัน ทำให้การวิเคราะห์ข้อมูลลำดับเบสของมะเร็งทำได้ยาก จำเป็นต้องใช้เทคนิคทางชีวสารสนเทศ และระบบคอมพิวเตอร์ประสิทธิภาพสูงมาช่วยในการวิเคราะห์

นอกจากการศึกษาในระดับจีโนมิกส์ทั้งจีโนมและเอ็กโซมแล้ว ข้อมูลโอมิกส์อื่น ๆ เช่น ข้อมูลในระดับทรานสคริปโตม (transcriptome) ที่วัดการแสดงออกของยีนทั้งจีโนมผ่านเทคโนโลยีอย่างอาร์เอ็นเอซีค (RNA-Seq) ข้อมูลในระดับโปรตีโอม (proteome) ที่วัดการแสดงออกของโปรตีน และข้อมูลเมตาโบโลม (metabolome) ที่วัดการแสดงออกของเมตาโบไลต์ผ่านเทคโนโลยี เช่น แมสสเปคโตรเมทรี (mass spectrometry) ข้อมูลการตอบสนองต่อยาของออร์แกนอยด์ต่าง ๆ ก็ถูกนำมาเชื่อมโยงประกอบการวินิจฉัยและรักษาโรคหรือออกแบบวิธีการรักษาที่จำเพาะ และเหมาะสมกับผู้ป่วยแต่ละรายได้มากขึ้น

ด้วยขนาดและความหลากหลายของข้อมูลโอมิกส์ ข้อมูลทางคลินิกและข้อมูลอื่น ๆ จากห้องปฏิบัติการ เช่น ผลเลือด การตอบสนองต่อยาของผู้ป่วย การตอบสนองต่อยาของออร์แกนอยด์ของผู้ป่วย เป็นต้น รวมทั้งความหลากหลายในวิธีการวิเคราะห์และการแสดงผลข้อมูลเหล่านี้ งานวิจัยซอฟต์แวร์แพลตฟอร์มเพื่อสนับสนุนการรักษาโรคมะเร็งแบบแม่นยำและจำเพาะ จึงมีความสำคัญและเป็นที่ต้องการเป็นอย่างมาก ทั้งในมิติของความสามารถในการบริหารจัดการข้อมูลที่หลากหลายและมีขนาดใหญ่ ความสามารถในการพัฒนาต่อยอดการวิเคราะห์และเชื่อมโยงข้อมูลในหลายมิติ เช่น การหาความสัมพันธ์ระหว่างการแปรผันของรหัสพันธุกรรมกับโรคมะเร็ง ความสัมพันธ์ของยีนกับยารักษาโรคหรือการเชื่อมโยงข้อมูลระหว่างออร์แกนอยด์ การแสดงออกของยีนและการตอบสนองของยา เป็นต้น รวมทั้งความสามารถในการพัฒนาต่อยอดการแสดงผลความสัมพันธ์ดังกล่าวในรูปแบบที่หลากหลาย เพื่อเป็นแนวทางในการรักษาโรค การเลือกใช้ยาและการปรับขนาดยาให้สอดคล้องกับผู้ป่วยแต่ละคนได้อย่างมีประสิทธิภาพ

1.2 วัตถุประสงค์ของงานวิจัย

วิจัยออกแบบและพัฒนาซอฟต์แวร์แพลตฟอร์ม โดยมีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อสนับสนุนการจัดการข้อมูลประวัติผู้ป่วย ข้อมูลทางคลินิก ข้อมูลชีววัตถุ ข้อมูลโอมิกส์ และข้อมูลจากฐานข้อมูลสาธารณะที่เกี่ยวข้อง
2. เพื่อสนับสนุนการบูรณาการข้อมูลที่สนับสนุนการวิเคราะห์ วินิจฉัยและออกแบบการรักษาโรคมะเร็งแบบแม่นยำจำเพาะ
3. เพื่อสนับสนุนให้นักพัฒนาสามารถเพิ่มโมดูลการวิเคราะห์ข้อมูล และการแสดงผลข้อมูลที่หลากหลายเข้าสู่ระบบได้โดยง่าย และเป็นระบบ

1.3 ขอบเขตงานวิจัย

1. ซอฟต์แวร์แพลตฟอร์มถูกพัฒนาเป็นเว็บที่ใช้งานภายในหน่วยงานหรือองค์กร (web-based standalone application) รองรับเว็บเบราว์เซอร์ได้แก่กูเกิลโครมเวอร์ชัน 60 และไฟร์ฟอกซ์เวอร์ชัน 54 ขึ้นไป
2. ผู้พัฒนาสามารถสร้างและติดตั้งคอมพิวเตอร์เน็ตการวิเคราะห์ใหม่บนแพลตฟอร์มอัตโนมัติผ่านทางคอมมานด์ไลน์ (command line)
3. ผู้ใช้สามารถจัดเก็บข้อมูลประวัติผู้ป่วย ข้อมูลทางคลินิก ข้อมูลตัวอย่างชีววัตถุของผู้ป่วยผ่านจียูไอได้ (Graphic User Interface: GUI)
4. การจัดการข้อมูลโอมิกส์ในเวอร์ชันนี้รองรับการจัดการข้อมูล 2 ชนิด ได้แก่ 1) ข้อมูลตำแหน่งที่เกิดการแปรผันของดีเอ็นเอที่เป็นผลจากการถอดรหัสและวิเคราะห์ข้อมูลเอ็กโซมหรือจีโนม ในรูปแบบไฟล์วีซีเอฟ (Variant Call Format: VCF) โดยสามารถเชื่อมโยงไปยังฐานข้อมูลจำเพาะของโรคมะเร็งฐานข้อมูลองโคเคบี (OncoKB) และ 2) ข้อมูลการแสดงออกของยีนจากเทคโนโลยีอาร์เอ็นเอซีคในรูปแบบไฟล์อีเอ็กซ์พี (Expression Matrix Format: EXP)
5. พัฒนาและติดตั้งตัวอย่างปลั๊กอินการวิเคราะห์ข้อมูลแบบบูรณาการไว้ในแพลตฟอร์ม ได้แก่
 - 1) การเปรียบเทียบการแปรผันของดีเอ็นเอจากตัวอย่างชีววัตถุ 2 ชนิดของผู้ป่วยรายเดียวกัน และ
 - 2) แสดงผลการวิเคราะห์ระดับการแสดงออกของยีนที่เกี่ยวข้องกับมะเร็งผ่านแผนที่ความร้อนด้วยคลัสเตอร์แกรมเมอร์ และแสดงการมีปฏิสัมพันธ์ระหว่างคู่โปรตีนจากฐานข้อมูลตรงด้วยไซโตสเคป (Cytoscape)

1.4 ขั้นตอนและวิธีการดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องและความเป็นไปได้ในการพัฒนาซอฟต์แวร์แพลตฟอร์ม
2. ศึกษาทฤษฎีที่เกี่ยวข้องที่ใช้สำหรับการวิเคราะห์ข้อมูลการรักษาโรคมะเร็งแบบแม่นยำ
3. ศึกษาตัวอย่างเทคนิคและเครื่องมือที่ใช้สำหรับการวิเคราะห์ข้อมูลการรักษาโรคมะเร็งแบบแม่นยำ
4. ออกแบบระบบการทำงานโดยครอบคลุมโครงสร้างระบบความสัมพันธ์ของข้อมูล และการสนับสนุนการนำข้อมูลไปวิเคราะห์ข้อมูลการรักษาโรคมะเร็งแบบแม่นยำ
5. สร้างต้นแบบของระบบ
6. ทดสอบการทำงานของระบบ
7. ตรวจสอบผลการทำงานของระบบ
8. ตีพิมพ์ผลงานวิจัย

9. สรุปผลและเรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ซอฟต์แวร์แพลตฟอร์มที่สนับสนุนการจัดการและบูรณาการข้อมูลประเภทต่าง ๆ เพื่อการวิเคราะห์ วินิจฉัย และออกแบบการรักษาโรคมะเร็งแบบแม่นยำจำเพาะได้
2. ได้ซอฟต์แวร์แพลตฟอร์มที่สนับสนุนการเพิ่มโมดูลการวิเคราะห์ข้อมูล และการแสดงผลข้อมูลที่หลากหลายเข้าสู่ระบบได้โดยง่าย และเป็นระบบ

1.6 บทความวิชาการที่ได้รับการตีพิมพ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตีพิมพ์เป็นผลงานวิชาการในหัวข้อเรื่องดังนี้

Neda Peyrone and Duangdao Wichadakul, “RUN-ONCO: A Highly Extensible Software Platform for Cancer Precision Medicine” ในรายงานการประชุมวิชาการนานาชาติ 2019 6th International Conference on Biomedical and Bioinformatics Engineering (ICBBE 2019), Shanghai, China, November 13-15, 2019.

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องกับการออกแบบและพัฒนางานวิจัยประกอบไปด้วย ลักษณะข้อมูลที่เกี่ยวข้อง และเทคโนโลยีหรือกระบวนการที่ได้มาซึ่งข้อมูลเหล่านี้ ประกอบด้วย การหาลำดับเบสดีเอ็นเอ การแปรผันทางพันธุกรรม ข้อมูลโอมิกส์ เช่น ลำดับเบสบนจีโนม ลำดับเบสบนเอ็กโซม สนิปส์ ไฟล์รูปแบบวีซีเอฟ และเทคโนโลยีที่ใช้ในการวัดการแสดงออกของยีน เป็นต้น กระบวนการสกัดข้อมูล การแปลงข้อมูล และการนำข้อมูลเข้าสู่โกดังข้อมูล สถาปัตยกรรม 3-เทียร์ สถาปัตยกรรมเชิงคอมโพเนนท์ และแบบรูปรูปแบบต่าง ๆ ที่ถูกนำมาใช้ในการออกแบบระบบ

2.1.1 การหาลำดับเบสดีเอ็นเอ (DNA sequencing)

เทคโนโลยีเอ็นจีเอส (Next Generation Sequencing: NGS) [6] เพิ่มความเร็วในการถอดรหัสพันธุกรรม ด้วยราคาที่ถูกลงอย่างมากประมาณ 1,000 ดอลลาร์สหรัฐต่อจีโนม ในปัจจุบันเอ็นจีเอสที่เป็นที่นิยมและใช้กันมากที่สุด ได้แก่ การหาลำดับเบสแบบ sequencing-by-synthesis (SBS) ของบริษัทอิลลูมินา (Illumina) โดยขั้นตอนการหาลำดับเบสดีเอ็นเอด้วยแพลตฟอร์มอิลลูมินา [7] มีดังนี้

การเตรียมดีเอ็นเอต้นแบบ (Template library preparation)

มีลำดับขั้นตอนการทำงานประกอบด้วย

1. ตัดสายดีเอ็นเอออกเป็นเส้นสั้น ๆ ขนาดเท่า ๆ กัน (DNA fragment) และเชื่อมปลายสายดีเอ็นเอที่ถูกตัดด้วยเอนไซม์ ที4 ดีเอ็นเอพอลิเมอเรส (T4 DNA polymerase) คลีนาว (klenow) และที4 โพลีนิวคลีโอไทด์ไคเนส (T4 polynucleotide kinase) [8]
2. ใส่เบสอะดีนีนที่ปลายสายดีเอ็นเอ เพื่อเชื่อมสายดีเอ็นเอกับอะแดปเตอร์ (adapter) โดยอะแดปเตอร์จะจับกับลำดับเบสคู่สมเรียกว่า ไพร์เมอร์ (universal primer)
3. เพิ่มจำนวนของสายดีเอ็นเอด้วยการทำพีซีอาร์ (polymerase chain reaction: PCR)
4. เลือกดีเอ็นเอที่เชื่อมกับอะแดปเตอร์ที่มีคุณภาพ ที่มีขนาดเท่า ๆ กัน เพื่อเข้าสู่ขั้นตอนการถอดรหัสลำดับเบสต่อไป

2.1.2 การถอดรหัสจีโนม (Whole Genome Sequencing: WGS)

ในการถอดรหัสพันธุกรรมจีโนม ดีเอ็นเอสายสั้นที่ได้จากการเตรียมดีเอ็นเอต้นแบบข้างต้นจะถูกนำมายึดติดกับแผ่นกระจกสไลด์ (flow cell) และป้อนเข้าเครื่องถอดรหัส ซึ่งผลลัพธ์ที่ได้คือ ชุดของลำดับเบสของดีเอ็นเอสายสั้นจำนวนมาก จะถูกนำไปเปรียบเทียบกับจีโนมอ้างอิงเพื่อวิเคราะห์หาการแปรผันที่เกิดขึ้นในบริเวณต่าง ๆ ของจีโนมต่อไป

2.1.3 การถอดรหัสเอ็กโซม (Whole Exome Sequencing : WES)

การถอดรหัสเอ็กโซมจะถอดรหัสพันธุกรรมเฉพาะบริเวณที่เป็นเอ็กซอนทั้งหมดในจีโนม ซึ่งเอ็กซอนเหล่านี้ [9-11] เป็นส่วนประกอบของยีนที่สามารถแปลรหัสต่อไปเป็นโปรตีน (protein coding gene) การหาลำดับเบสบนเอ็กโซมนี้ใช้วิธีการเดียวกับการถอดรหัสจีโนมข้างต้น โดยจะมีชุดสารเคมีจำเพาะที่ใช้การตัดรหัสพันธุกรรมเฉพาะบริเวณที่เป็นเอ็กซอน ซึ่งเป็นต้นแบบสร้างสายเอ็มอาร์เอ็นเอ (mRNA) สำหรับสังเคราะห์โปรตีน

การเปลี่ยนแปลงของลำดับเบสรวมทั้งการเปลี่ยนแปลงเชิงโครงสร้างทั้งในจีโนมและเอ็กโซมเป็นสาเหตุทำให้เกิดโรคทางพันธุกรรมหลายชนิด ดังนั้นการศึกษาการแปรผันเหล่านี้จึงมีความสำคัญมากและถูกนำมาใช้วิเคราะห์และวินิจฉัยความผิดปกติทางพันธุกรรม

เนื่องจากเอ็กโซมเป็นรหัสพันธุกรรมเฉพาะบริเวณที่เป็นยีนที่สามารถแปลรหัสต่อไปเป็นโปรตีน ทำให้สามารถเชื่อมโยงความแปรผันที่เกิดขึ้นกับฟังก์ชันของยีนได้ง่ายกว่า ด้วยราคาเวลาที่ใช้ในการถอดรหัสและจำนวนข้อมูลที่ผลิตออกมามีปริมาณน้อยกว่าการถอดรหัสจีโนมมาก งานวิจัยที่ผ่านมา ได้มีการถอดรหัสเอ็กโซมออกมามากมายและมีฐานข้อมูลเปิดสาธารณะอย่าง ExAC [12] ที่ผู้ใช้ทั่วไปสามารถเข้าไปสืบค้นข้อมูลได้

2.1.4 การแปรผันทางพันธุกรรม (Genetic variation)

การแปรผันทางพันธุกรรม [13, 14] สามารถแบ่งตามความถี่ของอัลลีลออกเป็น 2 ประเภท

1. อัลลีลที่มีความถี่ของการพบอย่างน้อย 1% ของประชากรทั้งหมด (common variation) หรือเรียกอีกชื่อหนึ่งว่าโพลิมอร์ฟิซึม (polymorphism)
2. อัลลีลที่มีความถี่ของการพบน้อยกว่า 1% ของประชากรทั้งหมด (rare variation)

2.1.4.1 ซิงเกิลนิวคลีโอไทด์โพลิมอร์ฟิซึม (Single Nucleotide Polymorphisms)

ซิงเกิลนิวคลีโอไทด์โพลิมอร์ฟิซึมหรือสเนปส์ (SNPs) (รูปที่ 1) เป็นการแปรผันทางพันธุกรรม (genetic variation) [15, 16] ที่พบบ่อยที่สุด โดยจะเกิดจากการแปรผันของลำดับเบสบนดีเอ็นเอเพียง 1 เบสในตำแหน่งต่าง ๆ ในจีโนมของมนุษย์ มีรายงานการค้นพบ

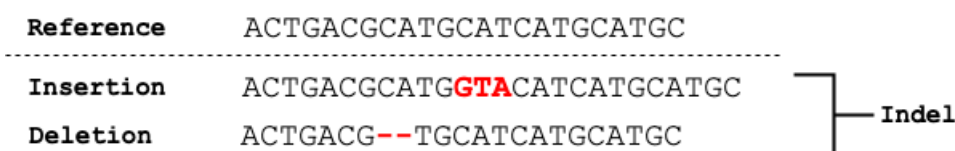
อย่างน้อย 11 ล้านสนิปส์ โดย 7 ล้านสนิปส์มีความถี่ของการพบมากกว่า 5% และสนิปส์ที่เหลือมีความถี่ระหว่าง 1-5% ของจำนวนประชากร [17] โดยปกติจะพบสนิปส์ในทุก ๆ 100-300 เบส และประมาณ 2 ใน 3 ของสนิปส์จะพบการแทนที่เบสไซโตซีน (cytosine: C) ด้วยไทมีน (thymine: T) ซึ่งสามารถนำไปประยุกต์ใช้ได้ทั้งการวินิจฉัย การรักษา การป้องกันโรค และการตอบสนองต่อยา



รูปที่ 1: ตัวอย่างสนิปส์ ตัดแปลงจาก [18] โดยมีการเปลี่ยนจากเบส T ในลำดับเบสอ้างอิงเป็น A ในชุดของลำดับเบสที่ถอดรหัสมาได้จากตัวอย่าง

2.1.4.2 อินเดล (Insertions and deletions: Indels)

อินเดล (indels) คือ การเพิ่มเข้ามาหรือการขาดหายไปของลำดับเบสที่อยู่ติดกันจำนวนน้อยกว่า 50 ตำแหน่ง [19] เมื่อเทียบกับจีโนมอ้างอิง รูปที่ 2 แสดงการเพิ่มเข้ามาของลำดับเบส GTA ในข้อมูลตัวอย่างโดยไม่พบในลำดับเบสอ้างอิง และลำดับเบส CA ที่หายไปจากข้อมูลตัวอย่าง

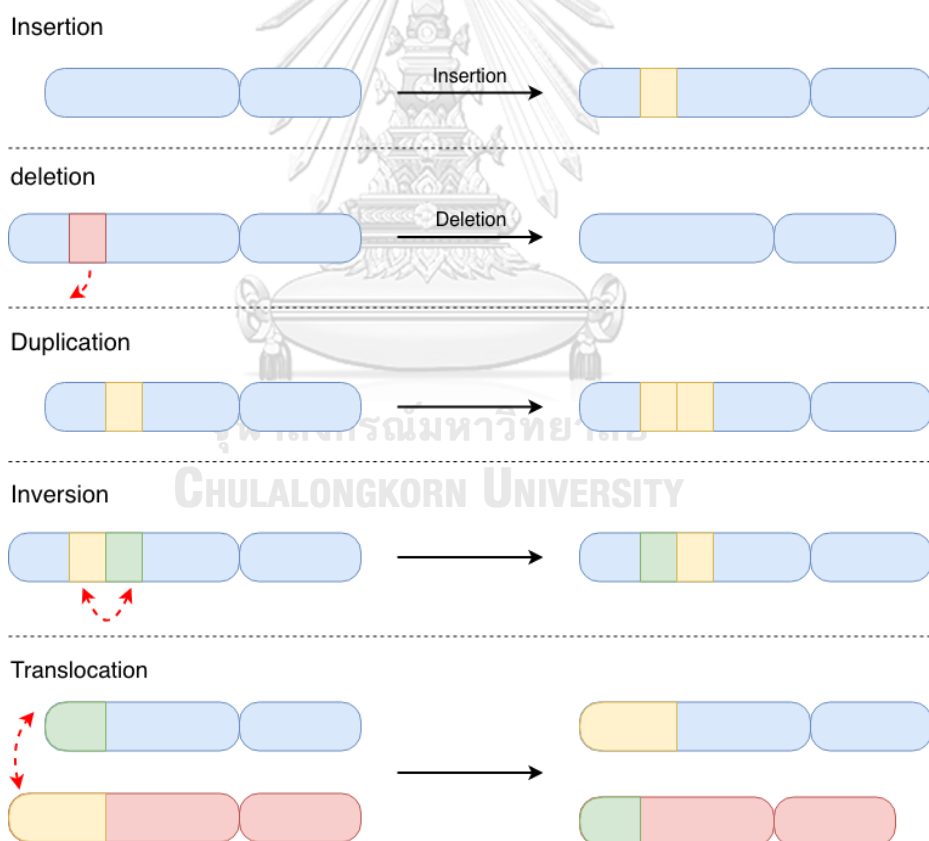


รูปที่ 2: ตัวอย่างอินเดล ตัดแปลงจาก [20] โดยมีการแทรกเพิ่มเข้ามาของลำดับเบส GTA และมีการขาดหายไปของลำดับเบส CA

2.1.4.3 การแปรผันเชิงโครงสร้าง (Structural variation)

การแปรผันเชิงโครงสร้าง (รูปที่ 3) เป็นการเปลี่ยนแปลงของลำดับเบสของจีโนมตัวอย่างมากกว่าเท่ากับ 50 ตำแหน่งที่อยู่ติดกัน [21] โดยมี 5 รูปแบบหลัก ๆ ดังต่อไปนี้

1. อินเสิร์ตชัน (insertion) คือ มีการเพิ่มขึ้นของลำดับเบสของจีโนมตัวอย่าง เมื่อเทียบกับจีโนมอ้างอิง
2. ดีลิตชัน (deletion) คือ การขาดหายไปของลำดับเบสของจีโนมตัวอย่าง เมื่อเทียบกับจีโนมอ้างอิง
3. แทนเดม ดูพลิเคชัน (tandem duplication) คือ ส่วนของดีเอ็นเอที่เกิดขึ้นซ้ำและแทรกเพิ่ม ตรงตำแหน่งเดิมของจีโนมตัวอย่าง เมื่อเทียบกับจีโนมอ้างอิง
4. อินเวอร์ชัน (inversion) คือ ส่วนของดีเอ็นเอที่มีการกลับด้านกันของลำดับเบสเมื่อเทียบกับจีโนมอ้างอิง
5. ทรานส์โลเคชัน (translocation) คือ ส่วนของดีเอ็นเอที่มีการย้ายตำแหน่งภายในโครโมโซม หรือจากโครโมโซมหนึ่งไปยังอีกโครโมโซมหนึ่งของจีโนมตัวอย่าง เมื่อเทียบกับจีโนมอ้างอิง

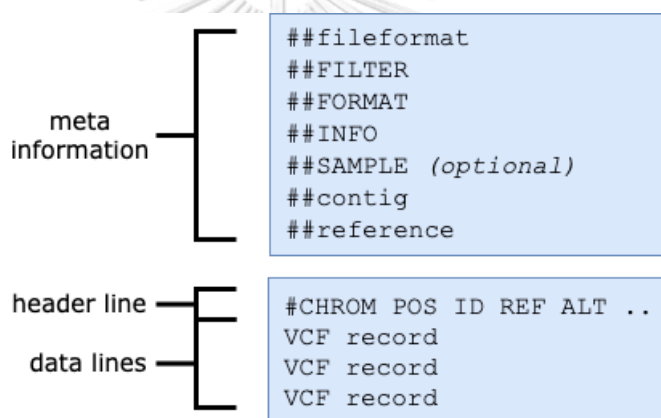


รูปที่ 3: รูปแบบการแปรผันเชิงโครงสร้างของจีโนมตัวอย่าง (ขวา) เมื่อเทียบกับจีโนมอ้างอิง (ซ้าย) ดัดแปลงจาก [22]

2.1.5 รูปแบบไฟล์วีซีเอฟ (Variant Call Format: VCF)

รูปแบบไฟล์วีซีเอฟเป็นมาตรฐานรูปแบบไฟล์ที่ใช้เก็บตำแหน่งที่เกิดการแปรผันของดีเอ็นเอ โดยโครงสร้างของไฟล์แสดงในรูปแบบที่ 4 และตัวอย่างข้อมูลภายในไฟล์ แสดงในรูปแบบที่ 5 ตามลำดับ ตารางที่ 1 แสดงคำอธิบายคอลัมน์ต่าง ๆ ภายในไฟล์ในส่วนของข้อมูล โดยไฟล์วีซีเอฟประกอบไปด้วย 3 ส่วนหลักดังนี้

1. ส่วนเมทาเดตาของข้อมูล (meta information) ทุกบรรทัดจะเริ่มต้นด้วย ##
2. ส่วนหัวของข้อมูล (header line) จะเริ่มต้นด้วย #CHROM ซึ่งเป็นคำอธิบายฟิลด์ต่าง ๆ ตามตารางที่ 1
3. ส่วนของข้อมูล (data lines) จะประกอบด้วยข้อมูลจีโนไทป์และตำแหน่งที่เกิดความแปรผันของเบสหรือลำดับเบส



รูปที่ 4: โครงสร้างไฟล์วีซีเอฟ ดัดแปลงจาก [23]

```

meta information {
  ##fileformat=VCFv4.1
  ##fileDate=20110413
  ##source=VCFtools
  ##reference=file:///refs/human_NCBI36.fasta
  ##contig=<ID=1,length=249250621,md5=1b22b98cdeb4a9304cb5d48026a85128,species="Homo Sapiens">
  ##contig=<ID=X,length=155270560,md5=7e0e2e580297b7764e31dbc80c2540dd,species="Homo Sapiens">
  ##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
  ##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
  ##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
  ##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
  ##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
  ##ALT=<ID=DEL,Description="Deletion">
  ##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
  ##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
}

header line {
  #CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE1 SAMPLE2
}

data lines {
  1 1 . ACG A,AT 40 PASS . GT:DP 1/1:13 2/2:29
  1 2 . C T,CT . PASS H2;AA=T GT 0|1 2/2
  1 5 rs12 A G 67 PASS . GT:DP 1|0:16 2/2:20
  X 100 . T <DEL> . PASS SVTYPE=DEL;END=299 GT:GQ:DP 1:12:.. 0/0:20:36
}

```

รูปที่ 5: ตัวอย่างไฟล์วีซีเอฟ ดัดแปลงจาก [24]

ตารางที่ 1: ฟิลด์หลักในไฟล์วีซีเอฟ

คอลัมน์	ชื่อคอลัมน์	ประเภทข้อมูล	คำอธิบาย
1	CHROM	String	ชื่อโครโมโซม
2	POS	Int	ตำแหน่งแรกของเบสที่เกิดการแปรผัน
3	ID	String	รหัสการแปรผันที่มีการรายงานมาก่อนในฐานะข้อมูลดีปซีนิปส์
4	REF	String	เบสหรือลำดับเบสบนจีโนมอ้างอิง
5	ALT	String	เบสหรือลำดับเบสของตัวอย่างที่นำมาเปรียบเทียบกับจีโนมอ้างอิงและเกิดการแปรผัน
6	QUAL	Int	คุณภาพของเบส
7	FILTER	String	สถานะของการกรองข้อมูล เช่น PASS, LowQual เป็นต้น
8	INFO	String	ส่วนที่ระบุข้อมูลเพิ่มเติม เช่น จำนวนอัลลีลทั้งหมดที่อยู่บนยีนตำแหน่งเดียวกัน (allele count: AC) ความถี่ของอัลลีล (allele frequency: AF) คือ การหาอัตราส่วนของอัลลีลชนิดนั้น ๆ ต่อจำนวนของอัลลีลทั้งหมดที่อยู่บนยีนตำแหน่งเดียวกัน เป็นต้น

2.1.6 ไฟล์ที่เก็บการแสดงออกของยีน หรือไฟล์อีเอ็กซ์พี (Expression Matrix Format: EXP)

รูปแบบไฟล์อีเอ็กซ์พีเป็นมาตรฐานรูปแบบไฟล์ที่ใช้จัดเก็บการแสดงออกของยีนจากเทคโนโลยีอาร์เอ็นเอซีค [25] โดยรูปที่ 6 แสดงโครงสร้างของไฟล์และตัวอย่างข้อมูลภายในไฟล์ และตารางที่ 2 แสดงคำอธิบายคอลัมน์ต่าง ๆ ภายในไฟล์ในส่วนของคุณข้อมูล โดยไฟล์อีเอ็กซ์พีประกอบไปด้วย 2 ส่วนหลักดังนี้

1. ส่วนหัวของข้อมูล (header line) จะเริ่มต้นด้วย Gene ID ซึ่งเป็นคำอธิบายฟิลด์ต่าง ๆ ตามตารางที่ 2
2. ส่วนของข้อมูล (data lines) จะประกอบด้วยข้อมูลรหัสของยีน ชื่อยีน ตำแหน่งของยีนและการแสดงออกของยีนที่ค่ามาตรฐานต่าง ๆ

header line	Gene ID	Gene Name	Reference	Strand	Start	End	Coverage	FPKM	TPM
data lines	ENSG00000237613	FAM138A	1	-	34554	36081	0.175923	0.028469	0.058367
	ENSG00000223972	DDX11L1	1	+	11869	14412	0.421951	0.068284	0.139993
	ENSG00000227232	WASH7P	1	-	14363	29806	63.059364	10.330416	21.179062
	ENSG00000243485	MIR1302-10	1	+	29554	31109	0.000000	0.000000	0.000000

รูปที่ 6: ตัวอย่างไฟล์อีเอ็กซ์พี

ตารางที่ 2: ฟิลด์ต่าง ๆ ในไฟล์อีเอ็กซ์พี

คอลัมน์	ชื่อคอลัมน์	ประเภทข้อมูล	คำอธิบาย
1	Gene ID	String	รหัสของยีน
2	Gene Name	String	ชื่อยีน
3	Reference	String	โครโมโซมอ้างอิง
4	Strand	String	แสดงทิศทางของเกลียวคู่ของสายดีเอ็นเอ เครื่องหมาย “+” หมายถึง สายดีเอ็นเอที่อ่าน จากซ้ายไปขวา และเครื่องหมาย “-” คือ สายดี เอ็นเอที่อ่านย้อนกลับจากขวามาซ้าย
5	Start	Int	ตำแหน่งเริ่มต้นของยีน ซึ่งมีค่าเริ่มจากดัชนีที่ 1
6	End	Int	ตำแหน่งสุดท้ายของยีน
7	Coverage	Float	ค่าเฉลี่ยของจำนวนรีดที่ครอบคลุมลำดับเบส หนึ่ง ๆ ในบริเวณที่เป็นยีน
8	FPKM	Float	ระดับการแสดงออกของยีนที่ปรับมาตรฐานด้วย ค่าเอฟพีเคเอ็ม (Fragments Per Kilobase Million: FPKM)
9	TPM	Float	ระดับการแสดงออกของยีนที่ปรับมาตรฐานด้วย ค่าทีพีเอ็ม (Transcripts Per Million: TPM)

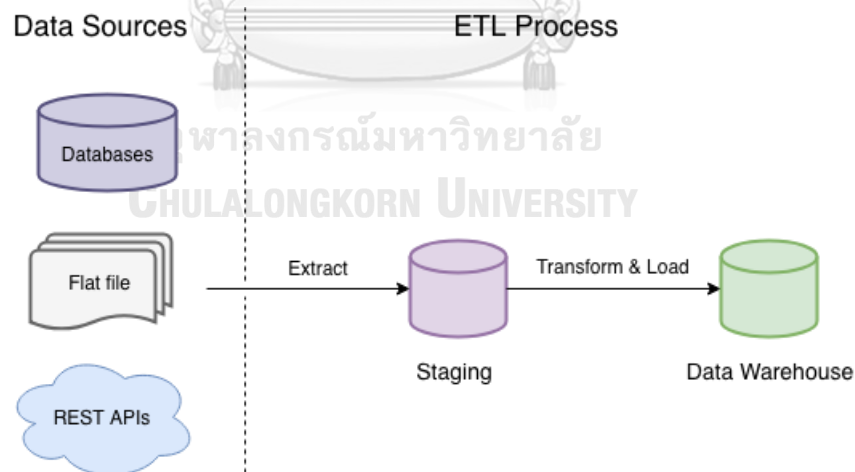
2.1.7 อีทีแอล (Extract-Transform-Load: ETL)

อีทีแอล [26] (รูปที่ 7) เป็นกระบวนการหลักในระบบโกดังข้อมูล (data warehouse) โดยเริ่มจากการสกัดข้อมูลจากแหล่งข้อมูลที่ต้องการ หลังจากนั้นทำการเปลี่ยนรูปข้อมูลแล้วทำการถ่ายโอนข้อมูลเข้าสู่โกดังข้อมูล ซึ่งแบ่งเป็น 3 ขั้นตอนดังนี้

1. การสกัดข้อมูล (extract) คือ กระบวนการดึงข้อมูลจากไฟล์ข้อมูลที่รวบรวมมา

2. การเปลี่ยนรูปข้อมูล (transform) คือ กระบวนการแปลงข้อมูลให้ตรงตามมาตรฐานของโกดังข้อมูลปลายทางที่จะนำเข้า และรวมไปถึงการทำความสะอาดข้อมูล (data cleansing) ซึ่งเป็นกระบวนการตรวจสอบและแก้ไขข้อมูลให้ถูกต้อง การคัดเลือกสดมภ์ (field selection) การกรองข้อมูลเฉพาะข้อมูลที่ต้องการ (filtering) การแปลงข้อมูลให้ตรงมาตรฐานของโกดังข้อมูลต้นทาง (encoding free form) เป็นต้น
3. การนำเข้าข้อมูล (load) คือ การนำเข้าข้อมูลที่เป็นผลลัพธ์จากข้อ 2 เข้าสู่โกดังข้อมูล

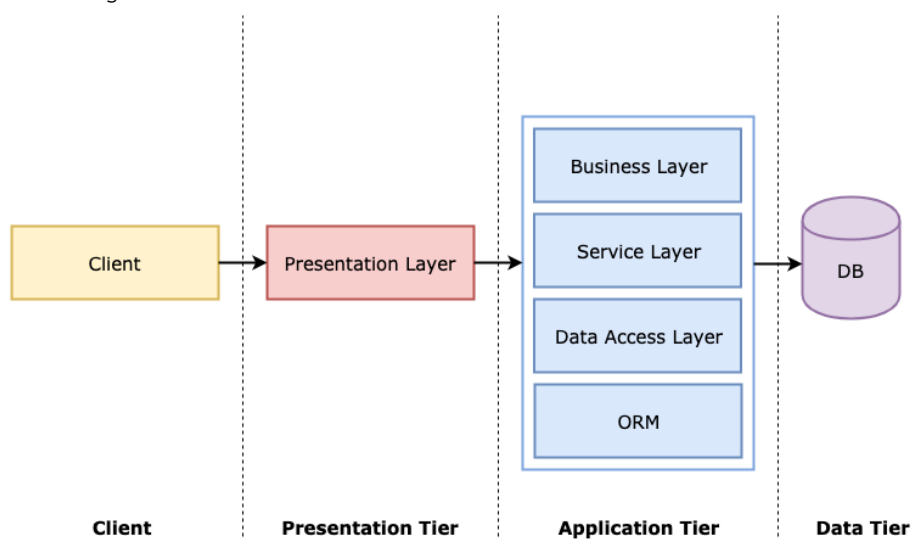
แหล่งข้อมูลที่จะนำเข้าสู่กระบวนการอีทีแอล ได้แก่ ข้อมูลจากฐานข้อมูลภายนอก ข้อมูลจากแฟ้มเดี่ยว (flat file) เช่น ไฟล์ซีเอสวี (comma-separated value: CSV) ไฟล์เอ็กเซล (excel file) ไฟล์อีเอ็กซ์ซีที และไฟล์วีซีเอฟ เป็นต้น รวมถึงการดึงข้อมูลจากฐานข้อมูลสาธารณะผ่าน REST (Representational State Transfer) ซึ่งเป็นสถาปัตยกรรมของระบบที่ออกแบบให้ทำงานบนระบบเครือข่ายในรูปแบบไคลเอนต์/เซิร์ฟเวอร์ โดยทุกการร้องขอของไคลเอนต์จะอ้างอิงด้วย URI (Uniform Resource Identifier) เช่น <http://onco-api/getAppVersion> เป็นต้น และจะต้องไม่มีข้อมูลของไคลเอนต์ถูกเก็บไว้ในเซิร์ฟเวอร์ (stateless) ข้อมูลจะถูกส่งไปเท่าที่จำเป็นเท่านั้น เป็นต้น



รูปที่ 7: โครงสร้างการทำงานของอีทีแอล

2.1.8 สถาปัตยกรรม 3-เทียร์ (3-Tier Architecture)

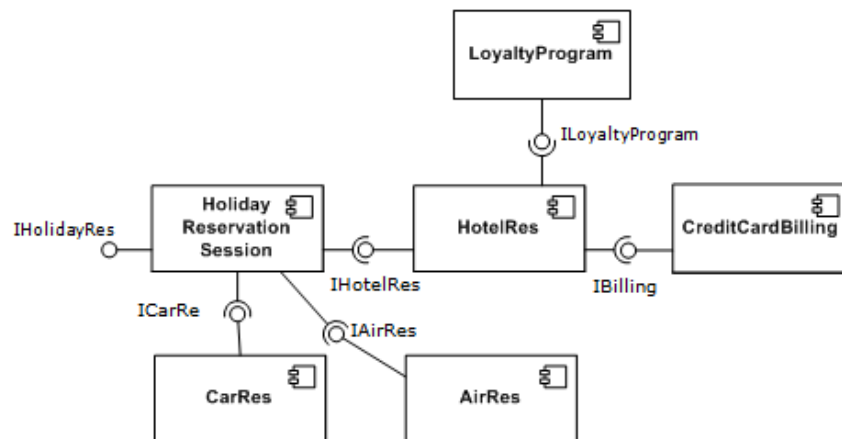
สถาปัตยกรรม 3-เทียร์ [27, 28] (รูปที่ 8) เป็นสถาปัตยกรรมแบบไคลเอนต์/เซิร์ฟเวอร์ที่มีการแยกส่วนการแสดงผล (presentation tier) ส่วนของการประมวลผล (application tier) และส่วนจัดการข้อมูล (data tier) ออกจากกัน ทำให้ระบบมีความยืดหยุ่นสูง สามารถควบคุมการเข้าถึงในแต่ละเทียร์ให้มีความปลอดภัยมากยิ่งขึ้น ลดความซับซ้อนของโค้ด ง่ายในการควบคุมเวอร์ชันของซอฟต์แวร์ และรองรับการทำบูรณาการพัฒนาซอฟต์แวร์อย่างต่อเนื่อง (continuous integration)



รูปที่ 8: โครงสร้างสถาปัตยกรรม 3-เทียร์

2.1.9 สถาปัตยกรรมเชิงคอมโพเนนต์ (Component-based Architecture: CBA)

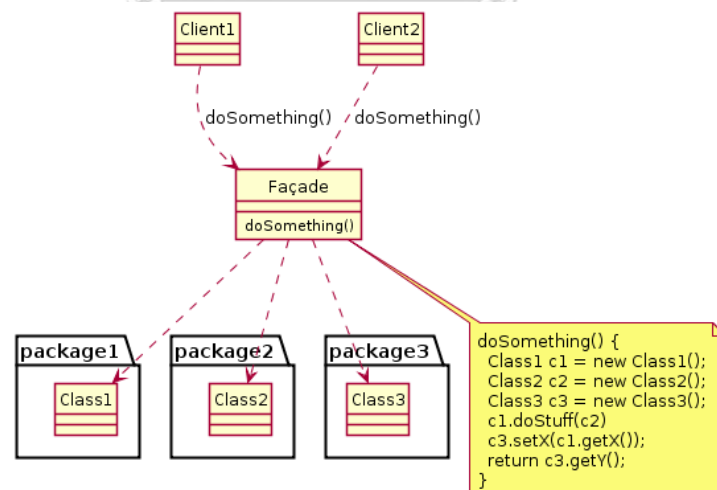
สถาปัตยกรรมเชิงคอมโพเนนต์ [29] (รูปที่ 9) เน้นความสามารถในการนำซอฟต์แวร์คอมโพเนนต์กลับมาใช้ซ้ำ (reuse) ซึ่งแต่ละคอมโพเนนต์มีชุดของฟังก์ชันที่ชัดเจน เช่น คอมโพเนนต์ในการนำเข้าสู่ข้อมูลโอมิกส์ คอมโพเนนต์ที่ใช้ในการแสดงผลในรูปแบบโครงข่ายของโปรตีน คอมโพเนนต์ที่ใช้ในการคำนวณค่าทีเอ็มบี (Tumor Mutational Burden: TMB) เป็นต้น โดยในการออกแบบและพัฒนาคอมโพเนนต์ต้องคำนึงถึงสิ่งต่อไปนี้ 1) ความเป็นอิสระต่อกัน 2) แต่ละคอมโพเนนต์สามารถถูกเรียกใช้งานผ่านอินเตอร์เฟซที่กำหนดไว้เป็นมาตรฐาน ซึ่งรายละเอียดของการทำงานภายในคอมโพเนนต์สามารถปรับเปลี่ยนได้ โดยไม่ส่งผลกระทบต่อซอฟต์แวร์ที่มาเรียกใช้งานคอมโพเนนต์นั้น ๆ และสามารถถูกแทนที่ได้ด้วยคอมโพเนนต์ที่มีอินเตอร์เฟซชุดเดียวกันและ 3) คอมโพเนนต์ควรมีฟังก์ชันทั่วไป เพื่อเพิ่มโอกาสในการนำไปใช้ซ้ำได้มาก ทั้งนี้ ส่วนประกอบต่าง ๆ ภายในสถาปัตยกรรม 3-เทียร์ ของซอฟต์แวร์แพลตฟอร์มที่นำเสนอ จะเน้นการออกแบบและพัฒนาให้อยู่ในรูปแบบสถาปัตยกรรมคอมโพเนนต์



รูปที่ 9: ตัวอย่างโครงสร้างสถาปัตยกรรมคอมโพเนนต์จาก [29]

2.1.10 แบบรูปพะซาด (Facade Pattern)

แบบรูปพะซาด [30] (รูปที่ 10) เป็นแบบรูปที่มีวัตถุประสงค์เพื่อลดความซับซ้อนของการเขียนโปรแกรม โดยยุบรวมโค้ดที่มีความซับซ้อนไว้ในเมธอดของคลาสพะซาด ซึ่งให้ง่ายต่อการเรียกใช้งาน พะซาดถูกนำมาใช้ในการออกแบบซอฟต์แวร์แพลตฟอร์มที่น่าเสนอ ก็เพื่อลดความซับซ้อนในบางชุดฟังก์ชันการทำงาน โดยสนับสนุนความสามารถในการพัฒนาต่ออย่างรวดเร็วในอนาคตได้โดยง่าย

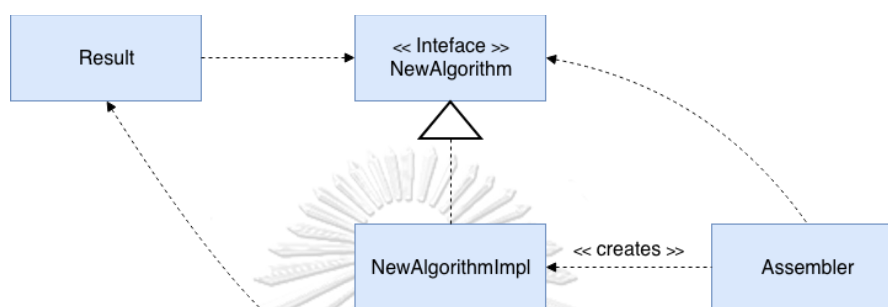


รูปที่ 10: ตัวอย่างคลาสไดอะแกรมของแบบรูปพะซาด [30]

2.1.11 แบบรูปดีเพนเดนซีอินเจกชัน (Dependency Injection: DI Pattern)

แบบรูปดีเพนเดนซีอินเจกชัน [31] (รูปที่ 11) เป็นวิธีการฉีดออบเจกต์เข้ามาผ่านทางคอนสตรัคเตอร์ของออบเจกต์ที่ให้บริการ โดยออบเจกต์ที่ให้บริการนี้จะถูกเรียกใช้งานตลอด

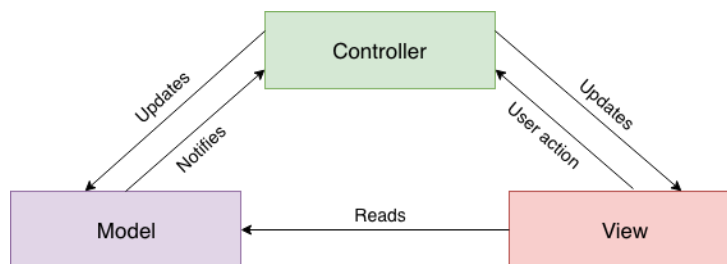
อายุการทำงานของออบเจ็กต์ไคลเอนต์ การใช้ดีเพนเดนซีอินเจกชันจะลดความขึ้นต่อกันระหว่างออบเจ็กต์ไคลเอนต์และออบเจ็กต์ที่ให้บริการที่ถูกฉีดเข้ามา โดยสามารถปรับเปลี่ยนพฤติกรรมให้บริการได้ในช่วงรันไทม์ ดีเพนเดนซีอินเจกชันถูกนำมาใช้ในการออกแบบซอฟต์แวร์แพลตฟอร์มที่นำเสนอเพื่อสนับสนุนการเพิ่มโมดูลการแสดงผลในรูปแบบต่าง ๆ เข้าสู่ระบบได้โดยง่าย



