Documentation for Clinical Region of Interest and Exome Tests
Updated July 2016

The Claritas Clinical Exome uses an innovative dual-capture, dual sequencing platform method that is unique in the industry. This "Orthogonal Approach" simultaneously confirms ~95% of all exome variants, providing the highest confidence clinical results. A description of the process can be found in the Open Access paper *Orthogonal NGS for High Throughput Clinical Diagnostics* that can be downloaded at http://www.nature.com/articles/srep24650.

The accompanying files are produced as part of the clinical test and are being returned to the ordering physician for research purposes. Please contact clientservices@claritasgenomics.com with any questions.

Note that for Region of Interest tests, we return whole exome data with no guarantees, warrantees or liability outside the region of interest ordered. Whole exome data is for research purposes only. Researchers are responsible for obtaining research consent from patients.

Platforms Used:
DNA from probands (affected individuals) is sequenced using both Illumina NextSeq and ThermoFisher Ion Proton. Parents are sequenced using Illumina NextSeq only.

File Descriptions:

1. Accession number key:
   <accession_number>_accession_key_clinical_trio.txt. This file provides the corresponding patient name for each Claritas Accession number generated for this test

2. Illumina VCF:
   <accession_number>.realigned.recalibrated.genotyped.filtered.phased.vcf. This is the variant call file (VCF) for the Illumina platform. This is ~27MB in size and the index ~7MB.

3. Illumina BAM: <accession_number>.realigned.recalibrated.bam & .bai: BAM file and the corresponding index from Illumina platform. The BAM is ~14 GB and the index is ~7MB in size.

4. Illumina QC metrics: These are generated by the Picard Tool. Details of the metrics can be accessed at the below links.
   <accession_number>.realigned.recalibrated.alignment_summary_metrics.txt

http://broadinstitute.github.io/picard/command-line-overview.html#CollectAlignmentSummaryMetrics

<accession_number>.realigned.recalibrated.pertarget_hsmetrics.txt
http://broadinstitute.github.io/picard/command-line-overview.html#CollectHsMetrics

5. Proton VCF: TSVC_variants.vcf: VCF file from Proton platform. This is ~27MB in size and will be present only for probands.

6. Proton BAM: IonXpress_<barcode>_rawlib_processed.bam: BAM file and the corresponding index from Proton platform. The BAM is ~7GB and the index is ~6MB and will be present only for probands.

7. Combined VCF: <sample_name>.combined.sorted.nounused.vcf. This is combined VCF from the Illumina and Proton platforms. This is ~30MB in size and will be present only for probands.

Details of Bioinformatics pipelines used to produce these files

Illumina BAM and VCF:

Processed using the GATK Best Practices pipeline which includes:

- Alignment using BWA mem
- GATK Indel Realignment
- GATK Base Quality Score Recalibration
- GATK Haplotype Caller

Two additional steps:

- Variant Filtration: Variants annotated as low quality based on select quality metrics in the VCF file.
- Read Backed Phasing: Variants phased to combine adjacent SNPs in cis into Multi Nucleotide Polymorphisms (MNPs)

Proton BAM and VCF:

Processed using Torrent Suite 4.4. Results are filtered using a method based on allele frequencies developed at Claritas.

Combined VCF:

VCF file containing consensus calls made by combining Proton and Illumina VCFs using Combinator software developed at Claritas. The FILTER column in the VCF file indicates the individual calls that were combined and the confidence in the call.

Data Download via URL

We will send you URLs for the zipped data and the corresponding MD5 files. These links are valid for a period of time specified by email. Each zip contains the files for a single sample: BAM, VCF, QC metrics, etc.

If downloading via a web browser on a Mac, **we recommend using Safari** rather than

Chrome, as Chrome is more sensitive to https sites and often renames the file extensions upon download.

One can use the following example to automate the download via the command line:

wget --no-check-certificate
"http://path-to-claritas.s3.amazonaws.com/file.zip.md5?AWSAccessKeyId=A
KIAJPDAMZR5&Expires=1444073489&Signature=WcXRwR0y6ayb" -O output-file-name

To check for md5 concordance, one can use a number of methods including the following example, which assumes the existence of files “file.zip.md5” and “file.zip” in the same directory:

cd /PATH/TO/ZIP md5sum file.zip > file.zip.my-md5sum.md5; cut -f 1
file.zip.my-md5sum.md5 -d " " | cmp -n 32 file.zip.md5