**GATK 3-day workshop summary**

This workshop will focus on the core steps involved in calling variants with the Broad’s Genome Analysis Toolkit, using the “Best Practices” developed by the GATK team. You will learn why each step is essential to the variant discovery process, what are the operations performed on the data at each step, and how to use the GATK tools to get the most accurate and reliable results out of your dataset. In the course of this workshop, we highlight key functionalities such as the GVCF workflow for joint variant discovery in cohorts, RNAseq-specific processing, and somatic variant discovery using MuTect2. We also preview capabilities of the upcoming GATK version 4, including a new workflow for CNV discovery, and we demonstrate the use of pipelining tools to assemble and execute GATK workflows.

**Program outline**

The workshop is composed of one day of lectures (including many opportunities for Q&A) and two optional days of hands-on training, structured as follows:

Day 1: lectures: **Rationale, theory and application of the GATK Best Practices for Variant Discovery in high-throughput sequencing data.**

Day 2 AM: opt hands-on: **Germline variant discovery (SNPs + Indels)**

Day 2 PM: opt hands-on: **Germline variant filtering (SNPs + Indels)**

Day 3 AM: opt hands-on: **Somatic variant discovery (SNPs + Indels + CNV)**

Day 3 PM: opt hands-on: **Pipelining on the cloud with WDL**

In the three optional hands-on sessions focused on analysis, we walk attendees through exercises that teach them how to manipulate the standard data formats involved in variant discovery and how to apply GATK tools appropriately to common use cases and data types. In the course of these exercises, we demonstrate useful tips and tricks for interacting with GATK and Picard tools, dealing with problems, and using third-party tools such as Samtools, IGV, RStudio and RTG Tools.

In the optional hands-on session on pipelining, we walk attendees through exercises that teach them to write workflow scripts using WDL, the Broad's new Workflow Description Language, and to execute these workflows locally as well as through publicly available cloud-based services.

**Target audience**

The lecture-based component of the workshop is aimed at a mixed audience of people who are new to the topic of variant discovery or to GATK, seeking an introductory course into the tools, or who are already GATK users seeking to improve their understanding of and proficiency with the tools. Attendees should already be familiar with the basic terms and concepts of genetics and genomics.

The hands-on component is aimed at novice to intermediate users who are seeking detailed guidance with GATK and related tools. Basic familiarity with the command line environment is required. Attendees will be expected to bring their own laptops with software preinstalled (detailed instructions will be posted two weeks before the course). Supported systems are Mac and Unix/Linux systems. MS Windows is NOT supported.

**Detailed agenda of the Best Practices lectures**

Introduction to variant discovery analysis and GATK Best Practices

*Coffee Break*

Marking Duplicates Indel Realignment Base Recalibration

*Lunch Break*

Variant Calling and Joint Genotyping Filtering variants with VQSR Genotype Refinement Workflow

*Coffee Break*

Callset Evaluation Somatic variant discovery with MuTect2

Preview of CNV discovery with GATK4

**Please note that this workshop is focused on human data analysis. The majority of the materials presented does apply equally to non-human data, and we will address some questions regarding adaptations that are needed for analysis of non-human data, but we will not go into much detail on those points.**