รูปที่ 11: ตัวอย่างแบบรูปดีเพนเดนซีอินเจกชัน ดัดแปลงจาก [31]

2.1.12 โมเดล-วิว-คอนโทรลเลอร์ (Model-View-Controller: MVC)

โมเดล-วิว-คอนโทรลเลอร์ (รูปที่ 12) เป็นสถาปัตยกรรมรูปแบบหนึ่งที่แยกการทำงานออกเป็น 3 ส่วนหลักคือ 1) แบบจำลองข้อมูล (data model) เป็นส่วนจัดการข้อมูลและตรรกะที่ใช้ในการคำนวณกระบวนการทางธุรกิจ (business logic) เช่น การคำนวณอายุจากวันเกิดที่ผู้ใช้ระบุไว้ในส่วนแสดงผล เป็นต้น 2) วิว (view) ส่วนที่จัดการเรื่องการแสดงผล และ 3) ส่วนควบคุม (controller) ทำหน้าที่ควบคุมการรับส่งข้อมูลระหว่างแบบจำลองข้อมูลกับวิว วิทยานิพนธ์นี้นำโมเดล-วิว-คอนโทรลเลอร์มาใช้ในการออกแบบส่วนติดต่อประสานงานกับผู้ใช้ (user interface) โดยเทมเพลต (template) ที่ใช้ในการแสดงผลบนเว็บเบราว์เซอร์ เขียนด้วยภาษาเอชทีเอ็มแอล (HTML) เมื่อผู้ใช้กระทำบางอย่างบนส่วนแสดงผล ส่วนควบคุมจะรับคำสั่งจากส่วนแสดงผลและส่งคำสั่งขอไปยังเอพีไอ และนำส่วนตอบกลับที่ได้รับจากเอพีไอแปลงให้อยู่ในรูปแบบของแบบจำลองข้อมูล และแสดงผลบนเทมเพลตต่อไป



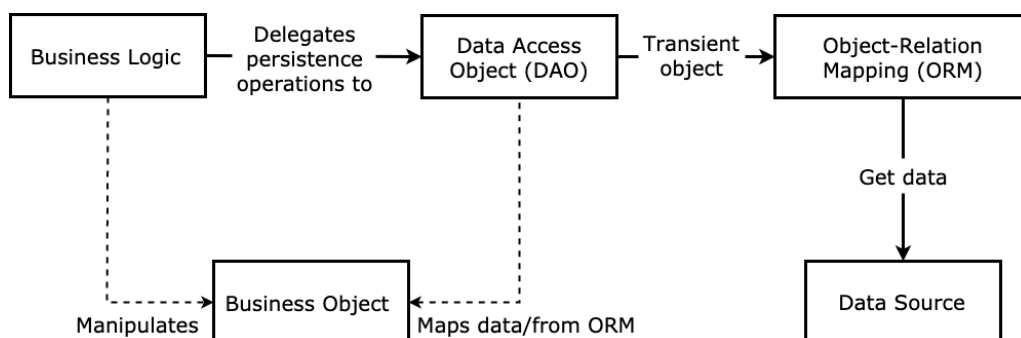
รูปที่ 12: โครงสร้างโมเดล-วิว-คอนโทรลเลอร์ ดัดแปลงจาก [32]

2.1.13 ตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ (Object-Relational Mapping: ORM)

ตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ เป็นตัวกลางที่เชื่อมความสัมพันธ์ระหว่างอ็อบเจกต์และตารางข้อมูล โดยคลาสในแบบจำลองวัตถุจะเทียบเท่ากับตารางข้อมูลในฐานข้อมูลเชิงสัมพันธ์ และแอตทริบิวต์ในคลาสใช้อ้างถึงฟิลด์ต่าง ๆ ในตารางข้อมูล ประโยชน์ของการใช้โมเดลเชิงวัตถุและเชิงสัมพันธ์คือ 1) ควบคุมเวอร์ชันของตารางข้อมูลในฐานข้อมูล โดยระบบสามารถตรวจสอบได้ทันทีว่าฟิลด์ใดบ้างที่มีการเปลี่ยนแปลง ตารางข้อมูลใดบ้างหายไปจากฐานข้อมูล เป็นต้น 2) ไม่ยึดติดกับฐานข้อมูลที่จำเพาะ 3) สนับสนุนการออกแบบด้วยแนวคิดโมเดลเฟิร์ส (model first) มองความสัมพันธ์ของข้อมูลเชิงวัตถุ ทำให้การออกแบบรวดเร็วและมีประสิทธิภาพ โดยระบบจะสร้างตารางข้อมูลในฐานข้อมูลอัตโนมัติจากคลาสที่สร้างไว้ และ 4) เขียนโค้ดได้สั้นลง ทำให้การบำรุงรักษาง่ายขึ้น

2.1.14 วัตถุการเข้าถึงข้อมูล (Data Access Object: DAO)

วัตถุการเข้าถึงข้อมูล [33, 34] (รูปที่ 13) คือ การสร้างคลาสสำหรับเก็บชุดคำสั่งหรือเมธอดที่ติดต่อกับฐานข้อมูล โดยมีวัตถุประสงค์เพื่อแยกชั้นข้อมูลออกจากชั้นกระบวนการทางธุรกิจ (business layer) เป็นการลดความซ้ำซ้อนของโค้ดคำสั่งการเชื่อมต่อฐานข้อมูล และสนับสนุนการกลับมาใช้ซ้ำ วัตถุการเข้าถึงข้อมูลถูกนำมาออกแบบและพัฒนาในส่วนของโคเชอร์วิส โดยเชื่อมต่อกับตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ ซึ่งทำหน้าที่เป็นตัวกลางติดต่อกับฐานข้อมูล



รูปที่ 13: ตัวอย่างการทำงานของวัตถุการเข้าถึงข้อมูลที่เชื่อมต่อกับตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ (ดัดแปลงจากรูปที่ 9.1 ของ [34])

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 จีโนมิกส์พอร์ทัลสำหรับโรคมะเร็งระบบศีรษะและลำคอ (GENIPAC: A Genomic Information Portal for Head and Neck Cancer Cell Systems)

จีโนมิกส์พอร์ทัลสำหรับโรคมะเร็งระบบศีรษะและลำคอหรือจีเนแพค [35] ถูกพัฒนาให้เป็นแหล่งรวบรวมข้อมูล เพื่อใช้ในการสำรวจข้อมูล แสดงผลด้วยภาพและวิเคราะห์ข้อมูลจีโนมและข้อมูลอื่น ๆ ที่เกี่ยวข้อง เช่น ข้อมูลการแสดงออกของยีน เป็นต้น สำหรับโรคมะเร็งระบบศีรษะและลำคอโดยเฉพาะ ระบบเก็บความสัมพันธ์ระหว่างยีนและความไวต่อยารักษาโรคมะเร็ง โดยมีการเชื่อมต่อกับฐานข้อมูลองโคเคบี (OncoKB) [36] งานวิจัยนี้ได้รวบรวมข้อมูล 44 เซลล์มะเร็งที่เกี่ยวข้องกับโรคมะเร็งระบบศีรษะและลำคอจาก 3 ชุดข้อมูล ได้แก่ ข้อมูลซีรีส์โออาร์แอล (ORL series) เป็นมะเร็งช่องปากชนิดสความัส (squamous cell carcinoma) 16 สายเซลล์ (cell lines) [37] ข้อมูลชุดโอพีซี-22 (OPC-22) เป็นข้อมูลจากงานวิจัยมะเร็งศีรษะและลำคอ 22 สายเซลล์ [38] และข้อมูลซีรีส์เอช (H series) เป็นสายเซลล์มะเร็งช่องปากชนิดสความัส จากมหาวิทยาลัยบริสตอล ประเทศอังกฤษ [39] ซึ่งได้จากการวิเคราะห์ข้อมูลอาร์เอ็นเอผ่านการถอดรหัสอาร์เอ็นเอ (RNA sequencing) เป็นหลัก

งานวิจัยนี้ได้พัฒนาจีเนแพคบนแพลตฟอร์มซีไบโอพอร์ทัล (cBioPortal) [40] พิเจอร์หลัก ๆ ที่ให้บริการ เช่น หน้าค้นหาแสดงในรูปที่ 14(A) ผู้ใช้สามารถเลือกชุดข้อมูลทางพันธุกรรมที่ต้องการค้นหา ได้แก่ ซีรีส์โออาร์แอล โอพีซี-22 และซีรีส์เอช ในส่วนรูปที่ 14(B) และรูปที่ 14(C) สามารถลิงก์ไปยังชุดข้อมูลที่สนใจ เพื่อแสดงรายละเอียดสรุปผลการศึกษา (study summary) และข้อมูลเพิ่มเติมจากชุดข้อมูลที่ศึกษาตามลำดับ สำหรับรูปที่ 14(D) ผู้ใช้สามารถระบุเงื่อนไขในการค้นหาโดยใช้รหัสของฮิวโก (human genome organisation: HUGO) หรือชื่อยีน และรูปที่ 14(E) ผู้ใช้สามารถดาวน์โหลดข้อมูล เพื่อวิเคราะห์ข้อมูลแบบออฟไลน์ได้

Figure 14: OncoPrint Query Interface and Results

Section A: Select Studies

- 1 studies selected (16 samples) Deselect all View summary
- Head and Neck (3) Select all listed studies (3)
- Head and Neck Squamous Cell Carcinoma
 - H Series
 - OPC-22
 - ORL Series
- View study summary (7 samples, 22 samples, 16 samples)

Section B: Select Genomic Profiles

- Mutations
- Putative copy-number alterations from GISTIC
- mRNA expression
- Enter a z-score threshold \pm 2.0

Section C: Select Patient/Case Set

- All Tumors (16)

Section D: Enter Gene Set

- User-defined List
- EGFR
- All gene symbols are valid.

Section E: Submit Query

Section C: Study Summary

Selected: 16 samples / 16 patients query genes - click to expand Query Select cases by IDs Add Chart

Mutated Genes (16 profiled samples)				CNA Genes (16 profiled samples)				
Gene	# Mut	#	Freq	Gene	Cytoband	CNA	#	Freq
TP53	17	16	100%	CCND1	11q13.3	AMP	12	75%
DHRX	35	16	100%	FGF3	11q13.3	AMP	12	75%
AMOTL2	19	16	100%	FGF4	11q13.3	AMP	12	75%
SLC30A9	31	16	100%	FGF19	11q13.3	AMP	12	75%
CAP1	48	16	100%	EGFR	7p11.2	AMP	12	75%
ARAP2	20	16	100%	FANCC	9q22.32	AMP	11	68.75%
LRRC8C	20	16	100%	PTCH1	9q22.32	AMP	11	68.75%
MAF1	16	16	100%	BCL6	3q27.3	AMP	10	62.50%
UBE2J1	16	16	100%	EIF4A2	3q27.3	AMP	10	62.50%
NSUN4	22	16	100%	EPHB3	3q27.1	AMP	10	62.50%
FN1	47	16	100%	ETV5	3q27.2	AMP	10	62.50%
CDKN1A	40	16	100%					

Section D: # of Samples Per Patient

16

Section E: With Mutation Data

16

Section F: With CNA Data

16

Section G: Mutation Count

Bar chart showing Mutation Count distribution: <=4060, 4130, 4270, 4410, 4550, >4550.

รูปที่ 14: หน้าจอค้นหาข้อมูลจีโนมสำหรับโรคมะเร็งระบบศีรษะและลำคอ (A) ตัวเลือกชุดข้อมูลทางพันธุกรรมเพื่อใช้ในการค้นหา ได้แก่ ซีรีส์โออาร์แอล โอพีซี-22 และซีรีส์เอช (B) ลิงก์สรุปผลชุดข้อมูลทางพันธุกรรม ได้แก่ ซีรีส์โออาร์แอล โอพีซี-22 และซีรีส์เอช (C) เมื่อคลิกชุดข้อมูลซีรีส์โออาร์แอล ระบบจะแสดงข้อมูลสรุปผลการศึกษาและส่วนข้อมูลเพิ่มเติมอื่น ๆ (D) พิมพ์ชื่อยีนในช่องค้นหา เช่น *EGFR* เป็นต้น และ (E) แท็บข้อมูลดาวน์โหลด ผู้ใช้สามารถนำออกชุดข้อมูลที่สนใจ เพื่อวิเคราะห์ข้อมูลแบบออฟไลน์ได้ (จากรูปที่ 1(A), 1(B), 1(C), 1(D), 1(E) ดัดแปลงจากรูปที่ 1 ของ [35])

รูปที่ 15 แสดงตัวอย่างข้อมูลการกลายพันธุ์ของยีนเปรียบเทียบกับฐานข้อมูลโปรตีนพีแฟม (Pfam) โดยมีการแสดงตำแหน่งของกรดอะมิโนที่เกิดการกลายพันธุ์ที่พบบ่อย (hotspot) ในสายเซลล์ นอกจากนี้ยังมีการแสดงข้อมูลอื่น ๆ เกี่ยวกับยีนนั้น ๆ เช่น การกลายพันธุ์ของยีนที่มีผลกับข้อมูลทางคลินิก ความถี่ของยีนที่พบในฐานข้อมูลคอสมิก (COSMIC) [4] เป็นต้น



รูปที่ 15: การกลายพันธุ์ของยีน *TP53* ภายในจีโนมแพคด้วยการเปรียบเทียบกับฐานข้อมูลโปรตีนพีแฟม (A) ตำแหน่งของกรดอะมิโนที่กลายพันธุ์ที่พบบ่อยในยีนนั้น ๆ ในรูปแบบกราฟลอลลิปล็อต (lollipop) แกน x คือตำแหน่งของกรดอะมิโน และแกน y คือความถี่ของการกลายพันธุ์ของยีนที่กรดอะมิโนตำแหน่งนั้น ๆ (B) ตารางรายละเอียดเกี่ยวกับยีนนั้น ๆ เช่น ยีนอยู่ในชุดข้อมูลใด พบยีนในฐานข้อมูลองโคเคปหรือไม่ว่าประเภทของการกลายพันธุ์ เป็นต้น (ที่มา: รูปที่ 3(A), 3(B) ของ [35])

นอกจากนี้ยังมีการแสดงอัตราส่วนกลุ่มเซลล์มะเร็งหรือสายเซลล์จากเซลล์มะเร็งในเชิงรูปภาพ โดยคัดเลือกยีนที่มีของการกลายพันธุ์สำคัญที่สุด 5 ลำดับแรกจากข้อมูลแผนที่จีโนมมะเร็ง (The Cancer Genome Atlas: TCGA) งานวิจัยนี้ยังได้นำระดับการแสดงออกของยีนในระดับเอ็มอาร์เอ็นเอมาร่วมวิเคราะห์กับการแปรผันของจำนวนชุดดีเอ็นเอ หรือซีเอ็นวี (copy

number variation: CNV) เพื่อเปรียบเทียบข้อมูลระดับการแสดงออกยีนระหว่างข้อมูลแผนที่จีโนมมะเร็งกับจีโนมปกติ เช่น คำนวณอัตราส่วนของการเปลี่ยนแปลงของยีนเนื่องจากเป็นมะเร็ง เพื่อหาว่ายีนใดมีการแสดงออกมากกว่าค่ามาตรฐาน (overexpression) หรือไม่ เป็นต้น

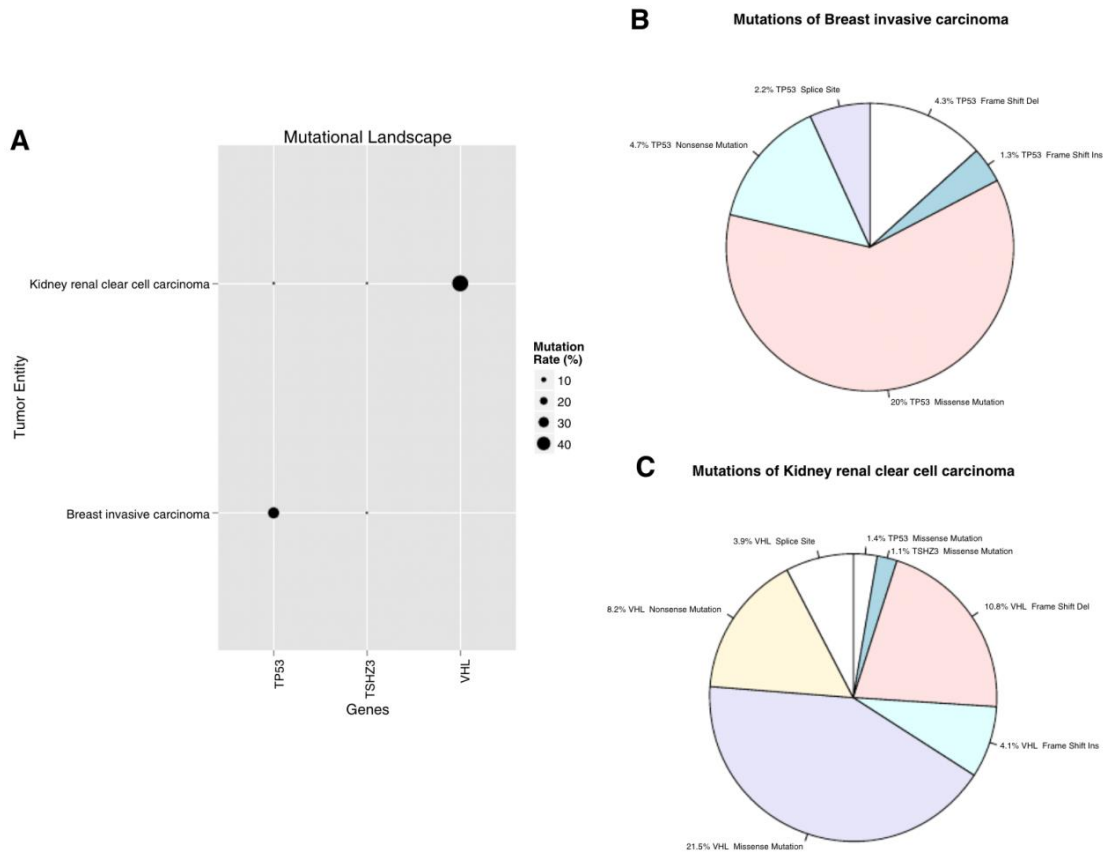
2.2.2 เว็บไซต์จีโนมแพลตฟอร์มออนไลน์สำหรับการวิเคราะห์โรคมะเร็งแบบบูรณาการในระดับ

โมเลกุล (Web-TCGA: an online platform for integrated analysis of molecular cancer data sets)

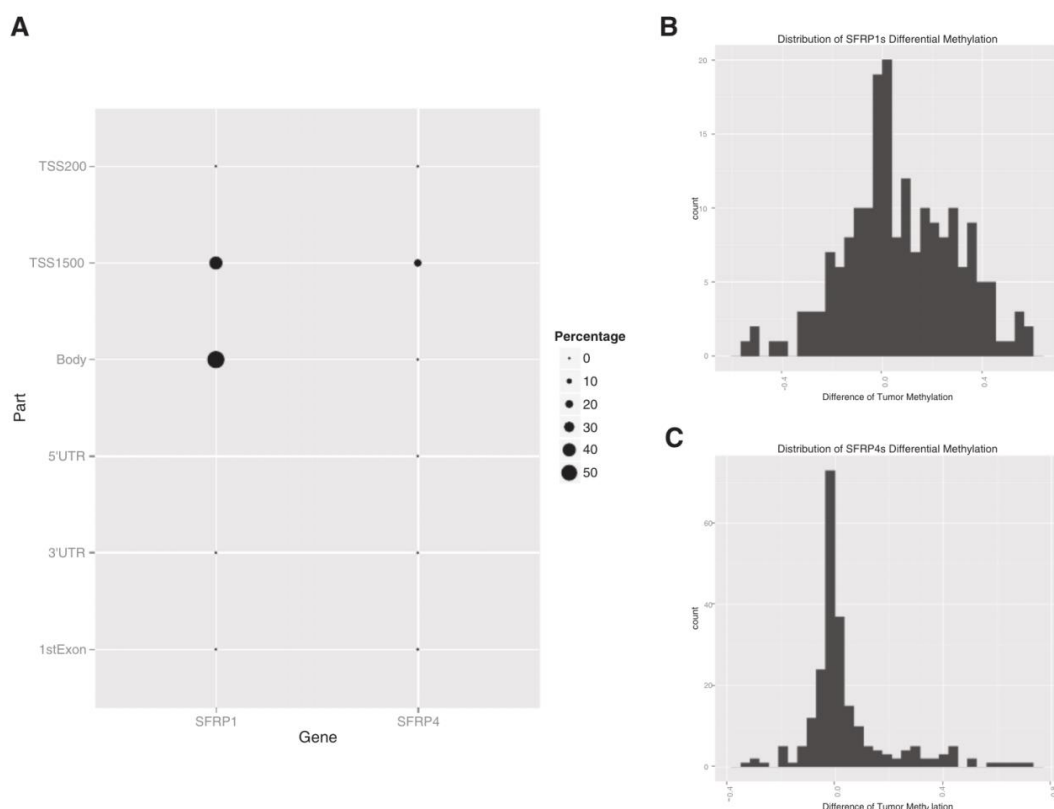
งานวิจัยนี้นำเสนอเว็บไซต์แพลตฟอร์มสำหรับวิเคราะห์โรคมะเร็งจากชุดข้อมูลที่จีโนมในระดับโมเลกุล [41] โดยถูกพัฒนาให้วิเคราะห์ข้อมูลยีนและเนื้องอกเป็นหลัก มีการแสดงผลการวิเคราะห์ข้อมูลในรูปแบบกราฟและตารางข้อมูลโดยมีฟีเจอร์หลัก ๆ ที่ให้บริการ เช่น หน้าแสดงความสัมพันธ์ระหว่างยีนและมะเร็ง 2 ชนิด จากรูปที่ 16(A) แสดงภูมิทัศน์ของยีนมะเร็ง 2 ชนิด พบว่าในมะเร็งเต้านม ยีน *TP53* มีค่าอัตราการกลายพันธุ์ (mutation rate) เท่ากับ 33.1% ในตัวอย่างมะเร็งเต้านม ในทางกลับกันมะเร็งไต จะพบยีน *VHL* มีค่าอัตราการกลายพันธุ์สูงเท่ากับ 40% และยีน *TSHZ3* มีค่าอัตราการกลายพันธุ์ที่ต่ำมากและยังคงตรวจไม่พบในมะเร็งทั้ง 2 ชนิด รูปที่ 16(B) และ 16(C) แสดงกราฟวงกลม (pie chart) โดยแบ่งวงกลมตามสัดส่วนการทำนายผลกระทบของการกลายพันธุ์ต่อลำดับสายโปรตีนในมะเร็งเต้านมและมะเร็งไตตามลำดับ

รูปที่ 17(A) แสดงสถานะเมธิลเลชันของยีน *SFRP1* และ *SFRP4* ในมะเร็งลำไส้ใหญ่ (colon adenocarcinoma: COAD) ซึ่งพบว่ายีน *SFRP1* เกิดเมธิลเลชันในบริเวณที่เป็นส่วนของยีน (body) สูงอย่างมีนัยสำคัญเมื่อเทียบกับบริเวณอื่น ๆ ในขณะที่ไม่พบความแตกต่างของการเกิดเมธิลเลชันในบริเวณต่าง ๆ ของยีน *SFRP4* ซึ่งถูกใช้เป็นยีนอ้างอิง ในรูปที่ 17(B) และ 17(C) แสดงกราฟค่าประมาณการกระจายตัวของค่าความแตกต่างของการเกิดเมธิลเลชันเปรียบเทียบระหว่างตัวอย่างปกติและตัวอย่างมะเร็ง โดยพิจารณาจากยีน *SFRP1* และ *SFRP4* ตามลำดับ แกน x แสดงค่าความแตกต่างของการเกิดเมธิลเลชันระหว่างตัวอย่าง และแกน y แสดงจำนวนตัวอย่าง

นอกจากนี้ยังมีกราฟแสดงความสัมพันธ์ของการแสดงออกของยีน (gene expression) ในมะเร็งอื่น ๆ ซึ่งคำนวณจากค่ามาตรฐาน z-score และนำมาแสดงความสัมพันธ์ในรูปแบบกราฟน้ำตก (waterfall plot) และในรูปแบบกราฟการลงจุดแบบกล่อง (box plot) เพื่อดูว่าการแสดงออกไปในทิศทางเดียวกันหรือไม่



รูปที่ 16: ภูมิทัศน์ของยีนมะเร็ง 2 ชนิด (A) เปรียบเทียบอัตราการกลายพันธุ์ของยีน เมื่อเทียบกับโรคมะเร็งต่างชนิดกัน (B) และ (C) แสดงกราฟวงกลมตามสัดส่วนการกลายพันธุ์ประเภทต่าง ๆ ในยีน *TP53* และ *VHL* ตามลำดับ (ที่มา: รูปที่ 1(A), 1(B), 1(C) ของ [41])



รูปที่ 17: กราฟผลการวิเคราะห์การเกิดเมธิลเลชันของยีน (A) เปรียบเทียบผลการเกิดเมธิลเลชันในบริเวณต่าง ๆ ของยีน *SFRP1* และ *SFRP4* (B) และ (C) กราฟแสดงค่าประมาณการกระจายตัวของค่าความแตกต่างของการเกิดเมธิลเลชันเปรียบเทียบระหว่างตัวอย่างปกติและตัวอย่างมะเร็งโดยพิจารณาจากยีน *SFRP1* และ *SFRP4* ตามลำดับ (ที่มา: รูปที่ 2(A), 2(B), 2(C) ของ [41])

2.2.3 เว็บพอร์ทัลแบบบูรณาการสำหรับจีโนมมะเร็ง (canEvolve: A Web Portal for Integrative Oncogenomics)

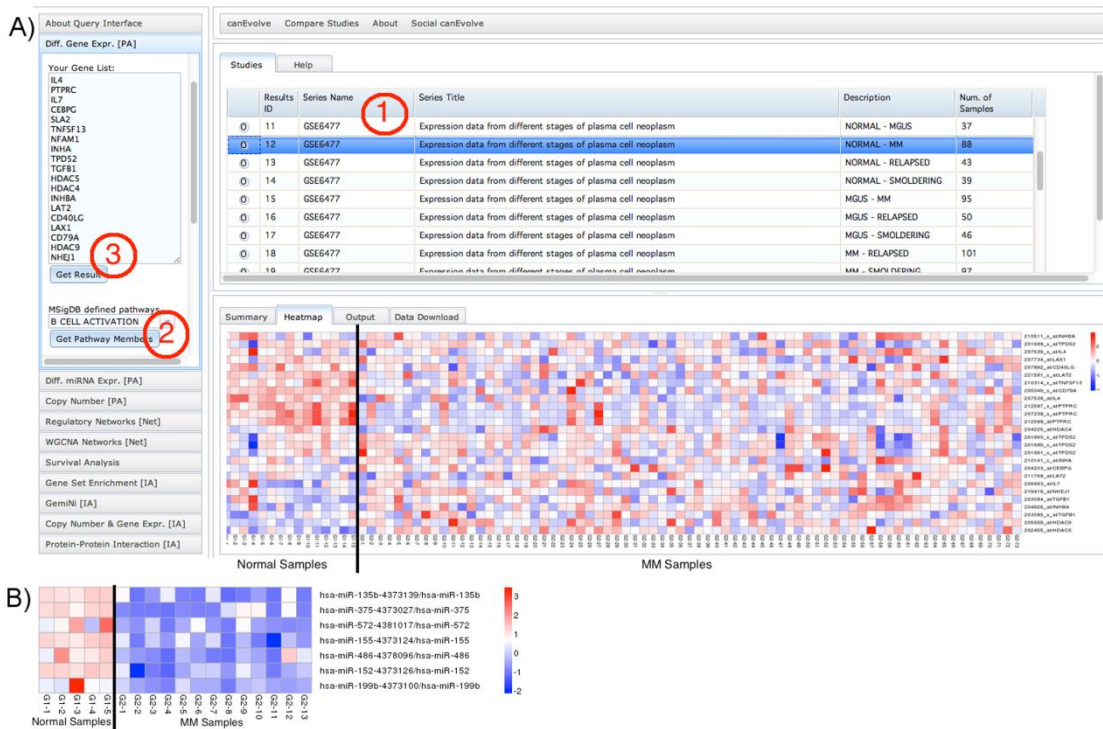
เว็บพอร์ทัลแบบบูรณาการสำหรับจีโนมมะเร็งหรือแคนอีโวลฟ์ [42] ถูกพัฒนาให้เป็นแหล่งรวบรวมจีโนมมะเร็ง ซึ่งจัดเก็บข้อมูลไมโครอาร์เอ็นเอ (microRNA) การแสดงออกของโปรตีน (protein expression profile) การแปรผันของจำนวนดีเอ็นเอ ข้อมูลการมีปฏิสัมพันธ์ระหว่างโปรตีน (protein-protein interaction) เพื่อช่วยนักวิจัยโรคมะเร็งในการวิเคราะห์ข้อมูลเบื้องต้น เช่น การวิเคราะห์แบบบูรณาการของการแสดงออกของไมโครอาร์เอ็นเอ รวมทั้งการวิเคราะห์การแปรผันของจำนวนชุดดีเอ็นเอด้วยวิธีการตรวจโครโมโซมอะเรย์ชนิดสนิปอะเรย์ (SNP array) และกราฟความสัมพันธ์ระหว่างการแสดงออกของยีนที่ทำให้เกิดลักษณะทางคลินิกและระยะรอดชีพของผู้ป่วยมะเร็ง แคนอีโวลฟ์ให้บริการข้อมูลจีโนมมะเร็ง 90 ประเภท จากผู้

ป่วยมากกว่า 10,000 ราย ซึ่งช่วยระบุเซลล์ที่ทำหน้าที่ควบคุมการตอบสนองทางภูมิคุ้มกันโรค
มะเร็ง ตารางที่ 3 แสดงประเภทและจำนวนข้อมูลที่อยู่ในระบบ

ตารางที่ 3: ประเภทและจำนวนชุดข้อมูลในแคณวีโวลฟ์

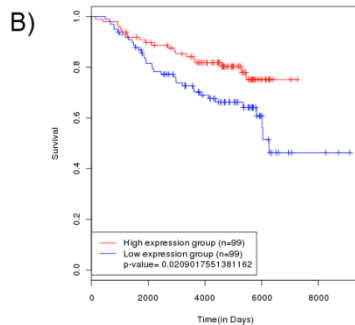
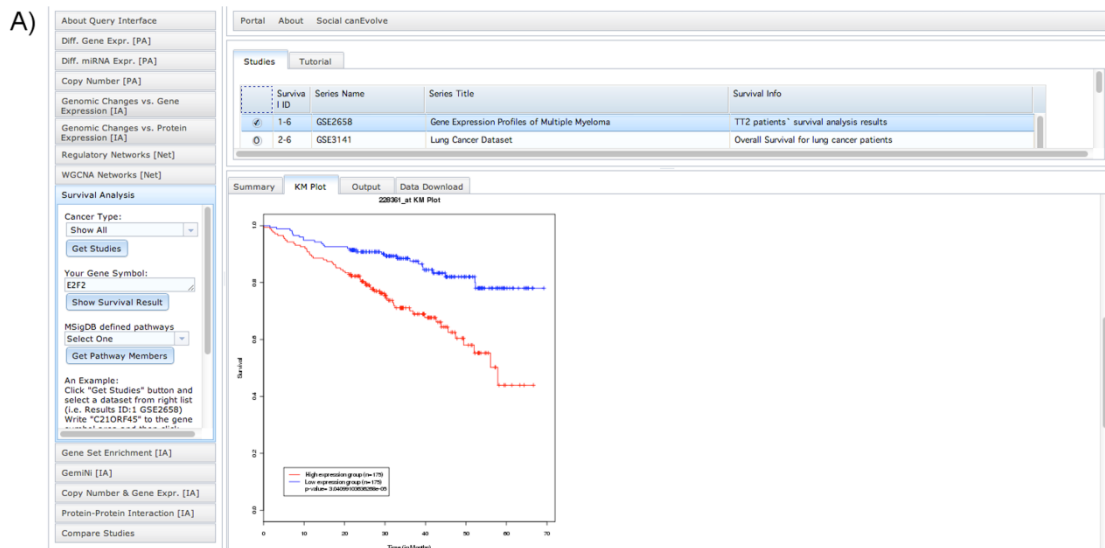
คอลัมน์	ประเภทข้อมูล	จำนวนชุดข้อมูล	จำนวนผู้ป่วย
1	Gene Expression	55	6,677
2	Copy Number	43	6,537
3	miRNA Expression	7	466
4	Mutation	14	2,867
5	Protein Expression	8	2,190
6	Protein-Protein Interaction	NA	NA

แคณวีโวลฟ์มีบริการหลัก ๆ เช่น การแสดงผลการวิเคราะห์ข้อมูลเบื้องต้นเชิงรูปภาพ รูป
ที่ 18(A) เป็นตัวอย่างการแสดงผลการเปรียบเทียบการแสดงออกของยีน เมื่อถูกกระตุ้นระหว่าง
เซลล์เม็ดเลือดขาวประเภทลิมโฟไซต์ (lymphocyte) ชนิดบีเซลล์ (B cell) และมะเร็งเม็ดเลือด
ขาวมัลติเพิลมายอีโลมา (multiple myeloma: MM) ด้วยชุดข้อมูล GSE6477 รูปที่ 18(A) 1)
แสดงรายการชุดข้อมูลที่จัดเตรียมสำหรับการวิเคราะห์ 2) พิมพ์กลุ่มยีนของมะเร็งในพาธเวย์ใน
ช่องค้นหา เพื่อดึงรายชื่อยีนภายใต้พาธเวย์ 3) รายการยีนที่สนใจ รูปที่ 18(B) ตัวอย่างการ
แสดงผลด้วยแผนที่ความร้อนเปรียบเทียบการแสดงออกของไมโครอาร์เอ็นเอในเซลล์ปกติกับ
เซลล์มะเร็งด้วยชุดข้อมูล GSE6477 ซึ่งข้อมูลที่นำมาวิเคราะห์เป็นชุดข้อมูลย่อย และไม่พบความ
ผิดปกติของเซลล์ในชุดข้อมูลดังกล่าว



รูปที่ 18: ผลการวิเคราะห์ข้อมูลเบื้องต้นเชิงรูปภาพ (A) 1) รายการชุดข้อมูลที่จัดเตรียมสำหรับการวิเคราะห์ 2) ช่องค้นหาเพื่อระบุกลุ่มยีนของมะเร็งในพยาธิที่สนใจ 3) รายการชื่อยีนที่สนใจ และ (B) แผนภูมิความร้อนเปรียบเทียบการแสดงออกของไมโครอาร์เอ็นเอในเซลล์ปกติกับมะเร็งเม็ดเลือดขาวมัลติเพิลมัยอีโลมา (ที่มา: รูปที่ 1(A), 1(B) ของ [42])

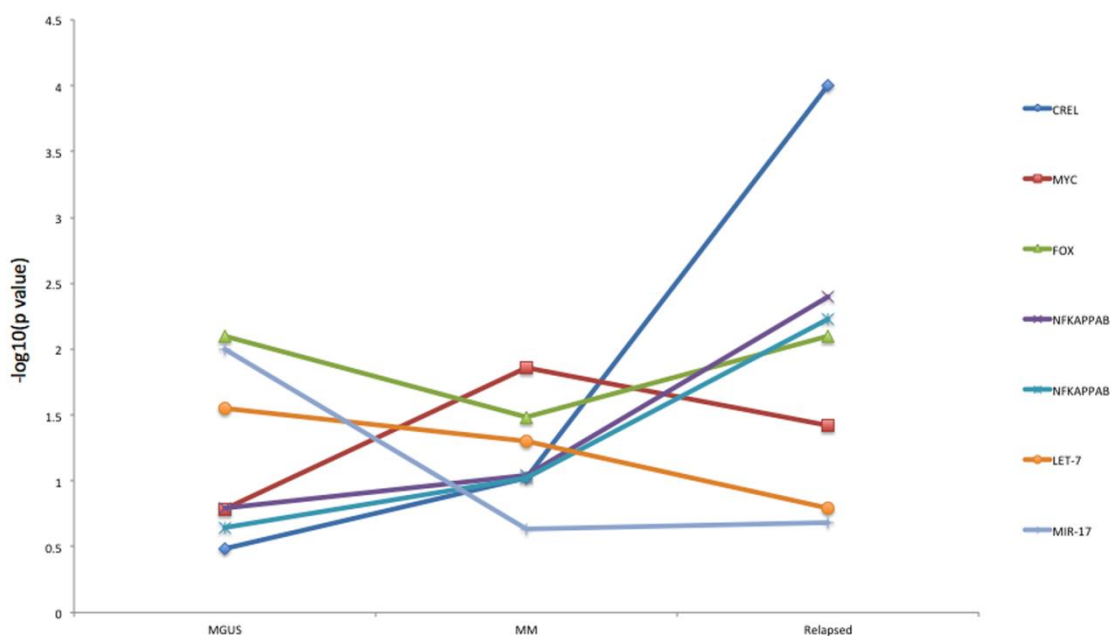
รูปที่ 19(A) แสดงผลการวิเคราะห์ความสัมพันธ์ระหว่างการแสดงออกของยีนและระยะรอดชีพของผู้ป่วยมะเร็งเม็ดเลือดขาวมัลติเพิลมัยอีโลมา โดยแสดงกราฟวิเคราะห์การรอดชีพ (survival analysis) ด้วยวิธีแคปแลนเมียร์ (kaplan-meier) ซึ่งแกน y คือ ความน่าจะเป็นของระยะรอดชีพในมะเร็งเม็ดเลือดขาวมัลติเพิลมัยอีโลมาเทียบกับแกน x ระยะเวลาตั้งแต่เริ่มรักษาถึงเสียชีวิต มีหน่วยเป็นจำนวนเดือน ด้านล่างขวาแสดงตัวอย่างความสัมพันธ์ระหว่างการแสดงออกของยีน *E272* และระยะรอดชีพ ด้วยชุดข้อมูล GSE6477 โดยแบ่งกลุ่มยีน *E2F2* ตามค่ามัธยฐานในทุกตัวอย่างออกเป็น 2 กลุ่ม และใช้สถิติ log-rank test คำนวณหาค่าพีแวลู (p-value) เพื่อเปรียบเทียบระยะเวลารอดชีพของกราฟทั้ง 2 เส้นว่ามีความแตกต่างกันหรือไม่ และรูปที่ 19(B) แสดงกราฟผลการวิเคราะห์แบบเดียวกันกับรูปที่ 19(A) ต่างกันที่แกน x มีหน่วยเป็นจำนวนวัน แต่เนื่องจากผลการวิเคราะห์ระหว่างรูปที่ 19(A) และ 19(B) ไม่สอดคล้องกัน เพราะจำนวนตัวอย่างที่นำมาวิเคราะห์ไม่เท่ากัน



รูปที่ 19: ผลการวิเคราะห์ความสัมพันธ์ระหว่างการแสดงออกของยีนและระยะรอดชีพของผู้ป่วยมะเร็งด้วยชุดข้อมูล GSE6477 (A) กราฟแสดงความน่าจะเป็นของระยะรอดชีพเทียบกับระยะเวลา (เดือน) และ (B) กราฟแสดงความน่าจะเป็นของระยะรอดชีพเทียบกับระยะเวลา (วัน) (ที่มา: รูปที่ 2(A), 2(B) ของ [42]) จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 20 แสดงการสร้างแบบจำลองทำนายการเกิดมะเร็งเม็ดเลือดขาวมัลติเพิลมีโยโลมา โดยหาความสัมพันธ์ระหว่างทรานสคริปชันแฟคเตอร์ (Transcription factors) และยีนเป้าหมายที่มีการแสดงออกที่เกี่ยวข้องกับการเจริญเติบโตอย่างผิดปกติจนกลายเป็นมะเร็ง เส้นกราฟแต่ละเส้นแสดงถึงระดับการแสดงออกของทรานสคริปชันแฟคเตอร์และยีนเป้าหมาย แกน x คือ ระยะของโรคมะเร็ง ได้แก่ ระยะการวินิจฉัยแยกโรคที่สำคัญ ซึ่งลักษณะทางคลินิกพบได้ตั้งแต่ไม่มีอาการ (Benign monoclonal gammopathy of undetermined significance: MGUS) จนกระทั่งเริ่มแสดงอาการที่บ่งบอกว่าเป็นโรคระยะมะเร็งเม็ดเลือดขาวมัลติเพิลมีโยโลมา (Multiple myeloma: MM) และระยะโรคมะเร็งย้อนกลับเป็นซ้ำ (relapsed stage) และ แกน y เป็นค่าที่ได้จากการคำนวณหาค่า $-\log_{10}(\text{p-value})$ ด้วยวิธีการวิเคราะห์ความสำคัญของกลุ่มยีน (gene set enrichment analysis: GSEA) โดยใช้ฐานข้อมูลจาก MSigDB (molecular

signatures) [31] จากกราฟพบว่าทรานสคริปชันแฟกเตอร์ ได้แก่ MYC, FOX, NF-kB [46], ไมโครอาร์เอ็นเอ MIR-17 และ LET-7 ซึ่งควบคุมการแสดงออกของยีนตอบสนองต่อการอักเสบมีความน่าจะเป็นที่ทำให้ผู้ป่วยระยะเริ่มแรกของโรคมะเร็งที่เปลี่ยนไปเป็นมะเร็งเม็ดเลือดขาวมัลติเพิลมายอีโลมาอย่างมีนัยสำคัญ และทรานสคริปชันแฟกเตอร์ CREL ซึ่งเป็นสมาชิกกลุ่ม NF-kB มีความน่าจะเป็นที่ทำให้ผู้ป่วยระยะมะเร็งเม็ดเลือดขาวมัลติเพิลมายอีโลมาเปลี่ยนไปเป็นระยะโรคมะเร็งย้อนกลับเป็นซ้ำอย่างมีนัยสำคัญ



รูปที่ 20: ตัวอย่างผลทำนายการเกิดมะเร็งเม็ดเลือดขาวมัลติเพิลมายอีโลมาโดยหาความสัมพันธ์ระหว่างทรานสคริปชันแฟกเตอร์และยีนเป้าหมายด้วยวิธีวิเคราะห์ความสำคัญของกลุ่มยีน (ที่มา: รูปที่ 7 ของ [42])

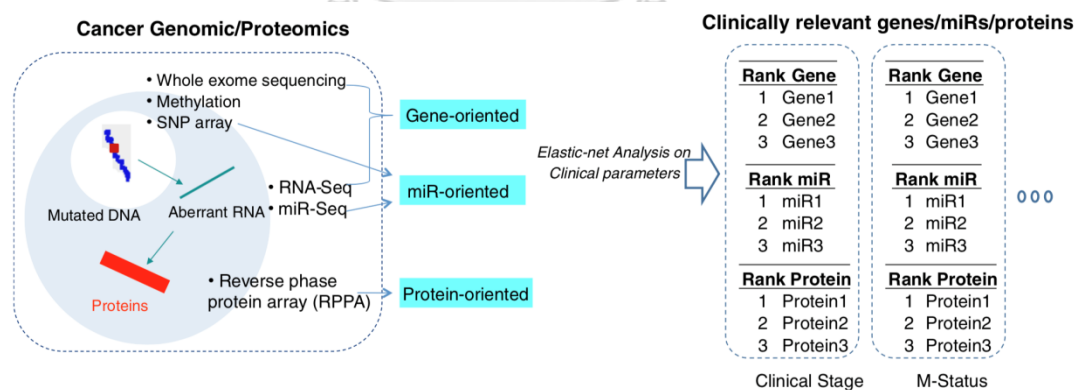
นอกจากนี้ยังมีวิธีวิเคราะห์เครือข่ายการควบคุมการแสดงออกของยีนด้วยอัลกอริทึม ARACNE [43] ของทรานสคริปชันแฟกเตอร์ที่มีต่อยีนเป้าหมายและแสดงผลเครือข่ายด้วยโปรแกรมไซโตสเคป [44] (<http://www.cytoscape.org>) การแสดงความสัมพันธ์ระหว่างความผิดปกติของจีโนมและการแสดงออกของยีนหรือโปรตีนด้วยรูปภาพ การแสดงผลการวิเคราะห์ข้อมูลการมีปฏิสัมพันธ์ของการควบคุมระหว่างทรานสคริปชันแฟกเตอร์หรือไมโครอาร์เอ็นเอกับยีนเป้าหมายด้วยวิธีวิเคราะห์ข้อมูลเชิงเครือข่าย dChip-GemiNI (gene and miRNA network-based integration) [45] และคำนวณค่าเอฟดีอาร์ (false discovery rate: FDR) เพื่ออธิบายความแตกต่างของการแสดงออกของยีนกลุ่มไมโครอาร์เอ็นเอระหว่างเซลล์ปกติและเซลล์

