NA12878 BRCA1 and BRCA2 Variants
Crystal Genetics, Inc.

www.crystalgenetics.com

Summary

This is a comparison of the variants (SNVs, insertions, deletions and block substitutions) of NA12878 for BRCA1 and BRCA2 genes between the calls made by Crystal Genetics versus NIST's GiaB V2.19 (herein referred to as the Benchmark). The variant data used for the Benchmark can be found at precisionFDA's website. The Benchmark provides one of the most popular public whole-genome variation calls for NA12878.

Results

The results are summarized in the associated Excel files:

- The variants that are not highlighted represent identical calls between Crystal Genetics and the Benchmark.
- The variants that are highlighted in green represent consistent calls, based on Crystal's visual analysis of the reads supporting the variants.
- The variants that are highlighted in red represent inconsistent calls, based on Crystal's visual analysis of the reads supporting the variants.
- The variants that are highlighted in yellow represent equivalent and consistent calls, based on Crystal's visual analysis of the reads supporting the variants.
- The variants that are highlighted in grey represent hard-to-assess calls, based on Crystal's visual analysis of the reads supporting the variants.

Please note that the consistent/inconsistent designation does not mean to reflect a correct/incorrect state for the calls. The consistent and inconsistent designation is meant to reflect the state of the calls as compared to Crystal's visual checking of the reads supporting the corresponding variants. Please also note that the visual analysis is not meant to be a replacement for a confirmation assay, which would be the ultimate way of validating the quality.

Raw Data

Crystal's call sets are based on the raw whole-genome sequencing (WGS) data, provided by Illumina in **http://www.illumina.com/platinumgenomes**. This data comprises a 50x human genome sequenced with Illumina's HiSeq2000 system.

Variant Calls

Both reported variants calls have used Build 37 of Human Reference Genome for reporting. Crystal Genetics' variant calls are obtained using its own proprietary algorithms --from raw reads (FASTQ) to variants (VCF)-- and without utilizing any consensus data or bioinformatics/other databases (e.g., dbSNP). It uses only the read data from one sequencing platform (Illumina) from that individual.

Reports

In order to provide a comprehensive view for the variant calls, a union of the reported positions in Crystal and the Benchmark was made. Subsequently, for each method, the reported loci were loaded from the VCF files into the corresponding positions.

For each of Crystal and the Benchmark, a column (named varcall) was added to show if the corresponding variant was found in the VCF file or not. If no variant was found in the VCF file, the other columns were filled with a dash. This was in attempt to line-up the variants of Crystal and the Benchmark. Please note that a dash should not be interpreted as no data, as it may, for instance, contain ref/ref calls or variants which are filtered out in one of the call sets.

Low-quality Calls

- Crystal's calls that are reported are all considered high-confidence calls.
- The Benchmark may have categories of low-confidence calls –If the filter designation is not equal to PASS. In our analysis, the low-confidence calls were marked as no-calls, prior to comparison.

Questions

- If you have any questions, please feel free to reach us via the Contact Tab at our website: www.crystalgenetics.com