



Programa del Taller Costa Rica GATK Workshop

Características generales:

Es un taller teórico-práctico de tipo presencial que abarca un total de 4 días dirigido a investigadores universitarios, nacionales, docentes universitarios, estudiantes de pregrado y posgrado. Incluye charlas magistrales con expertos internacionales sobre los tópicos de procesamiento de datos de secuenciación, identificación de variantes germinales, somáticas; el desarrollo y optimización de “pipelines” bioinformáticos que implementen los conceptos de manejo de datos de gran escala (big data genómica), buenas prácticas de análisis (“best practices of GATK”) y el desarrollo de código en el lenguaje WDL.

Requisitos del taller:

- Profesionales en áreas afines a la bioinformática y al análisis clínico humano.
- Estudiantes de pregrado intermedios o avanzados y de grado de las áreas de: Computación, Biología, Biotecnología, Microbiología y Química Clínica, Medicina y áreas afines.
- Estudiantes de posgrado de las áreas de: Computación, Bioinformática, Biología y Genética Humana, Biotecnología, Microbiología y Química Clínica, Medicina y áreas afines.
- Traer su propia laptop para la realización de las sesiones prácticas.
- Descargar el material que se brindará previo al taller.
- Asistencia obligatoria a todas las sesiones teóricas y prácticas del taller.

Descripción:

El software Genome Analysis Toolkit del GATK es desarrollado por la plataforma de Ciencias de Datos del Instituto Broad, que ofrece una amplia variedad de herramientas para enfocarse en el descubrimiento de las variantes genómicas. El poder de procesamiento y las características de desempeño computacionales permiten que el software sea capaz de adaptarse a cualquier tamaño de proyecto, por lo que lo hace ideal para el ambiente investigativo universitario. Además, en el taller se explicarán otras herramientas como MuTect, Cromwell + WDL, Docker y Big Query.

En el taller se desarrollarán dos tipos de dinámicas: la primera que abarca un componente teórico donde se realiza una introducción y explicación a los conceptos, escritura a “pipelines” y flujo de procesos y el componente práctico, donde se realizarán ejercicios didácticos enfocados en los temas expuestos por los expertos internacionales de cada área. Esto permite que los participantes tengan las bases necesarias para realizar las sesiones prácticas de cada tema expuesto. Asimismo, durante las sesiones prácticas los expertos brindarán apoyo logístico y aclaración de dudas de diversa índole en el flujo de trabajo relacionadas con cada temática.

El taller proporcionará los ejercicios y materiales para llevar a cabo el taller y se dará un certificado de participación.

Expositores internacionales:

Laura Gauthier
Rori Cremer
Mark Walker

Objetivo:

Capacitar a investigadores universitarios, nacionales, docentes universitarios, estudiantes de pregrado y posgrado en la implementación del software GATK y en el desarrollo de flujos de procesos mediante el lenguaje WDL utilizado en GATK para el análisis de los datos de secuenciación de NGS e identificación de variantes genómicas.

Organizadores:

Ricardo Chinchilla Monge MSc. Teléfono: 2511-3483 ricardo.chinchilla_m@ucr.ac.cr

Para consultas Teléfono: 2511-3322 recepcion.cicica@ucr.ac.cr

Contenidos:

FECHA	HORA Y DURACIÓN	DESCRIPCIÓN DE LA ACTIVIDAD (Indicar título de charlas o conferencias si corresponde)
6 de febrero de 2024	8:00 a.m. - 12:00 p.m.	-Introducción al taller GATK. -Introducción a los datos de Secuenciación. -Introducción a los datos de preprocesamiento. -Introducción al pipelining. -Terra por primera vez (demostración guiada).
06 de febrero de 2024	1:00 p.m. - 5:00 p.m.	-Descubrimiento de variantes germinales. -Haplotype Caller. -Ejercicios con HaplotypeCaller. -Joint Calling.
07 de febrero de 2024	8:00 a. m. - 12:00 p.m.	-gnomAD. -Casos de estudio: Análisis de variantes de la tetralogía de Fallot. -Introducción al filtrado de variantes. -Refinamiento y Evaluación: Refinamiento del genotipo y evaluación del llamado del set.
07 de febrero de 2024	1:00 p.m. - 5:00 p.m.	-Ejercicios con el uso de VETS. -Introducción al DRAGEN-GATK. -Lecturas largas: procesamiento de lecturas largas de secuenciación.
08 de febrero de 2024	8:00 a.m. -	-Ejercicios sobre las herramientas de refinamiento y evaluación.

	12:00 p.m.	-Introducción al descubrimiento de variantes somáticas. -Variantes de nucleótido simple (SNV) somáticas e Indels.
08 de febrero de 2024	1:00 p.m. - 5:00 p.m.	-Introducción a las Alteraciones en el número de copias (CNAs) somáticas. -Ejercicio sobre CNAs somático. -Variaciones en el Número de Copias Germinal.
09 de febrero de 2024	8:00 a.m. - 12:00 p.m.	-Datos de Mitocondria. -Variantes Estructurales: Introducción al llamado de variantes estructurales (SVs) con GATK-SV. -Ejercicios con el pipeline gCNV.
09 de febrero de 2024	1:00 p.m. - 5:00 p.m.	-Single-Cell RNA-seq Analysis: Introducción al análisis de single-cell RNA-seq. -Ejercicio sobre Single-cell RNA-seq. -Sesión de cierre para realizar la evaluación, aclaración de dudas y llenado de encuesta.

Recesos:

Cada día se contarán con los siguientes recesos:

En la mañana: 10:15 a.m. - 10:45 a.m.

Almuerzo: 12:00 p.m. - 1:00 p.m.

En la tarde: 2:30 p.m. - 3:00 p.m.

Bibliografía:

1) McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297-1303.

2) BigQuery—Analytics Data Warehouse | Google Cloud Platform. In: Google Cloud Platform [Internet]. [cited 19 Mar 2018]. Available: <https://cloud.google.com/bigquery/>