Somatic Analysis Services with Best Practice Workflows in a Cloud-Based Platform

JUNKO TSUJI, ALYSSA MACBETH, BRIAN GRANGER, MICAH RICKLES-YOUNG, TERA BOWERS, CARRIE CIBULSKIS, NIALL LENNON

Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge MA 02141

Challenges in Somatic Analysis

 Workflow with best practice methods
Detecting somatic variants is not a trivial task in terms of the technical aspects (e.g. filtering germline events and removing a variety of noises in tumor samples). It is also laborious to select a better performing tool from many software packages with similar functions for building a quality workflow.

Consistency and traceability of analysis
When a study (e.g. clinical trial) continues over extended periods of time, it becomes more difficult to keep consistent analysis methods and maintain these as functional workflows.

Computational resources
Large-scale cohort and integrative analyses require extensive computational resources to process these sequencing data.

Workflow Execution Infrastructure

We run best practice methods written in workflow description language (WDL) in a cloud system developed at the Broad, Terra (formerly FireCloud) for both external and internal scientific studies.

Analysis Workflows

We maintain 62 well-tested workflows with best practice methods. Since May 2017, our team has performed 18,003 analyses across a variety of experimental conditions shown below.

Somatic Analysis

Combined Caller Pipeline

### Quality Control

**SNP/INDEL**
-ualenomic analysis is performed using both WGS and WES with nonclassical CD14+ CD16++ cells.

**RNA**
- RNA sequencing is performed using the Illumina RNA-Seq protocol.

**DNA**
- DNA sequencing is performed using the Illumina HiSeq protocol.

**Bulk RNA**
- RNA sequencing is performed using the Illumina RNA-Seq protocol.

### Cell-free DNA (Liquid Biopsy) Analysis

**ULP-WGS**
- DNA sequencing is performed using the Illumina NextSeq 500 protocol.

**Targeted Panel Sequencing with UMI Adapters**
- RNA sequencing is performed using the Illumina NextSeq 500 protocol.

**ULP-WGS CNV Cohort Analysis**
- RNA sequencing is performed using the Illumina NextSeq 500 protocol.

**Bulk RNA Quantification and Differential Analysis**
- RNA sequencing is performed using the Illumina RNA-Seq protocol.

Acknowledgments

We thank Chip Stewart and Gaddy Getz from the Cancer Program, and Eric Banks, Lee Lichtenstein, David Benjamin, Samuel Lee, Geraldine Van der Auwera, Yossi Farjoun, Mark Fleharty, Madeleine Duran, Sarah Walker, and Christopher Kachulis from the Data Science Platform for their valuable inputs and their continued support.