มะเร็ง และการแสดงผลการวิเคราะห์อภิมาน (meta-analysis) ของหลาย ๆ ผลการศึกษาที่ได้จากแคนอีโวลฟ์เข้าด้วยกัน

2.2.4 เว็บและโมบายอินเตอร์เฟซสำหรับวิเคราะห์หาความสัมพันธ์ระหว่างข้อมูลทางคลินิกและจีโนม (The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations)

งานวิจัยนี้นำเสนอซอฟต์แวร์แพลตฟอร์มที่มุ่งเน้นการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลจีโนมและข้อมูลทางคลินิก [46] จากชุดข้อมูลที่ซีจีเอครอบคลุมโรคมะเร็ง 25 ชนิด และข้อมูลโอมิกส์อื่น ๆ เช่น ข้อมูลในระดับทรานสคริปโตมที่วัดการแสดงออกของยีนทั้งจีโนมผ่านเทคโนโลยีอย่างอาร์เอ็นเอซีค ข้อมูลในระดับโปรตีโอมที่วัดการแสดงออกของโปรตีน เป็นต้น โดยมีการนำความรู้ทางจีโนมิกส์และโปรตีโอมิกส์มาอธิบายอิทธิพลของยีนที่ทำให้เกิดลักษณะทางคลินิกและการตอบสนองต่อยา เพื่อสนับสนุนการแพทย์แม่นยำ

รูปที่ 21 แสดงการวิเคราะห์แบบการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ ซึ่งศึกษาความสัมพันธ์ของตัวแปรอิสระที่มีความสัมพันธ์สูง ๆ ในกลุ่มข้อมูลประเภทเดียวกัน ได้แก่ กลุ่มข้อมูลยีน กลุ่มข้อมูลไมโครอาร์เอ็นเอและกลุ่มข้อมูลโปรตีน เพื่อตรวจสอบว่าตัวแปรในข้อมูลทางคลินิกมีความสัมพันธ์กับยีน ไมโครอาร์เอ็นเอหรือโปรตีนจริงหรือไม่ โดยเรียกวธีนี้ว่าอีลาสติกเน็ต (elastic net)



รูปที่ 21: กระบวนการวิเคราะห์ข้อมูลทางคลินิกและจีโนมด้วยวิธีแบบอีลาสติกเน็ต (ที่มา: รูปที่ 1 ของ [46])

จากรูปที่ 22(A) ผู้ใช้สามารถระบุชนิดของตัวแปรในกลุ่มข้อมูลที่สนใจ เช่น ยีน ประเภทของมะเร็ง ไมโครอาร์เอ็นเอหรือโปรตีน เป็นต้น และระบุข้อมูลทางคลินิกที่ต้องการตรวจสอบความสัมพันธ์ จากรูปที่ 22(B) ผู้ใช้สามารถตรวจสอบการกลายพันธุ์ของโซมาติกใน

มะเร็งชนิดต่าง ๆ ว่ามีความเกี่ยวข้องกับข้อมูลทางคลินิกอย่างไร นอกจากนี้ผู้ใช้สามารถทดสอบสมมติฐานแบบ Knudson's two-hit hypothesis ที่มีแนวคิดว่าการกลายพันธุ์เกิดขึ้น 2 ครั้งในกลุ่มยีนต้านมะเร็ง (tumor suppressor gene) ซึ่งครั้งแรกเกิดการกลายพันธุ์ในเซลล์สืบพันธุ์ (germ cell) ได้รับการถ่ายทอดผ่านทางพันธุกรรม และครั้งที่สองเกิดการกลายพันธุ์ในเซลล์ร่างกาย ส่งผลให้ยีนนั้นทำงานผิดปกติก่อให้เกิดเป็นเซลล์มะเร็ง จากรูปที่ 22(C) ผู้ใช้สามารถระบุชนิดของมะเร็ง ระบุชื่อยีน ไมโครอาร์เอ็นเอหรือโปรตีน ที่เกิดการกลายพันธุ์ครั้งแรกและครั้งที่สองได้

รูปที่ 23 แสดงหน้าผลตัวอย่างการค้นหาความสัมพันธ์ระหว่างยีนและข้อมูลทางคลินิก จากตัวอย่างมีการระบุชื่อยีนเป็น *TP53* และชนิดของมะเร็งเป็นมะเร็งลำไส้ใหญ่และไส้ตรง (colorectal cancer: COADREAD) รูปที่ 23(A) แสดงตารางผลการคำนวณค่าพีแวลู เพื่อนำมาวิเคราะห์ว่ายีน *TP53* ส่งผลสำคัญต่อข้อมูลทางคลินิกตัวแปรตัวใดบ้าง โดยกำหนดค่าพีแวลูจะต้องน้อยกว่า 0.01 อธิบายได้ว่ายีน *TP53* มีความสัมพันธ์กับข้อมูลทางคลินิกตัวแปรนั้นจริง โดยเรียงลำดับความสำคัญจากมากไปน้อย ซึ่งแบ่งการแสดงผลข้อมูลเป็น 3 แท็บได้แก่ แท็บแสดงผลการวิเคราะห์ทั้งหมด (outcome summary) แท็บแสดงความถี่ในการกลายพันธุ์ (frequency-mutation) และแท็บแสดงความถี่ของซีเอ็นวี (frequency-copy number) รูปที่ 23(B) แสดงตัวอย่างตารางสรุปจำนวนกลุ่มข้อมูลยีน ไมโครอาร์เอ็นเอและโปรตีน ซึ่งนำมาวิเคราะห์หาความสัมพันธ์ระหว่างมะเร็งลำไส้ใหญ่และไส้ตรงกับข้อมูลทางคลินิกต่าง ๆ เช่น เพศ ประวัติครอบครัว เป็นต้น และรูปที่ 23(C) แสดงตารางสรุปจำนวนกลุ่มข้อมูลยีน ไมโครอาร์เอ็นเอและโปรตีน ที่ถูกนำมาวิเคราะห์เพื่อหาความสัมพันธ์ระหว่างยีนต่าง ๆ ที่เกี่ยวข้องกับมะเร็งลำไส้ใหญ่และไส้ตรงกับระยะของโรคตามลักษณะทางคลินิก (clinical stage)

A What Genes are associated with a Clinical Parameter of your interest? These genes are identified by Elastic-net analysis.

i Check if Gene/Protein/miR of your interest is associated with Clinical Parameters

Target: Gene Protein miR

ii What Cancer Type are you interested in?

iii What Clinical Parameter are you interested in?

B Query by Gene, miR, or Protein by Cancer Type and Clinical Parameter
(Ex: Genetic profiles of TP53 in Colorectal cancer by MSI status)

Target: Gene Protein miR

C Two-hit hypothesis test
(Ex: Mutation counts for TP53 and WRN from COADREAD)

Select First Target/Genomic pair

Target: Gene Protein miR

Select Second Target/Genomic pair

Target: Gene miR

รูปที่ 22: หน้าจอเพื่อการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทางคลินิก และจีโนมด้วยวิธีการอีลาสติกเน็ต (A) ผู้ใช้สามารถระบุชนิดของตัวแปรในกลุ่มข้อมูลที่สนใจได้ เช่น ยีน (B) สามารถตรวจสอบมะเร็งที่เกิดจากการกลายพันธุ์ของเซลล์ปกติในร่างกายว่าเกี่ยวข้องกับข้อมูลทางคลินิกอย่างไร และ (C) ระบุชนิดของมะเร็งเพื่อทดสอบสมมติฐานแบบ two-hit hypothesis (ที่มา: รูปที่ 2(A), 2(B), 2(C) ของ [46])

A. The list of clinical parameters that *TP53* is associated with

Outcome Summary		Frequency - Mutation		Frequency - Copy Number				
15 records per page		Search: <input type="text"/>						
Cancer Type	Clinical Parameter	Rank	Gene	Copy Number	mRNA expression	Methylation	Mutation	P-value
COADREAD	Country	40	TP53			48 ↓		0.000E+00
HNSC	ClinicalStage	41	TP53			67 ↑		1.912E-05
PRAD	N-Status	24	TP53			28 ↓		5.439E-04
STAD	MolecularSubtype	2	TP53				1 ↑	2.340E-16
STAD	MolecularSubtype	15	TP53				16 ↓	3.355E-08
UCEC	HistoGrade	4	TP53				5 ↑	2.765E-20
UCEC	HistoType	42	TP53				42 ↓	8.334E-22

B. Summary of COADREAD analysis

OverStage	Clarity	Family	Gene
#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49
HistoType	H-Status	M-Status	N-Status
#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49
Phenotypic	T-Status		
#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49		

C. Summary of clinical stage analysis

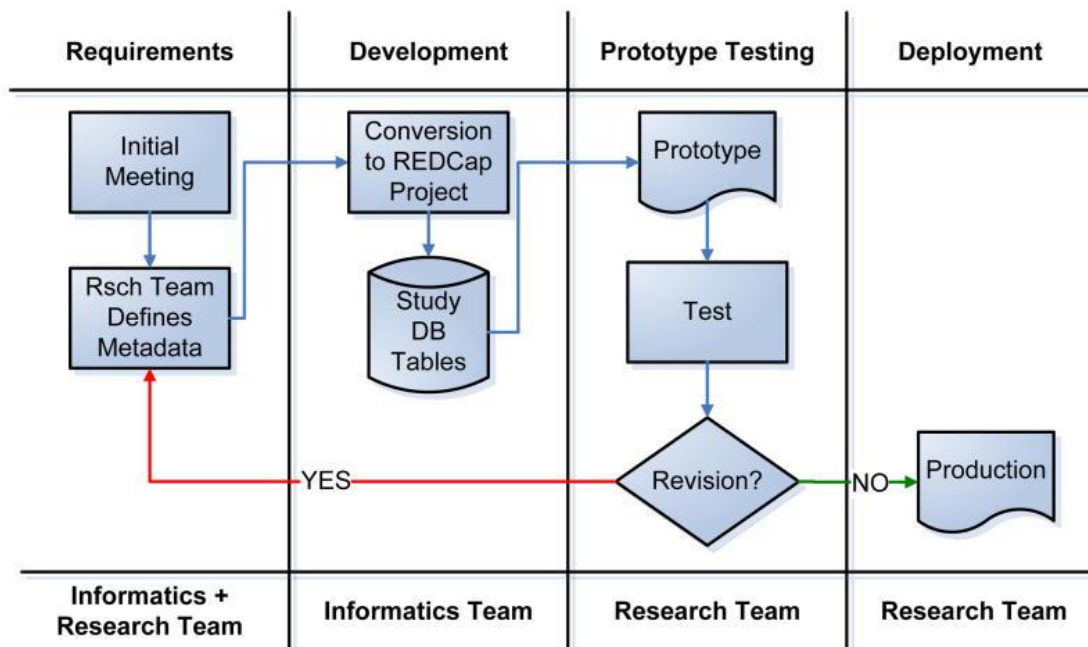
ACC	BLCA	BRCA	COADREAD
#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49
ESGA	HNSC	LIHC	LUSC
#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49
LGG	LIRC	LUSC	LUSC
#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49	#Candidates: 47 #Samples: 47 #Protein: 10 #Mutations: 49

รูปที่ 23: ผลลัพธ์ของการค้นหาความสัมพันธ์ระหว่างยีนและข้อมูลทางคลินิก (A) แสดงตารางผลการคำนวณค่าพีแวลูของยีน *TP53* ส่งผลสำคัญต่อข้อมูลทางคลินิกตัวแปรตัวใดบ้าง โดยเรียงลำดับความสำคัญ (B) แสดงตารางสรุปจำนวนกลุ่มข้อมูลที่นำมาวิเคราะห์หาความสัมพันธ์ระหว่างมะเร็งลำไส้ใหญ่และไส้ตรงกับข้อมูลทางคลินิก และ (C) แสดงตารางสรุปจำนวนกลุ่มข้อมูลนำมาวิเคราะห์หาความสัมพันธ์ระหว่างยีนต่าง ๆ ที่เกี่ยวข้องกับมะเร็งลำไส้ใหญ่และไส้ตรงกับระยะของโรคตามลักษณะทางคลินิก (ที่มา: รูปที่ 4 ของ [46])

2.2.5 เรดแคป (Research Electronic Data Capture: REDCap)

เรดแคป [47] คือ ระบบจัดการข้อมูลงานวิจัยด้านคลินิกซึ่งถูกพัฒนาในรูปแบบของเว็บไซต์ แอปพลิเคชัน เรดแคปมีขั้นตอนการทำงานแบบลำดับ รูปที่ 24 แสดงกระบวนการกำหนดโครงสร้างข้อมูลที่จะจัดเก็บไว้ในฐานข้อมูลเพื่องานวิจัยด้านคลินิก เริ่มจากเก็บรวบรวมความต้องการ (requirements) ที่จะจัดเก็บแอตทริบิวต์อะไรบ้างในระบบ กำหนดรูปแบบของพจนานุกรมข้อมูล (data dictionary) โดยการเพิ่มแอตทริบิวต์ต่าง ๆ ที่ต้องการในไฟล์เทมเพลต ทำการอัปเดตโครงสร้างตารางข้อมูลใหม่ด้วยไฟล์เทมเพลตนี้ในฐานข้อมูลเรดแคป หลังจากนั้นสร้างเว็บไซต์ต้นแบบ เพื่อให้ทีมนักวิจัยเข้ามาทดสอบ ถ้าการทำงานของระบบถูกต้อง ระบบใหม่จะถูกนำขึ้นใช้งานจริง แต่ถ้าไม่ถูกต้องกระบวนการเก็บรวบรวมความต้องการจะถูกทำซ้ำอีก

ครั้ง เรดแคปจะจัดเก็บข้อมูลงานวิจัยด้านคลินิกเพียงเท่านั้น ซึ่งระบบไม่สามารถหาความเชื่อมโยงทางพันธุกรรมกับข้อมูลทางคลินิกได้



รูปที่ 24: กระบวนการกำหนดโครงสร้างข้อมูลในเรดแคป (ที่มา: รูปที่ 1 ของ [47])

2.2.6 จีโนมิกส์คอมมอนส์ของเอ็นซีไอ ฐานข้อมูลและเครื่องมือสำหรับการแพทย์แม่นยำ (NCI Genomic Data Commons (GDC) as an engine for precision medicine)

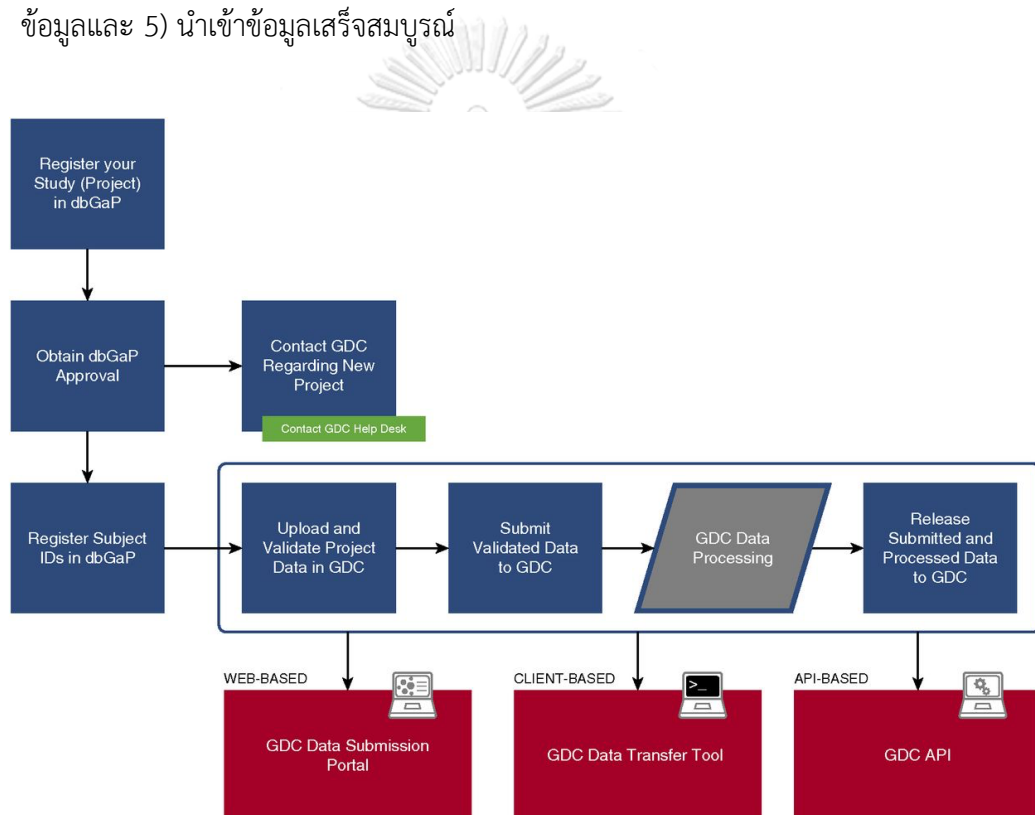
ข้อมูลจีโนมิกส์คอมมอนส์หรือจีดีซีของเอ็นซีไอ [48] เป็นแพลตฟอร์มที่เก็บรวบรวมข้อมูลจีโนมและข้อมูลทางคลินิก ทำการวิเคราะห์และนำผลที่ได้ ออกสู่สาธารณะเพื่อเป็นประโยชน์ต่อแพทย์เฉพาะทางที่ให้การรักษาโรคมะเร็ง โดยเอ็นซีไอพยายามผลักดันให้เกิดความร่วมมือระหว่างประเทศในการกำหนดมาตรฐานข้อมูลจีโนมและข้อมูลทางคลินิก

ลักษณะเด่นของแพลตฟอร์มจีดีซีคือ

1. จัดเก็บข้อมูลดิบ ซึ่งสามารถนำข้อมูลกลับมาคำนวณซ้ำใหม่ได้
2. ใช้เทคนิคไปป์ไลน์ในการวิเคราะห์ข้อมูลชีวสารสนเทศ เพื่อเพิ่มขีดความสามารถในการวิเคราะห์ข้อมูลที่มีความหลากหลายให้รวดเร็วและถูกต้อง
3. ออกแบบให้มีความยืดหยุ่นเพื่อรองรับข้อมูลที่เพิ่มขึ้นและมีความหลากหลาย
4. ข้อมูลจีโนมและข้อมูลทางคลินิกจะถูกจัดเก็บรักษาในจีดีซีทั้งหมด ซึ่งนักวิจัยสามารถนำข้อมูลไปใช้ในงานวิจัยได้โดยไม่เสียค่าใช้จ่ายใด ๆ
5. นักวิจัยทุกคนจะต้องปฏิบัติตามนโยบายของเอ็นไอเอช (National Institutes of Health: NIH)

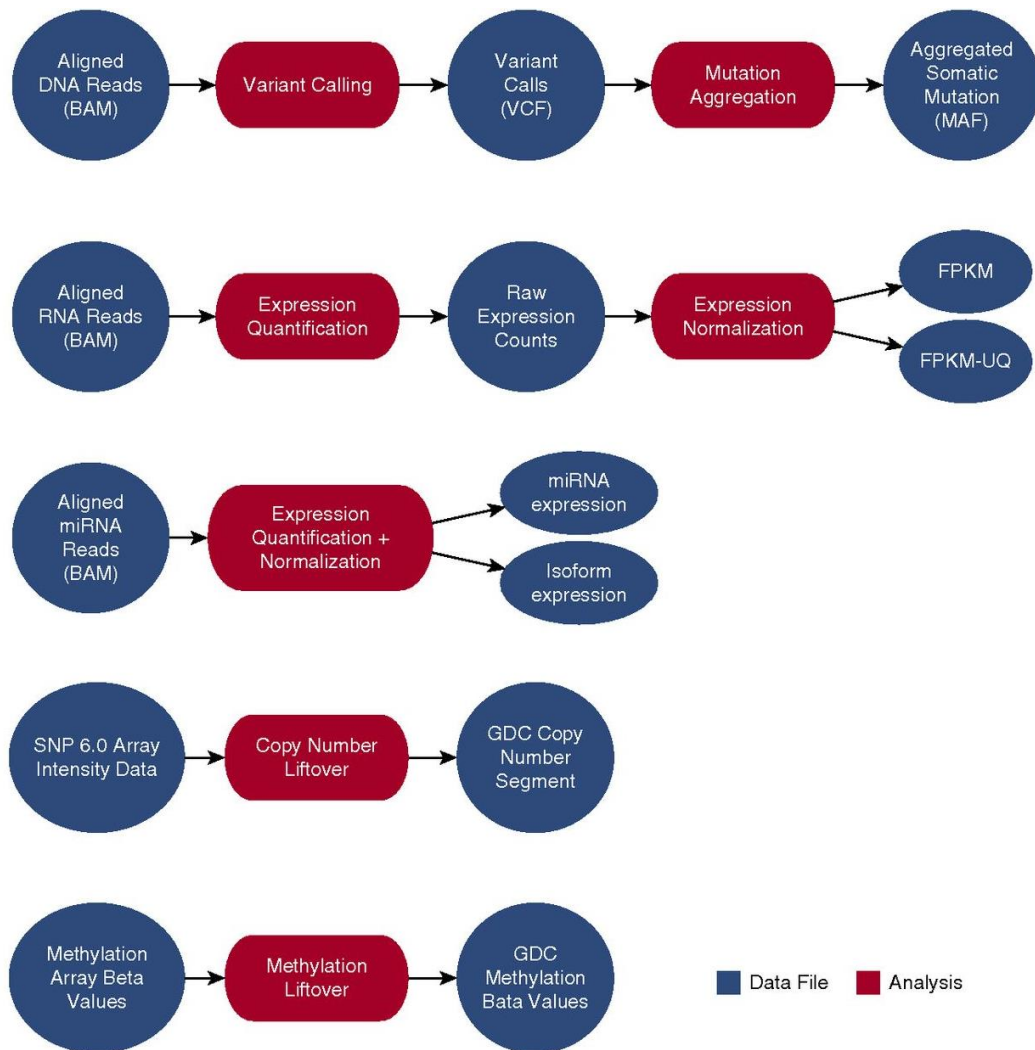
6. ข้อมูลในจีดีซีเติบโตขึ้นเรื่อย ๆ และเป็นฐานข้อมูลสำคัญที่ช่วยในการศึกษาวิจัยการแพทย์แม่นยำสำหรับโรคมะเร็ง

รูปที่ 25 แสดงการขออนุมัตินำข้อมูลเข้าระบบจีดีซี โดยมีขั้นตอนดังนี้ 1) ลงทะเบียนโครงการกับ dbGaP ซึ่งเป็นหน่วยงานที่เก็บรวบรวมข้อมูลจีโนมไทป์และพีโนมไทป์ โดย dbGaP จะทำการตรวจสอบข้อมูลทั้งหมดว่าได้รับการยินยอมจากผู้ป่วยจริงหรือไม่ เมื่อได้รับการอนุมัติจาก dbGaP 2) ลงทะเบียนข้อมูลผู้ป่วยและอัปโหลดเอกสารยินยอม 3) ทำการอัปโหลดข้อมูล ซึ่งขั้นตอนนี้จะมีการตรวจสอบความถูกต้องของข้อมูล ถ้าถูกต้องทั้งหมด 4) ทำการยืนยันการนำเข้าข้อมูลและ 5) นำเข้าข้อมูลเสร็จสมบูรณ์



รูปที่ 25: กระบวนการขออนุมัตินำข้อมูลเข้าระบบจีดีซี (ที่มา: รูปที่ 1 ของ [48])

รูปที่ 26 แสดงภาพรวมกระบวนการวิเคราะห์ข้อมูลชีวสารสนเทศด้วยเทคนิคไปป์ไลน์ จะเห็นว่าจีดีซีมีการประมวลผลข้อมูลหลากหลายรูปแบบ ได้แก่ การวิเคราะห์การแปรผันของดีเอ็นเอ การวิเคราะห์การแสดงออกของอาร์เอ็นเอและไมโครอาร์เอ็นเอ การค้นหาซีเอ็นวีจากข้อมูลสไนป์อาร์เรย์และการหาค่าระดับดีเอ็นเอเมทิลเลชัน

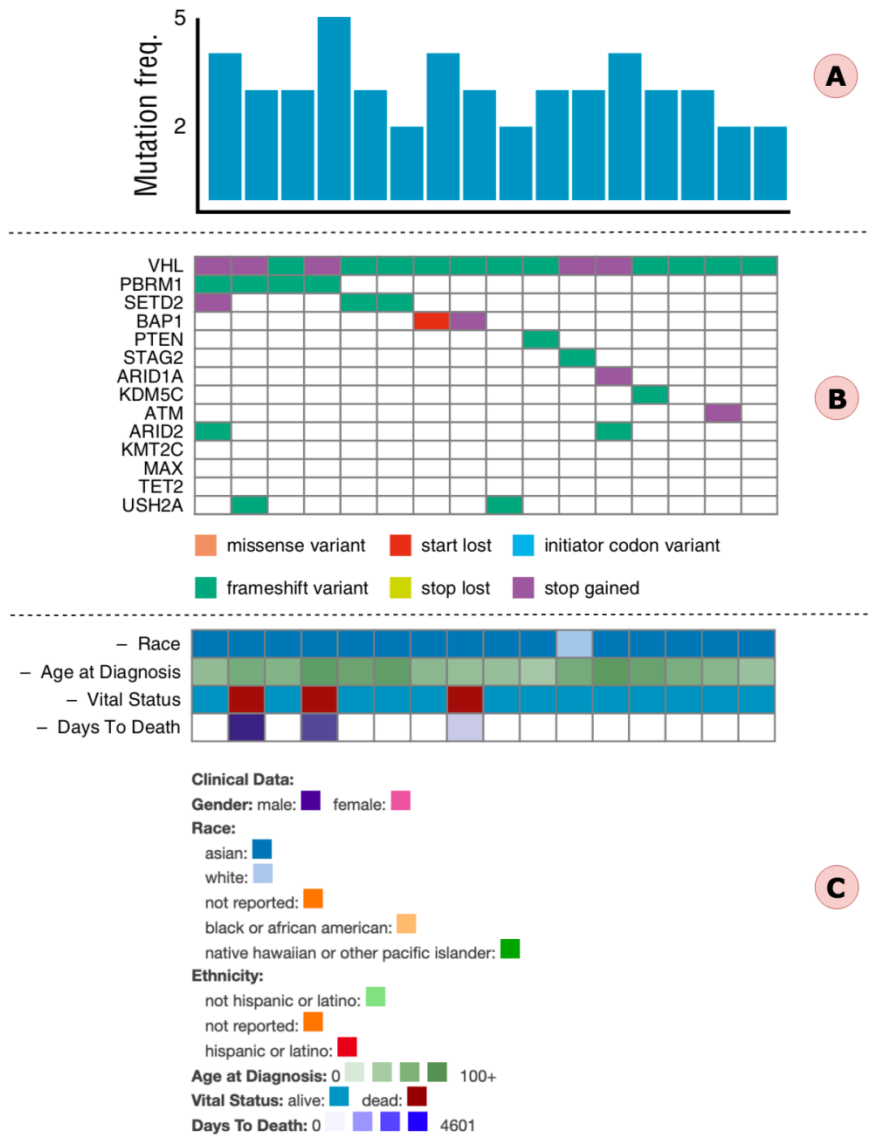


รูปที่ 26: ภาพรวมกระบวนการวิเคราะห์ข้อมูลชีวสารสนเทศด้วยเทคนิคไปป์ไลน์ (ที่มา: รูปที่ 2 ของ [48])

นอกจากนี้ผู้ใช้อย่างสามารถดาวน์โหลดข้อมูลจากจีดีซี เพื่อวิเคราะห์ข้อมูลแบบออฟไลน์ ประกอบด้วย 3 วิธีคือ 1) ดาวน์โหลดข้อมูลจากเว็บพอร์ทัล 2) ดาวน์โหลดข้อมูลจากแอปพลิเคชันที่ติดตั้งบนเครื่องผู้ใช้งาน และ 3) ดาวน์โหลดข้อมูลจากเอพีไอ ซึ่งจำเป็นต้องมีโทเค็น (token) เพื่อยืนยันตัวตน

ในการวิเคราะห์และการแสดงผลเชิงรูปภาพ จีดีซีได้มีพัฒนาการแสดงผลบนแพลตฟอร์มซีไบโอพอร์ทัล (cBioPortal) ซึ่งประกอบด้วยพีเจอร์หลักของจีดีซีของโคกริด (GDC OncoGrid) รูปที่ 27(A) กราฟฮิสโตแกรมแสดงการแจกแจงความถี่ของการแปรผันของเซลล์ร่างกาย แกน x คือการศึกษาเฉพาะราย (single case) และแกน y คือจำนวนการแปรผันของเซลล์ร่างกาย รูปที่ 27(B) ตารางกริดแสดงความสัมพันธ์ระหว่างยีน (แถว) กับตัวอย่างเนื้องอกของกรณีศึกษา

(คอสม์) และค่าสีในช่องคือ รูปแบบของการกลายพันธุ์ เช่น สีส้ม แสดงการแปรผันของเบส ส่งผลให้การถอดมีโนเปลี่ยนรูปแบบ (missense variant) เป็นต้น รูปที่ 27(C) ตารางกริดแสดง ความสัมพันธ์ระหว่างข้อมูลทางคลินิก (แถว) กับการศึกษาเฉพาะราย (คอสม์) และค่าสีในช่อง คือ ค่าต่าง ๆ ทางคลินิก ซึ่งที่ถูกละเลยมไว้ได้กราฟ



รูปที่ 27: ตัวอย่างการแสดงผลเชิงรูปภาพด้วยจีดีซี อองโคกริด (คัดเลือกมาบางส่วนจากรูปที่ 4 ของ [48])

2.2.7 เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

งานวิจัยที่กล่าวมาข้างต้นนำเสนอซอฟต์แวร์แพลตฟอร์มที่เป็นแหล่งรวบรวมข้อมูล เพื่อใช้ในการตรวจสอบข้อมูล การวิเคราะห์ข้อมูลและแสดงผลข้อมูลในรูปแบบแผนภาพด้วยชุดข้อมูลสาธารณะ (public dataset) เช่น ข้อมูลจากโครงการแผนที่จีโนมมะเร็ง เป็นต้น โดยเน้นการวิเคราะห์ข้อมูลจีโนมและข้อมูลโอมิกส์เป็นหลัก ข้อดีคือ แอตทริบิวต์ข้อมูลทางคลินิกน้อย และไม่มีส่วนจัดการข้อมูลทางคลินิก เมื่อเทียบกับเรดแคป (REDCap) ซึ่งเป็นระบบจัดการข้อมูลงานวิจัยด้านคลินิกโดยตรง แต่ข้อดีของเรดแคปคือ ไม่สามารถวิเคราะห์ความสัมพันธ์ระหว่างจีโนมและข้อมูลโอมิกส์ต่าง ๆ ได้ ตารางที่ 4 เปรียบเทียบชนิดข้อมูลที่ถูกนำไปใช้ในแต่ละแพลตฟอร์ม ตารางที่ 5 เปรียบเทียบฟีเจอร์หลักในแต่ละแพลตฟอร์มกับซอฟต์แวร์แพลตฟอร์ม RUN-ONCO ที่นำเสนอในวิทยานิพนธ์ฉบับนี้



ตารางที่ 4: เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

Platform	GENIPAC	Web-TCGA	canEvolve	TCGA Clinical Explore	REDCap	NCI GDC
cancer type						
All		X	X	X	X	X
clinical data						
Demographics	X			X	X	X
Outcomes	X			X	X	
Biological results	X			X	X	
Images					X	
Structured clinical research data	X			X	X	
-omics data						
mRNA expression	X			X		X
miRNA expression			X	X		X
SNPs						X
Copy number variations	X	X	X	X		X
DNA methylation		X		X		X
Protein/Phosphoprotein expression			X			

ตารางที่ 4: เปรียบเทียบงานวิจัยที่เกี่ยวข้อง (ต่อ)

Platform	GENIPAC	Web-TCGA	canEvolve	TCGA Clinical Explore	REDCap	NCI GDC
Structural rearrangements	X					X
external data links						
COSMIC	X		X	X		X
OncoKB	X					

ตารางที่ 5: เปรียบเทียบฟีเจอร์หลักในแต่ละแพลตฟอร์ม

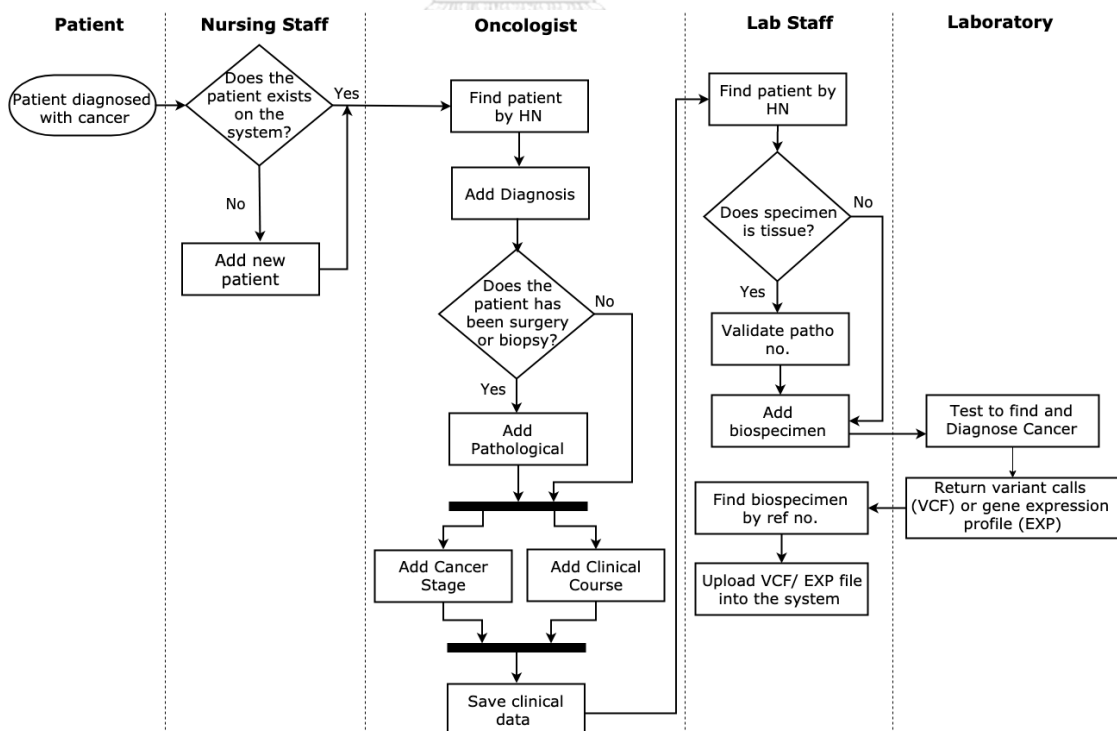
Platform	GENIPAC	Web-TCGA	canEvolve	TCGA Clinical Explorer	REDCap	NCI GDC	RUN-ONCO
public dataset	X	X	X	X			
manage data (add/edit/delete)							
clinical data					X		X
biospecimen					X		X
-omics data							X
data visualization	X	X	X	X		X	X
extensible component based architecture							X

บทที่ 3

วิธีการดำเนินงานวิจัย

3.1 ภาพรวมการทำงานของระบบ

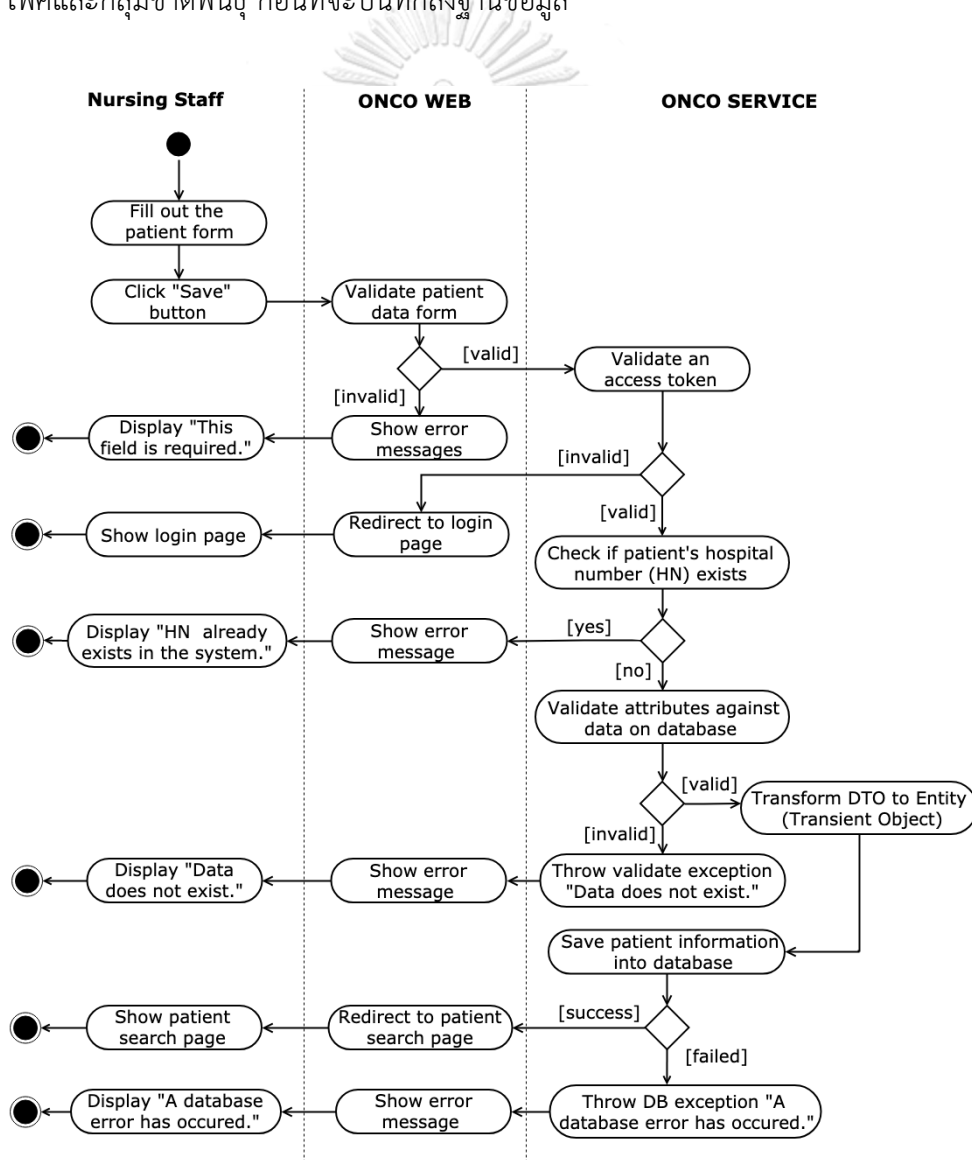
แพลตฟอร์มถูกออกแบบมาเพื่อลดความซับซ้อนของการวินิจฉัยและออกแบบการรักษาแบบการแพทย์แม่นยำในโรคมะเร็ง ซึ่งช่วยให้ผู้ใช้ระบบสามารถจัดการและวิเคราะห์ข้อมูลของตนเองได้ โดยมีขั้นตอนการทำงานแบบลำดับ จากรูปที่ 28 เริ่มจากนางพยาบาลนำประวัติผู้ป่วยที่ได้รับการวินิจฉัยว่าเป็นโรคมะเร็งบันทึกข้อมูลเข้าสู่ระบบ ทุกครั้งที่ผู้ป่วยเข้ามาได้รับการรักษา แพทย์ผู้ดูแลก็จะบันทึกข้อมูลทางคลินิกของผู้ป่วยรายดังกล่าว ได้แก่ ข้อมูลการวินิจฉัยโรค (diagnosis) การวินิจฉัยทางพยาธิวิทยา (pathological) ระยะของโรคมะเร็ง (cancer stage) และการดำเนินโรค (clinical course) โดยระบบจะทำการเชื่อมโยงกับประวัติผู้ป่วยให้อัตโนมัติ ถ้าครั้งใดมีการเก็บตัวอย่างตรวจ ได้แก่ เลือด เซลล์หรือเนื้อเยื่อจากผู้ป่วย เจ้าหน้าที่ห้องปฏิบัติการจะบันทึกข้อมูลรายละเอียดตัวอย่างเก็บตรวจ ได้แก่ ประเภทของตัวอย่างเก็บตรวจ วิธีการแช่เยือก วิธีการเก็บรักษา เป็นต้น และจะนำส่งตัวอย่างเก็บตรวจไปที่ห้องปฏิบัติการ เพื่อเข้าสู่กระบวนการถอดรหัสจีโนมและการถอดรหัสเอ็กโซมด้วยเทคโนโลยีโอมิกส์ ผลวิเคราะห์จะอยู่ในรูปแบบไฟล์วีซีเอฟหรือไฟล์อีเอ็กซ์พี โดยเจ้าหน้าที่ห้องปฏิบัติการจะนำไฟล์ดังกล่าวอัปโหลดเข้าสู่ระบบต่อไป



รูปที่ 28: ภาพรวมกระบวนการทำงานของระบบ

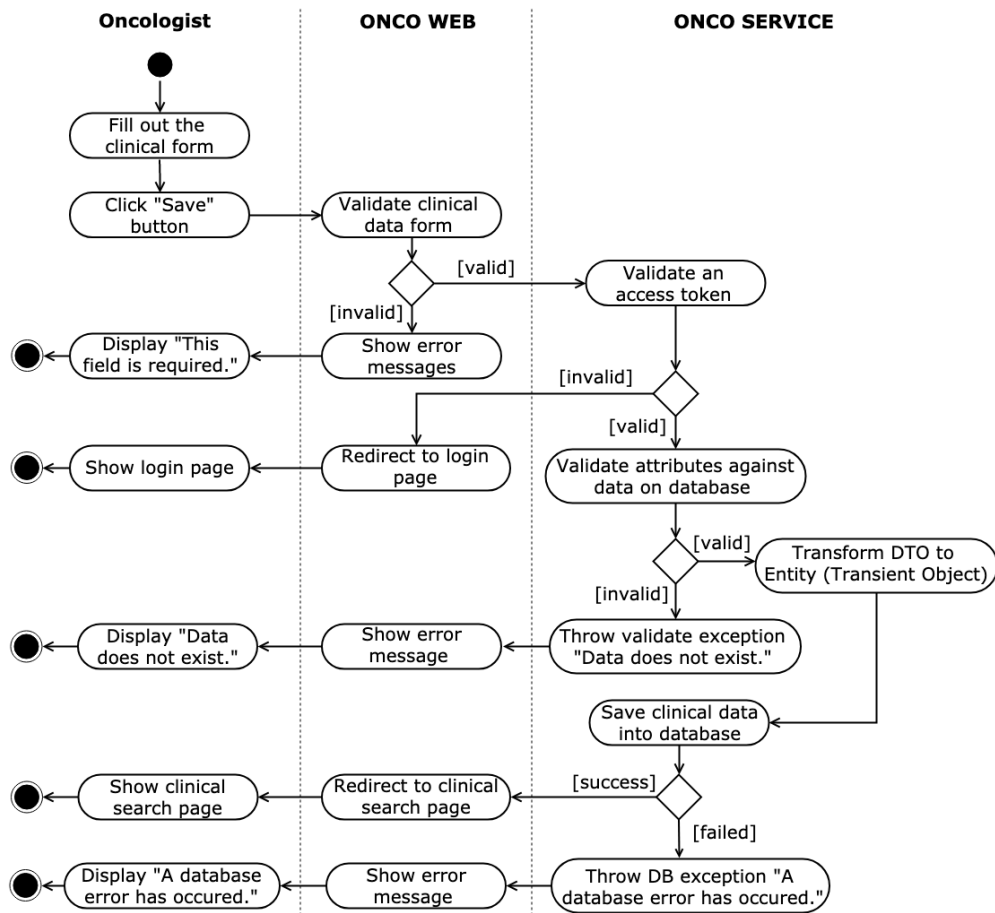
3.2 แผนภาพแสดงขั้นตอนการทำงานในระบบ

รูปที่ 29 แสดงขั้นตอนการจัดเก็บข้อมูลผู้ป่วยในครั้งแรก เริ่มต้นนางพยาบาลกรอกประวัติผู้ป่วย ได้แก่ ข้อมูลเลขที่เวชระเบียน (HN) ข้อมูลพื้นฐานของผู้ป่วย (demographic information) เช่น เพศ วันเกิด เชื้อชาติ น้ำหนัก ส่วนสูง เป็นต้น แล้วกดปุ่ม “Save” เพื่อบันทึกข้อมูล เว็บองค (เป็นส่วนหน้าของระบบ โดยมีรายละเอียดในหัวข้อที่ 3.3) จะตรวจสอบความครบถ้วนถูกต้อง ก่อนส่งคำร้องขอไปยังองคเซิร์ฟวิส (เป็นส่วนหลังบ้านของระบบ โดยมีรายละเอียดในหัวข้อที่ 3.3) เพื่อบันทึกประวัติผู้ป่วย โดยจะแนบโทเค้น (token) เพื่อยืนยันตัวตนกับองคเซิร์ฟวิสทุกครั้ง นอกจากนี้องคเซิร์ฟวิสยังมีขั้นตอนการตรวจสอบความถูกต้องของข้อมูลพื้นฐาน (master data) กับฐานข้อมูล ได้แก่ เพศและกลุ่มชาติพันธุ์ ก่อนที่จะบันทึกลงฐานข้อมูล



