

โครงการอบรมเชิงปฏิบัติการ
เรื่อง Genome Analysis Toolkit (GATK) Best Practices
ระหว่างวันที่ 8 - 9 ตุลาคม พ.ศ. 2562
จัดโดย ศูนย์วิทยาศาสตร์โอมิคส์และชีวสารสนเทศ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

1. ความเป็นมา/หลักการและเหตุผล

งานวิจัยทางด้านอณูชีววิทยา (molecular biology) จำเป็นต้องใช้ข้อมูล high-throughput sequencing (HTS) ในปริมาณ แต่เดิมยังไม่มีวิธีที่เป็นมาตรฐานในการวิเคราะห์ข้อมูลเหล่านี้ มีวิธีการวิเคราะห์ข้อมูลที่แตกต่างกันจำนวนมาก และทำให้ได้ผลลัพธ์แตกต่างกัน ปัจจุบันเริ่มมีการกำหนดวิธีการวิเคราะห์ข้อมูลที่เป็นมาตรฐาน (best practices) เพื่อให้ผลงานวิจัยเป็นที่ยอมรับและมีความน่าเชื่อถือ Broad Institute ซึ่งเป็นหน่วยงานที่เกิดจากความร่วมมือระหว่างมหาวิทยาลัย MIT และ Harvard และ Harvard-affiliated Hospitals ได้ออกมาตรฐานสำหรับการวิเคราะห์ข้อมูล high-throughput sequencing (HTS) ที่มีชื่อว่า Genome Analysis Toolkit หรือที่เรียกย่อ ๆ ว่า GATK Best Practices การวิเคราะห์ข้อมูลตามมาตรฐานของ GATK นั้น ประกอบด้วยขั้นตอนย่อยเป็นจำนวนมาก ดังนั้น GATK จึงออกแบบภาษาคอมพิวเตอร์ใหม่ชื่อว่า Workflow Description Language (WDL) สำหรับอธิบายขั้นตอนการวิเคราะห์ข้อมูลทั้งหมด อินพุต เอาพุต และซอฟต์แวร์ที่ใช้ประมวลผลในแต่ละขั้นตอน นอกจากนี้ GATK ยังถูกออกแบบให้ทำงานได้ทั้งบนคอมพิวเตอร์ส่วนบุคคล และทำงานได้บน Cloud Computing Services ด้วย

2. วัตถุประสงค์และขอบเขตของโครงการ

ผู้เข้าอบรมจะเรียนรู้ workflow ทั้งหมดของ GATK ดังนี้

- Data pre-processing
- Germline SNPs + Indel
- Somatic SNVs + Indels
- Germline CNVs
- Somatic CNVs

ในการอบรมจะสอนทฤษฎีของทุก workflow แต่เน้นการปฏิบัติเพียง workflow เดียวเท่านั้น เนื่องจากเวลามีจำกัดและทักษะที่ใช้ในการรัน workflow ทุกตัวเหมือนกัน ผู้เข้าอบรมจะได้ทดลองรัน workflow บนคอมพิวเตอร์ส่วนบุคคล บน Docker และบน Cloud Computing Services

3. หัวข้อการอบรม/สัมมนา

วันที่ 1:

08.30 - 09.00 น.	ลงทะเบียน
09.00 - 09.05 น.	กล่าวต้อนรับและพิธีเปิด โดย ผู้อำนวยการศูนย์ฯ
09.05 - 10.00 น.	Introduction to GATK Best Practices Workflow Description Language (WDL)
10.00 - 10.15 น.	พักรับประทานอาหารว่าง
10.15 - 12.00 น.	Workflow: Data pre-processing
12.00 - 13.00 น.	รับประทานอาหารกลางวัน
13.00 - 14.30 น.	Workflow: Germline SNPs + Indel
14.30 - 14.45 น.	พักรับประทานอาหารว่าง
14.45 - 16.30 น.	Workflow: Somatic SNVs + Indels

วันที่ 2:

8.30-9.00 น.	ลงทะเบียน
9.00-10.00 น.	Workflow: Germline CNVs
10.00-10.15 น.	พักรับประทานอาหารว่าง
10.15-12.00 น.	Workflow: Somatic CNVs
12.00-13.00 น.	รับประทานอาหารกลางวัน
13.00-14.30 น.	Docker
14.30-14.45 น.	พักรับประทานอาหารว่าง
14.45-16.30 น.	Cloud Computing Services

