**Introducing GATK4**

The Genome Analysis Toolkit or GATK, developed at the Broad Institute, is currently one of the most widely used software solutions for germline short variant discovery and genotyping in whole genome and exome data. GATK4 is the next generation of GATK; it runs faster and covers more ground, improving somatic SNVs and Indel discovery, and adds Copy Number Variation analysis to its variant discovery portfolio.

Given the breathtaking increase in the amount of genomic data produced in recent years, GATK development has come to tackle multiple challenges: a) reflect the cutting-edge of science in terms of depth as well as breadth of analyses offered, b) scale to petabytes of data and c) provide a way to navigate the tradeoff between turn-around-time and cost of analyses.

GATK4 addresses all of these challenges through a complete redesign and reimplementation of the original GATK software.

---

**A streamlined engine and Spark support make GATK4 much faster**

Designed with cloud infrastructure in mind (though it still runs on local infrastructure), GATK4 is implemented with built-in support for Apache Spark, makes key operations hundreds of times faster than previous generations of GATK and comes with end-to-end pipeline implementations that run out of the box on select cloud platforms. For the established tools and algorithms composing the GATK Best Practices pipelines for short variant discovery, GATK4 is a substantial upgrade that streamlines the tools’ operation, providing functionally equivalent results in a much faster timeframe.

---

**Copy Number Variation**

**Best Practices for Somatic CNV discovery**

The GATK Best Practices workflow for calling somatic CNVs in exomes consists of 4 phases:

1. Collection of proportional coverage profiles
2. Normalization to eliminate technical noise (requires an appropriate Panel of Normals)
3. Segmentation of contiguous copy ratio targets
4. Calling segmented amplified, deleted or neutral segments

Plots can be generated through an optional step. Generation of the Panel of Normals not shown.

---

**Learn more and ask questions at http://www.broadinstitute.org/gatk**

---

**SNVs and Indels**

**Best Practices for Somatic SNVs and Indels**

The GATK Best Practices workflow for identifying SNVs and Indels in whole genomes, exomes and targeted gene panels involves 3 phases: alignment and data cleanup; estimation of cross-sample contamination; and variant calling using MuTect2.

---

**Graph assembly enables discovery of large Indel events**

The graph assembly step performed from HaplotypeCaller is a key factor of MuTest's ability to call indels with great sensitivity. This step allows the caller to bypass technical limitations of the mapping algorithm (BWA MEM) such as discarding inserted sequence as soft-clipped bases.

---

**Somatic Mutation Discovery with GATK4**

Geraldine A. Van der Auwera for the Data Sciences & Data Engineering Group @ Broad Institute