Harmonization of GREGoR whole genome sequencing data

Author: Marsha Wheeler, Data Coordinating Center
Last updated: May 31, 2024

Methods

Harmonization of short-read, whole genome sequencing (srWGS) data was performed by the GREGoR Data Coordinating Center using the Whole Genome Germline Single Sample WARP pipeline, in DRAGEN-GATK mode (v3.1.6). This pipeline includes alignment to the GRCh38 genome reference using the DRAGEN DRAGMAP aligner and duplicate marking with Picard v2.26.10. The WARP pipeline performs single sample variant calling with GATK HaplotypeCaller. For GREGoR data this subtask was performed in DRAGEN mode using the Dragstr model and hard filtering parameters.

Joint variant calling for GREGoR srWGS was performed using the Genomic Variant Store (GVS). This pipeline was developed to perform joint calling at scale and is based on a schema designed for querying and rendering variants in which the variants are stored in GVS and rendered to an analyzable variant file format.

Functional annotation of the GREGoR joint callset was performed with Variant Effect Predictor v112.