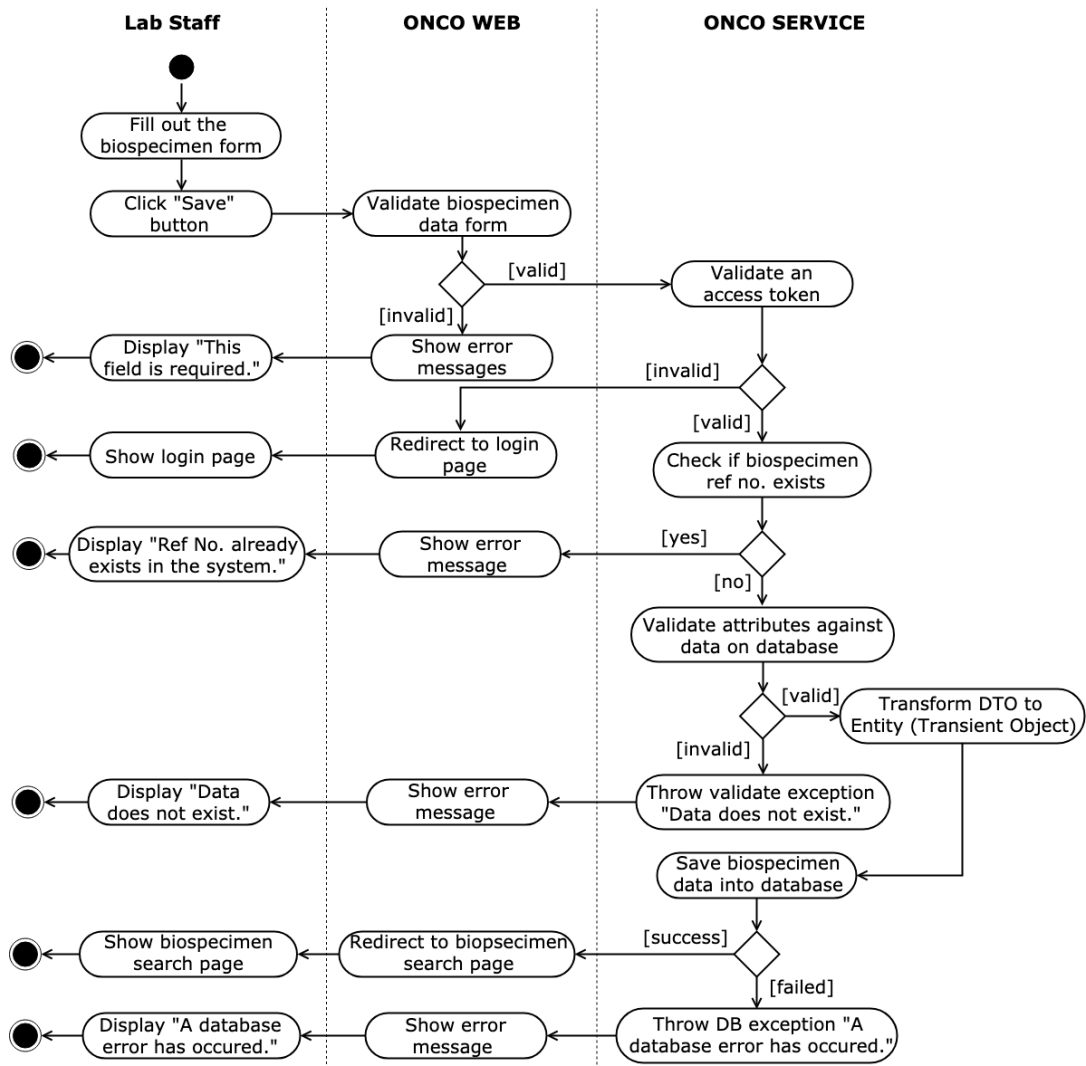
รูปที่ 29: แผนภาพกิจกรรมของส่วนจัดการข้อมูลประวัติผู้ป่วย

รูปที่ 30 แสดงขั้นตอนการจัดเก็บข้อมูลทางคลินิก ซึ่งข้อมูลทางคลินิกแบ่งออกเป็น 4 ประเภทหลัก 1) การวินิจฉัยโรค (patient diagnosis) โดยข้อมูลการวินิจฉัยหลักได้นำรหัสวินิจฉัยโรคในระบบไอซีดี-10 (10th revision of the International Statistical Classification of Diseases and Related Health Problems: ICD-10) [49] มาใช้งาน 2) การวินิจฉัยทางพยาธิวิทยา (pathological diagnosis) เป็นรายละเอียดของชิ้นเนื้อทางพยาธิวิทยา นอกจากนี้ได้นำรหัสวินิจฉัยโรคในระบบไอซีดี-โอ (International Classification of Diseases for Oncology: ICD-O) [50] มาอธิบายคุณลักษณะของชิ้นเนื้อ 3) ระยะของโรคมะเร็ง (cancer stage) ได้นำระบบ TNM ของ AJCC มาใช้งาน เพื่อจำแนกระยะของโรคแบบมาตรฐานสากลทั่วโลก [<http://cancerstaging.org>] และ 4) การดำเนินโรค (clinical course) เพื่อเฝ้าติดตามโรคมะเร็งย้อนกลับเป็นซ้ำ เมื่อแพทย์ผู้ดูแลกรอกข้อมูลทางคลินิกของผู้ป่วย แล้วกดปุ่ม “Save” เพื่อบันทึกข้อมูล เว็บไซต์จะตรวจสอบความครบถ้วนถูกต้อง ก่อนส่งคำร้องขอไปยังห้องไอซีเอชเอชเอชเพื่อบันทึกประวัติผู้ป่วย โดยจะแนบโทเค็น (token) เพื่อยืนยันตัวตนกับห้องไอซีเอชเอชเอชทุกครั้ง นอกจากนี้ห้องไอซีเอชเอชเอชยังมีขั้นตอนการตรวจสอบความถูกต้องของข้อมูลผู้ป่วยด้วยรหัสเวชทะเบียนและข้อมูลพื้นฐานข้อมูลอื่น ๆ ได้แก่ ประเภทของเนื้อเยื่อ (tissue type) และระดับของเนื้องอก (histologic grade) เป็นต้น ก่อนที่จะบันทึกลงฐานข้อมูล



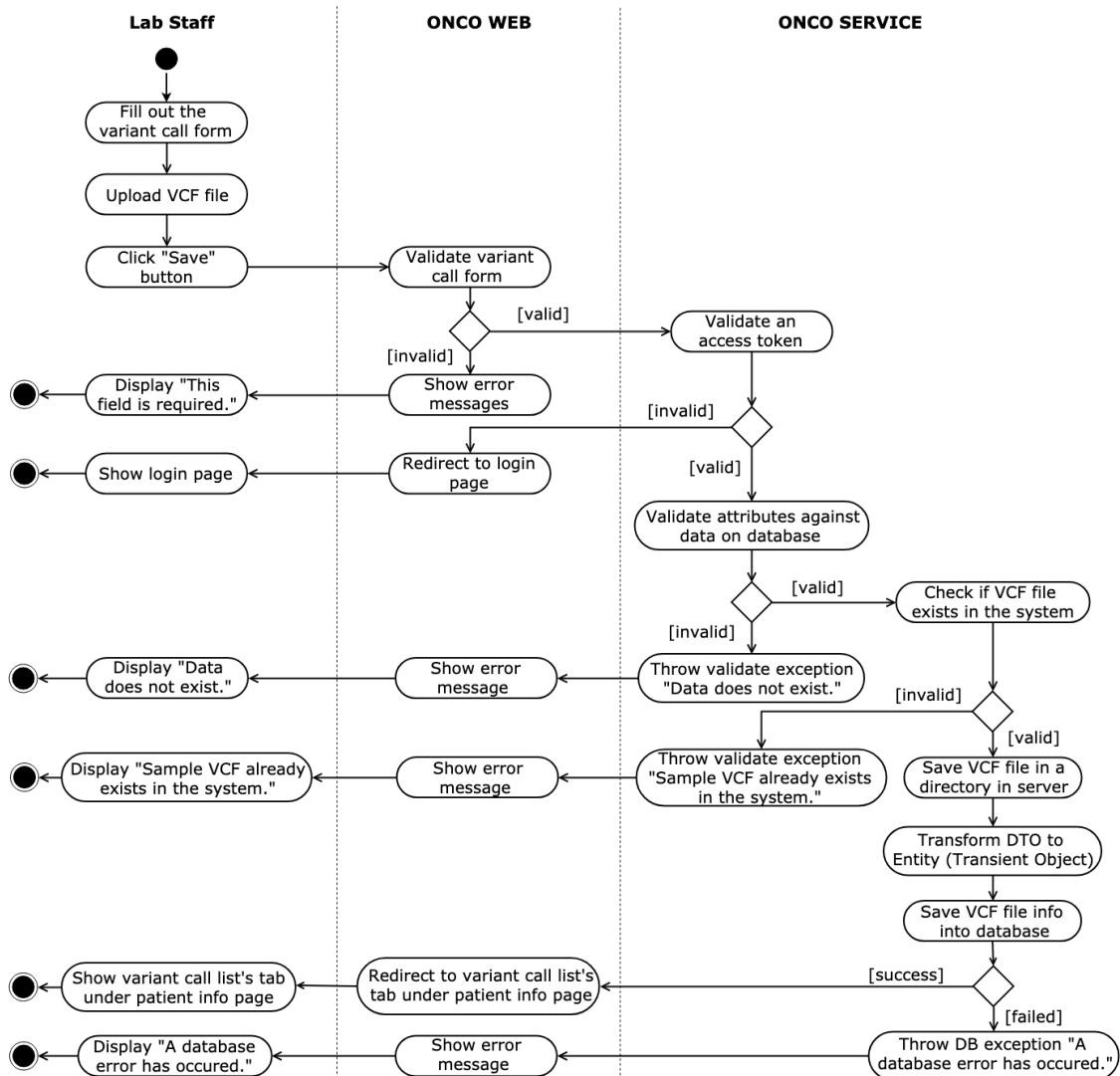
รูปที่ 30: แผนภาพกิจกรรมของส่วนจัดการข้อมูลทางคลินิก

รูปที่ 31 แสดงขั้นตอนการจัดเก็บข้อมูลชีววัตถุ (biospecimen) ประกอบไปด้วย ประเภทของตัวอย่างเก็บตรวจ (sample type) ได้แก่ เลือด (blood) เนื้อเยื่อ (tissue) และเซลล์ (cell) โดยเจ้าหน้าที่ห้องปฏิบัติการจะบันทึกข้อมูลรายละเอียดชีววัตถุที่เก็บตรวจ แล้วกดปุ่ม “Save” เพื่อบันทึกข้อมูล เว็บไซต์จะตรวจสอบความครบถ้วนถูกต้อง ก่อนส่งคำร้องขอไปยังห้องแล็บเพื่อบันทึกประวัติผู้ป่วย โดยจะแนบโทเค็น (token) เพื่อยืนยันตัวตนกับห้องแล็บทุกครั้ง นอกจากนี้ห้องแล็บยังมีขั้นตอนการตรวจสอบความถูกต้องของข้อมูลผู้ป่วยด้วยรหัสเวชทะเบียนและข้อมูลพื้นฐาน ได้แก่ ประเภทของตัวอย่างเก็บตรวจ (sample type) และประเภทของเนื้อเยื่อ (โดยจะตรวจสอบข้อมูลประเภทของเนื้อเยื่อ ก็ต่อเมื่อเลือกประเภทของตัวอย่างเก็บตรวจเป็นเนื้อเยื่อ) เป็นต้น ก่อนที่จะบันทึกลงฐานข้อมูล



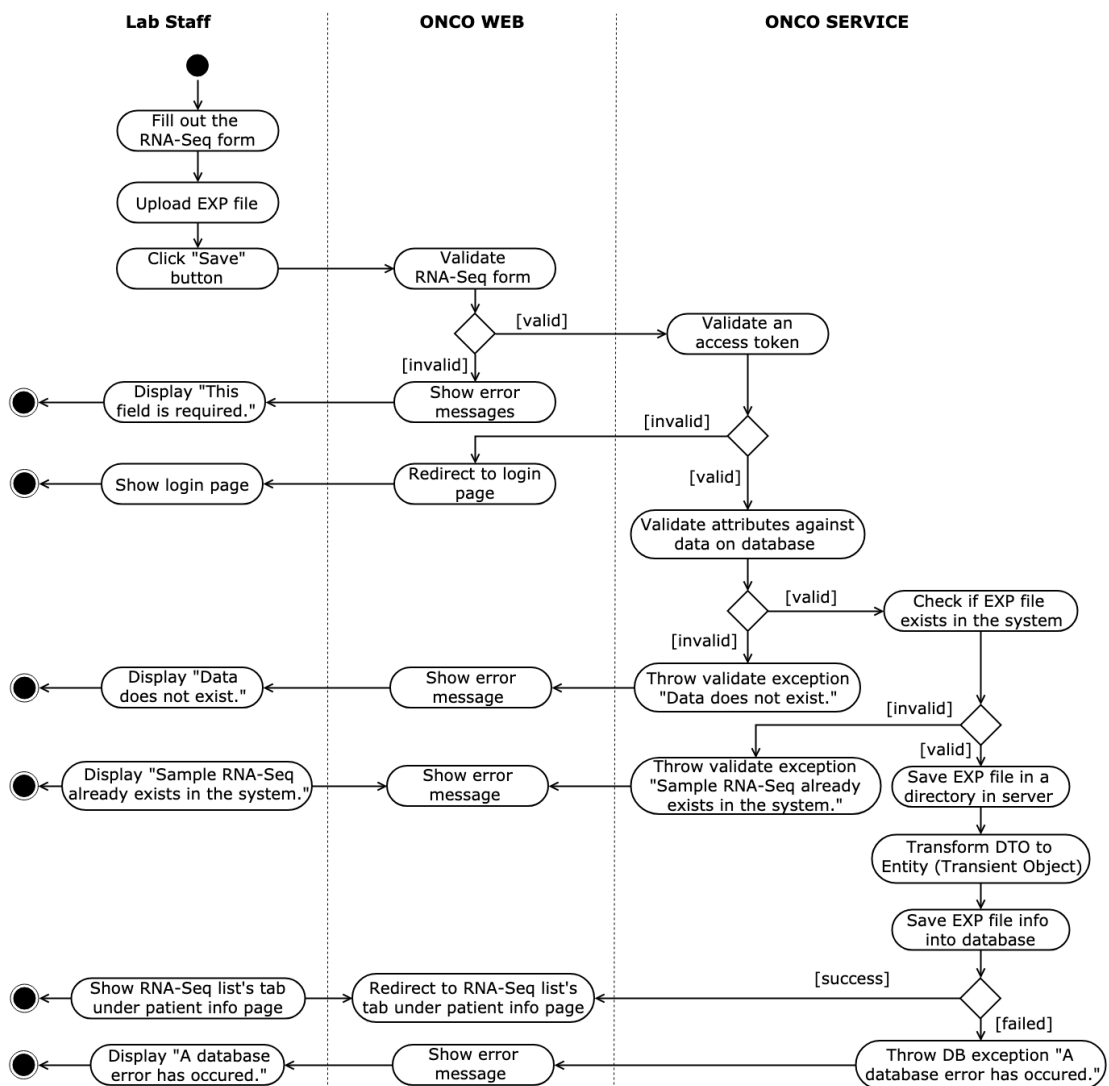
รูปที่ 31: แผนภาพกิจกรรมของส่วนจัดการข้อมูลชีววัตถุ

รูปที่ 32 แสดงขั้นตอนการนำเข้าข้อมูลการแปรผันของดีเอ็นเอ เมื่อได้รับผลการวิเคราะห์จากห้องปฏิบัติการในรูปแบบไฟล์ซีเอฟ เจ้าหน้าที่ห้องปฏิบัติการจะต้องเลือกประเภทการถอดรหัส (sequence type) และรหัสของชีววัตถุของผู้ป่วย (sample source) แล้วจึงอัปโหลดไฟล์ซีเอฟเมื่อกดปุ่ม “Save” เว็บองคจะตรวจสอบความครบถ้วนถูกต้อง ก่อนส่งคำร้องขอไปยังองคเซอร์วิสเพื่ออัปโหลดไฟล์ซีเอฟเข้าสู่ระบบ โดยจะแนบโทเค้น (token) เพื่อยืนยันตัวตนกับองคเซอร์วิสทุกครั้ง นอกจากนี้องคเซอร์วิสมีขั้นตอนการตรวจสอบความถูกต้องของรหัสชีววัตถุของผู้ป่วยว่าเป็นตัวอย่างเก็บตรวจของผู้ป่วยรายนี้จริง และไฟล์ซีเอฟดังกล่าวเคยนำเข้าสู่ระบบไปแล้วหรือไม่ ขั้นตอนนี้ไฟล์ซีเอฟจะถูกจัดเก็บไว้ในโฟลเดอร์ที่ถูกสร้างจัดเตรียมไว้บนเซิร์ฟเวอร์ โดยกำหนดให้มีสถานะ “DRAFT” เพื่อรอการทำงานของแบตซีในการสกัดข้อมูลจากไฟล์ซีเอฟต่อไป



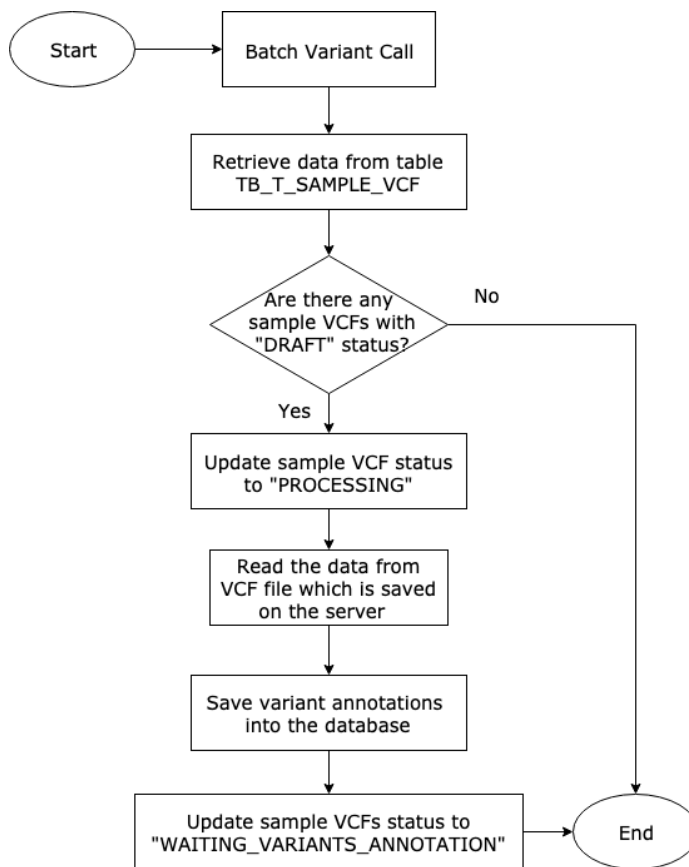
รูปที่ 32: แผนภาพกิจกรรมของส่วนนำเข้าไฟล์วีซีเอฟ

รูปที่ 33 แสดงขั้นตอนการนำเข้าไฟล์การแสดงผลของยีนจากเทคโนโลยีอาร์เอ็นเอซีค เมื่อได้รับผลการวิเคราะห์จากห้องปฏิบัติการในรูปแบบไฟล์อีเอ็กซ์พี เจ้าหน้าที่ห้องปฏิบัติการจะต้องเลือกรหัสชีววัตถุของผู้ป่วย (sample source) แล้วจึงอัปโหลดไฟล์อีเอ็กซ์พี เมื่อกดปุ่ม “Save” เว็บของโคจะตรวจสอบความครบถ้วนถูกต้อง ก่อนส่งคำร้องขอไปยังองโคเซอร์วิสเพื่ออัปโหลดไฟล์อีเอ็กซ์พีเข้าสู่ระบบ โดยจะแนบโทเค้น (token) เพื่อยืนยันตัวตนกับองโคเซอร์วิสทุกครั้ง นอกจากนี้องโคเซอร์วิสมีขั้นตอนการตรวจสอบความถูกต้องของรหัสชีววัตถุของผู้ป่วย ว่าเป็นชีววัตถุของผู้ป่วยรายนี้จริง และไฟล์อีเอ็กซ์พีดังกล่าวเคยนำเข้าสู่ระบบไปแล้วหรือไม่ ขั้นตอนนี้ข้อมูลจะจัดเก็บในรูปแบบไฟล์ไว้ในโฟลเดอร์ที่ถูกสร้างจัดเตรียมไว้บนเซิร์ฟเวอร์ โดยกำหนดให้มีสถานะ “DRAFT” เพื่อรอการทำงานของแบตช์ในการสกัดข้อมูลจากไฟล์อีเอ็กซ์พีต่อไป



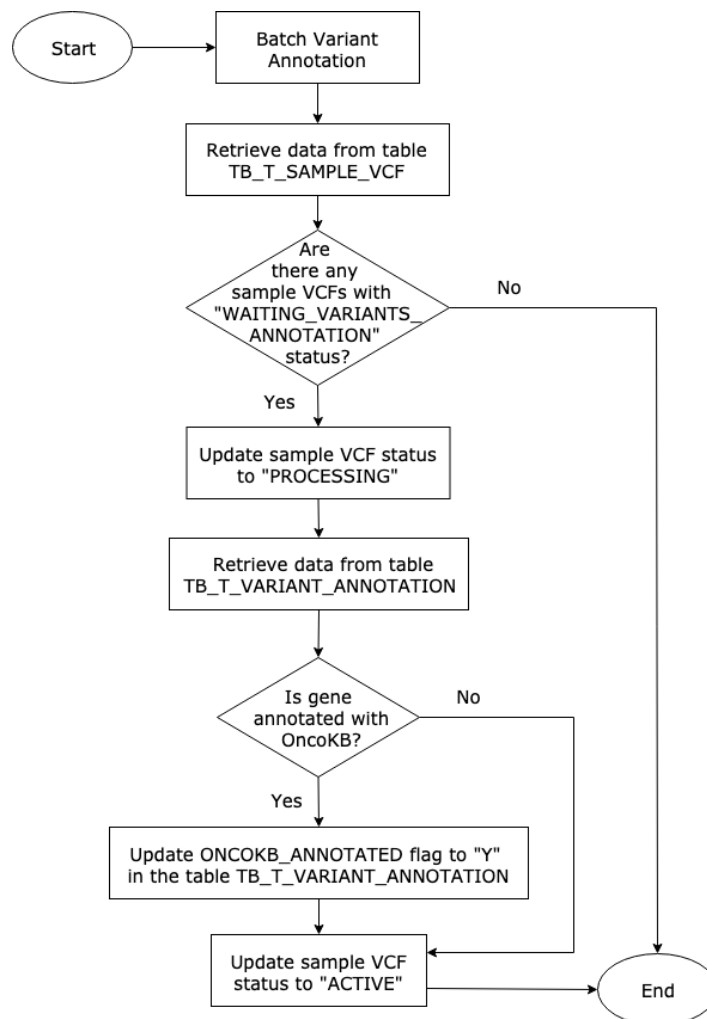
รูปที่ 33: แผนภาพกิจกรรมของส่วนนำเข้าไฟล์อีเอ็กซ์พี

รูปที่ 34 แสดงขั้นตอนการสกัดข้อมูลจากไฟล์วีซีเอฟ แบตช์จะทำงานทุก ๆ 5 นาที จะดึงข้อมูลรายละเอียดไฟล์วีซีเอฟเฉพาะที่มีสถานะ “DRAFT” เท่านั้น และอัปเดตสถานะของข้อมูลที่อ่านขึ้นมาให้เป็นสถานะ “PROCESSING” โดยอ่านพารของไฟล์วีซีเอฟดังกล่าวจากฐานข้อมูล แล้วจึงเริ่มสกัดข้อมูลจากไฟล์วีซีเอฟ และบันทึกข้อมูลทั้งหมดที่ได้ลงฐานข้อมูล นอกจากนี้จะอัปเดตสถานะของข้อมูลรายละเอียดไฟล์วีซีเอฟให้เป็นสถานะ “WAITING_VARIANTS_ANNOTATED” เพื่อรอแบตช์งานที่ทำการลิงก์ข้อมูลเข้ากับองค์ประกอบต่อไป



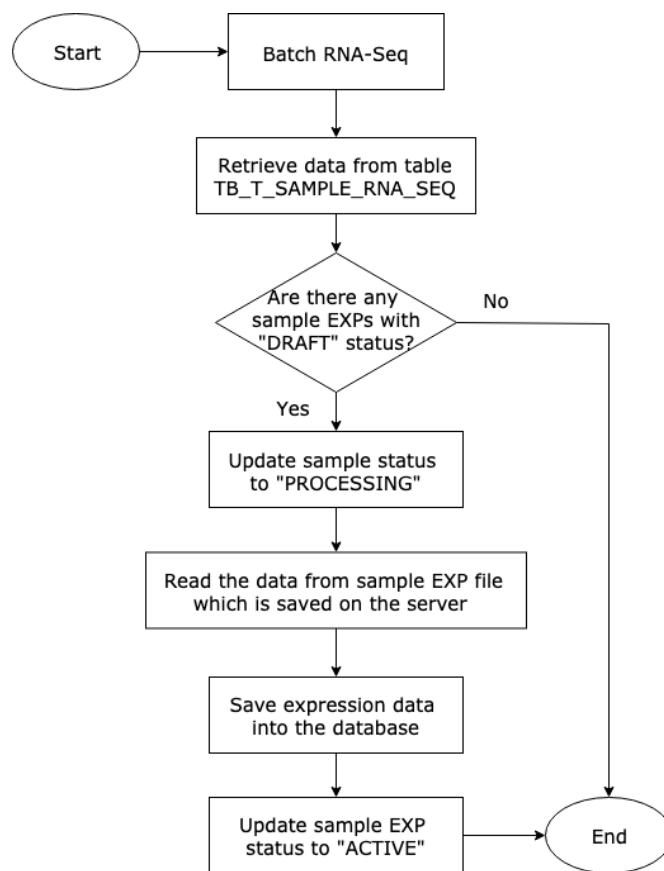
รูปที่ 34: แผนภาพแสดงลำดับขั้นตอนการทำงานของแบตช์สกัดข้อมูลจากไฟล์วีซีเอฟ

รูปที่ 35 แสดงขั้นตอนการตรวจสอบยีนที่เกี่ยวข้องกับมะเร็งที่สามารถรักษาได้ แบตช์จะทำงานทุก ๆ 10 นาที โดยจะดึงข้อมูลรายละเอียดไฟล์วีซีเอฟเฉพาะที่มีสถานะ “WAITING_VARIANTS_ANNOTATED” เท่านั้น และอัปเดตสถานะของข้อมูลที่อ่านขึ้นมาให้เป็นสถานะ “PROCESSING” หลังจากนั้นดึงข้อมูลจากตาราง TB_T_VARIANT_ANNOTATION ซึ่งเก็บตำแหน่งการแปรผันของดีเอ็นเอที่เกี่ยวข้องกับข้อมูลรายละเอียดไฟล์วีซีเอฟนั้น ๆ นำชื่อยีนในแต่ละเรคคอร์ดไปตรวจสอบกับข้อมูลในตาราง TB_M_ACTIONABLE_VARIANT ซึ่งเป็นข้อมูลจากอองโคเคบี ถ้าตรวจสอบยีนว่าเป็นยีนที่เกี่ยวข้องกับมะเร็งและรักษาได้ ฟิลด์ ONCOKB_ANNOTATED ของเรคคอร์ดดังกล่าว จะถูกอัปเดตค่าเป็น “Y” และเมื่อข้อมูลภายในตาราง TB_T_VARIANT_ANNOTATION ที่เกี่ยวข้องกับข้อมูลรายละเอียดไฟล์วีซีเอฟดังกล่าวได้รับการตรวจสอบครบถ้วน ระบบจะอัปเดตสถานะข้อมูลรายละเอียดไฟล์วีซีเอฟเป็น “ACTIVE”



รูปที่ 35: แผนภาพแสดงลำดับขั้นตอนการทำงานของแบตช์เพื่อตรวจสอบยืนยันที่เกี่ยวข้องกับมะเร็งที่สามารถรักษาได้จากข้อมูลองโคเคบี

รูปที่ 36 แสดงขั้นตอนการสกัดข้อมูลจากไฟล์อีเอ็กซ์พี แบตช์จะทำงานทุก ๆ 15 นาที จะดึงข้อมูลรายละเอียดไฟล์อีเอ็กซ์พีเฉพาะที่มีสถานะ “DRAFT” เท่านั้น และอัปเดตสถานะของข้อมูลที่อ่านขึ้นมาให้เป็นสถานะ “PROCESSING” โดยอ่านพารของไฟล์อีเอ็กซ์พีดังกล่าวจากฐานข้อมูล แล้วจึงเริ่มสกัดข้อมูลจากไฟล์อีเอ็กซ์พี และบันทึกข้อมูลทั้งหมดที่ได้ลงฐานข้อมูล จากนั้นอัปเดตสถานะของข้อมูลรายละเอียดไฟล์อีเอ็กซ์พีให้เป็น “ACTIVE”

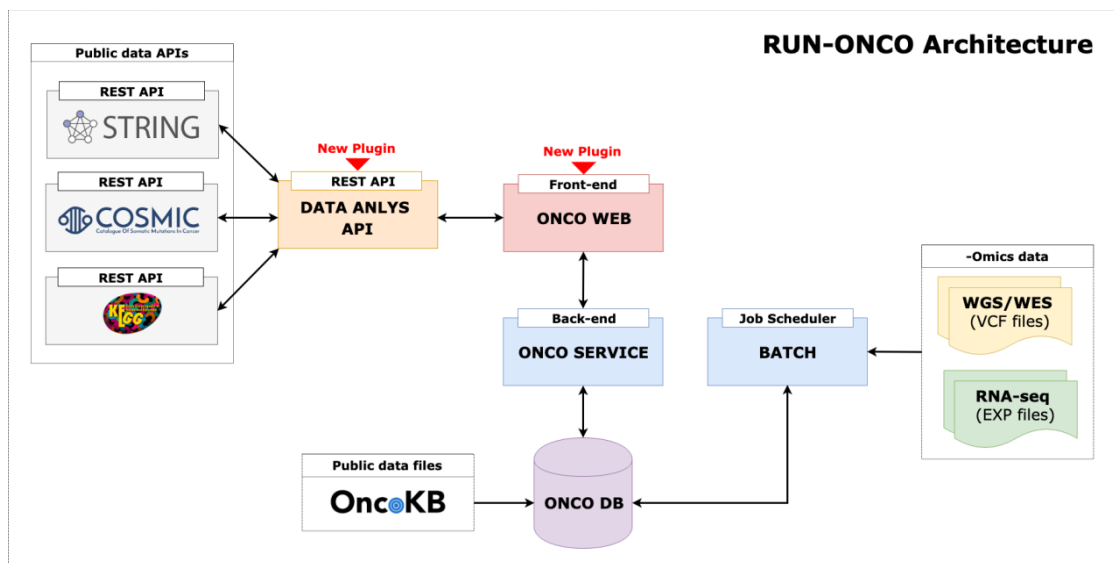


รูปที่ 36: แผนภาพแสดงลำดับขั้นตอนการทำงานของแบตช์สกัดข้อมูลจากไฟล์อีเอ็กซ์พี

3.3 สถาปัตยกรรมของระบบ

งานวิจัยนี้นำเสนอซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการรักษาโรคมะเร็งแบบแม่นยำ เนื่องจากการแพทย์แม่นยำเน้นการวินิจฉัยระดับพันธุกรรมเฉพาะบุคคล โดยระบบจะจัดเก็บข้อมูลทางคลินิกทุก ๆ ครั้งที่ผู้ป่วยเข้ารับการรักษา ร่วมกับการเก็บตัวอย่างชีววัตถุ ได้แก่ เลือด ชี้นเนื้อ มะเร็ง เพื่อนำมาตรวจในห้องปฏิบัติการด้วยเทคโนโลยีโอมิกส์ต่าง ๆ และนำไปวิเคราะห์ข้อมูลทางชีวสารสนเทศ (bioinformatics) ส่งผลให้ข้อมูลที่จัดเก็บในระบบค่อนข้างมีขนาดใหญ่และมีปริมาณมาก ซึ่งเติบโตไปพร้อม ๆ กับปริมาณผู้ป่วยที่เพิ่มขึ้น จากรูปที่ 37 ระบบประกอบด้วย 4 ส่วน ได้แก่ 1) เว็บองโค (ONCO WEB) ที่ใช้โต้ตอบหรือตอบสนองต่อผู้ใช้งาน นอกจากนี้ยังพัฒนาสามารถเพิ่มปลั๊กอินการวิเคราะห์และแสดงผลข้อมูลที่หลากหลาย 2) องโคเซอร์วิส (ONCO SERVICE) จะเป็นส่วนของการจัดการข้อมูล ไม่ว่าจะเป็นข้อมูลทางคลินิก ข้อมูลเนื้อเยื่อ การอัปโหลดไฟล์วีซีเอฟ การอัปโหลดไฟล์อีเอ็กซ์พี ซึ่งเป็นข้อมูลที่ถูกระมวลผลมาแล้ว 3) แบตช์ (BATCH) จะมีการทำงานโดยอัตโนมัติตามช่วงเวลา (job scheduler) โดยนำเข้าไฟล์ที่ถูกจัดเก็บไว้ก่อนหน้านี้ มาสกัดข้อมูลและนำข้อมูลที่ได้อัปโหลดลงในฐานข้อมูล เช่น ชุดไฟล์วีซีเอฟจำนวนมาก เป็นต้น 4) ดาต้าอานาไลซิสเอพีไอ

(Data Analysis API) เป็นส่วนที่อนุญาตให้นักพัฒนาหรือนักชีวสารสนเทศ สามารถเพิ่มอัลกอริทึมต่าง ๆ เพื่อการวิเคราะห์ข้อมูล เช่น อัลกอริทึมเค-มีนส์ (k-means cluster-ing) อัลกอริทึมการจัดกลุ่มทั้งแบบลำดับชั้น (hierarchical clustering) อัลกอริทึมที่ใช้คำนวณค่าทีเอ็มบี (Tumor Mutational Burden: TMB) ได้โดยง่าย พัฒนาด้วยภาษาไพธอน (python) และไลบรารีไซคิที-เลิร์น (scikit-learn) หรือภาษาอาร์ (R language) เป็นต้น

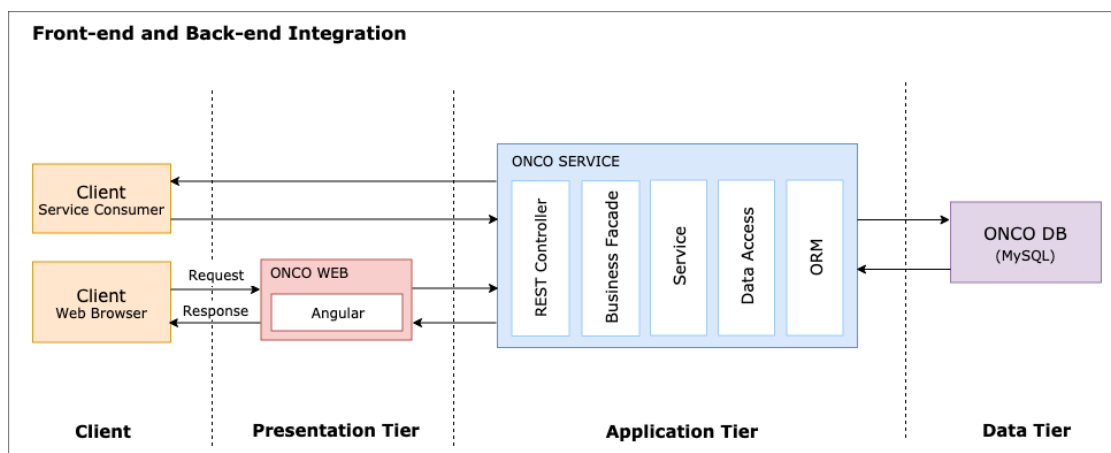


รูปที่ 37: ภาพรวมโครงสร้างซอฟต์แวร์แพลตฟอร์ม RUN-ONCO

3.4 การประยุกต์ใช้แบบรูปเพื่อการออกแบบระบบ

งานวิจัยนี้นำเสนอซอฟต์แวร์แพลตฟอร์มที่มีความยืดหยุ่นสูง เพื่อสนับสนุนการแพทย์แม่นยำ ซึ่งข้อมูลที่นำมาวิเคราะห์และการแสดงผลข้อมูลมีความหลากหลายเพื่อให้ง่ายต่อการพัฒนา และต่อยอดซอฟต์แวร์แพลตฟอร์มถูกออกแบบภายใต้แนวคิดการออกแบบแบบรูป (design pattern) หลัก ๆ ดังนี้ แบบรูปฟะซาด (facade pattern) แบบรูปดีเพนเดนซีอินเจกชัน (Dependency Injection: DI pattern) โมเดล-วิว-คอนโทรลเลอร์ (Model-View-Controller: MVC) ตัวส่งระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ (Object-Relational Mapping: ORM) และวัตถุการเข้าถึงข้อมูล (Data Access Object: DAO) จากหลักการออกแบบข้างต้น ทำให้ซอฟต์แวร์แพลตฟอร์มรองรับการทำงานที่ซับซ้อนและเพิ่มขยายระบบในอนาคต

จากรูปที่ 38 เป็นการออกแบบซอฟต์แวร์แพลตฟอร์มในส่วนขององโคเซอร์วิสตามสถาปัตยกรรม 3-เทียร์ ตารางที่ 6 แสดงคำอธิบายเลย์เออร์ต่าง ๆ ในองโคเซอร์วิส



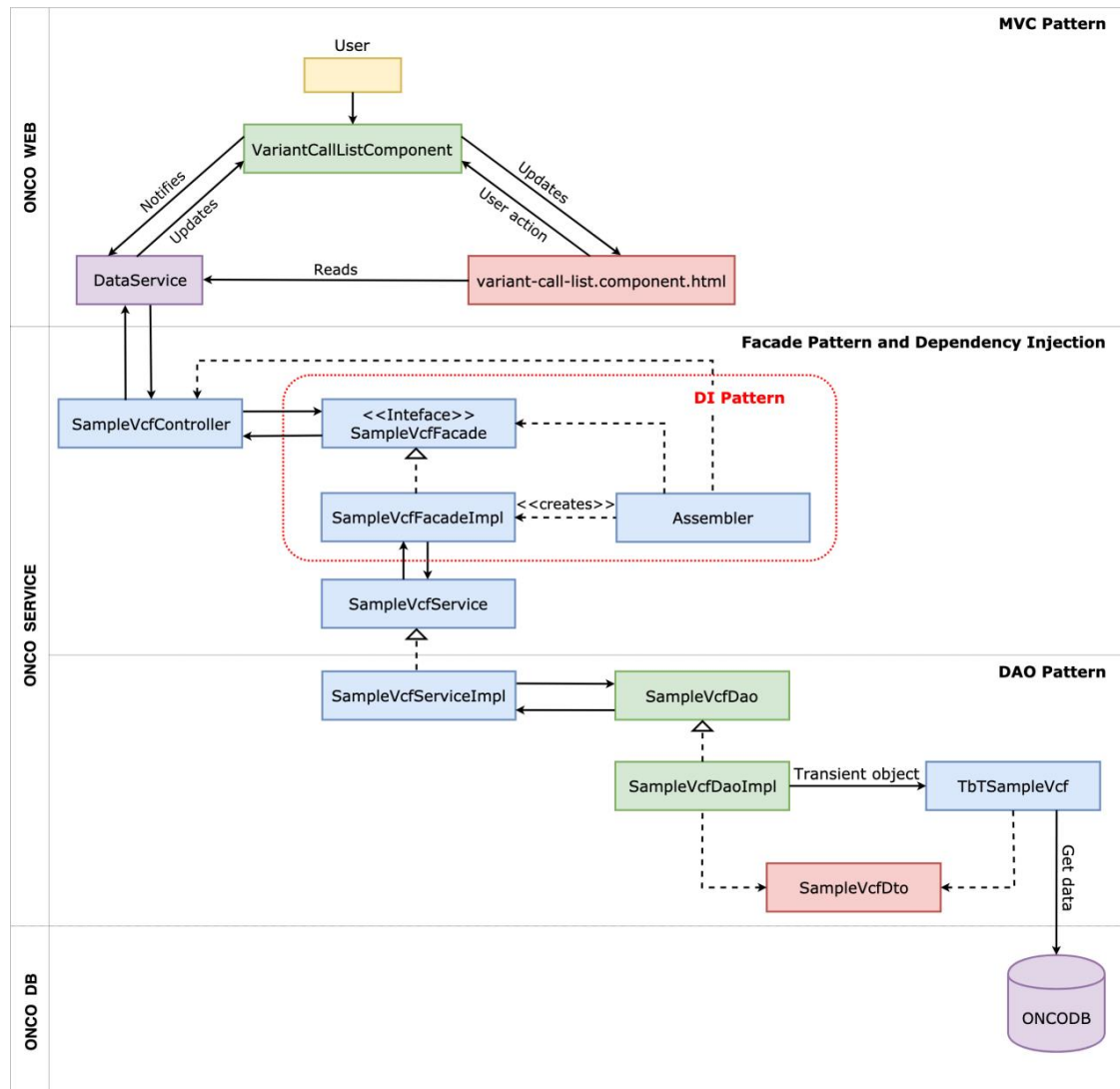
รูปที่ 38: โครงสร้างซอฟต์แวร์แพลตฟอร์ม RUN-ONCO ภายใต้แนวคิดสถาปัตยกรรม 3-เทียร์

ตารางที่ 6: เลเยอร์ต่าง ๆ ในองคเษอร์วิส

คอลัมน์	เลเยอร์	ความหมาย
1	REST Controller	ส่วนควบคุมในองคเษอร์วิส จะรับส่งข้อมูลระหว่างไคลเอนต์ โดยการรับส่งข้อมูลระหว่างส่วนควบคุมกับเว็บองคเษอร์วิสในรูปแบบเจสัน (JSON)
2	Business Facade	ส่วนจัดการข้อมูลและตรรกะที่ใช้ในการคำนวณกระบวนการทางธุรกิจ (business logic)
3	Service	ส่วนจัดการข้อมูลและตรรกะที่ใช้ในการเพิ่ม แก้ไข และลบข้อมูลในฐานข้อมูล
4	Data Access	คลาสสำหรับเก็บชุดคำสั่งหรือเมธอดที่ใช้ติดต่อกับฐานข้อมูล
5	ORM	เป็นตัวกลางที่เชื่อมความสัมพันธ์ระหว่างอ็อบเจกต์และตารางข้อมูลในฐานข้อมูล

จากรูปที่ 39 เมื่อผู้ใช้เรียกหน้าจอละเอียดข้อมูลไฟล์วีซีเอฟทั้งหมดของผู้ป่วยรายหนึ่ง ส่วนแสดงผล `variant-call-list.component.html` จะส่งคำสั่งไปยังคลาส `VariantCallListComponent` ซึ่งเป็นส่วนควบคุม และขอข้อมูลรายละเอียดของไฟล์วีซีเอฟผ่านคลาส `DataService` ไปยัง endpoint `/variant-call-list` ซึ่งคือคลาส `SampleVcfController` ในองคเษอร์วิส ส่วนนี้ได้ นำแบบรูปโมเดล-วิว-คอนโทรลเลอร์มาใช้งาน ในการเรียกข้อมูลรายละเอียดของไฟล์วีซีเอฟ คลาส `SampleVcfController` จะเรียกคลาส `SampleVcfFacade` ซึ่งประกอบไปด้วยคลาส `SampleVcfService` โดยนำแบบรูปพะชาดและตีเพนเดนซีอินเจกชันมาประยุกต์ใช้ ส่วนการเชื่อมต่อกับฐาน