หมายเหตุ: หัวข้อและเวลาอาจจะมีการปรับเปลี่ยนตามความเหมาะสม

4. วิธีการจัดอบรมและสัมมนา

เป็นการจัดอบรมเชิงปฏิบัติการ ศูนย์ฯ จัดเตรียมคอมพิวเตอร์เพื่อการอบรมให้แล้ว โดยแบ่งการอบรมเป็น ภาคบรรยายและภาคปฏิบัติ โดยในภาคปฏิบัตินั้น ผู้เข้าอบรมจะได้ลงมือปฏิบัติการรัน GATK workflows โดยใช้ข้อมูล high-throughput sequencing (HTS) ที่มีขนาดและจำนวน sample ไม่มาก เพื่อให้เหมาะสมกับเวลาฝึกอบรม

5. ระยะเวลาในการดำเนินงาน และการจัดฝึกอบรม/สัมมนา

- 5.1. ระยะเวลาในการเตรียมงาน ระหว่างวันที่ 1 กรกฎาคม 2561 ถึง 9 ตุลาคม 2562
- 5.2. ระยะเวลาในการจัดอบรม ระหว่างวันที่ 8 ถึง 9 ตุลาคม 2562

6. สถานที่

ห้อง Computer 1 ชั้น 5 อาคารมหาชิรณทิต คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

7. คุณสมบัติผู้เข้าร่วมอบรม/สัมมนา

7.1. อาจารย์ นักวิจัย นักวิชาการ นิสิต/นักศึกษา หรือผู้สนใจ ในสาขาวิทยาศาสตร์ที่ต้องวิเคราะห์ข้อมูล high-throughput sequencing (HTS) data เพื่อหา variants เช่น SNPs หรือ CNV ระหว่างประชากรกลุ่มทดลองและกลุ่มควบคุม

7.2 ผู้เข้าอบรมควรมีความรู้พื้นฐานเกี่ยวกับคอมพิวเตอร์ และการใช้งานระบบปฏิบัติการ Linux

8. จำนวนผู้เข้าร่วมอบรม และ ค่าลงทะเบียน

จำนวนผู้เข้าร่วมการอบรม 30 คน โดยมีระยะเวลาการลงทะเบียน และจัดเก็บค่าลงทะเบียน ดังนี้

ระยะเวลาการลงทะเบียน	อัตราการลงทะเบียน
วันที่ 1 สิงหาคม ถึง 22 กันยายน 2562	3,500 บาท (สำหรับนิสิต/นักศึกษา)
วันที่ 1 สิงหาคม ถึง 22 กันยายน 2562	4,500 บาท
วันที่ 23 ถึง 29 กันยายน 2562	4,500 บาท (สำหรับนิสิต/นักศึกษา)
วันที่ 23 ถึง 29 กันยายน 2562	5,500 บาท

* สิ้นสุดการรับลงทะเบียนในวันที่ 29 กันยายน 2562

** บุคลากรของรัฐจากส่วนงานและส่วนราชการต่างๆ เข้าร่วมอบรมได้โดยไม่ถือเป็นวันลาและมีสิทธิ์เบิกค่าลงทะเบียนตลอดจนมีสิทธิ์เบิกค่าใช้จ่ายต่างๆ ได้ตามระเบียบของราชการ

*** ศูนย์ฯ จัดเตรียมคอมพิวเตอร์สำหรับการ workshop ให้กับผู้เข้าร่วมอบรมทุกท่านแล้ว

9. ประโยชน์ที่คาดว่าจะได้รับ

9.1. ผู้เข้าร่วมอบรมสามารถหาวิเคราะห์หา variants เช่น SNPs และ CNVs จาก high-throughput sequencing (HTS) data ตาม protocols ของ GATK Best Practice ได้

9.2. ผู้เข้าร่วมอบรมมีความเข้าใจในขั้นตอนวิธีการวิเคราะห์ข้อมูล และวิธีทางสถิติที่ใช้ในแต่ละขั้นตอนของ GATK Best Practice

10. หน่วยงานที่รับผิดชอบ

ศูนย์วิทยาศาสตร์โอมิกส์และชีวสารสนเทศ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

11. รายชื่อวิทยากร

รองศาสตราจารย์ ดร.ชัชวีย์ อภรณ์เทวัญ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์