Variant Consensus Reporter: Increasing the Confidence of Variants in Whole Exome Sequencing via a Consensus Approach

Erich A. Peterson, Shweta S. Chavan, Michael A. Bauer, and Donald J. Johann

Introduction

Several tools now exist for the calling of variants from data produced by high throughput sequencing technologies. These tools, such as GATK and others, in conjunction with functional annotation tools, offer the promise of narrowing down the search for variants, which may directly affect a patient's cancer biology. However, more complex variants typically rely heavily on statistical models, and various approaches use differing algorithmic techniques. Thus, determining which tool's called variants may have the necessary confidence warranting further experimental validation can be tricky. Here we present an approach, called the Variant Consensus Reporter, for culling variants from several variant calling tools using a consensus method. The result is a report that can give the investigator more information and confidence concerning identified variants.

Methods

DNA from the Multiple Myeloma (MM) cell lines, RPMI-8226 and U266 BZ, were extracted and prepared for whole exome sequencing (WES) on an Illumina HiSeq 2500. A standardized computational pipeline for DNA processing has been designed and developed. In brief, it provides uniform methods and relies on BWA and STAMPY for alignment and BAM file creation. Variant calling is performed utilizing GATK as well as other methods, which result in a Variant Call File (VCF). Those VCFs were additionally functionally annotated using ANNOVAR. Finally, the two VCF files were used as input for the Variant Consensus Reporter, which analyzes the variant calls produced by each tool to create a consensus report.

Results

As a test-case, the Variant Consensus Reporter was designed and developed to report consensus variants, which contain the following functional annotations: i) Point mutations (both missense and nonsense); ii) Frame-shift mutations. These particular annotated variants were chosen because of their potential to be clinically actionable.