ข้อมูล คลาส *SampleVcfService* จะเรียกคลาส *SampleVcfDao* ซึ่งมีการทำงานแบบรูปวัตถุการเข้าถึงข้อมูล โดยแปลงข้อมูลที่ได้รับจากฐานข้อมูลให้อยู่ในรูปโมเดลเชิงวัตถุและเชิงสัมพันธ์



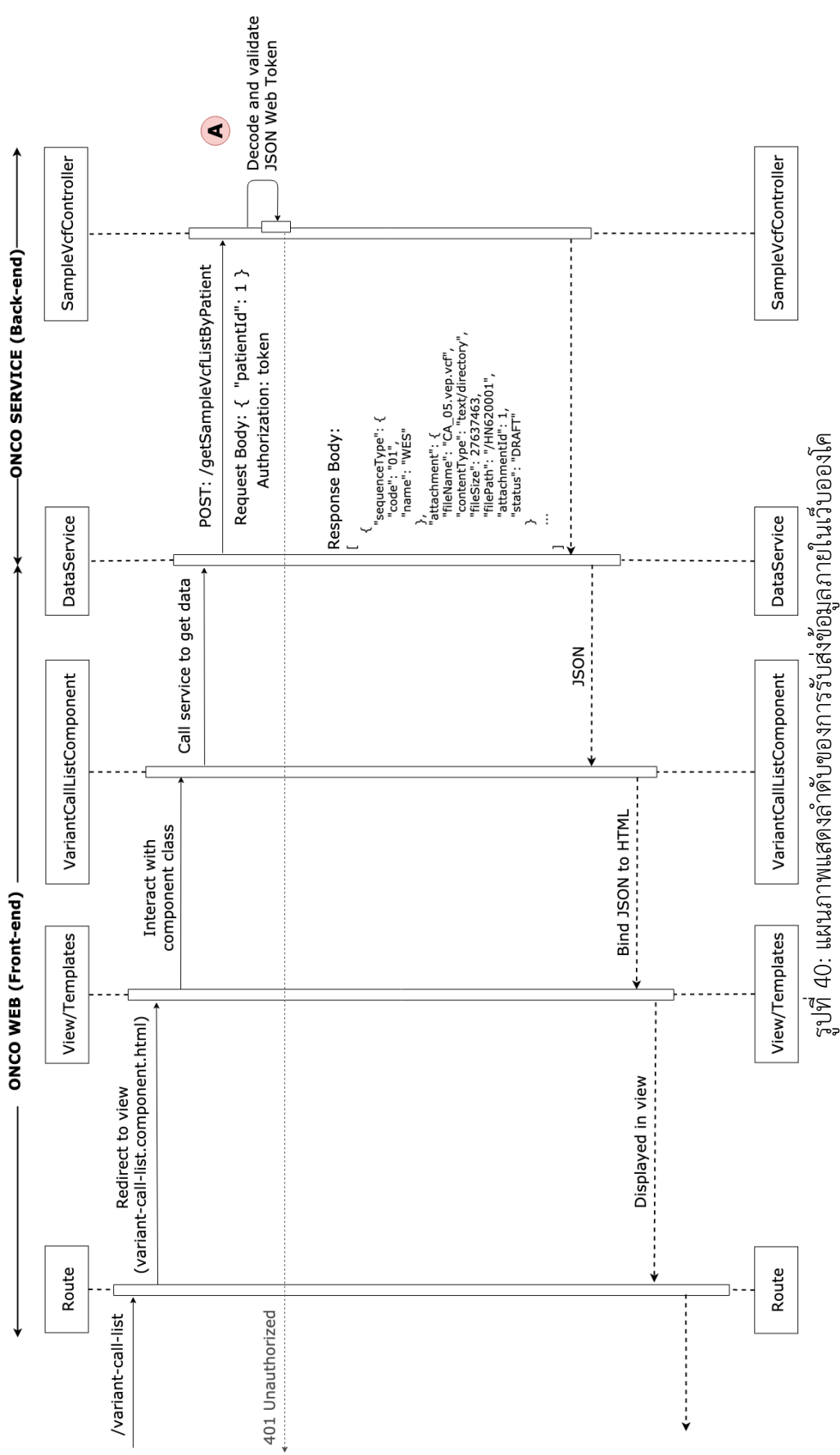
รูปที่ 39: ตัวอย่างการประยุกต์ใช้แบบรูปในแพลตฟอร์ม

3.5 การวิเคราะห์และออกแบบเซอร์วิสและรูปแบบข้อมูลที่ใช้รับส่งภายในระบบ

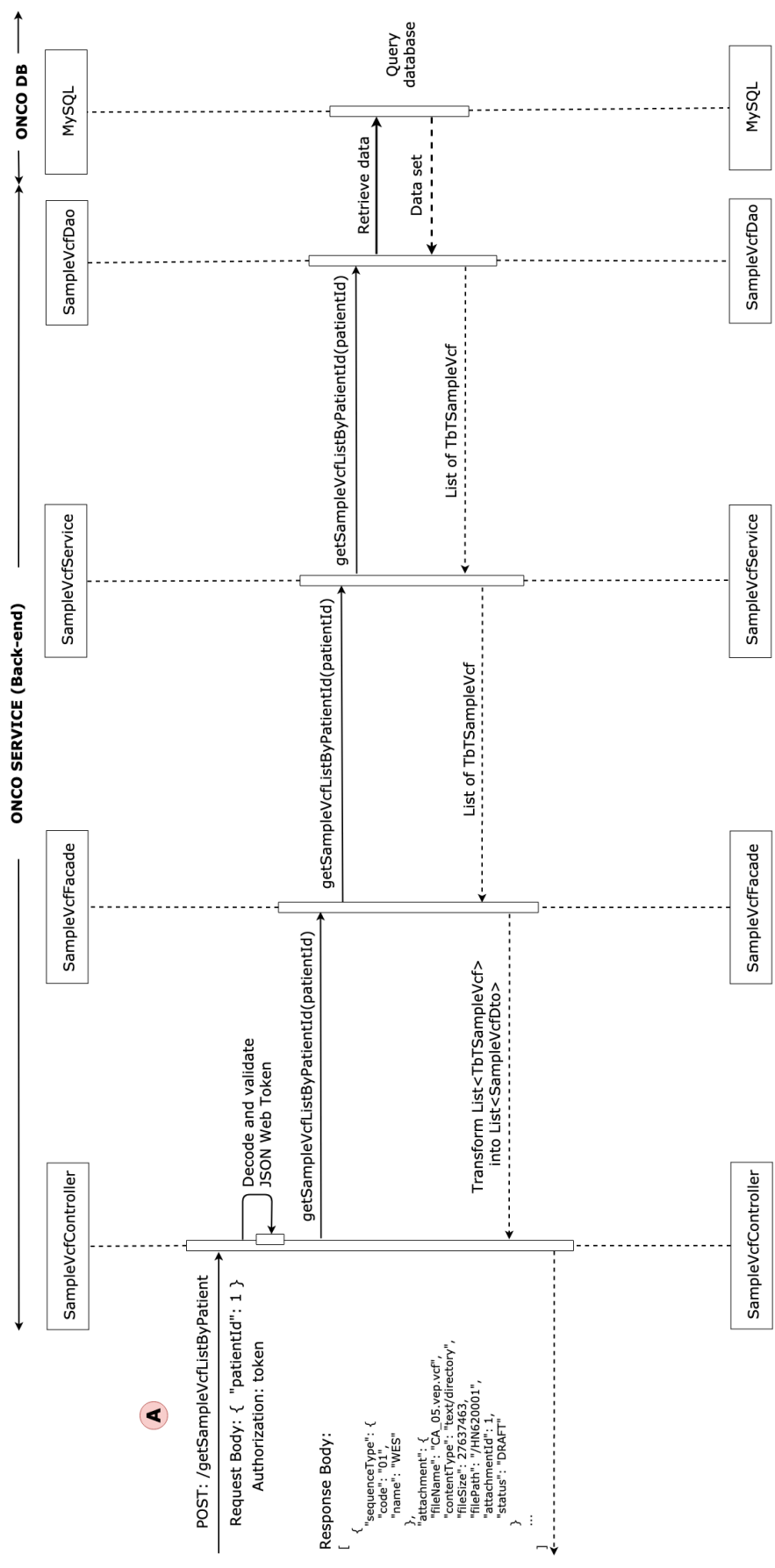
รูปที่ 40 แสดงแผนภาพแสดงลำดับของการรับส่งข้อมูลภายในเว็บองโค เมื่อผู้ใช้เรียกหน้าจอแสดงรายละเอียดข้อมูลไฟล์วีซีเอฟทั้งหมดของผู้ป่วยรายหนึ่ง ส่วนแสดงผล *variant-call-list.component.html* จะส่งคำสั่งไปยังคลาส *VariantCallListComponent* ผ่านคลาส *DataService* เพื่อร้องขอข้อมูลรายละเอียดของไฟล์วีซีเอฟไปยัง endpoint */getSampleVcfListByPatient* ซึ่งก็คือคลาส *SampleVcfController* ในองโคเซอร์วิส จากรูปที่ 41 คลาส *SampleVcfController* จะ

เรียกใช้งานคลาส *SampleVcfFacade* เพื่อตรวจสอบความถูกต้องของข้อมูลที่ร้องขอ ก่อนส่งต่อให้คลาส *SampleVcfService* เพื่อเปิดการเชื่อมต่อไปยังฐานข้อมูล และส่งคำสั่ง SELECT เพื่อดึงข้อมูลรายละเอียดของไฟล์วีซีเอฟจากฐานข้อมูลด้วยคลาส *SampleVcfDao* แล้วจึงแปลงข้อมูลที่ได้รับให้อยู่ในรูปแบบโมเดลเชิงวัตถุและเชิงสัมพันธ์



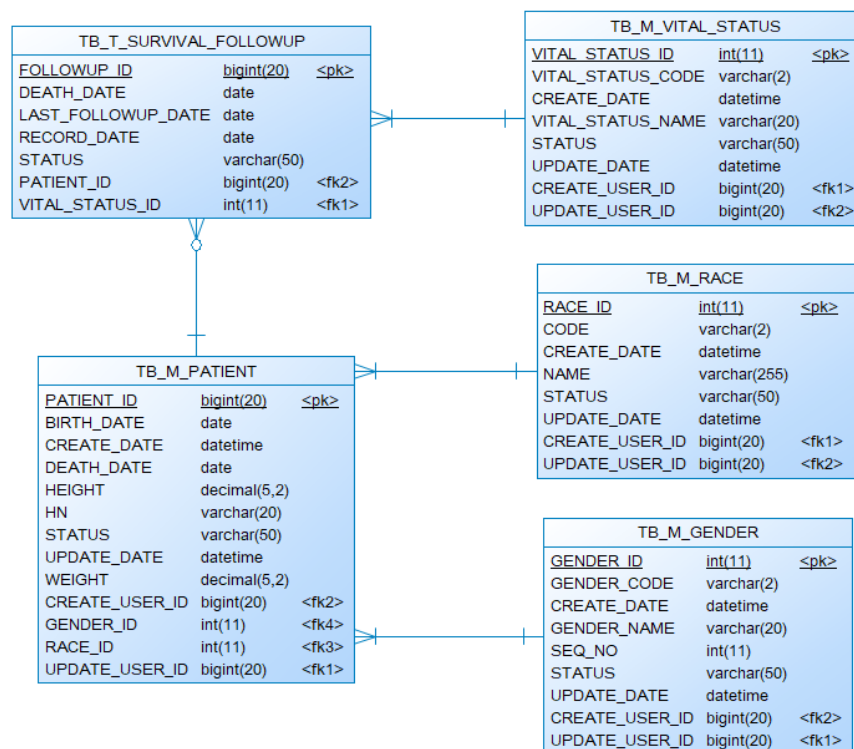


รูปที่ 40: แผนภาพแสดงลำดับของการรับส่งข้อมูลภายในเว็บของโค



รูปที่ 41: แผนภาพแสดงลำดับของการรับส่งข้อมูลภายในองเคเซอร์วิส

รูปที่ 44 แสดงโครงสร้างความสัมพันธ์ตารางข้อมูลประวัติผู้ป่วย ซึ่งจัดเก็บข้อมูลเลขที่เวชระเบียน ข้อมูลพื้นฐานของผู้ป่วย (demographic information) เช่น เพศ วันเกิด เชื้อชาติน้ำหนัก ส่วนสูง ในตารางข้อมูล TB_M_PATIENT และประวัติการติดตามการมีชีวิตของผู้ป่วย (survival follow-up) ไว้ในตาราง TB_T_SURVIVAL_FOLLOWUP นอกจากนี้จะเป็นตารางข้อมูลพื้นฐาน ได้แก่ เชื้อชาติ (TB_M_RACE) และเพศ (TB_M_GENDER)



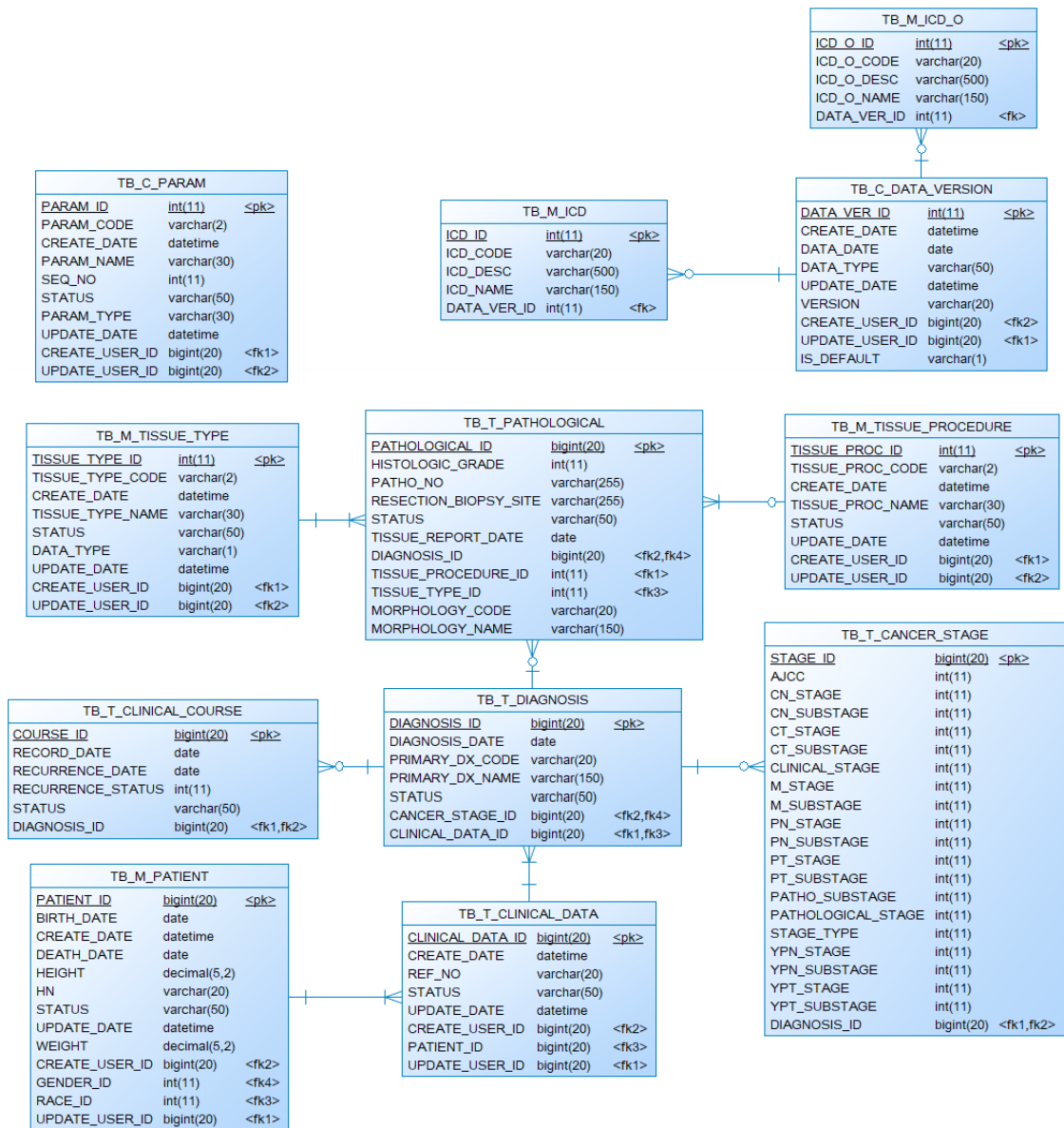
รูปที่ 44: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลผู้ป่วย

จากรูปที่ 45 แสดงโครงสร้างความสัมพันธ์ตารางข้อมูลทางคลินิก ซึ่งจัดเก็บข้อมูลในระหว่างการติดตามการรักษาได้ ซึ่งช่วยแพทย์ผู้เชี่ยวชาญสามารถออกแบบแผนการรักษาที่เหมาะสมปลอดภัย และมีความแม่นยำต่อผู้ป่วย ตารางข้อมูลหลักคือ TB_T_CLINICAL_DATA โดยจัดเก็บข้อมูลทางคลินิกออกเป็น 4 ประเภทด้วยกัน 1) การวินิจฉัยโรค ประกอบด้วย วันที่วินิจฉัยโรค (diagnosis date) รหัสสำหรับการวินิจฉัยหลัก (primary diagnosis) ประกอบด้วย รหัสไอซีดี-10 เช่น C18 ซึ่งใช้ในการจำแนกและระบุรายละเอียดของโรค malignant neoplasm of colon โดยจัดเก็บในตารางข้อมูล TB_T_DIAGNOSIS 2) การวินิจฉัยทางพยาธิวิทยา ถูกออกแบบให้จัดเก็บข้อมูลดังต่อไปนี้ รหัสพยาธิ (pathology number) วันที่ส่งตรวจชิ้นเนื้อ (date of tissue) วันที่รายงานทางพยาธิวิทยาของชิ้นเนื้อ (date of pathology report)

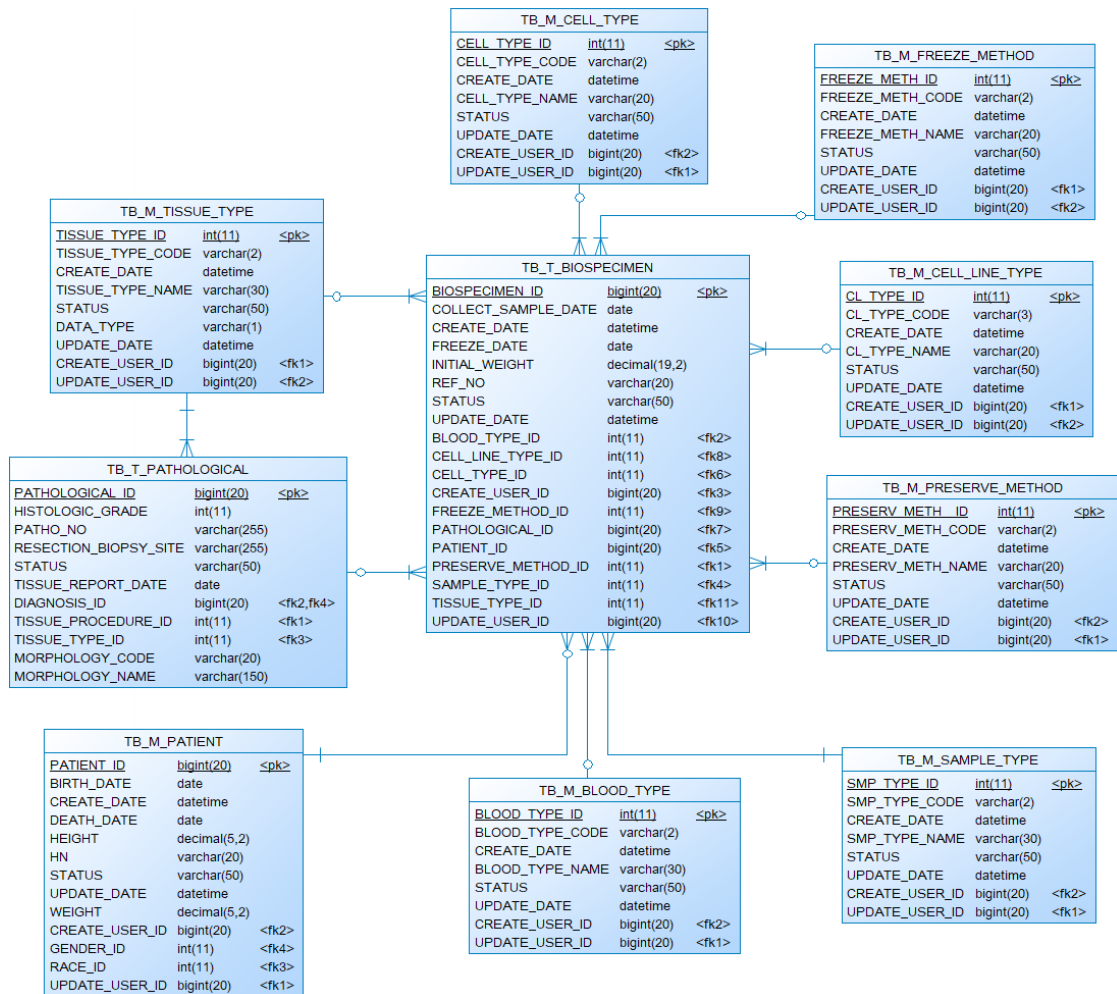
ตำแหน่งของการผ่าตัดเอาชิ้นเนื้อ (site of resection or biopsy) ลักษณะของเนื้องอกและคุณสมบัติทางชีววิทยาของเซลล์มะเร็ง (morphology) ประกอบด้วย รหัสไอซีดี-โอ ซึ่งใช้ในการอธิบายคุณลักษณะของชิ้นเนื้อ โดยจัดเก็บในตารางข้อมูล TB_T_PATHOLOGICAL 3) ระยะของโรคมะเร็ง ได้นำระบบ TNM ของ AJCC มาใช้งาน เพื่อจำแนกระยะของโรคแบบมาตรฐานสากลทั่วโลก ได้แก่ ประเภทระยะของโรคมะเร็ง (type of stage) ระยะของโรคมะเร็งตามลักษณะทางคลินิก (cT Stage) และระยะของโรคมะเร็งตามลักษณะทางพยาธิวิทยา (pT Stage) เป็นต้น นอกจากนี้ยังมีข้อมูลอื่น ๆ เช่น ระยะของโรคมะเร็งตามลักษณะทางคลินิก (clinical stage) โรคมะเร็งตามลักษณะทางพยาธิวิทยา (pathological stage) ซึ่ง 2 필ด์นี้ไม่ได้อยู่ระบบ TNM ค่าที่เป็นไปได้คือ I, II, III และ IV ข้อมูลในส่วนนี้จะถูกจัดเก็บไว้ในตารางข้อมูล TB_T_CANCER_STAGE และ 4) การดำเนินโรค เพื่อเฝ้าติดตามโรคมะเร็งย้อนกลับเป็นซ้ำ ข้อมูลที่ถูกจัดเก็บ ได้แก่ สถานะย้อนกลับเป็นซ้ำ (recurrence status) วันที่ย้อนกลับเป็นซ้ำ (date of recurrence) ซึ่งจะเก็บไว้ในตารางข้อมูล TB_T_CLINICAL_COURSE

รูปที่ 46 แสดงโครงสร้างความสัมพันธ์ตารางข้อมูลชีววัตถุ ประกอบไปด้วย ประเภทของตัวอย่างเก็บตรวจ (sample type) ได้แก่ เลือด (blood) เนื้อเยื่อ (tissue) และ เซลล์ (cell) ประเภทของเนื้อเยื่อ (tissue type) เช่น colon fresh tissue, lung fresh tissue, overy fresh tissue รหัสพาโร วิธีการแช่เยือกแข็ง (freezing method) เช่น -20° C, -20° C, liquid nitrogen วิธีเก็บรักษาตัวอย่างเก็บตรวจ (preservation method) เช่น RNAlater, snap freezing โดยจะถูกจัดเก็บไว้ในตารางข้อมูล TB_T_BIOSPECIMEN นอกจากนี้จะเป็นข้อมูลพื้นฐาน เช่น TB_M_SAMPLE_TYPE ซึ่งเก็บข้อมูลประเภทของตัวอย่างเก็บตรวจ เป็นต้น

ในกรณีชิ้นเนื้อที่เกิดจากการผ่าตัดโดยศัลยแพทย์จะต้องนำส่งตรวจโดยพยาธิแพทย์ เพราะฉะนั้นชิ้นเนื้อดังกล่าวจะมีรหัสพาโรเสมอ โดยจะเกิดขึ้นเฉพาะกรณีที่เป็นเนื้อเยื่อเท่านั้น โดยระบบจะเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลชีววัตถุกับการวินิจฉัยพยาธิวิทยาได้ จากรูปที่ 46 ตารางข้อมูลจะเชื่อมโยงกับตารางข้อมูล TB_T_PATHOLOGICAL ในทางกลับกัน ถ้าข้อมูลชีววัตถุไม่ได้ผ่านกระบวนการพยาธิวิทยา มา ข้อมูลจะเชื่อมโยงกับข้อมูลประวัติผู้ป่วยโดยตรง ตัวอย่างเช่น เลือด หมายเหตุเนื่องจากการตัดชิ้นเนื้อเพื่อการวิเคราะห์ทำได้หลายครั้ง ระบบจึงรองรับการเชื่อมโยงความสัมพันธ์ระหว่างชิ้นเนื้อกลุ่มดังกล่าวเข้ากับรหัสพาโรเดียวกันได้



รูปที่ 45: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลทางคลินิก



รูปที่ 46: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลรายละเอียดชีวิตวัตถุ

3.6.2 ศึกษาโครงสร้างข้อมูลภายในไฟล์วีซีเอฟ

ศึกษาข้อมูลในไฟล์วีซีเอฟ (รูปที่ 47) โดยนำไลบรารีวีซีเอฟพาเซอร์ (vcf-parser) ของฟาร์มเคจีบี (PharmKGB) [51] มาดัดแปลง เพื่อช่วยในการสกัดข้อมูลจากไฟล์วีซีเอฟ จากรูปที่ 48 แสดงโครงสร้างความสัมพันธ์ข้อมูลการแปรผันของดีเอ็นเอประกอบไปด้วย ข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอ (variant calls) ซึ่งได้จาก WGS/WES ระบบจะทำการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลโอมิกส์และข้อมูลรายละเอียดชีวิตวัตถุ สามารถนำข้อมูลดังกล่าวไปวิเคราะห์ร่วมกับข้อมูลอื่น ๆ เช่น ยีนก่อมะเร็ง (oncogene) ยีนที่เกี่ยวข้องกับมะเร็งที่มีผลต่อการรักษา (actionable variant) โดยองโคเคบี เป็นต้น

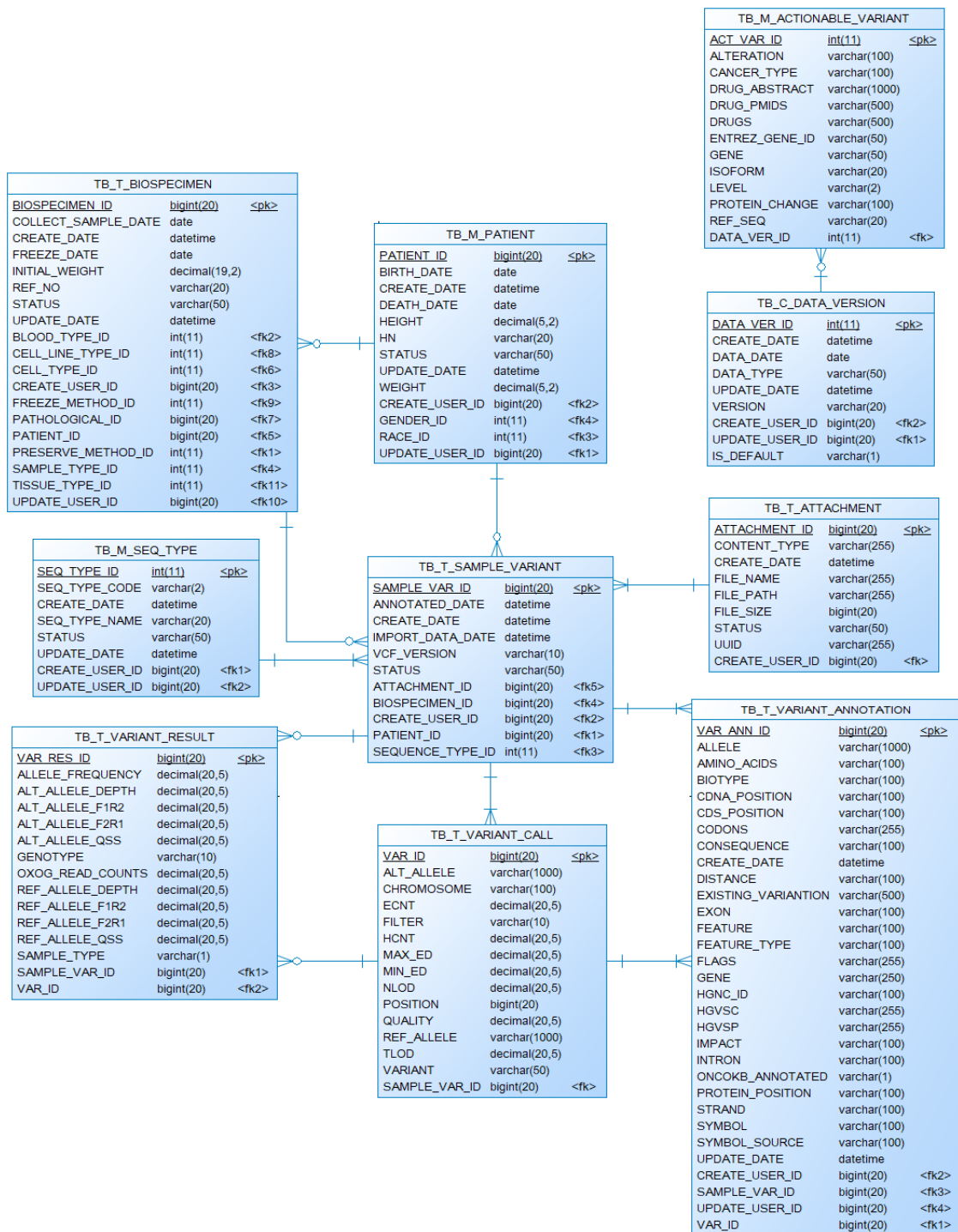
3.6.3 ศึกษาโครงสร้างข้อมูลภายในไฟล์อีเอ็กซ์พี

เทคโนโลยีอาร์เอ็นเอซีคถูกนำมาใช้ในการวัดระดับการแสดงออกของยีนในมะเร็งชนิดต่าง ๆ รูปที่ 49 แสดงตัวอย่างข้อมูลอาร์เอ็นเอซีค เนื่องจากการวิเคราะห์ระดับการแสดงออกของยีนรวมทั้งการแสดงผลข้อมูลการแสดงออกสามารถทำได้หลากหลายวิธีและรูปแบบ ระบบจึงถูกออกแบบให้การวิเคราะห์ข้อมูล หรือการแสดงผลข้อมูลโอมิกส์สามารถทำผ่านอองโคเซอร์วิส ในกรณีของข้อมูลอาร์เอ็นเอซีคนี้ ตัวแปรหลักที่สามารถดึงผ่านอองโคเซอร์วิสได้ประกอบด้วย รหัสยีน (Gene ID) และค่าการแสดงออกของยีนที่ปรับมาตรฐานด้วยค่าที่พีเอ็ม (Transcripts Per Million: TPM) เป็นต้น

รูปที่ 50 แสดงโครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลระดับการแสดงออกของยีนประกอบไปด้วย ข้อมูลระดับการแสดงออกของยีน ซึ่งได้จากไฟล์อีเอ็กซ์พี ระบบจะทำการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลโอมิกส์และข้อมูลรายละเอียดชีววัตถุ สามารถนำข้อมูลดังกล่าวไปวิเคราะห์ร่วมกับข้อมูลอื่น ๆ เช่น ข้อมูลการมีปฏิสัมพันธ์ระหว่างโปรตีน (protein-protein interaction) จากฐานข้อมูลสตริง หรือข้อมูลพาธเวย์ (pathway) จากฐานข้อมูล KEGG PATHWAY เป็นต้น

A	B	C	D	E	F	G	H	I	J	K
#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	TUMOR	NORMAL
chr1	13838	C	T	.	.	germline_risk	ECNT=1;HCNT=8;MAX_ED=.;MIN_ED=.;NLOD=0.556;TLOD=5.10;CSQ=T id	GT:AD:AF:AL	0/1:9,3:0.25	0/0:3,0:0.00
chr1	129285	G	A	.	.	germline_risk	ECNT=1;HCNT=1;MAX_ED=.;MIN_ED=.;NLOD=1.81;TLOD=4.29;CSQ=A up	GT:AD:AF:AL	0/1:13,3:0.2	0/0:10,0:0.00
chr1	827476	G	A	.	.	germline_risk	ECNT=1;HCNT=1;MAX_ED=.;MIN_ED=.;NLOD=0.903;TLOD=9.55;CSQ=A u	GT:AD:AF:AL	0/1:3,4:0.50	0/0:4,0:0.00
chr1	925965	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=2;MAX_ED=.;MIN_ED=.;NLOD=20.47;TLOD=4.92;CSQ=A u	GT:AD:AF:AL	0/1:104,2:0	0/0:97,0:0.00
chr1	926806	A	G	.	.	germline_risk	ECNT=1;HCNT=1;MAX_ED=.;MIN_ED=.;NLOD=0.903;TLOD=12.52;CSQ=G	GT:AD:AF:AL	0/1:1,3:0.75	0/0:6,0:0.00
chr1	931513	T	C	.	.	germline_risk	ECNT=1;HCNT=2;MAX_ED=.;MIN_ED=.;NLOD=0.903;TLOD=4.50;CSQ=C ir	GT:AD:AF:AL	0/1:0,2:1.00	0/0:4,0:0.00
chr1	962312	A	T	.	.	t_jod_fstar	ECNT=1;HCNT=1;MAX_ED=.;MIN_ED=.;NLOD=59.41;TLOD=4.04;CSQ=T u	GT:AD:AF:AL	0/1:269,2:8	0/0:271,0:0.00
chr1	1049516	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=13;MAX_ED=.;MIN_ED=.;NLOD=40.94;TLOD=4.26;CSQ=A	GT:AD:AF:AL	0/1:210,3:0	0/0:182,0:0.00
chr1	1050900	C	T	.	.	t_jod_fstar	ECNT=1;HCNT=2;MAX_ED=.;MIN_ED=.;NLOD=11.93;TLOD=5.21;CSQ=T ir	GT:AD:AF:AL	0/1:69,2:0.0	0/0:63,1:0.02
chr1	1087321	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=12;MAX_ED=.;MIN_ED=.;NLOD=26.44;TLOD=4.34;CSQ=A s	GT:AD:AF:AL	0/1:131,2:0	0/0:115,0:0.00
chr1	1180089	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=2;MAX_ED=.;MIN_ED=.;NLOD=20.71;TLOD=4.83;CSQ=A s	GT:AD:AF:AL	0/1:105,2:0	0/0:94,0:0.00
chr1	1205911	C	T	.	.	germline_risk	ECNT=1;HCNT=1;MAX_ED=.;MIN_ED=.;NLOD=1.20;TLOD=7.47;CSQ=T int	GT:AD:AF:AL	0/1:2,2:0.50	0/0:4,0:0.00
chr1	1242666	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=9;MAX_ED=.;MIN_ED=.;NLOD=42.35;TLOD=4.30;CSQ=A ir	GT:AD:AF:AL	0/1:210,2:0	0/0:207,0:0.00
chr1	1281504	A	G	.	.	t_jod_fstar	ECNT=1;HCNT=8;MAX_ED=.;MIN_ED=.;NLOD=18.03;TLOD=4.71;CSQ=G u	GT:AD:AF:AL	0/1:121,4:0	0/0:106,1:0.00
chr1	1299915	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=27;MAX_ED=.;MIN_ED=.;NLOD=12.94;TLOD=4.34;CSQ=A	GT:AD:AF:AL	0/1:65,3:0.0	0/0:56,0:0.00
chr1	1304112	C	A	.	.	t_jod_fstar	ECNT=1;HCNT=6;MAX_ED=.;MIN_ED=.;NLOD=56.89;TLOD=4.30;CSQ=A n	GT:AD:AF:AL	0/1:208,2:0	0/0:278,0:0.00
chr1	1313551	G	A	.	.	t_jod_fstar	ECNT=1;HCNT=9;MAX_ED=.;MIN_ED=.;NLOD=39.32;TLOD=4.35;CSQ=A n	GT:AD:AF:AL	0/1:193,2:0	0/0:199,1:7.1

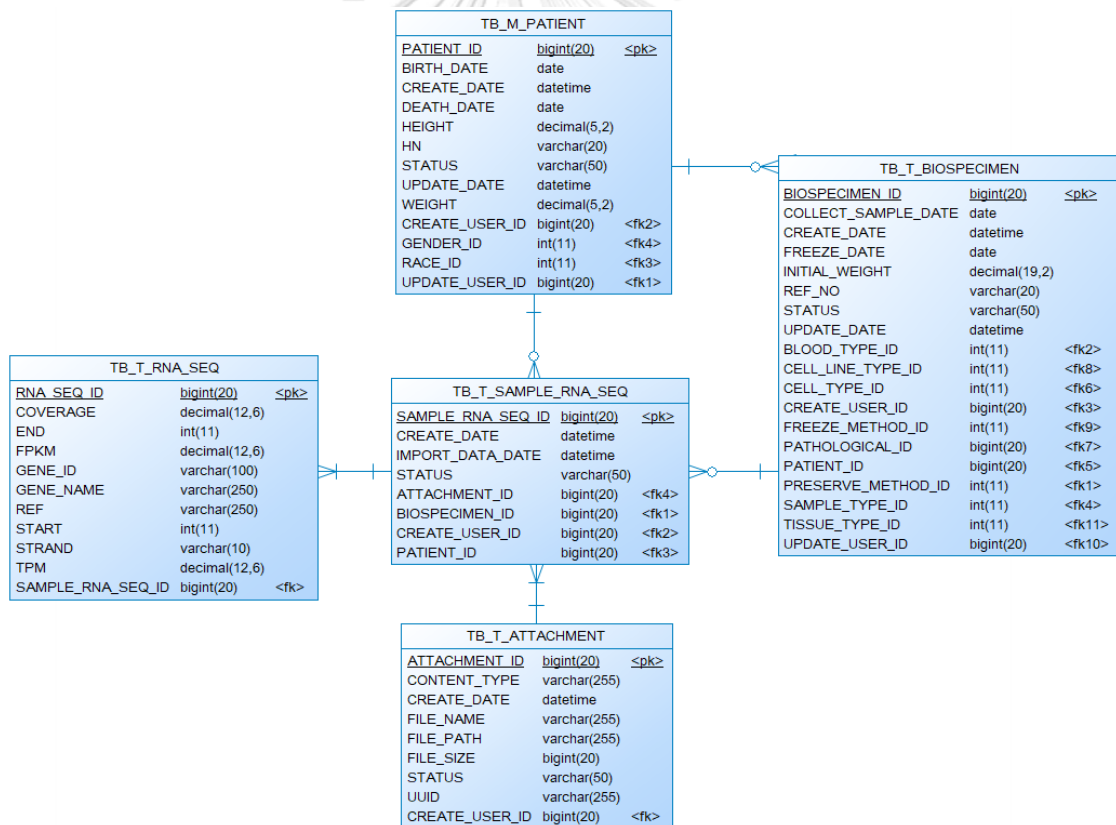
รูปที่ 47: ตัวอย่างไฟล์วีซีเอฟที่นำมาศึกษา



รูปที่ 48: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอ

A	B	C	D	E	F	G	H	I
Gene ID	Gene Name	Reference	Strand	Start	End	Coverage	FPKM	TPM
ENSG00000223972	DDX11L1		1 +	11869	14412	0.262214	0.043015	0.110758
ENSG00000227232	WASH7P		1 -	14363	29806	45.029705	7.610724	19.596645
ENSG00000243485	MIR1302-10		1 +	29554	31109	0.05191	0.02625	0.06759
SAMPLE1	-		1 +	30366	30503	0.25	0.041011	0.105599
SAMPLE1	-		1 -	14363	29370	19.56953	8.958896	23.068014
ENSG00000223181	RNU6-1199P		1 -	693613	693716	0	0	0
ENSG00000240618	RP11-206L10.5		1 -	694412	700305	16.225164	2.661656	6.853425
ENSG00000229905	RP11-206L10.4		1 +	696291	697369	0	0	0
ENSG00000228327	RP11-206L10.2		1 -	700237	714006	15.90947	2.609868	6.720078
SAMPLE1	-		1 -	694412	714006	13.178737	2.677485	6.894183
ENSG00000224956	RP11-206L10.1		1 -	661611	663527	19.332743	3.171439	8.16605
ENSG00000235373	RP11-206L10.3		1 -	677193	685396	0.675205	0.110764	0.285203
SAMPLE1	-		1 -	661611	685396	7.977602	1.783345	4.591886
ENSG00000237491	RP11-206L10.9		1 +	714150	745440	1.123616	0.39832	1.025622
ENSG00000230092	RP11-206L10.8		1 -	736259	745541	6.558317	1.075859	2.7702
ENSG00000269831	AL669831.1		1 -	738532	739137	0.263566	0.043237	0.111329
ENSG00000240453	RP11-206L10.10		1 -	745489	753092	1.435323	0.235457	0.606273
ENSG00000177757	FAM87B		1 +	752751	755214	0.617284	0.101262	0.260738
ENSG00000225880	LINC00115		1 -	761586	762902	2.181473	0.362806	0.934178

รูปที่ 49: ตัวอย่างไฟล์ข้อมูลที่เกี่ยวข้องการแสดงผลของยีน



รูปที่ 50: โครงสร้างความสัมพันธ์ตารางจัดเก็บข้อมูลระดับการแสดงผลของยีนจากเทคโนโลยีอาร์เอ็นเอซีค

3.6.1 ศึกษาโครงสร้างข้อมูลของโคเคบี (OncoKB)

ศึกษาข้อมูลไฟล์ของโคเคบี (รูปที่ 51) ซึ่งเป็นยีนที่เกี่ยวข้องกับมะเร็งที่สามารถรักษาได้ หรือมีแนวทางในการรักษา (clinically actionable) เพื่อช่วยในการวินิจฉัยโรคมะเร็ง เลือกใช้ยา และติดตามผลการรักษา ซึ่งสามารถเชื่อมโยงข้อมูลระหว่างฟิลด์ Symbol เช่น ฟิลด์ Gene ใน ไฟล์ของโคเคบี ซึ่งสามารถเชื่อมโยงกับ *EGFR* ในไฟล์จีโนม

Isoform	RefSeq	Entrez Gene	Gene	Alteration	Protein Change	Cancer Type	Level	Drugs(s)	PMDs for drug	Abstracts for drug
ENST00000275493	NM_005228.	1956	EGFR	Exon 20 inse	762_823ins	Non-Small Ce	R1	Afatinib, Erlot	23371856, 24065731, 19536777, 24353160,	
ENST00000275493	NM_005228.	1956	EGFR	T790M	T790M	Non-Small Ce	R1	Afatinib, Erlot	24478319, 2605 Yang et al. Abstract# O03.C	
ENST00000256078	NM_033360.	3845	KRAS	Oncogenic M	Oncogenic Muta	Colorectal Car	R1	Panitumumal	21228335, 20619739, 18316791, 20921465,	
ENST00000369535	NM_002524.	4893	NRAS	Oncogenic M	Oncogenic Muta	Colorectal Car	R1	Panitumumal	20619739, 24024844, 25110411, 24024839	
ENST00000257290	NM_006206.	5156	PDGFRA	D842V	D842V	Gastrointestir	R1	Imatinib	25905001, 17087936, 12949711, 24963404,	
ENST00000318560	NM_005157.	25	ABL1	BCR-ABL1 Fu	BCR-ABL1 Fusioi	Acute Lymphc		1 Dasatinib, Im	17496201, 11287973, 24441288, 20131302,	
ENST00000318560	NM_005157.	25	ABL1	BCR-ABL1 Fu	BCR-ABL1 Fusioi	Chronic Myelc		1 Nilotinib, Das	28095277, 28218239, 27217448, 11287973,	
ENST00000389048	NM_004304.	238	ALK	Fusions	Fusions	Non-Small Ce		1 Crizotinib, Bri	28586279, 29596029, 29668860, 28475456,	
ENST00000288602	NM_004333.	673	BRAF	V600E	V600E	Anaplastic Th		1 Dabrafenib +	29072975	
ENST00000288602	NM_004333.	673	BRAF	V600E	V600E	Melanoma		1 Vemurafenib,	29573941, 22735384, 28961848, 25265494,	
ENST00000288602	NM_004333.	673	BRAF	V600K	V600K	Melanoma		1 Dabrafenib +	29573941, 28891408, 27480103, 22663011,	
ENST00000288602	NM_004333.	673	BRAF	V600	V600	Non-Langerha		1 Vemurafenib	26287849, 29188284	
ENST00000288602	NM_004333.	673	BRAF	V600E	V600E	Non-Small Ce		1 Dabrafenib +	27283860, 28919011	
ENST00000357654	NM_007294.	672	BRCA1	Oncogenic M	Oncogenic Muta	Ovarian Canc		1 Rucaparib, Ni	27717299, 28882436, 28916367	
ENST00000380152	NM_000059.	675	BRCA2	Oncogenic M	Oncogenic Muta	Ovarian Canc		1 Rucaparib, Ni	27717299, 28882436, 28916367	
ENST00000275493	NM_005228.	1956	EGFR	A750P	A750P	Non-Small Ce		1 Afatinib, Erlot	28336552, 19625781, 25179728, 15737014	
ENST00000275493	NM_005228.	1956	EGFR	A763_Y764ir	A763_Y764insFC	Non-Small Ce		1 Erlotinib, Afal	23371856, 24353160, 23969006	
ENST00000275493	NM_005228.	1956	EGFR	E709K	E709K	Non-Small Ce		1 Afatinib, Erlot	28336552, 23566546, 18000506, 25179728,	
ENST00000275493	NM_005228.	1956	EGFR	E709_T710de	E709_T710delin:	Non-Small Ce		1 Erlotinib, Afal	21531810, 28336552, 26206867	
ENST00000275493	NM_005228.	1956	EGFR	Exon 19 dele	729_761del	Non-Small Ce		1 Erlotinib, Afal	22452895, 25589191, 23816960, 22285168,	
ENST00000275493	NM_005228.	1956	EGFR	Exon 19 dele	729_761indel	Non-Small Ce		1 Erlotinib, Afal	22452895, 25589191, 23816960, 22285168,	
ENST00000275493	NM_005228.	1956	EGFR	Exon 19 inse	729_761ins	Non-Small Ce		1 Afatinib, Erlot	28336552, 22190593	
ENST00000275493	NM_005228.	1956	EGFR	G719	G719	Non-Small Ce		1 Erlotinib, Afal	21531810, 28336552, 15118073, 26051236,	
ENST00000275493	NM_005228.	1956	EGFR	Kinase Doma	Kinase Domain C	Non-Small Ce		1 Erlotinib, Afal	28336552, 26286086	
ENST00000275493	NM_005228.	1956	EGFR	L747P	L747P	Non-Small Ce		1 Afatinib, Erlot	28336552, 22190593	

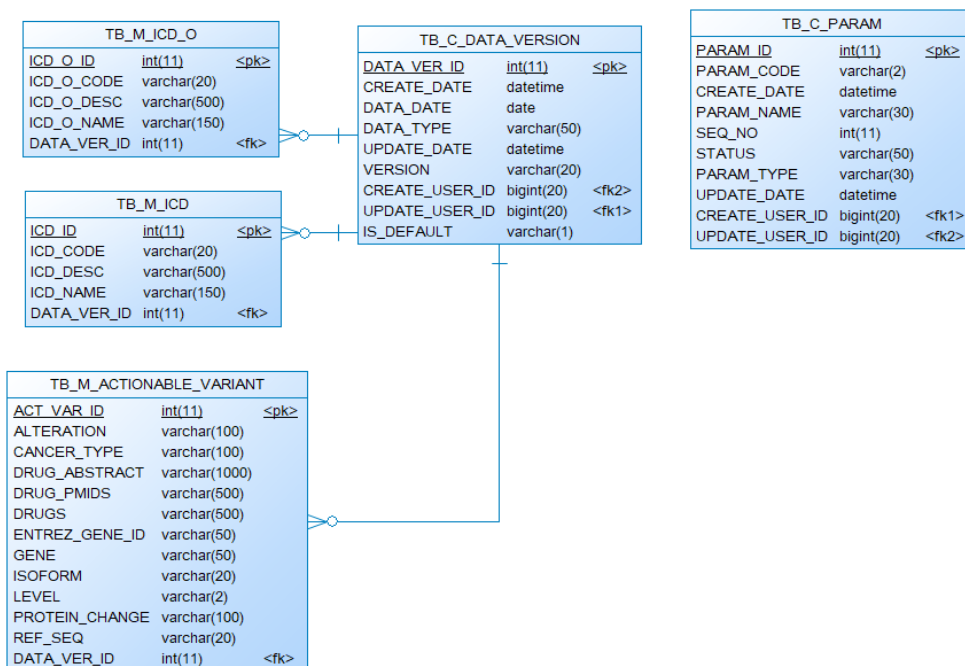
รูปที่ 51: ตัวอย่างไฟล์ข้อมูลของโคเคบีที่นำมาศึกษา

3.6.2 แนวทางออกแบบตารางข้อมูลให้รองรับข้อมูลจากฐานข้อมูลสาธารณะ

เนื่องจากแพลตฟอร์มต้องรองรับการเชื่อมโยงข้อมูลจากฐานข้อมูลสาธารณะหลากหลาย แหล่งข้อมูลเข้าด้วยกัน จึงได้สร้างตาราง TB_C_DATA_VERSION ดังรูปที่ 52 ซึ่งเก็บประวัติ เวอร์ชันของข้อมูลสาธารณะ เพื่อป้องกันผลกระทบที่เกิดจากการเปลี่ยนแปลงเวอร์ชันของข้อมูล สาธารณะที่มีต่อข้อมูลภายในแพลตฟอร์ม ในเวอร์ชันปัจจุบัน RUN-ONCO ได้นำรหัสวินิจฉัย โรคมะเร็งที่อยู่ในไอซีดี-10 มาบันทึกใช้ในการจัดเก็บข้อมูลทางคลินิกของผู้ป่วยแต่ละราย และถ้าใน อนาคตมีการเปลี่ยนรหัสวินิจฉัยโรคมะเร็งเป็นเวอร์ชันใหม่ เช่น ไอซีดี-11 แทน ก็จะไม่ส่งผลกระทบต่อ ข้อมูลเดิม นอกจากนี้ยังมีข้อมูลสาธารณะอื่น ๆ ได้แก่ ข้อมูลยีนที่เกี่ยวข้องกับมะเร็งที่มีผลต่อการ รักษา จากของโคเคบี (TB_M_ACTIONABLE_VARIANT) รหัสไอซีดี-10 (TB_M_ICD) และรหัส ไอซีดี-โอ (TB_M_ICD_O) ยกเว้นข้อมูลระยะของโรคมะเร็งที่ได้จากระบบ TNM ของ AJCC (TB_C_PARAM) ที่มีการเปลี่ยนแปลงข้อมูลน้อยมาก จึงไม่ต้องเก็บเป็นเวอร์ชันของข้อมูล

เมื่อมีการสร้างตารางข้อมูลใหม่ที่เป็นข้อมูลสาธารณะ ถ้าข้อมูลดังกล่าวมีการเปลี่ยนแปลงไปตามเวอร์ชัน จะต้องเชื่อมโยงข้อมูลไปยังตารางข้อมูล TB_C_DATA_VERSION เสมอ

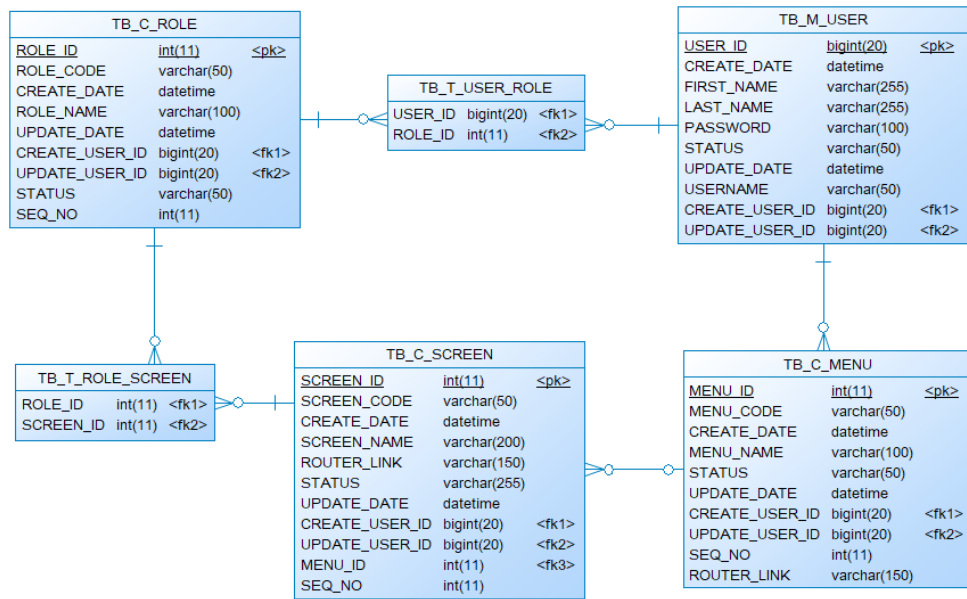
และถ้าต้องการใช้งานข้อมูลดังกล่าวทันที จะทำการอัปเดตฟิลด์ IS_DEFAULT ให้มีค่าเป็น Y และในเวอร์ชันนี้ RUN-ONCO รองรับการเพิ่มข้อมูลไอซีดี-10 และไอซีดี-โอผ่านหน้าจียูไอของเว็บของโคได้



รูปที่ 52: โครงสร้างตารางจัดเก็บข้อมูลที่รองรับการเปลี่ยนแปลงเวอร์ชันของข้อมูลสาธารณะ

3.6.3 แนวทางการออกแบบการกำหนดสิทธิ์การเข้าถึงหน้าจการทำงานและเมนูต่าง ๆ ของระบบ

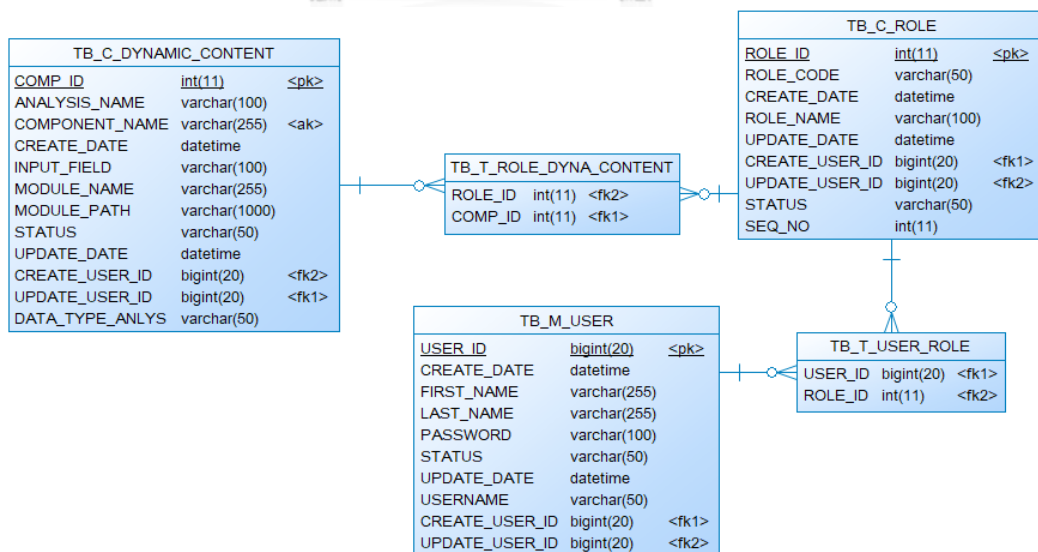
ระบบถูกออกแบบให้รองรับการกำหนดสิทธิ์ของผู้ใช้ระบบตามบทบาท (role) โดยผู้ใช้นี้หนึ่งคนสามารถมีได้มากกว่า 1 บทบาท และในแต่ละบทบาทจะกำหนดสิทธิ์การเข้าถึงตามหน้าจการทำงาน นอกจากนี้ระบบอนุญาตให้เพิ่มลดเมนูการทำงานได้ โดยกำหนดให้ภายใต้เมนูหนึ่ง ๆ สามารถมีได้หลายเมนูย่อย ซึ่งข้อมูลของเมนูย่อยจะใช้ตารางข้อมูลเดียวกันกับตารางข้อมูลของหน้าจการทำงาน ดังนั้นสิทธิ์การเข้าถึงเมนูย่อยจะใช้สิทธิ์เดียวกันกับสิทธิ์การเข้าถึงหน้าจการทำงาน ดังรูปที่ 53



รูปที่ 53: โครงสร้างตารางจัดเก็บข้อมูลการกำหนดสิทธิ์การเข้าถึงหน้าจการทำงานและเมนูต่าง ๆ ของระบบ

3.6.4 แนวทางการออกแบบการกำหนดสิทธิ์การเข้าถึงปลั๊กอินการวิเคราะห์

ในการออกแบบการกำหนดสิทธิ์การเข้าถึงปลั๊กอินการวิเคราะห์ ใช้หลักการเดียวกับการกำหนดสิทธิ์การเข้าถึงหน้าจการทำงาน ซึ่งใช้บทบาทในการกำหนดสิทธิ์การเข้าใช้งานปลั๊กอินต่าง ๆ ดังรูปที่ 54



รูปที่ 54: โครงสร้างตารางจัดเก็บข้อมูลการกำหนดสิทธิ์การเข้าถึงปลั๊กอินการวิเคราะห์

3.7 วิเคราะห์และออกแบบระบบเพื่อสนับสนุนการเพิ่มลดเซอร์วิสที่ใช้ในการวิเคราะห์ข้อมูลและแสดงผลข้อมูล

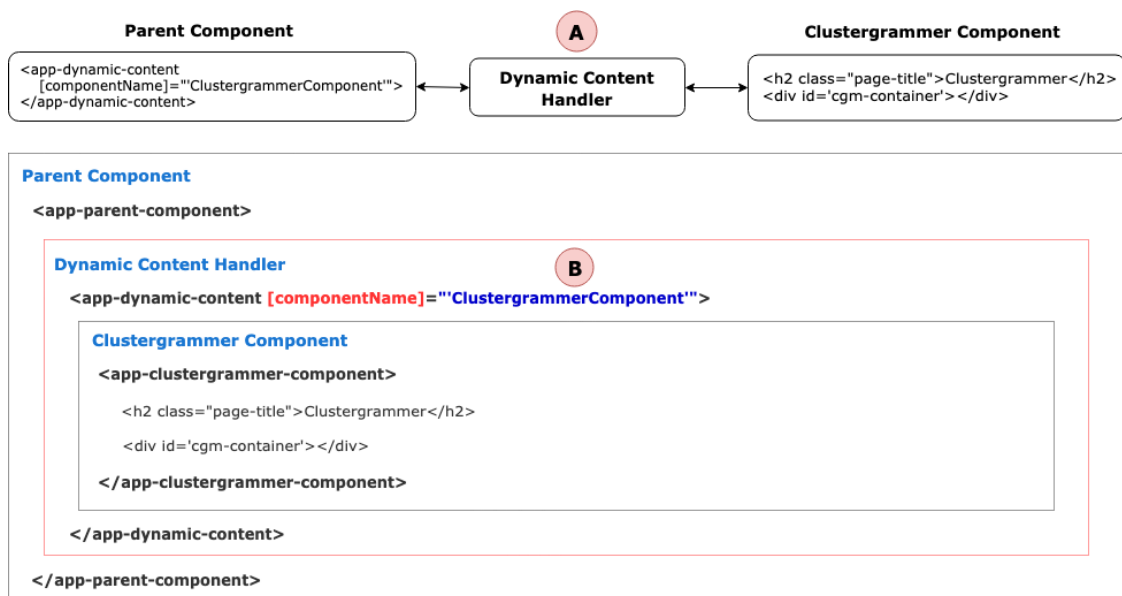
3.7.1 การเพิ่มเซอร์วิสที่ใช้ในการแสดงผลข้อมูล

จากแนวคิดของสถาปัตยกรรมเชิงคอมโพเนนต์ คอมโพเนนต์จะถูกสร้างให้มีคุณสมบัติการห่อหุ้ม (encapsulation) ซึ่งซ่อนกลไกการทำงานไว้ภายใน โดยกำหนดตัวแปรอินพุตเพื่อรับข้อมูลเข้ามาและตัวแปรเอาต์พุตเพื่อส่งข้อมูลออกไป วิทยานิพนธ์จึงนำคุณสมบัติดังกล่าวมาประยุกต์ใช้สร้างไดนามิกคอนเทนต์แฮนด์เลอร์ (dynamic content handler) ซึ่งเป็นยูไอคอนเทนเนอร์ (UI container) สำหรับปลั๊กอินคอมโพเนนต์ใหม่ ๆ เข้ามาบนเว็บของโคได้โดยง่าย เพื่อเพิ่มศักยภาพในการทำงานของคอมโพเนนต์หลัก (parent component) ด้วยการติดตั้งไดนามิกคอนเทนต์แฮนด์เลอร์ให้สามารถเพิ่มการแสดงผลวิเคราะห์ใหม่ ๆ โดยไม่ต้องแก้ไขโปรแกรมหรือรันโปรแกรมใหม่ ซึ่งผลวิเคราะห์จะทำงานทันทีในช่วงรันไทม์ เพื่อให้นักพัฒนาเพิ่มหรือดัดแปลงคอมโพเนนต์ได้ง่ายมากยิ่งขึ้น

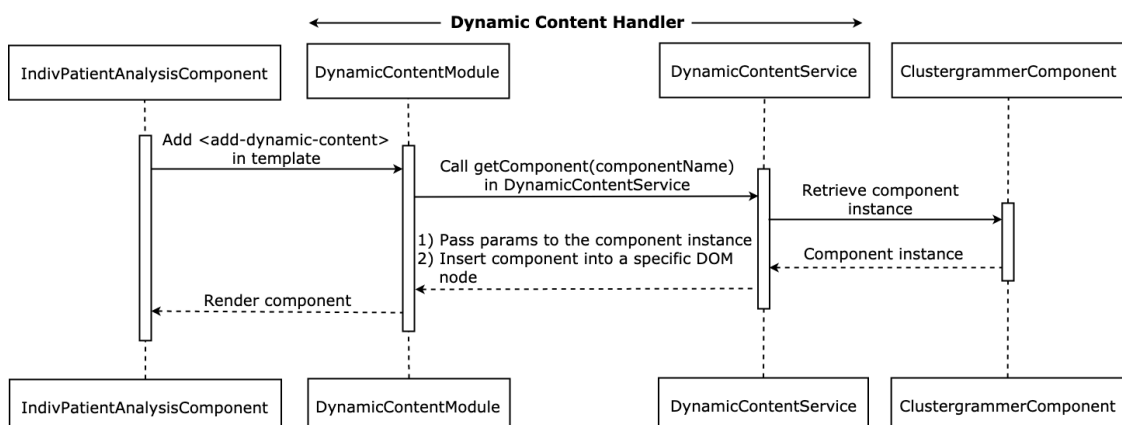
รูปที่ 55(A) แสดงภาพรวมกลไกการทำงานของไดนามิกคอนเทนต์แฮนด์เลอร์ โดยยกตัวอย่างการปลั๊กอินคลัสเตอร์แกรมเมอร์คอมโพเนนต์ (clustergrammer component) เพื่อแสดงผลการวิเคราะห์การจัดกลุ่มข้อมูลในรูปแบบแผนที่ความร้อน จากรูปที่ 55(B) จะเห็นว่าคอมโพเนนต์หลักส่งตัวแปรอินพุตชนิดสตริงชื่อคอมโพเนนต์ (component name) เช่น “ClustergrammerComponent” ไปยังไดนามิกคอนเทนต์แฮนด์เลอร์เพื่อสร้างอินสแตนซ์ของคลัสเตอร์แกรมเมอร์คอมโพเนนต์ขึ้นมา และแสดงผลการวิเคราะห์เชิงรูปภาพด้วยไลบรารีคลัสเตอร์แกรมเมอร์ภายใต้คอนเทนเนอร์ตัวเอง

ไดนามิกคอนเทนต์แฮนด์เลอร์ทำหน้าที่เป็นตัวกลางการรับส่งข้อมูลระหว่างคอมโพเนนต์หลักและปลั๊กอิน ดังรูปที่ 56 ประกอบด้วย 2 โมดูล 1) คลาส *DynamicContentService* ภายในไฟล์เรียกใช้ *SystemJs* ทำหน้าที่โหลดโมดูลและคอมไพล์ปลั๊กอินที่ลงทะเบียนไว้กับระบบ 2) คลาส *DynamicContentModule* ทำหน้าที่จัดเตรียมคอนเทนเนอร์ให้ปลั๊กอินแสดงผลการวิเคราะห์ ในการเปิดใช้งานปลั๊กอินแบบไดนามิกให้ส่งชื่อคอมโพเนนต์ เช่น “ClustergrammerComponent” คอมโพเนนต์จะแสดงผลในขณะรันไทม์ สำหรับคอมโพเนนต์ใดที่ต้องการสนับสนุนการแสดงผลวิเคราะห์ของปลั๊กอินแบบไดนามิกให้นำเข้าคลาส *DynamicContentModule* ไปยังโมดูลตัวเองและเพิ่มแท็ก *dynamic content* ในเทมเพลตดังแสดงในรูป 57 กระบวนการทำงานของไดนามิกคอนเทนต์แฮนด์เลอร์ มีดังนี้ 1) *IndivPatientAnalysisComponent* (parent component) ซึ่งเป็นคอมโพเนนต์หลักที่มีการติดตั้งการทำงานแบบไดนามิก ในการเปิดใช้งานปลั๊กอินต่าง ๆ ภายใต้คอมโพเนนต์นี้ จำเป็นต้องส่งชื่อคอมโพเนนต์ (component name) ทุกครั้ง (รูป 57) นอกจากนี้ยังสามารถส่งอ็อบเจกต์ *params* เพิ่มเติม เพื่อใช้ในการ

วิเคราะห์ผลโดยอยู่ขึ้นกับกระบวนการออกแบบโปรแกรมของนักพัฒนา นำส่งไปยังคลาส *DynamicContentModule* 2) คลาส *DynamicContentModule* จะเรียกใช้งาน *DynamicContentService* เพื่อสร้างอินสแตนซ์ของคลาสคอมโพเนนต์ และ 3) คลาส *DynamicContentService* จะส่งคืนอินสแตนซ์กลับไปยังคลาส *DynamicContentModule* แบบอะซิงโครนัส (asynchronous) เพื่อแทรกคอมโพเนนต์ดังกล่าวตรงตำแหน่งที่ได้กำหนดแท็ก *dynamic content* ไว้ในเทมเพลต



รูปที่ 55: กลไกการทำงานของไดนามิกคอนเทนต์แฮนด์เลอร์ (A) แผนภาพการเชื่อมต่อระหว่างคอมโพเนนต์หลักและคลัสเตอร์แกรมเมอร์คอมโพเนนต์ด้วยไดนามิกคอนเทนต์แฮนด์เลอร์ (B) โครงสร้างภาพการเรียกใช้คลัสเตอร์แกรมเมอร์ในคอมโพเนนต์หลักผ่านไดนามิกคอนเทนต์แฮนด์เลอร์



รูปที่ 56: แผนภาพแสดงลำดับของไดนามิกคอนเทนต์แฮนด์เลอร์

จากรูปที่ 57 แสดงตัวอย่างไดนามิกปลั๊กอินที่ติดตั้งใช้งานในระบบ RUN-ONCO โดยนำค่าที่พีเอ็ม จากผลวิเคราะห์การแสดงผลของยื่นไปจัดกลุ่มด้วยอัลกอริทึมเค-มีนส์ และนำผลวิเคราะห์ที่ได้ไปแสดงบนแผนที่ความร้อนด้วยไลบรารี clustergrammer.js [52] จาก [\[https://github.com/MaayanLab/clustergrammer\]](https://github.com/MaayanLab/clustergrammer)

ในการสร้างปลั๊กอินใหม่เป็นไปได้ว่ามีการนำไลบรารีจาวาสคริปต์จากภายนอกอย่าง clustergrammer.js มาช่วยในการแสดงผล ซึ่งขั้นตอนการสร้างไดนามิกปลั๊กอินและติดตั้งจาวาสคริปต์จากภายนอกมีด้วยกันดังต่อไปนี้ 1) เพิ่มไฟล์ clustergrammer.js ไปยังเฟรมเวิร์กของเว็บของโคและกำหนดพารของไฟล์ดังกล่าวไปที่ angular.json (รูปที่ 57(A)) 2) สร้างปลั๊กอินใหม่ที่มีคลาสชื่อ *ClustergrammerComponent* และประกาศตัวแปรเพื่อเรียกใช้คำสั่งที่มีอยู่ในไลบรารี clustergrammer.js ตัวอย่างการเรียกใช้ "declare var Clustergrammer: any" (รูปที่ 57(B)) 3) เพิ่มคลาสปลั๊กอินไปยังอาร์เรย์ declarations (รูปที่ 57(C1)) entryComponents (รูปที่ 57(C2)) และที่ตัวแปร mapping ซึ่งกำหนดอยู่ในคลาส *DynamicContentModule* (รูปที่ 57(C3)) 4) ติดตั้งคลาส *DynamicContentModule* ไปยังคลาส *IndivPatientAnalysisModule* (รูปที่ 57(D))

โดยคลาส *IndivPatientAnalysisModule* นี้เป็นโมดูลที่ระบบออกแบบไว้เป็นหน้ารวมปลั๊กอิน เพื่อการวิเคราะห์และแสดงผลข้อมูลในรูปแบบที่หลากหลาย

<p>angular.json A</p> <pre>"architect": { "build": { "options": { "scripts": [..., clustergrammer.js] } } }</pre>	<p>DynamicContentModule C</p> <pre>@NgModule({ declarations: [ClustergrammerComponent, // Insert new component here 1], entryComponents: [ClustergrammerComponent, // Insert new component here 2] }) export class DynamicContentModule { static mapping = { ClustergrammerComponent, // Insert new component here 3 }; }</pre>
<p>ClustergrammerComponent B</p> <pre>declare var Clustergrammer: any;</pre>	
<p>IndivPatientAnalysisComponent D</p> <pre>@NgModule({ imports: [..., DynamicContentModule] })</pre>	

รูปที่ 57: การกำหนดค่าในไฟล์ต่าง ๆ ที่จำเป็นสำหรับการติดตั้งปลั๊กอินแบบไดนามิกในแพลตฟอร์ม RUN-ONCO

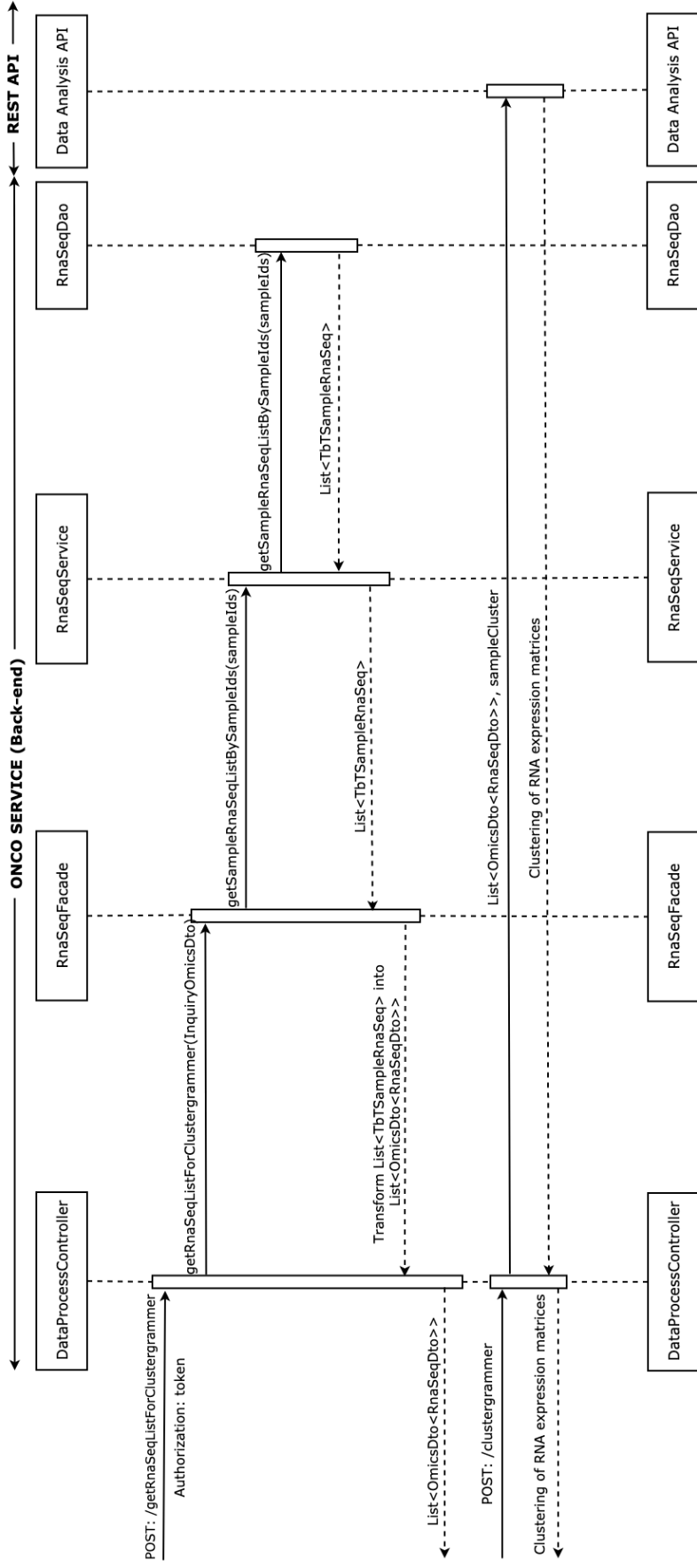
จากตัวอย่างปลั๊กอินคลัสเตอร์แกรมเมอร์ข้างต้น ภายในคอมโพเนนท์หลัก *IndivPatientAnalysisComponent* มีการส่งชื่อคอมโพเนนท์และออบเจกต์ `params` ผ่านแท็ก `app-`

ListForClustergrammer ซึ่งเป็น endpoint ของเมธอด `getRnaSeqListForClustergrammer` ภายใต้คลาส `DataProcessController` โดยคลาสดังกล่าวนี้จะเรียกใช้คลาส `RnaSeqFacade` ในการเรียกดูค่าการตรวจวัดการแสดงออกของยีนด้วยเลขที่ใช้ระบุไฟล์ตัวอย่าง (sample RNA-Seq IDs) ซึ่งภายในคลาสนี้จะตรวจสอบความถูกต้องของข้อมูลก่อนส่งไปให้ `RnaSeqService` เพื่อเปิดการเชื่อมต่อฐานข้อมูล และเรียกใช้ `RnaSeqDao` เพื่อสร้างคำสั่ง SELECT เลือกข้อมูลที่ต้องการจากฐานข้อมูล โดยนำส่งผลลัพธ์ที่ได้กลับไปยังคลาส `RnaSeqService` และที่ชั้นพะชาด จะทำการแปลงออบเจกต์ที่ได้ให้อยู่ในรูปแบบที่พร้อมถ่ายโอน (Data Transfer Objects: DTOs) เมื่อ `ClustergrammerComponent` ได้รายการค่าการแสดงออกของยีน จะทำการส่งคำขอ POST ไปยัง `/clustergrammer` ซึ่งเป็น endpoint ของเมธอด `clustergrammer` ภายใต้ไฟล์ `clustergrammer_api.py` (เป็นตัวอย่างดาต้าอานาไลซิสเอพีไอ) โดยจะจัดกลุ่มและวิเคราะห์ความสัมพันธ์ของการแสดงออกของยีนมะเร็ง โดยดาต้าอานาไลซิสเอพีไอนี้ได้จัดเตรียมอัลกอริทึม ซึ่งเป็นที่นิยมใช้กันมากในการจัดกลุ่ม ได้แก่ อัลกอริทึมเค-มีนส์ และอัลกอริทึมแบบลำดับชั้น โดยผลลัพธ์ที่ได้จะถูกแปลงให้อยู่ในรูปแบบออบเจกต์ที่ `clustergrammer.js` กำหนดไว้ เพื่อนำไปแสดงเป็นแผนที่ความร้อนบนคลาส `ClustergrammerComponent` ต่อไป

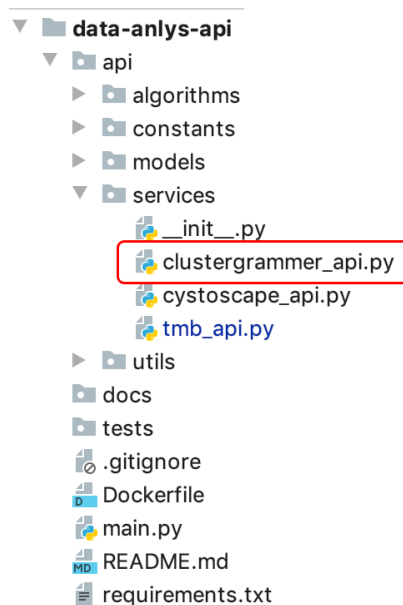
3.7.2 การเพิ่มเซอร์วิสที่ใช้ในการวิเคราะห์ข้อมูล

ดาต้าอานาไลซิสเอพีไอถูกออกแบบให้สนับสนุนการทำงานของปลั๊กอินแบบไดนามิก ไฟล์ `clustergrammer_api.py` ที่มีการอ้างอิงถึงในหัวข้อก่อนหน้านี้ เป็นตัวอย่างของดาต้าอานาไลซิสเอพีไอ โดยมีฟังก์ชันการจัดกลุ่มระดับการแสดงออกของยีนด้วยอัลกอริทึมการจัดแบ่งกลุ่ม ได้แก่ อัลกอริทึมเค-มีนส์ และอัลกอริทึมแบบลำดับชั้น นอกจากนี้ยังนำผลการวิเคราะห์ที่ได้จากการจัดแบ่งกลุ่มทำการแปลงข้อมูลให้อยู่ในรูปแบบคลัสเตอร์แกรมเมอร์กำหนด โดยขั้นตอนหลักในการเพิ่มเซอร์วิสมีดังต่อไปนี้ 1) สร้างไฟล์ชื่อ `clustergrammer_api.py` ภายใต้โฟลเดอร์ `api/services` (รูปที่ 60) 2) สร้างเมธอด `clustergrammer()` เพื่อใช้ในการคำนวณหาค่าที่เอ็มบี และกำหนด endpoint ชื่อ `clustergrammer` ในไฟล์ `clustergrammer_api.py` (รูปที่ 61) ตัวอย่างโดยละเอียดสำหรับการเพิ่มเซอร์วิสในการวิเคราะห์ข้อมูลอยู่ในหัวข้อที่ 4.7

สำหรับการเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างชีววัตถุ 2 ชนิด จะทำผ่านเอสคิวแอล (Structured Query Language: SQL) เป็นหลักโดยแสดงขั้นตอนการหาไว้ในภาคผนวก ข.



รูปที่ 59: แผนภาพแสดงลำดับการรับส่งข้อมูลของส่วนประกอบต่าง ๆ ที่เกี่ยวข้องกับการวิเคราะห์ที่การแสดงผลออกเป็นเมฆเริง



รูปที่ 60: ตำแหน่งที่เพิ่มไฟล์ clustergrammer_api.py ในดาต้าอานาไลซิสเอพีไอ

```

from flask import Blueprint, request, jsonify
from api.utils import app_util, data_preprocess as preprocessor

clustergrammer_api = Blueprint('clustergrammer_api', __name__)

@clustergrammer_api.route("/clustergrammer", methods=['POST'])
def clustergrammer():
    req = request.get_json()

    files, items = preprocessor.prepare_data(req['sampleOmics'])
    norm_gene, norm_sample, df = preprocessor.clustergrammer(files, items)

    sample_group = app_util.execute_module(norm_sample, req['sampleCluster'])
    gene_group = app_util.execute_module(norm_gene, req['geneCluster'])

    clustergrammer = app_util.clustergrammer(files, sample_group, gene_group, df)

    gene_id = df[['gene_id']]
    df_ideogram = preprocessor.ideogram(files, items, gene_id)
    container, datatable = app_util.ideogram(df_ideogram, gene_group)

    output = {
        "clustergrammer": clustergrammer,
        "ideogram": {
            "container": container,
            "datatable": datatable
        }
    }

    return jsonify(output), 200

```

Request (indicated by red arrows pointing to the request parameters in the code)

Response (indicated by a red box around the output dictionary)

รูปที่ 61: ตัวอย่างการเขียนโปรแกรมเพิ่มส่วนการวิเคราะห์ข้อมูลในไฟล์ clustergrammer_api.py

บทที่ 4

ผลการวิจัยและอภิปรายผล

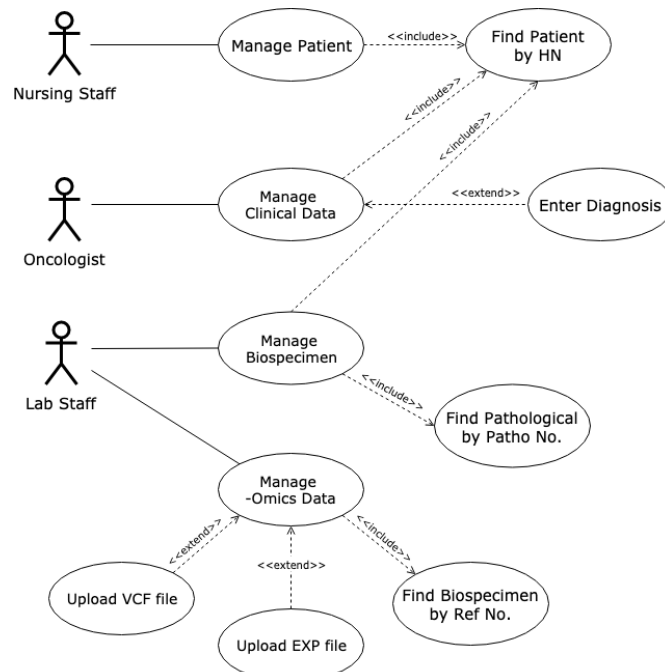
ในบทนี้ผู้วิจัยจะนำสิ่งที่ได้ศึกษาและขั้นตอนการออกแบบไว้ในบทที่ 3 มาพัฒนาซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการรักษาโรคมะเร็งแบบแม่นยำและจำเพาะ

4.1 ภาพรวมของระบบ

แพลตฟอร์มถูกออกแบบมาเพื่อลดความซับซ้อนของการรักษาแบบการแพทย์แม่นยำในโรคมะเร็ง ซึ่งช่วยให้ผู้ใช้ระบบสามารถจัดการและวิเคราะห์ข้อมูลของตนเองได้ จากรูปที่ 62 เมื่อผู้ป่วยเข้ามารับการรักษาคือครั้งแรกจำเป็นต้องส่งประวัติผู้ป่วยเข้าระบบ และจะจัดเก็บข้อมูลทางคลินิกทุกครั้งที่ของการรักษา ได้แก่ การวินิจฉัยโรค พยาธิวิทยา ระยะของโรคมะเร็ง และการดำเนินโรค โดยระบบจะทำการเชื่อมโยงกับประวัติผู้ป่วยให้โดยอัตโนมัติ ในผู้ป่วยบางรายแพทย์มีการผ่าตัดเนื้องอก ซึ่งจะนำตัวอย่างชิ้นเนื้อส่งห้องปฏิบัติการเพื่อตรวจวินิจฉัยการกลายพันธุ์ของยีน นอกจากนั้นสามารถที่จะแนะนำเซลล์หรือเจาะเลือดไปตรวจได้เช่นกันขึ้นอยู่กับชนิดและระยะของโรคมะเร็ง โดยกระบวนการตรวจวิเคราะห์อาจจะใช้การถอดรหัสจีโนม การถอดรหัสเอ็กโซมหรือการตรวจวัดการแสดงออกของยีนผ่านเทคโนโลยีอาร์เอ็นเอซีค แล็บจะส่งผลการตรวจในรูปแบบไฟล์วีซีเอฟหรือไฟล์อีเอ็กซ์พีตามลำดับ เมื่อได้รับผลผู้ใช้สามารถนำไฟล์ข้อมูลเข้าแพลตฟอร์ม

4.2 การจัดเก็บข้อมูลประวัติผู้ป่วย (Patient Information Management)

ระบบบริการส่วนจัดการข้อมูลประวัติผู้ป่วย โดยจัดเก็บข้อมูลเลขที่เวชระเบียน ข้อมูลพื้นฐานของผู้ป่วย เช่น เพศ วันเกิด เชื้อชาติ น้ำหนัก ส่วนสูง (รูปที่ 63(A)) และประวัติการติดตามการมีชีวิตของผู้ป่วย เช่น รูปที่ 63(B) และ รูป 63(C) ข้อมูลส่วนบุคคลและข้อมูลที่ละเอียดอ่อนเกี่ยวกับผู้ป่วย เช่น ชื่อ ที่อยู่ เลขที่บัตรประชาชน เป็นต้น จะไม่ถูกจัดเก็บไว้ในระบบ



รูปที่ 62: ภาพรวมของระบบ

Search > Patient Form

Patient Info

HN: HN620001 | Gender: Male | Date of birth: 2019-06-29 | Race: Asian

Weight (kg): 65.00 | Height (m): 1.75

Survival follow up | Variant call | RNA Seq

Vital Status	Date of Last Followup	Date of Death	Date of Record	Edit	Delete
Alive	29 Jun 2019		01 Jul 2019		

Survival follow up (Modal)

Vital Status: Alive | Date of Last Followup: 2019-06-29 | Date of Record: 2019-07-01

รูปที่ 63: ขั้นตอนการจัดเก็บข้อมูลประวัติผู้ป่วย (A) กรอกประวัติผู้ป่วยใหม่ (B) เพิ่มประวัติการติดตามการมีชีวิตของผู้ป่วย โดยการคลิกที่ปุ่ม "+ Add" และ (C) ข้อมูลประวัติการติดตามการมีชีวิตของผู้ป่วยจะอัปเดตในตารางข้อมูลให้อัตโนมัติ

4.3 การจัดเก็บข้อมูลทางคลินิก (Clinical Data Management)

ระบบบริการส่วนจัดการข้อมูลทางคลินิก ทำให้จัดเก็บข้อมูลในระหว่างการติดตามการรักษาได้ ซึ่งช่วยแพทย์ผู้เชี่ยวชาญสามารถออกแบบแผนการรักษาที่เหมาะสมปลอดภัย และมีความแม่นยำต่อผู้ป่วย ข้อมูลทางคลินิกแบ่งออกเป็น 4 ประเภทหลัก 1) การวินิจฉัยโรค (รูปที่ 64(A)) ประกอบด้วยวันที่วินิจฉัยโรค รหัสสำหรับการวินิจฉัยหลัก ประกอบด้วย รหัสไอซีดี-10 เช่น C18 ซึ่งใช้ในการจำแนกและระบุรายละเอียดของโรค malignant neoplasm of colon 2) การวินิจฉัยทางพยาธิวิทยา รูปที่ 64(B) และรูปที่ 64(C) ถูกออกแบบให้จัดเก็บข้อมูลดังต่อไปนี้ รหัสพาโร วันที่ส่งตรวจชิ้นเนื้อ วันที่รายงานทางพยาธิวิทยาของชิ้นเนื้อ ตำแหน่งของการผ่าตัดเอาชิ้นเนื้อ ลักษณะของเนื้องอก และคุณสมบัติทางชีววิทยาของเซลล์มะเร็ง ประกอบด้วย เช่น 8010/9 ซึ่งใช้ในการอธิบายประเภทของมะเร็ง carcinomatosis 3) ระยะของโรคมะเร็ง (รูปที่ 65) ได้นำระบบ TNM ของ AJCC มาใช้งาน เพื่อจำแนกระยะของโรค ได้แก่ ประเภทระยะของโรคมะเร็ง และถ้าเลือกประเภท c ระบบจะแสดงตัวเลือกระยะของโรคมะเร็งตามลักษณะทางคลินิก (cT Stage) และระยะย่อย (cT substage) ขึ้นมาให้ เมื่อเลือกครบทั้ง 2 ฟิลด์ ค่าที่ได้คือ cTxa นอกจากนั้นยังมีข้อมูลอื่น ๆ เช่น ระยะของโรคมะเร็งตามลักษณะทางคลินิก โรคมะเร็งตามลักษณะทางพยาธิวิทยา และ 4) การดำเนินโรค เพื่อเฝ้าติดตามโรคมะเร็งย้อนกลับเป็นซ้ำ ข้อมูลที่ถูกจัดเก็บ ได้แก่ สถานะย้อนกลับเป็นซ้ำ วันที่ย้อนกลับเป็นซ้ำ

4.4 การจัดเก็บข้อมูลชีววัตถุ (Biospecimen Management)

ระบบบริการจัดเก็บข้อมูลชีววัตถุ ประกอบไปด้วย ประเภทของตัวอย่างเก็บตรวจ ได้แก่ เลือด เนื้อเยื่อ และเซลล์ ประเภทของเนื้อเยื่อ เช่น colon fresh tissue, lung fresh tissue, ovary fresh tissue รหัสพาโร วิธีการแช่เยือกแข็ง เช่น -20°C , -20°C , liquid nitrogen วิธีเก็บรักษาตัวอย่างเก็บตรวจ เช่น RNAlater, snap freezing (รูปที่ 66)

RUN ONCO Patient Clinical data Biospecimen Omics Analysis Logout

Search > Diagnoses > Diagnosis Form

Diagnosis

Diagnosis Date: 2019-07-01 A

Primary Diagnosis (ICD-10): C18 Malignant neoplasm of colon

Pathological **Cancer Stage** Clinical Course C

Pathology No	Date of tissue	Site of Resection or Biopsy	Morphology (ICD-O)	Edit	Delete
62-0001	01 Jul 2019	colon	8010/9: Carcinomatosis		

[+ Add](#)

Pathological

Pathology Number: 62-0001

Date of tissue pathological report: 2019-07-01

Site of Resection or Biopsy: colon

Procedure (tissue assessment): Biopsy

Type of tissue: Pathology

Histologic grade: 1

Morphology (ICD-O): 8010/9 Carcinomatosis

B

[cancel](#) [save](#)

รูปที่ 64: ขั้นตอนการจับเก็บข้อมูลการวินิจฉัยทางพยาธิวิทยา (A) เพิ่มข้อมูลการวินิจฉัยโรคในหน้าจอ (B) เพิ่มการวินิจฉัยทางพยาธิวิทยา โดยการคลิกที่ปุ่ม "+ Add" และ (C) ข้อมูลการวินิจฉัยทางพยาธิวิทยาอัปเดตในตารางข้อมูลอัตโนมัติ

Pathological **Cancer Stage** Clinical Course

AJCC: 6th

Type of stage: c

cT stage (only neoadjuvant setting): Tx

cT substage: a

cN stage (only neoadjuvant setting): NO

cN substage: b

M stage: Mx

M substage: a

Clinical stage (only neoadjuvant setting): II

Pathological stage: II

Pathological substage: b

[Cancel](#) [Save & Next](#) [Save](#)

รูปที่ 65: แบบฟอร์มกรอกข้อมูลระยะของโรคมะเร็ง

RUN ONCO Patient Clinical data Biospecimen Omics Analysis Logout

Search > Biospecimen Form

Patient Info

ID :	HN620001	Gender :	Male	Date of Birth :	29 Jun 2019
Weight (kg) :	65	Height (m) :	1.75	Race :	Asian

Biospecimen

Biospecimen ID	Sample Type	Tissue Type	Pathology Number
0001	Tissue	Colon Fresh tissue	62-0001
Freezing Method	Preservation Method	Initial Weight (mg)	Date of Sample Collection
- 80 °C	RNALater	50.00	2019-07-01
Date of Freezing			
2019-07-01			

Cancel Save

รูปที่ 66: แบบฟอร์มกรอกข้อมูลชีววัตถุ

4.5 การจัดเก็บและการวิเคราะห์ข้อมูลโอมิกส์ (-Omics Data Management and Analysis)

ระบบบริการจัดเก็บข้อมูลโอมิกส์ ประกอบไปด้วย ข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอ (variant calls) ซึ่งได้จาก WGS/WES และผลวัดระดับการแสดงออกของยีนด้วยเทคนิค RNA-Seq สามารถนำไฟล์ข้อมูลทั้งหมดเข้าสู่แพลตฟอร์ม โดยไปยังหน้าจัดการข้อมูลประวัติผู้ป่วยแท็บที่สอง และสามารถตามลำดับ (รูปที่ 67) ระบบจะทำการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลโอมิกส์และชีววัตถุ ให้อัตโนมัติ สามารถนำข้อมูลดังกล่าวไปวิเคราะห์ร่วมกับข้อมูลอื่น ๆ เช่น ยีนก่อมะเร็ง (oncogene) การแปรผันเชิงพันธุกรรมที่เกี่ยวข้องกับมะเร็งที่มีผลต่อการรักษา (actionable variant) โดยอิงโคเคบี ข้อมูลการมีปฏิสัมพันธ์ระหว่างโปรตีน (protein-protein interaction) จากฐานข้อมูลสตริง หรือข้อมูลพาธเวย์ (pathway) จากฐานข้อมูล KEGG PATHWAY เป็นต้น โดยระบบให้บริการการวิเคราะห์ข้อมูลเชิงลึกแบบบูรณาการเพื่อช่วยในการวินิจฉัยโรคมะเร็ง และเลือกยาที่เหมาะสมกับผู้ป่วยแต่ละรายมากยิ่งขึ้น

RUN-ONCO ถูกออกแบบมาให้มีความยืดหยุ่นสูงต่อการติดตั้งปลั๊กอินใหม่ ๆ ที่ช่วยในการวิเคราะห์และแสดงผลเชิงรูปภาพ โดยโมดูลดังกล่าวจะถูกเพิ่มเข้าฐานข้อมูลและแสดงผลในช่วงรันใหม่ ปลั๊กอินที่มาพร้อมกับแพลตฟอร์ม ได้แก่ 1) การเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างชีววัตถุ 2 ชนิด และ 2) การวิเคราะห์ผลวัดระดับการแสดงออกของยีน

4.5.1 การเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างชีววัตถุ 2 ชนิด (Variant Comparison)

การเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างชีววัตถุสองชนิด (รูปที่ 67) ได้แก่ 0001 (Tissue) และ 0002 (Blood) ของผู้ป่วยรหัส HN620001 ที่มี actionable variant ยีนที่ค้น

พบในในฐานข้อมูลของโคเคบี ผู้ใช้ระบบสามารถเลือกไฟล์วีซีเอฟที่ต้องการ ซึ่งแสดงในรูปแบบที่ 67(A) และ 67(B) ยีน *MTOR* ถูกพบในเนื้อเยื่อแรก 0001_WES ตำแหน่งที่ 11293367 บน โครโมโซมที่ 1 อัลลีลอ้างอิง TA ถูกแทนที่ด้วยอัลลีล T ไม่พบการกลายพันธุ์ตำแหน่งนี้ในเลือด 0002_WES โดยยีน *MTOR* เป็นยีนที่มีกลายพันธุ์และมีข้อมูลในฐานข้อมูลของโคเคบี โดยมีระดับของหลักฐานระดับที่ 3A หลักฐานที่บ่งชี้ทางคลินิกสำหรับโรคมะเร็ง (compelling clinical evidence in this indication) และ 4 หลักฐานที่บ่งชี้ทางชีวภาพสำหรับโรคมะเร็ง (compelling biological evidence) ดังแสดงในรูปแบบที่ 67(C)

4.5.2 การวิเคราะห์ระดับการแสดงออกของยีนมะเร็ง (Integrative Analysis of Gene Expression Profiles)

การวิเคราะห์นี้นำข้อมูลอาร์เอ็นเอซีคของผู้ป่วยหลาย ๆ ราย มาจัดกลุ่ม (clustering) เพื่อวิเคราะห์ความสัมพันธ์ของการแสดงออกของยีนมะเร็ง จากรูปที่ 68(A) ผู้ใช้ระบบเลือกข้อมูลอาร์เอ็นเอซีคจากผู้ป่วยสองรายและกลุ่มยีนมะเร็งที่สนใจ เช่น OncoKB-52 โดย OncoKB-52 ประกอบไปด้วยยีนมะเร็ง 52 ยีนที่รักษาได้และมีวิธีการรักษาซึ่งได้จากฐานข้อมูลของโคเคบี นอกจากนี้ยังมีกลุ่มยีนมะเร็งอื่น ๆ ที่นำมาวิเคราะห์ใน RUN-ONCO ได้แก่ กลุ่มยีนมะเร็งที่มีความเสี่ยงต่อการเกิดมะเร็งลำไส้ใหญ่สูง (High-risk colon cancer) เช่น *APC*, *AXIN2*, *BMPR1A* กลุ่มยีนมะเร็งเต้านมและมะเร็งรังไข่ (Breast/Ovarian cancers) เช่น *BARD1*, *BRCA1*, *BRCA2* กลุ่มยีนมะเร็งศีรษะและคอ (Head and Neck cancers) เช่น *DKC1*, *RTEL1*, *TERT* กลุ่มยีนมะเร็งของต่อมไร้ท่อ เช่น *MAX*, *TMEM127*, *VHL* เป็นต้น นอกจากนี้ผู้ใช้สามารถเลือกอัลกอริทึมการจัดแบ่งกลุ่มได้ด้วยตัวเอง ได้แก่ อัลกอริทึมเค-มีนส์ และอัลกอริทึมแบบลำดับชั้น โดยกำหนดให้ค่าเริ่มต้นของการวิเคราะห์คือ อัลกอริทึมเค-มีนส์และค่าเค (K value) เท่ากับ 3 จากรูปที่ 68(B) แสดงผลวิเคราะห์บนแผนที่ความร้อนด้วย clustergrammer.js ตามระดับการแสดงออกของยีน ซึ่งตัดแบ่งออกเป็น 3 กลุ่ม ด้วยวิธีเค-มีนส์ จะสังเกตเห็นว่ากลุ่มที่ 2 จะมีระดับการแสดงออกสูงสุดเมื่อเทียบกับอีก 2 กลุ่ม จากรูปที่ 68(C) ผู้ใช้สามารถศึกษาการแสดงออกของยีนในมุมมองการมีปฏิสัมพันธ์ระหว่างโปรตีนจากฐานข้อมูลสตริง ซึ่งเก็บรวบรวมข้อมูลจากหลาย ๆ แหล่ง ได้แก่ Neighborhood, Gene Fusion, Experiments เป็นต้น ผู้ใช้สามารถเลือกแหล่งข้อมูลและยีนมะเร็งที่สนใจ จากตัวอย่างได้ทำการเลือกยีนมะเร็ง 4 ยีน จากกลุ่มที่ 2 ที่มีระดับการแสดงออกของยีนสูงสุด ได้แก่ *IDH1*, *IDH2*, *SRSF2*, *U2AF1* รูปที่ 68(D) แสดงการมีปฏิสัมพันธ์ระหว่างโปรตีนจากฐานข้อมูลสตริงในรูปแบบกราฟด้วย cytoscape.js [44] และพบการมีปฏิสัมพันธ์ระหว่างคู่โปรตีน 2 คู่ ได้แก่ 1) คู่ยีน *IDH1* และ *IDH2* และ คู่ยีน *SRSF2* และ *U2AF1* หรือ *U2AFBP*

Analysis Variant Comparison HN HN620001

Biospecimen A

Sequence type WES **Sample source** 0001 (Tissue), 0002 (Blood) 🔍

Variant Call

VCF 0001_WES, 0002_WES Update B

Variant Comparison

OncoKB	Variant	Chrom	Position	Ref. Allele	Alt. Allele	Gene	#1	#2
🔍	1:11172796 G / A	1	11172796	G	A	MTOR	✓	
🔍	1:11182261 GT / G	1	11182261	GT	G	MTOR	✓	
🔍	1:11187618 AT / A	1	11187618	AT	A	MTOR	✓	
🔍	1:11270556 GA / G	1	11270556	GA	G	MTOR	✓	
🔍	1:11272636 C / A	1	11272636	C	A	MTOR		✓
🔍	1:11293367 TA / T	1	11293367	TA	T	MTOR	✓	
🔍	1:115248250 CT / C	1	115248250	CT	C	NRAS	✓	
🔍	1:11525						✓	
🔍	1:11526						✓	
🔍	1:15681						✓	
🔍	1:15682						✓	
🔍	1:15683						✓	
🔍	1:15684						✓	
🔍	1:15684						✓	

Known oncogenic x

Isoform : ENST00000361445 C

Gene	Alteration	Cancer Type	Drugs	Level
MTOR	E2014K	Bladder Cancer	Everolimus	3A
	E2419K	Bladder Cancer	Everolimus	3A
	L1460P	Renal Cell Carcinoma	Temsirolimus	3A
	L2209V	Renal Cell Carcinoma	Temsirolimus	3A
	L2427Q	Renal Cell Carcinoma	Temsirolimus	3A
	Q2223K	Renal Cell Carcinoma	Everolimus	3A
	Oncogenic Mutations	All Solid Tumors	Temsirolimus, Everolimus	4

close

Showing 1 to 15 of 10

รูปที่ 67: ผลการเปรียบเทียบการแปรผันของดีเอ็นเอจากชีววัตถุ 2 ชนิดของผู้ป่วยรายเดียวกัน (A) เลือกรีวิววิเคราะห์ผล Variant Comparison และกรอกรหัสผู้ป่วย HN620001 เพื่อค้นหาชีววัตถุของผู้ป่วยรายดังกล่าว แล้วเลือกประเภทการวิเคราะห์ลำดับเบสเป็น WES โดยสามารถเลือกชีววัตถุหลากหลายชนิดในคราวเดียวกัน ได้แก่ 0001 (Tissue), 0002 (Blood) หลังจากนั้น (B) เลือกไฟล์วีซีเอฟดังนี้ 0001_WES และ 0002_WES และกดปุ่ม “Update” ผลลัพธ์จะปรากฏบนหน้าจอ และ (C) คลิกที่ไอคอนของอองโคเคบี สืบหาข้อมูลรายละเอียดการกลายพันธุ์ของยีน *MTOR* และยาที่เกี่ยวข้อง

Analysis
Clustergrammer A

RNA-Seq
RNA-Seq: RUN-CA05.exp, RUN-CA07.exp
Cancer Gene Group: OncoKB-52 Update

Settings (k=3 using K-means clustering)

Clustergrammer B

Cytoscape C

Data source
Experimental, Database, Textmining Update

Selected: experimental, database, textmining

neighborhood fusion phylogenetic coexpression experimental database textmining

Genes

	Gene ID	Gene Name	Reference	Group
<input type="checkbox"/>	ENSG00000174775	HRAS	11	3
<input checked="" type="checkbox"/>	ENSG00000138413	IDH1	2	2
<input checked="" type="checkbox"/>	ENSG00000182054	IDH2	15	2
<input type="checkbox"/>	ENSG00000096968	JAK2	9	1
<input type="checkbox"/>	ENSG00000147050	KDM6A	X	1
<input type="checkbox"/>	ENSG00000157404	KIT	4	1
<input type="checkbox"/>	ENSG00000133703	KRAS	12	1
<input type="checkbox"/>	ENSG00000169032	MAP2K1	15	1
<input type="checkbox"/>	ENSG00000135679	MDM2	12	1
<input type="checkbox"/>	ENSG00000105976	MET	7	1

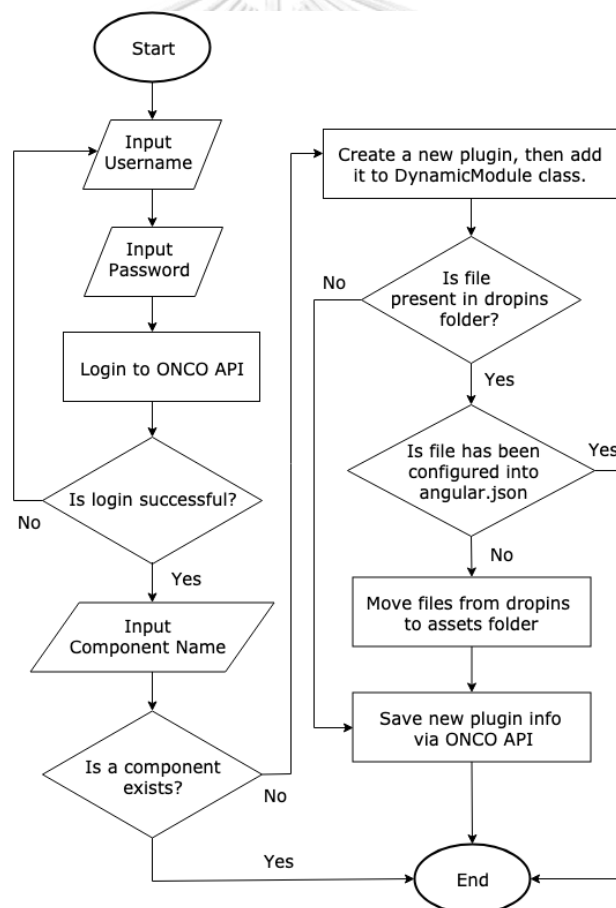
Showing 21 to 30 of 50 records Show: 10: « < 3 / 5 > »

Cytoscape D

รูปที่ 68: ผลการวิเคราะห์ระดับการแสดงออกของยีนมะเร็ง (A) เลือกการวิเคราะห์ผล cluster grammer (B) เลือกไฟล์อีเอ็กซ์พี RUN-CA05.exp, RUN-CA07.exp และเลือกกลุ่มยีนมะเร็งที่สนใจ และกดปุ่ม “Update” ผลลัพธ์การจัดกลุ่มจะปรากฏบนหน้าจอ (C) เลือกแหล่งข้อมูลจากข้อมูลสตริง โดยสามารถเลือกแหล่งข้อมูลหลาย ๆ แหล่งในคราวเดียวกัน แล้วคลิก “Update” และ (D) กราฟ แสดงการมีปฏิสัมพันธ์ระหว่างคู่โปรตีนของยีนก็จะปรากฏบนหน้าจอ

4.6 พัฒนาเครื่องมือที่ช่วยสร้างเทมเพลตปลั๊กอินส่วนแสดงผลบนเว็บของโคโดยอัตโนมัติ

หลักการการทำงานของเครื่องมือ (รูปที่ 69) มีดังนี้ 1) ในการสร้างเทมเพลตปลั๊กอินแต่ละครั้งจะต้องล็อกอินเพื่อยืนยันตัวตนผ่านองโคเซอร์วิส 2) เมื่อยืนยันตัวตนสำเร็จ จึงจะกรอกชื่อปลั๊กอินที่ต้องการสร้าง เช่น tmb-score 3) ถ้าปลั๊กอินดังกล่าวไม่เคยสร้างมาก่อนเครื่องมือจะทำการสร้างเทมเพลตพร้อมกับติดตั้งไปที่คลาส *DynamicContentModule* 4) ถ้านักพัฒนามีจาวาสคริปต์ภายนอกที่ต้องการติดตั้งบนเว็บของโค และไม่เคยนำเข้ามามาก่อน จะต้องนำไฟล์ดังกล่าวไปวางไว้ที่โฟลเดอร์ dropins เครื่องมือจะทำการย้ายไฟล์ไปยังโฟลเดอร์ assets พร้อมกับเพิ่มพาธไฟล์ในอาร์เรย์ scripts ภายในไฟล์ angular.json ให้อัตโนมัติ ในทางกลับกันถ้าไม่ต้องการติดตั้งไฟล์จาวาสคริปต์ใด ๆ ขึ้นตอนนี้ก็ข้ามไป 5) เมื่อขั้นตอนทุกอย่างถูกต้อง เครื่องมือจะส่งข้อมูลปลั๊กอินดังกล่าวไปยังองโคเซอร์วิส เพื่อบันทึกข้อมูลต่อไป



รูปที่ 69: แผนภาพแสดงลำดับการทำงานของเครื่องมือที่ช่วยสร้างเทมเพลตปลั๊กอินบนเว็บของโคอัตโนมัติ

4.7 แนวทางการพัฒนาต่อยอดปลั๊กอินการวิเคราะห์ใหม่ ๆ บนแพลตฟอร์ม

แพลตฟอร์ม RUN-ONCO ที่นำเสนอในวิทยานิพนธ์นี้ได้ให้บริการข้อมูลโอมิกส์ 2 ชนิดคือ 1) ข้อมูลการแปรผันของดีเอ็นเอ และ 2) ข้อมูลระดับการแสดงออกของยีน นักพัฒนาจะต้องมีบัญชีผู้ใช้บนแพลตฟอร์ม จึงจะสามารถเชื่อมต่อเพื่อขอใช้บริการข้อมูลดังกล่าวได้ และในตารางที่ 7 แสดงรายการบริการที่ทางแพลตฟอร์มเปิดให้ใช้งาน ซึ่งรายละเอียดต่าง ๆ สามารถดูเพิ่มเติมได้ที่ภาคผนวก ค.

ตารางที่ 7: รายการบริการที่ทางแพลตฟอร์มเปิดให้ผู้พัฒนาสามารถเรียกใช้งานได้

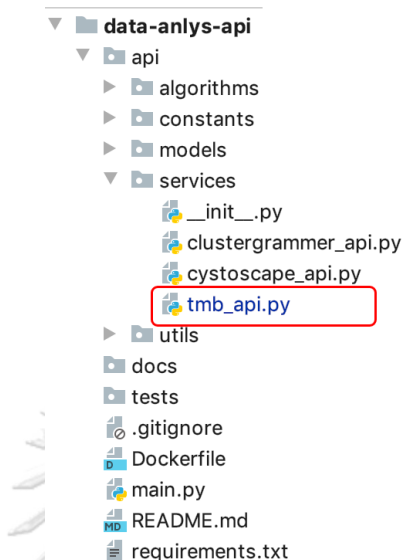
ลำดับ	ชื่อบริการ	ให้บริการ
1	getVariantCallsBySampleId	เรียกดูข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอ ด้วยเลขที่ตัวอย่าง (Sample VCF ID)
2	getRnaSeqListBySampleId	เรียกดูข้อมูลระดับการแสดงออกของยีนด้วยเลขที่ตัวอย่าง (Sample RNA-Seq ID)

ในส่วนต่อไปนี้จะยกตัวอย่างการสร้างบริการใหม่ชื่อทีเอ็มบี (Tumor Mutational Burden: TMB) ซึ่งเป็นตัวชี้วัดปริมาณการกลายพันธุ์ จึงจำเป็นต้องเขียนส่วนของการคำนวณเพิ่มเติมเข้ามาในแพลตฟอร์ม โดยนำข้อมูลการแปรผันของดีเอ็นเอจากข้อมูลเอ็กโซม (WES) มาใช้ในการวิเคราะห์

การสร้างเซอร์วิสการคำนวณค่าทีเอ็มบี มีขั้นตอนดังนี้

1. ไปยังโฟลเดอร์ data-anlys-api ซึ่งเก็บซอร์สโค้ดของการวิเคราะห์ข้อมูลโอมิกส์ โดยสร้างไฟล์ tmb_api.py ไว้ในโฟลเดอร์ api/services (รูปที่ 70)
2. สร้างเมธอด calc_tmb_score() เพื่อใช้ในการคำนวณค่าทีเอ็มบี และกำหนด endpoint ชื่อ calcTmbScore ในไฟล์ tmb_api.py (รูปที่ 71)
3. ไปยังโฟลเดอร์ onco-web ซึ่งเก็บโฟลเดอร์ซอร์สโค้ดของเว็บองโค โดยเริ่มจากโฟลเดอร์ราก (root folder) ของโปรเจกต์ โดยพิมพ์คำสั่ง `java -jar dynamic-component-jar-with-dependencies.jar` ที่คอมมานด์ไลน์ (รูปที่ 72) ซึ่งจาร์ไฟล์นี้ เป็นเครื่องมือที่ถูกพัฒนาขึ้นมาเพิ่มเติม เพื่อใช้สร้างคอมโพเนนต์ใหม่ที่ตั้งไว้บนแพลตฟอร์มให้โดยอัตโนมัติ โดยคอมโพเนนต์ใหม่ จะถูกสร้างขึ้นที่พาท `src/shared/components/dynamic-content` ภายใต้โฟลเดอร์ tmb-score (รูปที่ 73(A)) ประกอบไปด้วยไฟล์ `tmb-score.component.html` และ `tmb-score.component.ts` เป็นต้น และนอกจากนี้คอมโพเนนต์จะถูกเพิ่มเข้า

ไปยังคลาส *DynamicContentModule* ไฟล์ *dynamic-content.module.ts* โดยอัตโนมัติ (รูปที่ 73(B)) เพื่อเปิดการใช้งาน



รูปที่ 70: ตำแหน่งที่เพิ่มไฟล์ใหม่ของดาต้าอานาไลซิสเอพีไอ

```
import time
import json

from flask import Blueprint, request

from api.utils import data_preprocess as preprocessor

tmb_api = Blueprint('tmb_api', __name__)

Endpoint
@tmb_api.route('/calcTmbScore', methods=['POST'])
def calc_tmb_score():
    start_time = time.time()
    data = request.get_json()
    Request
    Response tmb_score_result = preprocessor.tmb_score(data['items'], data['exomeSize'])
    print("--- %s seconds ---" % (time.time() - start_time))
    json_string = json.dumps(tmb_score_result, default=lambda o: o.__dict__, ensure_ascii=False)
    return json_string, 200
```

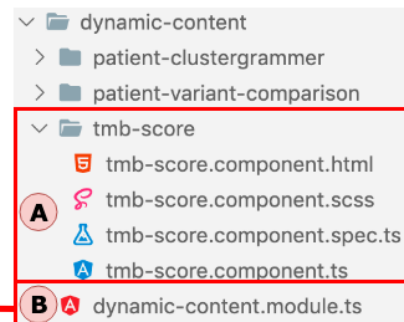
รูปที่ 71: ตัวอย่างการเขียนโปรแกรมเพิ่มส่วนการวิเคราะห์ข้อมูลในไฟล์ *tmb_api.py*

```
neda-MacBook-Pro:onco-web neda$ java -jar dynamic-component-jar-with-dependencies.jar ]
Welcome to RUN-ONCO Dynamic Component Generator.
Please Enter Username : admin
Please Enter Password : ]
Please Enter Component (e.g. tmb-score) : tmb-score
Generate component tmb-score successfully
neda-MacBook-Pro:onco-web neda$
```

รูปที่ 72: การเรียกใช้เครื่องมือเพื่อสร้างคอมโพเนนท์ใหม่และติดตั้งลงบนแพลตฟอร์มให้โดยอัตโนมัติ

```
import { TmbScoreComponent } from "../tmb-score/tmb-score.component";
```

```
@NgModule({
  imports: [
    CommonModule,
    ReactiveFormsModule,
    FormsModule,
    TranslateModule.forChild(),
    StepWizardModule,
    OmicsAnalysisFormModule,
    VariantAnalysisFormModule,
    RnaSeqAnalysisFormModule,
    SharedServicesModule,
    SharedDirectivesModule,
    SharedPipesModule,
    VariantComparisonModule,
    ClustergrammerModule,
    CytoscapeModule,
    IdeogramModule
  ],
  declarations: [
    PatientClustergrammerComponent,
    PatientVariantComparisonComponent,
    TmbScoreComponent
  ],
  entryComponents: [
    PatientClustergrammerComponent,
    PatientVariantComparisonComponent,
    TmbScoreComponent
  ]
})
export class DynamicContentModule {
  static mapping = {
    PatientClustergrammerComponent,
    PatientVariantComparisonComponent,
    TmbScoreComponent
  };
}
```



รูปที่ 73: ผลจากการใช้เครื่องมือ dynamic-component-jar-with-dependencies.jar โปรแกรมดังกล่าวจะสร้างคอมโพเนนท์ tmb-score และเพิ่มรายการ tmb-score เข้าไปในคลาส *DynamicContentModule* โดยอัตโนมัติ (A) ไฟล์ต่าง ๆ จะอยู่ที่พาส src/shared/components/dynamic-content ภายใต้โฟลเดอร์ tmb-score และ (B) ภายในคลาส *DynamicContentModule* จะถูกแทรกด้วยคลาส *TmbScoreComponent* เพื่อเปิดการใช้งาน

- เขียนโปรแกรมส่วนแสดงผลที่ไฟล์ tmb-score.component.html ซึ่งเป็นส่วนที่ตอบโต้กับผู้ใช้ ตัวอย่างเช่น เมื่อผู้ใช้กดปุ่ม “Update” (รูปที่ 74) เมธอด updateVariantParams ก็ทำงาน ซึ่งเมธอดนี้ได้กำหนดไว้ที่คลาส *TmbScoreComponent* ในไฟล์ tmb-score.component.ts จะอยู่ในขั้นตอนถัดไป

Variant Call

VCF

0003_WES Update Click "Update" button

Settings (Exome size(Mb)=37.5) [Download](#)

No. of all variants	14,571	388.56	
No. of variants in protein coding	10,839	289.04	
No. of variants w/ missense	1,993	51.15	●
No. of variants w/ missense w/ vaf>=1%	1,925	51.33	●
No. of variants w/ missense w/ vaf>=5%	299	7.97	●
No. of variants w/ stop_gained	111	2.96	●
No. of variants w/ stop_gained w/ vaf>=1%	109	2.91	●
No. of variants w/ stop_gained w/ vaf>=5%	16	0.43	●
No. of variants w/ frameshift_variant	174	4.64	●
No. of variants w/ frameshift_variant w/ vaf>=1%	166	4.43	●
No. of variants w/ frameshift_variant w/ vaf>=5%	50	1.33	●

tmb-score.component.html

```
<app-variant-analysis-form [isMultiSelect]="false" [criteria]="criteria"
(updatedChange)="updateVariantParams($event)">
</app-variant-analysis-form>
```

```
<table class="table table-bordered margin-top-10">
<tbody>
<tr *ngFor="let tmbScore of tmbScoreResults; trackBy: trackByFn;">
<td>{{ tmbScore.typeDesc }}</td>
<td class="text-right">
{{ tmbScore.numOfVariant | formatCurrency:true:0 }}
</td>
<td class="text-right">
{{ tmbScore.rateOfVariant | formatCurrency }}
</td>
<td class="text-center" style="width: 50px">
<div [hidden]="tmbScore.level === null"
class="circle" [ngClass]="tmbScore.level"
</div>
</td>
</tr>
</tbody>
</table>
```

รูปที่ 74: ตัวอย่างโปรแกรมส่วนแสดงผลในไฟล์ tmb-score.component.html

- เขียนโปรแกรมส่วนควบคุมเพื่อรองรับคำสั่งจากส่วนแสดงผลที่ไฟล์ tmb-score.component.ts เมื่อผู้ใช้เรียกผลการคำนวณค่าที่เอ็มบี ส่วนควบคุมจะเรียก endpoint /sec/getVariantCallsBySampleId ขององโคเซอร์วิสผ่านคลาส DataService เพื่อดึงข้อมูลการแปรผันของดีเอ็นเอก่อน และนำข้อมูลดังกล่าวส่งไปที่ endpoint /calcTmbScore ของดาต้าอานาไลซิสเอพีไอผ่านคลาส DataAnlysService (รูปที่ 75) ซึ่งผลลัพธ์ที่ได้จะถูกส่งไปยังส่วนควบคุม และส่วนแสดงผลตามลำดับ (รูปที่ 76)

tmb-score.component.ts

```
public updateVariantParams(criteria: any): void {
  this.variantParams = criteria;
  this.getVariantCallsBySampleId().then((res: any) => {
    if (res.responseStatus.responseCode === 'ONC-SCC000') {
      this.calcTmbScore(res.data, this.tmbConfig.exomeSize)
        .then((x: any) => {
          this.tmbScoreResults = x;
        });
    }
  });
}

private getVariantCallsBySampleId(): Promise<{}> {
  const p = new Promise((resolve) => {
    this.dataService.connect('sec', 'getVariantCallsBySampleId')
      .then((res: any) => {
        resolve(res);
      });
  });
  return p;
}

private calcTmbScore(variantCalls: any, exomeSize: number): Promise<{}> {
  const req: any = { exomeSize: exomeSize, items: variantCalls };
  const p = new Promise((resolve) => {
    this.dataAnlysService.connect('calcTmbScore', req)
      .subscribe((res: any) => {
        resolve(res);
      });
  });
  return p;
}
```

รูปที่ 75: ตัวอย่างโปรแกรมส่วนควบคุมในไฟล์ tmb-score.component.ts

tmb-score.component.html	tmb-score.component.ts
<pre> <app-variant-analysis-form [isMultiSelect]="false" [criteria]="criteria" (updatedChange)="updateVariantParams(\$event)"> </app-variant-analysis-form> <table class="table table-bordered margin-top-10"> <tbody> <tr *ngFor="let tmbScore of tmbScoreResults; trackBy: trackByFn;"> <td>{{ tmbScore.typeDesc }}</td> <td class="text-right"> {{ tmbScore.numOfVariant formatCurrency:true:0 }} </td> <td class="text-right"> {{ tmbScore.rateOfVariant formatCurrency }} </td> <td class="text-center" style="width: 50px"> <div [hidden]="tmbScore.level === null" class="circle" [ngClass]="tmbScore.level" </div> </td> </tr> </tbody> </table> </pre>	<pre> public updateVariantParams(criteria: any): void { this.variantParams = criteria; this.getVariantCallsBySampleId().then((res: any) => { if (res.responseStatus.responseCode === 'ONC-SCC000') { this.calcTmbScore(res.data, this.tmbConfig.exomeSize) .then((x: any) => { this.tmbScoreResults = x; }); } }); } </pre>

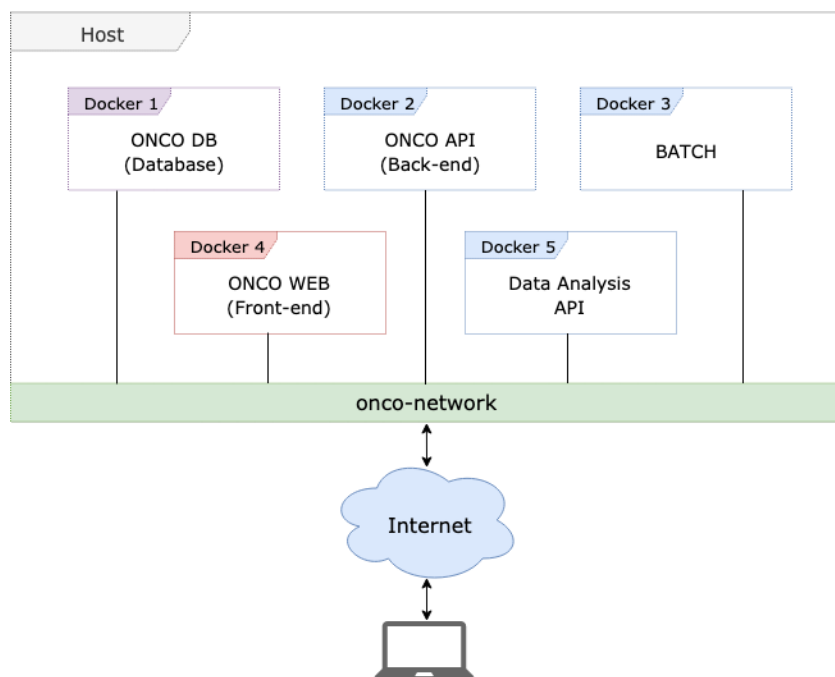
รูปที่ 76: การทำงานร่วมกันระหว่างส่วนแสดงผลและส่วนควบคุม

4.8 เครื่องมือที่ใช้ในการออกแบบและพัฒนาระบบ

ซอฟต์แวร์แพลตฟอร์มถูกพัฒนาในรูปแบบเว็บแอปพลิเคชัน โดยพิจารณาจากข้อดีดังนี้ 1) ผู้ใช้สามารถเข้าใช้งานระบบผ่านเว็บเบราว์เซอร์ ซึ่งติดตั้งมาพร้อมกับระบบปฏิบัติการ 2) ระบบจะถูกติดตั้งไว้ที่เซิร์ฟเวอร์ส่วนกลาง ซึ่งง่ายต่อการบริหารจัดการ เช่น การติดตั้งโปรแกรมชุดใหม่ การสำรองข้อมูล เป็นต้น นอกจากนี้ระบบถูกออกแบบให้แยกการทำงานของส่วนเว็บ (front-end) ออกจากส่วนเซิร์ฟเวอร์ (back-end) ช่วยลดความซับซ้อนในการพัฒนาระบบ และยังช่วยให้การติดตั้งระบบรวดเร็วขึ้น

เพื่อให้เว็บมีความยืดหยุ่นในการเพิ่มเซิร์ฟเวอร์วิเคราะห์ใหม่ จำเป็นต้องเลือกภาษาโปรแกรมที่สนับสนุนสถาปัตยกรรมเชิงคอมโพเนนต์ ได้แก่ แองกูลาร์ (angular) รีแอค (react) และ วิวเจเอส (vue.js) เป็นต้น แต่มีเพียงแองกูลาร์ที่สนับสนุนแบบรูปโมเดล-วิว-คอนโทรลเลอร์ จึงพัฒนาเว็บด้วยแองกูลาร์ เวอร์ชัน 7.2 ในการพัฒนาเซิร์ฟเวอร์ ได้นำสปริงบูท (spring boot) เวอร์ชัน 1.5.6 ซึ่งเป็นเฟรมเวิร์กที่พัฒนาต่อยอดจากสปริงเฟรมเวิร์ก (spring framework) โดยสนับสนุนแบบรูปดีเพนเดนซีอินเจกชัน นอกจากนี้ยังมีไลบรารีพื้นฐานที่ติดตั้งให้พร้อมใช้งาน (starter library) และการกำหนดค่าต่าง ๆ ให้อัตโนมัติ (auto configuration) เป็นต้น ซึ่งลดความยุ่งยากในการพัฒนาโปรแกรมเนื่องจากสปริงบูทเขียนด้วยภาษาจาวา (java) โปรแกรมจะทำงานบนเจวีเอ็ม (Java Virtual Machine: JVM) จึงปลอดภัย สำหรับตัวแปลงข้อมูลระหว่างโมเดลเชิงวัตถุและเชิงสัมพันธ์ ได้นำไฮเบอร์เนต (hibernate) เวอร์ชัน 5.2.10 ใช้งานร่วมกับสปริงบูท ซึ่งเป็นตัวกลางที่เชื่อมความสัมพันธ์ระหว่างอ็อบเจกต์และตารางข้อมูล โดยจะสร้างและประมวลผลเอสคิวแอล (Structured Query Language: SQL) แบบอัตโนมัติจากเมธอดต่าง ๆ ของคลาส ไฮเบอร์เนตยังเป็นเฟรมเวิร์กที่ได้รับความนิยม รองรับการเชื่อมต่อกับฐานข้อมูลหลายหลากยี่ห้อ การเปลี่ยนยี่ห้อของฐานข้อมูลจะไม่ส่ง

ผลกระทบใด ๆ ก็เกิดขึ้น โดยฐานข้อมูลเชิงสัมพันธ์ที่ใช้คือ ไมยเอสคิวแอล (MySQL) เวอร์ชัน 5.7.21 เป็นฐานข้อมูลที่ไม่ต้องเสียค่าใช้จ่ายและนิยมใช้งานอย่างแพร่หลาย โดยรองรับการจัดการทรานแซคชันและการควบคุมภาวะพร้อมกัน (concurrency) นอกจากนี้แพลตฟอร์มสามารถติดตั้งได้โดยง่ายผ่านดอกเกอร์ ดังรูปที่ 77



รูปที่ 77: ภาพรวมการติดตั้งแพลตฟอร์ม

บทที่ 5

สรุปผลการวิจัย

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอซอฟต์แวร์แพลตฟอร์มแบบบูรณาการเพื่อการแพทย์แม่นยำของโรคมะเร็ง ซึ่งได้รับการออกแบบและพัฒนาเพื่ออำนวยความสะดวกให้ศูนย์วิจัยโรคมะเร็งและโรงพยาบาล ด้วยบริการการจัดเก็บข้อมูลทางคลินิก ข้อมูลทางพยาธิวิทยา ข้อมูลชีววัตถุ และนำข้อมูลเหล่านี้มาวิเคราะห์ร่วมกับข้อมูลโอมิกส์ต่าง ๆ รวมทั้งข้อมูลจากฐานข้อมูลสาธารณะที่เกี่ยวข้อง เพื่อช่วยในการวินิจฉัยและออกแบบแนวทางการรักษาโรคมะเร็งที่มีความจำเพาะกับผู้ป่วยแต่ละราย โดยมีการติดตั้งปลั๊กอินการวิเคราะห์และแสดงผลเชิงรูปภาพบนแพลตฟอร์มด้วยสถาปัตยกรรมที่นำเสนอ เช่น ดาต้าอนาไลซิสเอพีไอ และการติดตั้งปลั๊กอินแบบไดนามิก ซึ่งยืดหยุ่นในการติดตั้งและสามารถปรับแต่งได้ง่าย ดังที่ได้แสดงไว้ในตัวอย่างการวิเคราะห์ข้อมูลประกอบด้วย การเปรียบเทียบการแปรผันของดีเอ็นเอจากชีววัตถุ 2 ชนิดของผู้ป่วยรายเดียวกันและการวิเคราะห์ระดับการแสดงออกของยีนมะเร็งด้วย clustergrammer.js และแสดงการมีปฏิสัมพันธ์ระหว่างคู่โปรตีนจากฐานข้อมูลสตริงด้วย cystoscope.js นอกจากนี้สามารถติดตั้งแพลตฟอร์มได้โดยง่ายด้วยด็อกเกอร์

5.2 แนวทางวิจัยในอนาคต

สำหรับการศึกษาต่อไปจะพัฒนาออกแบบฐานข้อมูลให้รองรับการปรับแต่งฟิลต์จากนักพัฒนาและความหลากหลายของข้อมูลโอมิกส์ รวมทั้งการตรวจสอบเวอร์ชันของจีโนมอ้างอิง (reference genome) ในข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอ นอกจากนี้การศึกษากฎหมายคุ้มครองข้อมูลส่วนบุคคลเพื่อพัฒนาออกแบบฐานข้อมูลให้รองรับเป็นสิ่งจำเป็น และนำเสนอสำหรับงานวิจัยเพิ่มเติม



ภาคผนวก ก.

ตารางฐานข้อมูล (Database) ของซอฟต์แวร์แพลตฟอร์ม RUN-ONCO

พจนานุกรมข้อมูล (Data Dictionary)

1. TB_C_CHANNEL

Column Name	Data Type	Description
CHANNEL_ID	INT(11)	Channel ID
CHANNEL_CODE	VARCHAR(50)	Channel code
CHANNEL_NAME	VARCHAR(100)	Channel name
STATUS	VARCHAR(50)	Status
SYSTEM_ID	INT(11)	System ID

2. TB_C_DATA_ANALYS

Column Name	Data Type	Description
PARAM_ID	INT(11)	Parameter ID
PARAM_CODE	VARCHAR(50)	The code of the parameter
PARAM_NAME	VARCHAR(100)	The name of the parameter
PARAM_TYPE	VARCHAR(30)	The type of the parameter
SEQ_NO	INT(11)	Sequence number
CREATE_USER_ID	BIGINT(20)	Create user
CREATE_DATE	DATETIME	Create date
UPDATE_USER_ID	BIGINT(20)	Update user
UPDATE_DATE	DATETIME	Update date

3. TB_C_DATA_VERSION

Column Name	Data Type	Description
DATA_VER_ID	INT(11)	Data version ID
DATA_DATE	DATE	Data date
DATA_TYPE	VARCHAR(50)	Data type

3. TB_C_DATA_VERSION (ต่อ)

Column Name	Data Type	Description
VERSION	VARCHAR(20)	Data version
IS_DEFAULT	VARCHAR(1)	Default flag
CREATE_USER_ID	BIGINT(20)	Create user
CREATE_DATE	DATETIME	Create date
UPDATE_USER_ID	BIGINT(20)	Update user
UPDATE_DATE	DATETIME	Update date

4. TB_C_DYNAMIC_CONTENT

Column Name	Data Type	Description
COMP_ID	INT(11)	Component ID
ANALYSIS_NAME	VARCHAR(100)	Analysis name
COMPONENT_NAME	VARCHAR(255)	Component name
INPUT_FIELD	VARCHAR(100)	Input field
MODULE_NAME	VARCHAR(255)	Module name
MODULE_PATH	VARCHAR(1000)	Module path
DATA_TYPE_ANALYS	VARCHAR(50)	Data type analysis
STATUS	VARCHAR(50)	Status
CREATE_USER_ID	BIGINT(20)	Create user
CREATE_DATE	DATETIME	Create date
UPDATE_USER_ID	BIGINT(20)	Update user
UPDATE_DATES	DATETIME	Update date

5. TB_C_MENU

Column Name	Data Type	Description
MENU_ID	INT(11)	Menu ID
MENU_CODE	VARCHAR(50)	Menu code
MENU_NAME	VARCHAR(100)	Menu name
ROUTER_LINK	VARCHAR(150)	Router link

5. TB_C_MENU (ต่อ)

Column Name	Data Type	Description
SEQ_NO	INT(11)	Sequence number
STATUS	VARCHAR(50)	Status
CREATE_USER_ID	BIGINT(20)	Create user
CREATE_DATE	DATETIME	Create date
UPDATE_USER_ID	BIGINT(20)	Update user
UPDATE_DATE	DATETIME	Update date

6. TB_C_PARAM

Column Name	Data Type	Description
PARAM_ID	INT(11)	Parameter ID
PARAM_CODE	VARCHAR(2)	Parameter code
PARAM_NAME	VARCHAR(30)	Parameter name
PARAM_TYPE	VARCHAR(30)	Parameter type
SEQ_NO	INT(11)	Sequence number
STATUS	VARCHAR(50)	Status
CREATE_USER_ID	BIGINT(20)	Create user
CREATE_DATE	DATETIME	Create date
UPDATE_USER_ID	BIGINT(20)	Update user
UPDATE_DATE	DATETIME	Update date

7. TB_C_ROLE

Column Name	Data Type	Description
ROLE_ID	INT(11)	Role ID
ROLE_CODE	VARCHAR(50)	Role code
ROLE_NAME	VARCHAR(100)	Role name
SEQ_NO	INT(11)	Sequence number
STATUS	VARCHAR(50)	Status

7. TB_C_ROLE (ต่อ)

Column Name	Data Type	Description
CREATE_USER_ID	BIGINT(20)	Create user
CREATE_DATE	DATETIME	Create date
UPDATE_USER_ID	BIGINT(20)	Update user
UPDATE_DATE	DATETIME	Update Date

8. TB_C_SCREEN

Column Name	Data Type	Description
SCREEN_ID	INT(11)	Screen ID
SCREEN_CODE	VARCHAR(50)	Screen code
CREATE_DATE	DATETIME	Create date
SCREEN_NAME	VARCHAR(200)	Screen name
ROUTER_LINK	VARCHAR(150)	Router link
STATUS	VARCHAR(255)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user
MENU_ID	INT(11)	Menu ID
SEQ_NO	INT(11)	Sequence number

9. TB_C_SYSTEM

Column Name	Data Type	Description
SYSTEM_ID	INT(11)	System ID
SYSTEM_CODE	VARCHAR(50)	System code
SYSTEM_NAME	VARCHAR(100)	System name
STATUS	VARCHAR(50)	Status

10. TB_M_ACTIONABLE_VARIANT

Column Name	Data Type	Description
ACT_VAR_ID	INT(11)	Actionable variant ID
ALTERATION	VARCHAR(100)	Alteration
CANCER_TYPE	VARCHAR(100)	Cancer type
DRUG_ABSTRACT	VARCHAR(1000)	Drug abstract
DRUG_PMIDS	VARCHAR(500)	Drug PMIDS
DRUGS	VARCHAR(500)	Drugs
ENTREZ_GENE_ID	VARCHAR(50)	Entrez gene ID
GENE	VARCHAR(50)	Gene
ISOFORM	VARCHAR(20)	Isoform
LEVEL	VARCHAR(2)	Level
PROTEIN_CHANGE	VARCHAR(100)	Protein change
REF_SEQ	VARCHAR(20)	Ref Seq
DATA_VER_ID	INT(11)	Data version ID

11. TB_M_BLOOD_TYPE

Column Name	Data Type	Description
BLOOD_TYPE_ID	INT(11)	Blood type ID
BLOOD_TYPE_CODE	VARCHAR(2)	Blood type code
CREATE_DATE	DATETIME	Create date
BLOOD_TYPE_NAME	VARCHAR(30)	The name of the blood type
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

12. TB_M_CANCER_GENE

Column Name	Data Type	Description
GENE_ID	INT(11)	Gene ID
CREATE_DATE	DATETIME	Create date
GENE_NAME	VARCHAR(250)	Gene name
UPDATE_DATE	DATETIME	Update date
GRP_GENE_ID	INT(11)	Group gene ID
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

13. TB_M_CANCER_GENE_GROUP

Column Name	Data Type	Description
GRP_GENE_ID	INT(11)	Cancer gene group ID
CREATE_DATE	DATETIME	Create date
GRP_GENE_NAME	VARCHAR(255)	The name of the cancer gene group
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

14. TB_M_CELL_LINE_TYPE

Column Name	Data Type	Description
CL_TYPE_ID	INT(11)	Cell line type ID
CL_TYPE_CODE	VARCHAR(3)	The code of the cell line type
CREATE_DATE	DATETIME	Create date
CL_TYPE_NAME	VARCHAR(20)	The name of the cell line type
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

15. TB_M_CELL_TYPE

Column Name	Data Type	Description
CELL_TYPE_ID	INT(11)	Cell type ID
CELL_TYPE_CODE	VARCHAR(2)	The code of the cell type
CREATE_DATE	DATETIME	Create date
CELL_TYPE_NAME	VARCHAR(20)	The name of the cell type
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

16. TB_M_FREEZE_METHOD

Column Name	Data Type	Description
FREEZE_METH_ID	INT(11)	Freezing method ID
FREEZE_METH_CODE	VARCHAR(2)	The code of freezing method
CREATE_DATE	DATETIME	Create date
FREEZE_METH_NAME	VARCHAR(20)	The name of freezing method
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

17. TB_M_GENDER

Column Name	Data Type	Description
GENDER_ID	INT(11)	Gender ID
GENDER_CODE	VARCHAR(2)	The code of the gender
CREATE_DATE	DATETIME	Create date
GENDER_NAME	VARCHAR(20)	The name of the gender
SEQ_NO	INT(11)	Sequence number

17. TB_M_GENDER (ต่อ)

Column Name	Data Type	Description
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

18. TB_M_ICD

Column Name	Data Type	Description
ICD_ID	INT(11)	ICD ID
ICD_CODE	VARCHAR(20)	ICD code
ICD_DESC	VARCHAR(500)	The description of the ICD
ICD_NAME	VARCHAR(150)	The name of the ICD
DATA_VER_ID	INT(11)	Date version ID

19. TB_M_ICD_O

Column Name	Data Type	Description
ICD_O_ID	INT(11)	ICD-O ID
ICD_O_CODE	VARCHAR(20)	The code of the ICD-O
ICD_O_DESC	VARCHAR(500)	The description of the ICD-O
ICD_O_NAME	VARCHAR(150)	The name of the ICD-O
DATA_VER_ID	INT(11)	Data version ID

20. TB_M_PATIENT

Column Name	Data Type	Description
PATIENT_ID	BIGINT(20)	Patient ID
BIRTH_DATE	DATE	Birth date
CREATE_DATE	DATETIME	Create date
DEATH_DATE	DATE	Death date
HEIGHT	DECIMAL(5,2)	Height

20. TB_M_PATIENT (ต่อ)

Column Name	Data Type	Description
HN	VARCHAR(20)	Hospital number
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
WEIGHT	DECIMAL(5,2)	Weight
CREATE_USER_ID	BIGINT(20)	Create user
GENDER_ID	INT(11)	Gender ID
RACE_ID	INT(11)	Race ID
UPDATE_USER_ID	BIGINT(20)	Update user

21. TB_M_PRESERVE_METHOD

Column Name	Data Type	Description
PRESERV_METH_ID	INT(11)	Preserve method ID
PRESERV_METH_CODE	VARCHAR(2)	The code of the Preserve method
CREATE_DATE	DATETIME	Create date
PRESERV_METH_NAME	VARCHAR(20)	The name of preserve method
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

22. TB_M_RACE

Column Name	Data Type	Description
RACE_ID	INT(11)	Race ID
CODE	VARCHAR(2)	The code of the race
CREATE_DATE	DATETIME	Create date
NAME	VARCHAR(255)	The name of the race
STATUS	VARCHAR(50)	Status

22. TB_M_RACE (ต่อ)

Column Name	Data Type	Description
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

23. TB_M_SAMPLE_TYPE

Column Name	Data Type	Description
SMP_TYPE_ID	INT(11)	Sample type ID
SMP_TYPE_CODE	VARCHAR(2)	The code of the sample type
CREATE_DATE	DATETIME	Create date
SMP_TYPE_NAME	VARCHAR(30)	The name of the sample type
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

24. TB_M_SEQ_TYPE

Column Name	Data Type	Description
SEQ_TYPE_ID	INT(11)	Sequencing type ID
SEQ_TYPE_CODE	VARCHAR(2)	The code of the sequencing type
CREATE_DATE	DATETIME	Create date
SEQ_TYPE_NAME	VARCHAR(20)	The name of the sequencing type
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

25. TB_M_TISSUE_PROCEDURE

Column Name	Data Type	Description
TISSUE_PROC_ID	INT(11)	Tissue procedure ID
TISSUE_PROC_CODE	VARCHAR(2)	The code of the tissue procedure
CREATE_DATE	DATETIME	Create date
TISSUE_PROC_NAME	VARCHAR(30)	The name of the tissue procedure
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

26. TB_M_TISSUE_TYPE

Column Name	Data Type	Description
TISSUE_TYPE_ID	INT(11)	Tissue type ID
TISSUE_TYPE_CODE	VARCHAR(2)	The code of the tissue type
CREATE_DATE	DATETIME	Create date
TISSUE_TYPE_NAME	VARCHAR(30)	The name of the tissue type
STATUS	VARCHAR(50)	Status
DATA_TYPE	VARCHAR(1)	Data type
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

27. TB_M_USER

Column Name	Data Type	Description
USER_ID	BIGINT(20)	User ID
CREATE_DATE	DATETIME	Create date
FIRST_NAME	VARCHAR(255)	First name
LAST_NAME	VARCHAR(255)	Last name
PASSWORD	VARCHAR(100)	Password

27. TB_M_USER (ต่อ)

Column Name	Data Type	Description
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
USERNAME	VARCHAR(50)	Username
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

28. TB_M_VITAL_STATUS

Column Name	Data Type	Description
VITAL_STATUS_ID	INT(11)	Vital status ID
VITAL_STATUS_CODE	VARCHAR(2)	The code of the vital status
CREATE_DATE	DATETIME	Create date
VITAL_STATUS_NAME	VARCHAR(20)	The name of the vital status
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
UPDATE_USER_ID	BIGINT(20)	Update user

29. TB_T_ATTACHMENT

Column Name	Data Type	Description
ATTACHMENT_ID	BIGINT(20)	Attachment ID
CONTENT_TYPE	VARCHAR(255)	Content type
CREATE_DATE	DATETIME	Create date
FILE_NAME	VARCHAR(255)	File name
FILE_PATH	VARCHAR(255)	File path
FILE_SIZE	BIGINT(20)	File size
STATUS	VARCHAR(50)	Status
UUID	VARCHAR(255)	UUID
CREATE_USER_ID	BIGINT(20)	Create user

30. TB_T_BIOSPECIMEN

Column Name	Data Type	Description
BIOSPECIMEN_ID	BIGINT(20)	Biospecimen ID
COLLECT_SAMPLE_DATE	DATE	The date of collect sample
CREATE_DATE	DATETIME	Create date
FREEZE_DATE	DATE	Freeze date
INITIAL_WEIGHT	DECIMAL(19,2)	Initial weight
REF_NO	VARCHAR(20)	Reference number
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
BLOOD_TYPE_ID	INT(11)	Blood type ID
CELL_LINE_TYPE_ID	INT(11)	Cell line type ID
CELL_TYPE_ID	INT(11)	Cell type ID
CREATE_USER_ID	BIGINT(20)	Create user
FREEZE_METHOD_ID	INT(11)	Freezing method ID
PATHOLOGICAL_ID	BIGINT(20)	Pathological ID
PATIENT_ID	BIGINT(20)	Patient ID
PRESERVE_METHOD_ID	INT(11)	Preserve method ID
SAMPLE_TYPE_ID	INT(11)	Sample type ID
TISSUE_TYPE_ID	INT(11)	Tissue type ID
UPDATE_USER_ID	BIGINT(20)	Update user

31. TB_T_CANCER_STAGE

Column Name	Data Type	Description
STAGE_ID	BIGINT(20)	Stage ID
AJCC	INT(11)	AJCC
CN_STAGE	INT(11)	cN stage
CN_SUBSTAGE	INT(11)	cN substage
CT_STAGE	INT(11)	cT stage
CT_SUBSTAGE	INT(11)	cT substage

31. TB_T_CANCER_STAGE (ต่อ)

Column Name	Data Type	Description
CLINICAL_STAGE	INT(11)	Clinical stage
M_STAGE	INT(11)	M stage
M_SUBSTAGE	INT(11)	M substage
PN_STAGE	INT(11)	pN stage
PN_SUBSTAGE	INT(11)	pN substage
PT_STAGE	INT(11)	pT stage
PT_SUBSTAGE	INT(11)	pT substage
PATHO_SUBSTAGE	INT(11)	Pathological substage
PATHOLOGICAL_STAGE	INT(11)	Pathological stage
STAGE_TYPE	INT(11)	Stage type
YPN_STAGE	INT(11)	ypN stage
YPN_SUBSTAGE	INT(11)	ypN substage
YPT_STAGE	INT(11)	ypT stage
YPT_SUBSTAGE	INT(11)	ypT substage
DIAGNOSIS_ID	BIGINT(20)	Diagnosis ID

32. TB_T_CLINICAL_COURSE

Column Name	Data Type	Description
COURSE_ID	BIGINT(20)	Course ID
RECORD_DATE	DATE	The date of record
RECURRENCE_DATE	DATE	The date of recurrence
RECURRENCE_STATUS	INT(11)	Recurrence status
STATUS	VARCHAR(50)	Status
DIAGNOSIS_ID	BIGINT(20)	Diagnosis ID

33. TB_T_CLINICAL_DATA

Column Name	Data Type	Description
CLINICAL_DATA_ID	BIGINT(20)	Clinical data ID
CREATE_DATE	DATETIME	Create date
REF_NO	VARCHAR(20)	Reference number
STATUS	VARCHAR(50)	Status
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
PATIENT_ID	BIGINT(20)	Patient ID
UPDATE_USER_ID	BIGINT(20)	Update user

34. TB_T_DIAGNOSIS

Column Name	Data Type	Description
DIAGNOSIS_ID	BIGINT(20)	Diagnosis ID
DIAGNOSIS_DATE	DATE	Diagnosis date
PRIMARY_DX_CODE	VARCHAR(20)	The code of the primary diagnosis
PRIMARY_DX_NAME	VARCHAR(150)	The name of the primary diagnosis
STATUS	VARCHAR(50)	Status
CANCER_STAGE_ID	BIGINT(20)	Cancer stage ID
CLINICAL_DATA_ID	BIGINT(20)	Clinical data ID

35. TB_T_PATHOLOGICAL

Column Name	Data Type	Description
PATHOLOGICAL_ID	BIGINT(20)	Pathological ID
HISTOLOGIC_GRADE	INT(11)	Histologic grade
PATHO_NO	VARCHAR(255)	Pathology number
RESECTION_BIOPSY_SITE	VARCHAR(255)	Resection biopsy site
STATUS	VARCHAR(50)	Status
TISSUE_REPORT_DATE	DATE	Tissue report date
DIAGNOSIS_ID	BIGINT(20)	Diagnosis ID

35. TB_T_PATHOLOGICAL (ต่อ)

Column Name	Data Type	Description
TISSUE_PROCEDURE_ID	INT(11)	Tissue procedure ID
TISSUE_TYPE_ID	INT(11)	Tissue type ID
MORPHOLOGY_CODE	VARCHAR(20)	The code of the morphology
MORPHOLOGY_NAME	VARCHAR(150)	The name of the morphology

36. TB_T_RNA_SEQ

Column Name	Data Type	Description
RNA_SEQ_ID	BIGINT(20)	RNA-Seq ID
COVERAGE	DECIMAL(12,6)	Coverage
END	INT(11)	End
FPKM	DECIMAL(12,6)	FPKM
GENE_ID	VARCHAR(100)	GENE ID
GENE_NAME	VARCHAR(250)	GENE name
REF	VARCHAR(250)	Reference
START	INT(11)	Status
STRAND	VARCHAR(10)	Strand
TPM	DECIMAL(12,6)	TPM
SAMPLE_RNA_SEQ_ID	BIGINT(20)	Sample RNA-Seq ID

37. TB_T_ROLE_DYNA_CONTENT

Column Name	Data Type	Description
ROLE_ID	INT(11)	Role ID
COMP_ID	INT(11)	Dynamic component ID

38. TB_T_ROLE_SCREEN

Column Name	Data Type	Description
ROLE_ID	INT(11)	Role ID
SCREEN_ID	INT(11)	Screen ID

39. TB_T_SAMPLE_RNA_SEQ

Column Name	Data Type	Description
SAMPLE_RNA_SEQ_ID	BIGINT(20)	Sample RNA-Seq ID
CREATE_DATE	DATETIME	Create date
IMPORT_DATA_DATE	DATETIME	Import date date
STATUS	VARCHAR(50)	Status
ATTACHMENT_ID	BIGINT(20)	Attachment ID
BIOSPECIMEN_ID	BIGINT(20)	Biospecimen ID
CREATE_USER_ID	BIGINT(20)	Create user
PATIENT_ID	BIGINT(20)	Patient ID

40. TB_T_SAMPLE_VARIANT

Column Name	Data Type	Description
SAMPLE_VAR_ID	BIGINT(20)	Sample variant ID
ANNOTATED_DATE	DATETIME	Annotated date
CREATE_DATE	DATETIME	Create date
IMPORT_DATA_DATE	DATETIME	Import date date
STATUS	VARCHAR(50)	Status
ATTACHMENT_ID	BIGINT(20)	Attachment ID
BIOSPECIMEN_ID	BIGINT(20)	Biospecimen ID
CREATE_USER_ID	BIGINT(20)	Create user
PATIENT_ID	BIGINT(20)	Patient ID
SEQUENCE_TYPE_ID	INT(11)	Sequencing type ID

41. TB_T_SEQUENCE

Column Name	Data Type	Description
SEQUENCE_ID	BIGINT(20)	Sequence ID
CREATE_DATE	DATE	Create date
DATA_TYPE	VARCHAR(1)	Data type

41. TB_T_SEQUENCE (ต่อ)

Column Name	Data Type	Description
SEQUENCE_NO	INT(11)	Sequence number
UPDATE_DATE	DATETIME	Update date
VERSION	INT(11)	Sequence version

42. TB_T_SURVIVAL_FOLLOWUP

Column Name	Data Type	Description
FOLLOWUP_ID	BIGINT(20)	Survival follow up ID
DEATH_DATE	DATE	Death date
LAST_FOLLOWUP_DATE	DATE	Last follow up date
RECORD_DATE	DATE	Record date
STATUS	VARCHAR(50)	Status
PATIENT_ID	BIGINT(20)	Patient ID
VITAL_STATUS_ID	INT(11)	Vital status ID

43. TB_T_USER_ROLE

Column Name	Data Type	Description
USER_ID	BIGINT(20)	User ID
ROLE_ID	INT(11)	Role ID

44. TB_T_VARIANT_ANNOTATION

Column Name	Data Type	Description
VAR_ANN_ID	BIGINT(20)	Variant annotation ID
ALLELE	VARCHAR(1000)	Allele
AMINO_ACIDS	VARCHAR(100)	Amino acids
BIOTYPE	VARCHAR(100)	Biotype
CDNA_POSITION	VARCHAR(100)	CDNA position
CDS_POSITION	VARCHAR(100)	CDS position
CODONS	VARCHAR(255)	Codons

44. TB_T_VARIANT_ANNOTATION (ต่อ)

Column Name	Data Type	Description
CONSEQUENCE	VARCHAR(100)	Consequence
CREATE_DATE	DATETIME	Create date
DISTANCE	VARCHAR(100)	Distance
EXISTING_VARIANTION	VARCHAR(500)	Existing variantion
EXON	VARCHAR(100)	Exon
FEATURE	VARCHAR(100)	Feature
FEATURE_TYPE	VARCHAR(100)	Feature type
FLAGS	VARCHAR(255)	Flags
GENE	VARCHAR(250)	Gene
HGNC_ID	VARCHAR(100)	HGNC ID
HGVSC	VARCHAR(255)	HGVSC
HGVSP	VARCHAR(255)	HGVSP
IMPACT	VARCHAR(100)	Impact
INTRON	VARCHAR(100)	Intron
ONCOKB_ANNOTATED	VARCHAR(1)	OncoKB annotated
PROTEIN_POSITION	VARCHAR(100)	Protein position
STRAND	VARCHAR(100)	Strand
SYMBOL	VARCHAR(100)	Symbol
SYMBOL_SOURCE	VARCHAR(100)	Symbol source
UPDATE_DATE	DATETIME	Update date
CREATE_USER_ID	BIGINT(20)	Create user
SAMPLE_VAR_ID	BIGINT(20)	Sample variant ID
UPDATE_USER_ID	BIGINT(20)	Update user
VAR_ID	BIGINT(20)	Variant call ID

45. TB_T_VARIANT_CALL

Column Name	Data Type	Description
VAR_ID	BIGINT(20)	Variant call ID
ALT_ALLELE	VARCHAR(1000)	Alt allele
CHROMOSOME	VARCHAR(100)	Chromosome
ECNT	DECIMAL(20,5)	ECNT
FILTER	VARCHAR(10)	Filter
HCNT	DECIMAL(20,5)	HCNT
MAX_ED	DECIMAL(20,5)	Max ED
MIN_ED	DECIMAL(20,5)	Min ED
NLOD	DECIMAL(20,5)	NLOD
POSITION	BIGINT(20)	Position
QUALITY	DECIMAL(20,5)	Quality
REF_ALLELE	VARCHAR(1000)	Reference allele
TLOD	DECIMAL(20,5)	TLOD
VARIANT	VARCHAR(50)	Variant
SAMPLE_VAR_ID	BIGINT(20)	Sample variant ID

46. TB_T_VARIANT_RESULT

Column Name	Data Type	Description
VAR_RES_ID	BIGINT(20)	Variant result ID
ALLELE_FREQUENCY	DECIMAL(20,5)	Allele frequency
ALT_ALLELE_DEPTH	DECIMAL(20,5)	The alteration's allele depth
ALT_ALLELE_F1R2	DECIMAL(20,5)	The alteration's allele F1R2
ALT_ALLELE_F2R1	DECIMAL(20,5)	The alteration's allele F2R1
ALT_ALLELE_QSS	DECIMAL(20,5)	The alteration's allele QSS
GENOTYPE	VARCHAR(10)	Genotype
OXOG_READ_COUNTS	DECIMAL(20,5)	OxoG read counts
REF_ALLELE_DEPTH	DECIMAL(20,5)	The reference's allele depth
REF_ALLELE_F1R2	DECIMAL(20,5)	The reference's allele F1R2

46. TB_T_VARIANT_RESULT (ต่อ)

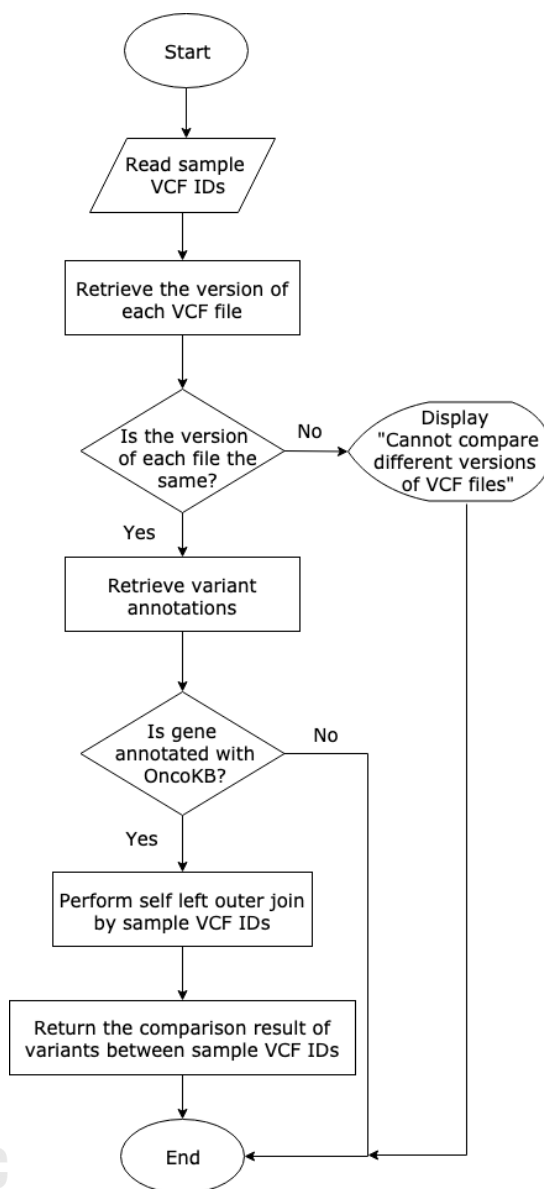
Column Name	Data Type	Description
REF_ALLELE_F2R1	DECIMAL(20,5)	The reference's allele F2R1
REF_ALLELE_QSS	DECIMAL(20,5)	The reference's allele QSS
SAMPLE_TYPE	VARCHAR(1)	Sample type
SAMPLE_VAR_ID	BIGINT(20)	Sample variant ID
VAR_ID	BIGINT(20)	Variant call ID



ภาคผนวก ข.

ขั้นตอนการเปรียบเทียบการแปรผันของดีเอ็นเอระหว่างซีวัตถุ 2 ชนิด

ในส่วนนี้จะอธิบายขั้นตอนการทำงานของตัวอย่างการวิเคราะห์ที่ได้จากข้อมูลการแปรผันของลำดับเบสบนดีเอ็นเอและระดับการแสดงออกของยีน ซึ่งผลการวิเคราะห์ที่ได้จะถูกนำมาแสดงผลข้อมูลแบบตารางและเชิงรูปภาพต่อไป รูปที่ ข.1 แสดงขั้นตอนการเปรียบเทียบการแปรผันของลำดับเบสบนดีเอ็นเอระหว่างซีวัตถุ 2 ชนิดขึ้นไป เริ่มจากดึงข้อมูลตัวอย่างเก็บตรวจด้วยเลขที่ตัวอย่างเก็บตรวจ (sample VCF ID) โดยจะตรวจสอบเวอร์ชันของไฟล์วีซีเอฟว่าตรงกันหรือไม่ ถ้าเวอร์ชันตรงกันระบบส่งคำสั่ง SELECT เพื่อดึงข้อมูลจากราย TB_T_VARIANT_ANNOTATION โดยมีเงื่อนไขว่าต้องตรวจพบข้อมูลยีนดังกล่าวบนฐานข้อมูลยีนโคเคบี และทำการเปรียบเทียบตัวอย่างเก็บตรวจทั้งสองชนิดว่ามีการแปรผันของลำดับเบสบนดีเอ็นเอตำแหน่งใดบ้าง ถ้าเวอร์ชันของไฟล์ข้อมูลที่นำมาเปรียบเทียบไม่ตรงกัน ระบบจะแสดงข้อความเตือน “Cannot compare different versions of VCF files.” รูปที่ ข.2 แสดงตัวอย่างคำสั่ง SELECT ที่ใช้ในการดึงข้อมูลเพื่อเปรียบเทียบการแปรผันของลำดับเบสบนดีเอ็นเอระหว่างซีวัตถุ 2 ชนิด



รูปที่ ข.1: แผนภาพแสดงลำดับขั้นตอนการเปรียบเทียบการแปรผันของลำดับเบสบนดีเอ็นเอระหว่าง

ชีววัตถุ 2 ชนิดขึ้นไป


```

SELECT 'Y' as ONCOKB_ANNOTATED,
CONCAT(var.CHROMOSOME,':',var.POSITION,' ', var.REF_ALLELE, ' / ', var.ALT_ALLELE) as VARIANT,
var.*
FROM (
SELECT CAST(SUBSTRING(var.CHROMOSOME, 4) as SIGNED) as CHROMOSOME,
var.POSITION, var.REF_ALLELE, var.ALT_ALLELE, var.SYMBOL,
v1.SAMPLE_VAR_ID as SPECIMEN_1,
v2.SAMPLE_VAR_ID as SPECIMEN_2
FROM (
SELECT
DISTINCT v.CHROMOSOME,
v.POSITION,
v.REF_ALLELE,
v.ALT_ALLELE, a.SYMBOL
FROM TB_T_VARIANT_CALL v
INNER JOIN TB_T_SAMPLE_VARIANT s ON s.SAMPLE_VAR_ID = v.SAMPLE_VAR_ID
INNER JOIN TB_T_VARIANT_ANNOTATION a ON a.VAR_ID = v.VAR_ID
WHERE s.SAMPLE_VAR_ID IN ( '1','2' ) AND a.ONCOKB_ANNOTATED = 'Y'
) var
LEFT JOIN TB_T_VARIANT_CALL v1 ON v1.CHROMOSOME = var.CHROMOSOME AND v1.POSITION = var.POSITION
AND v1.REF_ALLELE = var.REF_ALLELE AND v1.ALT_ALLELE = var.ALT_ALLELE AND v1.SAMPLE_VAR_ID = 1
LEFT JOIN TB_T_VARIANT_CALL v2 ON v2.CHROMOSOME = var.CHROMOSOME AND v2.POSITION = var.POSITION
AND v2.REF_ALLELE = var.REF_ALLELE AND v2.ALT_ALLELE = var.ALT_ALLELE AND v2.SAMPLE_VAR_ID = 2
) var
ORDER BY var.CHROMOSOME asc, var.POSITION asc

```

รูปที่ ข.2: ตัวอย่างคำสั่ง SELECT เพื่อดึงข้อมูลและเปรียบเทียบการแปรผันของลำดับเบสบนดีเอ็นเอ
ระหว่างชีวิตคู่ 2 ชนิด



ภาคผนวก ค.

รายการบริการที่ทางแพลตฟอร์มเปิดให้ผู้พัฒนาสามารถเรียกใช้งานได้

Service Name Get a set of variant calls by Sample VCF ID

URL /sec/getVariantCallsBySampleId

HTTP method POST

Interface Type REST

getVariantCallsBySampleId: Request

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
header	1	M	Object		
serviceName	2	M		Service name	getVariantCallsBySampleId
systemCode	2	M		System code	ONC
channelId	2	M		Channel ID	ONCO_WEB
referenceNo	2	M		Reference Number	Format: ONCYYYMMDDHHmmssSSSS Example: ONC201911291435310040
transactionDate	2	M		Transaction date	Format: YYYYMMDDHHmmss Example: 201911291435531

getVariantCallsBySampleId: Request (ต่อ)

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
data	1	M	Object		
sampleVcflids	2	M	Array		[3]

getVariantCallsBySampleId: Response

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
header	1	M	Object		
serviceName	2	M	String	Service name	getVariantCallsBySampleId
systemCode	2	M	String	System code	ONC
channelId	2	M	String	Channel ID	ONCO_WEB
referenceNo	2	M	String	Reference Number	ONC201911291435310040
transactionDate	2	M	String	Transaction date	201911291435531
responseStatus	1	M	Object		
responseCode	2	M	String	Status code	ONC-SCC000
responseMessage	2	M	String	Status message	Service Success.
data	1	M	Array		

getVariantCallsBySampleId: Response (ต่อ)

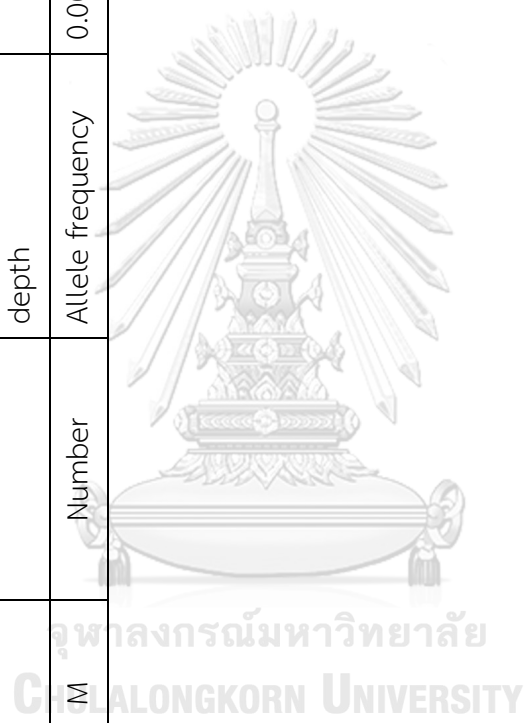
Parameter	Level	Mandatory/ Optional	Data type	Description	Example
chromosome	2	M	String	Chromosome	chr1
position	2	M	Number	Position	13838
refAllele	2	M	String	Reference allele	C
altAllele	2	M	String	Alternative allele	T
allele	2	M	String	Allele	T
consequence	2	M	String	Consequence	downstream_gene_variant
impact	2	M	String	Impact	MODIFIER
symbol	2	M	String	Gene name	LDLRAP1
gene	2	M	String	Gene ID	ENSG00000157978
featureType	2	M	String	Feature type	Transcript
feature	2	M	String	Feature	ENST00000484476
biotype	2	M	String	Biotype	processed_transcript
exon	2	O	String	Exon	
intron	2	O	String	Intron	
hgvsP	2	O	String	HGVSP	
hgvsC	2	O	String	HGVSC	

getVariantCallsBySampleId: Response (ต่อ)

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
cdnaPosition	2	O	String	CDNA position	
cdsPosition	2	O	String	CDS position	
proteinPosition	2	O	String	Protein position	
aminoAcids	2	O	String	Amino Acids	
codons	2	O	String	Codons	
existingVariation	2	O	String	Existing Variation	
distance	2	M	String	Distance	733
tumor	2	M	Object		
refAlleleDepth	3	O	Number	Reference allele depth	
altAlleleDepth	3	M	Number	Alternative allele depth	3.00000
alleleFrequency	3	M	Number	Allele frequency	0.02100
normal	2	M	Object		
refAlleleDepth	3	O	Number	Reference allele depth	

getVariantCallsBySampleId: Response (ต่อ)

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
altAlleleDepth	3	M	Number	Reference allele depth	0.00000
alleleFrequency	3	M	Number	Allele frequency	0.00000



Service Name Get a set of gene expression matrices by Sample RNA-Seq ID

URL /sec/getRnaSeqBySampleId

HTTP method POST

Interface Type REST

getRnaSeqBySampleId: Request

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
header	1	M	Object		
serviceName	2	M	String	Service name	getRnaSeqBySampleId
systemCode	2	M	String	System code	ONC
channelId	2	M	String	Channel ID	ONCO_WEB
referenceNo	2	M	String	Reference Number	Format: ONCYYMMDDHHmmsSSS Example: ONC201911291435310040
transactionDate	2	M	String	Transaction date	Format: YYMMDDHHmms Example: 20191129143531
data	1	M	Object		
sampleRnaSeqIds	2	M	Array		[1]

getRnaSeqBySampleId: Response

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
header	1	M	Object		
serviceName	2	M	String	Service name	getRnaSeqBySampleId
systemCode	2	M	String	System code	ONC
channelId	2	M	String	Channel ID	ONCO_WEB
referenceNo	2	M	String	Reference Number	ONC201912021902191810
transactionDate	2	M	String	Transaction date	20191202190219
responseStatus	1	M	Object		
responseCode	2	M	String	Status code	ONC-SCC000
responseMessage	2	M	String	Status message	Service Success.
data	1	M	Array		
geneId	2	M	String	Gene ID	ENSG00000223972
geneName	2	M	String	Gene name	DDX11L1
reference	2	M	String	Reference	1
start	2	M	Number	Start position	11869
end	1	M	Number	End position	14412
coverage	1	M	Number	Coverage	0.075624

getRnaSeqBySampleId: Response (ต่อ)

Parameter	Level	Mandatory/ Optional	Data type	Description	Example
strand	1	M	String	Strand	+
fpkm	1	M	Number	FPKM	0.011782
tpm	1	M	Number	TPM	0.023063



บรรณานุกรม

1. Strategy and Planning Division, O.o.P.S., *Public Health Statistic A.D.2017*, M.o.P. Health, Editor. 2018, Strategy and Planning Division: Nonthaburi, Thailand.
2. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation*. *Nucleic Acids Res*, 2001. **29**(1): p. 308-11.
3. Landrum, M.J., et al., *ClinVar: public archive of relationships among sequence variation and human phenotype*. *Nucleic Acids Res*, 2014. **42**(Database issue): p. D980-5.
4. Forbes, S.A., et al., *The Catalogue of Somatic Mutations in Cancer (COSMIC)*. *Curr Protoc Hum Genet*, 2008. **Chapter 10**: p. Unit 10.11.
5. Auton, A., et al., *A global reference for human genetic variation*. *Nature*, 2015. **526**(7571): p. 68-74.
6. Woollard, P.M., et al., *The application of next-generation sequencing technologies to drug discovery and development*. *Drug Discovery Today*, 2011. **16**(11): p. 512-519.
7. Ansorge, W.J., *Next-generation DNA sequencing techniques*. *New Biotechnology*, 2009. **25**(4): p. 195-203.
8. Son, M.S. and R.K. Taylor, *Preparing DNA libraries for multiplexed paired-end deep sequencing for Illumina GA sequencers*. *Curr Protoc Microbiol*, 2011. **Chapter 1**: p. Unit 1E.4.
9. Fickett, J.W. and C.S. Tung, *Assessment of protein coding measures*. *Nucleic Acids Res*, 1992. **20**(24): p. 6441-50.
10. Fickett, J.W., *The gene identification problem: An overview for developers*. *Computers & Chemistry*, 1996. **20**(1): p. 103-118.
11. Vaidyanathan, P.P. and B.-J. Yoon, *The role of signal-processing concepts in genomics and proteomics*. *Journal of the Franklin Institute*, 2004. **341**(1): p. 111-135.
12. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans*. *Nature*, 2016. **536**(7616): p. 285-91.

13. Frazer, K.A., et al., *Human genetic variation and its contribution to complex traits*. Nature Reviews Genetics, 2009. **10**: p. 241.
14. Altshuler, D.M., et al., *Integrating common and rare genetic variation in diverse human populations*. Nature, 2010. **467**(7311): p. 52-8.
15. Tattini, L., R. D'Aurizio, and A. Magi, *Detection of Genomic Structural Variants from Next-Generation Sequencing Data*. Front Bioeng Biotechnol, 2015. **3**: p. 92.
16. Huddleston, J., et al., *Discovery and genotyping of structural variation from long-read haploid genome sequence data*. Genome Res, 2017. **27**(5): p. 677-685.
17. Kruglyak, L. and D.A. Nickerson, *Variation is the spice of life*. Nature Genetics, 2001. **27**: p. 234.
18. Anita, S. *SNP phylogeny*. 1998 [cited 2018 Nov 22nd]; Available from: https://aschuerch.github.io/MolecularEpidemiology_AnalysisWGS/09-SNPphylo/index.html
19. Karakoc, E., et al., *Detection of structural variants and indels within exome data*. Nature methods, 2011. **9**(2): p. 176-178.
20. EMBI-EBI. *Types of genetic variation*. 2019 [cited 2019 Dec 1st]; Available from: <https://www.ebi.ac.uk/training/online/course/human-genetic-variation-introduction-2019/what-genetic-variation/types-genetic-variation>.
21. Stankiewicz, P. and J.R. Lupski, *Structural Variation in the Human Genome and its Role in Disease*. Annual Review of Medicine, 2010. **61**(1): p. 437-455.
22. Bodell, G. *Accidents affecting chromosome can cause disorders*. 2018 [cited 2018 Nov 2nd]; Available from: <http://bodell.mtchs.org/OnlineBio/BIOCD/text/chapter12/concept12.2.html>.
23. Geraldine, V. *What is a VCF and how should I interpret it*. 2012 [cited 2018 Nov 2nd]; Available from: <https://gatkforums.broadinstitute.org/gatk/discussion/1268/what-is-a-vcf-and-howshould-i-interpret-it>.
24. Danecek, P., et al., *The variant call format and VCFtools*. Bioinformatics, 2011. **27**(15): p. 2156-2158.
25. Pertea, M., et al., *Transcript-level expression analysis of RNA-seq experiments*

- with HISAT, StringTie and Ballgown. *Nature Protocols*, 2016. **11**: p. 1650.
26. Kimball, R. and J. Caserta, *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data*. 2004: John Wiley & Sons, Inc.
 27. Microsoft, *Microsoft Application Architecture Guide*. Second ed. 2009.
 28. Ramirez, A.O., *Three-Tier Architecture*. *Linux J.*, 2000. **2000**(75es): p. 7.
 29. Wikipedia. *Component-based software engineering*. 2009 [cited 2018 Nov 22nd]; Available from: https://en.wikipedia.org/wiki/Component-based_software_engineering#/media/File:Component-based-Software-Engineering-example2.png.
 30. Wikipedia. *Facade pattern*. 2016 [cited 2018 Nov 22nd]; Available from: https://en.wikipedia.org/wiki/Facade_pattern#/media/File:Example_of_Facade_design_pattern_in_UML.png.
 31. Martin, F. *Inversion of Control Containers and the Dependency Injection*. 2016 [cited 2018 Nov 22nd]; Available from: <https://martinfowler.com/articles/injection.html>.
 32. Wikipedia. *Model–view–controller*. 2010 [cited 2018 Nov 22nd]; Available from: <https://en.wikipedia.org/wiki/Model%E2%80%93view%E2%80%93controller#/media/File:MVC-Process.svg>.
 33. Nock, C., *Data Access Patterns: Database Interactions in Object-oriented Applications*. 2004: Addison-Wesley.
 34. Alur, D., et al., *Core J2EE Patterns (Core Design Series): Best Practices and Design Strategies*. 2003: Sun Microsystems, Inc. 650.
 35. Lee, B.K.B., et al., *GENIPAC: A Genomic Information Portal for Head and Neck Cancer Cell Systems*. *Journal of Dental Research*, 2018. **97**(8): p. 909-916.
 36. Chakravarty, D., et al., *OncoKB: A Precision Oncology Knowledge Base*. *JCO Precis Oncol*, 2017. **2017**.
 37. Fadlullah, M.Z., et al., *Genetically-defined novel oral squamous cell carcinoma cell lines for the development of molecular therapies*. *Oncotarget*, 2016. **7**(19): p. 27802-18.
 38. Zhao, M., et al., *Assembly and initial characterization of a panel of 85*

- genomically validated cell lines from diverse head and neck tumor sites.* Clin Cancer Res, 2011. **17**(23): p. 7248-64.
39. Prime, S.S., et al., *The behaviour of human oral squamous cell carcinoma in cell culture.* The Journal of Pathology, 1990. **160**(3): p. 259-269.
 40. Gao, J., et al., *Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal.* Science Signaling, 2013. **6**(269): p. pl1.
 41. Deng, M., et al., *Web-TCGA: an online platform for integrated analysis of molecular cancer data sets.* BMC Bioinformatics, 2016. **17**(1): p. 72.
 42. Samur, M.K., et al., *canEvolve: A Web Portal for Integrative Oncogenomics.* PLOS ONE, 2013. **8**(2): p. e56228.
 43. Margolin, A.A., et al., *ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context.* BMC Bioinformatics, 2006. **7 Suppl 1**(Suppl 1): p. S7.
 44. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
 45. Yan, Z., et al., *Integrative analysis of gene and miRNA expression profiles with transcription factor-miRNA feed-forward loops identifies regulators in human cancers.* Nucleic Acids Res, 2012. **40**(17): p. e135.
 46. Lee, H., et al., *The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations.* Genome Med, 2015. **7**: p. 112.
 47. Harris, P.A., et al., *Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.* J Biomed Inform, 2009. **42**(2): p. 377-81.
 48. Jensen, M.A., et al., *The NCI Genomic Data Commons as an engine for precision medicine.* Blood, 2017. **130**(4): p. 453-459.
 49. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th revision, 5th edition (2016).* 2015 [cited 2019 Aug 19th]; Available from: <https://apps.who.int/iris/handle/10665/246208>.
 50. World Health Organization. *International Classification of Diseases for Oncology (ICD-O)-3rd edition, 1st revision, 3rd edition.* 2013 [cited 2019 Aug 19th];

Available from: <https://apps.who.int/iris/handle/10665/96612>.

51. Hewett, M., et al., *PharmGKB: the Pharmacogenetics Knowledge Base*. *Nucleic Acids Res*, 2002. **30**(1): p. 163-5.
52. Fernandez, N.F., et al., *Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data*. *Scientific data*, 2017. **4**: p. 170151-170151.





จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

ชื่อ-สกุล	เนต้า เปอีโรเน
วัน เดือน ปี เกิด	1 ตุลาคม 2525
วุฒิการศึกษา	วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า สถาบันเทคโนโลยีพระ จอมเกล้าพระนครเหนือ
ที่อยู่ปัจจุบัน	เขตบึงกุ่ม จังหวัดกรุงเทพมหานคร



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY